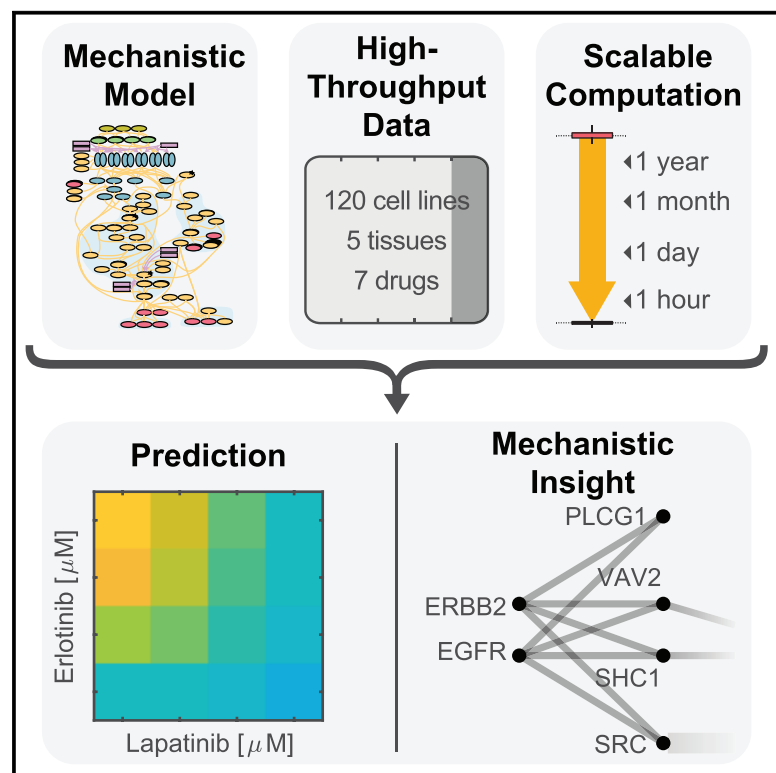**Article**

Efficient Parameter Estimation Enables the Prediction of Drug Response Using a Mechanistic Pan-Cancer Pathway Model

Graphical Abstract

Authors

Fabian Fröhlich, Thomas Kessler, Daniel Weindl, ..., Christoph Wierling, Bodo Lange, Jan Hasenauer

Correspondence

b.lange@alacris.de (B.L.),
jan.hasenauer@
helmholtz-muenchen.de (J.H.)

In Brief

Fröhlich et al. present a computational framework that allows training of mechanistic models at a previously infeasible scale. They apply the framework for drug response prediction in a cancer signaling context and demonstrate that combination treatment outcomes can be predicted from single drug treatment data.

Highlights

- Detailed, large-scale mechanistic model of cancer-related signaling pathways

- Speedup of over 10,000-fold enables data-driven modeling at unprecedented scales

- Pronounced parameter uncertainties do not imply pronounced prediction uncertainties

- Mechanistic models can predict response to drug combinations from single drug data

CellPress

OPEN ACCESS

Cell Systems

Article

CellPress

Efficient Parameter Estimation Enables the Prediction of Drug Response Using a Mechanistic Pan-Cancer Pathway Model

Fabian Fröhlich,[1,2,8] Thomas Kessler,[3,4,8] Daniel Weindl,[1] Alexey Shadrin,[6,7] Leonard Schmiester,[1,2] Hendrik Hache,[3] Artur Muradyan,[3] Moritz Schütte,[3] Ji-Hyun Lim,[3] Matthias Heinig,[1,2] Fabian J. Theis,[1,2] Hans Lehrach,[4,5] Christoph Wierling,[3,4] Bodo Lange,[3,4,*] and Jan Hasenauer[1,2,9,*]

[1]Institute of Computational Biology, Helmholtz Zentrum München, Neuherberg 85764, Germany
[2]Center for Mathematics, Chair of Mathematical Modeling of Biological Systems, Technische Universität München, Garching 85748, Germany
[3]Alacris Theranostics GmbH, Berlin 12489, Germany
[4]Max Planck Institute for Molecular Genetics, Berlin 14195, Germany
[5]Dahlem Centre for Genome Research and Medical Systems Biology, Berlin 12489, Germany
[6]KG Jebsen Centre for Psychosis Research, NORMENT, Institute of Clinical Medicine, University of Oslo, Oslo 0450, Norway
[7]Division of Mental Health and Addiction, Oslo University Hospital, Oslo 0450, Norway
[8]These authors contributed equally
[9]Lead Contact
*Correspondence: b.lange@alacris.de (B.L.), jan.hasenauer@helmholtz-muenchen.de (J.H.)
https://doi.org/10.1016/j.cels.2018.10.013

SUMMARY

Mechanistic models are essential to deepen the understanding of complex diseases at the molecular level. Nowadays, high-throughput molecular and phenotypic characterizations are possible, but the integration of such data with prior knowledge on signaling pathways is limited by the availability of scalable computational methods. Here, we present a computational framework for the parameterization of large-scale mechanistic models and its application to the prediction of drug response of cancer cell lines from exome and transcriptome sequencing data. This framework is over 10^4 times faster than state-of-the-art methods, which enables modeling at previously infeasible scales. By applying the framework to a model describing major cancer-associated pathways (>1,200 species and >2,600 reactions), we could predict the effect of drug combinations from single drug data. This is the first integration of high-throughput datasets using large-scale mechanistic models. We anticipate this to be the starting point for development of more comprehensive models allowing a deeper mechanistic insight.

INTRODUCTION

High-throughput experimental techniques are key for the comprehensive understanding of biological processes (Garnett et al., 2012; Marcotte et al., 2016; Seashore-Ludlow et al., 2015; The Cancer Genome Atlas Network, 2012). The analysis, integration, and interpretation of high-throughput data require computational methods. At the heart of this endeavor are usually mathematical models (Aldridge et al., 2006; Eduati et al., 2017). As widespread statistical models do not provide mechanistic insights, mechanistic models become increasingly important (Sanghvi et al., 2013). Mechanistic models featuring ordinary differential equations (ODEs) aim at a quantitative description of biological processes by systematic integration of prior knowledge and experimental data. These models have been used for the analysis of signal processing mechanisms (Bachmann et al., 2011), for the identification of drug targets (Schoeberl et al., 2009), as well as the development of prognostic signatures (Eduati et al., 2017; Fey et al., 2015). In the field of cancer research, mechanistic modeling has facilitated the study of oncogene addiction (Weinstein and Joe, 2006), synthetic-lethal phenotypes (Kaelin, 2005), and many other relevant phenomena (Zhang et al., 2009).

The development of a mechanistic model requires the definition of the model structure and the estimation of the unknown model parameters. The model structure is usually derived from prior knowledge available in scientific publications and databases. Models of varying detail have been developed and tailored to various pathways (Li et al., 2010), which ensures manageability of the development effort, but frequently neglects crosstalk. Large-scale pathway maps such as the Atlas of Cancer Signaling Network (ACSN) (Kuperstein et al., 2015) and Recon 2.2 (Swainston et al., 2016) address this issue by covering a majority of known molecular processes. These large-scale models contain thousands of biochemical reactions, the rates of which depend on unknown kinetic parameters. As precise literature values for many of these kinetic parameters are missing and predictions with imprecise literature values may lead to pronounced prediction uncertainties (Gutenkunst et al., 2007), these parameters must be estimated from experimental data. In principle, this can be achieved using optimization methods, which iteratively minimize an objective function, i.e., the distance between model simulation and experimental data (Raue et al., 2013; Villaverde et al., 2015). Yet, this requires repeated
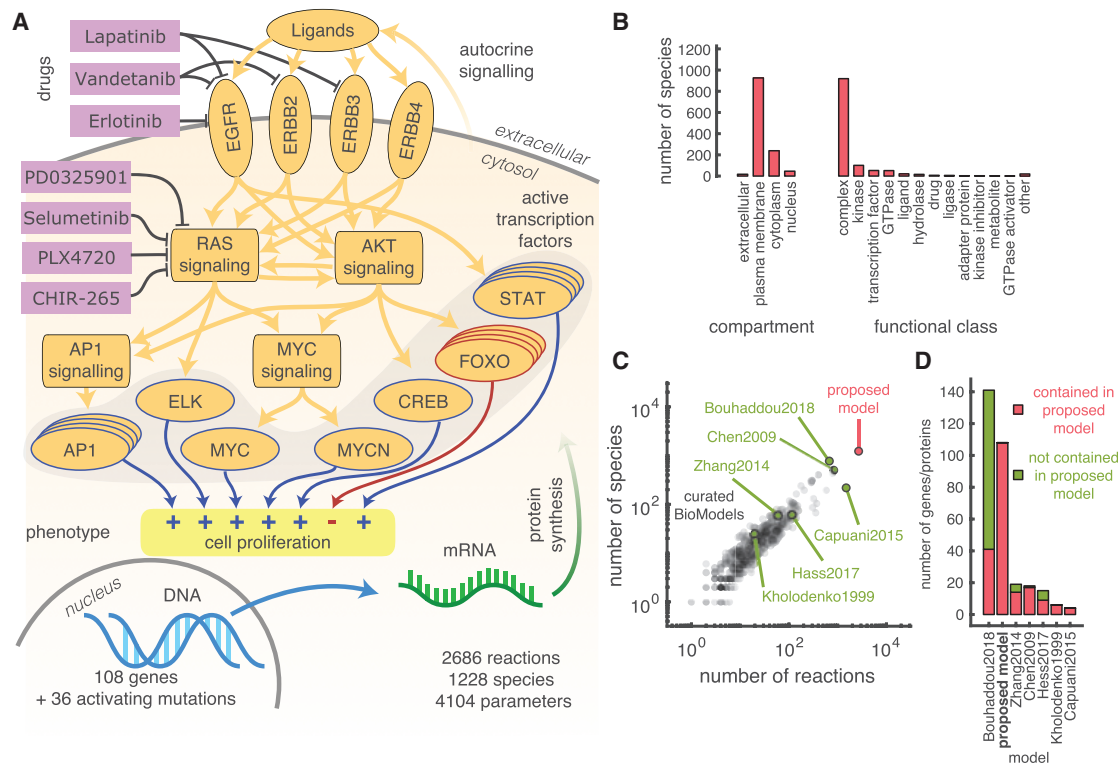
OPEN
ACCESS

CellPress

Figure 1. Model Structure and Properties
(A) Sketch of modeled signaling pathways. The developed model describes the synthesis and protein-protein interactions for protein products of 108 genes and 36 activating mutations. The visualization depicts drugs (purple), selected molecular species (orange), and cell viability as a phenotypic readout (yellow).
(B) Distribution of modeled molecular species on compartments and functional classes.
(C) Comparison of complexity of the proposed model (red circle) with curated models from the BioModels database (Li et al., 2010) (gray circles) and selected examples of models covering similar pathways (green circles).
(D) Comparison of number of genes or proteins in the proposed model with the selection of models from (C).
See also Figures S1 and S2 and Table S1.

numerical simulations. As—even for medium-scale models—millions of simulations may be necessary, the computational burden is often immense (Jagiella et al., 2017). Accordingly, parameterizing large-scale pathway models is often deemed intractable and has not been done in practice (Babtie and Stumpf, 2017). A scalable method for parameterization of large-scale mechanistic models is therefore essential for the community, as it enables the comprehensive integration of prior knowledge and experimental data.

Here, we introduce a computational framework for the parameterization of large-scale ODE models, which reduces computation time by multiple orders of magnitude compared to state-of-the-art methods. We demonstrate the applicability of this framework by considering the problem of predicting the drug response of cancer cell lines from somatic mutations and gene expression levels. To this end, we developed a large-scale mechanistic model of cancer signaling, which can be individualized using sequencing data. We demonstrate the parameterization of the model from thousands of drug assays from over 100 human cancer cell lines and validate the predictive power of the model. The mechanistic model provides several predictions that cannot be obtained using standard statistical models. Among others, we show that the parameterized model can accurately predict the effect of drug combinations from single drug data.

RESULTS

Large-Scale Mechanistic Model Integrates Knowledge of Cancer Signaling Pathways

To predict the drug response of cancer cell lines, we developed a mechanistic model integrating signaling modules reflecting the canonical human ERBB, RAS, and PI3K-AKT signaling pathways (Sanchez-Vega et al., 2018), as well as regulation of the transcription factors (TFs) MYC and AP1 (Wierling et al., 2015) (Figure 1A). The considered signaling modules are targeted by a large number of anti-cancer drugs. The model describes synthesis, degradation, translocation, complex formation, phosphorylation, and various other types of reactions for proteins, including functional variants (Figures 1B and S1). We assembled this model using the web-based platform PyBioS (Klipp et al., 2005; Wierling et al., 2007) (STAR Methods, section Model Development) and provide it as annotated systems biology markup language (SBML) file (Supplementary File 1). The model is based on curated information from ConsensusPathDB (Herwig et al., 2016), a meta-database integrating more than 20 public databases (e.g., DrugBank [Wishart et al., 2006], KEGG [Kanehisa et al., 2010], and Reactome [Croft et al., 2011]) and additional publications.

The model accounts for 108 genes and 36 activating mutations yielding a total of 1,228 molecular species in 4 compartments

(Figure 1B) involved in 2,686 reactions. The modeled mutations cover 7 of the 10 most frequent driver mutations reported by Rubio-Perez et al. (2015) and account for 22.1% of driver mutations observed in patient samples. The model describes the polypharmacologic action of 7 different small molecule kinase inhibitors (Table S1); this includes main and off-targets (in the pathways). Four of the considered inhibitors are FDA-approved. For 17 additional FDA-approved kinase inhibitors, one or more main targets are included in the model, but their action is currently not described. In total, the model covers main targets for 27.3% of FDA-approved targeted cancer therapies.

The drugs considered in this study, like most other drugs, interact with multiple targets (Davis et al., 2011; Klaeger et al., 2017). Our model captures multiple of these targets and possesses a good overlap with the drug-target interactions reported in the literature (Figure S2A and Table S1). The average number of implemented targets per drug is 4.3 (Figure S2A). Indeed, the model implements almost all drug-target interactions within the pathways considered in our model, reported by Davis et al. (2011) and Barret et al. (2008)—the manuscripts that primarily informed model construction. The model implements more drug targets than what is reported by Klaeger et al. (2017), who did not measure affinities for, e.g., multiple members of the ERBB-family. However, our study of drug-target interactions outside the model pathways revealed that only a fraction of the interactions is captured (Figures S2B and S2C). Again, Davis et al. (2011) report more interactions than Klaeger et al. (2017) as the latter study considers a smaller number of kinases targeted by the drugs. While some of the observed bindings might not influence the activity of the respective kinases, there is an overall reasonable agreement between binding affinities and kinase activity measured by Anastassiadis et al. (2011) (Figures S2D and S2E). Hence, although the proposed model provides a comprehensive description of the polypharmacology within the canonical human ERBB, RAS, and PI3K/AKT cancer signaling pathways, our analysis suggests that even larger models are required for a rigorous mechanistic description of the polypharmacology of anti-cancer drugs.

The mechanistic modeling of signal transduction and drug actions is complemented by a simple model for relative cell viability. The relative cell viability describes the number of healthy cells in a treated sample compared to an untreated control. This is the main readout of the two largest screening projects, the Cancer Cell Line Encyclopedia (CCLE) (Barretina et al., 2012) and the Genomics of Drug Sensitivity in Cancer (GDSC) (Yang et al., 2013) project. Originally employed by the CCLE, the GDSC now also adopted the CellTiterGlo assay as a viability assay—after a controversial discussion (Haibe-Kains et al., 2013; Haverty et al., 2016). The CellTiterGlo assay quantifies ATP levels as proxy for the number of healthy cells. Still, several concerns regarding reproducibility of results have been raised (Hafner et al., 2017; Niepel et al., 2017). The (relative) cell viability is the net sum of cell division and cell death, which are governed by a complex interplay of cellular signaling processes regulating, e.g., the balance between pro-growth and (anti-)apoptotic signals in response to extracellular stimuli or the presence of activating mutations within respective signal transduction cascades. A major function in the regulation of cell proliferation has been attributed to TF activation, e.g., of the MYC, AP1, and FOXO TFs, and regulation of target gene expression in response to extracellular or oncogenic stimuli. In the proposed model, we used the weighted sums of the simulated molecularly activated state of these TFs as a surrogate for viability (see STAR Methods, section Model Development). The relative cell viability is determined by the ratio of the cell viability in the treated and untreated conditions. This semi-mechanistic description provides a simple model of downstream regulatory processes.

To quantify the scale and comprehensiveness of our model, we compared it to curated mechanistic pathway models available in the BioModels database (Li et al., 2010) and also—in more detail—to a set of reference models that describe the ERBB, RAS, or AKT pathways. The reference set includes three models from the BioModels database, e.g., the models with the highest number of species (Chen et al., 2009) and the highest number of reactions (Capuani et al., 2015), and four additional models (Bouhaddou et al., 2018; Hass et al., 2017; Zhang and Liu, 2002), which were not published in the BioModels database. The proposed model describes more biochemical species and reactions than any other of the curated models (Figure 1C). Only the model by Bouhaddou et al. (2018) covers more unique protein or gene species (Figure 1D). The Bouhaddou model describes pathways, which, along with their constituting genes or proteins, are not included in our model. However, the comparably small overlap in included genes and proteins of the Bouhaddou model and the proposed model, as well as the lower number of included reactions (Figure 1C), suggests a lower comprehensiveness of the Bouhaddou model. Overall, most of the analyzed models do not describe the action of any drugs and do not include synthesis reactions for all modeled protein species, which prohibits the integration of transcriptomic data. In summary, the proposed model exceeds, to the best of our knowledge, all published models in terms of scale and comprehensiveness and is one of the few models that are suitable for the predictions of drug sensitivity from transcriptomic and genomic data.

Genomic Data Provides a Basis for Individualization of the Mechanistic Model

The mechanistic model provides a generic template for a subset of signaling processes in human cells and a link to cell viability. To obtain a model for a particular cancer cell line, we individualized the mechanistic model by incorporating gene expression levels as synthesis rates for proteins and their mutated functional variants (Figure 2A). We assumed that all other kinetic parameters, such as transport, binding, and phosphorylation rates, depend only on the chemical properties of the involved biochemical species. Accordingly, these parameters differ between proteins and their functional variants, while they are assumed to be identical across cell lines. This enables the simultaneous consideration of multiple cell lines and drugs for model parameterization, thereby increasing the available training sample size. Furthermore, this assumption allows us to predict the drug response of new cell lines solely from information about gene expression levels and functional variants.

In this study, we considered data for 120 human cancer cell lines from 5 tissues (breast, large-intestine, lung, pancreas, and skin) provided in the CCLE. We processed the included genetic characterization of cell lines in the untreated condition
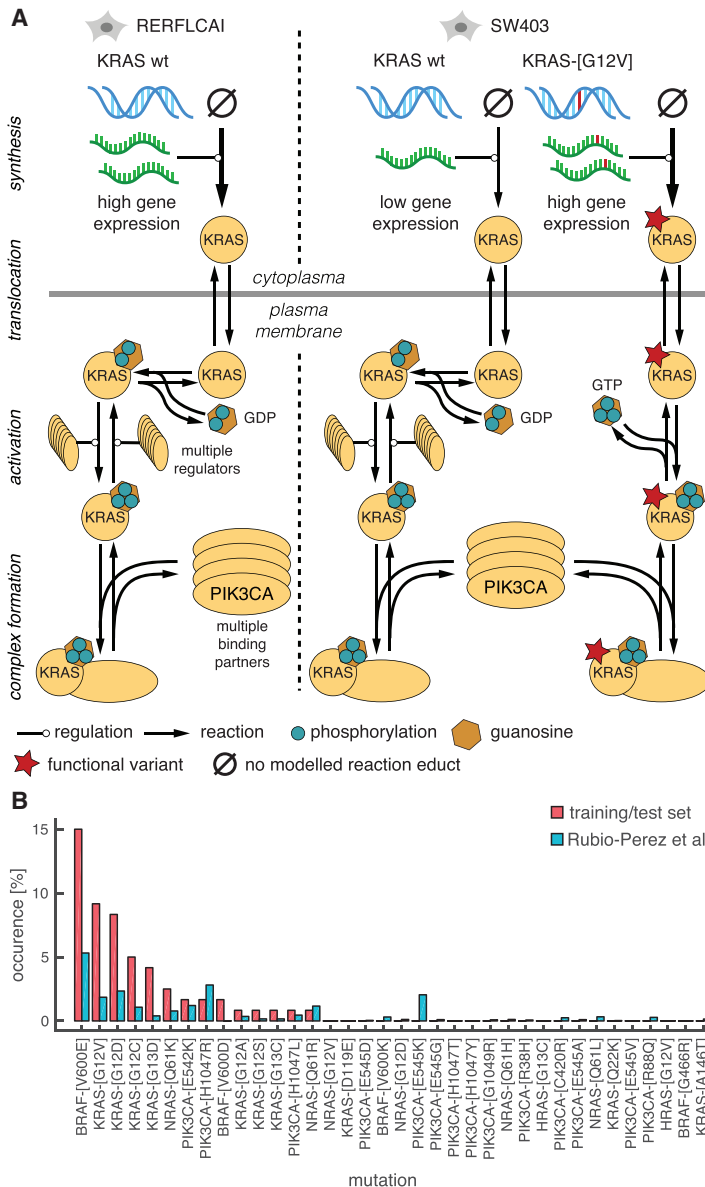
A

Figure 2. Individualization of the Model with Genomic and Transcriptomic Data
(A) Individualization of the generic mechanistic model for two exemplary cell lines: RERFLCAI (wild-type KRAS) and SW403 (wild-type and mutated KRAS). KRAS signaling model is illustrated from synthesis to complex formation. Degradation reactions are omitted.
(B) Comparison of the occurrence frequency of mutations included in the model between the training and test sets extracted from the Cancer Cell Line Encyclopedia and the InTOGen database by Rubio-Perez et al. (2015), which provides an extensive characterization of somatic mutations in human tumors.

B

tainty, we performed 5-fold cross-validation with 5 pairs of training (80%; 96 cell lines) and test datasets (20%; 24 cell lines).

To parameterize the model from the training data, we developed a computational framework. Similar to established approaches, this framework minimizes the sum of squared residuals of measured and simulated relative viability. This non-linear and non-convex ODE-constrained optimization problem was solved using multi-start local optimization, an efficient and reliable approach that outperformed global optimization methods in several studies (Hross and Hasenauer, 2016; Raue et al., 2013) (see STAR Methods, section Parameterization). As the optimization problem is high dimensional, state-of-the-art methods, such as forward sensitivity analysis (Raue et al., 2013), required 5×10^4 CPU hr (>6 CPU years) (Figure 3A) for a single evaluation of the objective function gradient. This is due to (1) the large-scale ODE model, (2) the large number of parameters, and (3) the large number of experimental conditions. As the gradient has to be evaluated hundreds of times for a single optimization, available toolboxes were not applicable.

To render parameterization tractable, we addressed challenges 1–3. First, we reduced the CPU time per model evaluation by using a sparse linear solver (Davis and Palamadai Natarajan, 2010) for ODE integration (0.5% non-zero entries in the Jacobian). Second, we employed a tailored variant of adjoint sensitivity analysis (Fröhlich et al., 2017) (see STAR Methods, section Adjoint Sensitivity Analysis), which improves scaling with the number of parameters. These two methodological advancements reduced the computation time over 37,000-fold (Figure 3A). Third, we established scalability with respect to the number of experimental conditions by parallelization on the level of cell lines (Figure 3A). Using 8 cores (7 workers), we observed a 6.4-fold acceleration. In total, our flexible and easily extendable parameterization framework reduced the expected wall time by over 240,000-fold.

Using 400 cores and a parallelization over local optimizations, our computational framework enabled the parameterization for all cross-validations in less than 1 week. In comparison, state-of-the-art approaches would have required hundreds of thousands of years. The local optimization achieved a substantial reduction of the sum of squared residuals within a few iterations, and then the curve flattened out (Figure 3B). To ensure a manageable computational cost, the optimization was stopped after 100 iterations. To filter insufficient optimization runs and to
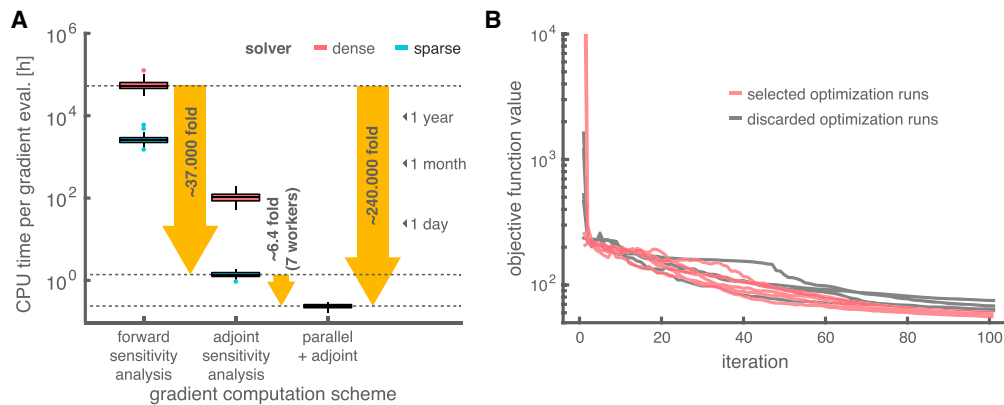
using a standardized bioinformatics pipeline (STAR Methods, section Experimental Model and Subject Details). Of the modeled driver mutations, 14 are present in more than one cell line (Figure 2B).

Scalable, Parallel Optimization Method Enables Model Parameterization

The mechanistic model includes more than 4,100 unknown parameters, i.e., kinetic constants and weighting factors. Most of these parameters are unknown and cannot be measured reliably under physiologically relevant conditions. To determine these unknown parameters such that available data and future experiments can be quantitatively described, we trained the model using measured viability data from 120 cell lines treated with 7 different drugs at up to 9 concentrations provided in the CCLE. In total, the selected CCLE subset comprises more than 6,900 experimental conditions. To assess the prediction uncer-

Figure 3. Parameterization of the Mechanistic Model

(A) Computation time for one evaluation of the objective function gradient, which determines the per-iteration time for a single local optimization step. For the non-parallelized evaluation, the time was computed based on representative samples (see STAR Methods, section Numerical Benchmark). The gradient evaluation time was dramatically reduced by using adjoint sensitivities, exploiting sparsity, and parallelization.

(B) Objective function traces for 10 different local optimization runs for the first cross-validation set. Initial conditions for the local optimization runs are sampled from a Latin hypercube. Although higher initial objective function values were observed, the corresponding axis was cropped at 10^4. The 5 best optimization runs are colored in red and were used for subsequent analysis.

improve robustness, we used ensemble averaging (see STAR Methods, section Ensemble Averaging) over the 5 optimization runs that achieved lowest objective function value in each cross-validation for all following analysis and prediction.

Mechanistic Model Yields Quantitative Description of Experimental Data and Generalizes to Test Data

The mechanistic model, parameterized using our scalable computational framework, describes the drug dose-dependent viability of cell lines. A comparison of model simulations with measured data revealed a good agreement and little variation in the prediction (Figure 4A). Overall, the dose response distribution across cell lines is captured well: the model accurately predicts that cell lines respond to drugs at concentrations that are higher than the experimentally measured kinase affinities (IC_{50}/K_D values) that were hard-coded into the model equations (Figure S3). The observed difference between kinase and growth inhibition efficiency could, e.g., be explained by the fact that tight inhibition of target kinases, which requires concentrations much higher than the IC_{50}, is often required for effective growth inhibition.

For a more quantitative analysis, we evaluated the model-data mismatch in terms of Pearson correlation coefficient (r) and root-mean-square error (RMSE). The correlation displayed a pronounced dependence on the drug concentration (Figure 4B), while the RMSE was effectively constant with respect to drug concentration. A careful inspection of measured data and model simulations (Figure S4) suggested that differences in correlations can be explained by differences in the dynamic range of measured relative viability across concentrations: if the dynamic range of measured relative viability was close to the level of measurement errors, as it is the case for low concentrations where cell-lines unanimously show little to no response to the drugs, we observed poor correlations despite good agreement between model simulations and measured data.

To avoid this problem, we aggregated measurements and simulations across concentrations and drugs and consistently found an overall high correlation (0.85 ± 0.01) and low RMSE

(0.14 ± 0.01) (Figure 4C). Our analysis of the dependence of RMSE and correlation on drug and tissue of origin revealed relatively small differences across drugs and tissues. We only found significant differences to both the overall correlation and overall RMSE for the drugs PD0325901 ($p_{corr} = 6.6 \times 10^{-4}$; $p_{RMSE} = 8.3 \times 10^{-3}$) and PLX4720 ($p_{corr} = 1.2 \times 10^{-2}$; $p_{RMSE} = 2.5 \times 10^{-2}$), and the pancreas tissue ($p_{corr} = 3.2 \times 10^{-4}$; $p_{RMSE} = 4.8 \times 10^{-3}$). We did not find any association between mutations or drug response pattern and the training error (Figure S5).

To assess the consistency of available datasets and the generalizability of the model predictions, we compared model simulations for cell lines in the training set to measured data from the GDSC database. Our analysis of the GDSC data revealed a substantially lower correlation (0.51 ± 0.01) and higher RMSE (0.25 ± 0.01) (Figure 5A) compared to the CCLE data. To provide a reference for these values, we computed RMSE and correlation between measured data from CCLE and GDSC. As the two databases employ different drug concentrations in their viability assays, we interpolated data from the CCLE using sigmoid interpolation. The computed correlation (0.49 ± 0.01) and RMSE (0.26 ± 0.01) are marginally weaker than what was achieved with the mechanistic model but significant with $p_{corr} = 1.8 \times 10^{-2}$ and $p_{RMSE} = 3.8 \times 10^{-3}$. This mediocre agreement of data from different databases was already characterized in previous studies (Haibe-Kains et al., 2013) and has been attributed to differences in assay protocols (Haverty et al., 2016), evaluation methods (Fallahi-Sichani et al., 2013; Pozdeyev et al., 2016), as well as shortcomings in the employed assay protocols (Hafner et al., 2017; Niepel et al., 2017). Accordingly, the slightly better performance of the mechanistic model compared to sigmoid interpolation is encouraging, but the low magnitude of differences is understandable as assay protocols are currently not considered in the mechanistic model.

To evaluate the predictive power of the parameterized mechanistic model, we turned to the test sets of the cross-validation (Figure 5B, left) as well as an additional independent test set containing 31 cell lines from four tissues not included in the training
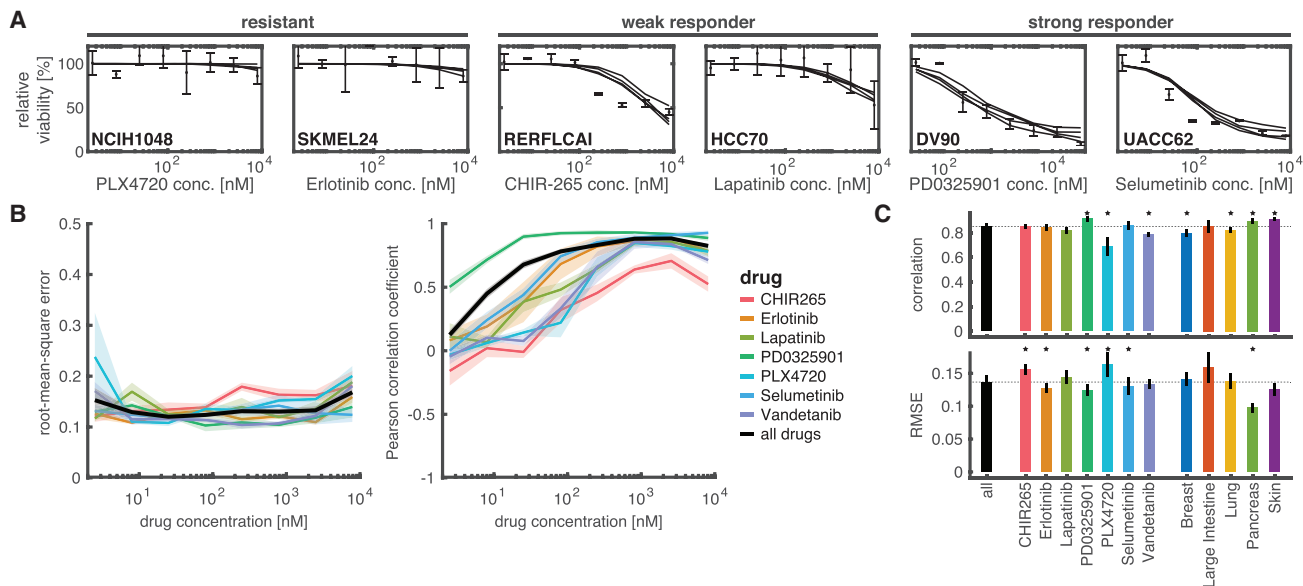
Figure 4. Analysis of Fitting Properties of the Model

(A) Representative examples of model simulations for six combinations of drugs (x-label) and cell lines (bold text on bottom, left). The five plotted lines are the median fit for the five best optimization runs for every cross-validation set.

(B) Pearson correlation and root-mean-square error (RMSE) of simulation and measured viability for the response at different drug concentrations. Shaded area indicates 95% confidence intervals over cross-validations.

(C) Correlation and RMSE statistics over cross-validations for individual drugs and tissues. Error bars indicate 95% confidence intervals. Statistically significant (p < 0.05) differences from the full distribution, according to a paired-sample t test after Bonferroni-Holm multiple testing correction, are indicated by stars.
See also Figures S3–S5.

data (kidney, soft tissue, ovary, stomach; Figure 5B right). Our analysis of correlation and RMSE revealed a good quantitative agreement of measured and predicted relative viability for the test set (r = 0.69 ± 0.09; RMSE = 0.19 ± 0.03) and the independent test set (r = 0.62 ± 0.03; RMSE = 0.17 ± 0.01) (Figure 5C). To assess the importance of parameterization for the agreement between model predictions and measurements, we also evaluated correlation and RMSE for 10 sets of randomly drawn parameters. Each set of 10 random parameter vectors was drawn uniformly from the specified parameter domain, and an ensemble model using the 5 best-performing parameter vectors was constructed. For the unparameterized model, correlation was consistently low (0.18 ± 0.05 to 0.22 ± 0.05) and RMSE was consistently high (0.25 ± 0.01 to 0.29 ± 0.01), which highlights that proper parameterization was essential for the predictive power of the mechanistic model.

To provide a reference for the performance of the parameterized mechanistic model, we trained several well-established statistical models on the same training set. The statistical models include a random forest (Breiman, 2001), sparse linear and nonlinear regression model (Tibshirani, 1996), as well as a network-constrained sparse regression model (with network derived from the mechanistic model) (Chen et al., 2015). These methods were among the top performers in previous drug sensitivity prediction DREAM challenges (Costello et al., 2014). The training of all statistical models was performed using state-of-the-art toolboxes (see STAR Methods, section Statistical Analysis). On the test set, the random forest model achieved the highest correlation (0.73 ± 0.05) and lowest RMSE (0.17 ±

0.02) (Figure 5D), which is statistically indistinguishable from the performance of the mechanistic model. On the independent test set, graph-constrained regularization for sparse generalized linear models (glmgraph) achieved the highest correlation (0.70 ± 0.01) and lowest RMSE (0.16 ± 0.01) (Figure 5E), which is marginally better than the mechanistic model (p_{corr} = 4.8 × 10^{-3}; p_{RMSE} = 7.4 × 10^{-4}). Accordingly, although this is the first time large-scale mechanistic models are used for prediction on this scale, they have already similar predictive power as long-established statistical models.

Pronounced Parameter Uncertainties Do Not Imply Pronounced Prediction Uncertainties

Experimental noise and data availability, which are ubiquitous features of biological experiments, limit the accuracy of the parameter estimates (Villaverde and Banga, 2014). This limitation is quantified in terms of structural and practical parameter identifiability as well as parameter uncertainty (Chis et al., 2011; Raue et al., 2009). As the parameter uncertainties can influence the prediction uncertainty, a rigorous quantification is desirable.

Profile likelihood calculation (Raue et al., 2009) and Markov-chain Monte Carlo sampling (Ballnus et al., 2018; Schmidl, 2012) are powerful methods for the rigorous quantification of parameter uncertainties. Yet, the largest mechanistic model, to which these methods have been successfully applied, possessed about 100 parameters and experimental conditions (Bachmann et al., 2011; Hug et al., 2013). The high computational cost currently prohibits the application of these methods
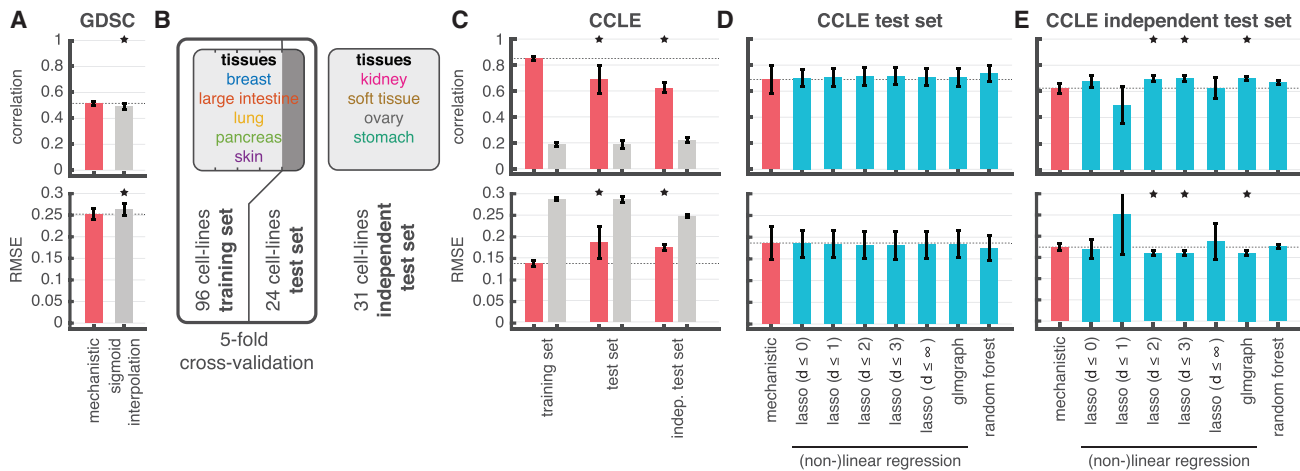
Figure 5. Validation of Model Prediction for Single Drug Treatment

(A, C, D, and E) Comparison of correlation coefficients and RMSE between measured and predicted viability for training, test, and independent datasets from CCLE and GDSC. Bar height indicates average across cross-validations; error bars show the 95% confidence interval of the mean. A dashed line indicates reference values; statistically significant ($p < 0.05$) differences from this reference, according to a paired-sample t test after Bonferroni-Holm multiple testing correction, are indicated by stars.

(A) Comparison of correlation and RMSE on GDSC data of the training set cell lines for the parameterized mechanistic model and a sigmoid interpolation.

(B) Overview of CCLE datasets used to evaluate the classification accuracy: (1) test set for cell lines originating from the same tissues on which the model was trained and (2) independent test set using cell lines from different tissues.

(C) Comparison of correlation and RMSE across training, test, and independent test set on CCLE data for the mechanistic model with optimized (red) and random parameters (gray).

(D and E) Comparison of correlation and RMSE for the mechanistic model and statistical models for test set and independent test set. For the Lasso approach, the parameter d defines up to which network distance interactions were considered.

to the considered model (Babtie and Stumpf, 2017). Therefore, we assessed the uncertainties using alternative methods and cautiously analyzed the results.

We performed an asymptotic uncertainty analysis using the Fisher information matrix (FIM)[5]. We computed the FIMs for the endpoints of the five best optimization runs of every cross-validation and evaluated their eigenvalue spectra (Figure 6A, top). The observed eigenvalue spectra were surprisingly similar across cross-validations and optimization runs, which suggested a rather consistent local topology of the objective function across the sampled parameter space. A large fraction of the eigenvalues was numerically indistinguishable from zero, which shows that a large fraction of parameters is structurally and thus also practically non-identifiable. The broad distribution of eigenvalues implies pronounced parameter sloppiness (Gutenkunst et al., 2007). Parameter sloppiness does not necessarily imply practical non-identifiability (Chis et al., 2016), but the low average value of non-zero eigenvalues suggests an even larger fraction of practically non-identifiable parameters than indicated by the numerically zero eigenvalues.

To follow up on this analysis, we computed the standard deviation of log-parameter estimates from the best 5 optimization runs across all cross-validations (Figure 6B). We found a large number of parameters that have a standard deviation close to zero. At first sight, this may suggest a good identifiability of these parameters. Yet, our asymptotic uncertainty analysis of the FIM suggested that many parameters are structurally non-identifiable. Accordingly, the geometry of the optimization problem needs to be considered when analyzing the distribution of local optimization results (Fröhlich et al., 2014): as the regularization

imposed by the employed interior point algorithm promotes a log-value of zero for non- or poorly identifiable parameters, the low standard deviation of values located at zero indicates poor identifiability. In fact, all parameters with a log parameter standard deviation smaller than 0.2 have log-parameter estimates that are statistically not different from 0 ($p < 0.005$; t test with Bonferroni-Holm multiple testing correction). This corroborates the hypothesis that most of the parameters are poorly identifiable.

To study the impact of poor parameter identifiability on predictions, we assessed the uncertainties of state variables and the cell viability output using the FIM (see STAR Methods, section Uncertainty Analysis). The analysis of the eigenvalue spectra (Figure 6A, middle, bottom) revealed a substantially higher fraction of non-zero eigenvalues and a higher average of non-zero eigenvalues. This suggests that low-uncertainty predictions are possible despite the pronounced poor identifiability. To test this, we analyzed the variance of predicted proliferation on the training and test set across the 5 considered cross-validations (Figure 6C). We found a generally low variance of model simulations, which increases with drug concentration. To put this into perspective, we decomposed the mean squared error (MSE) as the sum of prediction variance and squared prediction bias and computed the fraction that is explained by the prediction variance. Similar to variance, the respective fraction of the MSE increases with drug concentration but globally remains below 50%. This suggests that the observed MSE is predominantly due to a prediction bias and not due to prediction uncertainty. The source of the bias is currently unclear but overfitting and the fact that differences in cell lines might not be fully
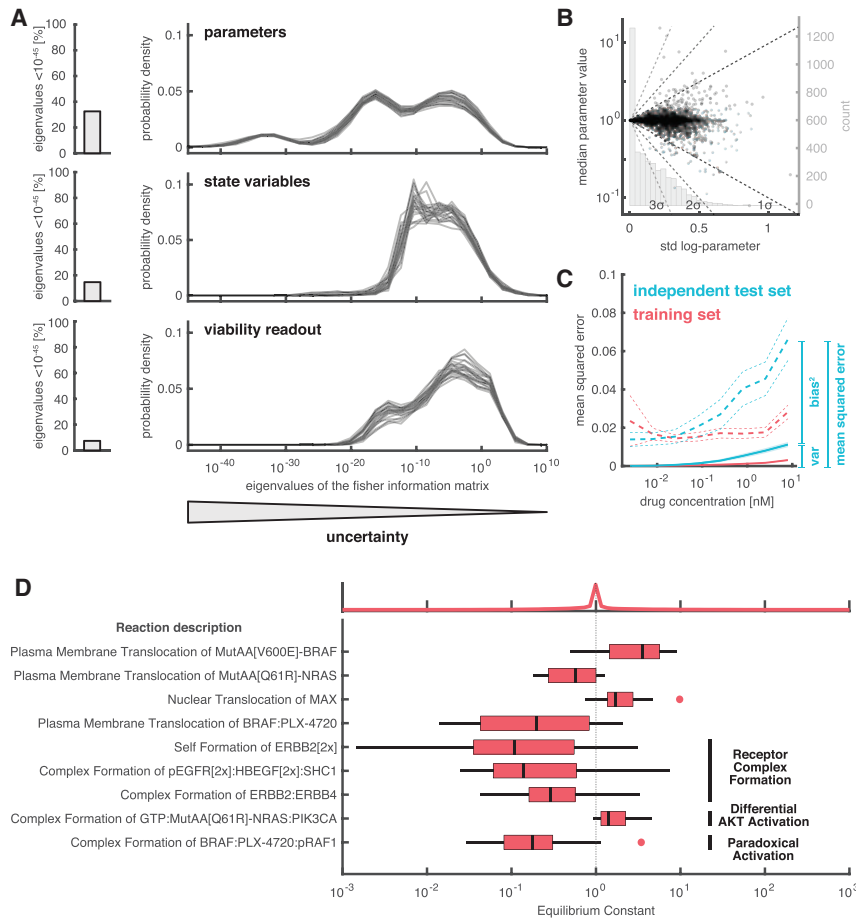
Figure 6. Parameter and Prediction Uncertainty Analysis

(A) Eigenvalues densities of the Fisher information matrix for parameters, state variables, and viability readouts. Small eigenvalues correspond to large uncertainties of readout combinations defined by respective eigenvectors. One distribution for each of the best 5 optimization runs for each of the 5 cross-validations is shown. Eigenvalues below 10^{-45} are not shown in the density plot, but the corresponding fraction of eigenvalues is indicated in the bar plot on the left.

(B) Histogram and scatterplot of mean and standard deviation of parameter estimates. Standard deviations and means were computed from the final values over the 5 best optimization runs for each of the 5 cross-validations. Dashed lines indicate 1σ, 2σ, and 3σ significance intervals (without multiple testing correction).

(C) Variance of ensemble model relative viability predictions at different drug concentrations for the training (red) and independent test set (blue). Lines indicate average value, and shaded areas indicate 95% confidence intervals computed over all data points. Reference mean squared error for both datasets is presented as dashed lines, where thick lines correspond to mean and thin lines indicate 95% confidence intervals computed over all data points. Decomposition of mean squared error into variance and $bias^2$ for the highest concentration in the independent test set is illustrated on the right edge of the figure.

(D) Distribution of parameter estimates for equilibrium constants differing substantially from zero. The background distribution of the parameter estimates is indicated on the top. For all displayed translocation reactions, the forward reactions are defined as the translocation from the cytoplasm to the nucleus or plasma membrane. For all displayed complex formation reactions, the forward reactions are defined as the binding of two constituting sub-complexes or species. The nominal value 1 is indicated as dashed line.

encoded in gene expression levels are plausible contributors to biased predictions.

Mechanistic Model Unravels Molecular Mechanisms for Sensitivity and Resistance from Viability Measurements

The parameters of the proposed mechanistic model are reaction rate constants. Unlike parameters of statistical models, these reaction rate constants provide information about the underlying biological reaction network. Yet our analysis revealed large parameter uncertainties (Figure 6). As previous studies on small-scale models suggested that products or fractions of parameters often have lower parameter uncertainties, we computed equilibrium constants as a ratio of on and off rates of reversible reactions (Vehlow et al., 2013). We assessed the uncertainty of the equilibrium constants and their difference from 1. As the employed interior-point optimization algorithm preferentially pushes non-identifiable parameters toward 1 (see STAR Methods, section Parameterization), consistent differences from 1 imply that the corresponding parameters are important to explain the observed drug sensitivity.

In total, nine equilibrium constants were estimated to be significantly different from 1 (Figure 6D; $p < 0.05$, t test with Bonferroni-Holm multiple testing correction). Three equilibrium constants are associated with receptor dimerization reactions, which are known to be important in cancer signal transduction (Chen et al., 2009). Four equilibrium constants are associated with localization to the nucleus and plasma membrane. One equilibrium constant is associated with the binding of activated NrasQ61R to Pik3ca, which leads to the activation of the AKT. This reaction enables differential, mutant-specific associations with downstream signaling transducers, which have previously been characterized to be important for Kras mutant variants in the non-small cell lung cancer context (Ihle et al., 2012). One equilibrium constant is associated with the binding of BRAF:PLX4720 and pRAF1, which is known to play an important role in the paradoxical activation through BRAF and CRAF homo- and heterodimerization, a well-known resistance mechanism in melanoma (Kholodenko, 2015.; Su et al., 2012).

Although most parameters were non-identifiable (Figure 6), our model-based analysis of viability data identified several molecular mechanisms important for drug response. Indeed, 5 out of 9 identified reactions have previously been reported to be important. This confirms that large-scale mechanistic modeling can extract molecular insights from phenotypic data, which is not possible using statistical modeling approaches.

Resistance and Susceptibility Markers Elucidate Drug Primary Mode of Action

The CCLE dataset has previously been used to identify biomarkers for drug resistance and susceptibility (Barretina et al., 2012). For logistic regression, negative regression coefficients indicate genes that convey drug resistance, positive coefficients indicate genes that convey drug susceptibility, and the absolute value of coefficients indicates the magnitude of influence. Genes with large positive coefficients are good candidates for biomarkers that can be used for patient stratification.

For mechanistic models, equivalent coefficients can be computed following the concept of model-based biomarkers introduced by Fey et al. (2015). The concept of model-based biomarkers interprets the kinetic model as a complex non-linear regression model, where respective regression coefficients can be computed as the sensitivity of output variables (cell viability) with respect to input variables (e.g., transcription levels). Here, we computed these coefficients for the mechanistic model using adjoint sensitivity analysis. Due to the non-linear nature of the mechanistic model, the computed coefficients are cell-line specific. This allows elucidation of genes that can act as resistance markers in some cell lines and susceptibility markers in other cell lines.

To identify primary markers for resistance and susceptibility, we tested for significant differences of the distribution of absolute sensitivity coefficients in a two-sided t test. As the number of genes with statistically significant p value ($p < 0.1$; Bonferroni-Holm multiple testing correction) was high, we selected 15 coefficients with the lowest p value for each drug, which yielded a set of 22 genes, including mutation variants. Using these candidate genes, we derived a skeletal signaling structure, which provides a high-level abstraction of the mechanistic model for the considered cell lines and drugs (Figure S6A): the signal is initiated at receptor tyrosine kinases (ERBB2, EGFR) and mediated to downstream signaling pathways through adapter proteins (SHC1), guanine exchange factors (VAV2), phospholipases (PLCG1), and kinases (SRC). Mutated species (NRAS, KRAS, BRAF) may activate downstream signaling independent of receptor signaling and activate the canonical Raf (BRAF, RAF1)-Mek (MAP2K1/2)-Erk (MAPK1/3) signaling pathway, which activates TFs such as MYC. The signal transduction can be inhibited by phosphatases (PPP2CA). To assess the impact of (1) the model structure and (2) the model training on the result, we repeated the same evaluation with the random initial parameters of each optimization run, i.e., untrained parameters. This yielded a substantially longer list of 36 genes, which included 16 genes from the analysis with the trained model. This suggests that the model structure already encodes information and narrows the list of relevant genes but that the training results in a re-weighting that is essential for generalization of genes that are relevant for resistance and susceptibility to different drugs.

This strong overlap of identified genes across drugs suggests that this skeletal structure defines a signaling program prototypes that, if active, can be targeted by drugs or convey resistance. To explore this hypothesis, we classified nodes as resistance markers if the average sensitivity for a specific drug was statistically significantly bigger than zero and as sensitivity markers if significantly smaller than zero. All remaining nodes, which have absolute value but not signed value significantly different from zero are classified as ambiguous markers. Comparison of the classification for receptor targeting drugs vandetanib, lapatinib, and erlotinib revealed high similarity of classifications across drugs as well as consistent classification of markers within pathways (Figures S6B–S6D): EGFR-SRC-STAT signaling was consistently identified as the primary pathway responsible for the heterogeneity in drug susceptibility, whereas MAPK and PI3K signaling were responsible for the heterogeneity in drug resistance. Interestingly, ERBB2 may convey both resistance and susceptibility to vandetanib and erlotinib, which have EGFR, but not ERBB2, as the main target. In contrast, ERBB2 is a clear susceptibility marker for lapatinib, which targets both EGFR and ERBB2. Linear logistic models cannot reproduce such ambiguous behavior, indicating that the mechanistic model facilitates a more fine-grained analysis. Furthermore, the mechanistic model provides additional information about the signal flow, which is not provided by statistical models.

Combination Treatment Outcomes Predicted from Single Drug Data

A key advantage of the mechanistic model is the capacity to predict latent variables, i.e., quantities that were not experimentally measured. To this end, we predicted the outcome of treatments with two drugs using the mechanistic model, which was solely trained on data from single drug data. Such predictions are interesting since the exhaustive experimental characterization of combination treatments is experimentally impractical due to the large number of required experiments. Accurate predictions of combination treatments from single treatment data would allow a cost-effective pre-selection of promising drug candidates for experimental follow-up, which is highly relevant for modern personalized medicine (Fitzgerald et al., 2006).

We considered the dataset published by O'Neil et al. (2016) reporting the response of cancer cell lines to single drugs as well as to combinations of two drugs at different concentrations (Figure 7A). The dataset includes 10 cell lines (NCIH1650, NCIH23, NCIH520, UACC62, RPMI7951, ZR751, SKMES1, HT144, MDAMB436, and SKMEL30) and 3 drugs (erlotinib, lapatinib, and PD0325901) contained in our training set. For the single drug viability measurements by O'Neil et al. (2016), the mechanistic model trained on the CCLE data achieves a correlation of 0.80 ± 0.02 and an RMSE of 0.16 ± 0.01 (Figures 7B and 7C). Accordingly, the agreement of the data by O'Neil et al. (2016) with the CCLE data is much higher compared to the previously analyzed agreement between GDSC data and CCLE data (Figure 5B). For every cross-validation, we only considered the measurements from cell lines, which were also contained in the training set.

Following the previous analysis for single drug treatments, we evaluated correlation and RMSE between measured and predicted relative viability. The mechanistic model displayed a good agreement with the drug combination data ($r = 0.75 \pm 0.03$), which is similar to the correlation observed for single drug treatments. Furthermore, the mechanistic model achieves a better correlation than established reference models based on Bliss independence (Bliss, 1939) (BI) ($r = 0.67 \pm 0.01$; $p = 0.04$) and highest single agent (Lehár et al., 2007) (HSA) model ($r = 0.67 \pm 0.01$; $p = 0.03$) (Figure 7A). As these reference models are usually used to define synergistic and antagonistic
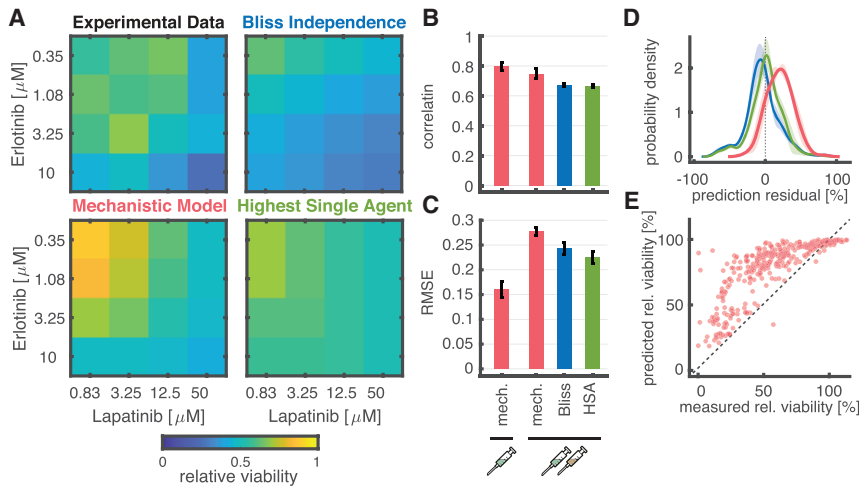
Figure 7. Prediction of the Effect of Drug Combinations from Single Treatment Data

(A) Experimental data and predictions from the mechanistic (mech, red), bliss independence (BI, blue), and highest single agent (HSA, green) models for the erlotinib-lapatinib combination in the NCIH1650 cell line.

(B and C) (B) Quantification of Pearson correlation and (C) root-mean-square error (RMSE) for single drug data and drug combination data for the models from (A). Bar height indicates average across cross-validations, error bars show the 95% confidence interval of the mean.

(D) Residual (measured relative viability—predicted relative viability) densities for the models from (A). Lines indicate the mean, and shaded areas the corresponding 95% confidence intervals across cross-validations.

(E) Scatterplot of model predictions versus experimental data for the mechanistic model for the first cross-validation.

interactions, a higher correlation of the mechanistic models means that synergy and antagonism were frequently correctly predicted. BI and HSA models achieve a better RMSE (Figure 7C), as the mechanistic model tends to overestimate the viability (Figures 7D and 7E). Yet, a consistent over-estimation of relative viability is acceptable, if not desirable, for the identification of promising drug combinations, as it results in a small false-positive rate. Importantly, the high correlation ensures proper prioritization for experimental follow-up. Accordingly, the mechanistic model provides accurate and valuable prediction for drug combination treatments based on the biochemical processes.

DISCUSSION

We generated a large-scale mechanistic model that integrates large amounts of prior knowledge and expands upon previous large-scale models by implementing kinetic rate laws, mutation variants of key regulators, and possibilities for individualization. As the model can be individualized to particular cell lines and covers many relevant driver mutations, the model provides a valuable resource for analysis of various cancer types and drug treatments. Furthermore, as the model already describes the action of drugs on on- and off-targets, it can—unlike models with tens to hundreds of species—contribute to a mechanistic understanding of polypharmacology. In particular, the inclusion of dose-response data for non-tumor cell lines could deepen our understanding of off-target toxicity. Conceptually, the model could also be integrated with other large-scale models, e.g., Recon 2.2 (Swainston et al., 2016), to provide a basis for a whole-cell computational model of human (cancer) cells and complement work on micro-organisms (Karr et al., 2012).

To parameterize this model, we developed a flexible computational framework that provides scalability with respect to the number of parameters and number of state variables and employs parallelization to handle the large number of experimental conditions. The final wall time requirement for all optimization runs ($\sim 4 \times 10^3$ hr) was more than one order of magnitude lower

than the wall time required for a single gradient evaluation using established methods ($\sim 6 \times 10^4$ hr). This allowed, to the best of our knowledge, for the first time, the parameterization of a large-scale mechanistic model from experimental data from over 100 cell lines, each under dozens of experimental conditions. The computational efficiency of the approach renders iterative rounds of optimization, hypothesis generation, and model refinement of large-scale mechanistic models from multiple high-throughput datasets feasible within a reasonable time frame. Our implementation of the model and method is available as Supplementary File 1 and can be freely reused by other research groups. The modularity of the method allows straightforward application to different models as well as datasets.

The assessment of cell viability—a key readout for cancer therapy—revealed that the parameterized model achieves a prediction accuracy comparable to established statistical models. Hence, the large-scale mechanistic model we derived and parameterized can predict the drug response of cancer cell lines from sequencing data. A broad advantage of mechanistic models is the possibility to predict latent variables, such as equilibrium constants of individual reactions or the response to combination drug treatments. The prediction of experimentally not measured, i.e., latent, quantities is generally not possible with conventional statistical methods.

Our uncertainty analysis of the model suggested that—despite pronounced parameter uncertainties—prediction uncertainties were low and accounted only for a minor fraction of errors in predicted viability. Still, high parameter uncertainties related to structural and practical non-identifiabilities impede the mechanistic interpretability and are likely to be more problematic for predictions at the molecular level. We expect that these parameter uncertainties can be substantially reduced by considering molecular data in the training process, e.g., proteomic and phosphoproteomic data. Such datasets are currently not available from published comprehensive screens but are being collected in ongoing efforts such as the Library of Integrated Network-Based Cellular Signature (Koleti et al., 2018) program. Complementary, scalable computational methods for the rigorous assessment of parameters and

prediction uncertainties need to be developed. Unlike profile calculation and sampling, uncertainty analysis based on asymptotic theory and optimization results provide only a part of the picture. Indeed, we found that the variability between optimizer runs and cross-validation results has to be interpreted with caution (Fröhlich et al., 2014).

Beyond parameter values, also the model structure is subject to uncertainty. While the model is based on extensive literature and was manually curated, interactions might still be missing or incorrect. Some methods have been demonstrated to detect such problems (Kondofersky et al., 2016; Penas et al., 2017); however, it is unlikely that these methods scale sufficiently well. An alternative might be to consider ensembles of models featuring different model structures and employ the concepts of bagging and boosting (Opitz and Maclin, 1999). First applications of this in the field of mechanistic modeling are available for models with dozens of biochemical species (Henriques et al., 2017).

We anticipate that the computational framework we introduced will address the growing demand (Babtie and Stumpf, 2017) for methods that allow the systematic integration of heterogeneous datasets with large-scale mechanistic models. So far, only statistical and machine learning methods were applicable, limiting the mechanistic insights and the exploitation of prior knowledge. High-quality mechanistic models—such as the one introduced in this manuscript—can identify actionable targets and guide personalized treatments and are thus of great relevance to researchers and clinicians in systems biology and systems medicine.

STAR★METHODS

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

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - CCLE Data
  - GDSC Data
- METHOD DETAILS
  - Model Development
  - Model Encoding in SBML
  - Numerical Simulation
  - Stability Analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Parameterization
  - Adjoint Sensitivity Analysis
  - Numerical Benchmarking of Forward and Adjoint Sensitivity Analysis
  - Ensemble Averaging
  - Uncertainty Analysis
  - Statistical Modeling
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and one table and can be found with this article online at https://doi.org/10.1016/j.cels.2018.10.013.

AUTHOR CONTRIBUTIONS

A.M., B.L., C.W., H.L., and T.K. designed the model. F.F. and J.H. designed the methods for numerical simulation, parameter optimization, and uncertainty analysis. C.W., F.F., F.J.T., J.H., M.H., and T.K. designed the numerical experiments. A.S., J.-H.L., H.H., H.L., M.S., and T.K. wrote and ran the code for model and data mapping and integrated and assembled model input data. D.W., F.F., J.H., and L.S. wrote and ran the code for the parameterization and assessment of the mechanistic model. M.H. wrote and ran the code for the parameterization and assessment of the statistical models. D.W., F.F., J.H., L.S., and M.H. analyzed output data. C.W., D.W., F.F., J.H., M.H., and T.K. wrote the manuscript. All authors discussed the results and implications and commented on the manuscript at all stages.

REFERENCES

Aldridge, B.B., Burke, J.M., Lauffenburger, D.A., and Sorger, P.K. (2006). Physicochemical modelling of cell signalling pathways. Nat. Cell Biol. 8, 1195–1203.

Anastassiadis, T., Deacon, S.W., Devarajan, K., Ma, H., and Peterson, J.R. (2011). Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity. Nat. Biotechnol. 29, 1039–1045.

Babtie, A.C., and Stumpf, M.P.H. (2017). How to deal with parameters for whole-cell modelling. J.R. Soc. Interface 14.

Bachmann, J., Raue, A., Schilling, M., Böhm, M.E., Kreutz, C., Kaschek, D., Busch, H., Gretz, N., Lehmann, W.D., Timmer, J., et al. (2011). Division of labor by dual feedback regulators controls JAK2/STAT5 signaling over broad ligand range. Mol. Syst. Biol. 7, 516.

Ballnus, B., Schaper, S., Theis, F.J., and Hasenauer, J. (2018). Bayesian parameter estimation for biochemical reaction networks using region-based adaptive parallel tempering. Bioinformatics 34, i494–i501.

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 483, 603–607.

Barrett, S.D., Bridges, A.J., Dudley, D.T., Saltiel, A.R., Fergus, J.H., Flamme, C.M., Delaney, A.M., Kaufman, M., LePage, S., Leopold, W.R., et al. (2008). The discovery of the benzhydroxamate MEK inhibitors CI-1040 and PD 0325901. Bioorg. Med. Chem. Lett. 18, 6501–6504.

Bliss, C.I. (1939). The toxicity of poisons applied jointly. Ann. Appl. Biol. 26, 585–615.

Bouhaddou, M., Barrette, A.M., Stern, A.D., Koch, R.J., DiStefano, M.S., Riesel, E.A., Santos, L.C., Tan, A.L., Mertz, A.E., and Birtwistle, M.R. (2018). A mechanistic pan-cancer pathway model informed by multi-omics data

interprets stochastic cell fate responses to drugs and mitogens. PLoS Comput. Biol. 14, e1005985.

Breiman, L. (2001). Random forests. Mach. Learn. 45, 5–32.

Capuani, F., Conte, A., Argenzio, E., Marchetti, L., Priami, C., Polo, S., Di Fiore, P.P., Sigismund, S., and Ciliberto, A. (2015). Quantitative analysis reveals how EGFR activation and downregulation are coupled in normal but not in cancer cells. Nat. Commun. 6, 7999.

Chen, L., Liu, H., Kocher, J.P., Li, H., and Chen, J. (2015). Glmgraph: an R package for variable selection and predictive modeling of structured genomic data. Bioinformatics 31, 3991–3993.

Chen, W.W., Schoeberl, B., Jasper, P.J., Niepel, M., Nielsen, U.B., Lauffenburger, D.A., and Sorger, P.K. (2009). Input–output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data. Mol. Syst. Biol. 5, 239.

Chis, O.T., Banga, J.R., and Balsa-Canto, E. (2011). Structural identifiability of systems biology models: A critical comparison of methods. PLoS One 6, e27755.

Chis, O.T., Villaverde, A.F., Banga, J.R., and Balsa-Canto, E. (2016). On the relationship between sloppiness and identifiability. Math. Biosci. 282, 147–161.

Costello, J.C., Heiser, L.M., Georgii, E., Gönen, M., Menden, M.P., Wang, N.J., Bansal, M., Ammad-ud-din, M., Hintsanen, P., Khan, S.A., et al. (2014). A community effort to assess and improve drug sensitivity prediction algorithms. Nat. Biotech. 32, 1202–1212.

Croft, D., O'Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., Caudy, M., Garapati, P., Gopinath, G., Jassal, B., et al. (2011). Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Res. 39, D691–D697.

Davis, M.I., Hunt, J.P., Herrgard, S., Ciceri, P., Wodicka, L.M., Pallares, G., Hocker, M., Treiber, D.K., and Zarrinkar, P.P. (2011). Comprehensive analysis of kinase inhibitor selectivity. Nat. Biotechnol. 29, 1046–1051.

Davis, T.A., and Palamadai Natarajan, E. (2010). Algorithm 907: KLU, a direct sparse solver for circuit simulation problems. ACM Trans. Math. Softw. 37, 36.

Eduati, F., Doldàn-Martelli, V., Klinger, B., Cokelaer, T., Sieber, A., Kogera, F., Dorel, M., Garnett, M.J., Blüthgen, N., and Saez-Rodriguez, J. (2017). Drug resistance mechanisms in colorectal cancer dissected with cell type–specific dynamic logic models. Cancer Res. 77, 3364–3375.

Fallahi-Sichani, M., Honarnejad, S., Heiser, L.M., Gray, J.W., and Sorger, P.K. (2013). Metrics other than potency reveal systematic variation in responses to cancer drugs. Nat. Chem. Biol. 9, 708–714.

Fey, D., Halasz, M., Dreidax, D., Kennedy, S.P., Hastings, J.F., Rauch, N., Munoz, A.G., Pilkington, R., Fischer, M., Westermann, F., et al. (2015). Signaling pathway models as biomarkers: patient-specific simulations of JNK activity predict the survival of neuroblastoma patients. Sci. Signal. 8, ra130.

Fitzgerald, J.B., Schoeberl, B., Nielsen, U.B., and Sorger, P.K. (2006). Systems biology and combination therapy in the quest for clinical efficacy. Nat. Chem. Biol. 2, 458–466.

Fröhlich, F., Kaltenbacher, B., Theis, F.J., and Hasenauer, J. (2017). Scalable parameter estimation for genome-scale biochemical reaction networks. PLoS Comput. Biol. 13, e1005331.

Fröhlich, F., Theis, F.J., and Hasenauer, J. (2014). Uncertainty analysis for non-identifiable dynamical systems: profile likelihoods, bootstrapping and more. In Computational Methods in Systems Biology, P. Mendes, J.O. Dada, and K.O. Smallbone, eds. (Springer International Publishing), pp. 61–72.

Garnett, M.J., Edelman, E.J., Heidorn, S.J., Greenman, C.D., Dastur, A., Lau, K.W., Greninger, P., Thompson, I.R., Luo, X., Soares, J., et al. (2012). Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature 483, 570–575.

Gutenkunst, R.N., Waterfall, J.J., Casey, F.P., Brown, K.S., Myers, C.R., and Sethna, J.P. (2007). Universally sloppy parameter sensitivities in systems biology models. PLoS Comput. Biol. 3, 1871–1878.

Hafner, M., Niepel, M., and Sorger, P.K. (2017). Alternative drug sensitivity metrics improve preclinical cancer pharmacogenomics. Nat. Biotechnol. 35, 500–502.

Haibe-Kains, B., El-Hachem, N., Birkbak, N.J., Jin, A.C., Beck, A.H., Aerts, H.J.W.L., and Quackenbush, J. (2013). Inconsistency in large pharmacogenomic studies. Nature 504, 389–393.

Hass, H., Masson, K., Wohlgemuth, S., Paragas, V., Allen, J.E., Sevecka, M., Pace, E., Timmer, J., Stelling, J., MacBeath, G., et al. (2017). Predicting ligand-dependent tumors from multi-dimensional signaling features. NPJ Syst. Biol. Appl. 3, 27.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). The Elements of Statistical Learning (Springer-Verlag).

Haverty, P.M., Lin, E., Tan, J., Yu, Y., Lam, B., Lianoglou, S., Neve, R.M., Martin, S., Settleman, J., Yauch, R.L., et al. (2016). Reproducible pharmacogenomic profiling of cancer cell line panels. Nature 533, 333–337.

Henriques, D., Villaverde, A.F., Rocha, M., Saez-Rodriguez, J., and Banga, J.R. (2017). Data-driven reverse engineering of signaling pathways using ensembles of dynamic models. PLoS Comput. Biol. 13, e1005379.

Herwig, R., Hardt, C., Lienhard, M., and Kamburov, A. (2016). Analyzing and interpreting genome data at the network level with ConsensusPathDB. Nat. Protoc. 11, 1889–1907.

Hindmarsh, A.C., Brown, P.N., Grant, K.E., Lee, S.L., Serban, R., Shumaker, D.E., and Woodward, C.S. (2005). SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers. ACM Trans. Math. Softw. 31, 363–396.

Hross, S., and Hasenauer, J. (2016). Analysis of CFSE time-series data using division-, age- and label-structured population models. Bioinformatics 32, 2321–2329.

Hug, S., Raue, A., Hasenauer, J., Bachmann, J., Klingmüller, U., Timmer, J., and Theis, F.J. (2013). High-dimensional Bayesian parameter estimation: case study for a model of JAK2/STAT5 signaling. Math. Biosci. 246, 293–304.

Ihle, N.T., Byers, L.A., Kim, E.S., Saintigny, P., Lee, J.J., Blumenschein, G.R., Tsao, A., Liu, S., Larsen, J.E., Wang, J., et al. (2012). Effect of KRAS oncogene substitutions on protein behavior: implications for signaling and clinical outcome. J. Natl. Cancer Inst. 104, 228–239.

Jagiella, N., Rickert, D., Theis, F.J., and Hasenauer, J. (2017). Parallelization and high-performance computing enables automated statistical Inference of multi-scale models. Cell Syst. 4, 194–206.e9.

Kaelin, W.G., Jr. (2005). The concept of synthetic lethality in the context of anticancer Therapy. Nat. Rev. Cancer 5, 689–698.

Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., and Hirakawa, M. (2010). KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Res. 38, D355–D360.

Karr, J.R., Sanghvi, J.C., Macklin, D.N., Gutschow, M.V., Jacobs, J.M., Bolival, B., Assad-Garcia, N., Glass, J.I., and Covert, M.W. (2012). A whole-cell computational model predicts phenotype from genotype. Cell 150, 389–401.

Khodayari, A., and Maranas, C.D. (2016). A genome-scale Escherichia coli kinetic metabolic model k-ecoli457 satisfying flux data for multiple mutant strains. Nat. Commun. 7, 13806.

Kholodenko, B.N. (2015). Drug resistance resulting from kinase dimerization is rationalized by thermodynamic factors describing allosteric inhibitor effects. Cell Rep. 12, 1939–1949.

Klaeger, S., Heinzlmeir, S., Wilhelm, M., Polzer, H., Vick, B., Koenig, P.A., Reinecke, M., Ruprecht, B., Petzoldt, S., Meng, C., et al. (2017). The target landscape of clinical kinase drugs. Science 358.

Klipp, E., Herwig, R., Kowald, A., Wierling, C., and Lehrach, H. (2005). Systems Biology in Practice (Weinheim, Germany: Wiley-VCH Press).

Koleti, A., Terryn, R., Stathias, V., Chung, C., Cooper, D.J., Turner, J.P., Vidović, D., Forlin, M., Kelley, T.T., D'Urso, A., et al. (2018). Data Portal for the Library of Integrated Network-based cellular Signatures (LINCS) program: integrated access to diverse large-scale cellular perturbation response data. Nucleic Acids Res. 46, D558–D566.

Kondofersky, I., Theis, F.J., and Fuchs, C. (2016). Inferring catalysis in biological systems. IET Syst. Biol. 10, 210–218.

Kuperstein, I., Bonnet, E., Nguyen, H.A., Cohen, D., Viara, E., Grieco, L., Fourquet, S., Calzone, L., Russo, C., Kondratova, M., et al. (2015). Atlas of

Cancer Signalling Network: a systems biology resource for integrative analysis of cancer data with Google Maps. Oncogenesis 4, e160.

Lehár, J., Zimmermann, G.R., Krueger, A.S., Molnar, R.A., Ledell, J.T., Heilbut, A.M., Short, G.F., Giusti, L.C., Nolan, G.P., Magid, O.A., et al. (2007). Chemical combination effects predict connectivity in biological systems. Mol. Syst. Biol. 3, 80.

Li, C., Donizelli, M., Rodriguez, N., Dharuri, H., Endler, L., Chelliah, V., Li, L., He, E., Henry, A., Stefan, M.I., et al. (2010). BioModels Database: an enhanced, curated and annotated resource for published quantitative kinetic models. BMC Syst. Biol. 4, 92.

Marcotte, R., Sayad, A., Brown, K.R., Sanchez-Garcia, F., Reimand, J., Haider, M., Virtanen, C., Bradner, J.E., Bader, G.D., Mills, G.B., et al. (2016). Functional genomic landscape of human breast cancer drivers, vulnerabilities, and resistance. Cell 164, 293–309.

Niepel, M., Hafner, M., Williams, E.H., Chung, M., Barrette, A.M., Stern, A.D., Hu, B., Gray, J.W., Birtwistle, M.R., Heiser, L.M., et al. (2017). A multi-center study on factors influencing the reproducibility of in vitro drug-response studies. bioRxiv. https://doi.org/10.1101/213553.

O'Neil, J., Benita, Y., Feldman, I., Chenard, M., Roberts, B., Liu, Y., Li, J., Kral, A., Lejnine, S., Loboda, A., et al. (2016). An unbiased oncology compound screen to identify novel combination strategies. Mol. Cancer Ther. 15, 1155–1162.

Opitz, D., and Maclin, R. (1999). Popular ensemble methods: an Empirical Study. jair 11, 169–198.

Penas, D.R., Henriques, D., González, P., Doallo, R., Saez-Rodriguez, J., and Banga, J.R. (2017). A parallel metaheuristic for large mixed-integer dynamic optimization problems, with applications in computational biology. PLoS One 12, e0182186.

Pozdeyev, N., Yoo, M., Mackie, R., Schweppe, R.E., Tan, A.C., and Haugen, B.R. (2016). Integrating heterogeneous drug sensitivity data from cancer pharmacogenomic studies. Oncotarget 7, 51619–51625.

Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., and Timmer, J. (2009). Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics 25, 1923–1929.

Raue, A., Schilling, M., Bachmann, J., Matteson, A., Schelker, M., Kaschek, D., Hug, S., Kreutz, C., Harms, B.D., Theis, F.J., et al. (2013). Lessons learned from quantitative dynamical modeling in systems biology. PLoS One 8, e74335.

Rubio-Perez, C., Tamborero, D., Schroeder, M.P., Antolín, A.A., Deu-Pons, J., Perez-Llamas, C., Mestres, J., Gonzalez-Perez, A., and Lopez-Bigas, N. (2015). In silico prescription of anticancer drugs to cohorts of 28 tumor types reveals targeting opportunities. Cancer Cell 27, 382–396.

Sanchez-Vega, F., Mina, M., Armenia, J., Chatila, W.K., Luna, A., La, K.C., Dimitriadoy, S., Liu, D.L., Kantheti, H.S., Saghafinia, S., et al. (2018). Oncogenic Signaling Pathways in The Cancer Genome Atlas. Cell 173, 321–337.e10.

Sanghvi, J.C., Regot, S., Carrasco, S., Karr, J.R., Gutschow, M.V., Bolival, B., Jr., and Covert, M.W. (2013). Accelerated discovery via a whole-cell model. Nat. Meth. 10, 1192–1195.

Schmidl, D. (2012) Bayesian Model Inference in Dynamic Biological Systems Using Markov Chain Monte Carlo Methods. (Dissertation: Fakultät für Mathematik).

Schoeberl, B., Pace, E.A., Fitzgerald, J.B., Harms, B.D., Xu, L., Nie, L., Linggi, B., Kalra, A., Paragas, V., Bukhalid, R., et al. (2009). Therapeutically targeting ErbB3: A key node in ligand-induced activation of the ErbB receptor-PI3K axis. Sci. Signal. 2, ra31.

Seashore-Ludlow, B., Rees, M.G., Cheah, J.H., Cokol, M., Price, E.V., Coletti, M.E., Jones, V., Bodycombe, N.E., Soule, C.K., Gould, J., et al. (2015). Harnessing connectivity in a large-scale small-molecule sensitivity dataset. Cancer Discov. 5, 1210–1223.

Smallbone, K., and Mendes, P. (2013). Large-scale metabolic models: From reconstruction to differential equations. Ind. Biotechnol. 9, 179–184.

Stapor, P., Fröhlich, F., and Hasenauer, J. (2018). Optimization and profile calculation of ODE models using second order adjoint sensitivity analysis. Bioinformatics 34, i151–i159.

Su, F., Viros, A., Milagre, C., Trunzer, K., Bollag, G., Spleiss, O., Reis-Filho, J.S., Kong, X., Koya, R.C., Flaherty, K.T., et al. (2012). RAS mutations in cutaneous squamous-cell carcinomas in patients treated with BRAF inhibitors. N. Engl. J. Med. 366, 207–215.

Swainston, N., Smallbone, K., Hefzi, H., Dobson, P.D., Brewer, J., Hanscho, M., Zielinski, D.C., Ang, K.S., Gardiner, N.J., Gutierrez, J.M., et al. (2016). Recon 2.2: from reconstruction to model of human metabolism. Metabolomics 12, 109.

The Cancer Genome Atlas Network (2012). Comprehensive molecular portraits of human breast tumours. Nature 490, 61–70.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. J.R. Stat. Soc. B 58, 267–288.

Vehlow, C., Hasenauer, J., Kramer, A., Raue, A., Hug, S., Timmer, J., Radde, N., Theis, F.J., and Weiskopf, D. (2013). iVUN: interactive Visualization of Uncertain biochemical reaction Networks. BMC Bioinformatics 14, S2.

Villaverde, A.F., and Banga, J.R. (2014). Reverse engineering and identification in systems biology: strategies, perspectives and challenges. J.R. Soc. Interface 11, 20130505.

Villaverde, A.F., Henriques, D., Smallbone, K., Bongard, S., Schmid, J., Cicin-Sain, D., Crombach, A., Saez-Rodriguez, J., Mauch, K., Balsa-Canto, E., et al. (2015). BioPreDyn-bench: a suite of benchmark problems for dynamic modelling in systems biology. BMC Syst. Biol. 9, 8.

Weinstein, I.B., and Joe, A.K. (2006). Mechanisms of Disease: oncogene addiction – a rationale for molecular targeting in cancer therapy. Nat. Clin. Pract. Oncol. 3, 448–457.

Wierling, C., Herwig, R., and Lehrach, H. (2007). Resources, standards and tools for systems biology. Brief Funct. Genomic Proteomic 6, 240–251.

Wierling, C., Kessler, T., Ogilvie, L.A., Lange, B.M.H., Yaspo, M.-L., and Lehrach, H. (2015). Network and systems biology: essential steps in virtualising drug discovery and development. SI Netw.-Based Discov. Syst. Biol. 15, 33–40.

Wishart, D.S., Knox, C., Guo, A.C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z., and Woolsey, J. (2006). DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res. 34, D668–D672.

Yang, W., Soares, J., Greninger, P., Edelman, E.J., Lightfoot, H., Forbes, S., Bindal, N., Beare, D., Smith, J.A., Thompson, I.R., et al. (2013). Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res. 41, D955–D961.

Zhang, J., Yang, P.L., and Gray, N.S. (2009). Targeting cancer with small molecule kinase inhibitors. Nat. Rev. Cancer 9, 28–39.

Zhang, W., and Liu, H.T. (2002). MAPK signal pathways in the regulation of cell proliferation in mammalian cells. Cell Res. 12, 9–18.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Drug-response data of CCLE	Barretina et al. (2012)	https://portals.broadinstitute.org/ccle/
Mutation data of CCLE	Barretina et al. (2012)	https://portals.broadinstitute.org/ccle/
RNAseq data of CCLE		https://portal.gdc.cancer.gov/legacy-archive
Post-processed drug response data of GDSC	Pozdeyev et al. (2016)	http://tanlab.ucdenver.edu/QAPC/
Software and Algorithms		
MATLAB (including the Optimization Toolbox, the Statistics Toolbox and the Symbolic Toolbox)	Mathworks	RRID: SCR_001622
Advanced MATLAB Interface for CVODES and IDAS (AMICI)	Fröhlich et al. (2017)	https://github.com/ICB-DCM/AMICI https://doi.org/10.5281/zenodo.579891
Parameter Estimation Toolbox (PESTO)	Stapor et al. (2018)	https://github.com/ICB-DCM/PESTO https://doi.org/10.5281/zenodo.579890
SBML model of canonic human ERBB, RAS and PI3K/AKT cancer signaling pathways	This paper	https://doi.org/10.5281/zenodo.1472794 https://doi.org/10.24433/CO.0756760d-cb46-4ef1-9d9a-07c4cef40baa
Tailored implementation of parameter estimation, uncertainty analysis and statistical modeling	This paper	https://doi.org/10.5281/zenodo.1472794 https://doi.org/10.24433/CO.0756760d-cb46-4ef1-9d9a-07c4cef40baa

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for software and algorithms should be directed to the Lead Contact Jan Hasenauer (jan.hasenauer@helmholtz-muenchen.de).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

CCLE Data
In this study, we employ previously published data for human cancer cell lines.

We downloaded RNAseq BAM-files for 780 CCLE cell lines from the Cancer Genomics Hub (https://cghub.ucsc.edu/) in April 2014. The same data, including additional cell lines is now available for download in the Cancer Genomics Cloud (https://cgc.sbgenomics.com/). Gene expression was quantified as Reads Per Kilobase Million (RPKM) using gene models from Ensembl Release 73. Mutation data was downloaded from the CCLE data portal (https://portals.broadinstitute.org/ccle/data/, file CCLE_hybrid_capture1650_hg19_NoCommonSNPs_NoNeutralVariants_CDS_2012.05.07.maf). RNA allele frequencies for the mutations were determined from the downloaded RNAseq BAM-files using SAMtools mpileup (http://www.htslib.org/). Drug response data were downloaded from: https://data.broadinstitute.org/ccle_legacy_data/pharmacological_profiling/CCLE_NP24.2009_Drug_data_2015.02.24.csv.

We identified 10 different tissues for which known driver mutations and respective pathways were implemented in the model and for which RNAseq data was collected in sufficiently many originating cell-lines. Cell lines from tissues that include many cell lines with RNAseq data were included in the training/test set while cell lines from tissues with few cell lines with RNAseq data were included in the independent set. This yielded a total of 123 cell lines originating from the tissues breast, large-intestine, lung, pancreas and skin, which were included in the training/test data-set and 31 cell lines originating from the tissues kidney, soft tissue, ovary and stomach which were included in the independent test set. For the training/test data we considered 120 of the 123 available cell lines to ensure equally sized training and test sets in all cross-validations.

To generate test and training datasets from the processed CCLE data, we performed 20-80% splits on the cell-line level, which yielded 5 training sets with 96 cell lines and test sets with 24 cell lines. The split was performed such that the tissue distribution in the individual training sets is maximally similar. The number of experimental conditions in the training sets varies from 5390 to 5403 due to incomplete data for individual cell lines.

GDSC Data

We used postprocessed drug response data provided by Pozdeyev et al. (2016) (http://tanlab.ucdenver.edu/QAPC/Downloads/Drug_Sensitvity_Metrics.zip). For the individualization of cell-line specific simulations we used the same postprocessed RPKM that we extracted for the analysis of CCLE drug response data.

METHOD DETAILS

Model Development

The mechanistic model was developed using PyBioS (Wierling et al., 2007), a web-based platform for modeling of complex molecular systems. We exploited several features of PyBioS, including the modular formulation of large-scale models based on individual pathways and their interactions. For model development we employed information from ConsensusPathDB (Herwig et al., 2016). The information was manually curated and implemented in the model using a standard operating procedure (SOP). The SOP ensured the model quality and the compatibility of different pathway models. The following paragraphs concisely describe the model development in accordance to these SOPs and outline the model topology.

Models for the ERBB, RAS and AKT, AP1 and MYC pathways were generated in PyBioS as modules to be merged into a consistent and comprehensive model of multiple cancer-related signaling pathways. Information from scientific literature and pathway databases were considered for incorporation into the pathway modules. PyBioS is using templates for the annotation of individual reaction types, such as protein biosynthesis, translocation, decay, (auto-/de-)phosphorylation, complex formation, GTP (de-)activation or cleavage reactions. The reaction kinetics used by the templates follow a mass action law structure for reversible or irreversible reactions. PyBioS is using ENSEMBL gene identifiers for the unique annotation of the model species such as genes and proteins. The ERBB module covers interaction of three different receptor tyrosine kinases (EGFR, ERBB2, ERBB4) and the non-kinase receptor ERBB3 with at least one of the eleven ligands that are known to interact with at least one of these receptors. Ligand receptor complexes undergo autophosphorylation reactions reflecting auto-activation on tyrosine residues of the receptor's cytoplasmic domain that serve as docking sites for distinct adapter molecules, e.g. GAB1, GRB2. Annotation of these phosphorylation sites in the model and relevant binding partners is based on literature search and annotations in databases, such as the UniProt database (http://www.uniprot.org/). Further complex formation reactions in the model reflect recruitment of proteins such as SOS1 or PIK3CA that in turn trigger activation of RAS and AKT signaling, well known events of ERBB signaling in cancer cells. AKT and RAS signaling activation in cancer can be triggered by several receptor tyrosine kinases as well as by activating mutations of key pathway components, e.g. HRAS, KRAS, NRAS, PIK3CA or AKT. Within the RAS module, activated RAS primarily triggers RAF/MEK/ERK kinase cascade activation and also triggers activation of the AKT signaling module via complex formation with PIK3CA and subsequent generation of the second messenger phosphatidylinositol-(3,4,5)-trisphosphate (PIP3). The functional effects of RAS or PIK3CA activating mutations are identical to the signaling elicited by the non-mutated species but modeled as independent of upstream activating events. Within the AKT module, PIP3 activates AKT1, AKT2 and AKT3 via binding to PDPK1 and subsequent phosphorylation of AKT isoforms. The functional effect of activated AKT is then modeled through phosphorylation reactions of distinct substrate proteins that are known to activate or repress other signaling pathways, e.g. MTOR (AKT1S1, RHEB), apoptosis (BAD) or the G1S cell cycle checkpoint (CDKN1A/B). Activation of transcription by MYC, AP1 and FOXO factors is modeled as phosphorylation reactions through ERK or AKT, and therefore phosphorylated species are used as primary readouts for pathway activation in absence of more complex modules reflecting transcriptional co-regulation.

Based on this ERBB/RAS/AKT model and CCLE drug screening data, we integrated drugs targeting key pathway components such as erlotinib (EGFR and ERBB2) or selumetinib (ERK) by generating reversible complex formation reactions of drugs with their (main and off-) target species, namely ligand receptor complexes in case of erlotinib or protein species in case of selumetinib. K_D values for these complex formation reactions were set according to kinase inhibitor screen data (Barrett et al., 2008; Davis et al., 2011). All values are reported in the SBML file.

The phosphorylation and complex formation reactions within the ERBB, RAS and AKT signaling modules provide links to other signaling pathways and thereby reflect crosstalk between pathways. However, modules secondarily affected by ERBB, RAS and AKT are of similar complexity and their inclusion into the model used in this study would have significantly extended the ODE system and thereby the number of parameters for optimization.

Based on the reaction details (substrates, products, enzymes, stoichiometries and kinetic laws) the reaction rate equations for the biochemical reaction network were generated. This provides a system of ordinary differential equations (ODEs) describing the model. Each pathway module was checked for structural consistency, i.e. for each module the relative behavior of different readout species of the respective modules were manually checked based on a systematic structural testing of key species such as ligands, receptors and other regulators which were defined during the annotation of the model. Activating simulations were performed for the external activators (ligands, if part of the respective model), knock-out simulations for receptors (if part of the respective model), and knock-out and over-expression in silico experiments for additional activator species were performed and the expected down-stream structural behavior based on the biological expectation or evidence was validated. These structural tests were performed with all kinetic parameters set to one. Finally all pathway modules, modules of mutated genes/proteins and drug/target interactions were merged into a comprehensive final model using PyBioS and exported in SBML format.

Model Encoding in SBML

The developed model is made publicly available as a supplement to this manuscript. The model features extensive annotation, including UniProt and Ensembl IDs. Phosphorylations are indicated in the name of the species by a preceding "P[$X;$Y;...]-" where $X and $Y specify the phosphorylation site using a one letter amino acid code, followed by the amino acid number. Mutations are indicated in the name of the species by a preceding "MutAA[$Z]-", where $Z specifies the mutation site using standard sequence variant nomenclature. Stoichiometric coefficients $N of complex constituents are indicated by a trailing "[$Nx]" for $N>1.

The SBML file encodes an ordinary differential equation (ODE) model of the form

$$\frac{dx}{dt} = S \cdot v(x, \theta, d, c), x(0) = x_0,$$

with concentration vector x and its initial condition x_0, stoichiometric matrix S and flux vector v. The parameter vector θ provides the reaction rates, e.g. binding affinities. The vector d provides the drug concentrations and the vector c provides the expression levels for the gene products and respective variants for a cell line. The flux vector $v(x, \theta, d, c)$ is derived according to the law of mass action. As the highest order reactions implemented in the model are complex formation reactions with two binding partners, $v(x, \theta, d, c)$ is linear in θ and up to quadratic in x. The dependency on d and c is explained in the following.

To consider different drug treatments and cell lines, only d and c have to be changed. In the SBML model, the species encoding drug concentrations have the same name as the respective drug and are located in compartment_2 (extracellular). To consider different cell lines the vector c has to be changed. In the SBML file, the Reads Per Kilobase of transcript per Million (RPKM) values are encoded as local parameters r$X_k_RPKM2protein of the respective synthesis reactions reaction_$X, where $X is the reaction id. The synthesis rate is defined as product of the RPKM value (r$X_k_RPKM2protein), a gene specific scaling constant (r$X_k_GeneSpecificScaling) and a species that encodes the presence of the respective gene (annotated with ENSEMBL ID). For all simulations, we always used the RPKM values of the untreated condition. In the current implementation the gene specific scaling constants are always set to 1 and not estimated during parameterization. The parameter vector θ is generic and can be used for different cell-lines and drug treatments. The SBML model implements a representative parameter estimate.

In the SBML model the viability output variable is specified as an assignment rule. The viability output is computed as fraction of weighted sums of concentrations of active forms of transcription factors

$$y^{c,d} = \frac{\sum_i^N \omega_i^{pos} x_{i,c,d}^{pos}}{1 + \sum_j^M \omega_j^{neg} x_{j,c,d}^{neg}},$$

in which $x_{i,c,d}^{pos}$ and $x_{i,c,d}^{neg}$ denote the concentrations of transcription factors, for a particular cell line and drug treatment combination, with a positive and negative influence on viability, respectively. The corresponding weights are denoted by ω_i^{pos} and ω_i^{neg} and were estimated during model parameterization. The model captures the effect of $N = 12$ transcription factors, which in their molecularly active state have positive influence on cell viability: P[S63;S73]-JUN[2x], P[S252;S265]-FOSL1:P[S63;S73]-JUN, P[T69;T71]-ATF2:P[S63;S73]-JUN, P[S374;T325;T331]-FOS:P[S63;S73]-JUN P[Y701]-STAT1[2x], P[Y705]-STAT3[2x], P[Y694]-STAT5A[2x], P[Y699]-STAT5B[2x], MAX-001:P[S62]-MYCN, MAX:P[S62]-MYC, P[S324;S383]-ELK1, and P[S133]-CREB1 (active states). Furthermore, $M = 4$ transcription factors, that in their molecularly active state have negative influence on cell viability, were included: FOXO1, FOXO3, FOXO4 and FOXO6 (inactive states). In all cases only the species with nuclear localization were considered to be active.

The model employs experimentally derived drug-target binding affinity (KD) values for the drugs CHIR-265, erlotinib, lapatinib, PLX4720, selumetinib, sorafenib and vandetanib, which were obtained from Davis et al. (2011). For PD0325901 the model employs the inhibitory concentrations (IC50), which was measured in a cell-free assay by Barrett et al. (2008).

We note that the model includes several components that were not used in the presented analysis. This includes the option to specify gene specific scaling constants to individually adjust synthesis rates. Furthermore, the small molecular kinase inhibitor sorafenib was modeled. However, as none of the considered cell lines responded to sorafenib and as sorafenib targets several components that are not captured by the model, the corresponding response data was not considered in this study.

Numerical Simulation

The compilation and numerical simulation of the model was performed using the MATLAB toolbox AMICI (Fröhlich et al., 2017) (https://doi.org/10.5281/zenodo.579891). AMICI employs the backward differentiation method implemented in the SUNDIALS solver package (Hindmarsh et al., 2005). We used the KLU linear solver with AMD reordering and relative and absolute error tolerance 10^{-8}.

As the viability measurements were taken after 72 to 84 hours, we assumed that the signaling had reached a steady state. Consequently, we simulated the model until all species reached steady state. To find the steady state for the untreated condition of a cell line, the forward simulation was initialized with zero initial conditions and run until a time t for which the maximal absolute value of the regularized relative derivative was smaller than 10^{-6},

$$\max_i \left| \frac{(S \cdot v(x(t, \theta, d), \theta, d, c))_i}{x_i + 10^{-6}} \right| < 10^{-6}$$

where x denotes the state of the system at time t or drug concentration vector $d = 0$ and cell line c. For all treated conditions of a cell line, the forward simulation was initialized with the steady state of the corresponding untreated condition and run until the same convergence criterion was met. As the model contains non-linear reaction rates, the existence of a unique stable steady state is theoretically not guaranteed. Interestingly, all simulations reached steady state for the considered model topology.

The forward and backward simulation of experimental conditions was parallelized using the MATLAB command *parfor*, which implements OpenMP parallelization. As our cluster infrastructure features 8 core nodes, we parallelized each gradient computation over 8 cores (1 master, 7 workers), thereby avoiding inter-node communication overhead. The different local optimizations were performed on different nodes.

Stability Analysis

In the context of large-scale ODE models of metabolic networks, previous studies have reported substantial problems with instability of steady states (Smallbone and Mendes, 2013). Instability of steady states may lead to problems in numerical simulations and can necessitate the premature termination of simulations (Khodayari and Maranas, 2016). For the proposed model we did not encounter diverging solutions or other numerical problems that could be attributed to the instability of steady states. In a retrospective analysis, we performed linear stability analysis for all estimated parameter values and found that the largest eigenvalue of the Jacobian is always strictly smaller than -10^{-4} (Figure S7). This corroborates the finding that the model is generally numerically well behaved.

QUANTIFICATION AND STATISTICAL ANALYSIS

Parameterization

To estimate the model parameters θ, we used the measurement data for the viability in the treated condition relative to the untreated condition, $y_m^{c,d}/y_m^{c,0}$, provided in the CCLE dataset. These data were fitted using a sum-of-squared-residuals objective function

$$J(\theta) = \frac{1}{2} \sum_{c \in C} \sum_{d \in D_c} \left(\frac{y_m^{c,d}}{y_m^{c,0}} - \frac{y^{c,d}(\theta)}{y^{c,0}(\theta)} \right)^2,$$

in which $c \in C$ is cell-line specific and $d \in D_c$ denotes the drug concentration. This objective function is equivalent to the negative log-likelihood function under the assumption of additive independent and identically distributed standard normally distributed measurement noise for the relative viability measurements. To minimize the objective function we used multi-start local optimization (Raue et al., 2013) implemented in the MATLAB toolbox PESTO (https://doi.org/10.5281/zenodo.579890). Parameters were constrained to a $[10^{-2}, 10^2]$ hypercube. For each local optimization run, parameters were drawn log-uniformly from this hypercube, followed by 100 optimization iterations of the MATLAB fmincon interior-point algorithm in logarithmic parameters. An exemplary implementation of these methods is provided at https://doi.org/10.5281/zenodo.1472794 in the file runSyntheticExample.m in folder estimation_example. The MATLAB interior-point implementation applies a regularization that is symmetric around the center of the parameter domain (here $10^0 = 1$) to every parameter. As the regularization term attains the lowest value at this center point, the optimization will preferentially estimate parameters to value 1, as long as this does not lead to worse objective function values. Parameter estimates with value 1 were consistently observed for parameters with no influence on model outputs, such as parameters of reactions that involve mutated species that were present in none of cell-lines in the training set.

A high-performance-computing-ready standalone executable was generated from the parameterization pipeline implemented in MATLAB using the MATLAB Compiler toolbox. For every cross-validation we performed 10 local optimization runs. As no communication was necessary between optimization runs, each could be submitted as a separate job to the cluster. In total we submitted 50 jobs using 8 cores each, resulting in a total parallelization over 400 cores.

Adjoint Sensitivity Analysis

To compute the gradient of the objective function, we employed adjoint sensitivity analysis (Fröhlich et al., 2017). Adjoint sensitivity analysis facilitates the computation of the gradient of a scalar function $\gamma(\theta)$, which typically is the objective function $\gamma(\theta) = J(\theta)$. However, the established approach assumes that the individual summands of the objective function can be assessed using a single simulation of a single model. For the objective function we formulated in the previous section, a single summand requires the simulation of two experimental conditions, one treated and the untreated condition for a particular cell line. Hence, single simulation would have to combine simulation of one cell-line in two experimental conditions, which would require a single ODE model with twice the number of state variables and repeated simulation of the control condition. While this would be feasible, it seemed impractical and computationally unnecessary demanding.

In this study, we exploited the fact that we only consider the viability output as a scalar observable, $y^{c,d}(\theta)/y^{c,0}(\theta)$, for every experimental condition. Instead of using adjoint sensitivities to calculate the objective function directly, we instead computed the gradients of $\gamma(\theta) = y^{c,d}(\theta)$ and $\gamma(\theta) = y^{c,0}(\theta)$. Using $\partial y^{c,d}(\theta)/\partial\theta$ and $\partial y^{c,0}(\theta)/\partial\theta$, the gradient is

$$\frac{\partial J(\theta)}{\partial \theta} = \sum_{c \in C} \sum_{d \in D_c} \left(\frac{y_m^{c,d}}{y_m^{c,0}} - \frac{y^{c,d}(\theta)}{y^{c,0}(\theta)} \right) \frac{\frac{\partial y^{c,d}(\theta)}{\partial\theta} y^{c,0}(\theta) - y^{c,d}(\theta) \frac{\partial y^{c,0}(\theta)}{\partial\theta}}{y^{c,0}(\theta)^2}.$$

This approach should reduce the computation time compared to the established approach at least by a factor of two. The effective model size for adjoint sensitivity analysis using this approach is equal to the size of the pathway model and simulation has to be performed only for a single experimental condition.

We note that the original adjoint sensitivity approach is also applicable to models with more than one output variable.

Numerical Benchmarking of Forward and Adjoint Sensitivity Analysis

To compare different methods for gradient evaluation, we assessed the computation time for a single gradient evaluation on the full training set. For sequential and parallel gradient evaluation using adjoint sensitivity analysis, we measured computation time. As this would have been too time consuming for forward sensitivities, we first assured that the computation time for individual experimental conditions is comparable and then extrapolated to all experimental conditions. The computation time was evaluated on the training set of the first cross-validation for 10 randomly sampled parameter vectors. For the difference between forward and adjoint sensitivity analysis and sparse and dense solvers, we only evaluated the simulation time for the untreated condition of a single cell-line. The performance was evaluated based on 100 samples with a randomly drawn parameter vector and a randomly drawn cell-line. The computation time was then normalized such that the median for the sparse adjoint approach matched the computation time for the full training set.

Ensemble Averaging

We used ensemble averaging to reduce the effect of overfitting and the variance of predictors. For the mechanistic model we used an ensemble model based on five optimization runs that achieved the lowest objective function value. The model was individualized to cell lines from the test and independent test set and simulated with parameter values from these optimization runs. Predictions were then computed as median over these five simulations. The ensemble averaging was solely based on results from the training set. The test and independent test set were only used for validation.

Uncertainty Analysis

To assess the uncertainty of parameters we considered the eigenvalue spectrum of the Fisher Information Matrix (FIM). Small eigenvalues indicate large uncertainties in the direction of the respective eigenvector while large eigenvalues indicate small uncertainties. The eigenvalue spectrum was evaluated for the best five optimization runs for every cross-validation.

The FIM was computed by summing the dyadic product of adjoint sensitivities over all experimental conditions

$$FIM_\theta = \sum_{c \in C} \sum_{d \in D_c} \frac{1}{y^{c,0}(\theta)^4} \left(\frac{\partial y^{c,d}(\theta)}{\partial \theta} y^{c,0}(\theta) - \frac{\partial y^{c,0}(\theta)}{\partial \theta} y^{c,d}(\theta) \right)^T \left(\frac{\partial y^{c,d}(\theta)}{\partial \theta} y^{c,0}(\theta) - \frac{\partial y^{c,0}(\theta)}{\partial \theta} y^{c,d}(\theta) \right).$$

As the number of experimental conditions ($\sim$5,400) exceeds the number of parameters ($\sim$4,100) the FIM_θ could theoretically have full rank.

For parameter derived readouts z, such as viability readouts as well as state variables, a similar quantification of the uncertainty is possible by considering a transformation FIM_z of FIM_θ. The transformation is obtained by multiplication with the respective parameter derivatives

$$FIM_z = \frac{\partial z}{\partial \theta} FIM_\theta \frac{\partial z}{\partial \theta}^T.$$

For state variables the formula for sensitivity of the steady state with respect to the model is according to the implicit function theorem

$$\frac{\partial x}{\partial \theta} = - \left(S \cdot \frac{\partial v}{\partial x} \right)^{-1} S \cdot \frac{\partial v}{\partial \theta},$$

assuming that the system is in steady state,

$$S \cdot v = 0.$$

For the assessment of the uncertainties of the steady state based on FIM_z, we only considered state variables with non-zero steady state. The state variables with steady state equal to zero correspond to molecular species that are not expressed.

Statistical Modeling

For the comparison of model performances, we trained a series of statistical models for the quantitative prediction of cell viability at different drug concentrations, based on the exact same training data sets, cross-validation setup and test data sets that were used for the mechanistic model. The model quality was evaluated for three sets of predictor variables: 1) mutation genotype data; 2) gene expression data; and 3) genotype and gene expression data. In addition, we also provided the network topology as input to some of the models. Model training was performed for each drug and each drug concentration independently by nested cross-validation, where the outer 5-fold cross validation loop split the data into training set (80% of the data) and test set (20% of the data). For each model and each training set we estimated the model parameters by optimizing the performance measured as root mean

OPEN
ACCESS
CellPress

squared error (RMSE) in the inner cross validation loop splitting the training data again into training and validation sets or by bootstrapping of the training data (random forest).

We used the R implementations of the following models: 1) linear regression with LASSO penalty (glmnet package); 2) Random Forest (randomForest package); 3) graph regularized linear regression (glmgraph package); and 4) linear regression with LASSO penalty on augmented data, including additional interaction terms. The interaction terms were defined based on the network topology used in the mechanistic model. The adjacency matrix was extracted from the Jacobian of the right-hand side of the differential equation. For a pair of genes, the Jacobian was reduced to rows or columns corresponding to species that include the corresponding proteins in any form (phosphorylated, cleaved or bound). Two genes were defined to be adjacent when the corresponding submatrix of the Jacobian has at least one non-zero entry. We augmented the data set either with all pairwise interactions (products of individual variables) between variables of the same type (genotype or gene expression) or interactions between genes that are connected by paths in the network not longer than $d = 1, 2$ or 3 steps. The optimal parameter λ for LASSO models, λ_1 and λ_2 or the graph regularized LASSO model were selected as the largest λ that produces an RMSE within one standard error of the minimal RMSE (Hastie et al., 2009) in an 8-fold inner cross validation. Random forest regression models were trained by selecting parameters that minimize the out-of-bag error (Breiman, 2001). We optimized over the number of variables randomly sampled as candidates for each split, the number of trees in the forest ranging from 50 to 500 and the maximum leaf node size criterion ranging from 1 to 1/3 of the data set. All models were then applied to the test set and performance was assessed as the RMSE or Pearson correlation coefficient between predicted and observed cell viability.

Finally, we used the models trained in each round of the cross validation and applied them to an independent test set. Performance was assessed as the RMSE or Pearson correlation coefficient between predicted and observed cell viability and averaged over the models trained in each of the five cross validation rounds.

DATA AND SOFTWARE AVAILABILITY

The model, code for the simulation and inference as well as the result files are available on Code Ocean: https://doi.org/10.24433/CO.0756760d-cb46-4ef1-9d9a-07c4cef40baa and Zenodo: https://doi.org/10.5281/zenodo.1472794.