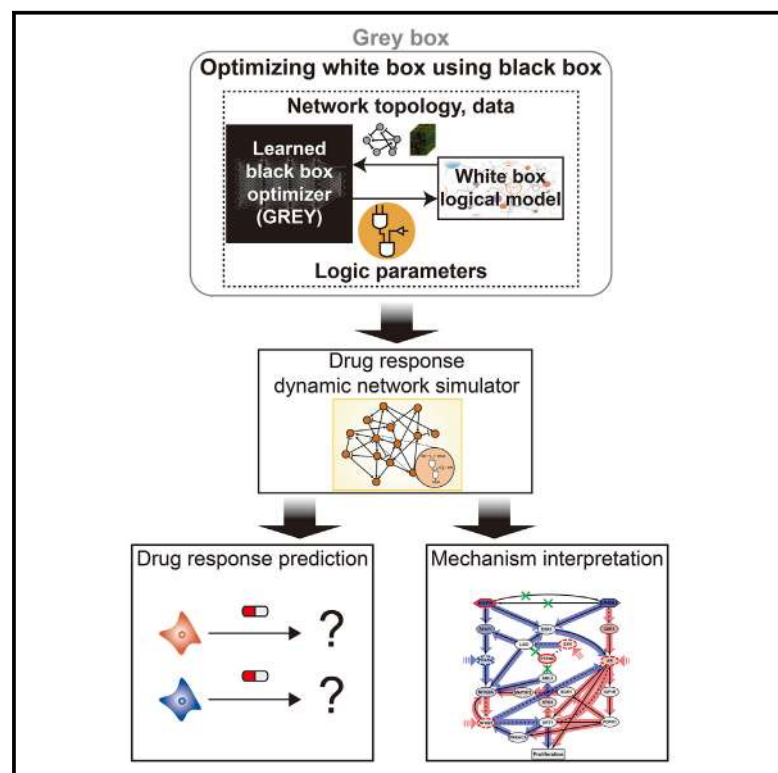


# A gray box framework that optimizes a white box logical model using a black box optimizer for simulating cellular responses to perturbations

## Graphical abstract



## Authors

Yunseong Kim, Younghyun Han, Corbin Hopper, ..., Junsoo Kang, Taeyoung Kim, Kwang-Hyun Cho

## Correspondence

ckh@kaist.ac.kr

## In brief

Kim et al. present a gray box framework that combines a white box logical model with a black box optimizer, addressing challenges in interpreting molecular regulatory dynamics. The gray box framework successfully predicts anti-cancer drug responses of cancer cells, while shedding light on the underlying molecular regulatory mechanisms.

## Highlights

- We propose a gray box framework with Boolean networks and a black box optimizer
- GREY is a learned optimizer specialized in Boolean network optimization
- Gray box framework can predict drug responses and reveal underlying mechanisms

## Article

# A gray box framework that optimizes a white box logical model using a black box optimizer for simulating cellular responses to perturbations

Yunseong Kim,<sup>1,3</sup> Younghyun Han,<sup>1,2,3</sup> Corbin Hopper,<sup>1</sup> Jonghoon Lee,<sup>1</sup> Jae Il Joo,<sup>1,2</sup> Jeong-Ryeol Gong,<sup>1</sup> Chun-Kyung Lee,<sup>1</sup> Seong-Hoon Jang,<sup>1</sup> Junsoo Kang,<sup>1</sup> Taeyoung Kim,<sup>1</sup> and Kwang-Hyun Cho<sup>1,4,\*</sup>

<sup>1</sup>Laboratory for Systems Biology and Bio-inspired Engineering, Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Korea

<sup>2</sup>Present address: biorevert, Inc., 193, Munji-ro, Yuseong-gu, Daejeon 34051, Korea

<sup>3</sup>These authors contributed equally

<sup>4</sup>Lead contact

\*Correspondence: [ckh@kaist.ac.kr](mailto:ckh@kaist.ac.kr)

<https://doi.org/10.1016/j.crmeth.2024.100773>

**MOTIVATION** The challenge of predicting cellular responses to perturbations amid the complex non-linearities of molecular interactions has spurred the development of machine learning-based models. However, interpreting these models in terms of molecular regulatory dynamics remains difficult. Conversely, logical network models like Boolean networks offer interpretability but struggle with large-scale networks due to high-dimensional search spaces. To overcome these hurdles, we introduce a scalable derivative-free optimizer, trained via meta-reinforcement learning, for Boolean network models. This approach enables prediction of anti-cancer drug responses in cancer cell lines while offering valuable insights into their molecular regulatory mechanisms.

## SUMMARY

Predicting cellular responses to perturbations requires interpretable insights into molecular regulatory dynamics to perform reliable cell fate control, despite the confounding non-linearity of the underlying interactions. There is a growing interest in developing machine learning-based perturbation response prediction models to handle the non-linearity of perturbation data, but their interpretation in terms of molecular regulatory dynamics remains a challenge. Alternatively, for meaningful biological interpretation, logical network models such as Boolean networks are widely used in systems biology to represent intracellular molecular regulation. However, determining the appropriate regulatory logic of large-scale networks remains an obstacle due to the high-dimensional and discontinuous search space. To tackle these challenges, we present a scalable derivative-free optimizer trained by meta-reinforcement learning for Boolean network models. The logical network model optimized by the trained optimizer successfully predicts anti-cancer drug responses of cancer cell lines, while simultaneously providing insight into their underlying molecular regulatory mechanisms.

## INTRODUCTION

Controlling cell fate with molecular regulatory perturbations is an important challenge in various areas of biology. However, the control effect of a perturbation on specific molecules is often counterintuitive due to the complexity of the cellular network arising from non-linear molecular regulatory interactions.<sup>1</sup> Molecular regulatory scale network structures, such as feedback loops and crosstalk, cause perturbation responses of the network to be highly dynamic processes. Therefore, predicting cellular response to a perturbation from static information alone, such as mutation profiles incorporated into the cellular network,

obscures the underlying mechanism.<sup>2</sup> In contrast, dynamic information regarding network state changes over time after a perturbation reveals the key regulatory mechanisms of the network response and aids the precision of response predictions. In this view, dynamic models of molecular regulation capable of simulating cellular responses to perturbations provide critical insight.<sup>3–5</sup>

Various machine learning (ML) models, including deep learning, have shown high accuracy in predicting non-linear dynamic processes of cellular networks, such as responses to molecular regulatory perturbations.<sup>6–10</sup> However, ML-based models cannot provide biological rationale for their predictions

since interpreting the internal workings of the model is difficult. To rectify this “black box” nature of ML models, some interpretable methods such as random forest and linear regression have been applied to predict cellular responses to perturbations,<sup>11–16</sup> which provide biological features highly correlated to the predictions. More recently, explainable artificial intelligence<sup>17,18</sup> and deep neural networks with architectures that mimic the relationships of biological entities, such as hierarchies of biological pathways,<sup>19</sup> subsystems,<sup>20,21</sup> or molecular interactions,<sup>22</sup> have been utilized. These methods suggest biological features related to the decision of the models and identify their relationships. However, due to the lack of dynamic information provided by these suggestions from the ML-based models, a dynamic model of molecular regulation is needed to simulate and predict perturbation responses.

In systems biology, a cellular network is considered as a non-linear dynamical system of interacting cellular components, such as genes and transcription factors.<sup>23,24</sup> Thus, the network can be abstracted to a system model containing molecular components and their regulatory parameters represented by nodes and links. Continuous dynamic models, such as ordinary differential equation network models, can be constructed by integrating multiple sources of biological data and *a priori* knowledge, and therefore can successfully recapitulate complex dynamic processes such as cellular responses to perturbations.<sup>25,26</sup> Reconstructing and simulating a dynamic network model reproduces molecular regulatory mechanisms, which not only predicts the perturbation responses but also enables interpretation of molecular regulation dynamics such as perturbation signal flows related with those responses. This approach is called “white box” modeling due to the interpretable nature of the predicted responses driven by the system. Despite this, estimating every kinetic parameter of the continuous dynamic model is only feasible for small and well-investigated networks. As the network size increases, accurately estimating parameters becomes more difficult due to the increase in the dimensionality of the search space.

Discrete logic-based models, such as Boolean network models, can partially overcome these problems by abstracting the network dynamics into simple but essential regulatory functions such as logical equations. These functions can be represented by network parameters or edge weights. The influence of neighboring nodes on a specific node in the network can be determined by considering the states of these neighboring nodes and the weights of the directed edges connecting them to the specific node. This influence is quantified by a weighted sum, calculated by multiplying the state value of each neighboring node by the corresponding weight of the edge connecting it to the specific node. If this weighted sum is greater than 0, the specific node becomes activated, influenced by the states of its neighboring nodes. Conversely, if the sum is less than 0, the node is deactivated, again due to the influence of the states of its neighboring nodes. Several studies have been made to estimate regulatory functions of these discrete models based on information theory<sup>27–29</sup> or meta-heuristic optimization<sup>30,31</sup> due to the gradient-free characteristics of discrete model optimization, but such approaches still have scalability problems with high computational complexity. Moreover, the growing demand for discrete dynamic models of *in silico* drug discovery for precision

medicine requires more efficient optimization algorithms, due to the complexity of estimating network parameters for each unique *in silico* model that predicts diverse drug response dynamics of patient samples. Therefore, a more sophisticated optimization algorithm is needed to efficiently estimate regulatory functions and scale up to larger networks.

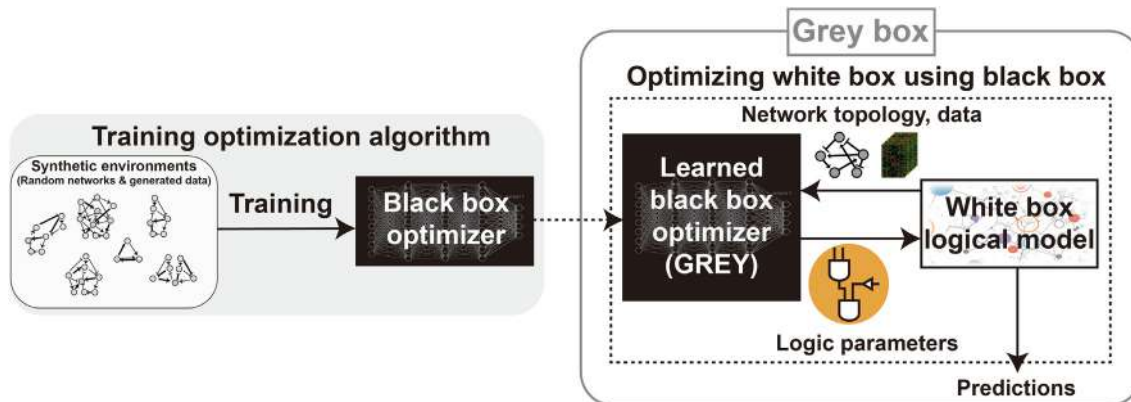
The learning to optimize (L2O)<sup>32</sup> framework is a methodology that utilizes ML to train an optimization algorithm. Whereas traditional ML trains a particular model, L2O trains an optimizer that can be subsequently applied to optimize several related models. By using L2O to train a specialized optimizer for a specific type of optimization problem, it has the potential to outperform classical optimization algorithms in terms of computational complexity.<sup>33,34</sup> This is because the trained optimizer tries to learn the characteristics of the search space for the class of problems at hand, which enables it to efficiently explore the space of a particular one of those problems. In this study, we present a “gray box framework” that optimizes discrete dynamic models of molecular regulatory networks using an optimizer trained for discrete models with the L2O approach. This gray box framework combines the interpretability of a white box model with the scalability of a black box optimizer, allowing it to be both interpretable and able to scale up to larger networks (Figure 1). We trained the black box optimizer, which we call GREY for gradient-free optimizer trained by meta-reinforcement learning to yield edge weights of Boolean networks, using synthetic optimization problems that imitate white box models consisting of Boolean networks.

GREY can learn a derivative-free exploration strategy to find an optimal solution in the high-dimensional discontinuous search space of discrete dynamical models such as Boolean networks. To validate the optimization ability of GREY for real biological perturbation data, we applied it to drug response prediction. We constructed a drug response core network (DRCN) structure for each individual drug extracted from a pan-cancer core network (PCCN) integrated from public databases. The DRCN optimized by GREY can successfully predict the drug responses of cancer cell lines by utilizing the area under the curve (AUC) values derived from the viability curve data in the database Genomics of Drug Sensitivity in Cancer 2 (GDSC2): <https://www.cancerrxgene.org/>.<sup>35</sup> GREY was able to optimize the logical parameters of the DRCN more efficiently than a state-of-the-art derivative-free optimization algorithm. Further mechanism analysis showed that the optimized network reveals key network motifs related with perturbation response dynamics, which were utilized to infer drug response biomarkers that match basket trial marker genes reported from previous experiments.

## RESULTS

### A gray box framework that optimizes a white box logical model using a black box optimizer

We developed a gray box framework by combining a black box optimizer with a white box modeling approach that generates biologically interpretable predictions. The framework constructs a Boolean network to model drug response at the cellular level, and optimizes the edge weights of the network to match existing biological data, such as cellular responses to drug perturbation.



**Figure 1. A schematic representation of the grey box framework**

In the gray box framework, the white box model provides topological information of intracellular networks to the black box model, and the black box model optimizes the edge weights of intracellular networks from biological data. More specifically, the black box optimizer is trained on synthetic environments generated by random networks with perturbation fold change simulations. By blending these modeling frameworks, the gray box model can predict precise drug responses that are biologically interpretable.

The gray box framework can be broken into three parts (Figure 2): (1) an extraction module for the DRCN structure, (2) a preprocessing module for the cell line profiles and drug response data, and (3) a gray box module that optimizes the logical relationships within the DRCN.

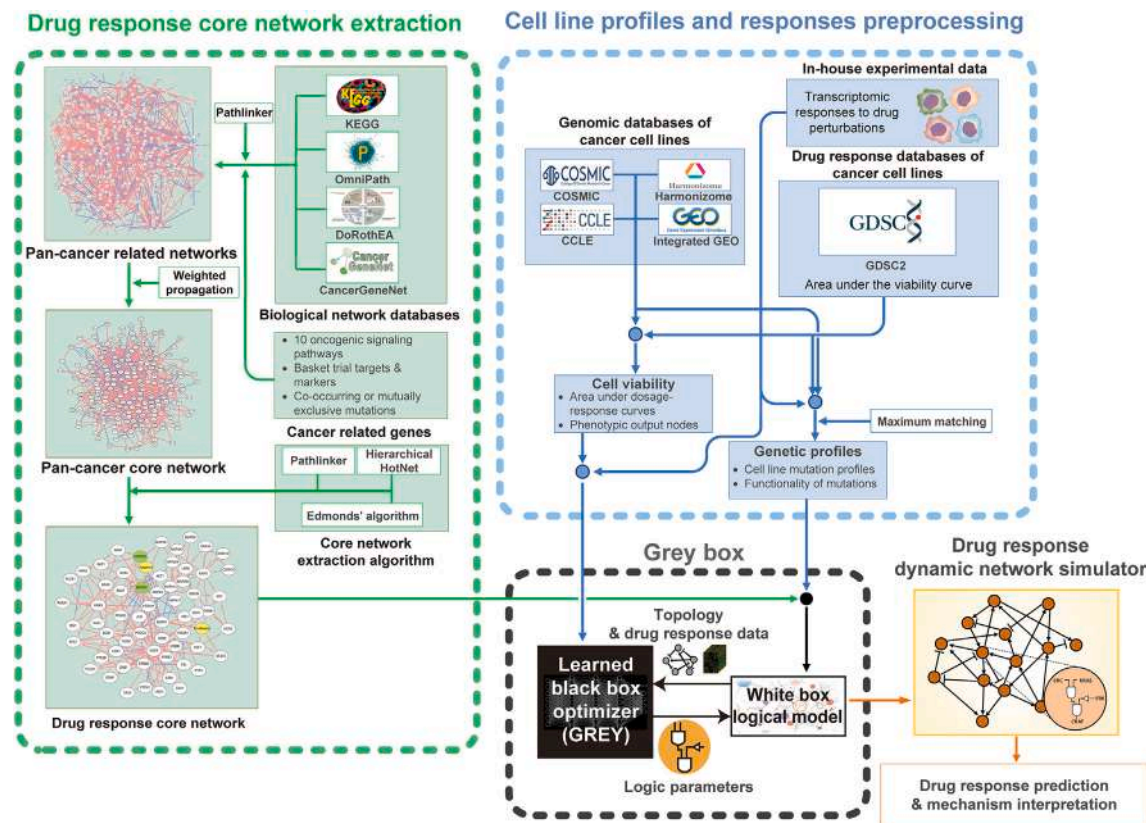
The DRCN is constructed to distill existing knowledge of regulatory interactions between genes and proteins related to drug responses in cancer into a minimal network. It begins with a pan-cancer related network that isolates genes contained in cancer-related biological networks, frequently mutated genes in cancer, and target genes of anti-cancer drugs. The links from a structural network database of intracellular signaling among genes and proteins, Omnipath,<sup>36</sup> that have both a direction (identified source and target nodes) and a sign (whether the source node activates or inhibits the target node) are chosen to connect the selected genes into the pan-cancer related network. The PCCN was then extracted by removing irrelevant inter-nodes in the pan-cancer related network that are not contained in a path between cancer-related genes. Next, we extracted the DRCN from the constructed PCCN. The DRCN is required to keep the network size as small as possible to minimize the computational requirements of the Boolean network simulations, while retaining all drug response-related genes and signaling pathways. To accomplish these requirements, a minimum number of genes were selected that were significantly related to the drug response of cancer cell lines in GDSC2. Then, a minimum number of links and intermediate nodes were selected by Edmonds' algorithm, and utilized to connect the selected genes to the DRCN. Finally, phenotypic "proliferation" and "apoptosis" nodes, and their links from the CancerGeneNet,<sup>37</sup> were added to the extracted DRCN as phenotypic markers. As a result, the constructed DRCNs incorporate both drug response-related genes and proteins as nodes, recapitulating the fundamental molecular regulatory dynamics of the selected drug responses (see STAR Methods).

The ability for the proposed gray box framework to predict drug responses was evaluated across three different DRCNs

to ensure robustness across a variety of network structures. We constructed three different DRCNs of US Food and Drug Administration-approved targeted cancer drugs from Molecular Analysis for Therapy Choice (MATCH) basket trials run by the National Cancer Institute (NCI-MATCH).<sup>38</sup> We selected afatinib (an inhibitor of *EGFR* and *ERBB2*), trametinib (an inhibitor of *MAP2K1* and *MAP2K2*, also known as *MEK1* and *MEK2*, respectively), and palbociclib (an inhibitor of *CDK4* and *CDK6*). Each of these drugs inhibits well-known and unique cancer-related machinery, and together they engage a large fraction of the nodes covering the PCCN structure (see STAR Methods). The afatinib DRCN comprised 57 nodes, whereas the trametinib and palbociclib DRCNs comprised 66 nodes and 192 nodes, respectively.

While the constructed three different DRCNs that GREY optimizes recapitulate essential regulatory network structures related with drug responses, each drug response of each individual cancer cell line can be simulated by applying the mutational profiles of each cell line from the GDSC2 database to the DRCN of each of the three drugs. To apply mutations from a cell line to the DRCN of a given drug, the corresponding nodes in the DRCN whose mutations correspond to the mutation profiles of that cell line in GDSC2, are set to a specific invariant state. The invariant node state is set to 1 if the mutation is either activation, amplification, or gain of function and set to  $-1$  if the mutation is suppression, deletion, or loss of function. Mutations that are in nodes outside of the DRCN are treated as a constant input signal that activates or inactivates any downstream nodes in the DRCN. The mutation profiles of each cell line were obtained from the Catalogue of Somatic Mutations in Cancer (COSMIC)<sup>39</sup> database and the effect of each mutation was determined from their oncogenic or tumor suppressive  $q$  values (see STAR Methods).<sup>40</sup> In this way, we mapped numerous mutation profiles of cell lines to a small number of nodes in our DRCNs. For example, 716 cell lines were represented identically by the afatinib DRCN composed of 57 nodes. Having incorporated mutation profiles and drug responses, the edge weights of the resulting DRCNs are then ready to be optimized by GREY.





**Figure 2. Overall procedure for constructing the gray box framework**

The extraction module (green dotted box) constructs the pan-cancer core network (PCCN) based on pan-cancer related networks gathered from cancer-related genes and network topology data and extracts drug response core networks (DRCNs). The preprocessing module (blue dotted box) gathers mutation profiles and drug response data of cell lines, and projects them onto the DRCN, generating DRCNs specific to each drug. The gray box module (black dotted box) optimizes edge weights using GREY based on the DRCN and preprocessed data. The ensemble of optimized DRCNs is utilized by the drug response dynamic network simulator to predict drug responses and generate mechanistic interpretations.

### Construction and optimization procedure of GREY

We developed GREY to optimize edge weights of DRCNs for the mutation profiles and the drug response data from GDSC2. We formalized DRCNs using weighted sum Boolean logic (see [STAR Methods](#)), and the edge weights of the DRCN were assigned to each link. In weighted sum Boolean logic, the activation of each node is determined by a weighted sum of values from its incoming connections, where the weights represent the regulatory influence of each link. Due to the non-continuous fitness space that results from the discrete states of the nodes, and the time-consuming attractor simulation processes of Boolean networks, we needed a derivative-free and sample-efficient optimizer. Thus, we trained GREY, which is an optimizer specialized for optimizing the edge weights of Boolean networks, through L2O, which employs multi-agent meta-reinforcement learning<sup>41–43</sup> for its implementation. An agent in reinforcement learning is an entity that learns to make optimal decisions by interacting with an environment and receiving feedback in the form of rewards. In multi-agent reinforcement learning, multiple agents must learn how to cooperate with each other to achieve their collective objectives.

In the current setting, the environment is an optimization problem (optimizing the edge weights of Boolean networks), and there is one agent for each edge weight. The agents explore the search space of the optimization problem (all possible edge weight values) to find good solutions that have high fitness. The edge weight values of each solution are utilized to simulate a Boolean network. Fitness is then defined as the Pearson's correlation coefficient between the simulated fold change of the node values and the ground truth extracted from a database or biological experiment. The reward for agents is proportional to the difference between its current fitness and its previous best fitness, and the agents are trained to maximize reward. In traditional reinforcement learning, an agent learns an exploration strategy to maximize its reward within a specific environment. For our problem, each drug corresponds to a single DRCN structure. These different DRCNs represent different environments for the optimizer. To accommodate multiple environments in the form of multiple different DRCN structures, we adopted meta-reinforcement learning, which allows agents to learn adaptable exploration strategies by training across a variety of different environments.

In a given environment, GREY receives the network structure of the environment with edge weights and reward as input, and produces new edge weights as output. Then, fitness is computed using these new edge weights, followed by a reward. The entire sequence from entering input to GREY to obtaining a reward constitutes a single step. The new edge weights and the reward obtained from the preceding step are used as input for the subsequent step. In the initial step, edge weights and reward used as input are initialized to random values and zero, respectively. An episode is a sequence of interactions between GREY and an environment, i.e., a sequence of steps, and GREY is trained from the cumulative experiences gained throughout the episode.

GREY was trained using 100,000 synthetic environments with random networks composed of 20–40 nodes, random edge weights, and a generated ground truth dataset (see [STAR Methods](#); [Figures 3A](#) and [S1](#)). The random network was formalized by weighted sum Boolean logic, and the ground truth dataset comprises the fold change of the average activity for nodes between the first attractors attained from random initial states and the second attractors reached from the first attractor after fixing some nodes to 1 or  $-1$ . The resulting GREY trained on randomly generated synthetic environments can be applied to optimize the edge weights of biological networks with any given structure and corresponding data ([Figure 3B](#)).

The architecture of GREY consists of three layers: the input, the submodule, and the main layer ([Figure S1A](#)). The input and the submodule layers serve an auxiliary role of providing information to the main layer. The input layer is composed of several graph neural networks (GNNs),<sup>44</sup> which are a type of neural network designed to process and learn from data structured as graphs by capturing relationships between nodes. GNNs in the input layer produce node, edge, and global (graph-level) features by aggregating information from neighboring nodes, edges, and the entire graph, and they represent underlying dependencies within the graph ([Figure 3C](#)). Those features are then reorganized for each edge by concatenating the edge feature with the node features of sender and receiver nodes, as well as the global features ([Figure 3D](#)). These reorganized features are then passed to the submodule and the main layers and processed by long short-term memory (LSTM) ([Figure 3E](#)). In the submodule and the main layers, one LSTM is assigned to each edge, and LSTMs in the main layer each act as one agent. Each agent produces a new weight for its corresponding edge, and the new weights from all agents collectively determine fitness, since fitness depends on the dynamics of the network as a whole. The submodule layer has a context estimator, which helps the agent adapt quickly to the environment by estimating information about the environment, and a global state estimator, which helps the agent assigned to each edge to cooperate by estimating the behavior of the other agents.

### Application of GREY to synthetic and biological environments

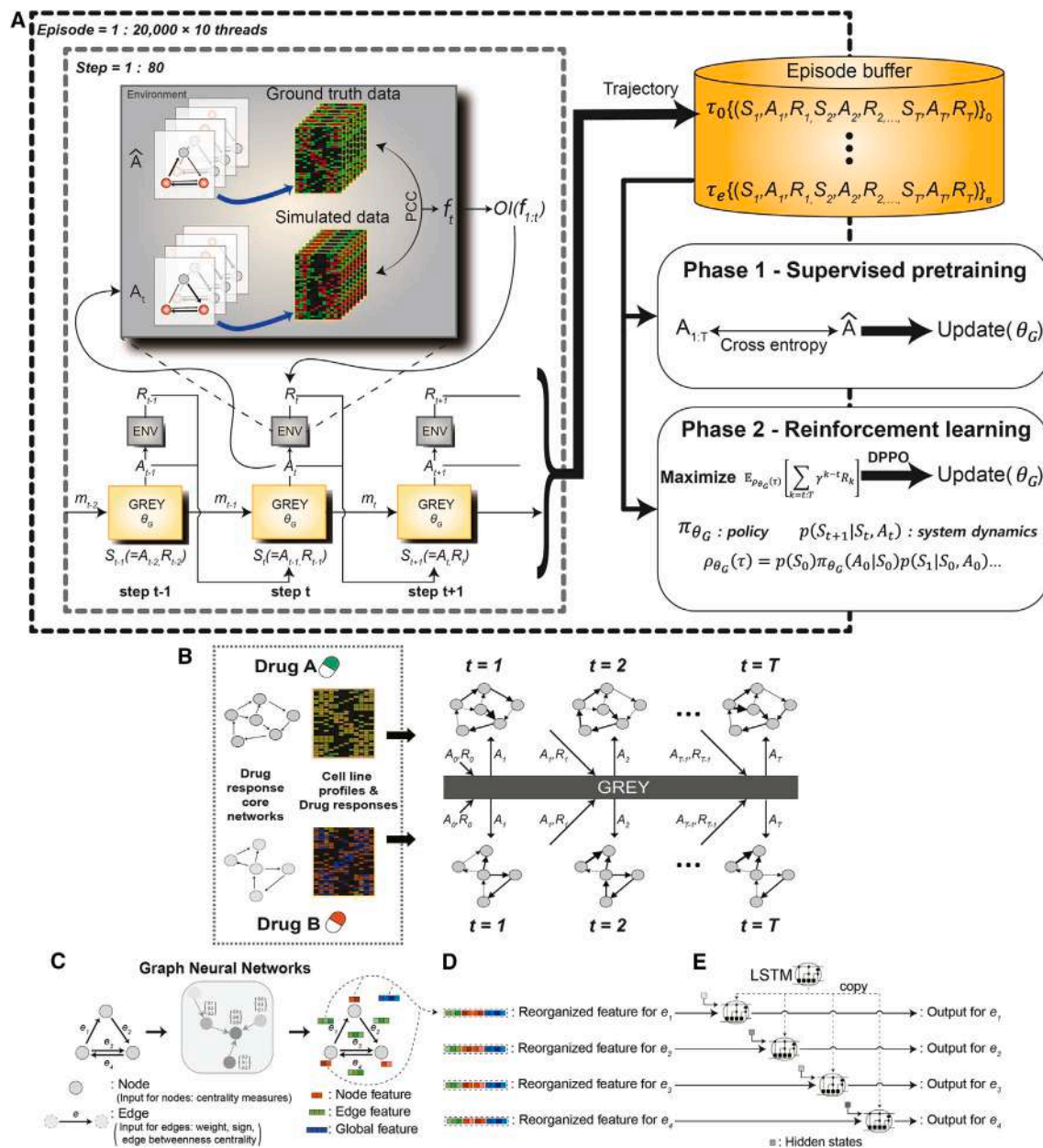
We assessed the optimization performance of GREY against unseen environments. In each application, GREY produces five new candidate edge weight sets for each step (i.e., generating

five outputs at each step) and randomly replaces elements (e.g., 50%) from each candidate edge weight set by the edge weights with the best fitness from all previous steps. Then, we evaluated the fitness of each candidate edge weight set (i.e., five evaluations per step) and used the candidate with the best fitness as an input for the next step. First, we checked that GREY can perform optimization on an example synthetic environment not used in training. As more evaluations were performed, GREY found edge weights with higher fitness ([Figure 4A](#)). Random agents, which are untrained GREY, showed no meaningful behavior. Thus, we confirmed that GREY can perform optimization properly.

Next, we compared the optimization performance of GREY to covariance matrix adaptation evolution strategy (CMA-ES),<sup>45</sup> which is considered the state-of-the-art derivative-free optimization method for challenging settings, such as non-convex and non-continuous problems. During each optimization step, evaluating fitness incurs the greatest computational cost for both GREY and CMA-ES, because of the time-consuming attractor simulation processes of Boolean networks. Thus, we compared the performance in terms of sample efficiency, aiming for higher fitness with fewer evaluations. GREY achieved statistically significant higher fitness for all scales of networks (20–200 links with 20 networks for each size interval) that were not used during training ([Figure 4B](#)). Furthermore, GREY outperformed CMA-ES when the number of evaluations per iteration step was the same ([Figure 4C](#)).

We then applied GREY to optimizing the DRCNs of three drugs, afatinib, trametinib, and palbociclib, with respect to their drug response data from GDSC2 ([Figure 4D](#)). When optimizing the edge weights of the DRCN for a certain drug, the fitness of the solution is the accuracy of DRCN drug-response predictions. More specifically, the DRCN predicts drug responses by calculating the difference between the average activity of phenotype nodes between the prior attractor and the new attractor. The prior attractor is obtained from random initial states with the given mutation profile for each cell line, and the new attractor is obtained from the prior attractor while maintaining drug target nodes at  $-1$ . The fitness is then defined as the Pearson's correlation coefficient between the predicted drug responses from DRCN and the ground truth, as before. However, unlike with the simulated datasets, the ground truth drug response of each node is not known. Instead, only the activity change of the phenotype nodes are measured and compared with the ground truth to determine fitness. The ground truth of the drug response is the measured area under the viability curve (AUC) value collected from GDSC2 (see [STAR Methods](#)).

Compared with CMA-ES, GREY consistently achieved better fitness when optimizing DRCNs for all three drugs. For the trametinib DRCN, GREY reached a fitness of 0.3 after an average of 403 evaluations, which is approximately 3 times more sample-efficient than an average of 1,102 evaluations of CMA-ES. GREY successfully optimized the palbociclib DRCN consisting of 655 edge weights, despite the search space size being significantly larger than that of the synthetic environments used in the training of GREY. These results suggest that GREY is sample efficient and highly scalable.



**Figure 3. Overall architecture of GREY and training procedure**

(A) An illustration of the training procedure.  $\hat{A}$  is the true edge weights of the environment, and  $A_t$  is the new edge weights at step  $t$ .  $f_t$  is the fitness at step  $t$ , and  $R_t$  is the observed improvement (OI) of the fitness (PCC, Pearson's correlation coefficient) as a reward at step  $t$ .  $m_t$  includes hidden states. After generating an episode by repeating 80 steps, the trajectory is stored in the buffer. Neural network parameters of GREY ( $\theta_G$ ) are updated whenever 64 episodes are saved, and then the buffer is cleared. Supervised pre-training to predict the true edge weights is performed until the initial 8,000 episodes are used and, from then on, reinforcement learning is performed with the DPPO algorithm to maximize the cumulative discounted rewards.

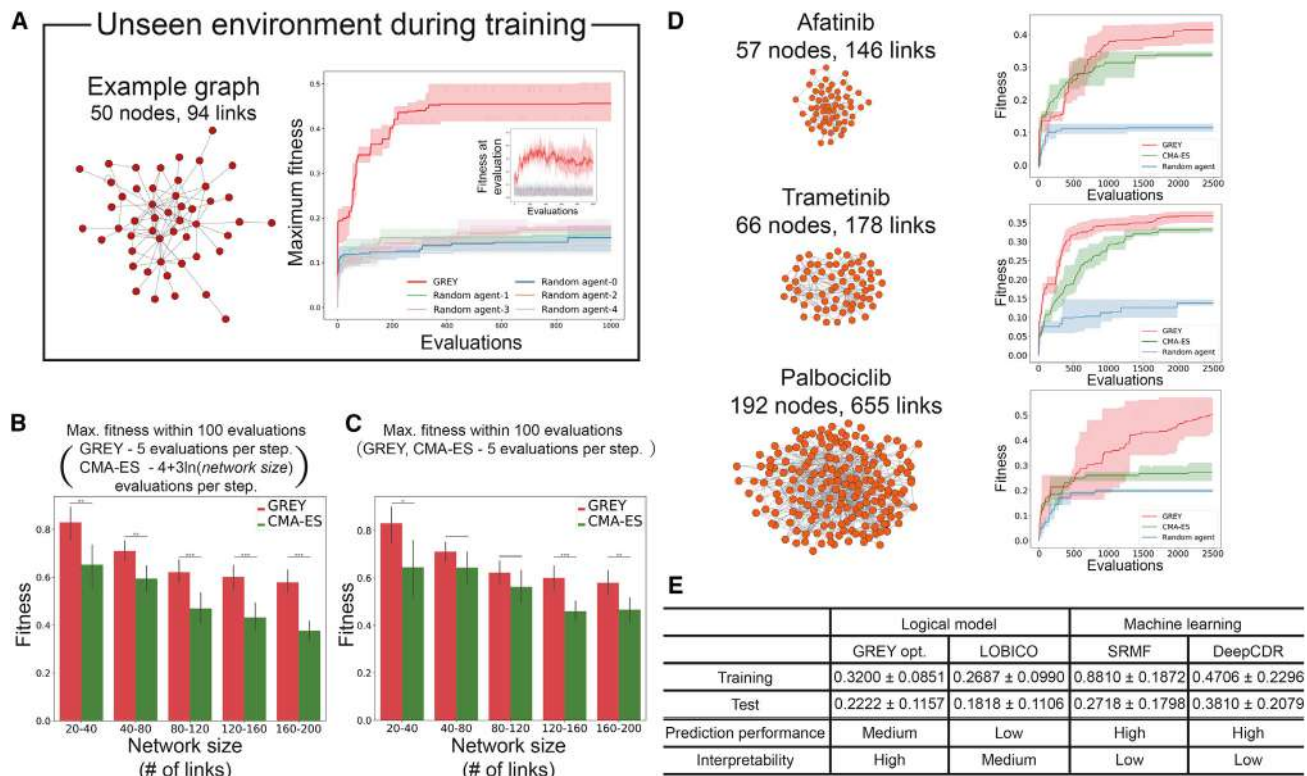
(B) Trained GREY can optimize the edge weights of the white box Boolean network model regardless of network structure. During each step of optimization, trained GREY iteratively receives network structure with preceding edge weights and its reward as input, and then attempts to find new edge weights with better fitness and returns those as output.

(C) Inputs are processed using graph neural networks and node, edge, and global features are generated.

(D) For each edge, a reorganized feature is formed by concatenating its edge feature, the node features of sender and receiver nodes, and the global feature.

(E) LSTMs are assigned to each edge. While the parameters of these LSTMs are shared, they maintain independent hidden states.





**Figure 4. Performance of GREY on synthetic environments and DRCNs**

For synthetic environments and DRCNs, we used 128 and 1,000 random initial states, respectively. Initial states are sampled for every step during optimization. During each step, 5 sets of edge weights are generated and evaluated, and the set of edge weights with the highest fitness becomes an input for the next step. (A) GREY optimization for a synthetic environment (50 nodes, 94 links) that was not used in training. The larger line plot shows the best fitness (y axis) across evaluations (x axis), while the inset plot shows fitness at each evaluation. Results of five random agents are shown for comparison. Shaded areas have 95% confidence intervals over three independent runs.

(B and C) Comparison of the results of CMA-ES on various scales. Initial standard deviation ( $\sigma_0$ ) of CMA-ES was chosen among (0.1, 0.3, 0.5, 1.0, 2.0) by selecting that which exhibited the highest average fitness. The number of evaluations per step (population size) of CMA-ES are  $(4 + 3 \ln(\text{network size}))$ , which is the default setting, in (B) and 5 in (C). The results depict the average maximum fitness after 100 evaluations. Error bars are 95% confidence intervals over 20 random networks ( $*p < 0.05$ ,  $**p < 0.01$ ,  $***p < 0.001$ , Student's t test, two-sided).

(D) Optimization performance for DRCNs of afatinib, trametinib, and palbociclib. We randomly selected 300 cell lines for each drug. Shaded areas are 95% confidence intervals over three independent runs.

(E) Median and median absolute deviations of F1 scores for 21 drugs in four methods.

We then compared prediction performance of optimized DRCNs over 21 drugs that have various modes of action (Table S1) with ML models SRMF<sup>46</sup> and DeepCDR,<sup>10</sup> and a logical prediction model LOBICO.<sup>47</sup> Each of the 21 drugs represents a corresponding group of drugs targeting the specific pathway according to the labels in the GDSC2 database. For each drug, 300 cell lines were randomly selected to form a training set and the remainder formed a test set. All comparison models were trained using the same set of training data with default settings. Drug-response values were binarized according to the rule defined by Knijnenburg et al.<sup>47</sup> Predicted drug-response values (Figure S2) from DRCNs are binarized using an optimal threshold, determined based on the highest F1 scores observed across the training set cell lines. F1 score (Equation 2) was calculated for each drug. Medians and median absolute deviations of all drugs are summarized in Figure 4E.

$$\text{Precision} = \frac{\text{True positive}}{\text{True positive} + \text{False positive}} \quad (\text{Equation 1})$$

$$\text{Recall} = \frac{\text{True positive}}{\text{True positive} + \text{False negative}}$$

$$F1 \text{ score} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (\text{Equation 2})$$

The prediction performances from the optimized DRCNs using GREY (GREY opt.) are lower than that of the ML models such as SRMF or DeepCDR. However, GREY opt. showed better performance than the logical prediction model LOBICO, and it has further strengths in terms of interpretability. For example, a marker predicted for afatinib by LOBICO is an *ERBB2* mutation,<sup>47</sup> which is a direct target of afatinib, but it does not explain



the underlying dynamics of how the mutation affects the phenotype. On the other hand, our model can suggest molecular regulatory dynamics by simulating the optimized DRCN with sensitive and resistant cell lines, as we describe in the following sections.

### Successful prediction of drug responses and molecular regulatory dynamics with the optimized DRCNs

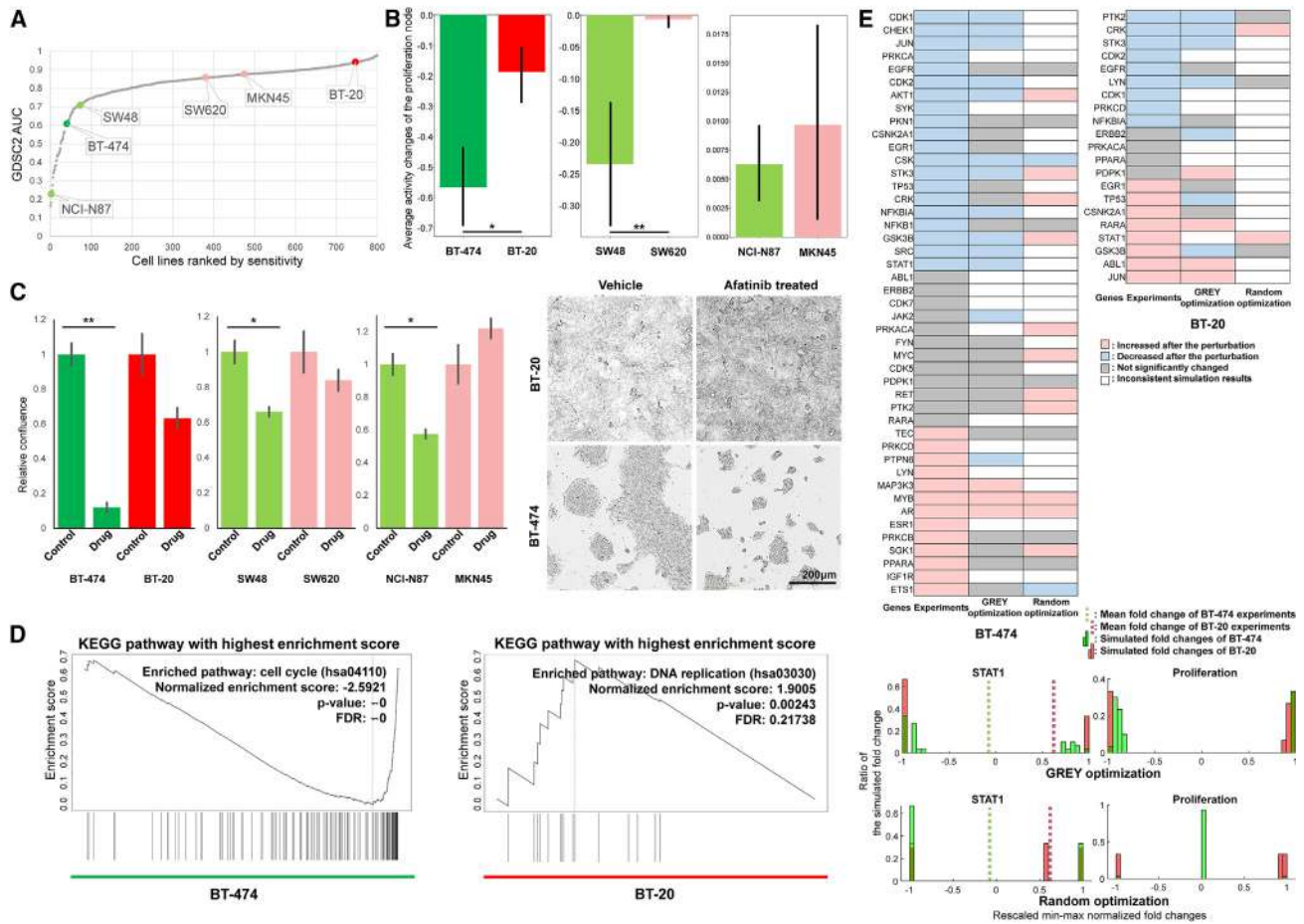
Interpretable drug responses depend on accurate inference of both the phenotypic response and the underlying molecular regulatory dynamics. We explored whether the DRCNs optimized through GREY can predict not only the phenotype-level drug responses but also the changes represented by fold changes in gene expression. To ensure the applicability of our gray box framework throughout the pan-cancer, we selected two cell lines, one sensitive and one resistant to afatinib treatments, respectively, for each of three frequently occurring cancer types in humans: breast, colon, and stomach cancer. Each of the six selected cell lines has been carefully chosen to serve as a representative model, demonstrating either a sensitive or resistant response to the selected drug, afatinib, based on the AUCs of the GDSC2 database (Figure 5A). However, within each cancer type, the sensitive and resistant cell lines share identical genetic markers for drug response predictions, as suggested by the MATCH basket trials conducted by the National Cancer Institute (NCI-MATCH). Therefore, classifying drug response differences between sensitive and resistant cell lines requires interpretation of the underlying molecular dynamics, which remain obfuscated by classical marker-based methods (see STAR Methods).

With the selected six cell lines, we firstly confirmed that GREY-optimized DRCNs are capable of reproducing the expected phenotypes in response to drug perturbations. To do so, we compared our in-house experimental data of drug responses in six cell lines with the drug responses simulated by optimized DRCNs. We mapped the genetic profile of each cell line onto the basal node activities of the afatinib DRCN, and GREY optimized the logical parameters to reproduce the viability of the cell lines in a robust manner. Using the optimized afatinib response dynamic network simulator, we simulated the proliferation response of each cell line, calculated as the average change of the proliferation node from 34 repeats (Figure 5B). Average activities of the proliferation nodes after afatinib treatment in two of the sensitive cell lines (BT-474 and SW48) were significantly decreased compared with those of two of the resistant cell lines (BT-20 and SW620). For the pair of stomach cancer cell lines, the average activity of the proliferation nodes did not decrease following afatinib treatment in the simulations. However, the average increase of activity in the sensitive stomach cancer cell line (NCI-N87) was smaller than that of the resistant stomach cancer cell line (MKN45). Thus, all of the predicted drug responses were consistent with the relative afatinib responses from the GDSC2 database. We assessed the effect of afatinib on proliferation of the cell lines in culture. The experimental results of afatinib responses were consistent with the response data from the GDSC2 database and with the simulated predictions of drug responses through DRCNs (Figure 5C). Detailed in-house experimental data of drug treatment in cancer cells are presented in Figure S3.

After confirming that the optimized model reproduces the expected phenotypic responses, we compared the model predictions of each gene in response to the drug perturbations with both existing databases and in-house biological data. To determine plausible gene sets that distinguish between sensitive and resistant pairs of cell lines at the cellular level, we first performed a pathway enrichment analysis with the web-based gene set analysis toolkit (WebGestalt)<sup>48</sup> based on the Kyoto Encyclopedia of Genes and Genomes pathways.<sup>49–51</sup> We selected BT-474 and BT-20 cell lines, because experimentally these cells had the most significantly different proliferation responses to afatinib. We calculated the log<sub>2</sub>-fold changes of differentially expressed genes (DEGs) before and after drug treatments in both cell lines by RNA sequencing data and analyzed the pathways with altered expressions (Figure 5D). The pathway with the highest enrichment score of DEGs in the afatinib-treated BT-474 was “decreasing cell cycle” (hsa04110), which is consistent with the response of this sensitive cell line. The pathway with the highest enrichment score of DEGs in the afatinib-treated BT-20 was “increasing DNA replication” (hsa03030), which is known to be related to drug resistance mechanisms,<sup>52,53</sup> but did not meet the threshold for false discovery rate, indicating that there were no significant changes in gene expression induced by afatinib in these cells.

We then preprocessed the expression fold change data of BT-20 and BT-474 cells to enable comparison with simulated afatinib treatment in GREY-optimized DRCNs. Although the expression fold changes in both BT-20 and BT-474 cell lines display a sufficiently large number of preprocessed DEGs exceeding threshold values to ensure quality control, the two cell lines exhibit a different number of significant DEGs. BT-20 cells had fewer genes with a base mean count over 100 (5,190 genes) than the BT-474 cells (15,583 genes), of which only 21 of 55 nodes were matched with the afatinib GREY-optimized DRCN (excluding the two phenotypic nodes). In the case of BT-474, 45 of 55 nodes in the afatinib DRCN were matched with the measured genes with a base mean count over 100. Both in-house expression data and DRCN simulation data are calculated in the form of log<sub>2</sub>-fold changes and min-max normalized. The log<sub>2</sub>-fold change values of each gene of the in-house experimental data and the average activity changes of each afatinib DRCN node computed by the drug treatment simulation were min-max normalized and compared. As the afatinib DRCN is an essential regulatory network of intricately interconnected genes and proteins that collectively influence drug response, it is imperative to validate the correlation between changes in gene expression levels and the resulting shifts in the activities of proteins within this network.<sup>54</sup> Consequently, the activity changes of proteins represented by nodes in the afatinib DRCN can be compared and validated with the RNA expression fold changes measured by our in-house experiments.

The insights into drug response differences between sensitive and resistance cell lines based on the in-house data were then compared with the predictions of GREY-optimized DRCNs. The optimized afatinib DRCN was simulated, and the average change in node activity due to drug perturbation was compared with the gene expression fold changes of the experimental data. The computed average activity change of each node was an



**Figure 5. In-house experimental validation of simulated changes in gene expression**

(A) Relative afatinib responses of six selected cell lines from the GDS2 drug responses database. Three afatinib-sensitive (in green) and three afatinib-resistant (in red) cell lines were well separated with respect to the area under the curve (AUC) of cell viability after various dosages of afatinib treatments given in the GDS2 database. Two breast cancer cell lines used for further analysis are colored with higher saturations.

(B) Simulated afatinib responses of six cell lines. Note that the first two graphs show a decrease in the average activity of the proliferation node (y axis < 0) and the third graph shows an increase in average activity of the proliferation node (y axis > 0) ( $p < 0.05$ ,  $**p < 0.01$ ,  $***p < 0.001$ , Student's t test, two-sided).

(C) In-house afatinib response data of six cell lines, quantified by relative confluence. Representative microscopic images for one pair of cell lines are shown ( $*p < 0.05$ ,  $**p < 0.01$ ,  $***p < 0.001$ , Student's t test, two-sided).

(D) The pathway enrichment analysis results of DEGs before and after each drug treatment. The most enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway of DEGs in afatinib-treated BT-474 is “decreasing cell cycle,” The most enriched pathway in afatinib-treated BT-20 is “increasing DNA replication.”

(E) Min-max normalized log<sub>2</sub>-fold change values of each gene of the in-house experimental data compared with min-max normalized average activity changes of each afatinib DRCN node computed by the drug treatment simulation. Parameter sets are optimized by a random agent and GREY, each repeated three times. Min-max normalized fold changes in the iterative simulation results of phenotype nodes and *STAT1* (downstream node of *EGFR*) computed by the optimized parameter sets found by GREY and a random agent are compared with experimental data in histograms.

average value of three simulations from three parameter sets optimized by GREY or by a random agent, which had the same structure as GREY but with randomly initialized weights. To check the consistency of fold changes in each node computed by the optimized parameter sets from GREY, we compared the results of 30 iterative afatinib treatment simulations, each with ten repetitions and different initial states for each of the three parameter sets, against the corresponding 30 simulation results computed by three parameter sets optimized by the random agent. Genes with moderate fold changes in experimental data and nodes with inconsistent fold changes, which depended on

the parameter sets or the initial state sets, were filtered according to log<sub>2</sub>-fold changes and fold change variance over 30 simulations, respectively. As a result, the fold change tendencies of 14 nodes simulated by the parameters of GREY-optimized afatinib DRCN and mutational profiles of BT-474 projected on the network were in accord with afatinib treatment experiments. Whereas in the random agent-optimized parameters, only 5 of 42 results matched the experiments. In BT-20, 7 of the simulated node fold changes optimized by GREY showed the same tendency as the experimental data, while only 1 node fold change optimized by the random agent was well in accord with the

experiments (Figure 5E). The correlation between the simulated node fold changes optimized by GREY and the experimental data is further validated with a statistical test and compared with the data of randomly shuffled node labels (Figure S4).

The major difference between the networks with parameters optimized by GREY and the random agent was that the former depends both on network topology and logic, whereas the latter depends on only network topology. Although random agent-optimized parameters were able to correctly predict some of the node fold changes, the simulated network converged into a single attractor. Note that the iterative fold change results of phenotypic nodes and STAT1 (downstream node of EGFR) simulated by GREY-optimized afatinib DRCNs were more spread out, indicating various attractors, while those simulated by random agent-optimized afatinib DRCNs were more concentrated in a few rescaled fold change intervals, indicating a single attractor (histograms in Figure 5E). The node states of the converged attractor with random agent-optimized afatinib DRCNs were predictable by topologically estimating the effect of the afatinib perturbation with the network diffusion algorithm suggested by Hofree et al.<sup>55</sup> The result infers that the dynamics of the network are simple, lacking any complex feedback or crosstalk dynamics, and hence inadequate for reproducing the behavior of a biological network. Moreover, the single attractor converged to by each of the optimized afatinib DRCNs using the random agent differed depending on the parameter sets. This implied that with additional optimized parameter sets, the simulated network dynamics would not converge into a single distribution. As a result, increasing the number of parameters sets cannot guarantee prediction performance improvement.

In stark contrast, the networks with the parameters optimized by GREY converged into a higher number of attractors, corresponding to varied regulatory dynamics, which exhibited similar phenotypic predictions. These results suggest that the networks optimized by GREY comprise not only topological information, but also dynamic properties including feedbacks and crosstalk, which are known to be important features to understand in order to overcome drug resistance mechanisms in cancer therapy.<sup>56</sup> Moreover, even though average responses of GREY and the random agent-optimized DRCNs are similar, GREY-optimized networks can be said to more properly recapitulate the biological nature of perturbation responses of regulatory networks. Biology regulatory networks exhibit stochasticity in their gene expressions, molecular activities, and probability of chemical reactions, which tend to produce a range of distinct attractors.<sup>57–60</sup> The multiple attractors exhibited by GREY-optimized DRCNs are hence more realistic. Moreover, the distributions of attractors from the network optimized by different parameter sets from GREY were more highly clustered than those of the random agents. This illustrates that the prediction performance of the drug-response networks optimized by GREY will not be derailed by a larger number of parameter sets. Thus, the networks optimized by GREY are more suitable than the networks optimized by the random agent for the purpose of cancer drug-response predictions.

### Interpreting regulatory mechanisms of drug resistance with the gray box framework

After establishing the ability of GREY to infer regulatory dynamics, we explored if those cellular-level insights could be in-

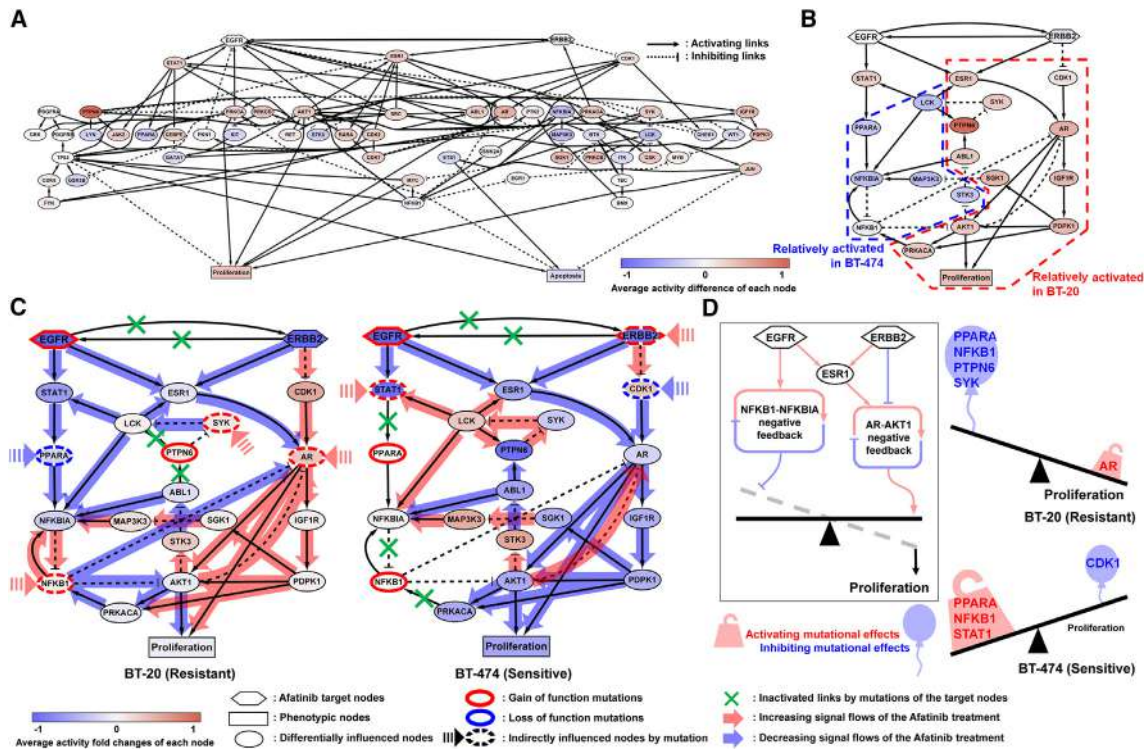
terpreted to reveal underlying mechanisms of drug resistance. We analyzed the simulated responses of BT-20 and BT-474 cells to discover network motifs behind the differences in afatinib response. We calculated the difference between the average activity fold change of afatinib-treated BT-20 and that of BT-474 by subtracting fold change values of the BT-474 afatinib core network from that of the BT-20 afatinib core network (Figure 6A). Note that some nodes had significantly different fold change values. With an assumption that the nodes with significant difference in fold change between the sensitive and the resistant cell lines cause the difference in afatinib responses, we selected the five nodes with the largest positive difference (*PTPN6*, *AR*, *PDPK1*, *IGF1R*, and *SGK1*, in decreasing order of absolute values) and five with the largest negative difference (*MAP3K3*, *LCK*, *NFKBIA*, *PPARA*, and *STK3*, in decreasing order of absolute values) in fold changes of simulated node activities.

To explore the interactions between these nodes in the context of drug resistance, we extracted an activity difference sub-network from the afatinib DRCN. This activity difference sub-network was formed by connecting the selected ten nodes with links and inter-nodes by PathLinker,<sup>61,62</sup> which explored up to 50 paths from the drug target nodes (*EGFR*, *ERBB2*) to the phenotypic node (proliferation) through the selected 10 nodes (Figure 6B). We found that the *ERBB2* cascade and *AR-AKT1* negative feedback are activated in BT-20, whereas the *EGFR* cascade and *NFKBIA-NFKB1* negative feedback are activated in BT-474 after afatinib treatment.

To interpret the mechanism underlying the different afatinib responses with respect to *EGFR* and *ERBB2* inhibition signals using the activity difference sub-network, mutational effects and afatinib perturbation signal flows were analyzed (Figure 6C). Signal flows from a node are defined by adding the incoming signal flow to the product of the node activity change with the sign (positive or negative) of the outgoing links.<sup>63</sup> The effect of a mutation on a node inside the activity difference sub-network was treated as fixing that node to a state, such that incoming links were deleted and a constant signal flows out from that node. The color of the arrows denotes whether the signal flow increases (red) or decreases (blue) after afatinib treatment. The effect of a mutation from outside the activity difference sub-network was incorporated by propagating the change in signal flow caused by a mutation through exclusive simple paths computed by the maximum matching algorithm. The propagated mutational profiles of two cell lines were incorporated into the sub-network as the indirect influence of mutation on the sub-network nodes (see STAR Methods).

The activity-difference sub-network reveals how drug responses differently affect each cell line. In BT-20, the decreasing signal flows from the afatinib-inhibited *EGFR* node through the *STAT1-PPARA-NFKBIA* cascade and the *ESR1-AR* cascade were offset by the increasing signal flows into the *NFKB1* node and the *AR* node, respectively. The *NFKB1* node and *AR* node were indirectly influenced by mutations in nodes outside the afatinib sub-network, but within the PCCN. The increasing signal from the afatinib-inhibited *ERBB2* node flows through the *CDK1-AR* cascade and is further enhanced by the increasing signal flow at the indirectly activated node, *AR*, increasing the proliferation node activity. The increasing signal flowing into





**Figure 6. Mechanistic analysis of afatinib treatment to identify the most influential nodes for predicting and controlling drug resistance**

(A) Afatinib DRCN. The color of each node represents the predicted relative activity changes of resistant and sensitive cell line models, and the color of the arrows denotes whether the signal flow increases (red) or decreases (blue) after afatinib treatment.

(B) Activity difference sub-network for afatinib response from BT-20 and BT-474 cells. This sub-network was extracted from the afatinib response network in (A). (C) Identification of context-dependent flows of afatinib treatment effects toward the proliferation node. The propagated mutational profile of BT-20 on the activity difference sub-network contained two nodes with activating mutations (*EGFR* and *PTPN6*), and three inputs from mutations outside this sub-network that activated (*NGKB1*, *SYK*, and *AR*), and one that is inactivated (*PPARA*). The propagated mutational profile of BT-474 on the activity difference sub-network had three nodes with activating mutations (*EGFR*, *PPARA*, and *NFKB1*), two nodes activated by input from mutations outside this sub-network (*STAT1* and *ERBB2*), and one node inactivated by external mutation (*CDK1*). Signal flow and node activity changes in activity difference sub-network of resistant cells and sensitive cells upon perturbation of *EGFR* and *ERBB2* (afatinib treatment) are illustrated.

(D) The balance between mutational effects on two essential negative feedback loops is primarily responsible for determining the resistance to afatinib treatment.

*SYK* from the mutations outside the afatinib DRCN also inactivated *LCK*, which dampened the ability for afatinib-inhibited *EGFR* to decrease the flow into *STAT1* and *NFKBIA*. The net outcome of the inhibitory effect of afatinib on *EGFR* and *ERBB2* is predicted to be an increase in proliferation node activity in BT-20.

In BT-474, the decreasing signal flow from the afatinib-inhibited *EGFR* node into the *STAT1* node is blocked by the *PPARA* mutation, but the afatinib-inhibited *EGFR* decreases *AR* node activity by reducing the signal flow through the *ESR1-AR* cascade. The increasing signal from the afatinib-inhibited *ERBB2* node is offset by decreasing the signal flow from the *CDK1* node, which is indirectly inhibited by mutations. As a result, the activity of the *AR* node is decreased after afatinib treatment, resulting in decreased activity of the proliferation node.

Together, the signal flow analyses of the activity difference sub-networks predicted a reason why BT-20 and BT-474 have different responses to afatinib treatment: the signal flows in the two cell lines differently affect the *NFKB1-NFKBIA*

and the *AR-AKT1* negative feedback regulations (Figure 6D). The underlying reason for the resistance to afatinib treatment for BT-20 can be summarized as the following: the mutational profile activates *AR-AKT1* negative feedback regulation and inactivates *NFKB1-NFKBIA* negative feedback regulation.

The genetic marker for the NCI-MATCH afatinib basket trial was mutated or amplified *EGFR*.<sup>64</sup> Both BT-20 and BT-474 cells have gain-of-function mutations in *EGFR*. Thus, using only that single genetic marker, one would expect the same response to afatinib for these two cell lines. On the other hand, our framework not only predicted the different responses of these two cell lines but also revealed their underlying mechanisms. The collection of genes involved in the two relevant feedbacks, *NFKB1-NFKBIA* and *AR-AKT1*, may constitute a new type of dynamic drug-response biomarker. A biomarker composed of a group of genes and their holistic dynamics, which respond differentially to drug treatment, may predict drug responsiveness more effectively than the classical mutation biomarkers of afatinib. Owing to the interpretable and simulatable afatinib DRCN, we can properly



predict the drug response, distinguishing resistant cell lines that need additional treatment. Moreover, by interpreting the resistance mechanisms, such as feedback and crosstalk, we can identify efficient combinatorial drug targets to compensate for those resistances. Additional validations for suggesting dynamic drug-response biomarkers are in [Figure S5](#). Detailed signal flow analyses of trametinib-treated SW620 and SW48 cell lines are discussed in [Figure S6](#).

## DISCUSSION

Predicting cellular response to a perturbation is of great interest in systems biology and ML.<sup>4,9,20,46</sup> Systems biology uses a white box approach based on mechanistic modeling of intracellular molecular regulatory networks, and can provide interpretable insights into regulatory dynamics, but is difficult to scale up to larger networks. ML uses a black box approach by modeling input-output relationships, and can offer high prediction performance, but it is difficult to interpret its underlying rationale. There are several works<sup>14,15,20,22,47,65</sup> that have attempted to merge these approaches to make interpretable predictions, but they fall short of achieving detailed insight into molecular regulatory dynamics about cellular responses after a perturbation. Therefore, to provide interpretations of molecular regulation dynamics, we constructed a gray box framework by combining the white box and the black box L2O approaches to model molecular regulatory dynamics. We devised a black box optimizer, GREY, to tackle the high computational complexity of parameter optimization for DRCNs. GREY learned efficient derivative-free search rules by identifying the relationship between the structure of a network and its dynamical response. To accomplish this, we adopted a GNN<sup>44</sup> to represent attributes of the network components (i.e., nodes and edges), and used meta-reinforcement learning to learn different dynamic responses to perturbations for each network. Remarkably, GREY optimized the edge weights of the DRCN for trametinib three times more efficiently than the state-of-the-art derivative-free optimization method, CMA-ES, with respect to sample efficiency. Moreover, GREY optimized the DRCN for palbociclib to a fitness value of 0.5, which is difficult to achieve by inferring the regulatory logic of a network of such scale (~600 edges) using any other optimization method. We applied GREY to optimize the DRCNs of 21 drugs. Constructing a DRCN for each drug involves numerous steps that influence the number of nodes and edges, and the optimal settings may vary for each drug. However, we applied the same settings to construct all DRCNs to allow for a systematic comparison of the prediction performance of optimized DRCNs. Consequently, certain drugs inevitably have a suboptimal DRCN, influencing our results. Investigating the impact of drug-specific settings on the prediction performance of DRCNs would be an intriguing avenue for further research.

To demonstrate the efficacy of the gray box framework on recapitulating cellular dynamics, we applied it to the drug-response predictions in cancer networks. Our results have shown the broad applicability of the framework with a wide range of cancer types and drug targets. As the mechanistic interpretation of DRCNs is not detached from drug-response

prediction performance, a question whether the molecular mechanism interpretation of afatinib drug responses is perfectly correct or cannot arise. However, since the Boolean networks with optimized weights not only fully reproduce the quantitative responses for various cell lines of several drugs, but also reproduce the activity changes after those drug treatments, we can assume that the regulatory logic recapitulated by the weighted Boolean logic of the optimized DRCNs accurately models the underlying molecular regulatory mechanism, at least for the scope of drug response. Moreover, the optimized DRCNs not only correctly predicted drug responses of BT-20 and BT-474 cell lines, but also suggested a biomarker based on regulatory dynamics that was directly related with drug-response mechanisms and thus more accurately predict the resistance. This result is well in accord with the research of Choi et al.,<sup>4</sup> where classifying cancer cell lines with similar changes in expression levels induced by drug treatment exhibits more similar drug responses than classifying cancer cell lines with similar mutation profiles. Therefore, we believe that the interpretability of the model can provide insight into biological mechanisms, such as drug resistance, which can subsequently help to overcome these hurdles in real biological applications.

Cancer is a patient-specific as well as tissue-specific disease, but there are recurrent regulatory architectures common to all cancer types.<sup>66,67</sup> Based on these recurrent architectures, we hypothesized the existence of a PCCN and developed a method for reconstructing such a core regulatory network from large datasets. Feltes et al.<sup>68</sup> also tried to curate core regulatory networks relevant to all cancer types.<sup>69,70</sup> However, these approaches might contain indirect interactions as curated links depend on the data. Due to the risk that few incorrect links in the network structure can vastly change the dynamics of the model, we instead focused on using a well-known network structure database such as Omnipath and cross-checked every link in the PCCN with other network databases: Pathway Commons, DoRothEA, and Cui's human signaling network.<sup>71–73</sup> Each drug has a different mode of action at the cellular level and thereby affects a different part of the intracellular network. Thus, we extracted the core structure responsible for each drug and optimized the parameters separately.

The gray box framework can be applicable to readily accessible perturbation data, while enabling deep interpretation of molecular regulatory dynamics. Single-cell sequencing of drug-induced cellular state transitions provides detailed dynamic information of drug responses. Using these data, a number of studies predict and interpret the response dynamics of molecular regulatory perturbations.<sup>74,75</sup> However, aside from the high financial costs, producing such data is a difficult and time-consuming task that requires experimentally well-prepared samples of cells undergoing specific dynamics and properly selected pre-processing and analyzing tools.<sup>76</sup> Under these limitations, the most commonly accessible data of molecular regulatory perturbations are drug responses of cancer cell lines in terms of cell viabilities. It is notable that discrete DRCNs optimized by GREY have successfully interpreted the drug-resistance mechanisms of afatinib in terms of molecular

regulatory dynamics without explicit molecular regulatory experimental data, but only with cell viability. Owing to the vastness of public drug-response viability data in various diseases caused by malfunctioning cell networks, the suggested framework can also be applicable to controlling fates of pluripotent stem cells<sup>77,78</sup> or macrophage differentiation.<sup>79</sup> Moreover, prediction performance and interpretation of molecular regulatory mechanisms of the developed gray box framework can be expected to further improve if data on molecular regulatory dynamics are also explicitly incorporated. Currently, GREY is trained with scale-free random networks for the specific task of optimizing biological networks, although many biological networks are not scale free.<sup>80</sup> Thus, there is room for improving the optimization performance of GREY by training it with more biologically plausible random networks. Furthermore, in the future, GREY could be trained with other types of random networks with different structural characteristics to optimize various discrete dynamic models.

### Limitations of the study

The DRCN formulates an "essential structure" connecting biological molecules tied to drug responses in the PCCN, aiming to minimize computational burdens of the optimizer by limiting number of links extracted between molecules in the PCCN. While this method efficiently identifies core network structures among molecules associated with drug responses, it overlooks long-distance bypass regulations involving multiple connections. Such pathways, despite potentially lower chances of affecting drug response dynamics, are critical but might be missed by our methodology. Dynamics influenced by these bypasses tend to be embedded through optimized parameters in shorter pathways among the selected molecules, omitting a few significant long-distance interactions. We utilized GREY to optimize DRCNs for 21 drugs. However, the prediction performance for certain drugs did not reach significance. Although constructing each DRCN involves varying optimal settings, we employed uniform settings for systematic performance comparison, potentially leading to suboptimal DRCNs for some drugs. In addition, GREY does not incorporate regularization in its optimization procedure for DRCNs. Incorporating regularization in future studies could enhance performance.

### STAR★METHODS

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

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead contact
  - Materials availability
  - Data and code availability
- **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**
  - Setup for in-house experimental validations
- **METHOD DETAILS**
  - Constructing DRCNs
  - Weighted sum Boolean logics
  - Synthetic environments for training GRAY
  - Neural network structure of GRAY

- Training procedure
- Selecting six cell lines for in-house experimental validations
- Integrating cell line mutation profiles with functional effect
- Maximum matching-based mutation data projection
- Incorporating mutations and drugs onto the simulation
- Signal flow analysis of DRCN perturbations
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

### SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.crmeth.2024.100773>.

### ACKNOWLEDGMENTS

This work was initially supported by Samsung Research Funding & Incubation Center of Samsung Electronics under project no. SRFC-IT1802-02. It was further supported by the National Research Foundation of Korea (NRF) funded by the Korea Government, the Ministry of Science and ICT (2023R1A2C3002619), and the Bio & Medical Technology Development Program of NRF funded by the Ministry of Science and ICT (2021M3A9I4024447). It was also supported by the Technology Development Program (RS-2023-00258764) funded by the Ministry of Small and Medium-sized Enterprises (SMEs) and Startups (MSS), Republic of Korea. The authors thank Sea Rom Choi for her critical reading and helpful comments. The authors thank Dongkwan Shin and Dongsan Kim for their helpful comments.

### AUTHOR CONTRIBUTIONS

K.-H.C. conceived the idea of the study. Y.K. and J.L. developed network construction and extraction algorithms. Y.H. developed GREY. J.I.J. and C.K.L. developed mutation data projection method. J.-R.G., S.-H.J., and T.K. performed experiments. Y.K., Y.H., C.H., J.L., J.I.J., C.-K.L., J.K., and K.-H.C. wrote the manuscript. K.-H.C. designed the project and supervised the study.

### DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 13, 2023

Revised: March 19, 2024

Accepted: April 19, 2024

Published: May 13, 2024

### REFERENCES

1. Kolch, W., Halasz, M., Granovskaya, M., and Kholodenko, B.N. (2015). The dynamic control of signal transduction networks in cancer cells. *Nat. Rev. Cancer* 15, 515–527. <https://doi.org/10.1038/nrc3983>.
2. Yuan, Z., Zhao, C., Di, Z., Wang, W.X., and Lai, Y.C. (2013). Exact controllability of complex networks. *Nat. Commun.* 4, 2447. <https://doi.org/10.1038/ncomms3447>.
3. Hopkins, A.L. (2008). Network pharmacology: the next paradigm in drug discovery. *Nat. Chem. Biol.* 4, 682–690. <https://doi.org/10.1038/nchembio.118>.
4. Choi, M., Shi, J., Zhu, Y., Yang, R., and Cho, K.H. (2017). Network dynamics-based cancer panel stratification for systemic prediction of anti-cancer drug response. *Nat. Commun.* 8, 1940. <https://doi.org/10.1038/s41467-017-02160-5>.
5. Park, S.G., Lee, T., Kang, H.Y., Park, K., Cho, K.H., and Jung, G. (2006). The influence of the signal dynamics of activated form of IKK on NF- $\kappa$ B and anti-apoptotic gene expressions: a systems biology approach. *FEBS Lett.* 580, 822–830. <https://doi.org/10.1016/j.febslet.2006.01.004>.

6. Xu, C., and Jackson, S.A. (2019). Machine learning and complex biological data. *Genome Biol.* 20, 76. <https://doi.org/10.1186/s13059-019-1689-0>.
7. Leung, M.K.K., Xiong, H.Y., Lee, L.J., and Frey, B.J. (2014). Deep learning of the tissue-regulated splicing code. *Bioinformatics* 30, i121–i129. <https://doi.org/10.1093/bioinformatics/btu277>.
8. Alipanahi, B., Delong, A., Weirauch, M.T., and Frey, B.J. (2015). Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat. Biotechnol.* 33, 831–838. <https://doi.org/10.1038/nbt.3300>.
9. Preuer, K., Lewis, R.P.I., Hochreiter, S., Bender, A., Bulusu, K.C., and Klambauer, G. (2018). DeepSynergy: predicting anti-cancer drug synergy with Deep Learning. *Bioinformatics* 34, 1538–1546. <https://doi.org/10.1093/bioinformatics/btx806>.
10. Liu, Q., Hu, Z., Jiang, R., and Zhou, M. (2020). DeepCDR: a hybrid graph convolutional network for predicting cancer drug response. *Bioinformatics* 36, i911–i918. <https://doi.org/10.1093/bioinformatics/btaa822>.
11. Bomane, A., Gonçalves, A., and Ballester, P.J. (2019). Paclitaxel Response Can Be Predicted With Interpretable Multi-Variate Classifiers Exploiting DNA-Methylation and miRNA Data. *Front. Genet.* 10, 1041. <https://doi.org/10.3389/fgene.2019.01041>.
12. Lind, A.P., and Anderson, P.C. (2019). Predicting drug activity against cancer cells by random forest models based on minimal genomic information and chemical properties. *PLoS One* 14, e0219774. <https://doi.org/10.1371/journal.pone.0219774>.
13. 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-task linear regression. *Bioinformatics* 33, i359–i368. <https://doi.org/10.1093/bioinformatics/btx266>.
14. He, L., Tang, J., Andersson, E.I., Timonen, S., Koschmieder, S., Wennerberg, K., Mustjoki, S., and Aittokallio, T. (2018). Patient-Customized Drug Combination Prediction and Testing for T-cell Prolymphocytic Leukemia Patients. *Cancer Res.* 78, 2407–2418. <https://doi.org/10.1158/0008-5472.CAN-17-3644>.
15. He, L., Bulanova, D., Oikkonen, J., Hakkinen, A., Zhang, K., Zheng, S., Wang, W., Erkan, E.P., Carpen, O., Joutsiniemi, T., et al. (2021). Network-guided identification of cancer-selective combinatorial therapies in ovarian cancer. *Briefings Bioinf.* 22. <https://doi.org/10.1093/bib/bbab272>.
16. Ogunleye, A.Z., Piyawajanusorn, C., Gonçalves, A., Ghislat, G., and Ballester, P.J. (2022). Interpretable Machine Learning Models to Predict the Resistance of Breast Cancer Patients to Doxorubicin from Their microRNA Profiles. *Adv. Sci.* 9, e2201501. <https://doi.org/10.1002/adv.202201501>.
17. Montavon, G., Binder, A., Lapuschkin, S., Samek, W., and Müller, K.-R. (2019). Layer-Wise Relevance Propagation: An Overview. In *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, W. Samek, G. Montavon, A. Vedaldi, L.K. Hansen, and K.-R. Müller, eds. (Springer International Publishing), pp. 193–209. [https://doi.org/10.1007/978-3-030-28954-6\\_10](https://doi.org/10.1007/978-3-030-28954-6_10).
18. Marco Singh, S., and Guestrin, C. (2016). Model-Agnostic Interpretability of Machine Learning. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1606.05386>.
19. Elmarakeby, H.A., Hwang, J., Arafeh, R., Crowdis, J., Gang, S., Liu, D., AlDubayan, S.H., Salari, K., Kregel, S., Richter, C., et al. (2021). Biologically informed deep neural network for prostate cancer discovery. *Nature* 598, 348–352. <https://doi.org/10.1038/s41586-021-03922-4>.
20. Kuenzi, B.M., Park, J., Fong, S.H., Sanchez, K.S., Lee, J., Kreisberg, J.F., Ma, J., and Ideker, T. (2020). Predicting Drug Response and Synergy Using a Deep Learning Model of Human Cancer Cells. *Cancer Cell* 38, 672–684.e6. <https://doi.org/10.1016/j.ccell.2020.09.014>.
21. 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. <https://doi.org/10.1038/Nmeth.4627>.
22. Fortelny, N., and Bock, C. (2020). Knowledge-primed neural networks enable biologically interpretable deep learning on single-cell sequencing data. *Genome Biol.* 21, 190. <https://doi.org/10.1186/s13059-020-02100-5>.
23. Kitano, H. (2002). Systems biology: a brief overview. *Science* 295, 1662–1664. <https://doi.org/10.1126/science.1069492>.
24. Seo, C.H., Kim, J.R., Kim, M.S., and Cho, K.H. (2009). Hub genes with positive feedbacks function as master switches in developmental gene regulatory networks. *Bioinformatics* 25, 1898–1904. <https://doi.org/10.1093/bioinformatics/btp316>.
25. Kitano, H. (2002). Computational systems biology. *Nature* 420, 206–210. <https://doi.org/10.1038/nature01254>.
26. Lee, H.S., Hwang, C.Y., Shin, S.Y., Kwon, K.S., and Cho, K.H. (2014). MLK3 Is Part of a Feedback Mechanism That Regulates Different Cellular Responses to Reactive Oxygen Species. *Sci. Signal.* 7, ra52. <https://doi.org/10.1126/scisignal.2005260>.
27. Gitter, A., Klein-Seetharaman, J., Gupta, A., and Bar-Joseph, Z. (2011). Discovering pathways by orienting edges in protein interaction networks. *Nucleic Acids Res.* 39, e22. <https://doi.org/10.1093/nar/gkq1207>.
28. Yu, J., Smith, V.A., Wang, P.P., Hartemink, A.J., and Jarvis, E.D. (2004). Advances to Bayesian network inference for generating causal networks from observational biological data. *Bioinformatics* 20, 3594–3603. <https://doi.org/10.1093/bioinformatics/bth448>.
29. Dunn, S.J., Martello, G., Yordanov, B., Emmott, S., and Smith, A.G. (2014). Defining an essential transcription factor program for naive pluripotency. *Science* 344, 1156–1160. <https://doi.org/10.1126/science.1248882>.
30. Tashkova, K., Korošec, P., Silc, J., Todorovski, L., and Dzeroski, S. (2011). Parameter estimation with bio-inspired meta-heuristic optimization: modeling the dynamics of endocytosis. *BMC Syst. Biol.* 5, 159. <https://doi.org/10.1186/1752-0509-5-159>.
31. Trairatphisan, P., Mizera, A., Pang, J., Tantar, A.A., and Sauter, T. (2014). optPBN: An Optimisation Toolbox for Probabilistic Boolean Networks (vol 9, e98001, 2014). *PLoS One* 9. <https://doi.org/10.1371/journal.pone.0105913>.
32. Chen, T., Chen, X., Chen, W., Heaton, H., Liu, J., Wang, Z., and Yin, W. (2021). Learning to Optimize: A Primer and A Benchmark. Preprint at arXiv. <https://doi.org/10.48550/arXiv.2103.12828>.
33. Chen, Y.T., Hoffman, M.W., Colmenarejo, S.G., Denil, M., Lillicrap, T.P., Botvinick, M., and de Freitas, N. (2017). Learning to Learn without Gradient Descent by Gradient Descent. *Pr Mach Learn Res* 70.
34. Hugo, S.G., Benjamin, L., and G, C. (2021). Meta Learning Black-Box Population-Based Optimizers. Preprint at arXiv. <https://doi.org/10.48550/arXiv.2103.03526>.
35. 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. <https://doi.org/10.1016/j.cell.2016.06.017>.
36. Turei, D., Korcsmaros, T., and Saez-Rodriguez, J. (2016). OmniPath: guidelines and gateway for literature-curated signaling pathway resources. *Nat. Methods* 13, 966–967. <https://doi.org/10.1038/nmeth.4077>.
37. Iannuccelli, M., Micarelli, E., Surdo, P.L., Palma, A., Perfetto, L., Rozzo, I., Castagnoli, L., Licata, L., and Cesareni, G. (2020). CancerGeneNet: linking driver genes to cancer hallmarks. *Nucleic Acids Res.* 48, D416–D421. <https://doi.org/10.1093/nar/gkz871>.
38. Chen, A.P., Eljanne, M., Harris, L., Malik, S., and Seibel, N.L. (2019). National Cancer Institute Basket/Umbrella Clinical Trials: MATCH, LungMAP, and Beyond. *Cancer J.* 25, 272–281. <https://doi.org/10.1097/PPO.0000000000000389>.
39. Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E., et al. (2019).

- COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* 47, D941–D947. <https://doi.org/10.1093/nar/gky1015>.
40. Tokheim, C.J., Papadopoulos, N., Kinzler, K.W., Vogelstein, B., and Karchin, R. (2016). Evaluating the evaluation of cancer driver genes. *Proc. Natl. Acad. Sci. USA* 113, 14330–14335. <https://doi.org/10.1073/pnas.1616440113>.
41. Buşoni, L., Babuška, R., and De Schutter, B. (2010). Multi-agent Reinforcement Learning: An Overview. In *Innovations in Multi-Agent Systems and Applications - 1*, D. Srinivasan and L.C. Jain, eds. (Springer Berlin Heidelberg), pp. 183–221. [https://doi.org/10.1007/978-3-642-14435-6\\_7](https://doi.org/10.1007/978-3-642-14435-6_7).
42. Hospedales, T.M., Antoniou, A., Micaelli, P., and Storkey, A.J. (2021). Meta-Learning in Neural Networks: A Survey. *IEEE Trans Pattern Anal Mach Intell* PP. <https://doi.org/10.1109/TPAMI.2021.3079209>.
43. Omidshafiei, S., Kim, D.K., Liu, M., Tesauro, G., Riemer, M., Amato, C., Campbell, M., and How, J.P. (2019). Learning to Teach in Cooperative Multiagent Reinforcement Learning. *Thirty-Third Aaai Conference on Artificial Intelligence/Thirty-First Innovative Applications of Artificial Intelligence Conference/Ninth Aaai Symposium on Educational Advances in Artificial Intelligence*, pp. 6128–6136.
44. Peter, J., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., Tacchetti, A., Raposo, D., Santoro, A., Faulkner, R., et al. (2018). Relational inductive biases, deep learning, and graph networks. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1806.01261>.
45. Nikolaus (2016). The CMA Evolution Strategy: A Tutorial. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1604.00772>.
46. Wang, L., Li, X., Zhang, L., and Gao, Q. (2017). Improved anticancer drug response prediction in cell lines using matrix factorization with similarity regularization. *BMC Cancer* 17, 513. <https://doi.org/10.1186/s12885-017-3500-5>.
47. Knijnenburg, T.A., Klau, G.W., Iorio, F., Garnett, M.J., McDermott, U., Shmulevich, I., and Wessels, L.F.A. (2016). Logic models to predict continuous outputs based on binary inputs with an application to personalized cancer therapy. *Sci. Rep.* 6, 36812. <https://doi.org/10.1038/srep36812>.
48. Wang, J., Duncan, D., Shi, Z., and Zhang, B. (2013). WEB-based GENE SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res.* 41, W77–W83. <https://doi.org/10.1093/nar/gkt439>.
49. Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. <https://doi.org/10.1093/nar/28.1.27>.
50. Kanehisa, M. (2019). Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* 28, 1947–1951. <https://doi.org/10.1002/pro.3715>.
51. Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., and Tanabe, M. (2021). KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* 49, D545–D551. <https://doi.org/10.1093/nar/gkaa970>.
52. Mansoori, B., Mohammadi, A., Davudian, S., Shirjang, S., and Baradaran, B. (2017). The Different Mechanisms of Cancer Drug Resistance: A Brief Review. *Adv. Pharmaceut. Bull.* 7, 339–348. <https://doi.org/10.15171/apb.2017.041>.
53. Vadlapatla, R.K., Vadlapudi, A.D., Pal, D., and Mitra, A.K. (2013). Mechanisms of drug resistance in cancer chemotherapy: coordinated role and regulation of efflux transporters and metabolizing enzymes. *Curr. Pharmaceut. Des.* 19, 7126–7140. <https://doi.org/10.2174/13816128113199990493>.
54. Liu, A., Trairatphisan, P., Gjerga, E., Didangelos, A., Barratt, J., and Saez-Rodriguez, J. (2019). From expression footprints to causal pathways: contextualizing large signaling networks with CARNIVAL. *NPJ Syst Biol Appl* 5, 40. <https://doi.org/10.1038/s41540-019-0118-z>.
55. Hofree, M., Shen, J.P., Carter, H., Gross, A., and Ideker, T. (2013). Network-based stratification of tumor mutations. *Nat. Methods* 10, 1108–1115. <https://doi.org/10.1038/nmeth.2651>.
56. Nguyen, L.K., and Kholodenko, B.N. (2016). Feedback regulation in cell signalling: Lessons for cancer therapeutics. *Semin. Cell Dev. Biol.* 50, 85–94. <https://doi.org/10.1016/j.semcdb.2015.09.024>.
57. Kitano, H. (2004). Biological robustness. *Nat. Rev. Genet.* 5, 826–837. <https://doi.org/10.1038/nrg1471>.
58. Hayford, C.E., Tyson, D.R., Robbins, C.J., 3rd, Frick, P.L., Quaranta, V., and Harris, L.A. (2021). An in vitro model of tumor heterogeneity resolves genetic, epigenetic, and stochastic sources of cell state variability. *PLoS Biol.* 19, e3000797. <https://doi.org/10.1371/journal.pbio.3000797>.
59. Brock, A., Chang, H., and Huang, S. (2009). Non-genetic heterogeneity—a mutation-independent driving force for the somatic evolution of tumours. *Nat. Rev. Genet.* 10, 336–342. <https://doi.org/10.1038/nrg2556>.
60. Brock, A., Krause, S., and Ingber, D.E. (2015). Control of cancer formation by intrinsic genetic noise and microenvironmental cues. *Nat. Rev. Cancer* 15, 499–509. <https://doi.org/10.1038/nrc3959>.
61. Ritz, A., Poirel, C.L., Tegge, A.N., Sharp, N., Simmons, K., Powell, A., Kale, S.D., and Murali, T.M. (2016). Pathways on demand: automated reconstruction of human signaling networks. *NPJ Syst Biol Appl* 2, 16002. <https://doi.org/10.1038/npjbsba.2016.2>.
62. Gil, D.P., Law, J.N., and Murali, T.M. (2017). The PathLinker app: Connect the dots in protein interaction networks. *F1000Res* 6, 58. <https://doi.org/10.12688/f1000research.9909.1>.
63. Lee, D., and Cho, K.H. (2018). Topological estimation of signal flow in complex signaling networks. *Sci. Rep.* 8, 5262. <https://doi.org/10.1038/s41598-018-23643-5>.
64. Flaherty, K.T., Gray, R., Chen, A., Li, S., Patton, D., Hamilton, S.R., Williams, P.M., Mitchell, E.P., Iafraite, A.J., Sklar, J., et al. (2020). The Molecular Analysis for Therapy Choice (NCI-MATCH) Trial: Lessons for Genomic Trial Design. *J. Natl. Cancer Inst.* 112, 1021–1029. <https://doi.org/10.1093/jnci/djz245>.
65. Masica, D.L., and Karchin, R. (2013). Collections of Simultaneously Altered Genes as Biomarkers of Cancer Cell Drug Response. *Cancer Res.* 73, 1699–1708. <https://doi.org/10.1158/0008-5472.Can-12-3122>.
66. Sanchez-Vega, F., Mina, M., Armenia, J., Chatila, W.K., Luna, A., La, K.C., Dimitriadou, S., Liu, D.L., Kantheti, H.S., Saghaforina, S., et al. (2018). Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell* 173, 321–337.e10. <https://doi.org/10.1016/j.cell.2018.03.035>.
67. Califano, A., and Alvarez, M.J. (2017). The recurrent architecture of tumour initiation, progression and drug sensitivity. *Nat. Rev. Cancer* 17, 116–130. <https://doi.org/10.1038/nrc.2016.124>.
68. Feltes, B.C., Poloni, J.d.F., Nunes, I.J.G., Faria, S.S., and Dorn, M. (2020). Multi-Approach Bioinformatics Analysis of Curated Omics Data Provides a Gene Expression Panorama for Multiple Cancer Types. *Front. Genet.* 11, 586602. <https://doi.org/10.3389/fgene.2020.586602>.
69. Emad, A., and Sinha, S. (2021). Inference of phenotype-relevant transcriptional regulatory networks elucidates cancer type-specific regulatory mechanisms in a pan-cancer study. *NPJ Syst Biol Appl* 7, 9. <https://doi.org/10.1038/s41540-021-00169-7>.
70. Wong, Y.H., Chen, R.H., and Chen, B.S. (2014). Core and specific network markers of carcinogenesis from multiple cancer samples. *J. Theor. Biol.* 362, 17–34. <https://doi.org/10.1016/j.jtbi.2014.05.045>.
71. Garcia-Alonso, L., Holland, C.H., Ibrahim, M.M., Turei, D., and Saez-Rodriguez, J. (2019). Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res.* 29, 1363–1375. <https://doi.org/10.1101/gr.240663.118>.
72. Cerami, E.G., Gross, B.E., Demir, E., Rodchenkov, I., Babur, O., Anwar, N., Schultz, N., Bader, G.D., and Sander, C. (2011). Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res.* 39, D685–D690690. <https://doi.org/10.1093/nar/gkq1039>.



73. Cui, Q., Ma, Y., Jaramillo, M., Bari, H., Awan, A., Yang, S., Zhang, S., Liu, L., Lu, M., O'Connor-McCourt, M., et al. (2007). A map of human cancer signaling. *Mol. Syst. Biol.* 3, 152. <https://doi.org/10.1038/msb4100200>.
74. Yuan, B., Shen, C.Y., Luna, A., Korkut, A., Marks, D.S., Ingraham, J., and Sander, C. (2021). CellBox: Interpretable Machine Learning for Perturbation Biology with Application to the Design of Cancer Combination Therapy. *Cell Systems* 12, 128. <https://doi.org/10.1016/j.cels.2020.11.013>.
75. Frohlich, F., Kessler, T., Weindl, D., Shadrin, A., Schmiester, L., Hache, H., Muradyan, A., Schutte, M., Lim, J.H., Heinig, M., et al. (2018). Efficient Parameter Estimation Enables the Prediction of Drug Response Using a Mechanistic Pan-Cancer Pathway Model. *Cell Syst* 7, 567–579.e566. <https://doi.org/10.1016/j.cels.2018.10.013>.
76. Lähnemann, D., Köster, J., Szczurek, E., McCarthy, D.J., Hicks, S.C., Robinson, M.D., Vallejos, C.A., Campbell, K.R., Beerenwinkel, N., Mahfouz, A., et al. (2020). Eleven grand challenges in single-cell data science. *Genome Biol.* 21, 31. <https://doi.org/10.1186/s13059-020-1926-6>.
77. Yachie-Kinoshita, A., Onishi, K., Ostblom, J., Langley, M.A., Posfai, E., Rossant, J., and Zandstra, P.W. (2018). Modeling signaling-dependent pluripotency with Boolean logic to predict cell fate transitions. *Mol. Syst. Biol.* 14, e7952. <https://doi.org/10.15252/msb.20177952>.
78. Collombet, S., van Oevelen, C., Sardina Ortega, J.L., Abou-Jaoudé, W., Di Stefano, B., Thomas-Chollier, M., Graf, T., and Thieffry, D. (2017). Logical modeling of lymphoid and myeloid cell specification and transdifferentiation. *Proc. Natl. Acad. Sci. USA* 114, 5792–5799. <https://doi.org/10.1073/pnas.1610622114>.
79. Palma, A., Jarrah, A.S., Tieri, P., Cesareni, G., and Castiglione, F. (2018). Gene Regulatory Network Modeling of Macrophage Differentiation Corroborates the Continuum Hypothesis of Polarization States. *Front. Physiol.* 9, 1659. <https://doi.org/10.3389/fphys.2018.01659>.
80. Broido, A.D., and Clauset, A. (2019). Scale-free networks are rare. *Nat. Commun.* 10, 1017. <https://doi.org/10.1038/s41467-019-08746-5>.
81. Rouillard, A.D., Gundersen, G.W., Fernandez, N.F., Wang, Z., Monteiro, C.D., McDermott, M.G., and Ma'ayan, A. (2016). The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database* 2016. <https://doi.org/10.1093/database/baw100>.
82. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. <https://doi.org/10.1093/bioinformatics/bts635>.
83. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>.
84. Cokelaer, T., Chen, E., Iorio, F., Menden, M.P., Lightfoot, H., Saez-Rodriguez, J., and Garnett, M.J. (2018). GDSCTools for mining pharmacogenomic interactions in cancer. *Bioinformatics* 34, 1226–1228. <https://doi.org/10.1093/bioinformatics/btx744>.
85. Reyna, M.A., Leiserson, M.D.M., and Raphael, B.J. (2018). Hierarchical HotNet: identifying hierarchies of altered subnetworks. *Bioinformatics* 34, i972–i980. <https://doi.org/10.1093/bioinformatics/bty613>.
86. Cui, Q. (2010). A network of cancer genes with co-occurring and anti-co-occurring mutations. *PLoS One* 5, e13180. <https://doi.org/10.1371/journal.pone.0013180>.
87. Hristov, B.H., Chazelle, B., and Singh, M. (2020). uKIN Combines New and Prior Information with Guided Network Propagation to Accurately Identify Disease Genes. *Cell Syst* 10, 470–479.e473. <https://doi.org/10.1016/j.cels.2020.05.008>.
88. Edmonds, J. (1968). Optimum branchings. *Part. Mathematics and the Decision Sciences* 1, 25.
89. van der Meer, D., Barthorpe, S., Yang, W., Lightfoot, H., Hall, C., Gilbert, J., Francies, H.E., and Garnett, M.J. (2019). Cell Model Passports—a hub for clinical, genetic and functional datasets of preclinical cancer models. *Nucleic Acids Res.* 47, D923–D929. <https://doi.org/10.1093/nar/gky872>.
90. Vanunu, O., Magger, O., Ruppin, E., Shlomi, T., and Sharan, R. (2010). Associating genes and protein complexes with disease via network propagation. *PLoS Comput. Biol.* 6, e1000641. <https://doi.org/10.1371/journal.pcbi.1000641>.
91. Font-Clos, F., Zapperi, S., and La Porta, C.A.M. (2018). Topography of epithelial-mesenchymal plasticity. *Proc. Natl. Acad. Sci. USA* 115, 5902–5907. <https://doi.org/10.1073/pnas.1722609115>.
92. Huang, S., Eichler, G., Bar-Yam, Y., and Ingber, D.E. (2005). Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys. Rev. Lett.* 94, 128701. <https://doi.org/10.1103/PhysRevLett.94.128701>.
93. Choi, M., Shi, J., Jung, S.H., Chen, X., and Cho, K.H. (2012). Attractor Landscape Analysis Reveals Feedback Loops in the p53 Network That Control the Cellular Response to DNA Damage. *Sci. Signal.* 5, ra83. <https://doi.org/10.1126/scisignal.2003363>.
94. Kim, Y., Choi, S., Shin, D., and Cho, K.H. (2017). Quantitative evaluation and reversion analysis of the attractor landscapes of an intracellular regulatory network for colorectal cancer. *BMC Syst. Biol.* 11, 45. <https://doi.org/10.1186/s12918-017-0424-2>.
95. Bollobas, B., Borgs, C., Chayes, J., and Riordan, O. (2003). Directed scale-free graphs. *Siam Proc S*, 132–139.
96. Brandes, U. (2001). A faster algorithm for betweenness centrality. *J. Math. Sociol.* 25, 163–177. <https://doi.org/10.1080/0022250x.2001.9990249>.
97. Golbeck, J. (2015). Analyzing networks. Elsevier, 221–235. <https://doi.org/10.1016/b978-0-12-801656-5.00021-4>.
98. Bonacich, P. (1987). Power and Centrality: A Family of Measures. *Am. J. Sociol.* 92, 1170–1182.
99. Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika* 18, 39–43. <https://doi.org/10.1007/BF02289026>.
100. Freeman, L.C. (1978). Centrality in social networks conceptual clarification. *Soc. Networks* 1, 215–239. [https://doi.org/10.1016/0378-8733\(78\)90021-7](https://doi.org/10.1016/0378-8733(78)90021-7).
101. Goh, K.I., Kahng, B., and Kim, D. (2001). Universal behavior of load distribution in scale-free networks. *Phys. Rev. Lett.* 87, 278701. <https://doi.org/10.1103/PhysRevLett.87.278701>.
102. Boldi, P., and Vigna, S. (2013). Axioms for Centrality. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1308.2140>.
103. Page, L., Brin, S., Motwani, R., and Winograd, T. (1999). The PageRank Citation Ranking: Bringing Order to the Web. Technical Report Stanford InfoLab. <http://ilpubs.stanford.edu:8090/422/>.
104. Jimmy, J.; Geoffrey (2016). Layer Normalization. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1607.06450>.
105. Humplik, J., Galashov, A., Hasenclever, L., Pedro, Y., and Heess, N. (2019). Meta reinforcement learning as task inference. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1905.06424>.
106. Clevert, D.-A.e., Unterthiner, T., and Hochreiter, S. (2016). Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). Preprint at arXiv. <https://doi.org/10.48550/arXiv.1511.07289>.
107. Foerster, J.N., Assael, Y.M., de Freitas, N., and Whiteson, S. (2016). Learning to Communicate with Deep Multi-Agent Reinforcement Learning. *Adv. Neur. In* 29.
108. Das, A., Gervet, T.e., Romoff, J., Batra, D., Parikh, D., Rabbat, M., and Pineau, J. (2020). TarMAC: Targeted Multi-Agent Communication. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1810.11187>.
109. Kim, W., Park, J., and Sung, Y. (2020). Communication in Multi-Agent Reinforcement Learning: Intention Sharing. In *International Conference on Learning Representations*.
110. Heess, N., Tb, D., Sriram, S., Lemmon, J., Merel, J., Wayne, G., Tassa, Y., Erez, T., Wang, Z., S., et al. (2017). Emergence of Locomotion Behaviours

# Cell Reports Methods

## Article



- in Rich Environments. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1707.02286>.
111. Béal, J., Montagud, A., Traynard, P., Barillot, E., and Calzone, L. (2018). Personalization of Logical Models With Multi-Omics Data Allows Clinical Stratification of Patients. *Front. Physiol.* *9*, 1965. <https://doi.org/10.3389/fphys.2018.01965>.
112. Hu, Y., Chen, C.H., Ding, Y.Y., Wen, X., Wang, B., Gao, L., and Tan, K. (2019). Optimal control nodes in disease-perturbed networks as targets for combination therapy. *Nat. Commun.* *10*, 2180. <https://doi.org/10.1038/s41467-019-10215-y>.
113. Glover, F. (1967). Maximum matching in a convex bipartite graph. *Nav. Res. Logist. Q.* *14*, 313–316. <https://doi.org/10.1002/nav.3800140304>.

**STAR★METHODS**

**KEY RESOURCES TABLE**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Chemicals, peptides, and recombinant proteins</b>		
Trametinib	APEXBIO	Cat#A3018
Afatinib	Selleckchem	Cat#S1011
Palbociclib	Selleckchem	Cat#S4482
<b>Critical commercial assays</b>		
RNA-spin kit	INTRON	Cat#17211
QuantSeq 3' mRNA-Seq Library Prep Kit	Lexogen	Cat#015.24
<b>Deposited data</b>		
Gene expression data	This paper	GSE184731
Harmonizome	Rouillard et al. <sup>81</sup>	SCR_016176
Kyoto Encyclopedia of Genes and Genomes	Kanehisa et al. <sup>49</sup>	SCR_012773
Genomics of Drug Sensitivity in Cancer 2	Iorio et al. <sup>35</sup>	SCR_011956
Catalog Of Somatic Mutations In Cancer	Tate et al. <sup>39</sup>	SCR_002260
Pathway Commons	Cerami et al. <sup>72</sup>	SCR_002103
<b>Experimental models: Cell lines</b>		
SW48	ATCC	CCL-231
SW620	ATCC	CCL-227
BT20	Korean Cell Line Bank(KCLB)	KCLB #: 60061
NCI-N87	Korean Cell Line Bank(KCLB)	KCLB #: 60113
MKN45	Korean Cell Line Bank(KCLB)	KCLB #: 80103
BT474	Korean Cell Line Bank(KCLB)	KCLB #: 60062
<b>Software and algorithms</b>		
GRAY	This paper	<a href="https://github.com/hanyh0807/GREY">https://github.com/hanyh0807/GREY</a> <a href="https://doi.org/10.5281/zenodo.10989715">https://doi.org/10.5281/zenodo.10989715</a>
Tensorflow version 1.15.5	Google	<a href="https://www.tensorflow.org/">https://www.tensorflow.org/</a>
Graph_nets python package	Peter et al. <sup>39,44</sup>	<a href="https://github.com/deepmind/graph_nets/tree/master">https://github.com/deepmind/graph_nets/tree/master</a>
SRMF	Wang et al. <sup>46</sup>	<a href="https://github.com/linwang1982/SRMF">https://github.com/linwang1982/SRMF</a>
DeepCDR	Liu et al. <sup>10</sup>	<a href="https://github.com/kimmo1019/DeepCDR">https://github.com/kimmo1019/DeepCDR</a>
LOBICO	Knijnenburg et al. <sup>47</sup>	<a href="https://github.com/TianxiaoNYU/LOBICO-Matlab">https://github.com/TianxiaoNYU/LOBICO-Matlab</a>
MATLAB R2019b	MathWorks	<a href="https://kr.mathworks.com/">https://kr.mathworks.com/</a>
pycma v3.1.0	Nikolaus et al. <sup>45</sup>	<a href="https://github.com/CMA-ES/pycma">https://github.com/CMA-ES/pycma</a>
STAR	Dobin et al. <sup>82</sup>	<a href="https://github.com/alexdobin/STAR">https://github.com/alexdobin/STAR</a>
DESeq2	Love et al. <sup>83</sup>	<a href="http://www.bioconductor.org/packages/release/bioc/html/DESeq2.html">http://www.bioconductor.org/packages/release/bioc/html/DESeq2.html</a>
Signal flow analysis	Lee et al. <sup>63</sup>	<a href="https://github.com/dwgoon/sfa">https://github.com/dwgoon/sfa</a>
GDSCTools	Cokelaer et al. <sup>84</sup>	<a href="http://github.com/CancerRxGene/gdsc_tools">http://github.com/CancerRxGene/gdsc_tools</a>
Hierarchical HotNet	Reyna et al. <sup>85</sup>	<a href="https://github.com/raphael-group/hierarchical-hotnet">https://github.com/raphael-group/hierarchical-hotnet</a>
The PathLinker app	Gil et al. <sup>62</sup>	<a href="https://apps.cytoscape.org/apps/pathlinker">https://apps.cytoscape.org/apps/pathlinker</a>

### RESOURCE AVAILABILITY

#### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Kwang-Hyun Cho ([ckh@kaist.ac.kr](mailto:ckh@kaist.ac.kr))

#### Materials availability

This study did not generate new unique reagents.

#### Data and code availability

- Gene expression data have been deposited at GEO and are publicly available as of the date of publication. Accession number is listed in the [key resources table](#).
- All original code has been deposited at GitHub and is publicly available as of the date of publication. The GitHub repository and DOI are listed in the [key resource table](#).
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

### EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

#### Setup for in-house experimental validations

The relevant cell cultures, SW48 (Colon), SW620 (Colon), BT20 (Breast), NCI-N87 (Breast), MKN45 (Stomach), and BT474 (Stomach), were obtained from Korean Cell Line Bank and the American Type Culture Collection (ATCC). All cell lines were cultured in Dulbecco's modified Eagle's medium (WeiGENE Inc., Gyeongsan, Republic of Korea) with 10% fetal bovine serum (FBS, WeiGENE Inc.) and antibiotics (100 U/ml of penicillin, 100  $\mu$ g/mL streptomycin, and 0.25  $\mu$ g/mL of Fungizone) (Life Technologies Corp., Carlsbad, CA) at 37°C in a humidified atmosphere containing 5% CO<sub>2</sub>.

To obtain reagents, trametinib (an inhibitor of *MAP2K1* and *MAP2K2*, also known as *MEK1* and *MEK2*, respectively) was purchased from APEX BIO (Boston, MA, USA). afatinib (an inhibitor of *EGFR* and *ERBB2*) and palbociclib (an inhibitor of *CDK4* and *CDK6*) were obtained from Selleckchem (Houston, TX, USA).

To conduct cell growth assays, cells were seeded into a 96-well plate at a density of 3–7 × 10<sup>3</sup> cells/well in growth medium, incubated for 24 h and then treated with the indicated drugs. Following incubation of the plates for 72–120 h, relative cell viability was measured. After seeding, cells were imaged using IncuCyte ZOOM. To assess cell growth, the average area of each cell was determined at each time point using the IncuCyte ZOOM analysis software. Images were captured at 3 h intervals from three separate regions per well with a 20× objective.

To conduct RNA isolation and bulk RNA-seq analysis, total RNA was extracted using an RNA-spin kit (INTRON), and RNA quality was assessed by Agilent 2100 bioanalyzer using the RNA 6000 Nano Chip (Agilent Technologies). RNA quantification was performed using ND-2000 Spectrophotometer (Thermo, Inc.), and 500 ng of RNA per sample were used for library construction. The library construction was performed using the QuantSeq 3' mRNA-Seq Library Prep Kit (Lexogen, Inc.), and the library was single-end sequenced on an Illumina NextSeq 500 (Illumina, Inc.). The reads were aligned using STAR,<sup>82</sup> and the read counts were estimated by featureCounts.

Differentially expressed gene (DEG) analysis was performed by using significance analysis of microarrays method implemented in the DESeq2 package. We calculated the differentially expressed genes in log<sub>2</sub>-fold changes of expression levels measured in both sensitive and resistant cell lines for each type of cancer (colon, breast and stomach) using the DESeq2<sup>83</sup> R package. The cutoff values were set with the 'adjusted *p*-value' under 0.05 and the base mean over 100.

### METHOD DETAILS

#### Constructing DRCNs

To efficiently elucidate the construction of the Pan-Cancer Core Network (PCCN) and the Drug Response Core Network (DRCN), we integrated a comprehensive list of cancer-related genes from multiple high-throughput data sources. These included oncogenic genes and compound targets from ten major signaling pathways identified in The Cancer Genome Atlas<sup>66</sup> (TCGA), such as *RTK/RAS*, *Nrf2*, *PI3K*, *TGFβ*, *Wnt*, *Myc*, *p53*, Cell cycle, Notch and Hippo. Additionally, we incorporated frequently amplified or deleted mutations,<sup>71</sup> the NCI-MATCH basket trial markers,<sup>38</sup> notably co-occurring or mutually exclusive mutations,<sup>86</sup> and genes from 'Pathways in cancer' (hsa05200) of the KEGG pathway.<sup>49–51</sup> Through these integrations, we identified a total of 2,822 pan-cancer genes.

Using the PathLinker algorithm,<sup>61,62</sup> these pan-cancer genes were each used as a source node to connect to all other genes, employing interconnecting pathways derived from the Omnipath network structure database.<sup>36</sup> The connected pathways, comprising 3,615 nodes after exclusion of paths with overlapping sources and targets, represented comprehensive interconnections among the pan-cancer related genes. To refine this network and extract the core structure, we pruned leaf nodes and redundant inter-nodes, smoothing the network using a network propagation algorithm,<sup>87</sup> and eventually extracted a core structure with 2,574 nodes and 5,992 links.



For the construction of the DRCN, we applied hierarchical HotNet<sup>85</sup> and Edmonds' algorithm<sup>88</sup> to extract a core network that captures the essential genes and pathways involved in drug responses. This involved gathering genomic features such as mutations, copy number alterations, and expression levels of genes from various databases including the Cell Model Passport,<sup>89</sup> OncoKB, and COSMIC. These features were used to compute z-scores in analysis of variance (ANOVA) of the GDSCTools<sup>84</sup> package, which then served as vertices in the hierarchical HotNet algorithm, identifying the most statistically significant clusters of genes. We connected these clusters using Edmonds' algorithm with link weights derived from inverse values of joint similarity scores calculated through HotNet. This process culminated in a DRCN that minimized network size while retaining critical drug response-related genes and pathways, further enriched with phenotypic markers such as 'proliferation' and 'apoptosis' nodes from the CancerGeneNet.<sup>37</sup>

To extract a core structure from the pan-cancer related network, we took the following four steps: first, the nodes that only have incoming links or outgoing links were pruned. Second, pan-cancer related genes were smoothed across the network by the network propagation algorithm shown in the Equation 3 suggested by Vanunu et al.<sup>90</sup>

$$P^{t+1} = (1 - r)WP^t + rP^0 \quad (\text{Equation 3})$$

Where  $P^t$  is a probability vector at time step  $t$  in which the  $i$ -th element holds the probability of a random walker that will be placed at that time step. The initial probability vector,  $P^0$ , is a vector with a probability of 1 for the cancer related genes and 0 for the other genes, which were normalized by a number of pan-cancer related genes, such that the vector sums to 1.  $W$  is a column-normalized adjacency matrix of the pan-cancer related network.  $r$  is a probability of restarting from the pan-cancer related genes at every time step, which was set as 0.3 in our extraction. Third, an iterative random walker with a limited number of transitions walked around the network. The number of iterations was set to 10,000 and the maximum number of transitions was set to 30. Fourth, the pan-cancer core network was constructed by selecting a sub-network that was visited by a larger number of walkers than a specific cutoff value, which was set to 12. As a result, the pan-cancer core network was constructed with 2,574 nodes and 5,992 links.

### Weighted sum Boolean logics

To simulate drug responses of a given DRCN, we adopted "weighted sum Boolean logic" by modifying the simple majority rule suggested by Font-Clos et al.<sup>91</sup> A state of a node  $k$  in a network with  $N$  nodes,  $s_k$ , can be either 1 (activated) or  $-1$  (inactivated). The state of a network at a discrete time point,  $t$ , can be defined as a state vector,  $V_s^t$ , which is composed of the  $N$  states of its nodes, such as  $s_k^t$ , at a certain time point. The tendency of each node state without any incoming signals can be represented as a basal level vector,  $V_b$ , which usually has a value of zero. Relative influences of each incoming link to a node can be a value in a specific range, which includes integers between  $-5$  and  $5$  in our work. Positive numbers represent activating regulations and negative numbers refer to inhibitory regulations. A relation matrix can be defined as a  $N$ -by- $N$  connectivity matrix of the network, wherein each element  $(M_R)_{ij}$  represents the relative influence of a node  $j$  on a node  $i$ . The network state at the next time point,  $V_s^{t+1}$ , can be decided by the sign of element-wise multiplication of a weighted sum vector,  $V_w$ , computed by the matrix operation shown in Equation 4. If some elements of the weighted sum vector are zero, the next state of the corresponding nodes follows the current state of each node. Network nodes are synchronously updated at each discrete time point:

$$V_s^{t+1} = \text{sign}(V_w) = \text{sign}(M_R \times V_s^t + V_b) \quad (\text{Equation 4})$$

The attractors, converged steady states of the network, can represent the biological phenotypes by analyzing average states of each node within the attractor states.<sup>92,93</sup> Average activity of each molecule represented by a node state of an attractor is computed as suggested by Kim et al.<sup>94</sup> The average state of each node is computed as the average of the node state within a specific attractor multiplied by the proportional ratio of initial states which converged into that attractor. If an attractor does not converge into a single state and cycles a set of states, the average state of each node in that attractor is computed as a sum of node states in the cycle divided by the period of the cycle. The converging ratio of each attractor is approximately measured by repeated enumeration of the network state starting from a random initial network state to a particular attractor.

### Synthetic environments for training GRAY

We designed a synthetic environment to train GRAY for Boolean network optimization through meta-reinforcement learning. A randomly generated scale-free network<sup>95</sup> of 20~40 nodes, with edge weights varying within  $[-5,5]$  except 0, were assigned to each environment. We generated 100,000 synthetic environments. At the start of each episode, one of 100,000 environments was randomly selected, and its virtual perturbation data was generated. The virtual perturbation data contains the fold change of the average activity for each node between the first attractors that were reached from 128 random initial states, and the second attractors that were reached starting from the first attractor after fixing several random nodes to 1 or  $-1$  to imitate a drug treatment. By varying the fixed nodes to compute the second attractor, 16 ground truth datasets were generated for each environment. The fitness is defined as the average of the Pearson's correlation coefficients (PCC) between the ground truth datasets and the fold change values computed by applying the edge weights ( $A^t$ , obtained edge weights at step  $t$ ) obtained from GRAY onto the network. The fixed node sets used to compute fold change values for  $A^t$  were the same as the ground truth datasets. The reward was given as the

observed improvement<sup>33</sup> as follows:

$$R^t = \max \left\{ f^t - \max_{i < t} (f^i), 0 \right\} \quad (\text{Equation 5})$$

Here,  $f^t$  is the fitness and  $R^t$  is the reward at step  $t$ . The goal of GRAY against a given environment is to maximize the discounted cumulative sum of reward.

### Neural network structure of GRAY

GRAY consists of three components: the input, the submodule, and the main layer (Figure S1A). The input layer has a graph module for embedding networks and a message module for communications between agents. The graph module consists of two graph neural networks (GNN), which is a class of neural networks tailored for processing input represented as graphs. The first ( $G$ ) is for the main layer and the global state estimator of the submodule layer, and the second ( $G^C$ ) is for the context estimator of the submodule layer. The message module ( $G^m$ ) is also a GNN and is used for all modules in both the submodule and the main layer.

In each step, GNNs in the graph module receives eight centrality measures for nodes (betweenness,<sup>96</sup> degree,<sup>97</sup> eigenvector,<sup>98</sup> katz,<sup>99</sup> closeness,<sup>100</sup> load,<sup>101</sup> harmonic,<sup>102</sup> pagerank<sup>103</sup>) and three values for edges (weight, sign of each edge in the network of the environment, and edge betweenness centrality<sup>96</sup>) as input, and updates nodes, edges, and the global feature by iteratively aggregating information from neighboring nodes, edges, and the entire graph. The weight in the input for the edge is from the output of the preceding step. The global feature represents a graph-level attribute. The message module initially receives a zero-vector input for all nodes and edges at the beginning of an episode, and generates nodes, edges, and the global feature. From the second step, the message module receives nodes, edges, global features, and edge weights generated from the preceding step as input, and updates its features. The outputs of the GNNs in the graph module and the message module are reorganized for each edge by concatenating the global feature, the edge feature, and the features of sender and receiver nodes of each edge (Figure 3D). Each reorganized feature represents the state of each edge. We use this as the feature vector to be fed into the subsequent modules. The reorganized feature of the  $i$ th edge from each of the three GNNs ( $G, G^C, G^m$ ) are as follows:

$$L_i = (n_{i,s}, n_{i,t}, l_i, g), L_i^C = (n_{i,s}^C, n_{i,t}^C, l_i^C, g^C), L_i^m = (n_{i,s}^m, n_{i,t}^m, l_i^m, g^m) \quad (\text{Equation 6})$$

$$L = L_{1:k}, L^C = L_{1:k}^C, L^m = L_{1:k}^m \quad (\text{Equation 7})$$

Here,  $L_i, L_i^C, L_i^m$  are the reorganized features of  $i^{\text{th}}$  edge from  $G, G^C, G^m$  respectively.  $n_{i,s}, n_{i,t}$  are the node features of the source and target node of  $i^{\text{th}}$  edge, respectively.  $l_i$  is the edge feature of  $i^{\text{th}}$  edge.  $g$  is the global feature of a network.  $L, L^C, L^m$  are the reorganized features for all edges (1 to  $k$ ) in the network from each module,  $G, G^C, G^m$ , respectively. Features without any superscript are from  $G$ , and features with superscript C and m are from  $G^C$  and  $G^m$ , respectively. Modules in the submodule and main layer receive the reorganized features as an input (Figure S1A).

A submodule layer consists of a context estimator module and a global state estimator module. Both modules are one-layer LSTMs with layer normalization.<sup>104</sup> In each module, there is one LSTM for each reorganized feature (i.e., one LSTM for each edge). LSTMs in each module share their parameters (weights of LSTM) but maintain independent hidden states. The context estimator is a module that helps the agent to quickly adapt to the selected environment by estimating information about the environment.<sup>105</sup> During training, the true edge weights of the synthetic network from the selected environment in each episode are known. Thus, we can use the true edge weights as a target for the context estimator.  $L^C, L^m$ , and the edge weight and the reward from the preceding step are set as inputs to the LSTM, and the 11-dimensional logits (edge weights are between  $-5$  and  $5$ ) are generated through two dense layers with exponential linear unit (ELU) activation functions.<sup>106</sup> Supervised training is done using cross entropy loss with respect to the true edge weight of the network from the selected environment in each episode. The global state estimator is a module used to provide information that can help learn cooperative policy in order to encourage the agents to cooperate and perform efficient exploration in high-dimensional search space, inspired by Foerster et al.,<sup>107</sup> Das et al.,<sup>108</sup> and Kim et al.<sup>109</sup> Since the output (edge weight) of an agent (main layer) for each edge collectively determines the fitness, the output of all the agents needs to be well-organized. We trained the communication protocol to gather  $L, L^m$ , and the edge weight of the current step into the global state estimator module, and then predict the feature ( $L$ ) of the next step. Since the input for the subsequent step is determined by the outputs of all agents, it is necessary to infer the plans of other agents in order to predict the features of the subsequent step. Therefore, the global state estimator provides inferred information of the outputs of the other agents. The global state estimator module predicts the average and standard deviation for the features of the subsequent step and is trained to increase the log probability for the actual value of that feature through supervised training.

In the main layer (agent),  $L, L^m$ , the edge weight and the reward from the preceding step, and the activations of the second-to-last layer of the two submodules are set as inputs to a two-layer LSTMs with layer normalization. Similar to the submodule layer, there is a two-layer LSTM for each reorganized feature. LSTMs in this layer share parameters, while maintaining independent hidden states. Output and value are generated through an independent dense layer with an ELU activation function. The output is 11-dimensional logits (to represent  $-5$  to  $5$ ) which estimates the next weight of an edge assigned to the agent, and the value is an expected discounted reward of the current input.

### Training procedure

To train a derivative-free search rule for optimizing the edge weights of Boolean networks, GRAY uses meta-reinforcement learning.<sup>32,42</sup> GRAY is trained on various synthetic environments, where the environment involves optimizing edge weights of random synthetic networks for ground truth datasets. At the beginning of each episode, an environment is sampled out of 100,000 synthetic environments and fixed until the end of the episode. Each episode consists of 80 steps. Hidden states of all LSTM in GRAY are set to the zero vector at the beginning of the episode. The initial state ( $S_1$ ) of the episode is a combination of a random parameter vector ( $A_0$ ) and zero ( $R_0$ ). The sign of each parameter in  $A_0$  is kept the same as that of the corresponding link in the network of the selected task. All episodes have 80 steps and each iteration of the input and output processes of an episode is saved to episode buffer when the episode is finished.

GRAY is trained using 10 workers and one master, where each worker spends 8,000 episodes in the supervised pretraining phase and 12,000 episodes in the reinforcement learning phase. Each worker independently collects episodes and computes gradients. The master collects gradients from each worker and then averages them. The averaged gradients are applied to the neural network of the master, and the updated parameters are redistributed to each worker.

The training procedure for the main layer consists of two phases: a supervised pre-training phase based on directly predicting the true edge weights of the selected environment and a reinforcement learning phase for maximizing the discounted cumulative sum of rewards using the distributed proximal policy optimization (DPPO) algorithm.<sup>110</sup> In the supervised pretraining phase, we randomly replaced 50% elements of the new parameters with the true parameters that have a probability of 0.05 for each step. Whenever 64 episodes are collected, GRAY is trained using Adam on 4 minibatch samples (4 episodes) for one or three epochs, and the buffer is cleared. In the supervised pretraining phase, all neural networks are trained using supervised learning. However, the parameters of  $G^C$  and the submodule layer are updated for three epochs for every 64 episodes and the remaining parameters ( $G$ ,  $G^m$ , main layer) are updated for one epoch for every 64 episodes. In the reinforcement learning phase,  $G^C$  and submodule layer are trained using supervised learning for one epoch for every 64 episodes. Moreover, the remaining parameters ( $G$ ,  $G^m$ , main layer) are updated using the DPPO algorithm for three epochs for every 64 episodes. Detailed hyperparameters are summarized in [Table S2](#).

### Selecting six cell lines for in-house experimental validations

To validate the molecular-scale interpretability of the DRCNs constructed by the gray box framework ensuring that the DRCNs can effectively capture significant drug response dynamics across a diverse range of cancer cell lines, with an affordable number of experimental validations, we have designed experiments by selecting specific cell lines based on a set of well-defined criteria that allows us to represent the broader landscape of drug response dynamics in pan-cancer settings.

Our selection criteria was carefully designed to ensure that the chosen cell lines could serve as representative samples for validating the predictive capabilities of our DRCNs. These criteria include the following four considerations. First, we selected three of the most commonly occurring cancer types—colon, breast, and stomach cancers. Each of these cancer types has unique gene families associated with them, which makes each cancer type having distinctive molecular machinery related with drug responses to others. Second, we chose three representative U.S. Food and Drug Administration (FDA)-approved targeted cancer drugs—afatinib (EGF signaling), trametinib (MEK signaling), and palbociclib (CDK signaling). These drugs act on distinct signaling pathways that are most commonly altered in various cancers, including our three selected types. Third, we selected two groups of cell lines corresponding to those ranked in the top 15 for sensitivity and resistance, respectively, to each of the three selected drugs. These rankings were based on the area under the dose-response curve (AUC) values from the GDSC2 database. Fourth, among the selected sensitive and resistant cell lines, we chose one cell line from each group that shared the same genetic markers as suggested by the Molecular Analysis for Therapy Choice (MATCH) basket trials conducted by the National Cancer Institute (NCI-MATCH).

As a result, we have used the mutational drug response markers from the EAY131 NCI-MATCH screening trial as follows: EGFR activating mutations and ERBB2 activating mutations for afatinib, BRAF fusions, BRAF V600E or V600K mutations, GNAQ or GNA11 mutations and NF1 mutations for trametinib, CCND1 or CCND2 or CCND3 amplifications, CDK4 or CDK6 amplifications for Palbociclib. For the six selected cancer cell lines, both BT-20 and BT-474 have the same mutation profiles for EGFR and ERBB2 (mutated EGFRs and not mutated ERBB2s), while they show different afatinib responses (sensitive BT-474 and resistant BT-20). Both NCI-87 and MKN45 have the same mutation profiles of no EGFR mutations, while they show different afatinib responses (sensitive NCI-87 and resistant MKN45). Both SW48 and SW620 have the same mutation profiles of no ERBB2 mutations, while they show different afatinib responses (sensitive SW48 and resistant SW620). Both SW48 and SW620 have no mutations on BRAF, GNA11, GNAQ and NF1, while they show different Trametinib responses (sensitive SW48 and resistant SW620). Both SW48 and SW620 have no mutations on CCND1, CDK4 and CDK6, while they show different Palbociclib responses (sensitive SW48 and resistant SW620). This approach allowed us to address the challenge of classifying drug response differences and interpret the resulting molecular dynamics, which remain obfuscated by classical marker-based methods. By adhering to these selection criteria, we aimed to ensure that our experiments could effectively represent and validate the predictive capabilities of the DRCNs optimized by the gray box framework. These experiments serve as representative examples of how our framework can predict drug responses at the molecular level across diverse cancer types and targeted drug categories, surpassing the predictions made by traditional marker-based methods.

In summary, we have been able to demonstrate the applicability of the gray box framework by strategically selecting representative drugs and cell lines for validation. This approach was chosen to avoid the need for experimental validation across all 300 cell lines used in our training set, as it would not provide additional insights beyond what has already been achieved with our carefully designed experiments.

### Integrating cell line mutation profiles with functional effect

The mutational profile of each cell line is curated from the preprocessed COSMIC database, which was processed by Béal et al.<sup>111</sup> The functional effect of each mutation is decided by comparing oncogenic and tumor suppressive q-values suggested by Tokheim et al.<sup>40</sup> If the oncogenic q-value of a specific gene is smaller than the tumor suppressive q-value, then the functional effect of the mutation on that gene is considered oncogenic. Therefore, the functional effect of the mutation on the gene is treated as a gain of function. In the case that the tumor suppressive q-value is smaller than the oncogenic q-value, the functional effect of the mutation on the gene is treated as a loss of function. These functional effects of mutations are then cross-validated with functionally associated genes in Harmonizome.<sup>81</sup>

### Maximum matching-based mutation data projection

Several mutated genes of cancer cell lines exist within the pan-cancer network but not in the DRCN. Thus, we needed to approximate the effects of each mutation on the DRCN by leveraging structural information. Hu et al.<sup>112,113</sup> defined the structural control configuration (SCC) of a network model to approximate the effects of certain drugs on the downstream genes. We employed this method to approximate the effects of mutations. An SCC of a network model is a maximally matched graph within the network, although there exist many different SCCs within a large network. SCC consists of elementary paths, which are cascades connected by maximum links, and elementary cycles, which are cycles connected by maximum links. Additional links indicate those that connect elementary paths and elementary cycles. We generated 1,000 SCCs until the proportion of new maximum matching links reached zero. Next, we extracted all possible simple paths consisting of maximum matching links and additional links from mutated genes to other genes within the DRCN of each SCC. We utilized the sign information of each link from the paths among the mutated genes as well as the genes within the DRCN to obtain the effects of each mutation on its downstream genes.

### Incorporating mutations and drugs onto the simulation

During simulations of drug responses of the core network, the effect of a mutation on a node inside the core network was reflected by fixing the state of the corresponding node to either 1 or -1 for a gain of function or a loss of function mutation, respectively. The effects of mutations on a node outside of the core network were projected into the core network using maximum matching. These effects were then applied onto their corresponding elements of the basal level vector  $V_b$ , 1 or -1 for a gain of function or a loss of function mutation, respectively. To estimate the inhibition effect of drugs on the core network, we applied a two-step drug response simulation. First, the network converged into attractors with the given mutation profile. Second, the network converged again from the prior attractor states to new attractors while the drug target nodes were additionally fixed to -1. If some of the prior attractor states were cycles, the first state of each cycle was treated as the prior attractor state in the second step of the drug response simulation. The difference between the average activity of the phenotype node in the two states (the new attractor and the prior attractor) were compared with the measured area under the viability curve (AUC) value collected from GDSC2.

### Signal flow analysis of DRCN perturbations

Signal flows after the DRCN perturbations are analyzed by an algorithm suggested by Lee et al.<sup>63</sup> Signal flows from a node are defined by adding the incoming signal flow to the product of the node activity change with the sign (positive or negative) of the outgoing links. Effects of mutations on signal flows are incorporated differently depending on whether the mutations were on a node inside the activity difference sub-network or not. The effect of a mutation on a node inside the activity difference sub-network was treated as pinning that node to a state, such that incoming links were deleted and a constant signal flows out from that node. The effect of a mutation from outside the activity difference sub-network was incorporated by propagating the change in signal flow caused by a mutation through exclusive simple paths computed by the maximum matching algorithm.

## QUANTIFICATION AND STATISTICAL ANALYSIS

For comparison between two groups, two-sided Student's t test was used. All pathway enrichment analyses were conducted using WebGestalt<sup>48</sup> and KEGG.<sup>49–51</sup> ANOVA was conducted using GDSCTools.<sup>84</sup> Statistical details can be found in the respective figure legends, results, and the [STAR Methods](#) section.