

Title: Dissecting cancer resistance to therapies with cell-type-specific dynamic logic models

Authors: Federica Eduati¹, Victoria Doldàn-Martelli^{1,2}, Bertram Klingner^{3,4}, Thomas Cokelaer¹, Anja Sieber^{3,4}, Fiona Kogera⁵, Mathurin Dorel^{3,4,6}, Mathew J Garnett⁵, Nils Blüthgen^{3,4,6*}, Julio Saez-Rodriguez^{1,7*}

One Sentence Summary: Signaling models trained to perturbation data untangle therapy resistance in colon cancer

Affiliations:

¹European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, UK

²Departamento de Física de la Materia Condensada, Condensed Matter Physics Center (IFIMAC) and Instituto Nicolás Cabrera, Facultad de Ciencias, Universidad Autónoma de Madrid, Madrid, Spain

³Institute of Pathology, Charité - Universitätsmedizin Berlin, Berlin, Germany

⁴Integrative Research Institute (IRI) Life Sciences and Institute for Theoretical Biology, Humboldt-Universität zu Berlin, Berlin, Germany

⁵Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, UK

⁶Berlin Institute of Health (BIH), Berlin, Germany

⁷Joint Research Centre for Computational Biomedicine (JRC-COMBINE), RWTH Aachen University, Faculty of Medicine, Aachen, Germany

* co-corresponding authors: nils.bluthgen@charite.de, saezrodriguez@gmail.com

Abstract:

Therapies targeting specific molecular processes, in particular kinases, are major strategies to treat cancer. Genomic features are commonly used as biomarkers for drug sensitivity, but our ability to stratify patients based on these features is still limited. As response to kinase inhibitors is a dynamic process affecting largely signal transduction, we investigated the association between cell-specific dynamic signaling pathways and drug sensitivity. We measured 14 phosphoproteins under 43 different perturbed conditions (combination of 5 stimuli and 7 inhibitors) for 14 colorectal cancer cell-lines, and built cell-line-specific dynamic logic models of the underlying signaling network. Model parameters, representing pathway dynamics, were used as features to predict sensitivity to a panel of 27 drugs. This analysis revealed associations between cell-specific signaling pathways and drug sensitivity for 14 of the drugs, 9 of which have no genomic biomarker. Following one of these associations, we validated a drug combination predicted to overcome resistance to MEK inhibitors by co-blockade of GSK3. These results underscore the value of perturbation-based studies to find biomarkers and combination therapies complementing those based on a static genomic characterization.

[Main Text:]

Introduction

Patient response to anticancer therapies is extremely variable and understanding the reasons for this variability is a major challenge in cancer research. One approach to address this problem is to identify biomarkers which correlate with therapy response. However, except for a few examples, no efficient biomarkers are available (1). The problem of finding biomarkers has become even more important with the advent of targeted cancer therapies that are designed to affect specific molecular changes in cancer cells which drive the cancer, and which provide a larger number of treatment options. However, biomarkers that can be used to stratify patients for targeted drugs also remain largely elusive (2).

The most common approaches for patient stratification are currently based on genomic biomarkers, typically consisting of either expression or mutation of specific genes. The popularity of those biomarkers has been favoured by the advancements of sequencing technology and their subsequent decrease in cost, and, for some drugs, they have shown strong potential (3–5). However, for many other cases no efficient genomic markers for patient stratification exist, and for those where markers exist, they have rather low power due to the complex nature of cancer. Furthermore, their actual clinical significance for precision oncology is questionable (1).

Many modern drugs target signaling molecules, and elicit a response of the signaling network. Therefore, investigating the functioning of signaling pathways can provide new insights into these mechanisms and may help to unveil new therapeutic strategies and biomarkers (6, 7). Understanding how drugs affect the signaling network as a whole is particularly important as signaling dynamic differs not only between tumors of different tissue, but also between tumors of the same tissue. For instance, for colon and liver cancer cell-lines, the dynamics of signaling networks have been shown to differ strongly between different cells of the same tissue of origin (8–10).

Furthermore, models parameterized and experimentally calibrated using a specific cell-line have been used to define pathway dynamic biomarkers of therapeutic

outcome, such as drug sensitivity (11) or patient survival (12), based on model simulations in breast cancer and neuroblastoma, respectively.

The understanding of mechanisms of cellular response (network structure and dynamic behaviour), can be also useful to tackle the development of compensatory signaling mechanisms of drug resistance (13), which is a recurrent problem for targeted therapy and is challenging to predict based only on genomic information. Signal transduction results from the integration of complex dynamic networks which can be modulated by mutations in oncogenes or tumor suppressors. The complex wiring confers a robustness to the cells that helps them to escape most single-agent-targeted treatments. Based on this idea, modeling of signaling pathways has been recently used to suggest combinatorial targeted therapies to effectively block multiple molecular pathways (8, 14, 15): by integration of prior pathway knowledge with experimental observations cell-line-specific models were built, which could be used to simulate and thus prioritize possible combinatorial perturbation experiments.

In this paper, we investigate to what extent dynamic interactions between different signaling pathways play a role in characterizing the specific cellular responses to drugs and suggesting targeted combinatorial therapies. Furthermore we were interested to assess how these dynamic features fare against static genomic traits as markers of drug response. We do so by characterizing cell-type-specific models for a panel of 14 different colorectal cancer (CRC) cell-lines that are integrated with a large-scale drug screening (16). We found associations with model parameters for 14 of the 27 drugs targeting our pathways of interests, for 9 of which there were no genomic biomarkers. These associations was used to define pathway dynamic biomarkers and interesting combinatorial therapies.

Results

To identify biomarkers of drug response from dynamic logic models, we proceed as follows (**Fig. 1**). First, cell-line-specific models were built for 14 colorectal cancer (CRC) cell-lines. Each model is a set of logic ordinary differential equations (17) based on a prior knowledge network (PKN) manually curated from literature (**Fig. 1A**) that is refined using CellNOpt (18) based on experimental data. Second, sensitivity data for our 14 cell-lines in response to drugs targeting nodes in the PKN

or first neighbours were retrieved from the Genomics of Drug Sensitivity in Cancer (GDSC) panel (16) (**Fig. 1B**). Third, we investigated correlations between model parameters and drug response using Elastic Net (19) to select strong associations (**Fig. 1C**). All steps will be detailed in the following sections.

Strategy for model optimization: a toy example

Our starting point is a prior knowledge network (PKN), which is generic in the sense that it contains information about different cell types and is hence not clear about how many of the contained interactions are functional in a given cell type. To identify cell-specific functional components, the PKN was trained as a logic model to cell-specific data. The dynamics of the system were modelled using a formalism based on logic ordinary differential equations (ODEs) (17), where a set of ODEs (one for each species in the model) was derived from the logic structure using a continuous update function (see **Materials and Methods**). Each species (i.e. node in the network) was characterized by a parameter τ_i (with $i=1, \dots, N$ where N is the number of species in the model) representing its life-time ($\tau=0$ meaning that the node is not functional), and each regulatory interaction was defined by a sigmoid function, where a parameter $k_{i,j}$ characterizes the strength of the regulation of species i dependent on species j ($k_{i,j}=0$ meaning no regulation).

In order to be able to deal with large-scale networks including the effect of multiple signaling pathways and their cross-talks, building on (18, 20) we developed a new calibration approach that uses L1 regularization to prune the network by inducing sparsity thus reducing the size of the model. The approach can be illustrated using the *in silico* model in **Fig. 2A**, where the PKN includes 9 nodes and 14 edges. A subnetwork (10 edges shown in black in **Fig. 2A**) was used to generate *in silico* data for 2 readouts (nodes shown in blue in **Fig. 2A**), consisting of 9 perturbations, which are some of the possible combinations of 2 stimuli (green nodes) and 2 inhibitors (targeting red nodes). *In silico* data were generated for 10 random sets of parameters (one example shown in **Fig. 2B**) and were then used to optimize the model by minimizing $Q_{LS}(T, k)$, which is defined as the sum of the squared difference between model predictions and true (*in silico*) values. Although the fit of the optimized models to the data was perfect, parameters could not be well estimated due to model redundancy and low identifiability. In order to improve parameter

estimates, we redefined the objective function Q introducing a L1 regularization term on the parameters τ_i to help the model remove the unconnected nodes (e.g. P4 and P5 in **Fig. 2A**). This was achieved by penalising the complexity of the model as:

$$Q(\tau, k) = Q_{LS}(\tau, k) + \lambda_{\tau} P(\tau) \text{ where } P(\tau) = \sum_i \tau_i \quad (1)$$

The term λ_{τ} controls the importance of the regularization term. We plotted the effect of increasing values of λ_{τ} for the *in silico* model on P and Q_{LS} (**Fig. 2C**), and on the accuracy of parameter estimates (as a sum of the square of the difference between estimated and true parameters) (**Fig. 2D**). As expected, for increasing values of λ_{τ} , Q_{LS} tends to increase (worse fit to the data) while P tends to decrease (sparser model). Good values of λ_{τ} are those on the elbow of the L-shaped curve in **Fig. 2C**, corresponding to the best compromise between good fit and sparse model.

Interestingly, these values (especially $\lambda_{\tau}=0.001$ which represents the most conservative choice in terms of regularization) also correspond to the best accuracy in estimating the parameters in **Fig. 2D**. Looking at the estimated parameters we could verify that for $\lambda_{\tau}=0.001$ the parameters τ_{P4} and τ_{P5} were indeed set to zero thanks to the regularization term in the objective function.

Generation of cell-line-specific models

The previously illustrated approach was applied to systematically characterize how signaling pathways mediate differential cellular responses to drugs in colorectal cancer. For this purpose we performed a large-scale perturbation screening consisting of 14 phosphoproteins measured in 14 genetically heterogeneous colorectal cancer (CRC) cell-lines in 43 different perturbed conditions. This sums up to a total of 8428 different experimental data points which are represented in **Fig. 3** (experimental details in **Materials and Methods**). Perturbations consist of combinations of 7 kinase specific inhibitors (targeting IKK, MEK, PI3K, BRAF, TAK1, TBK1, mTOR) and 5 stimuli (EGF, HGF, IGF1, TGFb, TNFa).

To characterize the signaling pathways we started from a comprehensive general prior-knowledge network (PKN) derived from the literature and public databases using OmniPath (21). The PKN was first compressed to reduce model complexity without affecting logic consistency as described in (22) (complete PKN in **fig. S1**), while maintaining interactions between the measured nodes (in blue in **Fig. 4A**), the inhibited nodes (in red) and the simulated nodes (in green). The definition of the

general model also took into consideration that one of the kinase inhibitors, PLX4720, works as selective BRAF inhibitor in cell-lines where BRAF is mutated in V600E (i.e. HT29 and SNUC5 in our panel), but induces a paradoxical activation of wild type BRAF cells, therefore- we modeled it as a stimulus for those cell-lines.

The following three-step optimization procedure was used for each cell-line (see **Materials and Methods** for more details). First, L1 regularization was applied to parameters τ_i to remove unconnected nodes, as described in Equation 1. Secondly, values of τ_i were fixed to the values estimated in Step 1 (shown in **Fig. 4C**) and optimization was repeated applying this time the L1 regularization to parameters k_{ij} to induce sparsity in the network. Thirdly, bootstrap distribution was then obtained for parameters k_{ij} by repeating the optimization on resampled experimental data with replacement (150 times). Median values for the k_{ij} parameters estimated at Step 3 are shown in **Fig. 4C** for all cell-lines. The resulting models were comparably sparse across cell-lines, with a percentage of values set to zero ranging between 12-28% for parameters τ_i and 51-73% for parameters k_{ij} . The optimized models showed good fit to the experimental data for all cell-lines, with mean squared error (MSE) within the range [0.013, 0.036] as shown in **Fig. 4B**.

Heterogeneity of the cell-line specific models

We further analysed the heterogeneity of the estimated parameters k_{ij} , in order to understand the variability in signal processing across the cell-lines. For each parameter, distributions across all cell-lines were compared using Kruskal-Wallis rank sum test (one-way analysis of variance on ranks) to test if estimated parameters (from bootstrap) stem from the same distribution for all cell-lines. For 46 out of 63 parameters the null hypothesis was rejected (p-value<0.05 after Bonferroni correction, and large effect size $w > 0.5$, see **Materials and Methods**), meaning that bootstrapped parameters come from a different distribution for at least one cell-line. The 17 Parameters for which the null hypothesis of equal mean rank was accepted showed a median value equal to zero for all cell-lines. For the remaining 46 parameters, post hoc pairwise tests (91 possible combination of 14 cell-lines) were performed using Wilcoxon rank sum test to additionally explore how many cell-lines have different parameter distributions (i.e. p-values Bonferroni corrected for multiple hypothesis testing < 0.05 and high r effect size > 0.5, see Online Methods). The

resulting percentages are shown in **Fig. 4A** as numbers next to each corresponding edge (only for the 46 edges passing the one-way analysis of variance test) and mapped as width of the edge, while edge colour intensity maps its mean value across cell-lines. As we might expect, genetic alterations alone are not representative of clusters of cell-lines neither at the level of phosphorylation response (**Fig. 3**) nor at the level of pathway models (**Fig. 4C**).

Among the 46 parameters k_{ij} showing different distributions across cell-lines, only 34 were different from zero in at least three cell-lines. For further analysis, we defined these parameters as highly heterogeneous. As for parameters τ_i , only 1 out of 25 parameters was set to zero across all cell-lines and 21 parameters were different from zero in at least three cell-lines. Among these, 12 were defined as highly heterogeneous, showing reasonable variability across cell-lines (standard deviation > 0.04). Although we are aware that the negative interaction between IKK and ERK might be artificial due to unspecific effect of the used IKK inhibitor, which causes an increase in ERK phosphorylation (8), we decided to keep the interaction in the model as we are interested in capturing all differential responses between cell-lines. From our analysis, we conclude that cell-lines are very different and hence it is important to characterize each cell-line with a specific signaling model.

Association of model parameters with drug response data

We next investigated if the previously defined highly heterogeneous model parameters can be associated with the efficacy of drugs. We reasoned that since the mode of action of multiple drugs involves affecting the functioning of the pathways we were modeling, their functional status should affect the efficacy of the drugs. Since all our 14 colorectal cancer cell-lines are part of the Genomics of Drug Sensitivity in Cancer (GDSC) panel (16), we could make use of the drug sensitivity data included in this comprehensive dataset in response to a large panel of drugs. In particular, in order to focus on biologically relevant associations, we selected only those drugs targeting nodes in our network (PKN) or first neighbours, i.e. targets that directly regulate a node in the PKN based on Omnipath (21), which is a curated ensemble of multiple pathway resources. We additionally excluded drugs showing a low variability in the response across our 14 cell-lines (less than three sensitive and three resistant cell-lines based on the classification used in (16)) and those for which

sensitivity data were missing for more than three of our 14 cell-lines. For the remaining 27 drugs, we investigated the association with our model parameters using cross-validated Elastic Net (details in **Materials and Methods**), using IC50s (half maximal inhibitory concentration) as a measure of drug sensitivity. We found a total of 146 robust associations (out of the 1242 possible ones), with 14 of our 27 drugs being explained by at least one model parameter. Top 50 associations are shown in **Fig. 5A** and all associations are shown in **fig. S2**. Resulting associations were then compared with those between drugs and genomic alterations (functional mutations and copy number alterations) acting on PKN nodes or first neighbor. Genetic biomarkers were identified using ANOVA as described in (16) (without correction for multiple hypothesis testing due to too large number of comparisons, see **fig. S3**) resulting in associations for 12 of our 27 drugs, five of which also showed associations with model parameters (corresponding genetic biomarkers are reported for the drugs in **Fig. 5A**).

Interpretation and validation of the dynamic biomarkers

We then focused on the 9 drugs whose efficacy was not significantly associated with mutations or copy number alterations but instead with pathway biomarkers, to explore how our approach provides insights in otherwise unexplained drug efficacy. Top associations are illustrated in **Fig. 5B**, visualized on the network (PKN) to show, for each drug, both the target(s) and the associated pathway parameter(s).

For example, multiple parameters related with MEK-ERK pathway showed association with EGFR signaling inhibitor (HG-5-88-01) and with BRAF inhibitors (SB590885 and Dabrafenib). Interestingly, many recent studies reported synergistic effects when combining MEK inhibitors with EGFR inhibitors (8, 23, 24) or with BRAF inhibitors (8, 25, 26) in different cancer types. In particular, focusing on colorectal cancer, combination of EGFR and MEK inhibitors was suggested to overcome resistance to MEK inhibitors in BRAF and KRAS mutants (8, 27) and it was suggested in (24) as potential therapy for CRC patients who become refractory to anti-EGFR therapies. A significant increase in response was also shown on a phase I study for CRC patients treated with a combination of three drugs targeting MEK, BRAF and EGFR respectively, in BRAF V600E mutants (28) and is currently in

Phase I/II study (NCT01750918 in clinicaltrials.gov). These examples underline the potential of our approach in suggesting targeted combinatorial therapies.

Another interesting example is the association of different MEK inhibitors (AZD6244, RDEA119, Trametinib, PD-0325901, CI-1040) with multiple model parameters related to GSK3 (τ_{GSK3} , $k_{RSKp90,GSK3}$, $k_{AKT,GSK3}$), as shown in **Fig. 6A-C**, with each parameter being associated with at least two of the inhibitors. The top association among these (which is overall ranked 8th in **Fig. 5A**), between AZD6244 and τ_{GSK3} , is shown in the centre of **Fig. 6D**. Our model predictions suggest that the combination with a GSK3 inhibitor could improve the sensitivity for cell-lines resistant to MEK inhibitors (i.e. high IC50) and with functional GSK3 (i.e. high τ_{GSK3}). On the contrary, cell-lines with nonfunctional GSK3 (i.e. low τ_{GSK3}) should not be affected by GSK3 inhibitors. In order to experimentally test this hypothesis, we combined a MEK inhibitor (Trametinib) with two GSK3 inhibitors (SB216763 and CHIR-99021) at increasing concentration (see **Materials and Methods**) as shown in the external plots in **Fig. 6D**. As expected, cell-lines with nonfunctional GSK3, based on the corresponding model parameter, do not show an increased sensitivity when treated in combination with any of the two tested GSK3 inhibitors, regardless of their response to MEK inhibitors (e.g. HT29 and SNUC2B for a sensitive and a more resistant case). On the contrary, cell-lines with more functional GSK3 tend to show an improved sensitivity under the combinatorial treatment (see HT115, SW1116, SW1463 and CCK81) even at the low drug concentrations tested.

Discussion

While it is broadly accepted that alterations in the functionality of signaling pathways largely determine the efficacy of kinase inhibitors used in the clinic, a complete understanding of their relationship is lacking. In this study we investigate this relationship and we show that the dynamic of signaling pathways can determine the efficacy of targeted drug treatments, in particular in cases where genomic data cannot. Additionally, we show that this information can be used to guide combinatorial therapies.

Our means towards this end were cell-specific mechanistic models of signal transduction for 14 colorectal cancer cell-lines trained with dedicated

phosphoproteomic data upon perturbations. We built our models based on differential equations to capture the continuous aspects of signal strength and time-dynamics, but using a logic formalism that allowed for straightforward interpretability of the model parameters in terms of life-time of each species in the network and strength of the regulatory interactions. We could then study how these model parameters, which determine the dynamic behaviour of the pathway, are related with the global cellular sensitivity to cancer drugs. We found some strong correlations between model parameters and drug sensitivity, such as between GSK3 life-time and MEK inhibitors, that suggest that the functionality of pathway interactions is indeed related to the efficacy of the drugs.

A key aspect of our study is the opportunity to analyze the heterogeneity of dynamic functionality (which is described by the model parameters) of signal transduction among a large panel of colorectal cancer cell-lines and how it relates to therapeutic drug response. The investigation of the heterogeneity of signaling pathways was empowered by the uniqueness of our dataset, consisting in 14 phospho-proteins measured under 43 differently perturbed conditions, which allowed us to characterize cell-line-specific models for 14 cell-lines derived from the same tissue of origin. We found that about half of the model parameters are highly heterogeneous in our panel of cell-lines and that heterogeneity cannot be explained by genomic alterations alone. Since these findings were obtained with a relatively limited sampling of the signaling status, we expect that many more associations can be found with a broader characterization. Mass-spectrometry based approaches, while still more costly and laborious, are under active development and likely to enable such an analysis in the near future.

Associations between model parameters and drug sensitivity were used to define interesting points of intervention to overcome resistance to specific drugs in specific cell-lines, thus paving the way for personalized combinatorial therapies. Some of the combinatorial therapies suggested using this approach, like the combination of MEK inhibitors with EGFR and/or BRAF inhibitors, are supported by literature (as reported in the Results) and are currently in clinical trials. We could also suggest novel combinatorial interventions, like the combination of MEK and GSK3 inhibitors, that we validated experimentally. Interestingly, the inhibition of GSK3 showed promising anticancer effect in recent studies on cancer cell-lines (29, 30) and its potential

interaction with MEK-ERK pathway has also been suggested in this context (30, 31) but, as far as we know, no proven synergistic effect had previously been reported.

Our findings underscore the value of studying signaling pathways dynamic to better understand tumor phenotypes and to exploit this knowledge to suggest new therapeutic strategies (32). Such an approach is fundamentally different from the currently common characterization of various ‘omics’ layers at the basal level and it can be exploited to prevent the dynamic adaptation mechanisms underlying drug resistance (14). Accordingly, we were able to find various pathway dynamic biomarkers for drugs with no genetic biomarkers. Hence, we believe that a perturbation-based strategy, even if restricted to fewer genetic backgrounds and only monitoring selected proteins, can provide a complementary strategy for biomarker discovery in cancer and beyond.

Materials and methods

Definition of logic ordinary differential equations

The logic ordinary differential equation formalism (17) is based on ordinary differential equations derived from logic models using a continuous update function B_i for each species x_i , which can assume continuous values $\{0, 1\}$. The differential equation for species i is defined as follow:

$$\dot{x}_i = \tau_i (B_i (f(x_{i1}), f(x_{i2}), \dots, f(x_{iN}))) - x_i)$$

Where τ_i is the life-time of species i . $\tau_i = 0$ means that the node is disconnected by the rest of the network and higher values of τ represent faster response, that can be interpreted as more functional species. $x_{i1}, x_{i2}, \dots, x_{iN}$ are the N regulators of x_i and each regulation is defined by a transfer function $f(x_{ij})$ with parameters n_{ij} and k_{ij} :

$$f(x_{ij}) = 1 - \frac{\frac{(1-x_{ij})^{n_{ij}}}{(1-x_{ij})^{n_{ij}+k_{ij}}}}{\frac{1}{1+k_{ij}}}$$

Where parameters n_{ij} were fixed to 3 and parameters k_{ij} were optimized (constrained between 0 and 1 included). As shown in **fig. S4**, $k_{ij} = 0$ means that the transfer function is equal to 0 for any value of the regulator j and can be interpreted as the edge j→i is not functional. Higher values of k_{ij} represent increasing strength of the regulatory interaction. For consistency with Boolean models (see (17) for more

details), the transfer function is defined to be $f(x_{ij}) = 1$ for $x_{ij} = 1$ (except for the $k_{ij} = 0$ case)

Optimization of parameters τ_i and k_{ij} was performed using MEIGO (33) and CNRode add-on of the CellNOptR package (18) using a modified version of the objective function to include the parameter regularization described in the paper, as well as an additional penalty to promote that the simulated values at $t=30$ min (corresponding to the measured time point) are at steady state in order to match the experimental assumptions.

Experimental setup of perturbation screen

Human colorectal cell-lines (CAR1, CCK81, COLO-320-HSR, DIFI, HCT116, HT115, HT29, SK-CO-1, SNU-C5, SW1116, SW1463, SW620 and SW837) are described in the COSMIC cell-lines project database (http://cancer.sanger.ac.uk/cell_lines). All cell-lines were cultured in either RPMI or DMEM/F12 medium with 10% FBS and were incubated at 37°C and 5% CO₂. Before perturbation commenced cells were starved overnight in serum free medium. At 90 minutes before lysis the cells were treated with inhibitors (or solvent control DMSO) and at 30 minutes before lysis cells were stimulated with ligand (or solvent control H₂O). This procedure was conducted for all single and combined inhibitor-ligand perturbations. The following inhibitors were used: MEKi AZD6244 (4µM), PI3Ki AZD6482 (10µM), mTORi AZD8055 (2µM), TBK1i BX-795 (10µM), IKKi BMS-345541 (25µM), BRAFi PLX4720 (5µM) and TAK1i 5Z-7-Oxozeanol (5µM). Concentrations were selected to inhibit target while minimizing off-target activity. We further used the following ligands: EGF (25ng/ml), HGF (50ng/ml), IGF1 (10ng/ml), TGFb (5ng/ml) and TNFa (10ng/ml). After treatment and incubation, lysates were collected and analyzed with the BioPlex Protein Array system (BioRad, Hercules, CA) as described earlier (8) using magnetic beads specific for AKT^{S473}, c-Jun^{S63}, EGFR^{Y1068}, ERK1/2^{T202,Y204/T185,Y187}, GSK3A/B^{S21/S9}, Ikbα^{S32,S36}, JNK^{T183,Y185}, MEK1^{S217,S221}, mTOR^{S2448}, p38^{T180,Y182}, PI3K^{Y458}, RPS6^{S235,S236}, p90RSK^{S380} and SMAD2^{S465,S467}. The beads and detection antibodies were diluted 1:3. For data acquisition, the BioPlex Manager software and R package lxb was used.

Drug combination experiments

Experiments were performed using MEKi Trametinib (0.005 μ M) as anchor drug and testing its effect on our 14 colorectal cancer cell-lines alone and in combination with 5 concentrations (from 10 μ M with 4-fold dilutions: 0.039 μ M, 0.156 μ M, 0.625 μ M, 2.5 μ M, 10 μ M) of two GSKi (SB 216763 and CHIR-99021). For available replicates median value was computed. Dose-response curves were normalised between 0 and 100 with 100 corresponding to the negative control (cells with no treatment) and 0 to the positive control (cells treated with Staurosporine 2 μ M). Same DMSO volume was maintained for all experiments (negative and positive controls, single drug and drug combination). Cell number was measured after 3 days of constant drug exposure using CellTiter-Glo reagent as described by the manufacturer (Promega).

Data preprocessing

For each cell-line, data were processed separately for each of the 14 measured phosphoproteins. The value of the control (unperturbed condition) was estimated as the median value of the 4 replicates (to decrease the risk of bias) and log₂ fold changes with respect to the control were then computed for each of the 42 perturbed conditions. As required by our logic formalism, data were then linearly scaled between 0 and 1, with 0.5 corresponding to the basal (control) condition.

Model optimization

The optimization procedure applied to each cell-line is the following:

1. L1 regularization was applied to parameters τ_i to remove unconnected nodes, as described in Equation 1, for five increasing values of λ_τ ($\lambda_\tau=0, 0.01, 0.1, 1, 10$). To increase the chances to obtain good solutions, 10 independent runs of the optimization were run in parallel. The best value for λ_τ was selected ($\lambda_\tau=0.1$) which provided the optimal balance between fitting precision and network size as described in the example, using the L-shaped curves (**fig. S5**).
2. Values of τ_i were fixed to the values estimated in Step 1 for $\lambda_\tau=0.1$ and optimization was repeated applying this time the L1 regularization to parameters k to induce sparsity in the network, with objective function defined as:

$$Q(k) = Q_{LS}(k) + \lambda_k P(k) \text{ where } P(k) = \sum_{ij} k_{ij}$$

As in step 1, optimization was repeated for increasing values of λ_k ($\lambda_k=0, 0.01, 0.1, 1, 10$) with 10 independent runs each, and best λ_k was selected based on L-shaped curves (**fig. S6**) to $\lambda_k=0.1$.

3. Bootstrap distribution was obtained for parameters k_{ij} (for $\lambda_k=0.1$ from Step 2) repeating the optimization by resampling experimental data 150 times with replacement.

Elastic net

In order to select only robust features, Elastic net training was iterated M times (where M is the total number of observations), each time leaving out one observation (sampling without replacement). For each iteration, leave-one-out cross-validation was applied to tune Elastic Net hyperparameters. Median values were computed for the weights of the features across the M iterations, in order to select only features appearing in at least half of the iterations.

Statistical Analysis

ANOVA was used to define genomic markers of drug sensitivity using the GDSC tools (<https://pypi.python.org/pypi/gdsctools/0.2.0>). Microsatellite instability was used as covariate and threshold on minimum size of the positive and negative population for each feature set to 3. The threshold for significance of the p-value was set to 0.05, but no correction for multiple hypothesis testing was applied.

Kruskal-Wallis rank sum test (which is a one-way analysis of variance on ranks) was used to test if estimated parameters (from bootstrap) derive from the same distribution for all 14 cell-lines (null hypothesis rejected if different for at least one group). Effect size w was computed as $\sqrt{\chi^2/N}$ where χ^2 is the statistics from the test and N is the number of observations. Effect size >0.5 is considered as large effect. P-values were Bonferroni corrected and threshold was set to 0.05. Wilcoxon rank sum test was then used for post hoc pairwise test on parameters passing the Kruskal-Wallis rank sum test, using as effect size Z/\sqrt{N} , where Z is the statistics from the test and N is the number of observations. Effect size >0.5 is considered as large effect. P-values were Bonferroni corrected and threshold was set to 0.05. Rank type test were preferred over parametric tests because they are highly robust against non-normality.

Supplementary Materials

Fig. S1. Extended version of the prior-knowledge network (PKN).

Fig. S2. Associations between model parameters and drug sensitivity.

Fig. S3. Genomic biomarkers of drug sensitivity.

Fig. S4. Transfer function used in logic ordinary differential equations.

Fig. S5. L-shaped curves for τ parameters.

Fig. S6. L-shaped curves for k parameters.

Data file S1 Optimised cell-line-specific models

Data file S2-S15 Phosphorylation perturbation data for the 14 cell-lines (MIDAS format)

References and Notes

1. V. Prasad, Perspective: The precision-oncology illusion, *Nature* **537**, S63 (2016).
2. A. de Gramont, S. Watson, L. M. Ellis, J. Rodón, J. Tabernero, A. de Gramont, S. R. Hamilton, Pragmatic issues in biomarker evaluation for targeted therapies in cancer, *Nat. Rev. Clin. Oncol.* **12**, 197–212 (2015).
3. B. J. Druker, F. Guilhot, S. G. O'Brien, I. Gathmann, H. Kantarjian, N. Gattermann, M. W. N. Deininger, R. T. Silver, J. M. Goldman, R. M. Stone, F. Cervantes, A. Hochhaus, B. L. Powell, J. L. Gabilove, P. Rousselot, J. Reiffers, J. J. Cornelissen, T. Hughes, H. Agis, T. Fischer, G. Verhoef, J. Shepherd, G. Saglio, A. Gratwohl, J. L. Nielsen, J. P. Radich, B. Simonsson, K. Taylor, M. Bacarani, C. So, L. Letvak, R. A. Larson, IRIS Investigators, Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia, *N. Engl. J. Med.* **355**, 2408–2417 (2006).
4. M. J. Garnett, E. J. Edelman, S. J. Heidorn, C. D. Greenman, A. Dastur, K. W. Lau, P. Greninger, I. R. Thompson, X. Luo, J. Soares, Q. Liu, F. Iorio, D. Surdez, L. Chen, R. J. Milano, G. R. Bignell, A. T. Tam, H. Davies, J. A. Stevenson, S. Barthorpe, S. R. Lutz, F. Kogera, K. Lawrence, A. McLaren-Douglas, X. Mitropoulos, T. Mironenko, H. Thi, L. Richardson, W. Zhou, F. Jewitt, T. Zhang, P. O'Brien, J. L. Boisvert, S. Price, W. Hur, W. Yang, X. Deng, A. Butler, H. G. Choi, J. W. Chang, J. Baselga, I. Stamenkovic, J. A. Engelman, S. V. Sharma, O. Delattre, J. Saez-Rodriguez, N. S. Gray, J. Settleman, P. A. Futreal, D. A. Haber, M. R. Stratton, S. Ramaswamy, U. McDermott, C. H. Benes, Systematic identification of genomic markers of drug sensitivity in cancer cells, *Nature* **483**, 570–575 (2012).
5. J. G. Paez, P. A. Jänne, J. C. Lee, S. Tracy, H. Greulich, S. Gabriel, P. Herman, F. J. Kaye, N. Lindeman, T. J. Boggon, K. Naoki, H. Sasaki, Y. Fujii, M. J. Eck, W. R. Sellers, B. E. Johnson, M. Meyerson, EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy, *Science* **304**, 1497–1500 (2004).
6. D. C. Kirouac, M. D. Onsum, Using network biology to bridge pharmacokinetics and pharmacodynamics in oncology, *CPT Pharmacometrics Syst Pharmacol* **2**, e71 (2013).

7. M. Niepel, M. Hafner, E. A. Pace, M. Chung, D. H. Chai, L. Zhou, B. Schoeberl, P. K. Sorger, Profiles of Basal and stimulated receptor signaling networks predict drug response in breast cancer lines, *Sci. Signal.* **6**, ra84 (2013).
8. B. Klinger, A. Sieber, R. Fritsche-Guenther, F. Witzel, L. Berry, D. Schumacher, Y. Yan, P. Durek, M. Merchant, R. Schäfer, C. Sers, N. Blüthgen, Network quantification of EGFR signaling unveils potential for targeted combination therapy, *Mol. Syst. Biol.* **9**, 673 (2013).
9. J. Saez-Rodriguez, L. G. Alexopoulos, M. Zhang, M. K. Morris, D. A. Lauffenburger, P. K. Sorger, Comparing signaling networks between normal and transformed hepatocytes using discrete logical models, *Cancer Res.* **71**, 5400–5411 (2011).
10. R. Merkle, B. Steiert, F. Salopiata, S. Depner, A. Raue, N. Iwamoto, M. Schelker, H. Hass, M. Wäsch, M. E. Böhm, O. Mücke, D. B. Lipka, C. Plass, W. D. Lehmann, C. Kreutz, J. Timmer, M. Schilling, U. Klingmüller, Identification of Cell Type-Specific Differences in Erythropoietin Receptor Signaling in Primary Erythroid and Lung Cancer Cells, *PLoS Comput. Biol.* **12**, e1005049 (2016).
11. D. C. Kirouac, J. Y. Du, J. Lahdenranta, R. Overland, D. Yarar, V. Paragas, E. Pace, C. F. McDonagh, U. B. Nielsen, M. D. Onsum, Computational modeling of ERBB2-amplified breast cancer identifies combined ErbB2/3 blockade as superior to the combination of MEK and AKT inhibitors, *Sci. Signal.* **6**, ra68 (2013).
12. D. Fey, M. Halasz, D. Dreidax, S. P. Kennedy, J. F. Hastings, N. Rauch, A. G. Munoz, R. Pilkington, M. Fischer, F. Westermann, W. Kolch, B. N. Kholodenko, D. R. Croucher, Signaling pathway models as biomarkers: Patient-specific simulations of JNK activity predict the survival of neuroblastoma patients, *Sci. Signal.* **8**, ra130 (2015).
13. J. B. Fitzgerald, B. Schoeberl, U. B. Nielsen, P. K. Sorger, Systems biology and combination therapy in the quest for clinical efficacy, *Nat. Chem. Biol.* **2**, 458–466 (2006).
14. A. Korkut, W. Wang, E. Demir, B. A. Aksoy, X. Jing, E. J. Molinelli, Ö. Babur, D. L. Bemis, S. Onur Sumer, D. B. Solit, C. A. Pratilas, C. Sander, Perturbation biology nominates upstream-downstream drug combinations in RAF inhibitor resistant melanoma cells, *Elife* **4** (2015), doi:10.7554/eLife.04640.
15. Å. Flobak, A. Baudot, E. Remy, L. Thommesen, D. Thieffry, M. Kuiper, A. Lægreid, Discovery of Drug Synergies in Gastric Cancer Cells Predicted by Logical Modeling, *PLoS Comput. Biol.* **11**, e1004426 (2015).
16. F. Iorio, T. A. Knijnenburg, D. J. Vis, G. R. Bignell, M. P. Menden, M. Schubert, N. Aben, E. Gonçalves, S. Barthorpe, H. Lightfoot, T. Cokelaer, P. Greninger, E. van Dyk, H. Chang, H. de Silva, H. Heyn, X. Deng, R. K. Egan, Q. Liu, T. Mironenko, X. Mitropoulos, L. Richardson, J. Wang, T. Zhang, S. Moran, S. Sayols, M. Soleimani, D. Tamborero, N. Lopez-Bigas, P. Ross-Macdonald, M. Esteller, N. S. Gray, D. A. Haber, M. R. Stratton, C. H. Benes, L. F. A. Wessels, J. Saez-Rodriguez, U. McDermott, M. J. Garnett, A Landscape of Pharmacogenomic Interactions in Cancer, *Cell* (2016), doi:10.1016/j.cell.2016.06.017.
17. D. M. Wittmann, J. Krumsiek, J. Saez-Rodriguez, D. A. Lauffenburger, S. Klamt, F. J. Theis, Transforming Boolean models to continuous models: methodology and application to T-cell receptor signaling, *BMC Syst. Biol.* **3**, 98 (2009).
18. C. Terfve, T. Cokelaer, D. Henriques, A. MacNamara, E. Goncalves, M. K. Morris, M. van Iersel, D. A. Lauffenburger, J. Saez-Rodriguez, CellNOptR: a flexible toolkit to train protein signaling networks to data using multiple logic formalisms, *BMC Syst. Biol.* **6**, 133 (2012).

19. H. Zou, T. Hastie, Regularization and variable selection via the elastic net, *J. R. Stat. Soc. Series B Stat. Methodol.* **67**, 301–320 (2005).
20. D. Henriques, M. Rocha, J. Saez-Rodriguez, J. R. Banga, Reverse engineering of logic-based differential equation models using a mixed-integer dynamic optimization approach, *Bioinformatics* **31**, 2999–3007 (2015).
21. D. Türei, T. Korcsmáros, J. Saez-Rodriguez, OmniPath: guidelines and gateway for literature-curated signaling pathway resources, *Nat. Methods* **13**, 966–967 (2016).
22. J. Saez-Rodriguez, L. G. Alexopoulos, J. Epperlein, R. Samaga, D. A. Lauffenburger, S. Klamt, P. K. Sorger, Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction, *Mol. Syst. Biol.* **5**, 331 (2009).
23. E. M. Tricker, C. Xu, S. Uddin, M. Capelletti, D. Ercan, A. Ogino, C. A. Pratilas, N. Rosen, N. S. Gray, K.-K. Wong, P. A. Jänne, Combined EGFR/MEK Inhibition Prevents the Emergence of Resistance in EGFR-Mutant Lung Cancer, *Cancer Discov.* **5**, 960–971 (2015).
24. S. Misale, S. Arena, S. Lamba, G. Siravegna, A. Lallo, S. Hobor, M. Russo, M. Buscarino, L. Lazzari, A. Sartore-Bianchi, K. Bencardino, A. Amatu, C. Lauricella, E. Valtorta, S. Siena, F. Di Nicolantonio, A. Bardelli, Blockade of EGFR and MEK intercepts heterogeneous mechanisms of acquired resistance to anti-EGFR therapies in colorectal cancer, *Sci. Transl. Med.* **6**, 224ra26 (2014).
25. J. Larkin, P. A. Ascierto, B. Dréno, V. Atkinson, G. Liskay, M. Maio, M. Mandalà, L. Demidov, D. Stroyakovskiy, L. Thomas, L. de la Cruz-Merino, C. Dutriaux, C. Garbe, M. A. Sovak, I. Chang, N. Choong, S. P. Hack, G. A. McArthur, A. Ribas, Combined Vemurafenib and Cobimetinib in BRAF -Mutated Melanoma, *N. Engl. J. Med.* **371**, 1867–1876 (2014).
26. C. Robert, B. Karaszewska, J. Schachter, P. Rutkowski, A. Mackiewicz, D. Stroiakovski, M. Lichinitser, R. Dummer, F. Grange, L. Mortier, V. Chiarion-Sileni, K. Drucis, I. Krajsova, A. Hauschild, P. Lorigan, P. Wolter, G. V. Long, K. Flaherty, P. Nathan, A. Ribas, A.-M. Martin, P. Sun, W. Crist, J. Legos, S. D. Rubin, S. M. Little, D. Schadendorf, Improved overall survival in melanoma with combined dabrafenib and trametinib, *N. Engl. J. Med.* **372**, 30–39 (2015).
27. C. Sun, S. Hobor, A. Bertotti, D. Zecchin, S. Huang, F. Galimi, F. Cottino, A. Prahallad, W. Grenrum, A. Tzani, A. Schlicker, L. F. A. Wessels, E. F. Smit, E. Thunnissen, P. Halonen, C. Lieftink, R. L. Beijersbergen, F. Di Nicolantonio, A. Bardelli, L. Trusolino, R. Bernards, Intrinsic resistance to MEK inhibition in KRAS mutant lung and colon cancer through transcriptional induction of ERBB3, *Cell Rep.* **7**, 86–93 (2014).
28. C. E. Atreya, E. Van Cutsem, J. C. Bendell, T. Andre, J. H. M. Schellens, M. S. Gordon, A. J. McRee, P. J. O'Dwyer, K. Muro, J. Tabernero, Others, in *ASCO Annual Meeting Proceedings*, (2015), vol. 33, p. 103.
29. Y. Yoshino, C. Ishioka, Inhibition of glycogen synthase kinase-3 beta induces apoptosis and mitotic catastrophe by disrupting centrosome regulation in cancer cells, *Sci. Rep.* **5**, 13249 (2015).
30. J. A. McCubrey, L. S. Steelman, F. E. Bertrand, N. M. Davis, M. Sokolosky, S. L. Abrams, G. Montalto, A. B. D'Assoro, M. Libra, F. Nicoletti, R. Maestro, J. Basecke, D. Rakus, A. Gizak, Z. N. Demidenko, L. Cocco, A. M. Martelli, M. Cervello, GSK-3 as potential target for therapeutic intervention in cancer, *Oncotarget* **5**, 2881–2911 (2014).
31. E. Lemieux, S. Cagnol, K. Beaudry, J. Carrier, N. Rivard, Oncogenic KRAS signalling

promotes the Wnt/ β -catenin pathway through LRP6 in colorectal cancer, *Oncogene* **34**, 4914–4927 (2015).

32. M. B. Yaffe, The Scientific Drunk and the Lamppost: Massive Sequencing Efforts in Cancer Discovery and Treatment, *Sci. Signal.* **6**, scisignal.2003684v1 (2013).

33. J. A. Egea, D. Henriques, T. Cokelaer, A. F. Villaverde, A. MacNamara, D.-P. Danciu, J. R. Banga, J. Saez-Rodriguez, MEIGO: an open-source software suite based on metaheuristics for global optimization in systems biology and bioinformatics, *BMC Bioinformatics* **15**, 136 (2014).

Acknowledgments: We thank members of the Saez laboratory for useful feedback, especially Luz Garcia-Alonso for help on mutations and Attila Gabor for fruitful discussion on model definition and David Henriques for support with CNORode. We thank Howard Lightfoot for help and assistance with drug combination data. **Funding:** N.B. acknowledges funding by the Berlin Institute of Health (BIH) and by the Federal Ministry of Research and Education (BMBF), grant OncoPath. F.E. thanks European Molecular Biology Laboratory Interdisciplinary Post-Docs (EMBL EIPOD) and Marie Curie Actions (COFUND) for funding. **Author contributions:** F.E. designed and performed the computational analysis and wrote the manuscript. V.D.M. and T.C. contributed to the computational analysis. B.K. Contributed to experimental design, the interpretation of the results and participated in writing the manuscript. A.S. and F.K. generated the experimental data. M.D. contributed to data preparation and analysis. M.G. Contributed to experimental design and experimental supervision. N.B. Supervised experiments, contributed to interpretation of the results, and participated in writing the manuscript. J.S.R Supervised the computational analysis and the overall project, and participated in writing the manuscript. M.G., N.B., and J.S.R. conceived the project. **Competing interests:** Authors declare no competing interests.

Figures and Tables

Figure 1

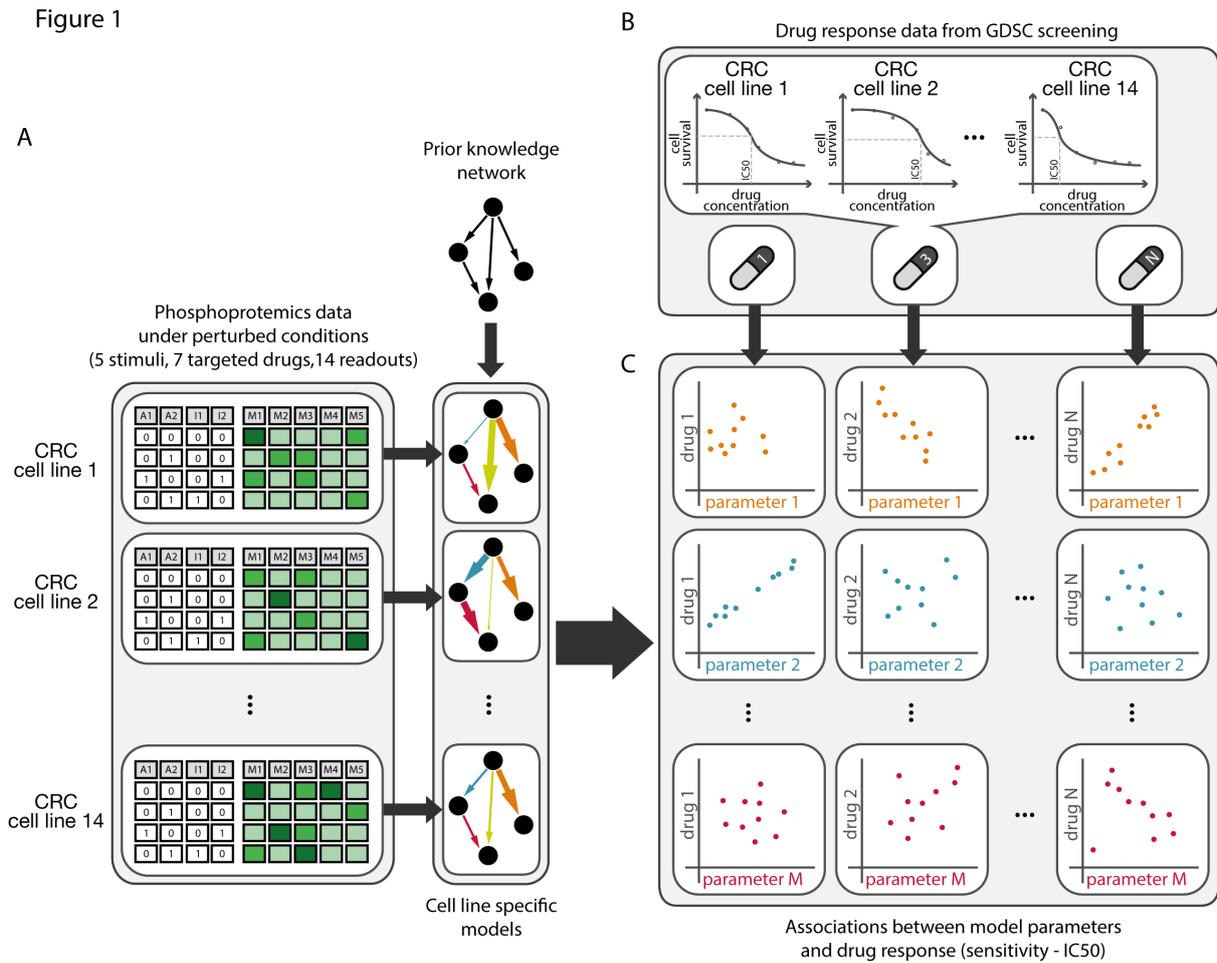


Fig. 1. Schematic of the approach used to define biomarkers of drug response from pathway models. A. Specific predictive models for 14 colorectal cancer (CRC) cell-lines were built from perturbation data and prior information about network structure (PKN). B. Drug sensitivity data for the same 14 CRC cell-lines were retrieved from the Genomics of Drug Sensitivity in Cancer (GDSC) panel in response to 27 different drugs targeting the PKN or first neighbour. C. Elastic Net was used to select strong associations between model parameters and drug sensitivity.

Figure 2

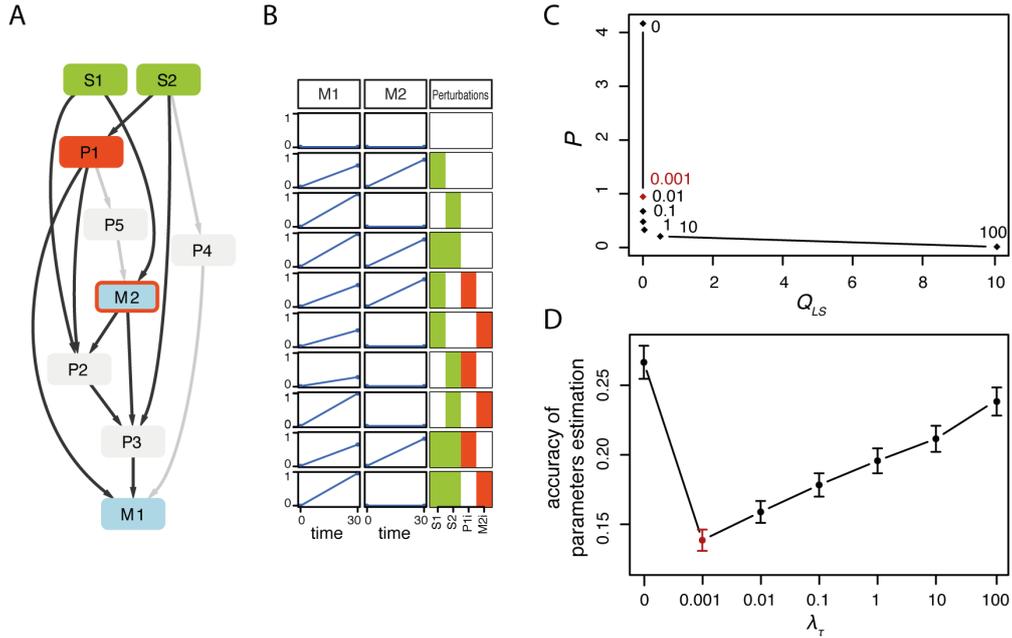


Fig. 2. Toy example to illustrate modeling approach. A. Toy general network (PKN) with edges used for data simulation highlighted in dark grey. B. One (out of 10) example of simulated perturbation data used to illustrate L1 regularization. C. Goodness of fit (Q_{LS} , i.e. sum of squared residuals) versus model sparsity (P , i.e. sum of estimated parameters) for different values of λ_T , ($\lambda_T=0, 0.001, 0.01, 0.1, 10, 100$) resulting in an L-shaped curve (each dot is the mean value obtained from the 10 *in silico* datasets). D. Mean accuracy (error bars represent standard deviation) of the estimated parameters (sum of the square of the difference between estimated and true parameters) as function of λ_T . Optimal λ_T is depicted in red.

Figure 3

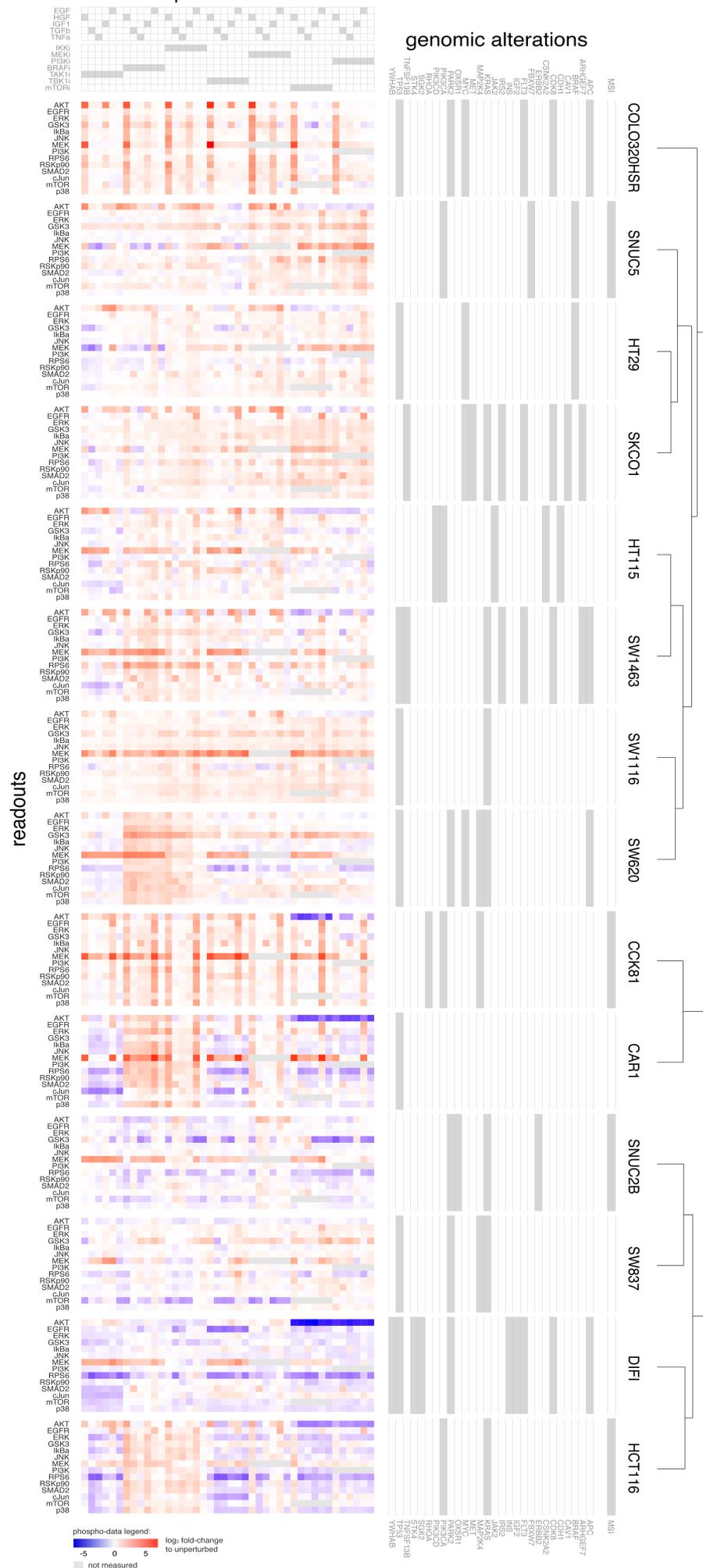


Fig. 3. Large-scale experimental perturbation dataset. Data for 14 colorectal cancer cell-lines in response to 43 different perturbed conditions with 14 measured phosphoproteins. Genomic alterations affecting investigated pathways are also shown and cell-lines are clustered based on phosphorylation profile across all perturbations.

Figure 4

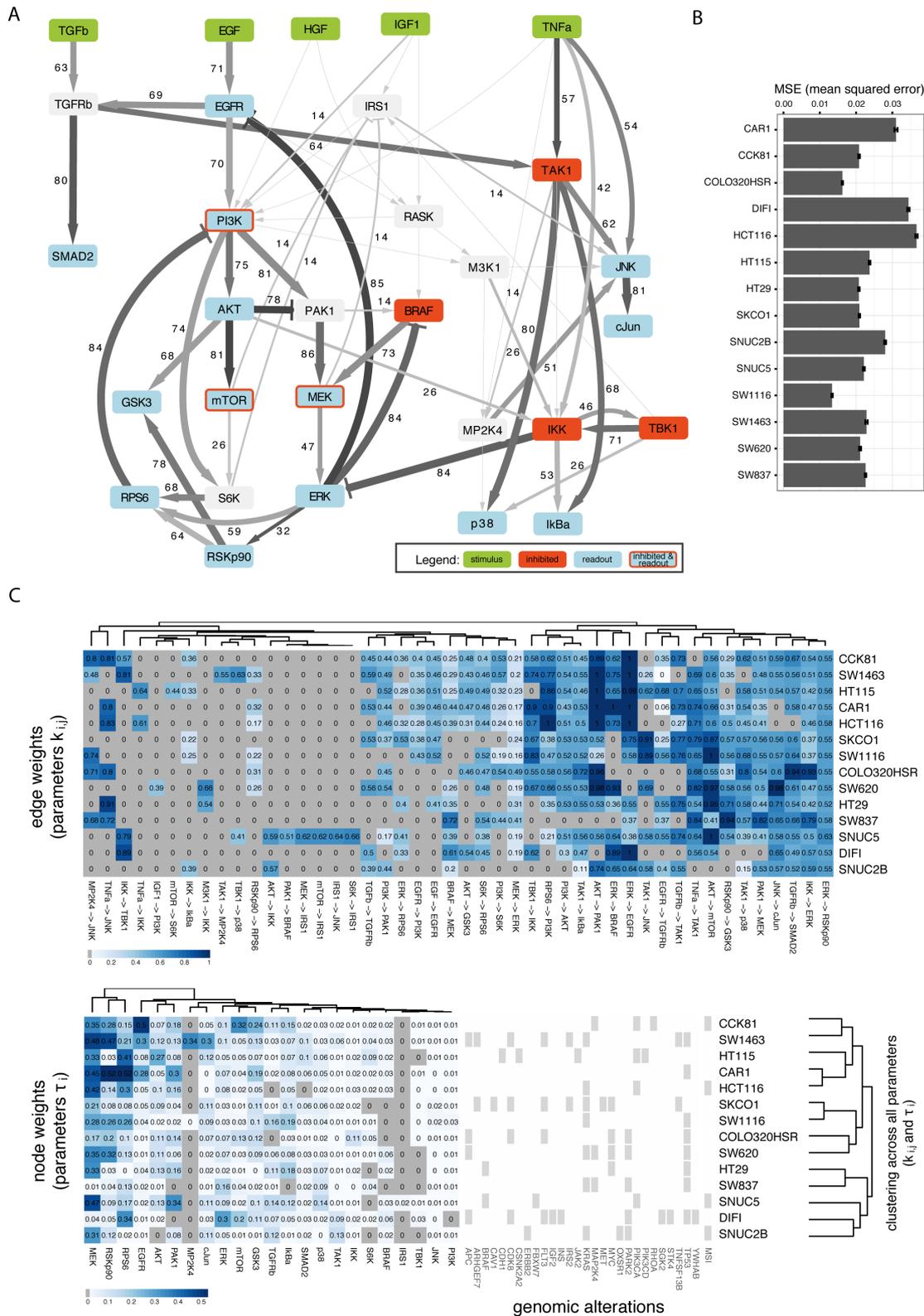


Fig. 4. Results of cell type specific model optimization. A. Compressed PKN with edge width and values representing the level of heterogeneity of the corresponding model parameter among cell-lines (percentage of pairwise tests between cell-lines for which null hypothesis of equal distribution was

rejected) and edge colour representing median value of the corresponding parameter across all cell-lines. B. Mean squared error (MSE) for each cell-line, error bars represent standard deviation derived from bootstrap. C. Estimated parameters k_{ij} (edge weight) and τ_i (node weight) which are different from zero in at least one cell-line. Clustering on the right is based on all estimated model parameters which does not correlate well with genomic alterations.

Figure 5

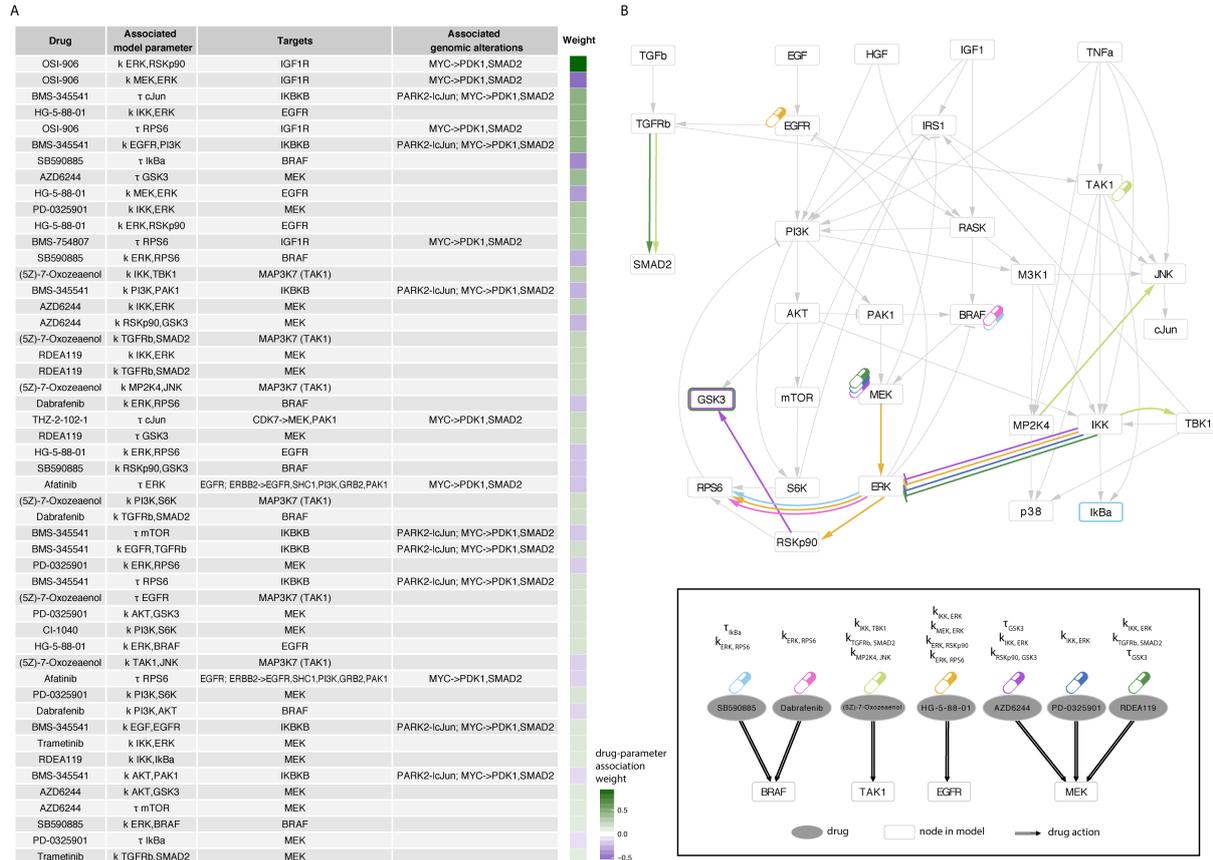


Fig. 5. Associations between model parameters and drug sensitivity. A. Top 50 associations resulting from Elastic Net selection between a drug (first column) and a model parameter (second column). For each drug, the corresponding targets are shown in third column and the associated genomic biomarker (if any) in the fourth column. Analysis is limited to drugs and mutations acting on PKN or first neighbour. For both drugs and mutations, corresponding PKN targets are indicated with an arrow (\rightarrow for positive regulation; $-$ for negative regulation) if mutations and/or drug targets are first neighbour regulating nodes in the PKN (which are named after the arrow). B. Among the top 25 parameter-drug associations, those involving drugs with no genetic biomarkers are schematically mapped to the signaling pathways. Drug targets for the 7 drugs listed as ellipses in the bottom panel, are shown in the compressed PKN (upper panel) using the corresponding coloured drug symbols. Corresponding model parameter associations for each drug are shown in the PKN using the same colour code to mark edges corresponding to regulator parameters k_{ij} (edge from i to j) and nodes border for corresponding life-time parameters τ_i (node i).

