PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Personalized logical models to investigate cancer response to BRAF treatments in melanomas and colorectal cancers

Jonas Béal[1,2,3☯*], **Lorenzo Pantolini**[1,2,3☯], **Vincent Noël**[1,2,3], **Emmanuel Barillot**[1,2,3], **Laurence Calzone**[1,2,3*]

1 Institut Curie, PSL Research University, Paris, France, **2** INSERM, U900, Paris, France, **3** MINES ParisTech, PSL Research University, CBIO-Centre for Computational Biology, Paris, France

☯ These authors contributed equally to this work.
* jonasbeal@protonmail.com (JB); laurence.calzone@curie.fr (LC)

Check for updates

OPEN ACCESS

Citation: Béal J, Pantolini L, Noël V, Barillot E, Calzone L (2021) Personalized logical models to investigate cancer response to BRAF treatments in melanomas and colorectal cancers. PLoS Comput Biol 17(1): e1007900. https://doi.org/10.1371/journal.pcbi.1007900

Editor: Joerg Stelling, ETH Zurich, SWITZERLAND

Received: April 16, 2020

Accepted: December 21, 2020

Published: January 28, 2021

Copyright: © 2021 Béal et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All data files are available from the Cell Model Passports (https://cellmodelpassports.sanger.ac.uk/downloads) and GDSC (https://www.cancerrxgene.org) portals. Code and data required to reproduce the study and all the plots are available in the following repositories: - Model-checking: https://github.com/sysbio-curie/MaBoSS_test - Data, model personalization and analyses: https://github.com/sysbio-curie/PROFILE_BRAF_Model The logical model can be found in the dedicated GitHub: https://github.com/sysbio-curie/PROFILE_BRAF_Model/tree/master/Models/BRAF_Model and is

Abstract

The study of response to cancer treatments has benefited greatly from the contribution of different omics data but their interpretation is sometimes difficult. Some mathematical models based on prior biological knowledge of signaling pathways facilitate this interpretation but often require fitting of their parameters using perturbation data. We propose a more qualitative mechanistic approach, based on logical formalism and on the sole mapping and interpretation of omics data, and able to recover differences in sensitivity to gene inhibition without model training. This approach is showcased by the study of BRAF inhibition in patients with melanomas and colorectal cancers who experience significant differences in sensitivity despite similar omics profiles.

We first gather information from literature and build a logical model summarizing the regulatory network of the mitogen-activated protein kinase (MAPK) pathway surrounding BRAF, with factors involved in the BRAF inhibition resistance mechanisms. The relevance of this model is verified by automatically assessing that it qualitatively reproduces response or resistance behaviors identified in the literature. Data from over 100 melanoma and colorectal cancer cell lines are then used to validate the model's ability to explain differences in sensitivity. This generic model is transformed into personalized cell line-specific logical models by integrating the omics information of the cell lines as constraints of the model. The use of mutations alone allows personalized models to correlate significantly with experimental sensitivities to BRAF inhibition, both from drug and CRISPR targeting, and even better with the joint use of mutations and RNA, supporting multi-omics mechanistic models. A comparison of these untrained models with learning approaches highlights similarities in interpretation and complementarity depending on the size of the datasets.

This parsimonious pipeline, which can easily be extended to other biological questions, makes it possible to explore the mechanistic causes of the response to treatment, on an individualized basis.

deposited to model repository in GINsim webpage http://ginsim.org/node/248.

Funding: LC, EB and VN acknowledge support from ANR in the context of the project ANR-FNR project AlgoReCell [ANR-16-CE12-0034]. This work was partially supported by the European Union's Horizon 2020 program (grant 826121, iPC project). EB holds a chair at Paris Artificial Intelligence Research Institute Institute (PRAIRIE), funded in part by the French government under management of Agence Nationale de la Recherche as part of the "Investissements d'avenir" program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

We constructed a logical model to study, from a dynamical perspective, the differences between melanomas and colorectal cancers that share the same BRAF mutations but exhibit different sensitivities to anti-BRAF treatments. The model was built from the literature and completed from existing pathway databases. The model encompasses the key proteins of the MAPK pathway and was made specific to each cancer cell line (100 melanoma and colorectal cell lines from public database) using available omics data, including mutations and RNAseq data. It can simulate the effect of drugs and show high correlation with experimental results. Moreover, the structure of the network confirms both the importance of the reactivation of the MAPK pathway through CRAF and the involvement of PI3K/AKT pathway in the mechanisms of resistance to BRAF inhibition.

The study shows that, because of the low number of samples, the mechanistic approach that we propose provides different insights than powerful standard machine learning methodologies would, showing the complementarity between the two approaches. An important aspect to mention is that the mechanistic approach presented here does not rely on training datasets but directly interprets and maps data on the model to simulate drug responses.

Introduction

In the age of high-throughput sequencing technologies, cancer is considered to be a genetic disease for which driver genes are constantly being discovered [1]. The study of mutational and molecular patterns in cancer patients aims to improve the understanding of oncogenesis. However, many of these gene alterations seem to be specific to certain cancer types [2] or exhibit different behaviors depending on the molecular context, particularly in terms of response to treatment. This prompted a shift from univariate biomarker-based approaches to more holistic methodologies leveraging the various omics data available.

To study these observed differences in drug response in various cancers, some approaches based on mathematical modeling were developed to explore the complexity of differential drug sensitivities. A number of machine learning-based methods for predicting sensitivities have been proposed [3], either without particular constraints or with varying degrees of prior knowledge; but they do not provide a mechanistic understanding of the response. Some other approaches focused on the description of the processes that might influence the response by integrating knowledge of the signaling pathways and their mechanisms and translated it into a mathematical model [4–6]. The first step of this approach implies the construction of a network recapitulating knowledge of the interactions between selected biological entities (driver genes but also key genes of signaling pathways), extracted from the literature or from public pathway databases, or directly inferred from data [7]. This static representation of the mechanisms is then translated into a dynamical mathematical model with the goal to not only understand the observed differences [5] but also to predict means to revert unexpected behaviors.

One way to address issues related to patient response to treatments is to fit these mechanistic models to the available data, and to train them on high-throughput cell-line specific perturbation data [4, 5, 8]. These mechanistic models are then easier to interpret with regard to the main drivers of drug response. They also enable the *in silico* simulations of new designs such as combinations of drugs not present in the initial training data [6].

However, these mechanistic models contain many parameters that need to be fitted or extracted from the literature. Some parsimonious mathematical formalisms have been

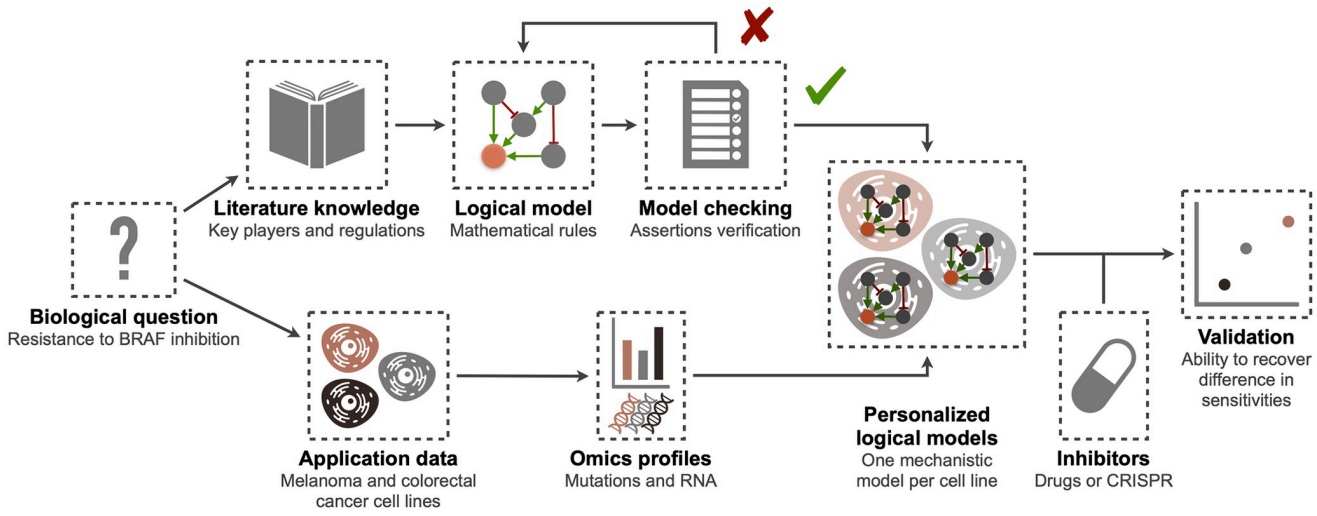

Fig 1. BRAF modeling flowchart: From a biological question to validated personalized logical models.

https://doi.org/10.1371/journal.pcbi.1007900.g001

developed to make up for the absence of either rich perturbation data to train the models or fully quantified kinetic or molecular data to derive the parameters directly from literature. One of these approaches is the logical modeling, which uses discrete variables governed by logical rules. Its explicit syntax facilitates the interpretation of mechanisms and drug response [9, 10] and despite its simplicity, semi-quantitative analyses have already been performed on complex systems [11] for both cancer applications [9, 12] and drug response studies [13, 14], and have proved their efficacy [15, 16].

The nature of this formalism has shown its relevance in cases where the model is not automatically trained on data but simply constructed from literature or pathway databases and where biological experiments focus on a particular cell line [17]. We propose here a pipeline based on logical modeling and able to go from the formulation of a biological question to the validation of a mathematical model on pre-clinical data, in this case a set of cell lines (Fig 1), and the subsequent interpretation of potential resistance mechanisms. The application of the mechanistic model to different cell lines is therefore done without any training of parameters but only on the basis of the automatic integration and interpretation of their omic features.

The construction of a mathematical model must be based first and foremost on a precise and specific biological problem, at the origin of the design of the model. Here, we choose to explore the different responses to treatments in diverse cancers that bear the same mutation. A well-studied example of these variations is the BRAF mutation and especially its V600E substitution. BRAF is mutated in 40 to 70% of melanoma tumours and in 10% of colorectal tumours, each time composed almost entirely of V600E mutations [18]. These tumors have shown poor responses to standard chemotherapy, prompting particular interest in targeted therapies [18]. In particular, many BRAF inhibitors have been developed such as Dabrafenib, Vemurafenib or Encorafenib. However, in spite of the molecular similarities between melanoma and colorectal cancers, BRAF inhibitors have experienced opposite results with improved survival in patients with melanoma [19] and significant resistance in colorectal cancers [20], suggesting drastic mechanistic differences.

Some subsequent studies have proposed context-based molecular explanations, often highlighting surrounding genes or signaling pathways, such as a feedback activation of EGFR

[21] or other mechanisms [22, 23]. All in all, several resistance mechanisms have been studied, pointing to the incomplete suppression of the MAPK pathway following inhibition [24]. These various findings support the need for an integrative mechanistic model able to formalize and explain more systematically the differences in drug sensitivities depending on the molecular context. The purpose of the study we propose here is not to provide a comprehensive molecular description of the response but to verify that the existence and functionality of the suggested feedback loops around the signaling pathway in which BRAF is involved [21] may be a first hint towards these differences. For a more thorough study of these cancers, we refer to other works [4, 25, 26]. The use of BRAF inhibitors in other cancers affected by the BRAV V600E mutation has also been explored in recent years, highlighting the value of mechanistic understanding tools to target and support these applications [27] and shed light on the relevance of the new therapeutic options proposed for colorectal cancers in particular [28].

A logical model summarizing the main molecular interactions at work in colorectal cancers and melanomas is thus built from the literature and completed with databases. As previously mentioned, the objective is to understand whether it is possible to model and explain differences in responses to BRAF inhibition in melanoma and colorectal cancer patients using the same regulatory network. The fact that the two cancers share the same network but differ from the alterations and expression of their genes constitute our prior hypothesis. We then use model checking tools to verify the consistency of this model with a series of qualitative assertions retrieved from literature. Finally, we use available public omics data from these cancer cell lines to transform the generic model into personalized cell-line models. This step is based on the previously published PROFILE method and extends it to the interpretation of treatment responses. The relevance of the personalized models is validated by their ability to recover the differences in BRAF inhibition sensitivities, from both drug and CRISPR screenings. The study therefore presents a global approach, from the biological question to the validation of personalized qualitative models, which links data and knowledge and integrates in particular two methodological innovations: a computer-readable model-checking approach for logical models in the MaBoSS formalism [11], and the interpretation of drug treatments using personalized models.

Materials and methods

Logical model principles and simulations

A concise overview of the main properties of logical modeling is provided and additional details may be found in dedicated reviews [29, 30]. A logical model can be represented by a regulatory graph where nodes are biological entities and edges are influences of one entity onto the others. Each node is considered as a discrete variable (0, 1 or more if required) corresponding to the activity level of the associated biological entity (0 is inactive or absent, 1 is active or present) (Fig 2A). Each entity (proteins, genes, etc.) can thus represent distinct biological states (e.g., expressed gene, phosphorylated protein, etc.) depending on the meaning that is given to each of the nodes. These entities are connected by edges representing positive (resp. negative) influences, i.e., activation (resp. inhibition) of the target node by the sourced node. Combinatorial outcome of influences on one node is defined by the logical rules assigned to the node and expressed with logical operators AND (&), OR (|) and NOT (!), as in Fig 2A. The dynamics of this mathematical model can be expressed using the state transition graph (STG) where the nodes of this graph represent the states of the model. In the STG, the edges show the possible transitions between the different model states according to the logical rules (Fig 2B). As these rules often allow for different transitions, either all of them can be performed at each time step (synchronous update) or performed sequentially by choosing how the priorities are defined (asynchronous updates) [29, 30]. In this study, only the asynchronous

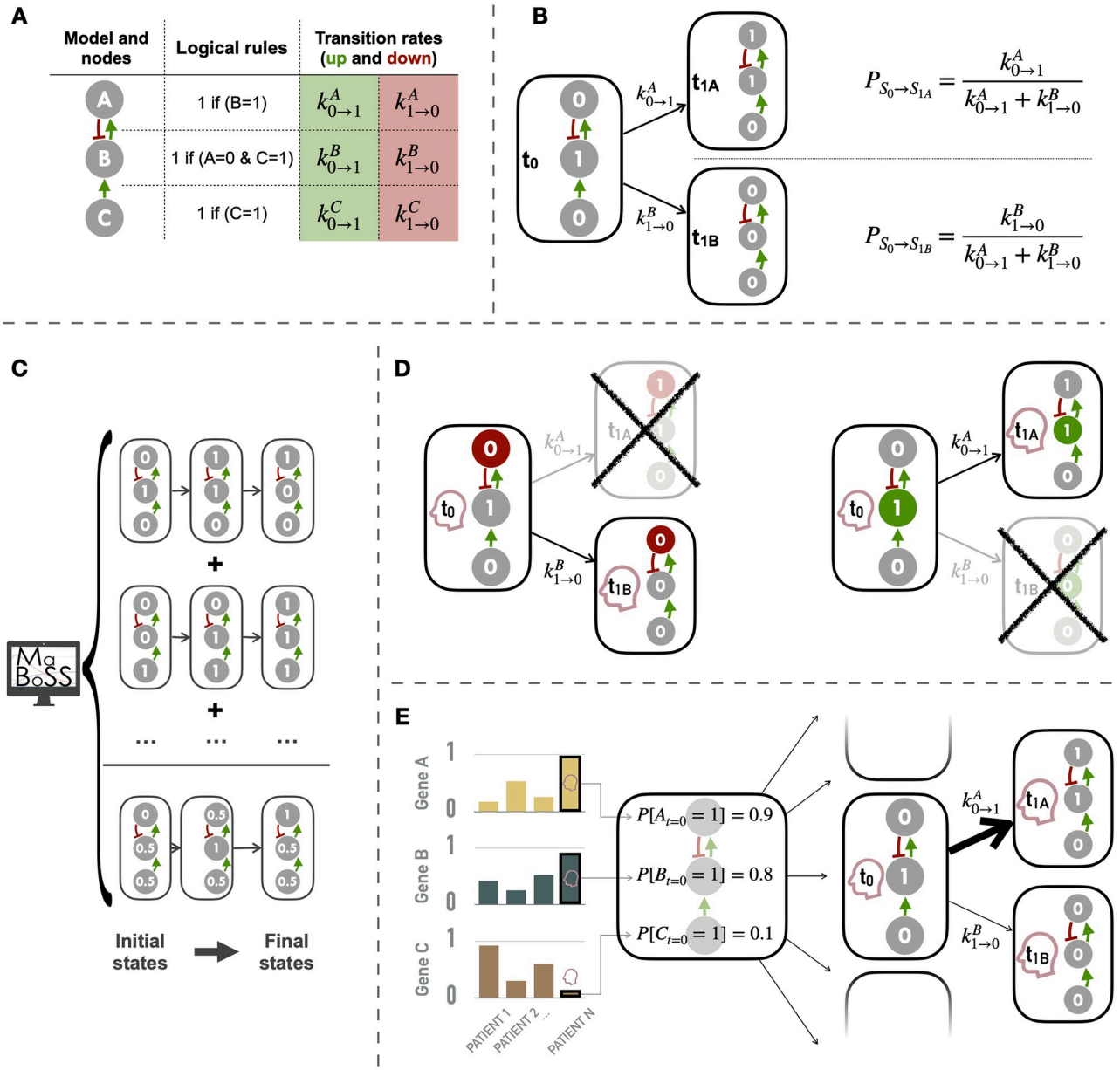

Fig 2. Logical modeling principles and personalization. (A) A logical model with three nodes: the regulatory graph, the corresponding logical rules and the transition rates as used in MaBoSS [31]. (B) Part of the state transition graph with the two possible transitions resulting from the given initial conditions and the probabilities of choosing stochastically one of them. (C) Schematic representation of a logical model simulation with MaBoSS: average trajectory obtained from the mean of many individual stochastic trajectories. (D) Personalization with discrete data (*e.g.*, mutations) with some nodes forced to 0 based on loss of function alteration (left) or 1 based on gain of function/constitutive activation (right). (E) Personalization with continuous data used to define the initial conditions of nodes and to influence the transitions rates and the subsequent probabilities of transition in asynchronous update; the graph on the left represents the normalized values of genes A, B and C for patients 1, 2 and N; the right side represents the personalization of logical model using values from patient N (red profile), first defining the initial probabilities of node activation (middle) and then influencing the probabilities of transitions from one state to another (right): here, since gene A is highly expressed in the red patient, the probability of activation of the corresponding node is increased (resp. probability of inhibition is decreased for gene B).

https://doi.org/10.1371/journal.pcbi.1007900.g002

update is considered. For the rest of the article, the term "node" will refer to those of the regulatory network.

However, it is necessary to differentiate between the activity of the node and that of the underlying biological entity. The active state of a node (Node = 1) can correspond to different

biological reality: if a node represents a gene, its activation can be interpreted biologically as the transcription of this gene; for a node representing a protein, its active and inactive states can be interpreted as the phosphorylated or non-phosphorylated states of the protein. Ideally, the level of node activation should reflect these different mechanisms and be inferred from the most appropriate data types. Here, in order to integrate data directly into logical models, we vary the speed of the reactions according to the data from the patients or cell lines as described below.

In the present article, all simulations are performed according to asynchronous updating with MaBoSS software [11, 31]. This algorithm, using continuous time Markov chain simulations on the Boolean network, provides a stochastic way to choose a specific transition among several possible ones. Each node is associated with transition rates, either for activation of the node $k_{0\to1}$ (or k_{up}) or inactivation $k_{1\to0}$ (or k_{down}) and the stochastic choice between the possible transitions is made based on these transition rates (Fig 2B). For our simulations, unless otherwise specified (cf. section about personalization of models), all transition states were initially assigned to 1. The exploration of all the state space of the model is then done by simulating a very large number of individual stochastic trajectories in order to aggregate them into a mean stochastic trajectory (Fig 2C). Each node is assigned a continuous initial value, between 0 and 1, which represents its probability of being initiated at 1 among all the simulated trajectories. In the absence of a more precise specification, the nodes are initiated at 0.5 and thus start randomly in active or inactive positions. To ensure a proper exploration of the state space, the number of computed stochastic trajectories should increase with the model complexity.

Automatized model-checking within unit-testing framework

The construction of a model is a daunting task, as each improvement is susceptible to change the dynamical properties of the model. To tackle this problem, we need a simple way to test these properties and detect if the model is still able to reproduce them. Software development knows similar challenges, where improvements can break existing functionalities. Software verification thus became an important part of software development, which assess whether a software meets a list of requirements. After each important modification, tests are run to verify that the software still produces the expected behavior. A similar framework can be applied for model construction to check the validity of the model for each iteration of the building process. First, the modeller must describe what is the expected behavior of the model for several conditions, based either on scientific literature or biological experiments. Some similar works using model checking to build and verify models, were recently published [32, 33].

In order to standardize this process, we developed a tool, called MaBoSS_test, to easily verify if a logical model was coherent with specific biological assertions. This tool is based on MaBoSS simulation software, which produces simulations describing the evolution of the probability of states with time. Inspired from python's unittest library, we developed an extension for MaBoSS simulations which tests the validity of the dynamical behavior of the model via assert methods. Each assertion is used to verify if the model satisfies a given type of biological statement.

The majority of the tests consist in altering the model, by changing the initial condition or introducing a set of modifications (e.g. inhibition or overactivation of a node), then observing how these alterations affect the probability of reaching a specific state with respect to the original simulation. An example of a biological assertion may be the reactivation of the MAPK (mitogen-activated protein kinase) pathway through EGFR signal after BRAF inhibition in

colorectal cancer [21]. To test if our model is consistent with this statement, we call the function:

$$.assertStateProbability(\{BRAF : OFF\}, IC_CRC, \{EGFR : 1\}, \text{“}increase\text{”}) \quad (1)$$

The arguments of this method are the following: the set of mutations to perform (BRAF: OFF), the predefined initial conditions of the simulation (IC_CRC), the state we wish to observe (EGFR:1), and a string to specify the behavior (increase). The function can be read as: "Assert that: after BRAF inhibition, using the initial condition for the colorectal cancer, *IC_CRC*, the probability of EGFR activation increases". If the "unit testing output" is selected, in the case that the assertion is not correct, the test will fail raising an exception. Otherwise, if the "detailed output" is selected, the result will be "True" or "False" and the probabilities to have EGFR active before and after BRAF inhibition will be displayed. This tool, its documentation and an example in the form of a Jupyter notebook are available on GitHub (https://github.com/sysbio-curie/MaBoSS_test).

Cell lines omics profiles

The omics profiles of colorectal and melanoma cell lines are downloaded from Cell Model Passports portal [34]. 64 colorectal cancer (CRC) cell lines and 65 cutaneaous melanoma (CM) cell lines are listed in the database, with at least mutation or RNA-seq data (59 CM and 53 CRC with both mutations and RNA-seq data).

Personalization of logical models with cell lines omics data

The PROFILE (PeRsonalization OF logIcaL ModEls) methodology transforms a generic logical model into as many personalized models as there are cell lines by using and integrating their omics profiles [35]. The general idea is to rely on the interpretation of the omics data and translate it into constraints of the mathematical model.

The method to integrate omics data are separated into two strategies: for discrete data (i.e., mutations, copy number alterations) and for continuous data (i.e., transcriptomics, (phospho) proteomics). The discrete strategy consists in setting the value of a node to 0 or 1 for the whole duration of the simulation. In the present work, this is done based on mutation data and functional effect inference. The mutations identified in the cell lines are interpreted using OncoKB database [36], an evidence-based repository with mutation annotations. Mutations referenced as loss-of-function (resp. gain-of function) are forced to 0 (resp. 1), thus constraining the possible transitions in the model as in Fig 2D, left (resp. right). Uninterpreted mutations, which are by far the majority, are not included in the models. The distribution of mutations in the four most frequently mutated genes is shown in Fig 3A.

The second strategy, preferably used with continuous data, is to modify the initial conditions and transition rates based on a continuous proxy for node activity. A cell line with a clear over-expression of a gene/protein, compared to the whole cohort of interest, will have the transition rate related to the activation (resp. inhibition) of the corresponding node favoured and made more (resp. less) likely (Fig 2E). The initial probability that the node will be activated (i.e., the probability to start at 1 among the 5000 stochastic trajectories) will also be modified accordingly, which is particularly important for input nodes that will not be regulated by the model. In the absence of a specification, these nodes remain randomly fixed in active or inactive position at the beginning of the simulation (initial condition 0.5). This method requires different conditions. First, a relevant node activity proxy has to be available in the data: it can be a level from transcriptomics, proteomics or even phospho-proteomics. Then, unlike mutations, the interpretation is not made absolutely but only in comparison to the other members

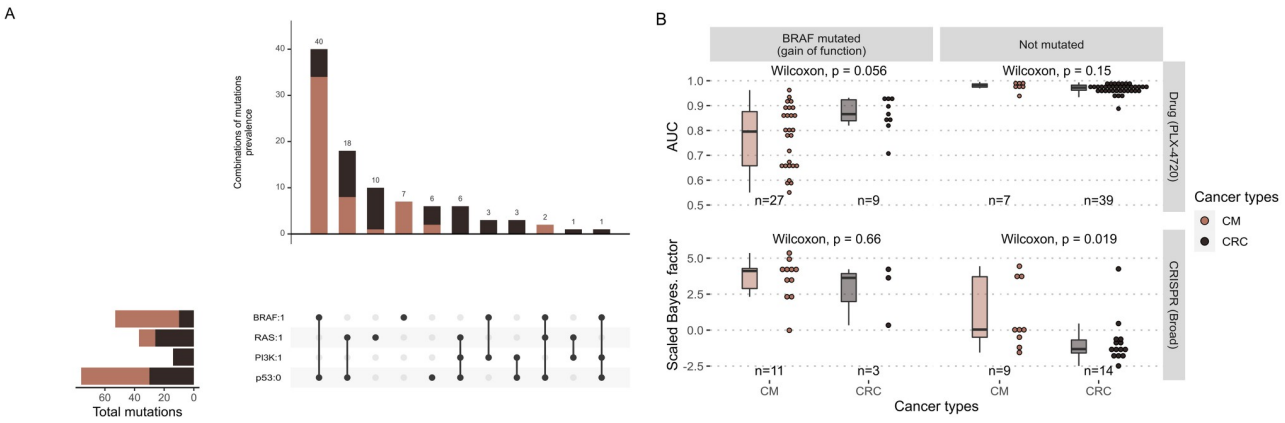

Fig 3. Cell lines data: Mutations and sensitivities to BRAF inhibition. (A) Distribution of the assigned mutations in the four most frequently mutated genes in the colorectal/melanoma cohort of cell lines [37]. (B) Differential sensitivities to BRAF inhibition by the drug PLX-4720 (upper panel) or by CRISPR inhibition (lower panel), depending on BRAF mutational status and cancer type. Numbers of cell lines in each category are indicated. Note that high sensitivities correspond to low AUC and high scaled Bayesian factors.

https://doi.org/10.1371/journal.pcbi.1007900.g003

of the cohort. In the present work, despite the proteic nature of most of the model nodes, RNA data will be used to perform continuous personalization and this for several reasons. First of all, RNA data are more frequently available in different databases, including GDSC, and offer a better coverage of the model nodes than RPPA data, which are often only available for a fraction of the nodes. Interpretation of the same RPPA data, especially on phosphosites, is difficult and may require adaptation of the method at each node. Although it is an imperfect proxy of pathway activity, RNA has already been used in modeling approaches [38, 39]. Note that the distribution of RNA levels is normalized between 0 and 1 on a gene-specific basis before being included in the model.

This method therefore consists of integrating patient data into a logical model with a predetermined structure in order to impose patient-specific constraints on it. The parameters of the model are therefore not optimized on the basis of an objective function and the quantitative treatment response data are not used in the process. Only generic mechanistic knowledge, not specific to individual cell lines, has been used beforehand in the model construction and model-checking steps.

Drug and CRISPR/Cas9 screening

In order to validate the relevance of personalized models to explain differential sensitivities to drugs, some experimental screening datasets are used. Drug screening data are downloaded from the Genomics of Drug Sensitivity in Cancer (GDSC) dataset [40] which includes two BRAF inhibitors: PLX-4720 and Dabrafenib. The cell lines are treated with increasing concentration of drugs and the viability of the cell line relative to untreated control is measured. The dose-response relative viability curve is fitted and then used to compute the half maximal inhibitory concentration (IC50) and the area under the dose-response curve (AUC) [41]. Since the IC50 values are often extrapolated outside the concentration range actually tested, we will focus on the AUC metric for all validation with drug screening data. AUC is a value between 0 and 1: values close to 1 mean that the relative viability has not been decreased, and lower values correspond to increased sensitivity to inhibitions. The results obtained with the two drugs are very strongly correlated (Pearson correlation of 0.91) and the analyses presented here will therefore focus on only one of them, PLX-4720.

Results of CRISPR/Cas9 screening are downloaded from Cell Model Passports [34]. Two different datasets from Sanger Institute [42] and Broad Institute [43] are available. We use scaled Bayesian factors to assess the effect of CRISPR targeting of genes. These scores are computed based on the fold change distribution of sgRNA [44]. The highest values indicate that the targeted gene is essential to the cell fitness. The agreement between the two databases is good [45] but we choose to focus on the Broad database, which is more balanced in terms of the relative proportions of melanomas and colorectal cancers.

Fig 3B illustrates both the relative quantities of cell lines for which drug or CRISPR screening data are available (depending on their BRAF status) as well as differences in sensitivity to BRAF inhibition. Concerning inhibition by the drug PLX-4720, a non-significant difference ($p = 0.056$) in sensitivity between melanomas and colorectal cancers, in favor of the former, is observed for BRAF mutated cell lines (Fig 3B, upper left panel), in line with the clinical description of the differences mentioned in the introduction. However, no difference is apparent for non-mutated BRAF cell lines (Fig 3B, upper right panel). Concerning BRAF inhibition by CRISPR, a difference in sensitivity is on the contrary observed in non-mutated BRAF, where melanomas appear to be more sensitive than colorectal cancers (Fig 3B, lower right panel); this difference is not visible in BRAF-mutated cell lines (Fig 3B, lower left panel). The statistical value of all these comparisons is nevertheless weakened by the small sample sizes, especially for CRISPR inhibition of BRAF-mutated colorectal cancers. For all types of cancers, BRAF-mutated cell lines are more sensitive to BRAF inhibition, but with a certain heterogeneity.

Validation of personalized models using CRISPR/Cas9 and drug screening

The validation of personalized logical models using these screening data is done with the following rationale. First, the models are personalized using omics data from the cell lines. Then, two separate simulations are performed for each personalized model, with and without the inhibition. The first is done by creating a new node representing the BRAF inhibitor and modifying the logical rule associated with the *BRAF* node: when the inhibitor is activated, the *BRAF* logical rule can no longer be satisfied and the node *BRAF* is therefore permanently deactivated. The ratio of the *Proliferation* phenotype obtained with inhibition and without inhibition is the proxy used to be compared with the different screening metrics each of which is also standardized (AUC calculated on relative viability for drugs and Bayes factor computed from fold-changes and then scaled).

Random forests

Random forests are used as an example of a machine learning approach to compare with mechanistic models [46] and are implemented with *randomForestSRC* R package. Random forests can be seen as an aggregation of decision trees, each trained on a different training set formed by uniform sampling with replacement of the original cohort. Prediction performances are computed using out-of-the bag estimates for each individual (*i.e.*, average estimate from trees that did not contain the individual in their bootstrap training sample) and summarized as percentage of variance explained by the random forest. It is also possible to compute the variable importance that assesses the contribution of variables to the overall performance. The solution adopted in this paper to measure it, and called VIMP in the package, consists in introducing random permutations between individuals for the values of a variable and quantifying the variation in performance resulting from this addition of noise. In the case of key variables for prediction, this perturbation will decrease the performance and will result in a high variable importance [47].

Results

A generic logical model for melanoma and colorectal cancers

The construction of the logical model aims at summarizing the current molecular understanding of BRAF gene and its molecular partners in both colorectal cancers and melanomas. The focus of this model is put on two important signaling pathways involved in the mechanisms of resistance to BRAF inhibition which are the ERK1/2 MAPK and PI3K/AKT pathways [48, 49].

The MAPK pathway encompasses three families of protein kinases: RAF, MEK, ERK. If RAF is separated into two isoforms, CRAF and BRAF, the other two families MEK and ERK are represented by a single node. When BRAF is inhibited, ERK can still be activated through CRAF, and BRAF binds to and phosphorylates MEK1 and MEK2 more efficiently than CRAF [50], especially in his V600E/K mutated form. When PI3K/AKT pathway is activated, through the presence of the HGF (Hepatocyte Growth Factors), EGF (Epidermal Growth Factors) and FGF (Fibroblast Growth Factors) ligands, it leads to a proliferative phenotype. The activation of this pathway results in the activation of PDPK1 and mTOR, both able to phosphorylate p70 (RPS6KB1) which then promotes cell proliferation and growth [51]. There has been some evidence of negative regulations of these two pathways carried out by ERK itself [52]: phosphorylated ERK is able to prevent the SOS-GRB2 complex formation through the activation of SPRY [53], inhibit the EGF-dependent GAB1/PI3K association [54] and down-regulate EGFR signal through phosphorylation [52]. The model also accounts for a negative regulation of proliferation through a pathway involving p53 activation in response to DNA damage (represented by ATM); p53 hinders proliferation through the activation of both PTEN, a PI3K inhibitor, and p21 (CDKN1A) responsible for cell cycle arrest.

The generic network presented in Fig 4 recapitulates the known interactions between the biological entities of the network. The network was first built from the literature, and then was verified and completed with potential missing connections using SIGNOR database [55]. More details about the model can be found in the GINsim annotation file of the model [56], available in Supporting Information. A node representing *Proliferation* is also defined from ERK, p70 and p21. It is a coarse-grained representation of the model phenotype that can be compared with experimental data. It was also used to estimate some MaBoSS simulation parameters of the model. For instance, in the present work, all simulations have been performed with 5000 trajectories after verifying that this number was sufficient to ensure very low variability in the final results: 100 different simulations of the generic model, with 5000 trajectories each, result in an average *Proliferation* score of 0.182 with a standard deviation of 0.005 across the 100 simulations. The scores obtained after each simulation correspond to the final asymptotic states, i.e., the average stochastic state reached by the model after a defined period of time. $t_{end} = 50$ was chosen because at this time, it was ensured that the solutions had reached their asymptotic state by comparing with values reached at later times (average *Proliferation* score of 0.182 also at $t_{end} = 100$ with 100 simulations of 5000 trajectories). The behavior of other nodes such as MEK or AKT, located further upstream, was checked in the same way.

We hypothesize that a single network is able to discriminate between melanoma and CRC cells. These differences may come from different sources. One of them is linked to the negative feedback loop from ERK to EGFR. As mentioned previously, this feedback leads to one important difference in response to treatment between melanoma and CRC: $BRAF^{(V600E)}$ inhibition causes a rapid feedback activation of EGFR, which supports continued proliferation. This feedback is observed only in colorectal since melanoma cells express low levels of EGFR and are therefore not subject to this reactivation [21]. Moreover, phosphorylation of SOX10 by ERK inhibits its transcription activity towards multiple target genes by interfering with the sumoylation of SOX10 at K55, which is essential for its transcriptional activity [57]. The absence of

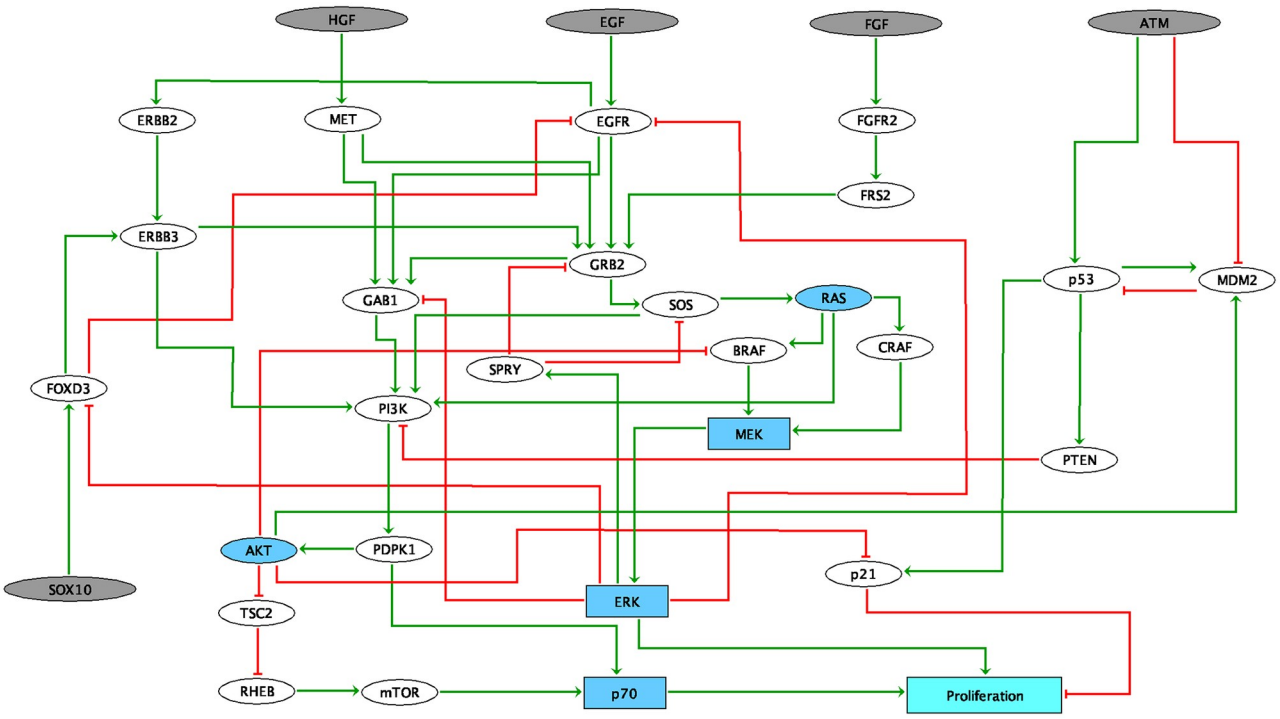

Fig 4. Logical model of signaling pathways around BRAF in colorectal and melanoma cancers. Grey nodes represent input nodes, which may correspond to the environmental conditions. Blue nodes accounts for families. Light blue node represents the output of the model. Square nodes represent multi-valued variable (MEK, ERK, p70 and Proliferation). Note that *Proliferation* is used as the phenotypic read-out of the model.

https://doi.org/10.1371/journal.pcbi.1007900.g004

ERK releases the activity of SOX10, which is necessary and sufficient for FOXD3 induction. FOXD3 is then able to directly activate the expression of ERBB3 at the transcriptional level, enhancing the responsiveness of melanoma cells to NRG1 (the ligand for ERBB3), and thus leading to the reactivation of both MAPK and PI3K/AKT pathways [57]. Furthermore, it has been shown that in colorectal cells, FOXD3 inhibits EGFR signal *in vitro* [58]. Interestingly, SOX10 is highly expressed in melanoma cell lines when compared to other cancer cells. In the model, we define SOX10 as an input because of the lack of information about the regulatory mechanisms controlling its activity. The different expression levels of SOX10 have been reported to play an important role in melanoma (high expression) and colorectal (low expression) cell lines.

The features and expected behaviors for both cancers were formulated as assertions of the model and verified at each step of the model construction (Table 1) through the automatized model-checking framework described in Methods. The tool enables to easily extend the model with new components, while ensuring that the constraints on which the model was built are maintained.

The logical model formalizes the knowledge compiled from different sources. It highlights the role of SOX10, FOXD3, CRAF, PI3K, PTEN and of EGFR in resistance to anti-BRAF treatments. The purpose here is not to suggest new pathways that may be responsible for resistance but to formally confirm what has been suggested and support hand-waving explanations with a mathematical model. The model can be further used to simulate drug experiments and suggest conditions for which the treatment may be efficient or not. Adapting the generic model to each cancer type or cancer cell line will allow to search for the samples that do not respond to

Table 1. List of assertions used to validate the logical model.

Assertions	Refs
1: *BRAF inhibition causes a feedback activation of EGFR in colorectal cancer and not in melanoma.*	[21]
2: *MEK inhibition stops ERK signal but activates the PI3K/Akt pathway and increases the activity of ERBB3.*	[52, 59]
3: *HGF signal leads to the reactivation of the MAPK and PI3K/AKT pathways, and resistance to BRAF inhibition.*	[60]
4: *BRAF inhibition in melanoma activates the SOX10/FOXD3/ERBB3 axis, which mediates resistance through the activation of the PI3K/AKT pathway.*	[57]
5: *Overexpression/mutation of CRAF results in constitutive activation of ERK and MEK also in the presence of a BRAF inhibitor.*	[61] [62]
6: *Early resistance to BRAF inhibition may be observed in case of PTEN loss, or mutations in PI3K or AKT.*	[61]
7: *Experiments in melanoma cell lines support combined treatment with BRAF/MEK + PI3K/AKT inhibitors to overcome resistance.*	[61]
8: *BRAF inhibition (Vemurafenib) leads to the induction of PI3K/AKT pathway and inhibition of EGFR did not block this induction.*	[63]
9: *Induction of PI3K/AKT pathway signaling has been associated with decreased sensitivity to MAPK inhibition.*	[63]

https://doi.org/10.1371/journal.pcbi.1007900.t001

treatment, suggest the possible reasons for this resistance, but still focusing on the model components (S1 Fig).

Differential sensitivities to BRAF targeting explained by personalized logical models

Once the logical model consistency has been validated, personalized models are generated for each cell line by integrating their interpreted genomic features directly as model constraints or parameters. Their sensitivity to BRAF inhibition is then compared to experimentally observed sensitivities (Fig 5). In all the following analyses, we focus on three different personalization strategies using: only mutations as discrete personalization (Fig 5A, upper row), only RNA as continuous personalization (Fig 5A, middle row) or mutations combined with RNA (Fig 5A, lower row). These choices reflect an interpretation of biological reality: mutations are much more drastic and permanent changes than RNA, whose expression levels are more subject to fluctuation and regulation. The objective is also to answer the following questions: What type of data is most likely to explain the differences in responses? Is it relevant to combine them? Fig 5 shows an example of the type of analyses possible with personalized models, zooming in more and more on the details from Fig 5A to 5C.

The first approach consists in using only mutations as discrete personalization (Fig 5A, upper row): the mutations identified in the dataset and that are present in the regulatory network are set to 1 for activating mutations and set to 0 for inactivating mutations. In this case, the *Proliferation* scores from personalized models significantly correlate with both BRAF drug inhibitors (PLX-4720 and Dabrafenib) and both CRISPR datasets (using Pearson correlations). Note that the opposite directions of the correlations for the drug and CRISPR datasets are due to the fact that cell lines sensitive to BRAF inhibition result in low AUCs, and high scaled Bayesian factors, respectively, and, if the models are relevant, to low standardized *Proliferation* scores. Looking more closely at the corresponding scatter plot for PLX-4720 (Fig 5B, upper left panel), it can be seen that this correlation results from the model's ability to recover the highest sensitivity of the BRAF-mutated cell lines that form an undifferentiated cluster on the left side. These cell lines are indeed relatively more sensitive than non-mutated BRAF cell lines. However, the integration of mutations alone does not explain the significant differences within this

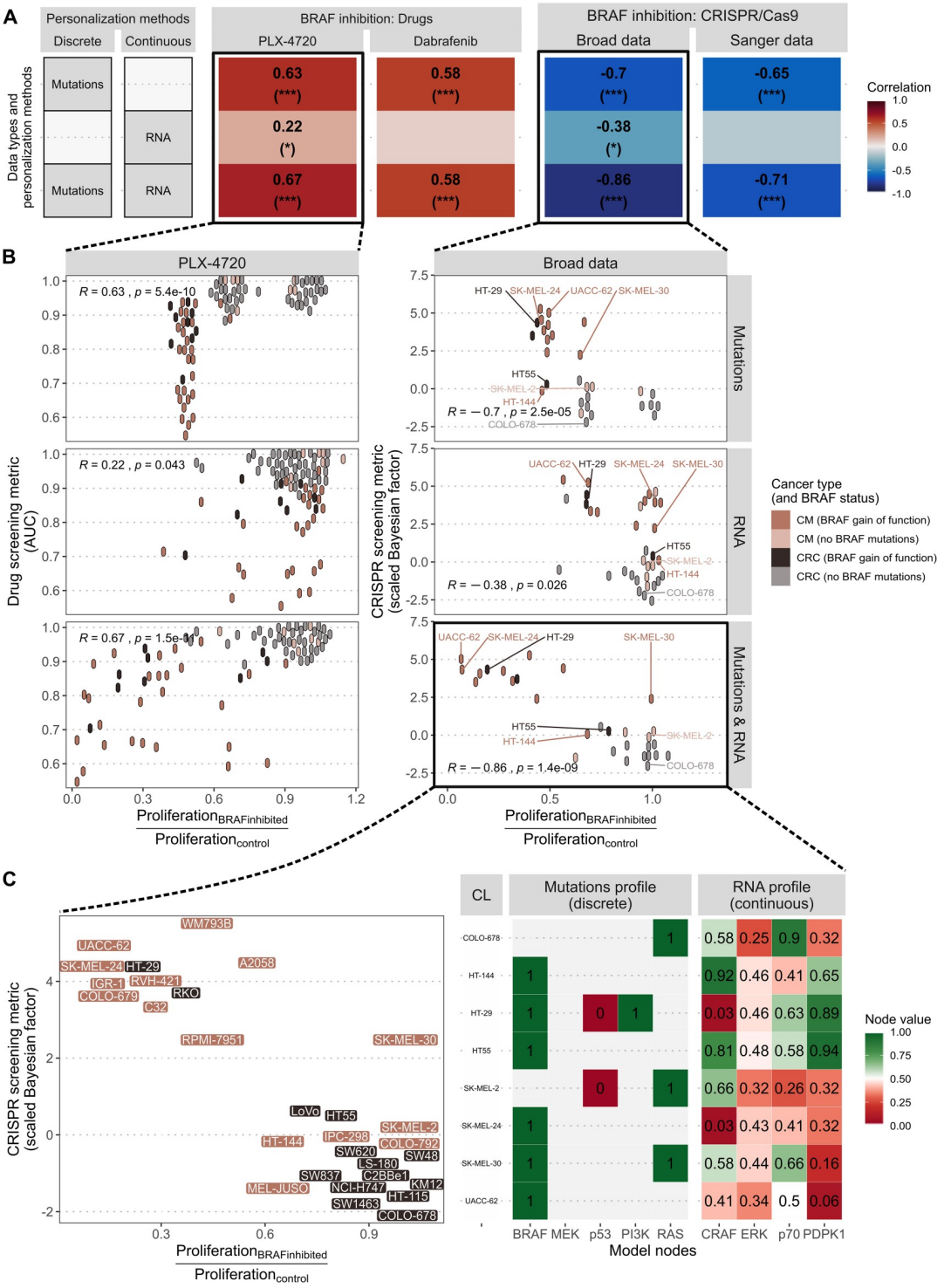

Fig 5. Validation of personalized models with cell lines data. (A) Pearson correlations between normalized *Proliferation* scores from personalized models and experimental sensitivities to BRAF inhibition by drug or CRISPR targeting; each row corresponds to a different personalization strategy; only the values for the significant correlations are displayed. (B) Scatter plots with non-overlapping points corresponding to correlations of panel A, with the three personalization strategies, focusing one one drug (PLX-4720) and one CRISPR dataset (Broad) only. (C) Enlargement of the scatter plot comparing model scores (personalized with mutations and RNA) and experimental sensitivity to CRISPR targeting of BRAF (left) with the corresponding table representing the omics profiles used for each cell line to explore the response mechanisms. This panel can be advantageously replaced by one of the interactive plots proposed in the provided code.

https://doi.org/10.1371/journal.pcbi.1007900.g005

subgroup (AUC between 0.55 and 0.9). A very similar behavior can be observed when comparing model simulations with CRISPR data (Fig 5B, upper right panel).

Using only RNA data as continuous personalization (Fig 5A and 5B, middle rows) is both less informative and more difficult to interpret. For continuous data such as RNAseq data, we normalize the expression values, following the rules described in Methods section and in [35], and set both the initial conditions and the transition rates of the model variables to the corresponding values. Correlations with experimental BRAF inhibitions appear weaker and more uncertain.

The key point, however, is that the combination of mutations and RNA, as depicted in Fig 5A and 5B lower rows, seems to be more relevant. This is partially true in quantitative terms, looking at the correlation in Fig 5 but it is even easier to interpret in the corresponding scatter plots. Comparing first the Broad CRISPR scatter plots using mutations only (Fig 5B, upper right) and using both mutations and RNA (Fig 5B, lower right), we can observe that non-responsive cell lines (scaled Bayesian factor below 0), grouped in the lower right corner and correctly predicted using only mutations stayed in the same area: these strong mutational phenotypes have not been displaced by the addition of RNA data. Other cell lines previously considered to be of intermediate sensitivity by the model (e.g., COLO-678 or SK-MEL-2) were shifted to the right, consistent with the lack of sensitivity observed experimentally. Finally, BRAF-mutated cell lines, previously clustered in one single column on the left using only mutations (with normalized *Proliferation* scores around 0.5), have been moved in different directions. Many of the most sensitive cell lines (scaled Bayesian factor above 2) have been pushed to the left in accordance with the high sensitivities observed experimentally (e.g., HT-29 or SK-MEL-24). It is even observed that the model corrected the position of the two BRAF mutated cell lines, but whose sensitivity is experimentally low (melanoma cell line HT-144 and colorectal cell line HT-55). Only one cell line (SK-MEL-30) has seen its positioning evolve counter-intuitively as a result of the addition of RNA in the personalization strategy: relatively sensitive to the inhibition of BRAF, it has, however, seen its standardised *Proliferation* score approach 1. All in all, this contribution of RNA data results in significant correlations even when restricted to BRAF-mutated cell lines only ($R = 0.69$, p.value = 0.006).

A similar analysis can be made of the impact of adding RNA data to personalization when comparing with the experimental response to PLX-4720 (Fig 5B, upper and lower left panels). Most of the non-sensitive cell lines (upper right corner) have not seen the behavior of the personalized models change with RNA addition. However, the numerous BRAF-mutated cell lines previously grouped around standardized *Proliferation* scores of 0.5, are now better differentiated and their sensitivity predicted by personalized models has generally been revised towards lower scores (i.e., higher sensitivity). Similar to the CRISPR data analysis, three sensitive cell lines have been shifted to the right and are misinterpreted by the model. As a result, the correlation restricted to BRAF-mutated cell lines is no longer significant ($R = 0.26$, p.value = 0.1).

An investigative tool

These personalized models are not primarily intended to be predictive tools but rather used to reason and explore the possible mechanisms and specificities of each cell line. By comparing the profiles of cell lines, it is possible to trace the origin of some of their differences of behaviors in a reverse engineering approach and within the framework of the mechanisms found in the literature and reported in the model. To continue on the previous examples, the two melanoma cell lines, HT-144 and SK-MEL-24, share the same mutational profiles but have very different sensitivities to BRAF targeting (Fig 5C). This inconsistency is partially corrected by the

addition of the RNA data, which allows the model to take into account the difference in CRAF expression between the two cell lines. In fact, CRAF is a crucial node for the network since it is necessary for the reactivation of the MAPK pathway after BRAF inhibition. Therefore, the high sensitivity of SK-MEL-24 may be explained by its low CRAF expression level, which makes the reactivation of the MAPK pathway more difficult for this cell line. Conversely, in HT-144, the high level of CRAF expression allows the signal to flow properly through this pathway even after BRAF inhibition, thus making this cell line more resistant. The importance of CRAF expression is also evident in HT-29, a CRC BRAF mutated cell line with other important mutations (PI3K activation and p53 deletion). However, it remains sensitive to treatment, due to its very low level of CRAF expression. This resistance mechanism related to CRAF is referenced in the literature [64].

Another interesting contribution of RNA appears in the melanoma cell line UACC-62, which is particularly sensitive to treatment. The model is able to correctly predict its response once RNA levels are integrated. In this case, the reason for sensitivity seems to be due to the low level of PDPK1, which makes it difficult to activate p70 and thus trigger the resistance linked to PI3K/AKT pathway activation. Similarly, the CRC resistant cell line, HT55, which carries only the BRAF mutation, expresses high levels of PDPK1, in addition to high levels of CRAF, supporting the idea that the presence of both MAPK and PI3K/AKT pathways may confer resistance to BRAF inhibition treatments. We can also mention a cluster of RAS mutated cell lines, usually NRAS mutated for melanomas (e.g., SK-MEL-2) and KRAS for colorectal cancers (e.g., COLO-678), which are classified by the model as resistant. The importance of PDK1 in resistance mechanisms has already been emphasized in the literature [65]. Interestingly, in these cell lines, a low level of CRAF is not enough to block the signal of the MAPK pathway, which is stronger in the model because of the simulation of the RAS mutation (RAS is set to 1).

Only SK-MEL-30 appears to be incorrectly classified and is observed to be more sensitive than the other cell lines with a similar mutation profile. This could be due to the fact that our network is incomplete and not able to account for some alterations responsible for this cell line sensitivity. The problem may also come from the fact that this cell line contains a frameshift mutation of RPS6KB2 (p70 node) not referenced in OncoKB and therefore not included in the simulation.

The versatility of the logical formalism makes it possible to test other node inhibitions as in Fig 6, but remains limited by the scope of the model. Since the present model has been designed around BRAF, its regulators have been carefully selected and implemented, which is not necessarily the case for other nodes of the model. Therefore, these personalized models can be used to study how comprehensive the descriptions of the regulation of other nodes or parts of the model are. Thus, model simulations show that response trends to TP53 inhibition are consistently recovered by the model (Fig 6B) but the simple regulation of p53 in the model results in coarse-grained patterns, although slightly improved by addition of RNA data. Similar analyses regarding the targeting of PIK3CA (in CRISPR data) simulated, in the model, by the inhibition of PI3K node, can be performed (Fig 6C). Low correlations are an indication highlighting the insufficient regulation of the node. In the same way, it is possible to apply the same pipeline to other cancers, for example by adding thyroid cancer cell lines, sometimes BRAF-mutated, to the analysis. Applicability can also be extended using logical models already published. The use of another cancer logical model, from [17], thus makes it possible to study the response to MEK inhibitors, insufficiently described in the previously studied model, but does not make it possible to recover the responses to BRAF inhibitors, in particular because of a less precise description, with a single *RAF* node representing both BRAF and CRAF genes (see supplemental information in the GitHub of the study).

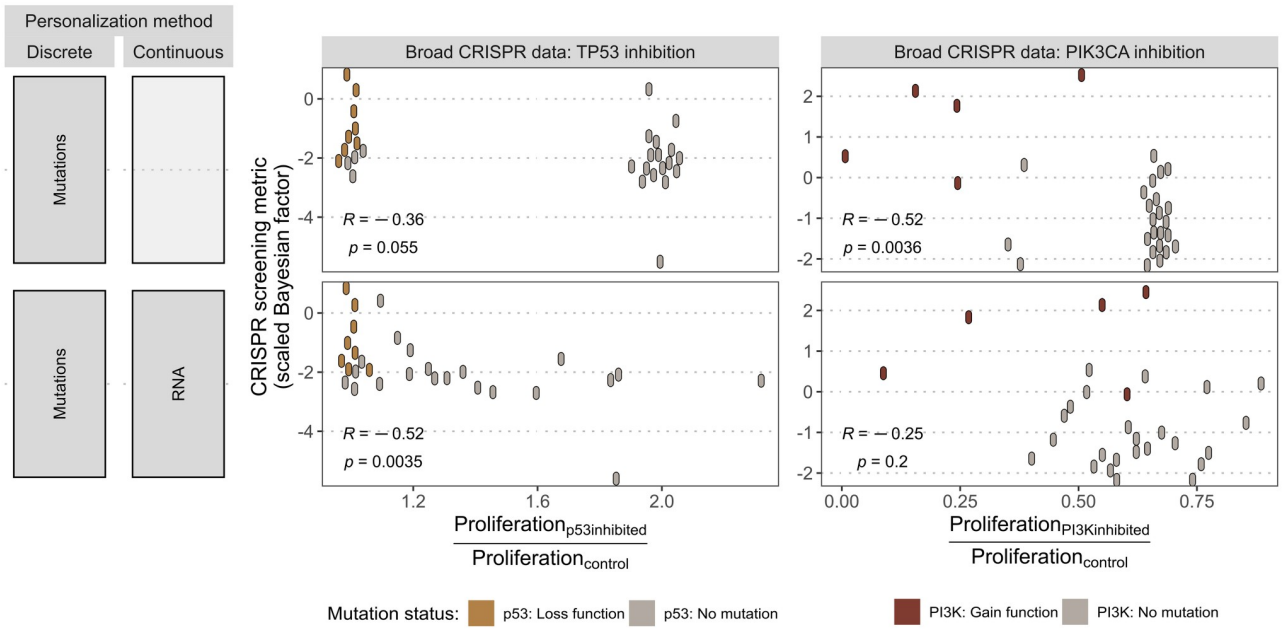

Fig 6. Application of personalized models to other CRISPR targets. (A) Personalization strategies using either mutations only (as discrete data) or combined with RNA (as continuous data) with their corresponding scatter plots in panels B and C. (B) Scatter plot comparing normalized *Proliferation* scores of p53 inhibition in the models with experiment sensitivity of cell lines to TP53 CRISPR inhibition, indicating p53 mutational status as interpreted in the model. Pearson correlations and the corresponding p-values are shown. (C) Similar analysis as in panel B with PI3K model node and PIK3CA CRISPR inhibition.

https://doi.org/10.1371/journal.pcbi.1007900.g006

Comparison of the mechanistic approach with machine learning methods

In order to provide comparison elements unbiased by prior knowledge or by the construction of the model, we performed some simple machine learning algorithms. Random forests have been fitted with inputs (mutations and/or RNA data) and outputs (sensitivities to drug or CRISPR BRAF inhibition) similar to those of logical models and the corresponding predictive accuracies are reported in S2 Fig. The first insight concerns data processing. The percentages of variance explained by the models are similar (around 70% of explained variance for drug sensitivity prediction) in the following three cases: unprocessed original data (thousands of genes), unprocessed original data for model-related genes only (tens of genes), and processed profiles of cell lines (tens of genes). This supports the choice of a model with a small number of relevant genes, which appear to contain most of the information needed for prediction. Second, the absolute level of performance appears much lower for CRISPR (between 30 and 50%) probably suffering from the lower number of samples, especially in cases where the number of variables is the highest. This tends to reinforce the interest of mechanistic approaches that do not use any training on the data for smaller datasets, less suitable for learning. Finally, while mutations and RNA data seem to provide the same predictive power (especially for drugs), using the two together does not necessarily result in a better performance.

Variable importance in these different random forests are reported in S3 Fig and are consistent with the analysis of mechanistic models. The mutational status of BRAF is definitely the most important variable followed by mutations in RAS or TP53. Concerning RNA levels, the most explanatory variables seem to be FOXD3 or PTEN, in line with model definitions.

Discussion

The emergence of high-throughput data has made it easier to generate models for prognostic or diagnostic predictions in the field of cancer. The numerous lists of genes, biomarkers and predictors proposed have, however, often been difficult to interpret because of their sometimes uncertain clinical impact and little overlap between competing approaches [66]. Methods that can be interpreted by design, which integrate *a priori* biological knowledge, therefore appear to be an interesting complement able to reconcile the omics data generated and the knowledge accumulated in the literature.

These benefits come at the cost of having accurate expert description of the problem to provide a relevant basis to the mechanistic models. This is particularly true in this work since the personalized models all derive from the same structure of which they are partially constrained versions. It is therefore necessary to have a generic model that is both sufficiently accurate and broad enough so that the data integration allows the expression of the singularities of each cell line. If this is not the case, the learning of logical rules or the use of ensemble modeling could be favoured, usually including perturbation time-series data [67]. It should also be noted that, in the BRAF model presented here, the translation of biological knowledge into a logical rule is not necessarily deterministic and unambiguous. The choices here have been made based on the interpretation of the literature only. And the presence of certain outliers, i.e., cell lines whose behavior is not explained by the models, may indeed result from the limitations of the model, either in its scope (important genes not integrated), or in its definition (incorrect logical rules). More global or data-driven approaches to define the model would be possible but would require different training/validation steps and different sets of data.

The second key point is the omics data used. For practical reasons, we have focused on mutation and RNA data. The legitimacy of the former is not in doubt, but their interpretation is, on the other hand, a crucial point whose relevance must be systematically verified. The omission or over-interpretation of certain mutations can severely affect the behavior of personalized models. Validation using sensitivity data provides a good indicator in this respect. However, the question is broader for RNA data: are they relevant data to be used to personalize models, i.e., can they be considered good proxies for node activity? The protein nature of many nodes in the model would encourage the use of protein level data instead, or even phosphorylation levels if they were available for these data. One perspective could even be to push personalization to the point of defining different types of data or even different personalization strategies for each node according to the knowledge of the mechanisms at work in the corresponding biological entity. A balance should then be found to allow a certain degree of automation in the code and to avoid overfitting.

Despite these limitations, the results described above support the importance of combining the integration of different types of data to better explain differences in drug sensitivities. There was no doubt about this position of principle in general [68], and in particular in machine learning methods [3, 69]. The technical implementation of these multi-omic integrations is nevertheless more difficult in mechanistic models where the relationships between the different types of data need to be more explicitly formulated [8]. The present work therefore reinforces the possibility and value of integrating different types of data in a mechanistic framework to improve relevance and interpretation and illustrates this by highlighting the value of RNA data in addition to mutation data in predicting the response of cell lines to BRAF inhibition. It is also interesting to note that the most powerful algorithm for predicting drug sensitivities, the Bayesian multitask MKL, combines this consideration of the multi-omics nature of the problem, through multi-view learning, with multi-task learning. This feature allows sharing information by learning about different drugs simultaneously. This can be

a solution to the small sample sizes mentioned above. In the same way, the study of different drugs simultaneously through a mechanistic model is also possible and can be fruitful: it is then especially the biological structure of the model that represents this shared information that facilitates generalization. In addition, one piece of data that could be further exploited is that of the specific behavior of the drugs or inhibitors studied, since for instance some BRAF inhibitors have affinities that vary according to mutations in the BRAF gene itself. PLX-4720 was shown to have a paradoxical activation effect on wild-type BRAF despite its inhibitory effect on mutated BRAF [70]. This mechanism is more complex than the one modeled here and also involves the status of the RAS gene. Thus, the simplistic representation of PLX-4720 as an unconditional inhibitor of BRAF remains a valid approximation in a qualitative model designed to highlight differences in sensitivities, but the quantitative study of the magnitude of these differences would require a more sophisticated model. The integration of truly precise data on the nature of the drug is nevertheless limited by logical formalism and is more often found in less constrained approaches [71].

To conclude, we provide a comprehensive pipeline from clinical question to a validated mechanistic model which uses different types of omics data and adapts to dozens of different cell lines. This work, which is based only on the interpretation of data and not on the training of the model, continues some previous work that has already demonstrated the value of mechanistic approaches to answer questions about response to treatment, especially using dynamic data [72], and sometimes about the same pathways [8]. In this context, our approach proves the interest of logical formalism to make use of scarce and static data facilitating application to a wide range of issues and datasets in a way that is sometimes complementary to learning-based approaches.

Supporting information

S1 Fig. Mapping of the data in the RNAseq data on the regulatory network. The expression data from both melanoma and colorectal cell lines used in this study are mapped onto the network. The scores correspond to the difference in the mean expression of the normalized data (using PROFILE method [35]). Red nodes show higher gene expression in melanomas and blue nodes higher expression in colorectal cancer cell lines. If most active nodes are equivalent to phosphorylated data, the mapping of RNAseq data informs on the gene status and the possibilities to activate the nodes. Thus, conclusions should be made with this fact in mind. At the gene level, then, genes such as SOX10, FOXD3, AKT, p21 and SPRY tend to have a higher expression in melanomas confirming their role in response to the treatment, whereas genes such as EGFR, ERBB2, MET, PTEN and FGFR2 are more relavant to colorectal cancers. This figure reinforces the idea that the mechanisms related to the response to anti-BRAF treatment may have different outcomes in both cancers bascule of a different gene context.
(TIF)

S2 Fig. Performances of random forests for BRAF sensitivity prediction. Random forests algorithms are trained with different omics types (mutations, RNA or both) and data processing (original data or processed data) to predict sensitivity to BRAF inhibition, through drug or CRISPR screening. Performances are expressed as percentage of explained variance by the fitted random forests.
(TIF)

S3 Fig. Variable importance for BRAF sensitivity prediction by random forests. Variable importance for inhibition of BRAF by drugs (first row in S2 Fig), when random forests algorithms are trained with different omics types (mutations, RNA or both) and data processing

(original data or processed data). Higher values of variable importance correspond to higher decrease in prediction performance when the variable is disturbed by permutation and therefore to variables with a positive contribution to predictive performance.
(TIF)

Acknowledgments

We would like to thank Celine Hernandez for fruitful discussions and inspiration during the elaboration of the verification tool, MaBoSS_test.

Author Contributions

Conceptualization: Jonas Béal, Emmanuel Barillot, Laurence Calzone.

Data curation: Jonas Béal, Lorenzo Pantolini.

Formal analysis: Jonas Béal, Lorenzo Pantolini, Laurence Calzone.

Funding acquisition: Emmanuel Barillot, Laurence Calzone.

Investigation: Jonas Béal, Lorenzo Pantolini.

Methodology: Jonas Béal, Lorenzo Pantolini, Vincent Noël.

Software: Jonas Béal, Lorenzo Pantolini, Vincent Noël.

Supervision: Emmanuel Barillot, Laurence Calzone.

Validation: Jonas Béal, Lorenzo Pantolini, Vincent Noël, Laurence Calzone.

Visualization: Jonas Béal, Lorenzo Pantolini.

Writing – original draft: Jonas Béal, Lorenzo Pantolini, Laurence Calzone.

Writing – review & editing: Jonas Béal, Lorenzo Pantolini, Vincent Noël, Emmanuel Barillot, Laurence Calzone.

References

1. Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. Nature medicine. 2004; 10(8):789–799. https://doi.org/10.1038/nm1087 PMID: 15286780
2. Hoadley KA, Yau C, Hinoue T, Wolf DM, Lazar AJ, Drill E, et al. Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. Cell. 2018; 173(2):291–304. https://doi.org/10.1016/j.cell.2018.03.022 PMID: 29625048
3. Costello JC, Heiser LM, Georgii E, Gönen M, Menden MP, Wang NJ, et al. A community effort to assess and improve drug sensitivity prediction algorithms. Nature biotechnology. 2014; 32(12):1202. https://doi.org/10.1038/nbt.2877 PMID: 24880487
4. Eduati F, Doldàn-Martelli V, Klinger B, Cokelaer T, Sieber A, Kogera F, et al. Drug Resistance mechanisms in colorectal cancer dissected with cell type–specific dynamic logic models. Cancer research. 2017; 77(12):3364–3375. https://doi.org/10.1158/0008-5472.CAN-17-0078 PMID: 28381545
5. Jastrzebski K, Thijssen B, Kluin RJ, de Lint K, Majewski IJ, Beijersbergen RL, et al. Integrative modeling identifies key determinants of inhibitor sensitivity in breast cancer cell lines. Cancer research. 2018; 78(15):4396–4410. https://doi.org/10.1158/0008-5472.CAN-17-2698 PMID: 29844118
6. Fröhlich F, Kessler T, Weindl D, Shadrin A, Schmiester L, Hache H, et al. Efficient parameter estimation enables the prediction of drug response using a mechanistic pan-cancer pathway model. Cell Systems. 2018; 7(6):567–579. https://doi.org/10.1016/j.cels.2018.10.013 PMID: 30503647
7. Verny L, Sella N, Affeldt S, Singh PP, Isambert H. Learning causal networks with latent variables from multivariate information in genomic data. PLOS Computational Biology. 2017; 13(10):1–25. https://doi.org/10.1371/journal.pcbi.1005662 PMID: 28968390

8. Klinger B, Sieber A, Fritsche-Guenther R, Witzel F, Berry L, Schumacher D, et al. Network quantification of EGFR signaling unveils potential for targeted combination therapy. Molecular systems biology. 2013; 9(1). https://doi.org/10.1038/msb.2013.29 PMID: 23752269
9. Zañudo JGT, Scaltriti M, Albert R. A network modeling approach to elucidate drug resistance mechanisms and predict combinatorial drug treatments in breast cancer. Cancer convergence. 2017; 1(1):5. https://doi.org/10.1186/s41236-017-0007-6
10. Iorio F, Knijnenburg TA, Vis DJ, Bignell GR, Menden MP, Schubert M, et al. A landscape of pharmacogenomic interactions in cancer. Cell. 2016; 166(3):740–754. https://doi.org/10.1016/j.cell.2016.06.017 PMID: 27397505
11. Stoll G, Caron B, Viara E, Dugourd A, Zinovyev A, Naldi A, et al. MaBoSS 2.0: an environment for stochastic Boolean modeling. Bioinformatics. 2017; 33(14):2226–2228. https://doi.org/10.1093/bioinformatics/btx123 PMID: 28881959
12. Remy E, Rebouissou S, Chaouiya C, Zinovyev A, Radvanyi F, Calzone L. A modeling approach to explain mutually exclusive and co-occurring genetic alterations in bladder tumorigenesis. Cancer research. 2015; 75(19):4042–4052. https://doi.org/10.1158/0008-5472.CAN-15-0602 PMID: 26238783
13. Knijnenburg TA, Klau GW, Iorio F, Garnett MJ, McDermott U, Shmulevich I, et al. Logic models to predict continuous outputs based on binary inputs with an application to personalized cancer therapy. Scientific reports. 2016; 6(1):1–14. https://doi.org/10.1038/srep36812 PMID: 27876821
14. Eduati F, Jaaks P, Wappler J, Cramer T, Merten CA, Garnett MJ, et al. Patient-specific logic models of signaling pathways from screenings on cancer biopsies to prioritize personalized combination therapies. Molecular systems biology. 2020; 16(2). https://doi.org/10.15252/msb.20188664
15. Cohen DPA, Martignetti L, Robine S, Barillot E, Zinovyev A, Calzone L. Mathematical Modelling of Molecular Pathways Enabling Tumour Cell Invasion and Migration. PLoS Computational Biology. 2015; 11(11):e1004571. https://doi.org/10.1371/journal.pcbi.1004571 PMID: 26528548
16. Collombet S, van Oevelen C, Sardina Ortega JL, Abou-Jaoudé W, Di Stefano B, Thomas-Chollier M, et al. Logical modeling of lymphoid and myeloid cell specification and transdifferentiation. Proceedings of the National Academy of Sciences. 2017; 114(23):5792–5799. https://doi.org/10.1073/pnas.1610622114 PMID: 28584084
17. Flobak Å, Baudot A, Remy E, Thommesen L, Thieffry D, Kuiper M, et al. Discovery of drug synergies in gastric cancer cells predicted by logical modeling. PLoS computational biology. 2015; 11(8). https://doi.org/10.1371/journal.pcbi.1004426 PMID: 26317215
18. Cantwell-Dorris ER, O'Leary JJ, Sheils OM. BRAFV600E: implications for carcinogenesis and molecular therapy. Molecular cancer therapeutics. 2011; 10(3):385–394. https://doi.org/10.1158/1535-7163.MCT-10-0799 PMID: 21388974
19. Chapman PB, Hauschild A, Robert C, Haanen JB, Ascierto P, Larkin J, et al. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. New England Journal of Medicine. 2011; 364(26):2507–2516. https://doi.org/10.1056/NEJMoa1103782 PMID: 21639808
20. Kopetz S, Desai J, Chan E, Hecht J, O'dwyer P, Lee R, et al. PLX4032 in metastatic colorectal cancer patients with mutant BRAF tumors. Journal of Clinical Oncology. 2010; 28(15_suppl):3534–3534. https://doi.org/10.1200/jco.2010.28.15_suppl.3534
21. Prahallad A, Sun C, Huang S, Di Nicolantonio F, Salazar R, Zecchin D, et al. Unresponsiveness of colon cancer to BRAF (V600E) inhibition through feedback activation of EGFR. Nature. 2012; 483(7387):100–103. https://doi.org/10.1038/nature10868 PMID: 22281684
22. Poulikakos PI, Persaud Y, Janakiraman M, Kong X, Ng C, Moriceau G, et al. RAF inhibitor resistance is mediated by dimerization of aberrantly spliced BRAF (V600E). Nature. 2011; 480(7377):387. https://doi.org/10.1038/nature10662 PMID: 22113612
23. Sun C, Wang L, Huang S, Heynen GJ, Prahallad A, Robert C, et al. Reversible and adaptive resistance to BRAF (V600E) inhibition in melanoma. Nature. 2014; 508(7494):118–122. https://doi.org/10.1038/nature13121 PMID: 24670642
24. Yaeger R, Corcoran RB. Targeting alterations in the RAF–MEK pathway. Cancer discovery. 2019; 9(3):329–341. https://doi.org/10.1158/2159-8290.CD-18-1321 PMID: 30770389
25. Baur F, Nietzer SL, Kunz M, Saal F, Jeromin J, Matschos S, et al. Connecting cancer pathways to tumor engines: a stratification tool for colorectal cancer combining human in vitro tissue models with boolean in silico models. Cancers. 2020; 12(1):28. https://doi.org/10.3390/cancers12010028
26. Cho SH, Park SM, Lee HS, Lee HY, Cho KH. Attractor landscape analysis of colorectal tumorigenesis and its reversion. BMC systems biology. 2016; 10(1):96. https://doi.org/10.1186/s12918-016-0341-9 PMID: 27765040

27. Hyman DM, Puzanov I, Subbiah V, Faris JE, Chau I, Blay JY, et al. Vemurafenib in multiple nonmelanoma cancers with BRAF V600 mutations. New England Journal of Medicine. 2015; 373(8):726–736. https://doi.org/10.1056/NEJMoa1502309 PMID: 26287849

28. Ducreux M, Chamseddine A, Laurent-Puig P, Smolenschi C, Hollebecque A, Dartigues P, et al. Molecular targeted therapy of BRAF-mutant colorectal cancer. Therapeutic advances in medical oncology. 2019; 11:1758835919856494. https://doi.org/10.1177/1758835919856494 PMID: 31244912

29. Saadatpour A, Albert R. Boolean modeling of biological regulatory networks: a methodology tutorial. Methods. 2013; 62(1):3–12. https://doi.org/10.1016/j.ymeth.2012.10.012 PMID: 23142247

30. Abou-Jaoudé W, Traynard P, Monteiro PT, Saez-Rodriguez J, Helikar T, Thieffry D, et al. Logical modeling and dynamical analysis of cellular networks. Frontiers in genetics. 2016; 7:94. https://doi.org/10.3389/fgene.2016.00094 PMID: 27303434

31. Stoll G, Viara E, Barillot E, Calzone L. Continuous time Boolean modeling for biological signaling: application of Gillespie algorithm. BMC systems biology. 2012; 6(1):116. https://doi.org/10.1186/1752-0509-6-116 PMID: 22932419

32. Sedghamiz H, Morris M, Craddock TJA, Whitley D, Broderick G. Bio-ModelChecker: Using Bounded Constraint Satisfaction to Seamlessly Integrate Observed Behavior With Prior Knowledge of Biological Networks. Frontiers in Bioengineering and Biotechnology. 2019; 7:48. https://doi.org/10.3389/fbioe.2019.00048 PMID: 30972331

33. Hernandez C. Dynamical modelling of T cell inhibitory mechanisms in the immune response to cancer. Paris Sciences et Lettres; 2019. Available from: https://www.theses.fr/s175716.

34. van der Meer D, Barthorpe S, Yang W, Lightfoot H, Hall C, Gilbert J, et al. Cell Model Passports—a hub for clinical, genetic and functional datasets of preclinical cancer models. Nucleic acids research. 2019; 47(D1):D923–D929. https://doi.org/10.1093/nar/gky872 PMID: 30260411

35. Béal J, Montagud A, Traynard P, Barillot E, Calzone L. Personalization of logical models with multi-omics data allows clinical stratification of patients. Frontiers in physiology. 2019; 9:1965. https://doi.org/10.3389/fphys.2018.01965 PMID: 30733688

36. Chakravarty D, Gao J, Phillips S, Kundra R, Zhang H, Wang J, et al. OncoKB: a precision oncology knowledge base. JCO precision oncology. 2017; 1:1–16. https://doi.org/10.1200/PO.17.00011 PMID: 28890946

37. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics. 2017; 33(18):2938–2940. https://doi.org/10.1093/bioinformatics/btx364 PMID: 28645171

38. Fey D, Halasz M, Dreidax D, Kennedy SP, Hastings JF, Rauch N, et al. Signaling pathway models as biomarkers: Patient-specific simulations of JNK activity predict the survival of neuroblastoma patients. Sci Signal. 2015; 8(408):ra130–ra130. https://doi.org/10.1126/scisignal.aab0990 PMID: 26696630

39. Lim CY, Wang H, Woodhouse S, Piterman N, Wernisch L, Fisher J, et al. BTR: training asynchronous Boolean models using single-cell expression data. BMC bioinformatics. 2016; 17(1):355. https://doi.org/10.1186/s12859-016-1235-y PMID: 27600248

40. Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic acids research. 2012; 41(D1):D955–D961. https://doi.org/10.1093/nar/gks1111 PMID: 23180760

41. Vis DJ, Bombardelli L, Lightfoot H, Iorio F, Garnett MJ, Wessels LF. Multilevel models improve precision and speed of IC50 estimates. Pharmacogenomics. 2016; 17(7):691–700. https://doi.org/10.2217/pgs.16.15 PMID: 27180993

42. Behan FM, Iorio F, Picco G, Gonçalves E, Beaver CM, Migliardi G, et al. Prioritization of cancer therapeutic targets using CRISPR–Cas9 screens. Nature. 2019; 568(7753):511. https://doi.org/10.1038/s41586-019-1103-9 PMID: 30971826

43. Meyers RM, Bryan JG, McFarland JM, Weir BA, Sizemore AE, Xu H, et al. Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. Nature genetics. 2017; 49(12):1779–1784. https://doi.org/10.1038/ng.3984 PMID: 29083409

44. Hart T, Moffat J. BAGEL: a computational framework for identifying essential genes from pooled library screens. BMC bioinformatics. 2016; 17(1):164. https://doi.org/10.1186/s12859-016-1015-8 PMID: 27083490

45. Dempster JM, Pacini C, Pantel S, Behan FM, Green T, Krill-Burger J, et al. Agreement between two large pan-cancer CRISPR-Cas9 gene dependency data sets. Nature Communications. 2019; 10(1):1–14. https://doi.org/10.1038/s41467-019-13805-y PMID: 31862961

46. Breiman L. Random forests. Machine learning. 2001; 45(1):5–32. https://doi.org/10.1023/A:1010933404324

47. Ishwaran H, et al. Variable importance in binary regression trees and forests. Electronic Journal of Statistics. 2007; 1:519–537. https://doi.org/10.1214/07-EJS039

48. Ursem C, Atreya CE, Van Loon K. Emerging treatment options for BRAF-mutant colorectal cancer. Gastrointestinal cancer: targets and therapy. 2018; 8:13. https://doi.org/10.2147/GICTT.S125940 PMID: 29628780

49. Rossi A, Roberto M, Panebianco M, Botticelli A, Mazzuca F, Marchetti P. Drug resistance of BRAF-mutant melanoma: Review of up-to-date mechanisms of action and promising targeted agents. European journal of pharmacology. 2019; p. 172621. https://doi.org/10.1016/j.ejphar.2019.172621 PMID: 31446019

50. Wellbrock C, Karasarides M, Marais R. The RAF proteins take centre stage. Nature reviews Molecular cell biology. 2004; 5(11):875–885. https://doi.org/10.1038/nrm1498 PMID: 15520807

51. Consortium U. UniProt: a worldwide hub of protein knowledge. Nucleic acids research. 2019; 47(D1): D506–D515. https://doi.org/10.1093/nar/gky1049

52. Lake D, Corrêa SA, Müller J. Negative feedback regulation of the ERK1/2 MAPK pathway. Cellular and Molecular Life Sciences. 2016; 73(23):4397–4413. https://doi.org/10.1007/s00018-016-2297-8 PMID: 27342992

53. Edwin F, Anderson K, Ying C, Patel TB. Intermolecular interactions of Sprouty proteins and their implications in development and disease. Molecular pharmacology. 2009; 76(4):679–691. https://doi.org/10.1124/mol.109.055848 PMID: 19570949

54. Lehr S, Kotzka J, Avci H, Sickmann A, Meyer HE, Herkner A, et al. Identification of major ERK-related phosphorylation sites in Gab1. Biochemistry. 2004; 43(38):12133–12140. https://doi.org/10.1021/bi049753e PMID: 15379552

55. Perfetto L, Briganti L, Calderone A, Cerquone Perpetuini A, Iannuccelli M, Langone F, et al. SIGNOR: a database of causal relationships between biological entities. Nucleic acids research. 2016; 44(D1): D548–D554. https://doi.org/10.1093/nar/gkv1048 PMID: 26467481

56. Naldi A, Hernandez C, Abou-Jaoudé W, Monteiro PT, Chaouiya C, Thieffry D. Logical modeling and analysis of cellular regulatory networks with ginsim 3.0. Frontiers in physiology. 2018; 9:646. https://doi.org/10.3389/fphys.2018.00646 PMID: 29971008

57. Han S, Ren Y, He W, Liu H, Zhi Z, Zhu X, et al. ERK-mediated phosphorylation regulates SOX10 sumoylation and targets expression in mutant BRAF melanoma. Nature communications. 2018; 9(1):1–14. https://doi.org/10.1038/s41467-018-03710-1

58. Li K, Guo Q, Yang J, Chen H, Hu K, Zhao J, et al. FOXD3 is a tumor suppressor of colon cancer by inhibiting EGFR-Ras-Raf-MEK-ERK signal pathway. Oncotarget. 2017; 8(3):5048. https://doi.org/10.18632/oncotarget.13790 PMID: 27926503

59. Gopal YV, Deng W, Woodman SE, Komurov K, Ram P, Smith PD, et al. Basal and treatment-induced activation of AKT mediates resistance to cell death by AZD6244 (ARRY-142886) in Braf-mutant human cutaneous melanoma cells. Cancer research. 2010; 70(21):8736–8747. https://doi.org/10.1158/0008-5472.CAN-10-0902 PMID: 20959481

60. Wroblewski D, Mijatov B, Mohana-Kumaran N, Lai F, Gallagher SJ, Haass NK, et al. The BH3-mimetic ABT-737 sensitizes human melanoma cells to apoptosis induced by selective BRAF inhibitors but does not reverse acquired resistance. Carcinogenesis. 2013; 34(2):237–247. https://doi.org/10.1093/carcin/bgs330 PMID: 23087082

61. Manzano JL, Layos L, Bugés C, de los Llanos Gil M, Vila L, Martinez-Balibrea E, et al. Resistant mechanisms to BRAF inhibitors in melanoma. Annals of translational medicine. 2016; 4(12). https://doi.org/10.21037/atm.2016.06.07 PMID: 27429963

62. Johannessen CM, Boehm JS, Kim SY, Thomas SR, Wardwell L, Johnson LA, et al. COT drives resistance to RAF inhibition through MAP kinase pathway reactivation. Nature. 2010; 468(7326):968–972. https://doi.org/10.1038/nature09627 PMID: 21107320

63. Corcoran RB, Ebi H, Turke AB, Coffee EM, Nishino M, Cogdill AP, et al. EGFR-mediated reactivation of MAPK signaling contributes to insensitivity of BRAF-mutant colorectal cancers to RAF inhibition with vemurafenib. Cancer discovery. 2012; 2(3):227–235. https://doi.org/10.1158/2159-8290.CD-11-0341 PMID: 22448344

64. Montagut C, Sharma SV, Shioda T, McDermott U, Ulman M, Ulkus LE, et al. Elevated CRAF as a potential mechanism of acquired resistance to BRAF inhibition in melanoma. Cancer research. 2008; 68(12):4853–4861. https://doi.org/10.1158/0008-5472.CAN-07-6787 PMID: 18559533

65. Scortegagna M, Lau E, Zhang T, Feng Y, Sereduk C, Yin H, et al. PDK1 and SGK3 contribute to the growth of BRAF-mutant melanomas and are potential therapeutic targets. Cancer research. 2015; 75(7):1399–1412. https://doi.org/10.1158/0008-5472.CAN-14-2785 PMID: 25712345

66. Domany E. Using high-throughput transcriptomic data for prognosis: a critical overview and perspectives. Cancer research. 2014; 74(17):4612–4621. https://doi.org/10.1158/0008-5472.CAN-13-3338 PMID: 25183786

67. Razzaq M, Paulevé L, Siegel A, Saez-Rodriguez J, Bourdon J, Guziolowski C. Computational discovery of dynamic cell line specific Boolean networks from multiplex time-course data. PLoS computational biology. 2018; 14(10):e1006538. https://doi.org/10.1371/journal.pcbi.1006538 PMID: 30372442

68. Azuaje F. Computational models for predicting drug responses in cancer research. Briefings in bioinformatics. 2017; 18(5):820–829. https://doi.org/10.1093/bib/bbw065 PMID: 27444372

69. Aben N, Vis DJ, Michaut M, Wessels LF. TANDEM: a two-stage approach to maximize interpretability of drug response models based on multiple molecular data types. Bioinformatics. 2016; 32(17):i413–i420. https://doi.org/10.1093/bioinformatics/btw449 PMID: 27587657

70. Poulikakos PI, Zhang C, Bollag G, Shokat KM, Rosen N. RAF inhibitors transactivate RAF dimers and ERK signalling in cells with wild-type BRAF. Nature. 2010; 464(7287):427–430. https://doi.org/10.1038/nature08902 PMID: 20179705

71. Manica M, Oskooei A, Born J, Subramanian V, Sáez-Rodríguez J, Rodríguez Martínez M. Toward explainable anticancer compound sensitivity prediction via multimodal attention-based convolutional encoders. Molecular Pharmaceutics. 2019. https://doi.org/10.1021/acs.molpharmaceut.9b00520 PMID: 31618586

72. Saez-Rodriguez J, Blüthgen N. Personalized signaling models for personalized treatments. Molecular Systems Biology. 2020; 16(1). https://doi.org/10.15252/msb.20199042 PMID: 32129942