

Logic modeling in quantitative systems pharmacology

Pauline Traynard^{*1}, Luis Tobalina^{*2}, Federica Eduati^{*3}, Laurence Calzone^{1,#}, Julio Saez-Rodriguez^{2,3,#}

* Co-first authors

Co-senior authors

Affiliations

¹ Institut Curie, PSL Research University, Mines Paris Tech, Inserm, U900, F-75005, Paris, France.

² RWTH Aachen University, Faculty of Medicine, Joint Research Centre for Computational Biomedicine, MT12 Wendlingweg 2 D-52074 Aachen, Germany

³ European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus CB10 1SD Hinxton, United Kingdom

Introduction

Here we present logic modeling as an approach to understand deregulation of signal transduction in disease and to characterize a drug's mode of action. We discuss how to build a logic model from literature and experimental data and how to analyze the resulting model to obtain insights of relevance for systems pharmacology. Our workflow uses the free tools *OmniPath* (network reconstruction from literature), *CellNOpt* (model fit to experimental data), *MaBoSS* (model analysis), and *Cytoscape* (visualization).

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1002/psp4.12225

Finding effective drugs and understanding how they work still pose considerable challenges with profound effects on human health. Several options exist to find compounds that could be used to treat a specific disease, such as target-based or phenotypic screening approaches.¹ More recent developments in experimental techniques such as shRNA² and CRISPR-Cas9³ or microfluidics^{4,5} promise to ease the discovery of new targets and drugs by increasing the number of targets that can be tested at once and decreasing the amount of biological material necessary to perform the experiments. However, a target discovered with such techniques may not be necessarily actionable in the clinical setting.^{6,7} A compound that showed promising results in the laboratory may fail to be effective on patients. Understanding how a drug acts at the systems level is key to increasing the likelihood of success. Likewise, understanding the molecular basis driving a disease can be of great help in the search for its cure.

Mechanistic models can help to better understand a drug's mode of action and predict the behavior of a biological system in response to drugs. The nascent field of Quantitative Systems Pharmacology aims to address these challenges by combining mechanistic models with pharmacokinetic modeling.⁸⁻¹⁰ It is known that many diseases involve alterations in the signaling pathways used by cells to interpret the cues from their environment.¹¹⁻¹⁴ Likewise, many drugs are designed to target components of these signaling pathways. These different pathways are not merely linear signaling conduits activated by different stimuli, but are also interconnected via crosstalk mechanisms to regulate each other, giving rise to signaling networks.

Regulatory Networks

The structure and function of signaling networks are complex, and they are differently deregulated in different biological contexts in non-trivial ways. Previous clinical studies have showed that inhibiting the same oncogenes can vary in efficacy, depending on the patient. The best example is the treatment of BRAF mutations in melanoma compared to colon carcinomas. In melanoma, a particular BRAF activating mutation, V600E, leads to the activation of signaling pathways involved in proliferation. Treatments targeting specifically the mutation, such as vemurafenib, show an efficient immediate response in melanomas. However, colon cancer patients with the same mutation do not respond to the treatment.¹⁵ The difference between these two responses to the drug is due to different cellular contexts. Only present in colon cancer, a feedback loop on EGFR activation leading to the activation of the mitogen-activated protein kinases (MAPK) pathway, through RAS, may be responsible for the poor outcome. This example highlights the importance of accurately describing the networks regulating these signaling pathways and their crosstalks. To ensure efficiency of the drug treatments, a good knowledge of these complex interactions and how patient mutations affect the cellular fate is necessary.

The regulatory networks used in these models are generally extracted from databases and literature. These sources represent the current knowledge available about interactions involving the proteins of interest. However, little information is known about which regulations are specific to a given cell type or a particular biological context. This information is important to highlight the specific mechanisms of regulation in different contexts. For example, comparing signaling pathways of healthy and diseased cells allows investigating which mechanisms are deregulated as a cause or an effect of the disease.^{16,17} Similarly the

comparison of pathway activity in different cells from the same cancer type allows investigating mechanism of resistance to drugs, which can be exploited to suggest personalized therapies.^{18–20} If experimental data are available, optimization procedures can be used to refine the initial networks to be cell line or context specific.

Basics of Logic Modeling

Mathematical modeling can help to decipher these complex mechanisms.¹⁸ Several modeling formalisms exist to deal with complex gene regulatory and signaling network structures.^{11,21,22} Because models can also provide insights into a drug's mode of action, scientists are also working towards building bridges between drug design and network modeling techniques.^{23–25}

Among the modeling techniques, logic(al) modeling has proven to be very versatile and able to provide useful biological insights.^{26–28} It has been applied for studying several biological phenomena (e.g. developmental processes,²⁹ haematopoiesis³⁰ or cell fate decision³¹) but primarily used in signaling and gene regulation.²⁷ Compared to other modeling techniques commonly used to describe biological systems, such as ordinary differential equations based on chemical kinetics, logic models are better suited to describe medium or large scale networks, where detailed biological knowledge is often incomplete and where a more schematic representation of the system can improve the overall interpretability of the results. However, it is necessary to keep in mind that the formulation of logic rules implies a simplification of the described biological system and might not be able to fully capture the complexity of the underlying system.³²

Within the logic formalism, signaling networks are modeled by defining a signed and directed causal regulatory network as well as rules to update the state of its components.^{22,26,28} The sign of the interaction depends on whether it is an activation or an inhibition, and the direction indicates which node acts over another one (e.g. in a kinase-substrate interaction where the kinase activates its substrate, the edge will go from the kinase to the substrate and its sign will be positive). In these networks, proteins are represented by nodes and interactions between proteins are represented by edges. The state of a node can represent its activation status. This way, the nodes can take different values: 0 (False) or 1 (True), in the case of Boolean models; or discrete values: 0, 1, 2 or higher, in the case of multi-valued models for representing distinct activation states. These states are updated following defined update rules, which take into account the state of input nodes and the type of interaction (positive or negative). Several different variations have been developed ranging from Petri nets³³ to logic-based Ordinary Differential Equations (ODEs).³⁴ An overview on some of the different formalisms available to model signaling networks (interaction graphs, logical models and logic-based ordinary differential equations - ODE - models) and the types of analyses that can be achieved with them is reviewed elsewhere.³⁵ By including nodes corresponding to the targets of different drugs in these models, their values can be manipulated in the simulations to find out what would be the effect of these compounds when applied to the system.

Biological Applications of Logic Modeling

The examples where logic modeling has been applied to understand biological mechanisms are numerous and go from the study of T-cell receptor signaling^{36,37} to cell-fate decision,³⁸ mammalian cell cycle³⁹ or host immune response.⁴⁰ In particular, cancer research has

motivated many of the applications of logic modeling and fostered several methodological developments, because it requires to study large signaling and regulatory networks.⁴¹ For example, discrete logical modeling has been used to describe the DNA damage response signal transduction network in human epithelial cells and predict candidate target proteins for sensitization of carcinomas to DNA-damaging agents.⁴¹ Also, the ErbB network was explored with a model built from prior knowledge and protein phosphorylation data to understand the drug resistance mechanisms of breast cancer cell lines.⁴² Logic modeling was applied to bladder cancer to investigate the effects of gene alterations leading to invasiveness.⁴³ However, the use of logic modeling tools is not limited to cancer applications, and research in other diseases also benefit from these techniques. For instance, Boolean modeling was applied to the study of Systemic Lupus Erythematosus, to stratify patients and find the best matching treatment.⁴⁴ It has also been used to study the immunological response to infections,⁴⁵ or to understand apoptosis given its relevance in diseases like Alzheimer and Parkinson.³¹ These and other examples^{46,47} illustrate the value of logic modeling to enhance our understanding of the systemic effect of therapies. The models provide a formal tool to quickly evaluate *in-silico* the effect of targeting one specific component of the model or explore the effects of possible drug combinations.^{20,48–50} It is also possible to assess how changes in the network (e.g. missing or inactive receptor, etc.) may affect the effect of a drug.

In this tutorial, we show the different steps involved in logic modeling on a prostate cancer example that involves some of the key phosphorylation pathways of this malignancy, and that we use to predict cell survival in different conditions (**Fig 1**). The steps include the building of the signaling network, its improvement using available data and its simulation and analysis geared towards obtaining useful biological insights and predictions. We also present several tools that are useful for such a workflow, namely *Omnipath*⁵¹ to construct the signaling network, *CellNOpt*⁵² to build a model trained to data, *MaBoSS*⁵³ to simulate and predict treatment response, and *Cytoscape*⁵⁴ to visualize some results of the simulations. It should be noted that many other excellent tools to model and analyze signaling networks exist, such as BoolSim,⁵⁵ BoolNet,⁵⁶ GINsim,⁵⁷ etc. Many of these are compatible via the SBML-qual format,⁵⁸ and their development is coordinated by the CoLoMoTo initiative.⁵⁹

Selection of a system to study

The decision to use logic models is usually spurred by the wish to gain a mechanistic understanding of a biological system for which the size, the lack of knowledge, or both, precludes more refined approaches such as reaction-based modeling. The system of study might be a specific pathway, a cell line or a disease, for example. The available data and publications related to the chosen system are usually a starting point for the following steps.⁴³ In other cases, new experiments are designed and carried out to provide the necessary data for modeling.¹⁶ For this tutorial, we decided to study a small signaling network in prostate cancer, and do so using published data describing the phosphorylation response of key proteins in prostate cancer cell lines in response to the addition of several ligands and inhibitors.⁶⁰

The study providing the data used in this tutorial⁶⁰ takes a data-driven approach based on multivariate regression analysis to predict prostate cancer cell survival from the phosphorylation levels of 14 key proteins. These proteins are related to core signaling pathways that drive cell growth in three prostate cancer cell lines in response to various

treatments by ligands or kinase inhibitors. The study focuses on the MAPK (Mitogen-activated protein kinases), PI3K and IKK pathways. Correlations between phosphoproteins and cell survival were discussed in relation with their known roles in signaling pathways, with the goal of studying the sources of castration resistance between the cell lines. The results suggest that, in prostate cancer, androgen-independent growth and androgen-mediated signaling are largely driven via MAPK and PI3K signaling.

We aim to study this system with a modeling approach, focusing on the data measured in LNCaP cell line.

Construction of a regulatory network

Constructing a regulatory network is the first step in the modeling process. A first network, commonly referred to as Prior Knowledge Network (PKN), gathers the biological knowledge already known for the main components involved in the process being studied as a signed and directed graph. The PKN is then used for downstream modeling.

When building a PKN, it is critical to know what the essential elements that should be included in the model are. The nodes often describe primarily proteins, but they can be considered as other types of elements such as genes that affect protein functionality. Elements to include in a cancer model comprise proteins related to the most frequently mutated genes, differentially expressed genes, drug targets, and phenotypic read-outs (e.g., cyclins for cell cycle, caspases for apoptosis, EMT regulators for EMT, etc.). Many of these usually surface during the explorative data analysis preceding the mechanistic modeling. The major players of signaling pathways known to be deregulated, such as EGFR, RAS, MEK for the MAPK pathway, or TP53, MDM2, CASP8, or/and CASP3 for the apoptosis pathway, also need to be considered. Finally, input nodes that account for the microenvironment (presence or absence of growth factors, nutrients or ligands, hypoxia conditions, etc.) or for the treatments performed in the experiments and used for the modeling have to be added. The list of nodes can, of course, evolve throughout the construction of the network, through the exploration of published experiments where new genes or proteins are identified as playing an important role, or the inclusion of intermediary nodes required to link two processes.

Once defined, the selected nodes are linked with edges that represent direct interactions or indirect regulations, identified in the literature. When available, the representation of detailed mechanisms is however challenged by the requirement to keep a reasonable size for the network, and as a consequence a reduced computation cost for simulations. A good balance has to be found between detailed mechanistic descriptions of the key signaling components and more schematic representations for less important processes reduced to the necessary players. This process often requires careful consideration of the mechanisms defined as essential for the question at hand. In addition, some simplification can be done once the model is built in an automatic manner. Automatic reduction tools exist to help the process while preserving the dynamics in the logic formalism,^{50,61} such as removing intermediary nodes in linear pathways.

Another complementary approach to sketching the network from literature knowledge is to use database information. There exist some pathway databases that depict, in great details, literature curated signaling pathways. Among these databases, KEGG,⁶² Reactome,^{63,64} BioCarta,⁶⁵ Wikipathways,⁶⁶ Signor,⁶⁷ Signalink,⁶⁸ ACSN,⁶⁹ etc. can be cited. Retrieving

information from them is faster than the tedious process of reviewing the literature. However, the modeler should keep in mind the limits of such resources: because they are not always up-to-date, specific but important information not included in the databases might be missed. To facilitate the retrieval of information from all these databases, we use *Omnipath*,⁵¹ a comprehensive collection of existing manually curated pathway databases. *Omnipath* comes with a Python tool called *pypath*, developed to query its content, manipulate and filter it easily. This allows us to extract the interactions that concern only the predefined list of proteins, and execute complex queries such as retrieving all possible paths between two proteins.

There are different ways in which the network can be represented.⁷⁰ For simplicity, we present our example with a type of diagrams referred to as 'entity relations' in the standard format SBGN,⁷¹ where nodes correspond to genes, proteins, modified proteins, complexes, or phenotypes, and edges are directed arrows representing activation or inhibition of one node over the other. This representation provides a simplified mechanistic view of the processes involved in the disease where the details of the reactions (degradation, synthesis, phosphorylation, etc.) are not explicitly represented but rather pictured as each node having a positive or a negative influence onto other nodes of the network.

In our case study, we build a simple network representing the signaling cascades involved in the survival of prostate cancer cells (**Fig 2**). We aim at validating the network discussed in the initial publication of the data⁶⁰ where two versions are provided. The first one is an undirected map containing the proteins whose phosphorylation status are measured in response to treatment with ligands and inhibitors (Erk1, Erk2, Akt1, Akt2, Akt3, RPS6, GSK3 α , GSK3 β , p38, JNK1, JNK2, JNK3, HSP27, Stat3), the ligands (EGF, IGF1, IL6, TNF α and DHT) that we select as inputs, and key components of the PI3K/mTOR and MAPK pathways (PI3K, mTOR, RPS6, β -Catenin, Jak, RAS, MEK, Rac and IKK). The second network suggests some possible links between the androgen receptor (AR) and the PI3K signaling pathway, RPS6 or cell cycle targets.

Using *Omnipath* signed interactions between the complete list of proteins used in the two initial networks were retrieved in order to connect them into a single network.

For instance, the query related to an interaction between PI3K (PIK3CA) and AKT (AKT1) in *Omnipath* returns an activation of AKT1 by PIK3CA from four databases (SignalLink3, Laudanna_effects, Wang, Signor), and an inhibition of AKT1 by PIK3CA from one database (Laudanna_effects). We can therefore include in the network an activation of AKT by PI3K as the most confident interaction. Looking back at the four literature references associated with the interaction also confirms the sign and direction. A detailed description of the network is reported in Supplementary Materials, with the python code used with *pypath*.

In general, one or more phenotypes of interest will be modeled as additional nodes (e.g. survival, apoptosis, proliferation, etc.) and linked to the rest of the network by appropriate edges, to facilitate model predictions. In our case study, we predict cell survival, available along with phosphorylation data,⁶⁰ as a phenotypic outcome of the model. Therefore, the resulting network was extended to describe roughly the regulation of the cell cycle and the apoptotic pathway, both of which influence an output node called "Survival". Aiming at simplicity for this tutorial, we chose a few prominent components among the high number of possible apoptotic and proliferative factors that could be included in the PKN. Thus, Caspase 8, p53 and Caspase 9 represent the possible modes of activation of the apoptotic pathway,

whereas the proliferative pathways are depicted by MYC and a generic node named Cell_cycle. For more information on this extension, see Supplementary Materials.

The resulting PKN is displayed in **Fig 2**. Complex crosstalks exist between pathways and are usually taken into account in the detailed model. For instance, the apoptotic regulation modeled in another work³⁸ integrates three highly intertwined pathways activated by death receptors. For this tutorial, however, the network is kept simple to focus on core signaling cascades.

Translation into a logic model

As shown on **Fig 1**, the PKN, or interaction graph, can be directly studied with structural analysis, or translated into a logic model to extend the scope of the analysis. Structural analysis include topological analyses such as degree and centrality measures, pathway searches, feedback and feedforward loop identification or characterization of minimal cut and intervention sets.⁷² *Cytoscape*⁵⁴ is a graph visualization and analysis tool that provides access to many of these methods. The translation of the interaction graph into a logic model allows to formally study how signaling information flows through the network and how perturbations of its components affect this flow.³⁵

In a logic model, nodes of the regulatory network are assigned a state: active/true or inactive/false in the case of Boolean models. A set of logic rules are assigned to the edges, that determine how the state of each node will be updated as a function of the state of the nodes that influence them. When a node has two upstream activators, if its activation depends on the simultaneous presence of its two upstream nodes, the two inputs will be linked by an AND gate. If its activation depends on either of the two upstream nodes, the two input nodes will be linked by an OR gate. Inhibition of a node by an upstream node is described with a NOT gate. Combining AND, OR and NOT gates, we can express any logical function as sum of products, also known as disjunctive normal form.

In some cases, when a node has several upstream regulators, it may be unclear if they act in combination or alone. In other words, they may act through any combination of logic gates. In those cases, we can instantiate the different possible rules and use experimental data to find the appropriate representation of the interaction.⁷³ It is thus possible to find an appropriate translation of the regulatory network to a logic model that can be used for simulation purposes.

An example of a complex logic rule (**Fig 2B**) in our case is the one regulating the “Survival” node. The upstream nodes for “Survival” are “Cell cycle”, “MYC”, “Caspase8” and “Caspase9” nodes. To activate “Survival”, the “Cell cycle” node needs to be active, but the presence of either caspase 8 or caspase 9 can inhibit survival. However, if MYC is also active, both caspases need to be active in order to inhibit survival.

This can be written in the following way:

Survival =1 if (Cell_cycle AND ((NOT Caspase8 AND NOT Caspase9) OR (MYC AND (NOT Caspase8 OR NOT Caspase9))))

Altered variants (often referred as ‘mutants’) of the model can then be generated by replacing the logic rule defining the regulation of a component by a constant (0 or 1 in the

Boolean case), to model inhibiting or activating perturbations, such as drug treatments or gene mutations.

The Boolean version of logic modeling might be somewhat restrictive in terms of the scenarios it can represent due to its limitation to binary states. Extensions have been developed to overcome this limitation. In this tutorial, we will show two of these alternatives with two selected tools: *CellNOpt*,⁵² which implements logic-based ODEs, and *MaBoSS*,⁵³ a tool for continuous time Boolean modeling.

The logic-based ODE approach deals with the conversion of logic functions that only accept binary inputs to continuous functions that replicate their behavior. By allowing continuous values for state and time, quantitative time course data can be matched. There are several approaches to build such ODEs,^{34,55} and we use here the one of Wittman and colleagues.³⁴ Because the number of variables and parameters for a given system is much smaller than ODE systems derived from a mechanistic (biochemical) description, logic-based ODE models can include a larger number of biological components. Therefore, logic-based ODEs are particularly useful when dealing with: 1) medium-scale network where only a qualitative biological knowledge of the system under investigation is available, which can be easily interpreted with the logic formalism, and 2) quantitative (and possibly time-course) data resulting from experiments which can be well described with the ODE formalism. The software *CellNOpt*⁵² provides the necessary tools to transform a logic model into a logic-based ODE model and find a set of good parameters by fitting the model to the available data.

MaBoSS is a C++ software that simulates continuous time Markov processes on a Boolean network.⁵³ In contrast to deterministic continuous approaches such as logic-based ODEs, this formalism handles the asynchronous updates of the states of the nodes in a stochastic way and generates a population of trajectories as sequences of Boolean states. Transition rates can be associated with each node, and probabilities of network states can be estimated given a set of initial probabilities (see **Fig 4**). The main result consists in the Boolean state distribution after reaching an asymptotic behavior, and the implicit time provided by sequences steps allows to consider transient dynamics. The size of the population of trajectories can be adapted to the scale of the network, under the control of probability errors, which makes *MaBoSS* useful even for very large networks.

In the next section, we take the regulatory network built previously, with the logic gates already selected based on literature information, and we use some of the available experimental data to find the appropriate weights for an ODE version of the logic model. These weights will also be used as some of the transition rates of the continuous time Markov process model.

Training with data

The purpose of using data to train a network is to obtain a better representation of our system, since information available in the literature and databases may be incomplete, and specify a generic model to a particular biological system. For example, different cell types may share the same signaling networks but with some differences in their wiring, and this may not be reported in the literature. That way, networks can be refined to be cell line or context specific by comparing model simulations with experimental data and refining model

parameters until simulations are able to reproduce the measurements. This process is named model training and corresponds to solving an optimization problem where an objective function that captures how well the model can describe the data is minimized. This optimization can be performed automatically using different algorithms.

The most useful data to train a model are measurements of the states of model variables under different conditions, if possible tracing their dynamic behavior. In networks, nodes typically represent proteins or complexes and a commonly used measurable proxy of their activation or deactivation are post-translational modifications such as phosphorylation. Perturbations of the system by stimulating the pathway with ligands and/or inhibiting the proteins with targeted drugs or antibodies, provide information about the dynamics, which is not available from static data.

Protein phosphorylation upon perturbation is commonly measured using mass-spectrometry or antibody-based techniques. For a detailed description of the advantages and disadvantages of these techniques, we refer to Saez-Rodrigues et al.¹¹ Briefly, with antibody-based techniques, only a maximum of few dozen phosphorylation sites can be measured at the same time. Additionally, antibodies are selected to be specific for a phosphosite, which implies making assumptions on which phosphorylation site represents the activity of each node. With mass-spectrometry techniques, this issue is overcome by the possibility of measuring thousands of phosphorylation sites at the same time. However, the application of mass-spectrometry to the investigation of dynamic signaling pathways has been so far limited by the relatively low number of experimental conditions, due to the expensive and laborious nature of this technology when compared with antibody-based techniques. Accordingly, modeling efforts with this type of data have been scarce,⁷⁴ but this is likely to change in the close future with the rapid development of the technology.⁷⁵

In the data considered for our example⁶⁰ the LNCaP prostate cancer cell line was perturbed with combinations of ligands (EGF, IGF1, IL6, TNF, DHT – in green in **Fig 2**) and kinase inhibitors targeting nodes in the network (PI3K, MEK, IKK, mTOR, p38 – in red in **Fig 2**) for a total of 44 different conditions. The study also included data from perturbation with Docetaxel, which we did not consider as little variation in the phosphoproteome was reported in Docetaxel condition as compared to controls. Docetaxel is known to target β -Tubulin which is related to the cytoskeleton and has no clear effect on our network of interest. However, it is known that cytotoxic treatments can also rewire cell signaling networks.⁷⁶ More measurements in docetaxel condition could allow to study cell signaling rewiring with logic models through network optimization.

Data were measured at three time points (30 min, 4 hours and 24 hours) using an antibody-based technique to measure key phosphosites of 8 proteins in the network (AKT, RPS6, GSK3, ERK, AKT, p38, JNK, HSP27, Stat3 – in blue in **Fig 2**). Since the antibody used for GSK3 measures an inhibitory site, the sign of the regulatory interactions to and from GSK3 were inverted. We choose to consider only the first two time points (30 min and 4 hours) as signaling through phosphorylation changes works on a fast time scale (affecting a wide part of the cell in the order of minutes) and is expected to reach a semi-steady state within the first few hours. Considering a longer time scale might also require taking into account slower effects, such as transcription regulation, that would lead to changes in the levels of proteins

and thereby to a rewiring of the network. We use these data to train cell line and context-specific logic model as detailed in the next paragraphs.

Training the logic-based ODE model with CellNOpt

First, the PKN is interpreted as a logic-based ODE as described in Wittmann et al.³⁴ and implemented in the R package *CellNOptR*.⁵² The dynamic of each node is described by an ODE as function of its regulators. For example, for the rate of change of the variable representing AKT activity the corresponding ODE is:

$$dAKT/dt = \tau_{AKT} \cdot (f(PI3K, k_{PI3K \rightarrow AKT}) - AKT)$$

where parameter τ_{AKT} represents the life-time of the node AKT. When $\tau_{AKT} = 0$, $dAKT/dt$ will also be $=0$, meaning that AKT remains at the basal value, i.e. its initial condition. The transfer function $f(PI3K, k_{PI3K \rightarrow AKT})$ represents the regulation exerted by PI3K, and is a monotonically increasing sigmoid in the variable PI3K (in the range $\{0,1\}$). The increase rate depends on the parameter $k_{PI3K \rightarrow AKT}$ defining the strength of the regulatory interaction. When

$k_{PI3K \rightarrow AKT} = 0$ the dynamic of AKT is independent from PI3K.²⁰

Second, some of the nodes and edges in the logic model are not identifiable (i.e. corresponding parameters cannot be precisely estimated) based on the available experimental data and are automatically compressed, as described in Saez-Rodriguez et al.⁷³ and implemented in *CellNOptR*⁵², in order to reduce the size of the model. These include the so-called non-controllable elements which are the ones that are not affected by any of the perturbed or measured species (e.g. the node “Stress” and its regulation to Rac) and the non-observable elements which are the ones downstream of all the measured or perturbed nodes (e.g. the node “Caspase8” and its regulation by TNFR).

Third, data are stored using the MIDAS format⁷⁷ (see Supplementary Materials) and then normalized between 0 and 1 to be in the same range of model simulations. Here, the normalization is performed separately for each measured species (different antibodies can have different affinities and corresponding data should therefore in general be normalized separately), by computing the \log_2 of the fold change of each perturbed condition with respect to the basal (unperturbed) state, and then linearly scaling the resulting values between 0 and 1 with 0.5 corresponding to the basal state (i.e., the 0 in the \log_2 fold change). In this way, all initial conditions are set to 0.5 (which is the basal state, or the state at time 0) and data at time 30 min and 4 hours can remain at the unperturbed state or show an increased or a decreased activity. More complex normalization procedures, involving saturation effects are implemented in *CellNOptR* and can be used, for example, in case of strong outliers in the data which would otherwise mask smaller effects.

Fourth, the model is trained to the experimental data by looking for the parameter set which minimizes the discrepancy between model simulation and experimental data in terms of sum of squares of the difference between measured and simulated data (namely RSS, residual sum of squares). We assume that the wave of activation of signaling pathways upon stimulation reaches semi-steady state within the first few hours. Hence, an additional term (SS_{penalty}) is included to penalize simulations that do not reach steady-state within the time

range observed in the experimental data. The resulting optimized cost function (Q) can be schematically written as:

$$Q = \sum_{conditions} \sum_{species} (measured - stimulated)^2 + SS_{penalty}$$

This optimization problem (code available as Supplementary Materials) is solved using a global population-based optimization method based on enhanced scatter search, as implemented in the MEIGO software.⁷⁸ As the optimization problem is non-convex, we solved the problem 10 times with different initial random guesses for the parameter values in order to further reduce our chances of ending the solution process in a bad local optimum (the local minima reached were actually very similar; coefficient of variation across the 10 runs = 0.009).

Fifth, we consider as the best model the one that better fits the experimental data in terms of minimum RSS across the 10 runs (**Fig 3A**) and verify that experimental data are well described (**Fig 3B**) (best model is available as Supplementary Materials). If the fit is not good, the initial PKN has to be refined as there might be missing interactions that are supported by the experimental data. This refinement can be performed with a revision of information from literature or databases, or using a combination of data-driven approaches.⁷⁹ After refinement of the PKN, model training must be repeated. In our case, the fit is good, as shown in **Fig 3 B-C**, where measured values are plotted against model simulations showing a good agreement for most of the measured conditions. Importantly, the model is able to represent in the simulations most of the conditions where data show a strong increase or decrease with respect to the basal value (i.e., 0.5).

Additionally, we performed bootstrapping to assess the variability of the optimal model when the optimization is repeated (300 times) using for training the data resampled with replacement. In order to assess the statistical significance of the trained models, we also repeated the optimization in two types of randomized conditions: 1. data randomization (data shuffling across all time points, measured species and conditions), while keeping as scaffold the network derived from prior knowledge (repeated 300 times), 2. network randomization (using *BiRewire*⁸⁰ to preserve network properties), while maintaining for training the measured data (repeated 100 times). As shown in **Fig 3D**, both randomizations showed a significant (p -values, one-sided t -test $< 10^{-26}$) decrease in performance when compared to the bootstrapped distribution or to the best model, regardless of the metric used for comparison (mean squared error, $MSE = RSS/N$ with N size of the training data; coefficient of determination, $COD = 1 - SS_{res}/SS_{tot}$, where SS_{res} is the residual sum of squares and SS_{tot} is the total sum of squares; Pearson Correlation, r). Thus these results prove that experimental data and the prior knowledge network are indeed informative and that our best model performs significantly better than random. The results of bootstrap analysis can also be used to assess the variability of the optimized parameters, highlighting which parameters or parts of the network are not well constrained, and possibly suggesting new targeted experiments which would better constrain the problem. Other approaches to address this problem consist in the use of regularization techniques to induce sparsity of the network, thus improving the identifiability of essential parameters.²⁰

After assessing the quality of our optimized model, we can use it for further analyses. One thing we observe is that IGF1-R, TNFR and IKK are associated with null τ parameters, which

means that they become independent from their regulator activities. The system therefore does not depend on the inputs IGF-1 and TNF α . Moreover, IL6-R is associated with a very small value for T_{IL6-R} , suggesting that the dynamics of the system will predominantly depend on pathways activated by EGF.

Simulation of a logical model in different conditions

We have just shown how it is possible to use experimental data to train and refine our PKN modeled using logic-based ODEs. Some nodes and interactions, however, were omitted because no data was available to determine their underlying parameters. In this section, we will use continuous time Markov processes on a Boolean network to predict the effect of perturbations in a semi-quantitative manner, including those that we could not include in the logic-based ODE model. In particular, we will focus on the probability of survival that can be predicted and compared to experimental observations.

We use the software *MaBoSS*⁵³ to compute time trajectories in this graph: continuous time Markov processes generate a population of sequences (1000 in this case) of asynchronous transitions between states of the system. Such stochastic simulations can exploit continuous values in addition to the Boolean formalism: the proportion of trajectories found in a state represents the probability of reaching this state, and converges toward an asymptotic value. Initial probabilities of inputs (values between 0 and 1) represent environmental conditions and ligand treatments. The relative probability needed to reach the activated output node “Survival” in a perturbed condition compared to the unperturbed condition can be related to the relative proportion of surviving cells after treatment compared to the control experiment.

Moreover, as previously mentioned, *MaBoSS* allows us to associate transition rates with each variable activation or inactivation. Although we keep the inactivation transition rates to the default value of 1, we can assign the values given by the parameters τ obtained in the optimized logic-based ODE model to the transition rates associated with node activations. Interactions associated with a null strength, an extreme situation, can be removed from the model: it is the case for the activation of JNK by Rac and the activation of AR by AKT whose strengths are close to 0 ($k_{Rac \rightarrow JNK}=6.7e-05$, $k_{AKT \rightarrow AR}=8.4e-04$). Finally, treatment with targeted inhibitors can be encoded as alterations of the model variables. In the next section, these alterations will be refined to take into account dose-response.

To validate the model, we first aim at verifying the main experimental observations reported in Lescarbeau and Kaplan⁶⁰ on the global behaviors of LNCaP cells across the different treatments conditions,⁶⁰ indicated on the first row of **Table 1**. For this purpose, the model is simulated until an asymptotic state is reached. Initial states are random for all nodes, including inputs, with the exception of “Stress” which is not related to the treatments and is therefore kept at the value of 0. We introduce several perturbations to the model to reproduce the effect of drug treatments. These simulations are explained below. **Fig 4** details the simulation outputs for the unperturbed model (control condition) and the final survival probabilities in different conditions are summarized in **Table 1**. *MaBoSS* also computes an error associated with each probability, which remains smaller than 0.016 in each condition, showing that the differences between results in different conditions are statistically significant. Detailed simulations with probability errors are given in Supplementary Materials.

Androgen treatment with DHT was experimentally shown to cause a 38% increase in survival as compared to the control condition. This survival advantage was essentially abrogated when treated in combination with the PI3K inhibitor: the slightly decreased survival caused by PI3K inhibitor treatment was only increased by 25% with the addition of DHT. To show that the model is able to reproduce experimental results, we run simulations with an initial probability of 0.5 for the control, or 1 for the DHT input, and with or without an inhibition of PI3K. Initial values for other inputs are kept random. We compare the survival probability when the system has reached asymptotic solutions. We find a survival probability of 0.68 for the control condition (unperturbed trained model, random values for inputs including DHT), and an increase of probability in the condition DHT=1 (probability of survival is 0.87). We also observe a decrease of survival probability with respect to control when the model is perturbed with a complete inhibition of PI3K (survival probability: 0.33). Consistent with experimental observations, the survival probability in this condition is slightly increased with the addition of DHT (survival probability: 0.63), but is not reestablished to the control value.

Interestingly, performing the same simulations on the Boolean model derived from the prior knowledge network shows no survival advantage induced by DHT on the untrained model, suggesting that there is an advantage in training the model to experimental data.

The decreased survival probability for the trained model with inhibited PI3K in the condition of random DHT (the probability of survival is 0.33) compared to the condition DHT=1 (the probability of survival is 0.63) suggests a greater dependency of androgen-independent survival to the PI3K pathway rather than the MAPK pathway in the trained model. This can be verified by simulating the model with inhibited PI3K in the condition DHT=0. In that case the survival probability is very low (probability 0.06). Therefore, androgen-independent survival is dependent on PI3K pathway, while androgen-mediated growth also relies on the MAPK pathway.

The experimentally observed effect of mTOR inhibitor treatment on survival was lower than that of PI3K inhibitor treatment, with a relative increase of cell survival between these two conditions of 31%. Our simulations with the trained model reproduce this trend, with a higher survival probability (0.53) predicted when the model is perturbed with a mTOR knock-out. In contrast, inhibiting PI3K or mTOR has the same effect on survival probability with the untrained model.

Finally, the relative survival measured in experiments in androgen depleted conditions was reported to represent more than 50% of the one in normal growth media condition (control). Here again, simulating the model with DHT=0 confirms this observation with a predicted survival probability of 0.53, 76% of the probability in the unperturbed model.

We have seen in the previous section that although the PKN encompasses the different pathways discussed in the work by Lescarbeau and Kaplan,⁶⁰ training the model on the experimental data points toward a predominant role for the pathways activated by EGFR. This finding is validated by the stochastic simulations of the trained model. Indeed, simulations in different conditions predict survival probabilities that reproduce the main dynamical observations reported on LNCaP cells.⁶⁰

Simulations of drug treatments and application to patient data

Beyond the validation of mechanistic hypotheses, the model can be used to predict the effect of new inhibitions, which can suggest new drug treatments or combinations of drugs.

The simulation of partial inhibitions that can be related to dose-response data can be encoded with the introduction of new nodes inhibiting directly possible drug targets in the model. For instance, the inhibition of PI3K can be enforced with a new input node “Anti_PI3K” which inhibits PI3K. The probability chosen for “Anti_PI3K” represents the strength of inhibition on PI3K. It allows us to assess the effect of low concentrations of inhibitors.

We compare the survival probability predicted by simulations of the model with random input values when inhibiting each node of the model separately, with three levels of inhibition: 0.1, 0.5 and 1 (full inhibition). The results are displayed on **Fig 5**. As expected, the inhibitions of MYC, PI3K or AKT have the most effect on survival for each inhibition level.

We note that the inhibition of p38 is predicted to increase survival, consistent with experimental observations.^{60,81,82} However, the observation that Erk and Stat3 have an important role in LNCaP cells⁶⁰ is not verified, as it has no influence on survival in the model. This suggests that the network should be further extended with a description of the effect of Stat3 on downstream proteins in order to account for this experimental observation.

Beyond single drug predictions, predicting efficacy and specificity of drug treatments and drug combinations within specific tumor contexts and for individual patients is a major challenge, especially for diseases characterized by a high heterogeneity, such as prostate cancer.⁸³ Multiple perturbations are easily introduced in logic models and can address this problem.

As a possible first step in this direction, we suggest to systematically compare the effect between single and double perturbations.⁸⁴ A drug-drug interaction denoting synergy or antagonism is found when the phenotypic quantitative effect of a double perturbation deviates from the effect predicted by the simple linear additive combination of single perturbations, and can suggest candidate targets for drug combinations. Each interaction is thus associated with a score:

$$s_{\phi}(A, B) = \frac{f_{\phi}^{AB} - (f_{\phi}^A + f_{\phi}^B)}{f_{\phi}}$$

where f_{ϕ}^A and f_{ϕ}^B are phenotype ϕ (here survival) fitness values (asymptotic probabilities) of single perturbations, f_{ϕ}^{AB} is the phenotype fitness of the double perturbation, and f_{ϕ} is the phenotype fitness of the unperturbed model.

The synergistic or antagonistic interactions associated with the highest score (absolute value higher than 2.3), are represented as a graph on **Fig 6**, and the table of synergistic scores is available in Supplementary Materials.

Blue interactions denote a negative effect on survival (synergy) and predict efficient drug combinations, while antagonistic interactions (in green) have a positive effect on survival and show possible resistance.

The intensity of the color is proportional to the score value. Non-symmetric interactions are represented by an arrow. In that case, the source node is responsible for the predominant

effect: it is the most efficient single treatment in synergistic interactions, and it confers the resistance in antagonistic interactions.

Computational predictions of drug-drug interactions is especially useful for larger networks, since they can be used to inform high-throughput screenings thereby decreasing the scope and hence cost of experiments.

Here, the graph contains several synergistic interactions between loss-of-function perturbations on the Ras and PI3K pathways, which characterize their parallel influence on the survival probability and predict that combinations of drugs targeting both pathways have a major impact on cell survival.

Moreover, such interactions can also be interpreted as genetic interactions between mutations (knock-out or over-expression), associated with epistasis scores. Genetic interactions provide insights into relationships between different biological functions, and highlight mutation properties such as sensitivity or resistance predicted by the model. Patient genetic profiles can then be exploited to predict resistance mechanisms and identify personalized treatments. For example, **Fig 6** shows that an activating mutation of RAS induces a sensitivity to p53 targeting. In contrast, a gain-of-function mutation of AKT provides a resistance to PI3K-targeted treatment.

Conclusion

In this tutorial, we outline a methodology using logic modeling to better understand and predict the way a system responds to different perturbations. The workflow we present is generic and flexible enough to be adapted to many different other cases. Once we have chosen a system to model, the first step is to build a network based on what is known about the system (PKN). Here, public databases and resources like *Omnipath* help to gather and relate known information. When appropriate experimental data are available, they can be used to refine the PKN. Tools like *CellNOpt* help us in this process. Finally, with a functional model, using tools like *MaBoSS*, we can simulate different cellular and experimental conditions and predict the effect of pharmacological interventions.

In all these processes, the construction of the network is crucial in accurately predicting drug effects.

Although not all the tools used here work under the same programming environment, the use of standard formats, in particular SBML-qual,⁸⁵ facilitate the communication between them. To get the most out of *Omnipath*, the use of its related *Pyopath python* package is recommended. While the advanced features of *CellNOpt* are available in *R* (a wrapper for *python* is also available), *MaBoSS* is a command line software and *Cytoscape* is mainly a standalone application.

To help with finding the correct network structures, experimental data coming from technology such as antibody-based arrays and/or mass spectrometry is of invaluable help. But having access to these data is not always feasible. Both of these approaches require a high amount of cell material, which is not a problem when dealing with cell lines or patient material available in large amount (e.g. blood samples), but becomes a limiting factor when only a small amount of cells are available such as in the case of biopsies or resections from solid tumors. Approaches for functional *ex vivo* screening of patient samples are currently being developed and recent works focus in particular on the investigation of new fast

reporters⁸⁶ and microfluidic-based technologies that exploit small volumes to test a large number of drug combinations.⁸⁷ These methods are so far limited to the investigation of the activity of only one node in the network. On the other hand, single time point omics data measured on patient samples provide constraints on many nodes for model training, with mutations integrated as perturbations and expression data as initial conditions or steady state goals.

While new technology is being developed to predict drug response in patients, cell lines and *in silico* systems are being successfully exploited to show how differential responses to drugs can derive from a different wiring of the signaling network and how systems pharmacology approaches provide useful tools for personalized medicine. In particular, a number of recent studies have successfully studied logic models to investigate signaling pathways and suggest effective drug combinations which were then validated *in vitro* and/or *in vivo*.^{18,20,46,48} Mathematical models calibrated using cell lines have also been proved effective in predicting clinical patient outcomes.⁸⁸

Overall, network-based models allow us to formalize this reasoning into a mechanistic computational model, and infer conclusions about drug's effects from quantitative simulations in a principled manner. For this purpose, logic modeling is a useful approach to capture biological mechanisms in a simplified manner.²³ While this simplicity can render models unable to accurately describe important molecular mechanisms,⁸⁹ it allows us to model larger signaling networks than more detailed approaches such as reaction-based differential equations. Due to this scalability, we expect logic modeling to become an increasingly used approach in systems pharmacology to gain valuable insights, powered by new developments in data acquisition techniques.

Acknowledgements

This work has received funding from the European Union Horizon 2020 research and innovation programme under grant agreement No 668858. F.E. thanks European Molecular Biology Laboratory Interdisciplinary Post-Docs (EMBL EIPOD) and Marie Curie Actions (COFUND) for funding.

Supplementary materials

Supplementary materials can be found on GitHub at https://github.com/saezlab/CPT_QSPtutorial

Supplementary files include: model files, codes for *CellNOptR* and for *pypath*, results of analyses (logic-based ODE parameters, *MaBoSS* simulations and genetic interactions), and a description of a step-by-step model extension using *pypath*.

References

1. Swinney, D. C. & Anthony, J. How were new medicines discovered? *Nat. Rev. Drug Discov.* **10**, 507–519 (2011).
2. Iorns, E., Lord, C. J., Turner, N. & Ashworth, A. Utilizing RNA interference to enhance cancer drug discovery. *Nat. Rev. Drug Discov.* **6**, 556–568 (2007).
3. Shi, J. *et al.* Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat. Biotechnol.* **33**, 661–667 (2015).
4. Neuži, P., Giselbrecht, S., Länge, K., Huang, T. J. & Manz, A. Revisiting lab-on-a-chip technology for drug discovery. *Nat. Rev. Drug Discov.* **11**, 620–632 (2012).
5. Shembekar, N., Chaipan, C., Utharala, R. & Merten, C. A. Droplet-based microfluidics in drug discovery, transcriptomics and high-throughput molecular genetics. *Lab Chip* **16**, 1314–1331 (2016).
6. Hopkins, A. L. & Groom, C. R. The druggable genome. *Nat. Rev. Drug Discov.* **1**, 727–730 (2002).
7. Santos, R. *et al.* A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discov.* **16**, 19–34 (2017).
8. Sorger, P. K. & Allerheiligen, S. Quantitative and Systems Pharmacology in the Post-genomic Era: New Approaches to Discovering Drugs and Understanding Therapeutic. 1–48 (2011).
9. Hasselt, J. G. C. van & Graaf, P. H. van der Towards integrative systems pharmacology models in oncology drug development. *Drug Discov. Today Technol.* **15**, 1–8 (2015).
10. Gadkar, K., Kirouac, D. C., Mager, D. E., Graaf, P. H. van der & Ramanujan, S. A Six-Stage Workflow for Robust Application of Systems Pharmacology. *CPT Pharmacometrics Syst Pharmacol* **5**, 235–249 (2016).
11. Saez-Rodriguez, J., MacNamara, A. & Cook, S. Modeling Signaling Networks to Advance New Cancer Therapies. *Annu. Rev. Biomed. Eng.* **17**, 150814155959008 (2015).

12. Laplante, M. & Sabatini, D. M. mTOR Signaling in Growth Control and Disease. *Cell* **149**, 274–293 (2012).
13. Logan, C. Y. & Nusse, R. The Wnt signaling pathway in development and disease. *Annu. Rev. Cell Dev. Biol.* **20**, 781–810 (2004).
14. Clevers, H. & Nusse, R. Wnt/ β -Catenin Signaling and Disease. *Cell* **149**, 1192–1205 (2012).
15. Prahallad, A. *et al.* Unresponsiveness of colon cancer to BRAF(V600E) inhibition through feedback activation of EGFR. *Nature* **483**, 100–103 (2012).
16. Saez-Rodriguez, J. *et al.* Comparing signaling networks between normal and transformed hepatocytes using discrete logical models. *Cancer Res.* **71**, 5400–5411 (2011).
17. Merkle, R. *et al.* Identification of Cell Type-Specific Differences in Erythropoietin Receptor Signaling in Primary Erythroid and Lung Cancer Cells. *PLoS Comput. Biol.* **12**, e1005049 (2016).
18. Klinger, B. *et al.* Network quantification of EGFR signaling unveils potential for targeted combination therapy. *Mol. Syst. Biol.* **9**, 673 (2013).
19. Silverbush, D. *et al.* Cell-specific computational modeling of the PIM pathway in acute myeloid leukemia. *Cancer Res.* (2016).doi:10.1158/0008-5472.CAN-16-1578
20. Eduati, F. *et al.* Drug resistance mechanisms in colorectal cancer dissected with cell type-specific dynamic logic models. *Cancer Res.* (2017).doi:10.1158/0008-5472.CAN-17-0078
21. Kholodenko, B., Yaffe, M. B. & Kolch, W. Computational Approaches for Analyzing Information Flow in Biological Networks. *Sci. Signal.* **5**, re1–re1 (2012).
22. Le Novère, N. Quantitative and logic modelling of molecular and gene networks. *Nat. Rev. Genet.* 1–13 (2015).
23. Kirouac, D. C. & Onsum, M. D. Using network biology to bridge pharmacokinetics and pharmacodynamics in oncology. *CPT Pharmacometrics Syst Pharmacol* **2**, e71 (2013).
24. Albert, R., DasGupta, B. & Mobasher, N. Some perspectives on network modeling in

- therapeutic target prediction. *Biomed Eng Comput Biol* **5**, 17–24 (2013).
25. Hasselt, J. G. C. van & Graaf, P. H. van der Towards integrative systems pharmacology models in oncology drug development. *Drug Discov. Today Technol.* **15**, 1–8 (2015).
 26. Morris, M. K., Saez-Rodriguez, J., Sorger, P. K. & Lauffenburger, D. A. Logic-based models for the analysis of cell signaling networks. *Biochemistry* **49**, 3216–3224 (2010).
 27. Abou-Jaoudé, W. *et al.* Logical Modeling and Dynamical Analysis of Cellular Networks. *Front. Genet.* **7**, 94 (2016).
 28. Wang, R.-S., Saadatpour, A. & Albert, R. Boolean modeling in systems biology: an overview of methodology and applications. *Phys. Biol.* **9**, 055001 (2012).
 29. Gonzalez, A., Chaouiya, C. & Thieffry, D. Logical modelling of the role of the Hh pathway in the patterning of the Drosophila wing disc. *Bioinformatics* **24**, i234–i240 (2008).
 30. Bonzanni, N. *et al.* Hard-wired heterogeneity in blood stem cells revealed using a dynamic regulatory network model. *Bioinformatics* **29**, i80–8 (2013).
 31. Schlatter, R. *et al.* ON/OFF and beyond--a boolean model of apoptosis. *PLoS Comput. Biol.* **5**, e1000595 (2009).
 32. Le Novère, N. Quantitative and logic modelling of molecular and gene networks. *Nat. Rev. Genet.* **16**, 146–158 (2015).
 33. Chaouiya, C. Petri net modelling of biological networks. *Brief. Bioinform.* **8**, 210–219 (2007).
 34. Wittmann, D. M. *et al.* Transforming Boolean models to continuous models: methodology and application to T-cell receptor signaling. *BMC Syst. Biol.* **3**, 98 (2009).
 35. Samaga, R. & Klamt, S. Modeling approaches for qualitative and semi-quantitative analysis of cellular signaling networks. *Cell Commun. Signal.* **11**, 43 (2013).
 36. Saez-Rodriguez, J. *et al.* A Logical Model Provides Insights into T Cell Receptor Signaling. *PLoS Comput. Biol.* **3**, e163 (2007).
 37. Zhang, R. *et al.* Network model of survival signaling in large granular lymphocyte leukemia. *Proceedings of the National Academy of Sciences* **105**, 16308 (2008).

38. Calzone, L. *et al.* Mathematical modelling of cell-fate decision in response to death receptor engagement. *PLoS Comput. Biol.* **6**, e1000702 (2010).
39. Fauré, A., Naldi, A., Chaouiya, C. & Thieffry, D. Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. *Bioinformatics* **22**, e124–31 (2006).
40. Thakar, J., Piliore, M., Kirimanjeswara, G., Harvill, E. T. & Albert, R. Modeling systems-level regulation of host immune responses. *PLoS Comput. Biol.* **3**, e109 (2007).
41. Poltz, R. & Naumann, M. Dynamics of p53 and NF-κB regulation in response to DNA damage and identification of target proteins suitable for therapeutic intervention. *BMC Syst. Biol.* **6**, 125 (2012).
42. Heyde, S. von der *et al.* Boolean ErbB network reconstructions and perturbation simulations reveal individual drug response in different breast cancer cell lines. *BMC Syst. Biol.* **8**, 75 (2014).
43. Remy, E. *et al.* A Modeling Approach to Explain Mutually Exclusive and Co-Occurring Genetic Alterations in Bladder Tumorigenesis. *Cancer Res.* **75**, 4042–4052 (2015).
44. Ruiz-Cerdá, M. L. *et al.* Towards patient stratification and treatment in the autoimmune disease lupus erythematosus using a systems pharmacology approach. *Eur. J. Pharm. Sci.* **94**, 46–58 (2016).
45. Thakar, J. & Albert, R. Boolean models of within-host immune interactions. *Curr. Opin. Microbiol.* **13**, 377–381 (2010).
46. Silverbush, D. *et al.* Cell-specific computational modeling of the PIM pathway in acute myeloid leukemia. *Cancer Res.* (2016).doi:10.1158/0008-5472.CAN-16-1578
47. Morris, M. K., Clarke, D. C., Osimiri, L. C. & Lauffenburger, D. A. Systematic Analysis of Quantitative Logic Model Ensembles Predicts Drug Combination Effects on Cell Signaling Networks. *CPT Pharmacometrics Syst Pharmacol* **5**, 544–553 (2016).
48. Flobak, Å. *et al.* Discovery of Drug Synergies in Gastric Cancer Cells Predicted by Logical Modeling. *PLoS Comput. Biol.* **11**, e1004426 (2015).
49. Fumiã, H. F. & Martins, M. L. Boolean network model for cancer pathways: predicting

- carcinogenesis and targeted therapy outcomes. *PLoS One* **8**, e69008 (2013).
50. Saez-Rodriguez, J. *et al.* Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction. *Mol. Syst. Biol.* **5**, 331 (2009).
 51. Türei, D., Korcsmáros, T. & Saez-Rodriguez, J. OmniPath: guidelines and gateway for literature-curated signaling pathway resources. *Nat. Methods* **13**, 966–967 (2016).
 52. Terfve, C. *et al.* CellNOptR: a flexible toolkit to train protein signaling networks to data using multiple logic formalisms. *BMC Syst. Biol.* **6**, 133 (2012).
 53. Stoll, G., Viara, E., Barillot, E. & Calzone, L. Continuous time Boolean modeling for biological signaling: application of Gillespie algorithm. *BMC Syst. Biol.* **6**, 116 (2012).
 54. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
 55. Di Cara, A., Garg, A., Micheli, G. de, Xenarios, I. & Mendoza, L. Dynamic simulation of regulatory networks using SQUAD. *BMC Bioinformatics* **8**, 462 (2007).
 56. Müssel, C., Hopfensitz, M. & Kestler, H. A. BoolNet--an R package for generation, reconstruction and analysis of Boolean networks. *Bioinformatics* **26**, 1378–1380 (2010).
 57. Naldi, A. *et al.* Logical modelling of regulatory networks with GINsim 2.3. *Biosystems*. **97**, 134–139 (2009).
 58. Chaouiya, C. *et al.* SBML qualitative models: a model representation format and infrastructure to foster interactions between qualitative modelling formalisms and tools. *BMC Syst. Biol.* **7**, 135 (2013).
 59. Naldi, A. *et al.* Cooperative development of logical modelling standards and tools with CoLoMoTo. *Bioinformatics* **btv013** (2015).
 60. Lescarbeau, R. M. & Kaplan, D. L. Quantitative analysis of castration resistant prostate cancer progression through phosphoproteome signaling. *BMC Cancer* **14**, 325 (2014).
 61. Naldi, A., Remy, E., Thieffry, D. & Chaouiya, C. Dynamically consistent reduction of logical regulatory graphs. *Theor. Comput. Sci.* **412**, 2207–2218 (2011).
 62. Kanehisa, M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.*

- 28, 27–30 (2000).
63. Fabregat, A. *et al.* The Reactome pathway Knowledgebase. *Nucleic Acids Res.* **44**, D481–7 (2016).
 64. Croft, D. *et al.* The Reactome pathway knowledgebase. *Nucleic Acids Res.* **42**, D472–7 (2014).
 65. Nishimura, D. BioCarta. *Biotech Software & Internet Report* **2**, 117–120 (2001).
 66. Kutmon, M. *et al.* WikiPathways: capturing the full diversity of pathway knowledge. *Nucleic Acids Res.* **44**, D488–94 (2016).
 67. Perfetto, L. *et al.* SIGNOR: a database of causal relationships between biological entities. *Nucleic Acids Res.* **44**, D548–D554 (2015).
 68. Fazekas, D. *et al.* Signalink 2 – a signaling pathway resource with multi-layered regulatory networks. *BMC Syst. Biol.* **7**, 7 (2013).
 69. Kuperstein, I. *et al.* Atlas of Cancer Signalling Network: a systems biology resource for integrative analysis of cancer data with Google Maps. *Oncogenesis* **4**, e160 (2015).
 70. Le Novère, N. *et al.* The Systems Biology Graphical Notation. *Nat. Biotechnol.* **27**, 735–741 (2009).
 71. Le Novère, N. *et al.* The Systems Biology Graphical Notation. *Nat. Biotechnol.* **27**, 735–741 (2009).
 72. Klamt, S., Saez-Rodriguez, J., Lindquist, J. A., Simeoni, L. & Gilles, E. D. A methodology for the structural and functional analysis of signaling and regulatory networks. *BMC Bioinformatics* **7**, 56 (2006).
 73. Saez-Rodriguez, J. *et al.* Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction. *Mol. Syst. Biol.* **5**, 331 (2009).
 74. Terfve, C. D. A., Wilkes, E. H., Casado, P., Cutillas, P. R. & Saez-Rodriguez, J. Large-scale models of signal propagation in human cells derived from discovery phosphoproteomic data. *Nat. Commun.* **6**, 8033 (2015).
 75. Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and

- function. *Nature* **537**, 347–355 (2016).
76. Lee, M. J. *et al.* Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. *Cell* **149**, 780–794 (2012).
77. Saez-Rodriguez, J. *et al.* Flexible informatics for linking experimental data to mathematical models via DataRail. *Bioinformatics* **24**, 840–847 (2008).
78. Egea, J. A. *et al.* MEIGO: an open-source software suite based on metaheuristics for global optimization in systems biology and bioinformatics. *BMC Bioinformatics* **15**, 136 (2014).
79. Eduati, F., De Las Rivas, J., Di Camillo, B., Toffolo, G. & Saez-Rodriguez, J. Integrating literature-constrained and data-driven inference of signalling networks. *Bioinformatics* **28**, 2311–2317 (2012).
80. Iorio, F. *et al.* Efficient randomization of biological networks while preserving functional characterization of individual nodes. *BMC Bioinformatics* **17**, 542 (2016).
81. Shimo, T. *et al.* Specific inhibitor of MEK-mediated cross-talk between ERK and p38 MAPK during differentiation of human osteosarcoma cells. *J. Cell Commun. Signal.* **1**, 103–111 (2007).
82. Hotokezaka, H. *et al.* U0126 and PD98059, specific inhibitors of MEK, accelerate differentiation of RAW264.7 cells into osteoclast-like cells. *J. Biol. Chem.* **277**, 47366–47372 (2002).
83. Mitrofanova, A. *et al.* Predicting Drug Response in Human Prostate Cancer from Preclinical Analysis of In Vivo Mouse Models. *Cell Rep.* **12**, 2060–2071 (2015).
84. Calzone, L., Barillot, E. & Zinovyev, A. Predicting genetic interactions from Boolean models of biological networks. *Integr. Biol.* **7**, 921–929 (2015).
85. Chaouiya, C. *et al.* SBML qualitative models: a model representation format and infrastructure to foster interactions between qualitative modelling formalisms and tools. *BMC Syst. Biol.* **7**, 135 (2013).
86. Montero, J. *et al.* Drug-Induced Death Signaling Strategy Rapidly Predicts Cancer Response to Chemotherapy. *Cell* **160**, 977–989 (2015).

87. Eduati, F. *et al.* Rapid identification of optimal drug combinations for personalized cancer therapy using microfluidics. *bioRxiv* 093906 (2016).
88. Fey, D. *et al.* Signaling pathway models as biomarkers: Patient-specific simulations of JNK activity predict the survival of neuroblastoma patients. *Sci. Signal.* **8**, ra130 (2015).
89. Birtwistle, M. R., Mager, D. E. & Gallo, J. M. Mechanistic Vs. Empirical Network Models of Drug Action. *CPT: Pharmacometrics & Systems Pharmacology* **2**, e72 (2013).

Figure Legends

Fig 1. Workflow suggested when applying logic modeling to the study of a biological question. In this tutorial we use *Omnipath*⁵¹ for signaling database mining, *CellNOpt*⁵² for model fitting, *MaBoSS*⁵³ for simulations and *Cytoscape*⁵⁴ for visualization and network analysis. Different steps of the pipeline include (1) selecting a system and a question of interest and building a first version of the network, (2) choosing a modeling formalism and improving the model with data and (3) analyzing the model, making predictions and comparing them to experimental data. The dashed arrow indicates a comparison between the results of the analysis and the experimental data. Dotted arrows represent feedback of the results into the modeling pipeline. Rounded boxes represent elements that can be considered part of the different types of the analysis. PKN: Prior Knowledge Network.

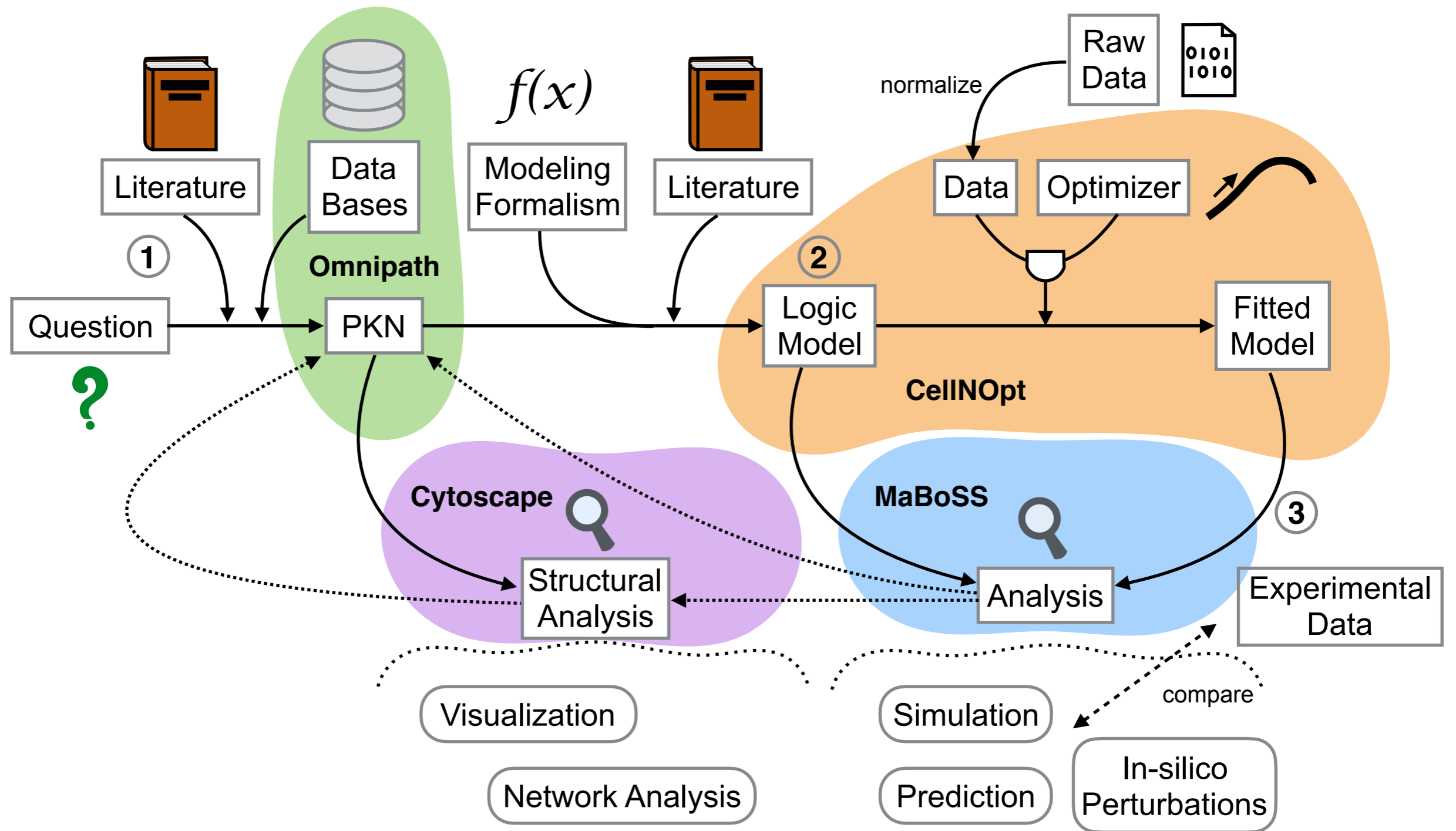
Fig 2. A. Prior knowledge network (PKN) derived from public resources, including interactions connecting nodes which are measured (in blue) or perturbed (stimulated in green and inhibited in red) in the experimental data.⁶⁰ The network was further expanded to include more components from the apoptotic pathway (p53, Caspase8 and Caspase 9) and Myc for the cell cycle activation and their regulation of Survival. Network layout is generated with *Cytoscape*.⁵⁴ B. Examples of logic rules used to convert the network to a logic model. All other nodes in the model with more than one input edge are modeled with a simple OR gate.

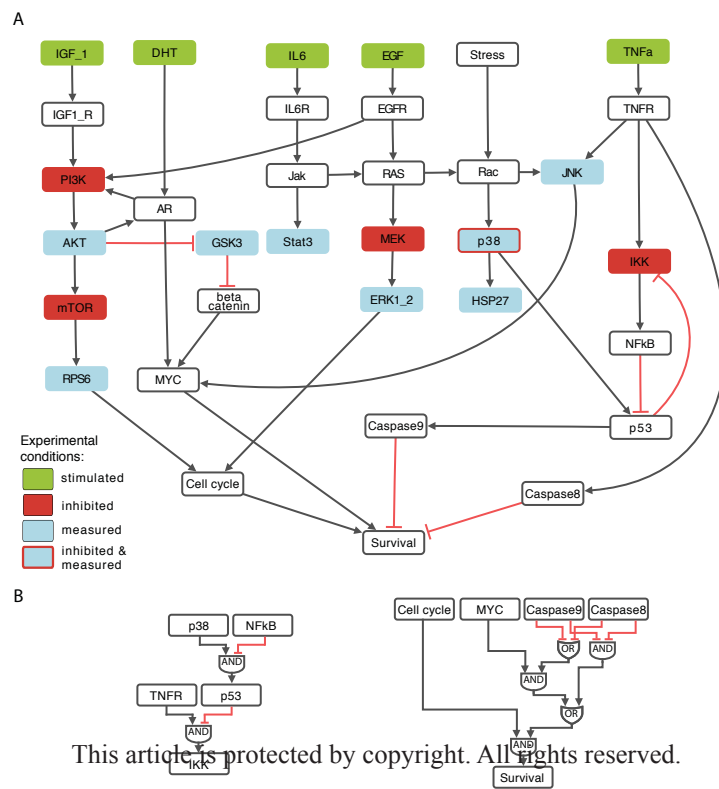
Fig 3. A. Optimized model with node and edge parameter values represented in grayscale. Dotted lines correspond to compressed nodes and edges which are removed before training the model, as not identifiable from the experimental data. B. Top panels show four examples of fit of optimal model simulation to experimental values. For each measured phosphoprotein in each experimental condition, colour scale is used to represent the mean squared error (MSE). C. Scatter plot of simulations using the optimal model with respect to experimental data, showing good correlation. D. Comparison of best model with the results of model optimization after bootstrap (repeated 300 times), network randomization (100 times) and data randomization (300 times) using different scoring metrics, i.e. MSE, coefficient of determination (COD), Pearson correlation (r).

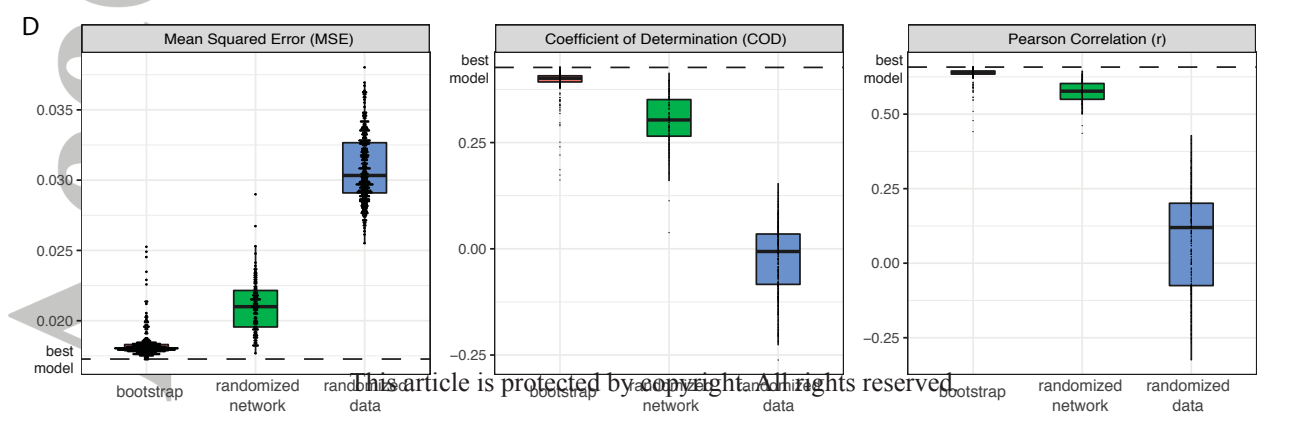
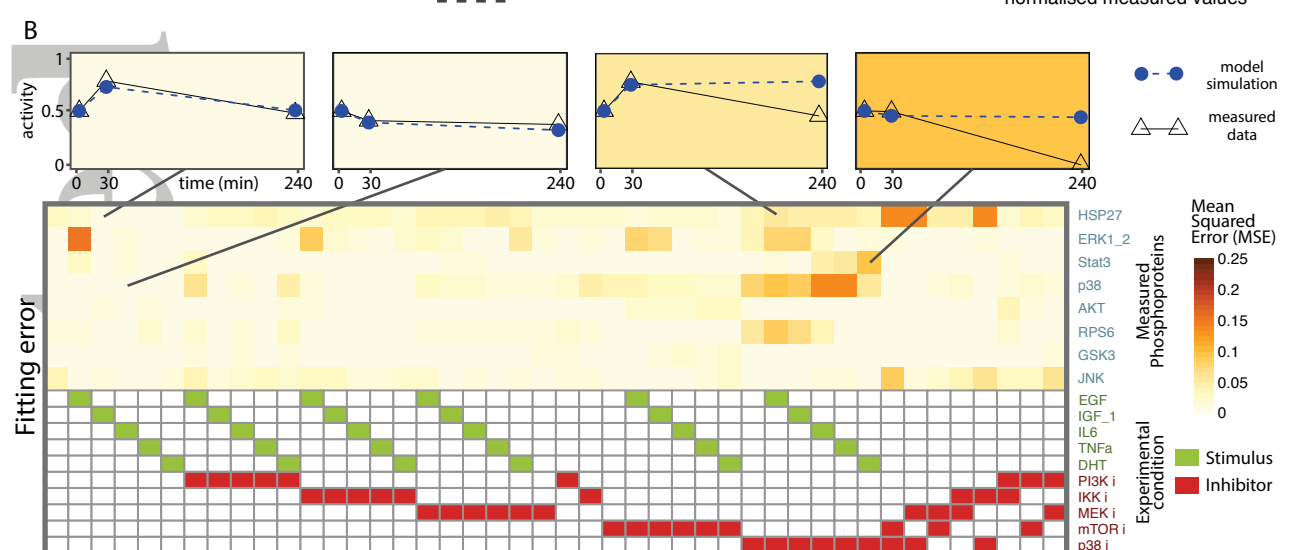
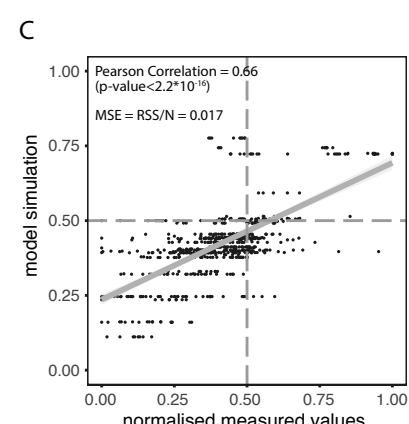
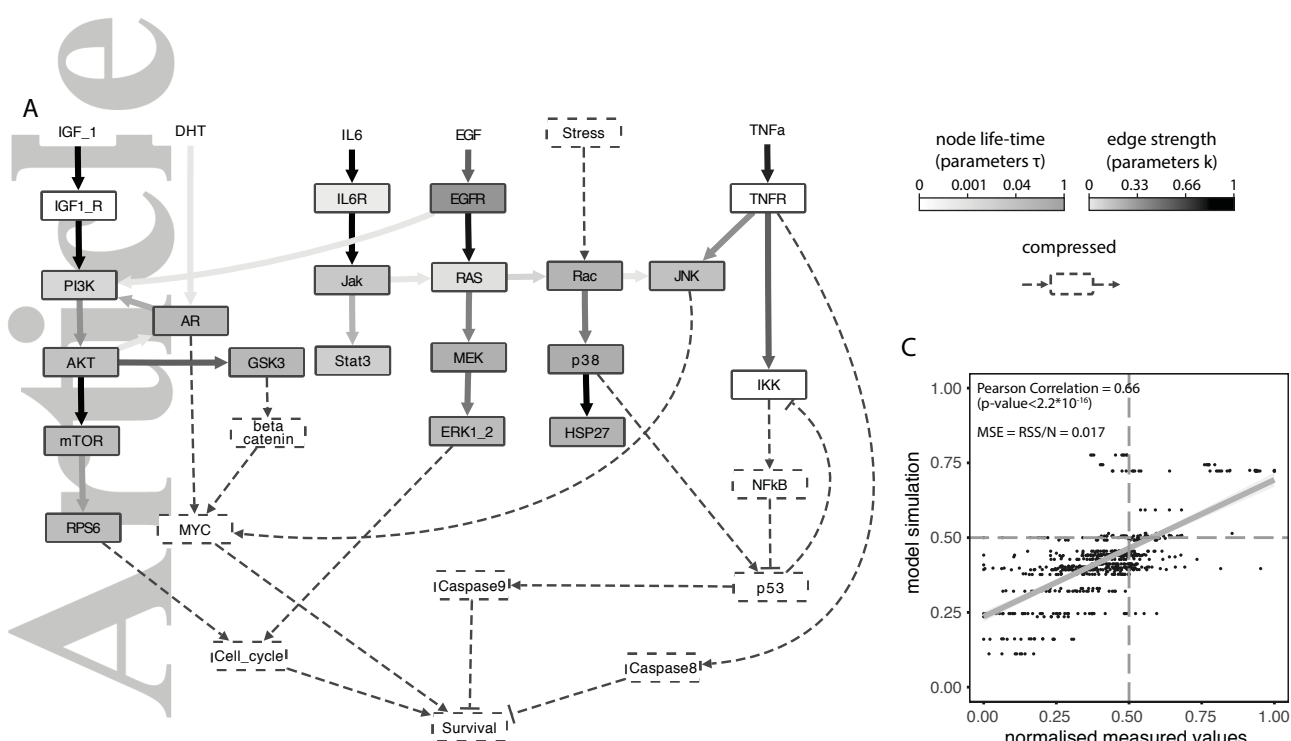
Fig 4. Outputs of *MaBoSS* simulations with random initial states. A. Time trajectories of unperturbed model (WT) or model treated with PI3K inhibitor (iPI3K) and mTOR inhibitor (imTOR), with arbitrary time units. B. Barplot of final state distribution for the unperturbed model. The probability of 7 final model states are shown (*Caspase 8-Myc* state means that the two variables are present, all the others are 0).

Fig 5. Probability of the node “Survival” predicted by the model for different node inhibitions. Survival probability in the control case (unperturbed model) is marked with a grey line.

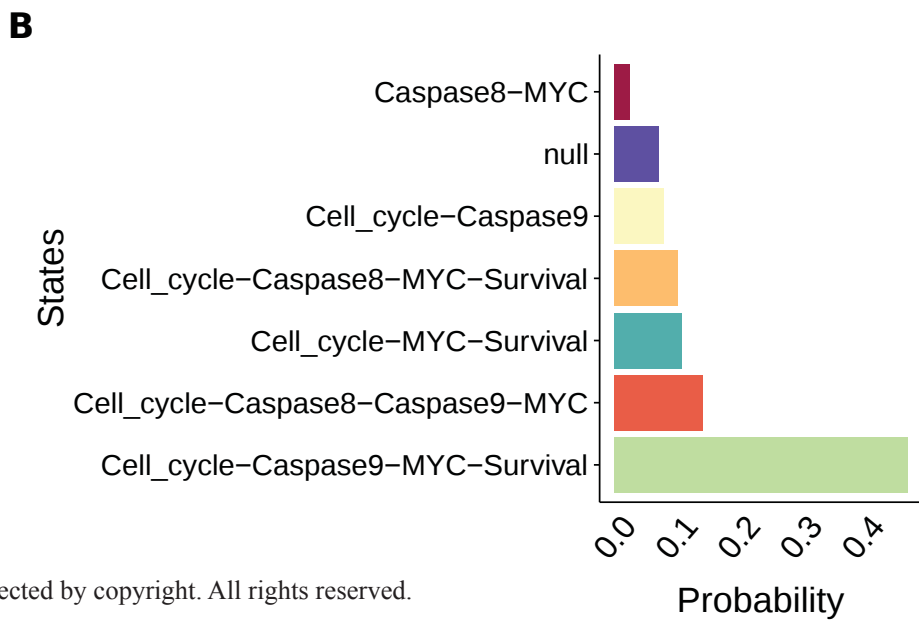
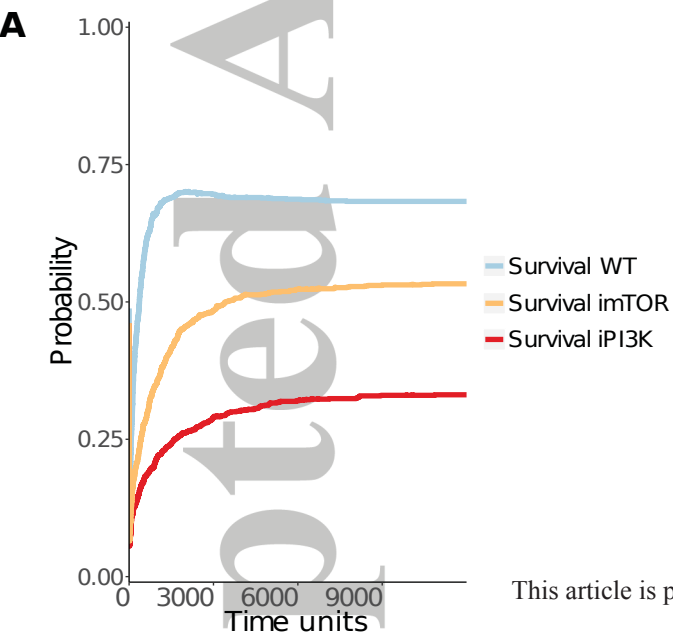
Fig 6. Network of synergistic and antagonistic interactions computed for the trained model, with random initial conditions (except for Stress=0), with Survival as quantitative phenotype. Red triangles represent gain of function alterations and green glyphs represent loss of function alterations. Edges between two alterations show that a combined alteration has a drastic decreasing (in blue) or increasing (in green) effect on the Survival probability when compared to single alterations.



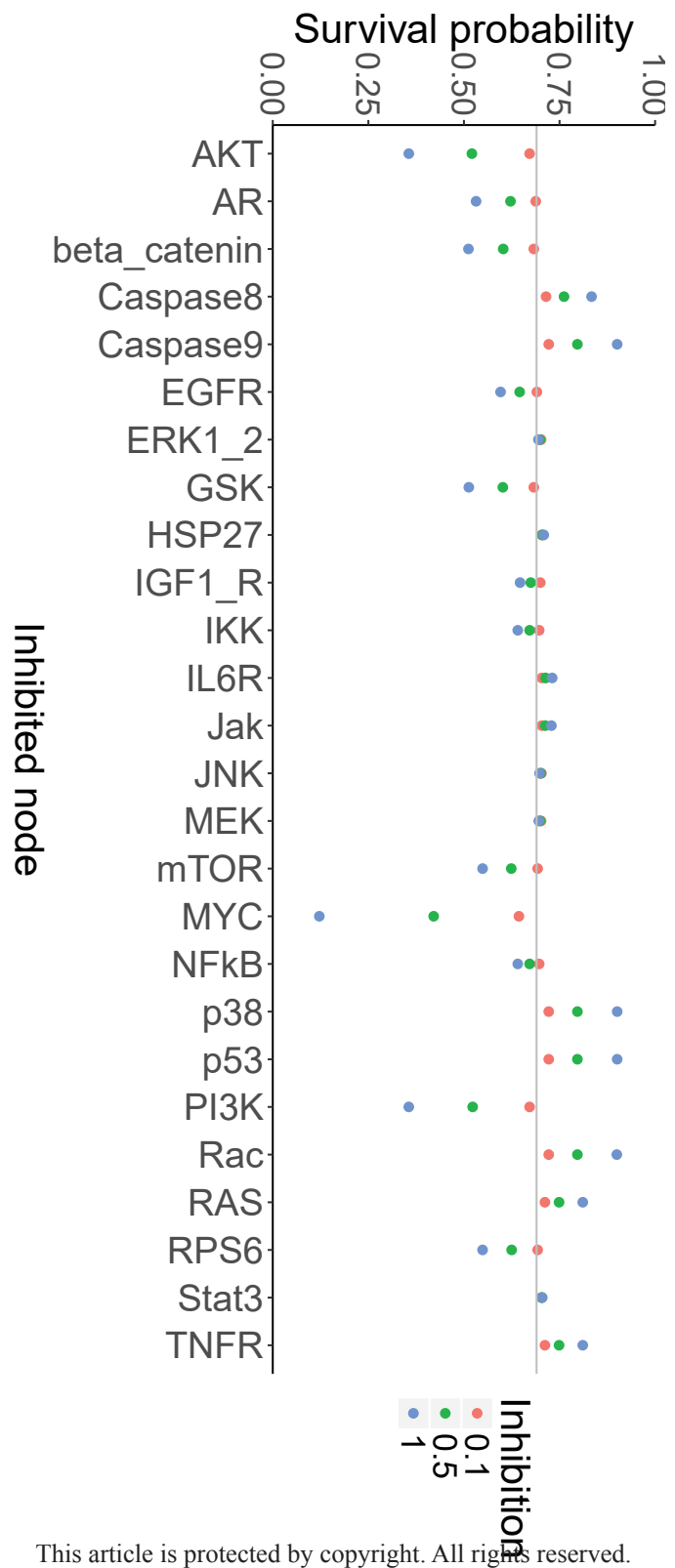




This article is protected by copyright. All rights reserved.



This article is protected by copyright. All rights reserved.



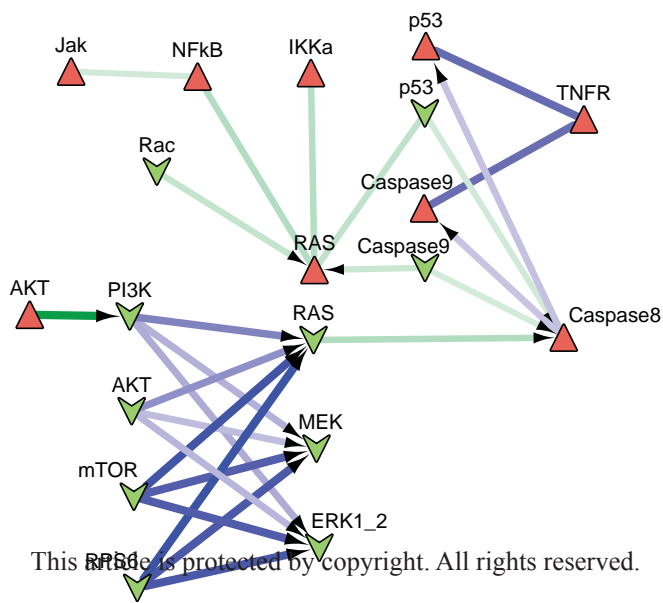


Table 1. Comparison between experimental data and model simulations for both untrained and trained model.

	Control	DHT=1	DHT=1 and PI3Ki	PI3Ki	mTORi	Androgen depleted (DHT=0)	DHT=0 and PI3Ki
Experimental observations	1.00	1.38	0.75	0.60	0.78	0.50	NA
Survival probability, untrained model	0.86	0.86	0.61	0.61	0.59	0.74	0.59
Survival probability, trained model	0.68	0.86	0.63	0.33	0.53	0.53	0.06

First row: proportion of surviving cells measured in the data⁶⁰ (NA: not available). Second row: Survival probabilities for simulations in different conditions, with random inputs, with the untrained model (model derived from PKN). Third row: Survival probabilities for the same simulations with the trained model.

Accepted Article