Interpretable Machine Learning for Perturbation Biology

Bo Yuan[1,2,3*#], Ciyue Shen[1,2,3*#], Augustin Luna[1,2,3], Anil Korkut[4], Debora S. Marks[3,5], John Ingraham[6], Chris Sander[1,2,3#]

1 Dept. of Cell Biology, Harvard Medical School

2 cBio Center, Dept. of Data Sciences, Dana-Farber Cancer Institute

3 Broad Institute of MIT and Harvard

4 Dept. of Bioinformatics & Computational Biology, The University of Texas MD Anderson Cancer Center

5 Dept. of Systems Biology, Harvard Medical School

6 MIT Computer Science & Artificial Intelligence Lab

* Joint first authors

Correspondence to mathcellbox@gmail.com reaches the principal authors

Abstract

Systematic perturbation of cells followed by comprehensive measurements of molecular and phenotypic responses provides an informative data resource for constructing computational models of cell biology. Models that generalize well beyond training data can be used to identify combinatorial perturbations of potential therapeutic interest. Major challenges for machine learning on large biological datasets are to find global optima in an enormously complex multi-dimensional solution space and to mechanistically interpret the solutions. To address these challenges, we introduce a hybrid approach that combines explicit mathematical models of dynamic cell biological processes with a machine learning framework, implemented in Tensorflow. We tested the modelling framework on a perturbation-response dataset for a melanoma cell line after drug treatments. The models can be efficiently trained to accurately describe cellular behavior, as tested by cross-validation. Even though completely data-driven and independent of prior knowledge, the resulting *de novo* network models recapitulate some known interactions. The main predictive application is the identification of combinatorial candidates for cancer therapy. The approach is readily applicable to a wide range of kinetic models of cell biology.

Introduction

The emergence of resistance to single anticancer agents has highlighted the importance of developing combinations of agents as a more robust therapeutic approach to cancer treatment[1–4]. However, experimental screening of all possible pairwise or higher order combinations of currently available agents is practically unrealistic. The space of potential new therapeutic targets is even larger and more challenging to explore experimentally. To efficiently narrow down the search space and nominate promising sets of experimentally testable candidates, computational models have been used to predict cellular responses based on sets of perturbation experiments[5–7], but these have been limited in scope. The ability to model cell biology at a larger scale and to infer causal mechanisms to generalize to unobserved perturbations is critical in facilitating the search for combinatorial, potentially therapeutic candidates.

In order to understand cell behavior, various experimental approaches have been used to profile cellular responses under different perturbations. Biochemical and cell biological experiments testing relationships of particular protein-protein pairs have for many years been successfully used to identify signaling cascades[8–10], but one-by-one experiments are laborious and the resulting models, while insightfully descriptive, typically are limited in quantitatively predicting both detailed molecular and system-level cell responses. Phenotypic screening collects high-throughput information on whole-cell responses with univariate readouts such as cell viability or growth rate[11–15]. In order to resolve intracellular interactions and provide mechanistic insights, systematic methods have been developed to profile post-perturbational molecular responses, e.g. changes in transcript[16–18] and protein[19,20] levels. These rich datasets challenge computational methods to potentially discover mechanisms and accurately model cell responses.

Various computational methods have been developed to predict cellular responses[21]. Static models use, e.g., differential expression analysis[22], regression or co-expression models [23–25], maximum entropy networks[26,27], or mutual information[28], to correlate cellular responses with perturbations and/or molecular measurements[29,30]. On the other hand, dynamic models, such as Boolean network models[31], fuzzy logic models[32], dynamic Bayesian networks[33] and ordinary differential equation (ODE) network models[34,35], can provide mechanistic insight in terms of propagation of cellular signals to phenotypic response over time, but typically require prior knowledge of interaction parameters and thus currently only work for small systems[34,36]. New algorithms to parametrize large-scale mechanistic models, while computationally efficient, have required prior knowledge of a set of relevant interactions[37]. Insufficient prior knowledge is one of the major constraints for modeling large systems, e.g., in that prior information is not available for all components or is aggregated from disparate experimental sources and thus lacks uniform context. A more rigorous approach is to use uniform datasets generated in systematic experiments in one experimental context and then perform de novo structure inference of an interaction network valid for that context. Given such data for large systems, the computational challenge is to search for optimal interaction parameter sets in a complex multi-dimensional solution space. Previous dynamic optimization approaches such as Monte Carlo（MC）methods and belief propagation (BP) algorithms, have been used to construct data-driven network models[19,36,38–40], but these may not efficiently scale to larger systems (e.g., MC) or may require excessive approximations for the chosen mathematical model to facilitate efficient exploration of solution space (e.g., independent row approximation in BP)[19,40]. Therefore, to achieve good accuracy of parameter inference for larger systems and to gain the ability to generalize to more sophisticated kinetic models, a more general and potentially more powerful data-driven modeling framework would be very useful.

Recently, deep learning has become an effective data-driven framework capable of generating predictions for large and complex systems. Gradient descent implemented

with automatic differentiation, which has been broadly used in training graphical models, allows efficient parameter optimization in complex network systems. This framework has been successfully applied to many domains of biomedical research, from pathology image classification[41,42] to sequence motif detection[43]. While predictive power of deep learning models is often impressive, their interpretation, which is crucial for providing understandable and therefore more trustable predictions, remains challenging. The complex multi-layer network architecture of most deep learning models lacks explicit representations and therefore direct interpretation. This difficulty is sometimes called the "black box" problem[44]. To address this problem, we apply a deep learning optimization approach to learn a data-driven model (called "Cellbox") that incorporates an explicitly interpretable network of interactions between cellular components, instead of a black-box neural network, while aiming to maintain a high level of learning performance.

Cellbox is designed to be a framework for computational modeling of cellular response to perturbations that i) links perturbations to molecular and phenotypic changes in a unified computational model; ii) quantifies time-dependent (dynamic) cellular responses; iii) promises training efficiency and scalability for large-scale systems; iv) is interpretable in terms of interactions that can be compared to established models of molecular biology, such as signaling pathways. Here, we construct a non-linear ordinary differential equations (ODE) based model that represents a biological network of 99 components connecting perturbations, protein response, and phenotypes to simulate dynamic cellular behavior. The network connections are directly learned from post-perturbational data under 89 experimental conditions with the objective of accurately reproducing the cellular and molecular responses on training data and withheld data. To reach this objective, we implemented gradient descent with automatic differentiation to infer interaction parameters in the ODE network, which can then be exposed to novel perturbations. The key performance criterion for the data-driven model trained with a relatively small set of experiments is whether the model is able to provide reasonably accurate predictions on a large set of unseen perturbation conditions.

Anticipating the availability of increasingly informative perturbation-response data sets in diverse areas of cell biology, we present Cellbox as a generally applicable framework for modeling a broad range of dynamic cell behavior.

Results

Cellbox model of perturbation biology

In order to construct a data-driven model to predict the dynamics of molecular and cellular behavior under combinations of drug treatments, the perturbation data has to have 1) paired measurements of changes in protein levels and cellular behavior for a set of perturbations; and 2) training and withheld data to test model performance. Here, we use a perturbation dataset for the melanoma cell line SK-Mel-133[19], which contains molecular and phenotypic response profiles of cells treated with 12 different drugs and their pairwise combinations (Figure 1a). For each of the 89 perturbation conditions, levels of 82 selected proteins and phosphoproteins were measured in cell lysates before and 24 hours after perturbation on antibody-based Reverse Phase Protein Arrays (RPPA). In parallel, cellular phenotypes were assayed, including cell cycle progression and cell viability. With parallel measurements of proteomic and phenotypic responses to a systematic set of perturbations, this dataset provides sufficient information to construct network models that quantitatively link molecular changes to cellular responses.

We used a set of differential equations (ODEs with a non-linear envelope) (Figure 1b) to model the dynamic responses of the system to drug perturbations (See Methods). The parameters of the ODEs (w_{ij}, ~10,000 in total) are the interaction strengths between the entities in the network model. These parameters were randomly initialized and updated throughout the model training process, with the objective to minimize a prediction performance loss function. For the loss function, we chose the Euclidean distance between experimental data and the results of the numerical simulation of the ODE model, plus an L1 regularization penalty on network density to avoid overfitting (See Methods). We used Heun's ODE solver[45] to numerically simulate the ODE system and the Adam optimizer[46] with automatic differentiation to minimize the loss function.

Taken together, we constructed an ODE model of a cell biological system trained using perturbation data, which we named Cellbox.

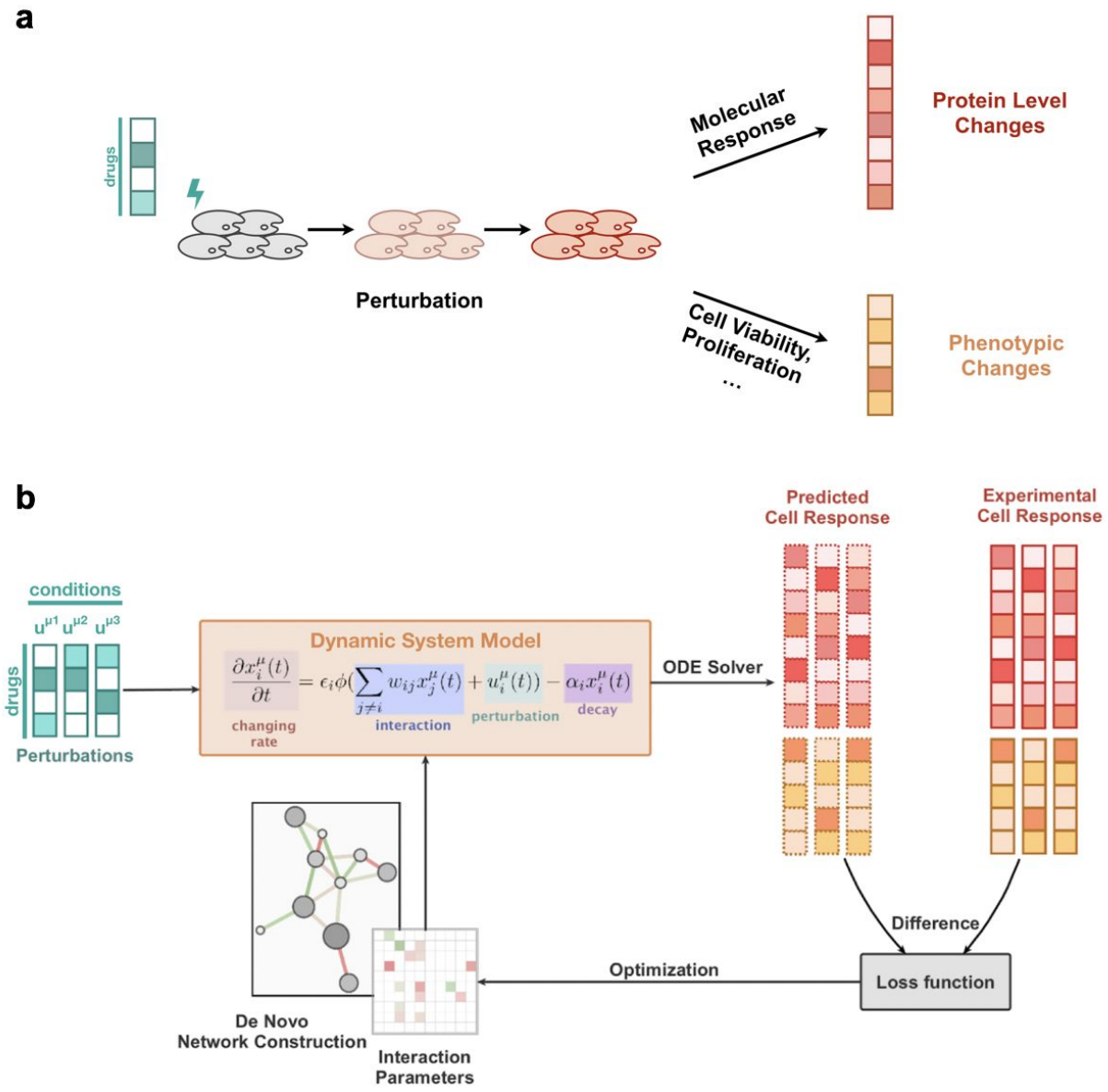


Figure 1. Cellbox: dynamic modeling of cellular systems with perturbation data.

a. Perturbations such as drugs are used to disturb the cellular system. The cell responses, including protein and phosphoprotein level changes, and phenotypic changes, were measured to provide information for model construction. **b.** Systematic responses of the cellular system under various drug perturbations were used to

construct an interpretable machine learning model. Cellbox models system behavior in terms of interaction parameters among system variables using a differential equation system. Cellbox was trained iteratively by changing interaction parameters to fit the numerically simulated system response to experimental observations. After training on pairwise data of input perturbation and output system behaviors, the Cellbox model can be used to predict the cellular response to arbitrary perturbation conditions.

Cellbox can be trained on perturbation data to accurately predict cell response.

In order to test the prediction performance of this training scheme, we randomly selected 70% of the perturbation data (n = 62 conditions) for training and withheld the rest 30% (n = 27 conditions) for testing. 20% of the training data was used as a validation set to stop model training when the performance on the validation set did not further improve. We manually fine-tuned the hyperparameters, including learning rate, regularization, and ODE simulation time, to increase the training efficiency (Figure S1; Supplementary Note 1). At the end of the training, the numerical solutions of the ODE model converged efficiently to experimental data (Figure 2a, 2b). We repeated the modeling scheme with random data partitions to construct 1,000 models for each partition. The average predictions on test sets across all models and all conditions correlate with experimental data with a Pearson's correlation coefficient of 0.94 (Figure 2c). A more refined analysis of individual perturbation conditions showed that the model trains equally well for all conditions and does not bias any particular condition (Figure 2d, Figure S3). The results illustrate that the Cellbox model can be efficiently trained with perturbation data to accurately predict cell response to experimentally applied perturbations.

Even though ~70% of the models reached steady solutions of the ODEs (Figure 2b), some models converged to oscillatory solutions (Figure S2a). In order to test whether the oscillation is an artifact of data partitioning during model training, we re-trained the models with the same train-test data partitioning but multiple different random seeds for the computational optimizer (See Methods). For each individual partition of training data, both steady and oscillatory solutions can result (Figure S2). Based on the assumption that the population average of cell response reaches a stable and non-oscillating steady state 24 hours after drug treatment, we excluded the oscillatory models in the following analysis (See Methods). Taken together, these results indicate that Cellbox, a data-driven ODE-based cellular system model, can be trained to

accurately predict dynamics of cell response, without any requirement of prior knowledge about the relationship between particular protein levels and phenotypes.

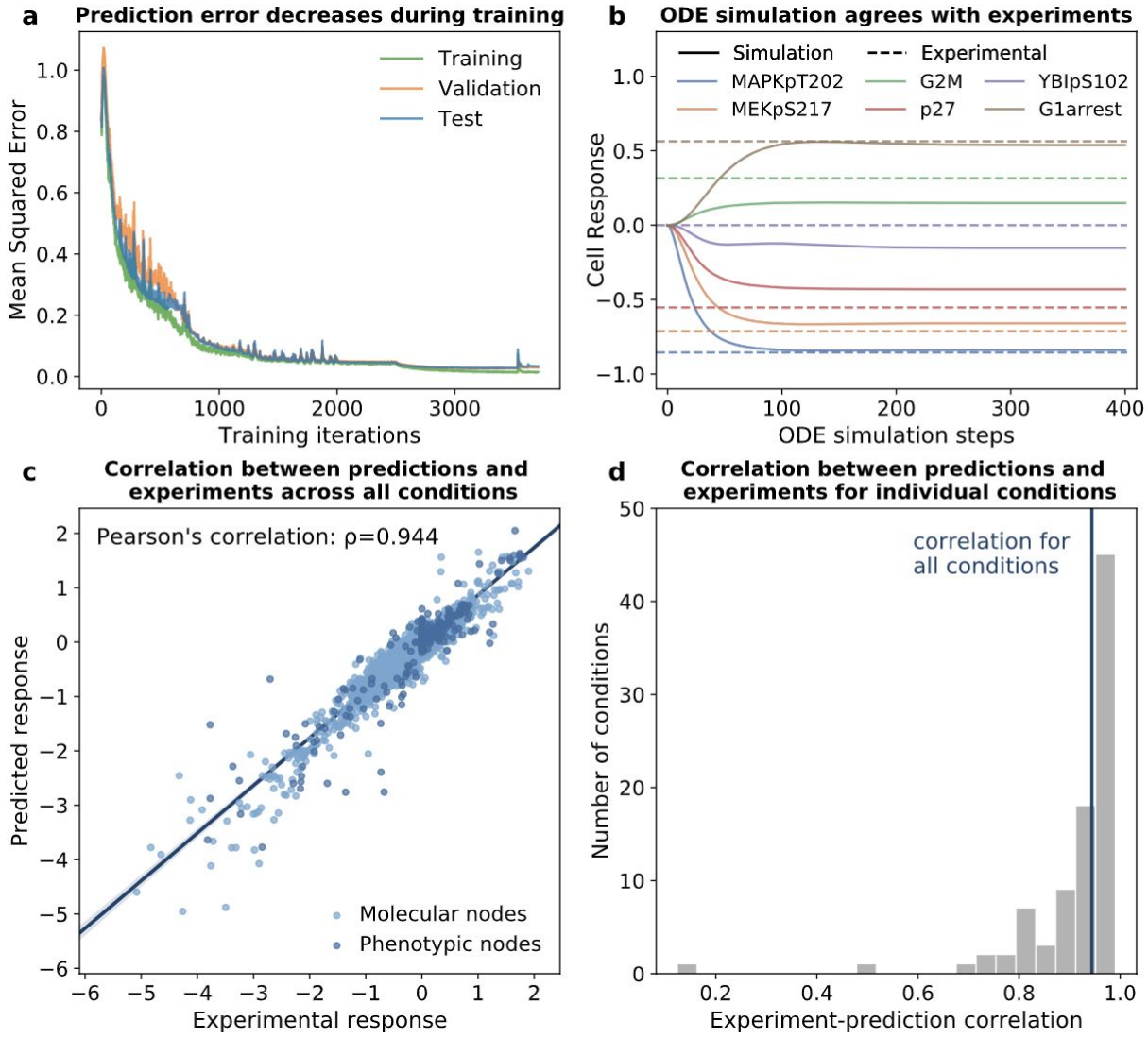


Figure 2. Cellbox convergence and prediction accuracy on randomly partitioned training/test datasets.

a. Over training iterations, the mean squared error on the training set (56% of the entire dataset), validation set (14%) and test set (30%) decreased nearly monotonically and the models converged at the end of the training. **b.** The predicted molecular and phenotypic responses at the steady state of the ODE simulations agree with the

experimental data on the test set. A subset of molecular measurements (MAPKpT202, YB1pS102, MEKpS217, and p27) and phenotypic measurements (G2M and G1arrest) are shown. Cell response is defined as log ratio of post- and pre-perturbation measurements. The annotations and the full set of measurements are in Supplementary Table S1. **c.** Across 1,000 models trained with different data partitions, the average predicted responses correlate with experimental observations (Pearson's correlation ρ = 0.944, regression line in dark blue with 95% confidence interval). Each point represents one measurement, either molecular or phenotypic, in one perturbation condition. **d.** Nearly all predictions for individual conditions have high correlations with experimental measurements.

Cellbox model predicts cell response for single-to-combo and leave-one-drug-out cross-validations.

Even though the model makes accurate predictions with different training data, data partitioning, especially random partitioning, raises the concern of information sharing between training and test datasets. Combinatorial conditions in both datasets might share the same drugs such that the test set might not be truly independent of training and therefore is suboptimal for rigorous evaluation of the model performance. Moreover, the ability to predict the combinatorial effect of a drug, e.g. dominant, additive, synergistic, when none of its combinations has been seen by the model, is a non-trivial challenge in the context of making accurate predictions of experimentally untested drug combinations.

In order to address these points, rather than training the model with random data partitioning, we instead designed more rigorous tasks: single-to-combo (Figure 3a) and leave-one-drug-out cross-validation (Figure 3b, 3c) for each drug. In single-to-combo analysis, all single-drug treatment conditions were used for training and prediction was

tested on all combinatorial drug conditions. In leave-one-drug-out cross-validation, all the combination conditions containing treatment of a particular drug with or without the corresponding single drug conditions were withheld while the rest of the conditions were used for training. In these more stringent tests, we found that the predicted values for withheld data were still highly correlated with the experimental observations (average Pearson's correlation: 0.93 for single-to-combo; 0.94 for leave-one-drug-out with single conditions, similar to that of the training with random partition; 0.79 for complete leave-one-drug-out). Under all three scenarios, on this dataset, Cellbox outperforms the belief propagation (BP) dynamic model approach also used in perturbation biology[19] in terms of predictive accuracy. These results indicate that the Cellbox model can be trained with a relatively small set of perturbation data and that its predictions can be generalized to unseen combinatorial perturbations.

Cellbox models are dynamic network models of a cell biological system. To test whether Cellbox increases model predictive power, we compared the results to those of a static biological network model and a deep neural network model. The static network model was constructed by learning co-expression correlation for each pair of protein nodes (Co-exp) while the deep neural network model was trained to directly regress phenotypic changes against parameterized perturbations (NN) (see Methods). In all three tasks, the static network models had lower accuracy relative to the dynamic Cellbox. The NN had comparable performance to Cellbox in the cross-validation for individual drugs, but its performance dropped significantly in the single-to-combo analysis (Figure 3a). Note that the NN was also unable to generalize to unseen targets whose information is completely excluded from training (Figure 3c). Taken together, due to the lack of mechanistic and dynamic information, static network or direct regression models appear to be less suitable for facilitating the search of combinatorial targets.

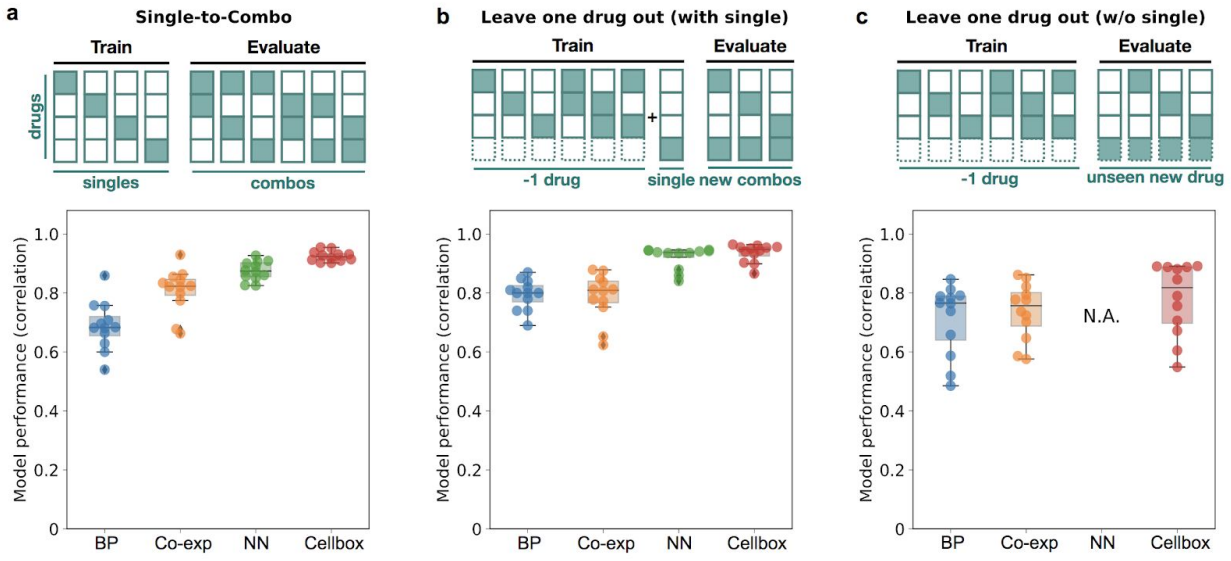


Figure 3. Cellbox models are accurately predictive of cell response for single-to-combo and leave-one-drug-out cross-validations.

a. When only single conditions were used for training (single-to-combo), the Cellbox models predict the effects of combinatorial conditions with good accuracy and outperform the dynamic network models inferred using belief propagation (BP), the static co-expression network model (Co-exp), and a neural network regression model (NN) trained on the same data. **b.** When combinatorial conditions associated with one drug were withheld from training, the Cellbox models retain high accuracy for predicting the effects of unseen drug pairs. **c.** When all conditions associated with one drug were withheld from training, the ODE network models predict the effects of the withheld drug with reduced accuracy but direct regression models such as NN cannot generalize to unseen targets at all. For each model type, performance was evaluated by Pearson's correlation between predicted cell response and experimental cell response.

Model performance is robust against noise and reduced training set size

To examine model robustness of the Cellbox models against reduction in training data, we tested the stability of model performance when the data quality or quantity is compromised. To test the former, we introduced different levels of Gaussian noise (see Methods) into the input molecular and cellular response data and trained models on the resultant noisy datasets. When comparing the predicted response in test sets to the experimental data, we found that the predictions from training on the noisy data retain similarly high correlations to experimental data as those trained on the original data, even with the addition of 5% Gaussian noise (Figure 4a). As the magnitude of the noise increases, the model performance gradually decreases in terms of both convergence (Figure S4) and predictive power (Figure 4a). We concluded that model performance is stable in the presence of moderate experimental errors.

To test the dependency of model performance on data quantity, we trained the model on subsamples of the experimental dataset. We trained models with varying amounts of data (from 10% to 90% in steps of 10%) and found that the models could make accurate predictions of withheld data with as little as 40% of the complete dataset (Figure 4b). We found that increasing the size of the training set further has diminishing returns in terms of model performance. This implies, on the data set used here with an interaction network of ~100 components, that a comparatively small number of perturbation conditions (40-100, rather than directly testing all ~3,000 possible combinations) are sufficient for constructing reasonably predictive models. This example may be a useful guide for power calculations for systems with hundreds of measured components, which would be of considerable interest.

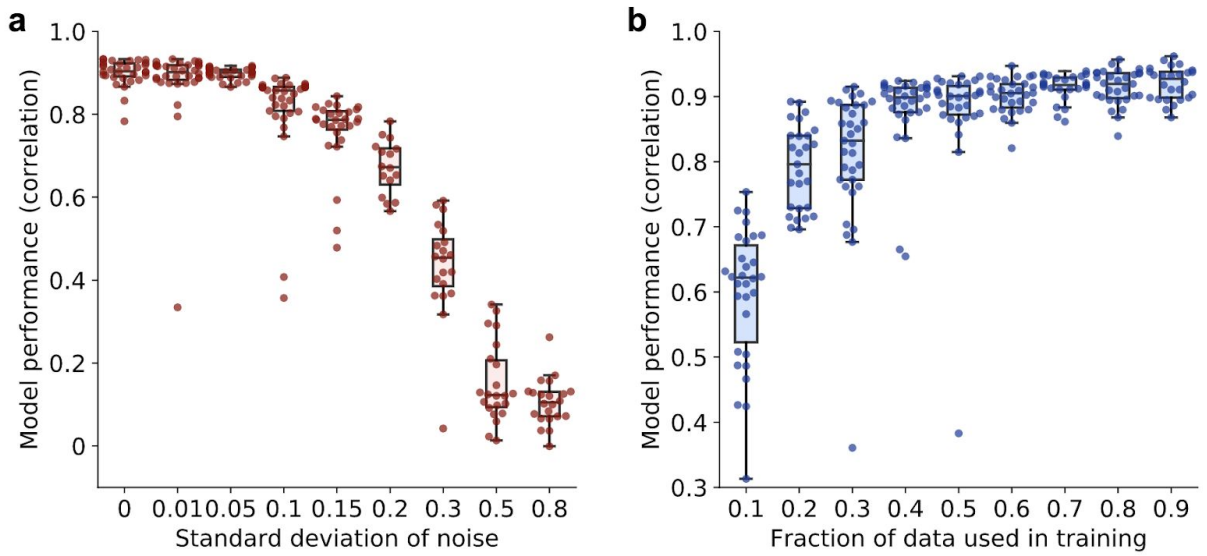


Figure 4. Model performance is stable against data noise and data reduction.

a. Correlation between predicted responses and experimental responses in the test set decreases as an increased level of noise is added to the training data (each dot represents one model). **b.** Correlation between predicted responses and experimental responses in the test set increases with increasing quantity of data used for model training. For the current dataset, the correlation plateaus when 40% of the original dataset is used.

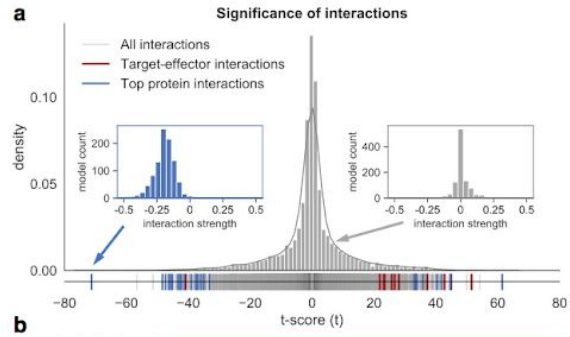
The network model gives interpretable results in biological contexts

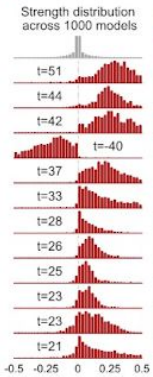
We used ordinary differential equations as the core framework of the current version of the Cellbox mathematical model. Each parameter in the model represents the strength and direction of a biological interaction. In order to investigate whether the inferred interactions are consistent with current knowledge of biology, we used the entire dataset as training data to generate 1000 full models and examined the resulting *de novo* network edges learned from training. We used a t-score (see Methods) as an indication of the statistical significance of each interaction, where a higher absolute value indicates

higher interaction strength and lower variance across the models (Fig. 5a). Using the drugs' primary targets as the ground truth, we first examined the interactions between the drug-activity nodes and their downstream effectors. We found that all 12 drug-activity nodes had significant edge connections to their primary downstream protein effectors with the interaction directions consistent with their expected effects (Fig. 5b), suggesting the models were able to capture the literature-provided interactions between the drug target and their downstream effectors.

To further investigate how much the network represents known pathway interactions, we compared the interactions in the Cellbox models to corresponding molecular interactions extracted from the Pathway Commons (PC) resource[10], which is an aggregation of ~20 curated publicly available pathway interaction databases. In this comparison between Cellbox models and prior knowledge, model interactions can in principle be found in PC either as one-step (A-B), two-step (A-X-B), or logical (>2 steps) interactions, where the logical interactions can be usefully predictive or perhaps erroneous. Direct interactions (A-B) in PC represent one component affecting the expression, phosphorylation, or state-of-change of the other component (B) (Figure. 5c, 5d, Supplementary Table S2). For example, models and prior knowledge agree for the phosphorylation of the AKT protein kinase (AKTpS473) that negatively affects the insulin receptor substrate 1 (IRS1)[47] and, for the activation of the mitogen-activated protein kinase (MAPKpT202) that affects p27 protein levels[48,49]. Some model interactions can be found as indirect two-step interactions (A-X-B) in PC, meaning at least one path with one intermediate component (X) can be found between the two components (A and B). For example, retinoblastoma protein (Rb1) affects cyclinD through p21[50,51]; and mitogen-activated protein kinase kinase (MEK1) indirectly interacts in two steps with the transcription factor c-Myc by a phosphorylation mechanism via ERK1/2[52–55]. Additional evidence of functional links between proteins identified in the models include protein-protein edges in the STRING database[56], such as links that represent expression (mRNA) correlation, and PubMed-derived links from

text-mining[57]. The remaining model interactions are between components that are more than two pathway steps away or cannot be connected by any path in PC. These interactions can be interpreted as either logical interactions important for predictive purposes, or potential new physical interactions that are yet to be discovered in molecular experiments. Since the Cellbox network models are constructed in a completely data-driven way without any a priori intention to recapitulate known molecular interactions, the partial agreement between the model-inferred molecular interactions and the experimentally known ones from the literature is evidence of validity of the modelling approach.





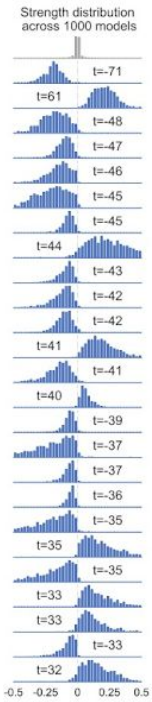
Drug target-effector interactions

Drug target	Effector	Mean strength	Expected
amTOR	S6pS240	0.29	positive
aPI3K	AKTpS473	0.26	positive
aMEK	MAPKpT202	0.39	positive
aHDAC	p27	-0.23	negative
aBRAFm	MAPKpT202	0.27	positive
aCDK4	RbpS807	0.21	positive
aSTAT3	STAT3pY705	0.09	positive
aJAK	GSK3abpS21	0.11	positive
aMDM2	4EBP1pS65	0.08	positive
aSRC	GSK3abpS21	0.08	positive
aPKC	S6pS240	0.12	positive
aAKT	AKTpS473	0.18	positive

c

Top protein-protein interactions

Upstream	Downstream	Mean strength	Interaction in PC
S6pS235	STAT5pY694	-0.20	-
S6pS235	PLK1	0.21	-
AKTpS473	AKT	-0.18	self
AKTpS473	IRS1	-0.10	direct
COX2	IGF1R-beta	-0.15	indirect
COX2	STAT3pY705	-0.21	direct
AKTpT308	STAT6pY641	-0.07	direct
4EBP1pS65	4EBP1pT37	0.25	self
4EBP1pS65	EGFRV	-0.10	-
4EBP1pT37	ACCpS79	-0.13	-
AKTpT308	BCL2	-0.11	direct
GSK3abpS21	TSC2pT1462	0.21	direct
S6pS240	Fibronectin	-0.15	-
GSK3abpS21	a-tubulin	0.07	indirect
S6pS240	CyclinE1	-0.07	-
MAPKpT202	p27	-0.20	direct
GSKpS9	EGFRV	-0.07	indirect
4EBP1pT37	ELK1pS383	-0.06	-
YBIpS102	p27	-0.22	-
YBIpS102	c-Myc	0.20	-
MAPKpT202	BIM	-0.18	direct
RbpS807	CyclinD1	0.18	indirect
GSKpS9	4EBP1pT37	0.13	direct
AKT	BIM	-0.08	direct
IRS1	S6pS240	0.15	indirect



d

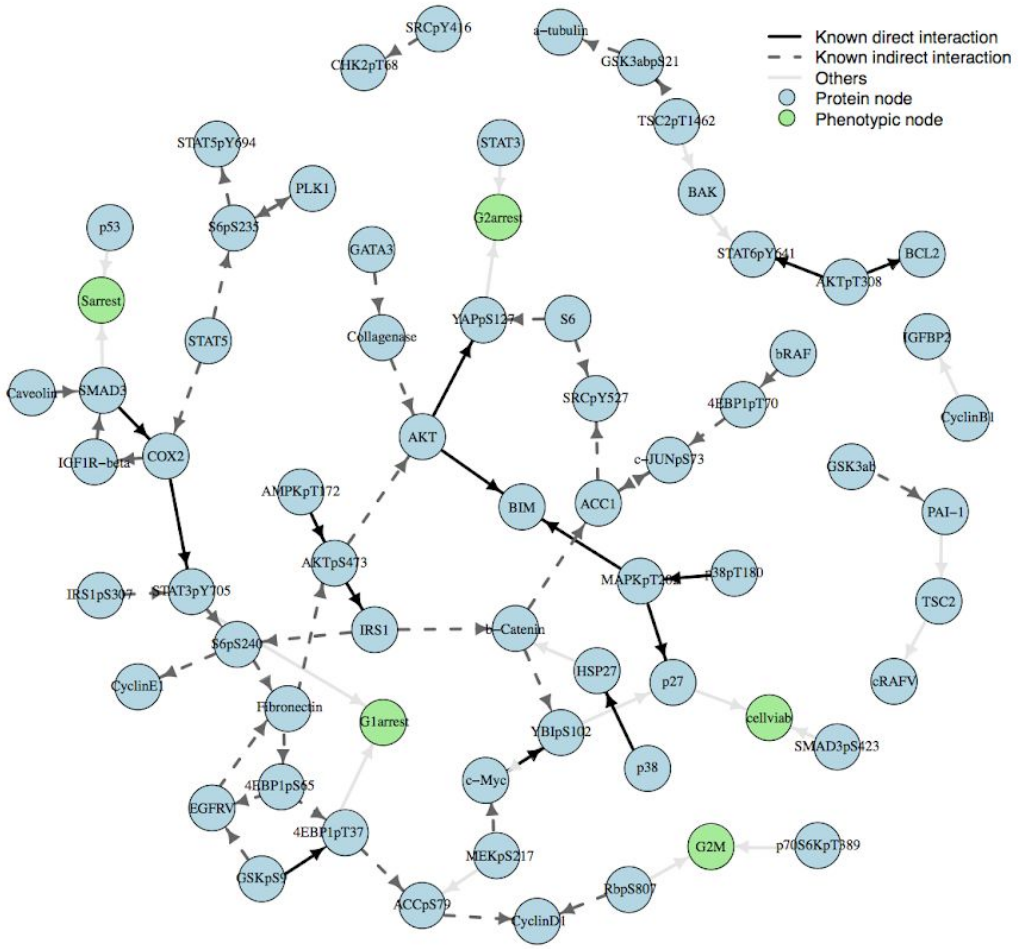


Figure 5. Interpretation of interactions in the network model in biological contexts

a. The t-score distribution of all interactions across 1000 full models suggests that a small fraction of interaction strengths is significantly different from zero. Insets are two examples of interaction strength distributions across models. **b.** All 12 interactions between drug target (drug activity nodes) and their downstream effectors (red bars in A) are significant, and the interaction directions are consistent with the literature. **c.** Most of the top significant protein-protein interactions (blue bars in A) can be found as direct or indirect interactions in Pathway Commons (PC). The distributions of interaction strength across 1000 models for each interaction in the two tables with corresponding colors are centered away from zero, in contrast to the background distributions of aggregated interactions across models (gray, all interactions with drug activity nodes in **a**, all protein-protein interactions in **c**). All other interactions with t-scores and PC information are included in Supplementary Table S2. **d.** A network visualization of the top interactions (top two interactions acting from each node and onto each node) highlights the level of agreement between model-inferred interactions and those in PC.

Predictions of unseen perturbations give candidates for drug combinations

Our results so far indicate that the Cellbox model can be efficiently trained on a relatively small set of experimental data to parametrize the differential equations that model the behavior of the entire system of nodes and interactions at a reasonable level of predictive accuracy. This model can then predict cell responses to a full range of single and combinatorial unseen perturbations, that would be laborious and costly to test exhaustively by experiment. In order to nominate effective drug combinations for a much reduced number of focused experiments, we used simulations of the 1,000 full models to quantitatively predict the dynamic cell responses to ~160,000 in silico perturbations, including different dosages of single perturbations on each protein node as well as all pairwise combinations (see Methods). For each perturbation condition, we averaged the predictions across all models and ranked the perturbations by predicted phenotypic changes (Figure 6a).

Previous models on the same dataset, using the same differential equations but parametrized using belief propagation, had predicted that two drug pairs, MEKi+c-Myc and RAFi+c-Myc, would increase G1 cell cycle arrest and this prediction was confirmed by experiments[19]. We found that the Cellbox model predicts similar effects for these two drug pairs (Figure 6b). In order to identify additional therapeutic candidates, we examined the effects of all possible single and pairwise perturbations on cell cycle arrest (Figure 6b, 6c). The top-ranked candidates included dominant anti-proliferative inhibition (uniform colors in rows or columns) of proteins in the Wnt, MAPK, and ERK/MEK pathways, known to be cancer-related. Besides strong single candidates, synergistic drug pairs are of potential therapeutic interest (Figure 6b, 6c departure from uniform colors). Inhibitory perturbations predicted to have pro-proliferation effects, which are undesirable as such, can also lead to effective anti-proliferative candidates via indirectly activating perturbations (Figure 6c, top left corner). For example, protein nodes can in principle be activated by reducing upstream inhibition or degradation. As the Cellbox model is completely data-driven, the *de novo* predictions represent system-specific predictions independent of prior knowledge.

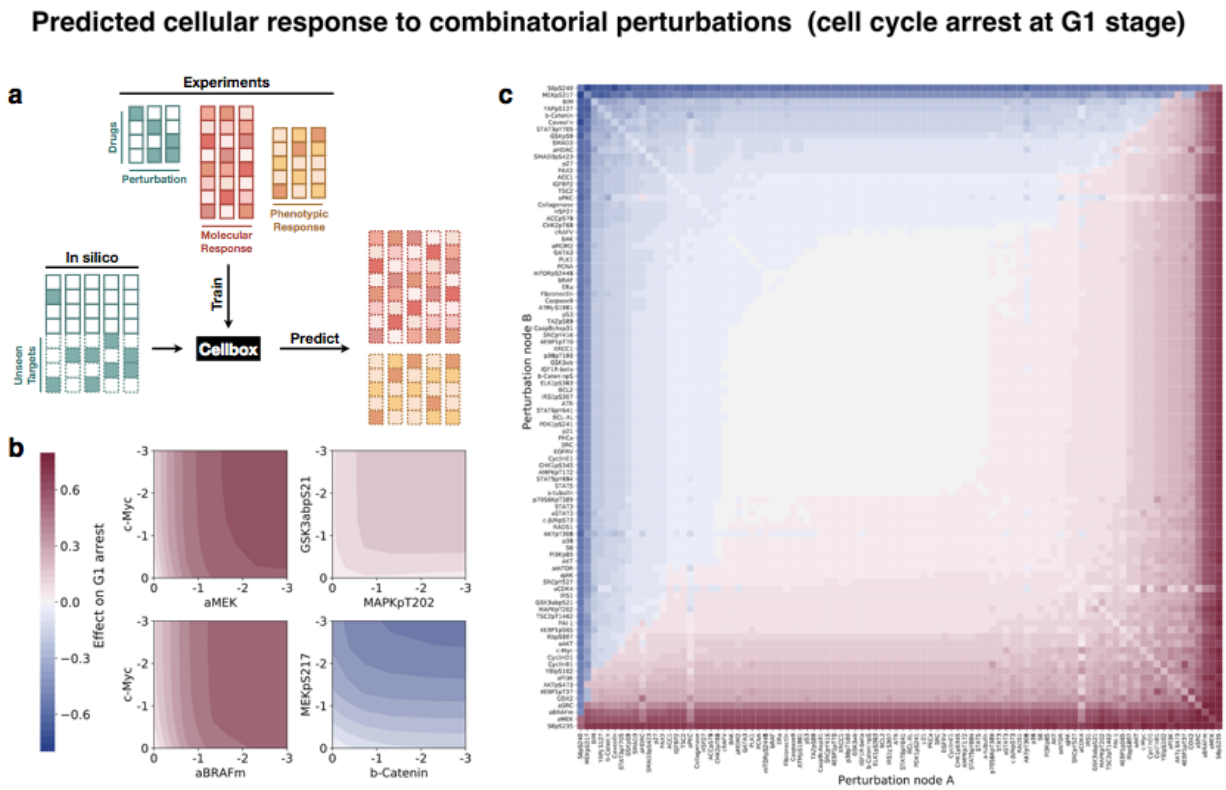


Figure 6. Cellbox provides testable predictions of cell phenotype under synthetic perturbations.

a. For each (phospho)protein node in the network, we simulated the inhibition effect of all single and paired inhibitions and used Cellbox to predict the phenotypic change. The phenotypic effects are the average prediction of 1,000 independent models trained on the full datasets. Probing: experimentally perturbed using drug treatment; profiled: measured with RPPA and cellular assay; untested: profiled but not perturbed. **b**. We more closely examined the anti-proliferation effect of two perturbation pairs whose effects on cell cycle arrest have been experimentally tested (left two panels, c-Myc+MEKi, and c-Myc+RAFi), as well as two other in silico conditions (right two panels, GSK3p+MAPKp and MEKp+b-Catenin), by simulating with combinatorial perturbation strengths. **c**. The effect on cell cycle arrest of pairwise combinatorial perturbation of all (phospho)proteins in the network were simulated and used to nominate effective pharmaceutical candidates. These in silico inhibitory perturbations

can result in anti-proliferation effects (red, bottom right) or pro-proliferation effects (blue, top left).

Discussion

Quantitative models that are predictive of dynamic cellular responses can be used to design combination therapies in cancer. To provide predictions with sufficient accuracy and potential mechanistic insight, we integrated machine learning methods with dynamic modeling: we applied an optimization algorithm used in deep learning to a biologically interpretable differential equation (ODE) system. Our model can be trained efficiently and independently of prior knowledge to predict molecular and phenotypic responses to unseen perturbations with high accuracy. Although trained on a relatively small set of experiments, the model is capable of simulating cell responses to numerous arbitrary combinatorial perturbations and dosages applied to nodes repeatedly measured under different perturbation conditions. Ranking of cellular responses to the in silico combinatorial perturbations by the desired phenotypic outcome, such as decreased proliferation, then leads to specific therapeutic hypotheses.

Interpretability of models that are to be used for practical decisions, such as the design of combination therapy, helps increase confidence and facilitates the design of focused validation experiments and is therefore as important as accuracy[58]. Cellbox features interpretability in two aspects: transparency and traceability. *Transparency*: by using a well-defined mathematical model, Cellbox is designed to be explicitly interpretable. In the current ODE model, each individual parameter represents a directed and quantitative interaction between cellular components or interaction with phenotypic quantities. *Traceability*: given a perturbation of the cellular system, the ODE simulation indicates how the effects of perturbations propagate throughout the directed network in a time-dependent manner. The models can therefore provide mechanistic hypotheses of how the perturbations cause the observable cellular responses. Our model aims to meet the emerging demands of explainable artificial intelligence (XAI), which is

essential for a socially acceptable application of machine learning models in biomedicine[59,60].

In principle, Cellbox is generalizable to other types of systems and larger systems. Other types of models will presumably benefit from automatic differentiation (AD) combined with stochastic gradient descent that performs optimization directly for any given mathematical ansatz and, therefore, can avoid oversimplified approximations[61]. The flexible AD framework allows the models to be easily adapted to various forms of cellular kinetics and dynamics. The ability to model larger systems depends both on the availability of larger datasets and scalable modeling methods. Larger datasets can be obtained by measuring diverse types of molecular data, for example, transcriptomic, epigenomic and metabolomic changes[62,63], by measuring a larger number of molecular or phenotypic observables, such as protein levels by mass spectrometry, or by multiplexing. A major opportunity for larger datasets may arise from recent cell barcoding techniques that significantly increase perturbation throughput relative to arrayed experiments[17,64] by measuring transcript levels or antibody levels labelled by oligonucleotide or isotopes at the single cell level[65–67]. As Cellbox is implemented in the Google TensorFlow framework, it can make use of various advanced machine learning techniques, such as dropout, mini-batching, and GPU boosting[68,69] to improve training efficiency, which partially addresses the issue of scalability.

A particular advantage of our modelling approach is its complete independence from prior knowledge. This is in contrast, e.g., to other frameworks for interpretable models[37] that incorporate prior knowledge by pre-defining the connections in the network model based on large-scale curation of molecular and biochemical pathways. While our current implementation of the model is completely data-driven, prior knowledge of cellular interactions could also be included by adding a penalty to the optimization function, which quantifies the disagreement between inferred and prior values for each parameter. Another particular advantage, compared to related algorithms[37], is the ability

to predict changes in molecular levels after perturbation, not just cellular phenotypic outcomes. It would be of interest to explore whether one can combine the efficiency advantage in[37] via adjoint analysis[70] with the advantages of the current implementation of Cellbox.

For broad translational applicability a tantalizing but challenging prospect is to incorporate individual tumor background via genetic perturbations and to propose optimal, personalized combinations of targeted therapeutics. We envision the systems biology approach described here to be broadly applicable to other areas of biology, such as developmental biology or synthetic biology, provided that suitable perturbation-response data becomes available. Key future challenges are therefore the design of experiments for each biological context of interest and the further development of transferable and scalable machine learning methods.

Methods

Perturbation dataset overview

The Cellbox models were trained using a perturbation-response dataset of the SK-Mel-133 melanoma cell line[19]. The cells were treated with 12 different single drugs each at two different concentrations and 66 pairwise combinations of these drugs at IC40 concentrations. 24 hours after drug treatment, Reverse Phase Protein Arrays (RPPA) were used to measure the level of 45 proteins and 37 phosphoproteins of interest. Cell cycle progression, including G1 arrest, G2 arrest, G2/M transition, and S arrest was measured by flow cytometry. Cell viability was measured 72 hours after drug treatment by the resazurin assay. The dataset was initialized with 12 drug activity nodes representing the inhibition strengths of different drugs to their targets[71]. The resulting dataset has 89 perturbation conditions and 99 observed nodes (82 protein and phosphoproteins, 5 phenotypes, and 12 drug activity). A more detailed description of the experimental data set is available in Korkut et al[19].

Model configuration

The models were constructed using Python 3.6 and Google Tensorflow (version = 1.9.0, https://www.tensorflow.org/about/bib). The molecular and phenotypic changes are linked in a unified biological network model using a system of ordinary differential equations

$$\frac{\partial x_i^{\mu}(t)}{\partial t} = \epsilon_i \phi\left(\sum_{j \neq i} w_{ij} x_j^{\mu}(t) - u_i^{\mu}(t)\right) - \alpha_i x_i^{\mu}(t) \tag{1}$$

where $x_i^{\mu}(t)$ represents the $\log_2$-normalized relative change of each (phospho)protein or phenotype levels relative to control levels under condition μ. $u_i^{\mu}(t)$ quantifies the

strength of the perturbation on target (i). Here the drug effect is assumed to be constant and therefore $u(t) = u$ for $t > t_0$. α_i characterizes the effect of decay, meaning the tendency of protein i to return to the original level before perturbation. The interaction parameters w_{ij} indicate interactions between network node j on network node i, assumed to be a constant property of the pair of molecules in this given cellular setting. We constrain the interaction parameters w_{ij} by disallowing three classes of interactions:

i) ingoing connections for drug nodes (drugs cannot be acted upon by any other node)

ii) outgoing connections for phenotypic nodes (phenotypes cannot act on any other nodes)

iii) self-interaction (nodes cannot act on themselves)

We use a sigmoid function $\varphi(\cdot) = tanh(\cdot)$, to model the saturation effect of the interaction term so that it is bounded by the value of ε_i.

The biological network interactions were constructed *de novo* without any prior knowledge input, meaning the network was fully connected with interaction parameters. The interaction parameters were randomly initialized and the ODE system was numerically solved using Heun's method[45] (eqn. 2, time steps $N_t = 400$, supplementary Figure S1), which is an improved variant of Euler's method. Model performance was evaluated by disagreement between the experimental cell responses and the numerical steady state levels.

$$\tilde{x}_i(t+h) = x_i(t) + hf\big(t, x_i(t)\big)$$

$$x_i(t+h) = x_i(t) + \frac{h}{2}\left[f\big(t, x_i(t)\big) + f\big(t_{i+1}, \tilde{x}_i(t+h)\big)\right] \quad (2)$$

Where h is the step size, $f\big(t, x(t)\big) = x'(t)$, $y(t_0) = \log(1) = 0$.

The loss function $L(w)$ is defined as a weighted sum of prediction error and complexity penalty in order to avoid overfitting. Here a mean squared error (MSE) and an L1-loss

regularization term are used, as defined in (3). The interaction parameters were optimized end-to-end using the Adam optimizer[46], with the objective of minimizing the loss function.

$$L(\mathbf{w}) = \sum_{\mu} \sum_{i} ||\hat{x_i^{\mu}}(\tau^*) - x_i^{\mu^*}||_2^2 + \lambda ||\mathbf{w}||_1 \qquad (3)$$

$\hat{x_i^{\mu}}$ is calculated as the converged value of the numerical simulation of the ODE with defined simulation timestep N_t. The dataset was divided into training, validation, and test sets, in order to optimize parameters, provide an indication for stopping training, and for testing model performance, respectively. Optimization was conducted with an initial learning rate for the Adam optimizer (lr=0.1) and regularization strength (λ=0.01). It has been shown that gradually decreasing learning rate is helpful for model convergence[68]. The model training was stopped when the loss function of the validation set does not further decrease for a continuous of 20 iterations (stopping patience).

The model was trained with mini-batching: a random 80% portion of the training set was used to optimize parameters for each iteration. Models that failed to converge (MSE for training set > 0.05) were excluded as unsuitable.

Model training with random data partitions

For initial model training and analysis of model performance, the cell line perturbation-response dataset was randomly partitioned into training, validation, and test set in the proportion of 56% (n=50 conditions), 14% (n=12 conditions), and 30% (n=27 conditions). 1500 models were generated on 1500 independently random-partitioned datasets.

The models were examined and categorized into non-oscillating and oscillating solutions based on time derivatives at the final time step of the ODE simulation. The non-oscillating solutions are defined as those with the average absolute value of time

derivatives of all nodes and conditions in the training set smaller than δ, i.e. $\frac{1}{m}\sum_{i=1}^{m}|\frac{\partial x_i^{\mu}(t)}{\partial t}| < \delta\,;\delta = 1e-03$. In each category, twenty models were randomly selected and each re-trained with the original data partitioning but forty different random seeds, covering all the random processes in the training, including parameter initialization and mini-batching sampling (Figure S2). Oscillating solutions comprise about 30 percent of all models. In the following analysis, models that converged to oscillating solutions were excluded.

Single-to-combo and leave-one-drug-out cross validation

To evaluate model performance by cross-validation for each drug, the data was partitioned into training (n=78 conditions) and test (n=11 conditions) sets where each test set contains all the drug combination conditions with the particular drug. 20% of training conditions (n=15) are used as a validation set. The predictions on the test set were averaged over 100 models. In the single-to-combo task, all single-drug conditions were allocated to the training set (n=23 conditions), and the combination perturbation conditions were randomly distributed among the validation and test set (n=53 conditions) in a 20/80 ratio.

The Belief Propagation (BP) models for both cross-validation and single-to-combo prediction were performed as in our earlier publication (https://github.com/korkutlab/pertbio). The predictions on the test set were averaged over 100 models. The deep neural network model (NN) network has 2 densely connected hidden layers (hidden layer H1: 20 neurons, H2: 100 neurons) that are densely connected, connecting the parameterized perturbation tensor with the cell response tensor. The NN models were constructed in the Tensorflow framework in Python and optimized using the same optimization methods (Adam optimizer). The co-expression static model (Co-exp) was constructed in a python environment using the sklearn (version = 0.21.3, https://scikit-learn.org/stable/) LinearRegression module. The model was trained to use the changes in levels of each pair of protein nodes to predict

the rest of the proteins and phenotypes. To compare the four different predictive models, Cellbox, BP, NN, and Co-exp, a t-test on two related samples was used to analyze the significance of the difference of model predictions and to assign a p-value.

Sensitivity analysis with noise and reduced training set size

We conducted a sensitivity analysis of our model to evaluate the robustness of its prediction in response to noise. We added varying levels of Gaussian noise to the input molecular and phenotypic data (eqn. 4).

$$x_i(t) = x_i^*(t) \cdot N(1, \sigma) \quad (4)$$

The scaling factor for each node and each condition was independently drawn from a Gaussian distribution $N(1, \sigma)$, with a mean of 1 and standard deviation of σ. For each noise level, we evaluated 15 different training/validation/test partitioning and each with 5 independent random noise patterns. Model training was performed on noisy training and validation sets, while the model performance evaluation was performed on the original, noise-free test data. For each noise level, the percentage of successful models, defined as those that converged in terms of both MSE and oscillation filters, was recorded.

We examined model sensitivity to training and validation set size. We reduced the combined size of the training and validation set, from 90% to 10% in steps of 10%, while keeping their relative size constant, 4:1. The remaining data was allocated to the test set. For each training set size, the percentage of successful models, defined as those that converged in terms of both MSE and oscillation filters, was reported.

Biological interpretation of the network model

The entire dataset was used to generate 1000 full successful models, each with an independent data partitioning of training (n=71 conditions) and validation (n=18 conditions). For each interaction (w_{ij}) between two nodes, a t-score ($\frac{\bar{w}_{ij}}{s\sqrt{m}}$), was calculated as an indication of the confidence level of obtaining a value different from

zero, where $\bar{w}_{ij}$ is the average interaction strength across models, s is the standard deviation, m is the number of models (m = 1,000) (Supplementary Table S2).

In order to compare the model inferred interactions to those present in prior-knowledge pathway databases, all the proteins and phosphoproteins nodes were identified by their corresponding gene names (Supplementary Table S1). The interactions were compared against the Pathway Commons (PC) database (current version at https://www.pathwaycommons.org/archives/PC2/v11) using the paxtoolsr[72] software. The database was filtered down to direct interactions, which include "controls expression of", "controls phosphorylation of", "controls state change of", "controls production of", "controls transport of", and "controls transport of chemical". All the direct interactions were converted to a directed PC graph using igraph (version = 1.2.4.1, https://igraph.org/r/). Distance was calculated as the length of the shortest path(s) between two components in the PC graph (Supplementary Table S2, column: distance). The number of the shortest paths (npath) between the two components was also reported. Additional detailed information on protein links in Homo sapiens was obtained from the STRING database, including combined score (string) and individual-channel scores (co-expression, experimental, database, text-mining) (https://stringdb-static.org/download/protein.links.detailed.v11.0/9606.protein.links.detailed.v11.0.txt.gz)[56]. The co-citation score was calculated as the number of papers mentioning both components in a customized paper collection aggregating all PubMed papers that refer to at least two and no more than five gene names[57].

Model predictions for a large number of unseen perturbations

We used the models trained with the full (non-partitioned) dataset to simulate responses of novel, experimentally unobserved, in silico perturbation conditions. These conditions included different doses of single perturbations (different levels of perturbation strength $u \in [0,3]$ on all individual (phospho)protein and drug activity nodes within the network and as well as all pairwise combinations ($n_{all} = 94 \times 6 + C_{94}^{2} = 157,920$ conditions). The

cell responses were dynamically simulated with u as the input perturbation with the same number of steps as in training ($N_t = 400$). For each perturbation condition, predictions for cell responses were averaged across 1,000 different models. Perturbations were nominated as therapeutic candidates by ranking the predicted magnitude of the phenotypic change in terms of cell cycle arrest.

Code and data availability

Cellbox code and data are available at https://github.com/dfci/CellBox.

Acknowledgements

We thank Alexandra Franz, Frank Poelwijk, Laura Kleiman, Nicholas Gauthier, Haozhe Shan, William Yuan, Han Altae-Tran and members of the Sander and Marks labs for constructive discussions. Funding support from DFCI, NHGRI (U41HG006623) and NIGMS (P41GM103504).

Supplementary materials

During the training of each model, the training loss decreased along training time together with the test loss (Stage 1). The model parameters started to fluctuate around a local optimum after N = 2,000 training iterations. It has been shown that a decreasing learning rate is helpful for model convergence[68]. Decreasing the learning rate to 0.1x allowed the model to escape local minima and continue learning (Stage 2). The loss function stopped decreasing again after (up to) N = 4,000 iterations, when the magnitude of the MSE loss was comparable to that of the regularization loss. To further improve training and decrease MSE loss mainly, we decreased the regularization strength by loosening the L1 constraints on the parameters (Stage 3). MSE decreased further while the numerical range of the parameters (interaction strengths) started to increase. Continuous decreasing of the learning rate did not further change the loss (Stage 4). ODE simulation of the model indicated that a steady state had not been fully reached. Therefore, ODE simulation time was doubled twice (Stage 5 and Stage 6) while the learning rate and regularization were kept the same as Stage 4. The training and testing loss together with the ODE trajectory indicated that, at the final stage, the models had converged, optimization made no further improvements, and the ODE simulation has reached a steady state. We then stopped the training and examined the results closely on the test dataset.

Figure S1 Multi-step fine-tuning of hyperparameters facilitates model training.
Models were trained in six individual stages with varying learning rate, regularization strength, and ODE simulation time. As the training proceeds, 1. the distribution of differences between predicted and experimental values in the test set narrowed around zero; 2. the distribution of interaction strengths widened as the L1 regularization was weakened; 3. both the training and test loss decreased; 4. ODE simulation reached steady state as simulation time increased (Supplementary Note 1).

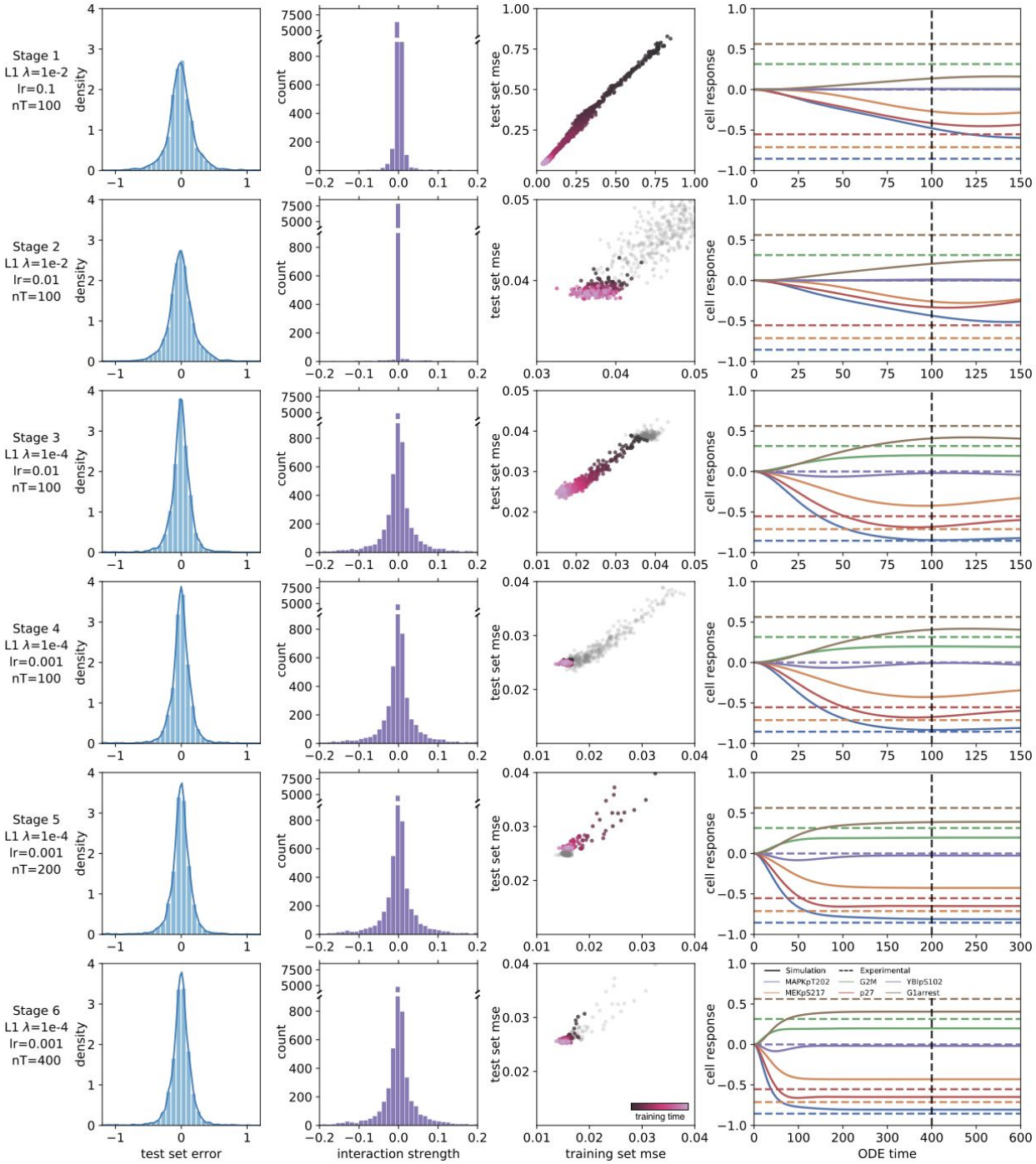

Stage 1
L1 λ=1e-2
lr=0.1
nT=100
Stage 2
L1 λ=1e-2
lr=0.01
nT=100
Stage 3
L1 λ=1e-4
lr=0.01
nT=100
Stage 4
L1 λ=1e-4
lr=0.001
nT=100
Stage 5
L1 λ=1e-4
lr=0.001
nT=200
Stage 6
L1 λ=1e-4
lr=0.001
nT=400
density
count
test set mse
cell response
test set error
interaction strength
training set mse
ODE time
training time
Simulation
Experimental
MAPKpT202
G2M
YBIpS102
MEKpS217
p27
G1arrest

Figure S2. Oscillatory models from stochastic training are independent of data partitioning.

a. Models were examined and categorized into non-oscillatory and oscillatory solutions based on the ODE simulation trajectories. **b-d.** For each data partitioning of training and test set (rows) in the two categories, different seeds for random processes in model training (columns). The models were examined for their performance in terms of average derivatives of each variable at the end of the ODE simulation (**b**), average mean squared error in the training set (**c**), and Pearson's correlation between prediction and experimental data (**d**). Therefore, the solution oscillation and the model convergence are independent of the data partition.

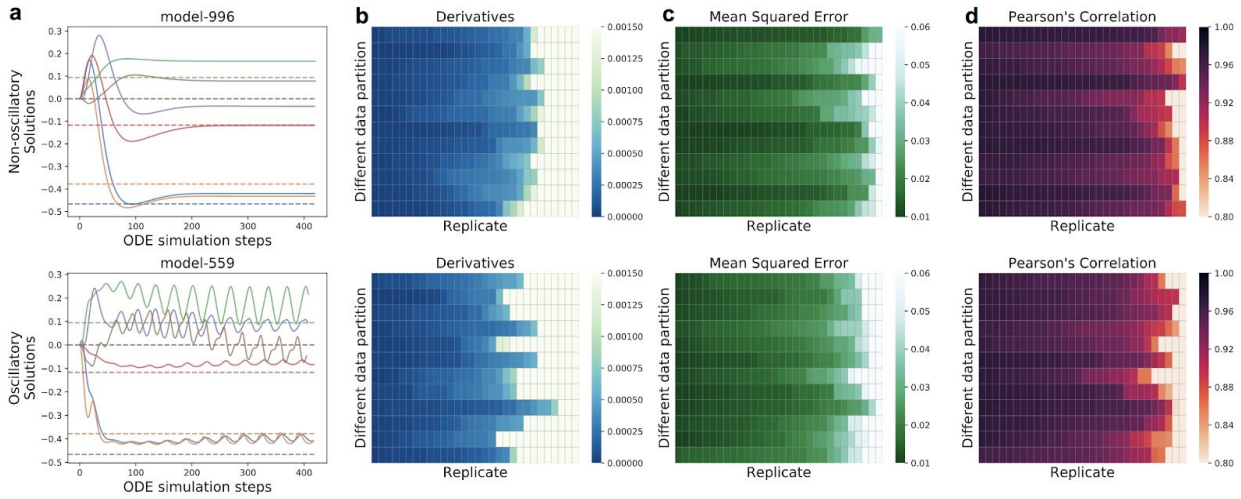
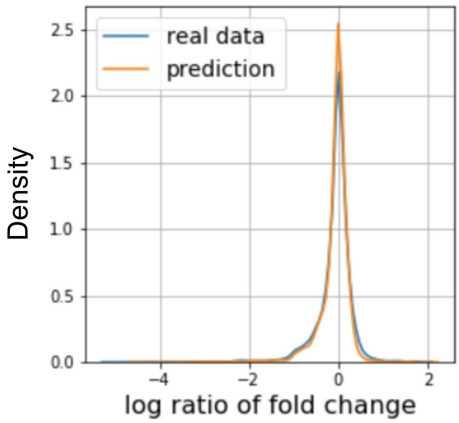


Figure S3. Correlations between predicted and experimental data were consistently high across different perturbation conditions.

a. Model prediction and experimental data had a similar range and distribution without skewing and extreme predictions. **b**. In addition to overall performance, the model predictions for each perturbation condition were examined. The prediction of cell response under each individual condition reached a similar high correlation (median Pearson's correlation 0.95) with experimental data. Meshes indicate the range of real data. Models generally performed better for conditions with larger data range.

a



b

901|1.5 ρ=0.95
901|1.5,HN|6 ρ=0.94
901|1.5,NT|3 ρ=0.95
901|1.5,P6|20 ρ=0.96
901|1.5,PLX|60 ρ=0.96
901|1.5,RO|3.5 ρ=0.93
901|1.5,SR|2.4 ρ=0.94
901|1.5,ST|0.6 ρ=0.95
901|1.5,Tm|0.3 ρ=0.99
901|1.5,ZS|0.6 ρ=0.96
901|3 ρ=0.84
AK|10 ρ=0.95
AK|5 ρ=0.97
AK|5,901|1.5 ρ=0.98
AK|5,HN|6 ρ=0.99
AK|5,NT|3 ρ=0.98
AK|5,P6|20 ρ=0.98
AK|5,PLX|60 ρ=0.98
AK|5,RO|3.5 ρ=0.97
AK|5,SR|2.4 ρ=0.99
AK|5,ST|0.6 ρ=0.98
AK|5,Tm|0.3 ρ=0.98
AK|5,ZS|0.6 ρ=0.96
HN|12 ρ=0.85
HN|6 ρ=0.93
HN|6,NT|3 ρ=0.93
HN|6,RO|3.5 ρ=0.80
HN|6,SR|2.4 ρ=0.93
HN|6,Tm|0.3 ρ=0.98
NT|3 ρ=0.52
NT|3,RO|3.5 ρ=0.97
NT|3,SR|2.4 ρ=0.75
NT|3,Tm|0.3 ρ=0.97
NT|6 ρ=0.68
P6|20 ρ=0.79
P6|20,HN|6 ρ=0.90
P6|20,NT|3 ρ=0.81
P6|20,RO|3.5 ρ=0.97
P6|20,SR|2.4 ρ=0.88
P6|20,ST|0.6 ρ=0.75
P6|20,Tm|0.3 ρ=0.98
P6|40 ρ=0.12
PLX|120 ρ=0.89
PLX|60 ρ=0.90
PLX|60,HN|6 ρ=0.96
PLX|60,NT|3 ρ=0.94
PLX|60,P6|20 ρ=0.90
PLX|60,RO|3.5 ρ=0.95
PLX|60,SR|2.4 ρ=0.95
PLX|60,ST|0.6 ρ=0.92
PLX|60,Tm|0.3 ρ=0.99
RO|3.5 ρ=0.85
RO|3.5,Tm|0.3 ρ=0.95
RO|7 ρ=0.89
RY|2 ρ=0.94
RY|2,901|1.5 ρ=0.94
RY|2,AK|5 ρ=0.95
RY|2,HN|6 ρ=0.93
RY|2,NT|3 ρ=0.97
RY|2,P6|20 ρ=0.95
RY|2,PLX|60 ρ=0.96
RY|2,RO|3.5 ρ=0.90
RY|2,SR|2.4 ρ=0.96
RY|2,ST|0.6 ρ=0.98
RY|2,Tm|0.3 ρ=0.98
RY|2,ZS|0.6 ρ=0.96
SR|2.4 ρ=0.79
SR|2.4,RO|3.5 ρ=0.96
SR|2.4,Tm|0.3 ρ=0.98
SR|4.8 ρ=0.83
ST|0.6 ρ=0.80
ST|0.6,HN|6 ρ=0.83
ST|0.6,NT|3 ρ=0.75
ST|0.6,RO|3.5 ρ=0.96
ST|0.6,SR|2.4 ρ=0.89
ST|0.6,Tm|0.3 ρ=0.98
ST|1.2 ρ=0.80
Tm|0.3 ρ=0.98
Tm|0.6 ρ=0.98
ZS|0.6 ρ=0.88
ZS|0.6,HN|6 ρ=0.97
ZS|0.6,NT|3 ρ=0.98
ZS|0.6,P6|20 ρ=0.97
ZS|0.6,PLX|60 ρ=0.98
ZS|0.6,RO|3.5 ρ=0.97
ZS|0.6,SR|2.4 ρ=0.98
ZS|0.6,ST|0.6 ρ=0.98
ZS|0.6,Tm|0.3 ρ=0.98
ZS|1.2 ρ=0.94
Prediction
Real data

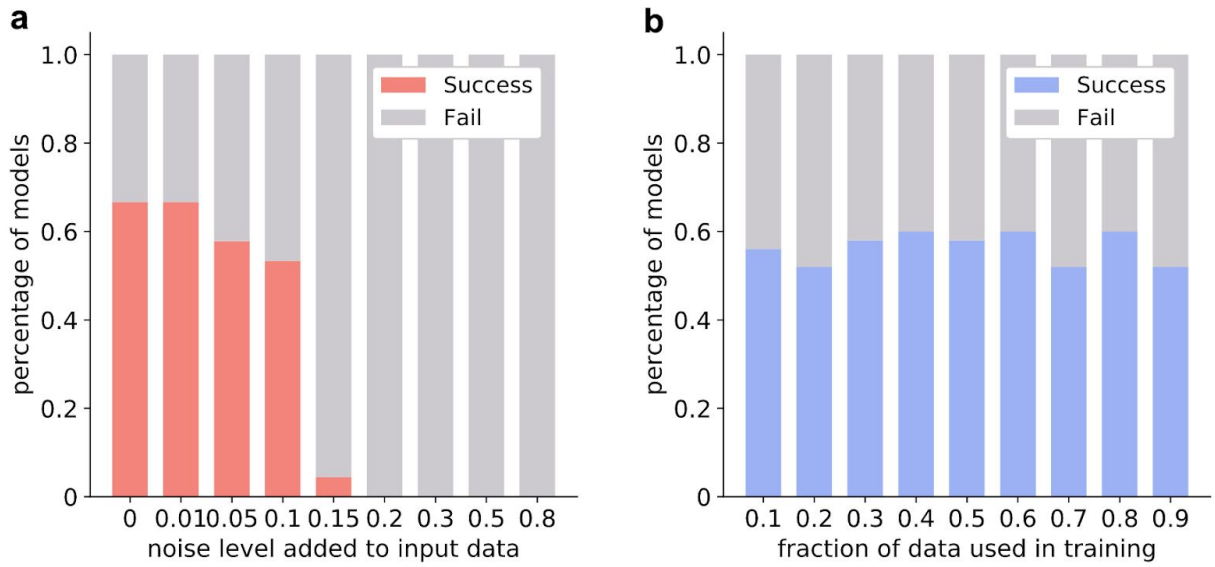


Figure S4. Model convergence against noise and reduced training set

a. The percentage of models that successfully converged, defined by MSE of training set below a threshold of 0.05, decreased as increased level of noise was added into training data. **b.** The percentage of successful models stayed the same as moref data was used for model training.

Supplementary Tables

Table S1. Annotation of nodes in the network.

Table S2. Model-inferred molecular interactions and their pathway database information

References

1. Mansoori, B., Mohammadi, A., Davudian, S., Shirjang, S. & Baradaran, B. The Different Mechanisms of Cancer Drug Resistance: A Brief Review. *Adv Pharm Bull* **7**, 339–348 (2017).
2. Mokhtari, R. B. *et al.* Combination therapy in combating cancer. *Oncotarget* vol. 8 (2017).
3. Garraway, L. A. & Jänne, P. A. Circumventing cancer drug resistance in the era of personalized medicine. *Cancer Discov.* **2**, 214–226 (2012).
4. Fitzgerald, J. B., Schoeberl, B., Nielsen, U. B. & Sorger, P. K. Systems biology and combination therapy in the quest for clinical efficacy. *Nat. Chem. Biol.* **2**, 458–466 (2006).
5. Azmi, A. S., Wang, Z., Philip, P. A., Mohammad, R. M. & Sarkar, F. H. Proof of concept: network and systems biology approaches aid in the discovery of potent anticancer drug combinations. *Mol. Cancer Ther.* **9**, 3137–3144 (2010).
6. Cheng, F., Kovács, I. A. & Barabási, A.-L. Network-based prediction of drug combinations. *Nat. Commun.* **10**, 1197 (2019).
7. Ryall, K. A. & Tan, A. C. Systems biology approaches for advancing the discovery of effective drug combinations. *J. Cheminform.* **7**, 7 (2015).
8. Wrzodek, C., Büchel, F., Ruff, M., Dräger, A. & Zell, A. Precise generation of systems biology models from KEGG pathways. *BMC Syst. Biol.* **7**, 15 (2013).
9. Croft, D. *et al.* The Reactome pathway knowledgebase. *Nucleic Acids Res.* **42**,

D472–7 (2014).

10. Cerami, E. G. *et al.* Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res.* **39**, D685–90 (2011).
11. Lamb, J. *et al.* The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313**, 1929–1935 (2006).
12. Cheung, H. W. *et al.* Systematic investigation of genetic vulnerabilities across cancer cell lines reveals lineage-specific dependencies in ovarian cancer. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 12372–12377 (2011).
13. Tsherniak, A. *et al.* Defining a Cancer Dependency Map. *Cell* **170**, 564–576.e16 (2017).
14. McDonald, E. R., 3rd *et al.* Project DRIVE: A Compendium of Cancer Dependencies and Synthetic Lethal Relationships Uncovered by Large-Scale, Deep RNAi Screening. *Cell* **170**, 577–592.e10 (2017).
15. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–84 (2014).
16. Niepel, M. *et al.* Common and cell-type specific responses to anti-cancer drugs revealed by high throughput transcript profiling. *Nat. Commun.* **8**, 1186 (2017).
17. Dixit, A. *et al.* Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **167**, 1853–1866.e17 (2016).
18. Norman, T. M. *et al.* Exploring genetic interaction manifolds constructed from rich phenotypes. *bioRxiv* 601096 (2019) doi:10.1101/601096.
19. Korkut, A. *et al.* Perturbation biology nominates upstream-downstream drug

combinations in RAF inhibitor resistant melanoma cells. *Elife* **4**, (2015).

20. Hill, S. M. *et al.* Context Specificity in Causal Signaling Networks Revealed by Phosphoprotein Profiling. *Cell Syst* **4**, 73–83.e10 (2017).
21. Vanhaelen, Q., Aliper, A. M. & Zhavoronkov, A. A comparative review of computational methods for pathway perturbation analysis: dynamical and topological perspectives. *Mol. Biosyst.* **13**, 1692–1704 (2017).
22. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).
23. Carter, S. L., Brechbühler, C. M., Griffin, M. & Bond, A. T. Gene co-expression network topology provides a framework for molecular characterization of cellular state. *Bioinformatics* **20**, 2242–2250 (2004).
24. Wang, K. *et al.* Genome-wide identification of post-translational modulators of transcription factor activity in human B cells. *Nat. Biotechnol.* **27**, 829–839 (2009).
25. Babur, O., Demir, E., Gönen, M., Sander, C. & Dogrusoz, U. Discovering modulators of gene expression. *Nucleic Acids Res.* **38**, 5648–5656 (2010).
26. Locasale, J. W. & Wolf-Yadlin, A. Maximum entropy reconstructions of dynamic signaling networks from quantitative proteomics data. *PLoS One* **4**, e6522 (2009).
27. Lezon, T. R., Banavar, J. R., Cieplak, M., Maritan, A. & Fedoroff, N. V. Using the principle of entropy maximization to infer genetic interaction networks from gene expression patterns. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 19033–19038 (2006).
28. Meyer, P. E., Lafitte, F. & Bontempi, G. minet: A R/Bioconductor package for

inferring large transcriptional networks using mutual information. *BMC Bioinformatics* **9**, 461 (2008).

29. Şenbabaoğlu, Y. *et al.* A Multi-Method Approach for Proteomic Network Inference in 11 Human Cancers. *PLoS Comput. Biol.* **12**, e1004765 (2016).
30. Yi, S. *et al.* Functional variomics and network perturbation: connecting genotype to phenotype in cancer. *Nat. Rev. Genet.* **18**, 395–410 (2017).
31. D'haeseleer, P., Liang, S. & Somogyi, R. Genetic network inference: from co-expression clustering to reverse engineering. *Bioinformatics* **16**, 707–726 (2000).
32. Aldridge, B. B., Saez-Rodriguez, J., Muhlich, J. L., Sorger, P. K. & Lauffenburger, D. A. Fuzzy logic analysis of kinase pathway crosstalk in TNF/EGF/insulin-induced signaling. *PLoS Comput. Biol.* **5**, e1000340 (2009).
33. Zou, M. & Conzen, S. D. A new dynamic Bayesian network (DBN) approach for identifying gene regulatory networks from time course microarray data. *Bioinformatics* **21**, 71–79 (2005).
34. Gardner, T. S., di Bernardo, D., Lorenz, D. & Collins, J. J. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science* **301**, 102–105 (2003).
35. Nyman, E. *et al.* Perturbation biology links temporal protein changes to drug responses in a melanoma cell line. doi:10.1101/568758.
36. Klinger, B. *et al.* Network quantification of EGFR signaling unveils potential for targeted combination therapy. *Mol. Syst. Biol.* **9**, 673 (2013).

37. Fröhlich, F. *et al.* Efficient Parameter Estimation Enables the Prediction of Drug Response Using a Mechanistic Pan-Cancer Pathway Model. *Cell Syst* **7**, 567–579.e6 (2018).

38. Nelander, S. *et al.* Models from experiments: combinatorial drug perturbations of cancer cells. *Mol. Syst. Biol.* **4**, 216 (2008).

39. Hug, S. *et al.* High-dimensional Bayesian parameter estimation: case study for a model of JAK2/STAT5 signaling. *Math. Biosci.* **246**, 293–304 (2013).

40. Bruggeman, F. J., Westerhoff, H. V., Hoek, J. B. & Kholodenko, B. N. Modular response analysis of cellular regulatory networks. *J. Theor. Biol.* **218**, 507–520 (2002).

41. Esteva, A. *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **542**, 115–118 (2017).

42. Hou, L. *et al.* Patch-based Convolutional Neural Network for Whole Slide Tissue Image Classification. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* **2016**, 2424–2433 (2016).

43. Zhou, J. & Troyanskaya, O. G. Predicting effects of noncoding variants with deep learning-based sequence model. *Nat. Methods* **12**, 931–934 (2015).

44. Montavon, G., Samek, W. & Müller, K.-R. Methods for interpreting and understanding deep neural networks. *Digital Signal Processing* vol. 73 1–15 (2018).

45. Süli, E. & Mayers, D. F. *An Introduction to Numerical Analysis*. (Cambridge University Press, 2003).

46. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. (2014).

47. Chandarlapaty, S. *et al.* AKT inhibition relieves feedback suppression of receptor tyrosine kinase expression and activity. *Cancer Cell* **19**, 58–71 (2011).

48. Osaki, L. H. & Gama, P. MAPK signaling pathway regulates p27 phosphorylation at threonin 187 as part of the mechanism triggered by early-weaning to induce cell proliferation in rat gastric mucosa. *PLoS One* **8**, e66651 (2013).

49. Donovan, J. C., Milic, A. & Slingerland, J. M. Constitutive MEK/MAPK activation leads to p27(Kip1) deregulation and antiestrogen resistance in human breast cancer cells. *J. Biol. Chem.* **276**, 40888–40895 (2001).

50. Carreira, S. *et al.* Mitf cooperates with Rb1 and activates p21Cip1 expression to regulate cell cycle progression. *Nature* **433**, 764–769 (2005).

51. Lei, W., Liu, F. & Ness, S. A. Positive and negative regulation of c-Myb by cyclin D1, cyclin-dependent kinases, and p27 Kip1. *Blood* **105**, 3855–3861 (2005).

52. Butch, E. R. & Guan, K. L. Characterization of ERK1 activation site mutants and the effect on recognition by MEK1 and MEK2. *J. Biol. Chem.* **271**, 4230–4235 (1996).

53. Aoki, K., Yamada, M., Kunida, K., Yasuda, S. & Matsuda, M. Processive phosphorylation of ERK MAP kinase in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 12675–12680 (2011).

54. Gupta, S. & Davis, R. J. MAP kinase binds to the NH2-terminal activation domain of c-Myc. *FEBS Lett.* **353**, 281–285 (1994).

55. Sears, R. Multiple Ras-dependent phosphorylation pathways regulate Myc protein stability. *Genes & Development* vol. 14 2501–2514 (2000).

56. von Mering, C. *et al.* STRING: known and predicted protein-protein associations,

integrated and transferred across organisms. *Nucleic Acids Res.* **33**, D433–7 (2005).

57. Maglott, D., Ostell, J., Pruitt, K. D. & Tatusova, T. Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.* **39**, D52–7 (2011).
58. Lipton, Z. C. The mythos of model interpretability. *Communications of the ACM* vol. 61 36–43 (2018).
59. Gunning, D. & Aha, D. DARPA's Explainable Artificial Intelligence (XAI) Program. *AI Magazine* vol. 40 44–58 (2019).
60. Thelisson, E. Towards Trust, Transparency and Liability in AI / AS systems. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence* (2017) doi:10.24963/ijcai.2017/767.
61. Baydin, A. G., Pearlmutter, B. A., Radul, A. A. & Siskind, J. M. Automatic Differentiation in Machine Learning: a Survey. *J. Mach. Learn. Res.* **18**, 1–43 (2018).
62. Brown, R., Curry, E., Magnani, L., Wilhelm-Benartzi, C. S. & Borley, J. Poised epigenetic states and acquired drug resistance in cancer. *Nat. Rev. Cancer* **14**, 747–753 (2014).
63. Zaal, E. A. & Berkers, C. R. The Influence of Metabolism on Drug Response in Cancer. *Frontiers in Oncology* vol. 8 (2018).
64. Adamson, B. *et al.* A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell* **167**, 1867–1882.e21 (2016).

65. Mimitou, E. P. *et al.* Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. *Nat. Methods* **16**, 409–412 (2019).

66. Wroblewska, A. *et al.* Protein Barcodes Enable High-Dimensional Single-Cell CRISPR Screens. *Cell* **175**, 1141–1155.e16 (2018).

67. Frei, A. P. *et al.* Highly multiplexed simultaneous detection of RNAs and proteins in single cells. *Nat. Methods* **13**, 269–275 (2016).

68. Bengio, Y. Practical Recommendations for Gradient-Based Training of Deep Architectures. *Lecture Notes in Computer Science* 437–478 (2012) doi:10.1007/978-3-642-35289-8_26.

69. Liang, J. & Liu, R. Stacked denoising autoencoder and dropout together to prevent overfitting in deep neural network. *2015 8th International Congress on Image and Signal Processing (CISP)* (2015) doi:10.1109/cisp.2015.7407967.

70. Fröhlich, F., Kaltenbacher, B., Theis, F. J. & Hasenauer, J. Scalable Parameter Estimation for Genome-Scale Biochemical Reaction Networks. *PLoS Comput. Biol.* **13**, e1005331 (2017).

71. Molinelli, E. J. *et al.* Perturbation biology: inferring signaling networks in cellular systems. *PLoS Comput. Biol.* **9**, e1003290 (2013).

72. Luna, A., Babur, Ö., Aksoy, B. A., Demir, E. & Sander, C. PaxtoolsR: pathway analysis in R using Pathway Commons. *Bioinformatics* vol. 32 1262–1264 (2016).