Designing logical rules to model the response of biomolecular networks with complex interactions: an application to cancer modeling

Carito Guziolowski, Sylvain Blachon, Tatiana Baumuratova, Gautier Stoll, Ovidiu Radulescu, and Anne Siegel

Abstract—We discuss the propagation of constraints in eukaryotic interaction networks in relation with model prediction and identification of critical pathways. In order to cope with post-translational interactions we consider two types of nodes in the network, corresponding to proteins and to RNA. Micro-array data provides very lacunar information for such types of networks because protein nodes, although needed in the model, are not observed. Propagation of observations in such networks leads to poor and not significant model predictions, mainly because rules used to propagate information – usually disjunctive constraints – are weak. Here we propose a new, stronger type of logical constraints that allow us to strengthen the analysis of the relation between micro-array and interaction data. We use these rules to identify which nodes are responsible for a phenotype, in particular for cell cycle progression. We use as benchmark an interaction network describing major pathways implied in the Ewing tumor development.

Index Terms—Systems biology, regulatory networks, post-translational effects, in-silico analysis, automatic reasoning, cancer.

1 Introduction

Formal methods have been proved to be useful to study the complexity of biological systems by modeling the dynamics of small or middle scale gene regulatory networks. Discrete dynamical frameworks have been first introduced in the 70. as logical models, *i.e.* Boolean networks [1]. In such models one considers that genes are either active or inactive, and logical rules allow deciding what will be the state of a system at time $n+1$ from its state at time n. Later, it appeared that such dynamical deterministic modeling is too crude to handle the non-linearity of biological systems, as some trajectories were missed. Eventually, these methods evolved to multi-valued logical models (or their cousins, the piecewise linear models) which allow efficient description of the dynamics of gene regulatory networks of tens of products [2], [3]. Exploring another – less dynamical – direction, the original Boolean formalisms are still very useful to study global properties of large-scale networks by using probabilistic approaches [4], as it was illustrated on the yeast transcriptional network [5].

Recently, as a consequence of the growing amount of gene expression micro-array data, a new strategy of using Boolean rules has appeared. Several studies have used logical disjunctive rules to project gene expression levels onto regulatory networks [6], [7], [8], [9].

In these works authors do not study the dynamics of a network – as it was done in [4] – but they reason with logical rules on variations of products during stress experiments or gene perturbations. In short, first, observations are propagated over a network topology. Then, various consistency tests are applied either to detect contradictions between experimental data and a network or to predict the variation of unobserved components in the network. The detection of contradictions leads to the detection of gaps in the networks. The prediction of unobserved components leads to new insights on the behavior of the regulatory network. Another application of such logical rules focuses on network reconstruction, as shown in [10].

From a theoretical point of view, the main idea underlying these approaches is not to consider the trajectories of systems but to set up constraints on variations between different states. As discussed in [11], [12], [13], a disjunctive logical rule is actually well adapted to describe steady state or gene perturbation effects. It is

- *C. Guziolowski is with the INRIA Rennes Bretagne Atlantique, Campus de Beaulieu, Rennes 35042 France, and Hamamatsu Tissue Imaging and Analysis (TIGA) Center, Institute for Medical Biometry and Informatics, University of Heidelberg, Heidelberg, Germany.*
  E-mail: carito.guziolowski@irisa.fr, carito.guziolowski@bioquant.uni-heidelberg.de
- *S. Blachon is with the INRIA Rennes Bretagne Atlantique, Campus de Beaulieu, Rennes 35042 France, and Bioinformatics Group, Max Planck Institute of Molecular Plant Physiology Wissenschaftspark Golm, Am Mhlenberg 1, 14476 Potsdam - Golm, Germany.*
  E-mail: sylvain.blachon@gmail.fr
- *T. Baumuratova is with University of Luxembourg, 162 A, avenue de la Faencerie, L-1511 Luxembourg.*
- *G. Stoll is with Institut Curie and INSERM, U900, Paris, F-75248 France; and Mines ParisTech, Fontainebleau, F-77300 France.*
- *O. Radulescu is with Université de Montpellier 2, DIMNP - UMR 5235 CNRS/UM1/UM2, and INRIA Rennes Bretagne Atlantique, Campus de Beaulieu, Rennes 35042 France.*
- *A. Siegel is with CNRS, Université Rennes 1 UMR 6074 IRISA, and INRIA Rennes Bretagne Atlantique, Campus de Beaulieu, 35042 Rennes France.*

also quite natural from the perspective of monotone subsystems [14]. This point of view provides a new application of the original Boolean rules of [4] in a considerably modified context that ensures their validity (variations during shifts instead of dynamical simulation). Basing on this idea, we have previously proposed a formal method to investigate the global consistency of a regulatory model with large-scale differential gene expression data. A general interaction logical rule (similar to a logical disjunction) was used to address the question of consistency. We proved its efficiency on large-scale prokaryotic networks, such as the *E. coli* transcriptional regulatory network. This approach allowed identifying errors in input data and highlighting sub-graphs where interactions needed to explain the global behavior of the network were missing [15].

Regulatory networks in eukaryotic cells hold, however, more complex interactions, since numerous post-translational interactions are involved in all critical biological functions. By post-translational interactions we understand all direct or indirect interactions between proteins that occur after their production, including post-translational modifications, protein-complex interactions, competition between pathways, sequestration, releasing, and complex inactivation.

However, it appears that constraints imposed by the – commonly used and biologically realistic – generic disjunctive rule on networks with post-translational interactions are too weak to detect defects in the model. Their prediction power is also low. The main reason for this weakness is the following: in networks with post-translational interactions protein activity cannot always be directly read from RNA variations measured by micro-arrays [16], [17]. This activity results from complex biochemical interactions of many components (including RNA). Therefore, in some cases, micro-array observations cannot propagate to protein nodes. Consequently, the values of the protein nodes are less constrained by expression level variations, and can be freely chosen in order to satisfy the constraints.

In this paper we address the question of such complex eukaryotic networks. In order to build more realistic models of regulatory networks in eukaryotes, we propose to decouple active protein or complexes from RNA levels and then reformulate the effect of complex interactions acting on protein nodes as new, specific logical rules. We illustrate how such specific logical rules accommodate less compatible valuations and therefore introduce stronger constraints. The benchmark model is an interaction network describing major pathways implied in the Ewing tumor development [18].

Concretely, we have introduced specific ternary logical rules to model the effect of post-translational processes on the variation of certain products; we deduced them from biological information found in the literature. These rules were incorporated to an automatic reasoning framework based on dependency graphs. This automatic framework allows propagating information along the network topology and reasoning over contradictions between observed variations and variations that may be expected from the knowledge on the system. We have used this framework to constrain the global behavior of the network when confronted with qualitative effects obtained from transcriptome time series data. These effects reflect gene profiles responding to the inactivation and reactivation of EWS-FLI1, the main oncogene responsible of Ewing tumor growth.

Our results show that using the new logical rules for post-translational interactions globally increases the number predictions on the qualitative change in expression of unobserved molecules, and the statistical validity of such predictions. We conclude that the use of specific logical rules in network models increases sensitivity and prediction power of constraint-based approaches.

Finally, we used these rules to reason over the interactions that target the cell cycle node in the network. On that account, we developed two algorithms that found which nodes of the network were responsible for triggering the cell cycle inhibition and reactivation. The global *in silico* analysis using specific post-translational rules proposed network molecules that may be interesting targets in order to manipulate the Ewing tumor phenotype.

As already mentioned, annotating interaction networks with transcription data and performing large-scale consistency analysis is an idea developed by many other authors [10], [19]. The main advances presented here, with respect to previous studies (including ours), are the introduction of new logical rules for post-translational interactions together with efficient automatic algorithms for finding sub-graphs responsible for phenotypes. This produces more realistic models and increases the significance of the analyses.

2 METHODS

2.1 Confronting a dataset with a network topology

2.1.1 Input data

Our method confronts regulatory networks with experimental datasets using a consistency criterion based on causal rules. Such formalism requires input data having the following characteristics.

The **network regulations** represent direct influences over the production or activity of a product. That is, if the influence $A \to B$ exists, this means that the increase or decrease in the quantity of A directly influences the increase or decrease in the production rate of B. Depending on our knowledge of the regulatory model, we can classify the regulations in at least three types: '**+**', for transcription, activation, or production; '**–**', for inhibition or inactivation; and '**?**', for unknown regulations. It is also possible, however, to add new regulatory rules in order to explain complex molecular regulations.

We denote by $\mathcal{N} = (V, E)$, the oriented and signed network to be analyzed, where V is the set of nodes and E the set of edges.

The **experimental dataset** must reflect variations of network products in expression or activity. The system is supposed to be at steady state; it is then perturbed by a stress, and eventually reaches a new steady state. Variations reflect changes between the beginning and the end of the perturbation. Intermediate time series variations are not taken into account. Therefore, some network products are classified into: '**+**', standing for a global increasing variation during the experiment (up-regulated genes, or proteins that switch from inactive to active configurations); and '**–**', standing for a decreasing variation (down-regulated genes or proteins that switch from active to inactive configurations).

We denote by $\mu = \{(n, s)\} \subset V \times \{+, -\}$, the dataset of partial significant variations of the nodes in $\mathcal{N}$. With this notation at hand, $dom(\mu)$ is the set of nodes which variation is observed.

Our final goal is to confront the knowledge provided by a regulatory network with the observations provided by the experimental dataset. The most natural confrontation is to decide whether these information do not contradict each other. To perform this task we add a new concept to the input data: **consistency rules**. A consistency rule describes which effects on a target node are considered to be compatible with the variations of the nodes that directly influence it: the predecessors of the target node.

We denote by $\sigma = \{F_T \mid T \in V\}$, the set of regulatory influences (functions), that each node receives. To describe more in detail F_T, let us assume that a node $T \in V$ is directly influenced by a set of nodes $\{S_1, S_2, \ldots, S_p\} \in V$. Let $t \in \{+, -\}$ (resp. s_k) represent the variation of node T (resp. S_k). The function $F_T : \{+, -\}^p \rightarrow \{+, -, ?\}$ represents how the combined variations of the predecessors of T affect t. In some cases this function may output an '**?**' value, which means that t cannot be deduced.

The F_T functions are a full part of the model. In Section 2.2 we will detail these functions, also we will discuss that a generic rule may often be applied; it states that the variation of a target node has to be explained by the influence of at least one predecessor. The aim of this paper is to discuss whether more constraining rules can be designed and used to perform stronger analyses of the network.

2.1.2 Constraint framework to model network and data confrontation

Once a network and an experimental dataset are provided, one evaluates their mutual **consistency**. In short, a network $\mathcal{N}$ will be consistent with a dataset μ if it is possible to assign $\{+, -\}$ variations to all the nodes of $\mathcal{N}$ in a way that will not contradict the influence paths provided by the network topology and described by σ.

This problem can be formalized by representing a network as a system of constraints where each constraint represents the influences that a node in the network receives. For example, assume that a node $T \in V$ is directly influenced by a set of nodes $\{S_1, S_2, \ldots, S_p\} \in V$. Let $t \in \{+, -\}$ (resp. s_k) represent the variation of node T (resp. S_k). We write the following constraint over t:

$$t \quad \simeq \quad F_T(s_1, s_2, \cdots, s_p) \tag{1}$$

The consistency relation $\simeq$ is satisfied if a positive variation of a node does not receive a total negative influence from its predecessors, or vice versa; or if F_T is indeterminate ('**?**') (see Table 1).

If this constraint is satisfied (true), then it states that the variation of product T is *consistent* with the combined influence of its predecessors: F_T.

2.1.3 Consistency check and predictions

Given a network $\mathcal{N}$, an experimental dataset μ, and a set of consistency rules σ, one builds a system of contraints as follows: the variables of the system of constraints represent the change in expression or activity of a network product. Constraints such as Eq. (1) establish a relation between variations of products and their predecessors. The experimental dataset μ is used to constraint this system by fixing some of its variables.

The **consistency** of $\mathcal{N}$ wrt μ and σ will be evaluated in terms of finding at least one $\{+, -\}$ assignment (solution) for all the variables of the system that satisfies all the system constraints.

We may conclude that the observations provided in the experimental dataset μ satisfy the network topology of $\mathcal{N}$. In this case the method generates new long term effects over certain network products that explain the observed values. We call **predictions** of the system of constraints, the set of nodes, other than the observed ones (data), that have the same $\{+, -\}$ value in all the consistent assignments. These nodes are constrained to unique values by data in μ, rules in σ, and Eq. (1).

Alternatively, we may conclude that the dataset μ is inconsistent with the network. In this case we compute a sub-network of $\mathcal{N}$, which explains the contradictory observations or regulations. Concretely, this sub-network – **minimal inconsistent sub-graph** – contains the interactions and observations that have to be corrected in order to make the network and dataset globally consistent. This task is called **diagnosis**.

2.1.4 Implementation

Several heuristics were proposed for the resolution of qualitative systems, such as the design of a complete set of rules based on Gaussian elimination [20]. Nevertheless, they cannot be used for solving biological qualitative systems of constraints because such algorithms need backtracking, which increases computation time. We have used a computational method introduced in [21], which is based on a representation of the sign algebra in $\mathbb{Z}/3\mathbb{Z}$. It uses dependency graphs to combine conjunction operations and variable elimination [22], [15]. Using $\mathbb{Z}/3\mathbb{Z}$ as a model for the sign algebra has

a quite long history in signal theory. It allows us to represent the '**?**' value by an independent variable and therefore to gain in variability in the rules that one may use in the system of constraints.

With this representation at hand, deciding the consistency, diagnosing the inconsistent constraints, and computing the predictions of large-scale systems are performed in a very efficient way [15]. These algorithms are implemented as a Python library BioQualiV2. The main functions we used here are:

- BQ.consistency($\mathcal{N}$,μ) returns **true** if the network is consistent with the dataset, or **false** if not.
- BQ.predictions($\mathcal{N}$,μ)$=\ \{(n,s)|\, n \in V \setminus dom(\mu),\ s \in \{+,-\}\}$, is the set of predictions generated when $\mathcal{N}$ is consistent with μ.
- BQ.inconsistentSubgraph($\mathcal{N}$,μ) outputs a minimal inconsistent sub-graph $\mathcal{I} = (V_I, E_I, \sigma_I)$, where $V_I \subset V$, $E_I \subset E$, and $V_I \cap dom(\mu) \neq \emptyset$. Its nodes and edges gather the origin of the contradictions between network and observations.

This library also includes a functionality allowing easy addition of new regulatory functions by coding them in a specific module. Basic, already defined, sign operators can be used to build a wide range of functions. BioQualiV2 is publicly available on our supplementary Webpage [23].

2.2 Modeling shift variations with constraints

2.2.1 Operators

The functions of type F_T represent how the variations of the predecessors of T affect the global variation of T. As we shall detail it now, these functions can be written by using several basic sign operators: $\otimes$ is the *qualitative product*, $\oplus$ is the *qualitative sum*, $\wedge$ and $\vee$ are the logical disjunction and conjunction. A $\neg$ operator was also added to output the opposite sign of a variation *i.e.* $\neg t = t \otimes$ **–**. These sign operators are shown in Table 1.

TABLE 1
Consistency relation $\simeq$ and sign tables for the $\otimes$, $\oplus$, $\wedge$, and $\vee$ sign operators. The $\simeq$ relation states the consistency answer of each constraint, T stands for true, whereas F for false.

$\simeq$	+	–	?
+	T	F	T
–	F	T	T
?	T	T	T

$\otimes$	+	–	?
+	+	–	?
–	–	+	?
?	?	?	?

$\oplus$	+	–	?
+	+	?	?
–	?	–	?
?	?	?	?

$\wedge$	+	–	?
+	+	–	?
–	–	–	–
?	?	–	?

$\vee$	+	–	?
+	+	+	+
–	+	–	?
?	+	?	?

In this way the functions AND_T and OR_T, representing one type of total influence over a node T, where $\{s_1, \ldots, s_p\}$ stand for the variations of its predecessors, are given by:

$$AND_T(s_1, s_2, \cdots, s_p) = s_1 \wedge s_2 \wedge \ldots \wedge s_p \quad (2)$$

$$OR_T(s_1, s_2, \cdots, s_p) = s_1 \vee s_2 \vee \ldots \vee s_p \quad (3)$$

Other type of total influence over T is given by the, already introduced in [12], GEN_T generic function (see Section 2.2.2 for details). More specific logical functions, built using the operators in Table 1, were designed to model complex regulatory phenomena that appeared in the EWS-FLI1 network. In particular, four mechanisms were detected that deserved to be modeled in a different way than by using the GEN function. Notice, that instead of the generic function GEN that is always satisfied, the new functions that we introduced are designed according to the biological literature on the products implied in the network that we are currently considering. Some rules may be quite generic and used in other contexts, but this deserves a specific mathematical study which is not our purpose presently.

2.2.2 Generic rule

The generic qualitative function $F_T = GEN_T$ describes the influences of the p predecessors of T by taking into account their regulation sign.

$$\begin{aligned} &GEN_T(s_1, \cdots, s_p) \\ &= (s_1 \otimes I_{S_1T}) \oplus (s_2 \otimes I_{S_2T}) \oplus \ldots \oplus (s_P \otimes I_{S_pT}), \end{aligned} \quad (4)$$

where $I_{S_jT} = sign(S_j \to T)$, represents the sign of the edge from S_j to T in the network and it is a value in $\{+, -, ?\}$.

GEN_T represents the influences that arrive over a product T in a generic way. If all the contributions $(s_j \otimes I_{S_jT})$ targeting T will be positive, F_T will be '**+**'. It is enough that a single contribution will have an opposite sign to obtain $F_T =$ '**?**'.

As a general rule, it is difficult to determine the priority or weight of biological regulations; however, in some cases it is possible to model F_T precisely. In the following subsections we provide a list of functions used to model some of the post-translational regulations in the EWS-FLI1 network.

2.2.3 Protein-complex formation

According to [24], in mammalian cells Cdks are constitutively expressed and present in excess to D-type cyclins even at times of maximal cyclin induction. This allows us to model Cyclin D-Cdk4,6 complexes by using only the cyclin D variation. Thus, the combined influence that the Cyclin D-Cdk4,6 complex receives is determined by the function in Eq. (5).

$$F_{CCND-CDK} = ccnd \quad (5)$$

2.2.4 Strong inhibitor

Let us take the RB inhibition of cell cycle as a canonical example of strong inhibitor. The RB protein family is a canonic member of tumor-suppressors. They act as default inhibitors of E2F protein family, which are transcription factors implied in cell cycle progression. According to [25], RB proteins act by sequestrating E2F. When RB proteins are phosphorylated, E2F proteins are released. This releasing triggers cell division, in particular transition to the S-phase. An illustration of these regulations is given in Fig. 1A, where the valuations of E2F, RB, and cell_cycle_S are consistent with respect to the literature. Consequently, the inhibition of E2F by RB was modeled using a *Strong-inhibitor* function, Eq. (6), adapted to all default inhibitors.

$$F_{cell\_cycle\_S} = e2f \wedge \neg rb1 \quad (6)$$

2.2.5 Complex inactivation

Most of the protein-complex formations in our network are modeled using the GEN function. However, some proteins can hamper the protein-complex formation. As discussed in [24], this is the case for the complex CCNE-CDK2: WEE1 phosphorylates CDK2 on tyrosine and threonine residues, causing the complex inactivation. In Fig. 1B we illustrate this case and we model it using the *Complex-inactivation* function shown in Eq. (7).

$$F_{CCNE\_CDK2} = \neg wee1 \wedge (ccne \oplus cdk2) \quad (7)$$

2.2.6 Complex inactivation-reactivation

In some cases a protein-complex may be under an inactivation influence that is itself inhibited. For example, as shown in [26], CCNA forms a complex with CDK2, which is inhibited by cyclin-dependent-kinase-inhibitor CDKN1A. This inhibition is reverted by CCND-CDK4-a, the active complex formed by Cyclin D and cyclin-dependent-kinase CDK4. This situation is illustrated in Fig. 1C. We model it using the *Complex-inactivation-reactivation* function shown in Eq. (8).

$$F_{CCNA\_CDK2} = (ccna \oplus cdk2) \wedge (\neg cdkn1a \vee ccnd\_cdk4\_a) \quad (8)$$

2.3 Algorithms

We used the BioQualiV2 functionalities (see Section 2.1.4) to provide two methods that, given a network $\mathcal{N}$ and a dataset μ of observations, obtained a subset of interactions of $\mathcal{N}$ and a subset of observations of μ that minimally explained a known fact.

A first method is given in **Algorithm 1**. It recovers a minimal sub-graph that connects observed values with a node of interest n. The idea underlying it is to find the pathways that activate/inhibit n. This algorithm is applied when we do not have precise information concering the priority order or strength of the regulators of

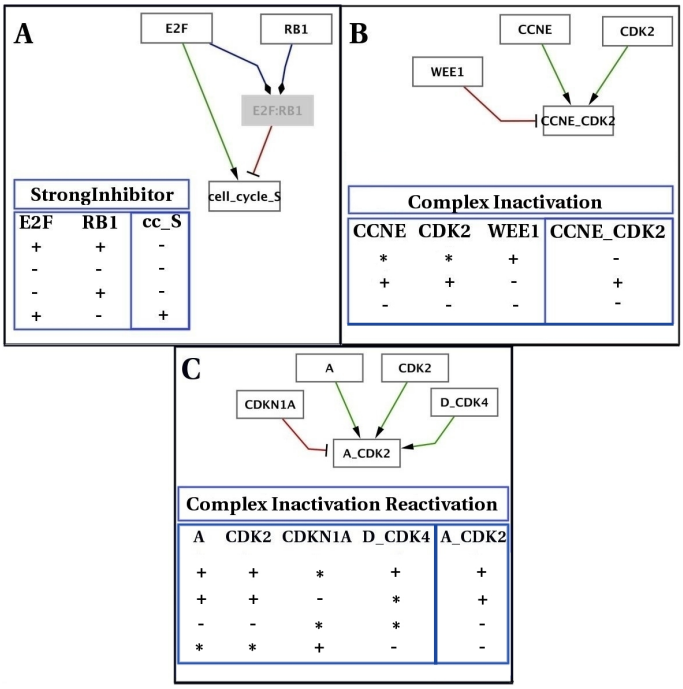

Fig. 1. Regulations from the EWS-FLI1 network modeled using specific logical functions. In each table we show all the possible $\{+,-\}$ variations of the predecessors of a node, so that the node will be predicted to a $\{+,-\}$ value. The * symbol refers to either a '+' or '–' variation.

n, that is, when n is modeled using the generic function GEN. The main steps of this algorithm are:

1) Create a temporal node for each influence that n receives. Model the new influences coming from the temporal layer of nodes using the generic function: GEN_n.
2) Run a consistency test with the dataset of observations.
3) If any node of the temporary layer is predicted, then its predicted value is reverted. This will generate an inconsistency, and the inconsistent sub-graph associated can be easily found.
4) The final graph will be a merge of all the inconsistent sub-graphs found.

The inconsistent sub-graph generated when the sign of an original prediction was changed, proposes the minimal paths, connecting the observed nodes, which are necessary to explain the original prediction.

It may happen that none of the nodes in the temporal layer will be predicted when using Algorithm 1. This means that the model, given by the dataset of observations and the rules that modeled the network topology, cannot propagate the effect of the observed gene variations over a node of interest (n). In that case, we propose a second method that finds which network (unobserved, unpredicted) product has to be *artificially fixed* (*e.g.* inhibited by a sh-RNA) to explain a known (experimentally observed) change of n without contra-

Algorithm 1 Find the minimal sub-graph that explains the $\{+,-\}$ sign of a node in the network. BQ refers to the BioQualiV2 Python package

Require: $\mathcal{N} = (V, E, \sigma)$: an oriented graph with V vertices, E edges, and $\sigma = \{F_T | T \in V\}$ as the set of combined influences for each network node
$\mu = \{(n, s) | n \in V, s \in \{+,-\}\}$: a partial valuation of the nodes of $\mathcal{N}$ (dataset of observations)
$n \in V$: a node of interest
BQ.consistency($\mathcal{N}$,μ) is **true**
Ensure: $T = (V_T, E_T), V_T \subset V, E_T \subset E$
$T \Leftarrow \emptyset$
$\mathcal{N}2 \Leftarrow$ addFuncPredec($\mathcal{N}, n$,GEN)
$fixed \Leftarrow$ BQ.predictions($\mathcal{N}2$,μ)
for p in $\mathcal{N}2$.predecessors(n) **do**
if $p \in dom(fixed)$ **then**
$\overline{\mu} \Leftarrow \mu$; $\overline{\mu}(p) \Leftarrow \neg fixed(p)$
$i \Leftarrow$ BQ.inconsistentSubgraph($\mathcal{N}2$,$\overline{\mu}$)
T.add(i)
end if
end for

Algorithm 2 Find a list of pairs $\{(m, s) | m \in V, s \in \{+,-\}\}$ that explain a fixed $\{+,-\}$ variation of a node of interest n. Find the sub-graph T connecting each node in the list to n. BQ refers to the BioQualiV2 Python package

Require: $\mathcal{N} = (V, E, \sigma)$: an oriented graph with V vertices, E edges, and $\sigma = \{F_T | T \in V\}$, the set of combined influences for each network node
$\mu = \{(n, s) | n \in V, s \in \{+,-\}\}$: a partial valuation of the nodes of $\mathcal{N}$ (dataset of observations)
$n \in V$: a node of interest
$sign \in \{+,-\}$: the sign of n to be explained
BQ.consistency($\mathcal{N}$,μ) is **true**
Ensure: $\{(m, s)\}$, $T = (V_T, E_T), V_T \subset V, E_T \subset E$
$\{(m, s)\} \Leftarrow \emptyset$; $T \Leftarrow \emptyset$
$pred1 \Leftarrow$ BQ.predictions($\mathcal{N}$,μ)
if $sign$ is '+' **then**
$\mathcal{N}2 \Leftarrow$ addFuncPredec($\mathcal{N}$,n,AND)
else
$\mathcal{N}2 \Leftarrow$ addFuncPredec($\mathcal{N}$,n,OR)
end if
$\mu_2 \Leftarrow \mu$; $\mu_2(n) \Leftarrow sign$
$pred2 \Leftarrow$ BQ.predictions($\mathcal{N}2$,μ_2)
$\{(m, s)\} \Leftarrow pred2 \setminus pred1$
for $(\overline{m}, \overline{s})$ in {(m,s)} **do**
$\mu_3 \Leftarrow \mu_2$; $\mu_3(\overline{m}) \Leftarrow \neg\overline{s}$
$i \Leftarrow$ BQ.inconsistentSubgraph($\mathcal{N}2$,μ_3)
T.add(i)
end for

dicting the observations. For that reason, we impose variations over the regulators of n and we run the consistency test in order to see if they are contradictory with our model. In the case the model is consistent with our imposed changes, the origin of these changes is tracked down. This method is detailed in **Algorithm 2**. The main steps of this algorithm are:

1) Run a first consistency test with the dataset of observations and store the predictions.
2) Create a temporal node for each influence that n receives. Model the new influences coming from the temporal layer of nodes using a non-generic function: AND_n or OR_n, depending on the sign of n. This will fix the signs of the nodes in the temporal layer to the same sign of n.
3) Run a second consistency test with the dataset of observations and store the predictions.
4) Create a subset of the newly obtained predictions (different from those obtained in Step 1).
5) For each node in the list of new predictions, revert its predicted value and find out the associated inconsistent sub-graph.
6) The final graph will be a merge of all the inconsistent sub-graphs found. The *artificially fixed* nodes are those obtained in Step 4.

Both algorithms are provided with the BioQualiV2 Python library on our supplementary Webpage [23].

2.4 Estimating the significance of consistency and prediction

The usual protocol when confronting data and interactions is first to test the consistency, then to predict the behavior of unobserved nodes. This analysis is built on top of statistical tests for differential expression. Prior to it, signs of variations have been assigned to some of the nodes, by using p-values computed from the micro-array transcription data. In the present study, we have combined p-values obtained from linear regression and from the Student test for time series expression data (see our supplementary Webpage [23]). Even by considering time series and linear regression, the sign assignment suffers from the well-known statistical drawback of micro-arrays that generates many false-positives when thousands of genes are tested.

Our consistency test represents a meta-analysis technique allowing to asses false discovery rates. It is complementary to other meta-analysis methods used to test significance of combined results [27]. Similar consistency tests have been independently discussed in [28]. However, we believe that our examples of analysis of consistency [15], [13], based on rigorous mathematical results [11] and quick, efficient solvers for qualitative equations [21], [22], supersede in speed and accuracy similar attempts [28] for automatic reasoning. As will be discussed next, the introduction of new logical rules further increases the performance of our algorithms.

In order to evaluate the significance of a consistency diagnosis and the significance of a prediction on a given node, we introduce indicators based on the predictions obtained from random datasets. For a given dataset μ

the set of mRNA nodes V_R in a network model can be divided into three subsets: $V_R = S_+ \cup S_- \cup S_0$, where $S_+ = \mu^{-1}(+)$, $S_- = \mu^{-1}(-)$ denote the sets of nodes observed with positive and negative variation respectively – these are RNA nodes whose variations are significant (have small p-value) in the gene micro-array – S_0 denotes a set of RNA nodes whose variation is not significant (have large p-values). The random datasets are obtained by randomly shuffling the elements of V_R, while keeping the cardinalities $n_+ = |S_+|$, $n_- = |S_-|$, $n_0 = |S_0|$ fixed. The proportion of consistent datasets in all the random datasets (the probability to obtain consistency by chance) represents the significance of the consistence test and will be denoted by $p_{consistent}$. This probability depends on the cardinalities n_+, n_-, n_0, as well as on the topology and signs of the interaction network.

By propagating constraints on the network, our method can predict the values of some protein and mRNA nodes in the network $\mathcal{N}$. Next, we investigate the probability that a specific prediction on a node n happens by chance. To that purpose, we define the *Prediction probability* $P_{pred}(n, s)$ that a node n in the network $\mathcal{N}$ is predicted with a $s \in \{+, -\}$ value as the ratio between the number of random datasets predicting the value s for n and the number of consistent random datasets.

In other words, the prediction probability indicates the degree of confidence of a given prediction. With this point of view, predictions with smallest prediction probability shall be considered as the more informative on the model and have the top priority to be experimentally confirmed.

3 Effects of Logical Rules on Consistency and Predictions

We applied our methods on a benchmark model that describes the effects of a fusion oncogene EWS-FLI1 on cell cycle, inducing young adult cancers [29].

3.1 Benchmark data

Time series data and annotated gene regulatory network were produced at Institute Curie[1] [18].

3.1.1 Time series micro-array data

The A673 cell line derived from a Ewing tumor was modified as follows: after induction by doxocyclin, a sh-RNA targeting EWS-FLI1 is produced, inactivating EWS-FLI1. This stops cell division. The gene response to EWS-FLI1 inactivation was investigated by using an Affymetrix HG-U133 Plus 2.0 micro-array during 17 days on two independent clones. At day 11, cells from both clones were washed and harvested in a solution without doxocyclin. When EWS-FLI1 is reactivated, the cell division restarts. The gene response to EWS-FLI1 reactivation was investigated by using the same micro-array between day 11 and day 17.

1. http://bioinfo-out.curie.fr/projects/sitcon/

3.1.2 Experimental dataset - encoding gene variation

Time series data on Ewing inducible cell lines were analyzed to select genes which show a significant response on both inhibition and reactivation of EWS-FLI1.

We consider a gene to correlate with EWS-FLI1 behavior if: (1) the gene is inhibited upon the EWS-FLI1 inhibition (Day0-Day11) and reactivated with reactivation of the oncogene, and (2) the responses are significant. Likewise, if a gene is up-regulated at EWS-FLI1 inhibition and down-regulated with EWS-FLI1 reactivation, showing the significant variation of its response, such gene is considered as anti-correlated with the oncogene.

Two statistical tests, linear regression and Student's t-test, were applied to the micro-array data corresponding to inhibition and reactivation of EWS-FLI1. If a gene had the following characteristics: (1) P-value $< 5\%$ for at least one of the statistical tests, (2) P-value $< 5\%$ for both parts of its response curves (inhibition and reactivation), and (3) opposite behavior during the inhibition and reactivation of the oncogene, it was considered to be correlated or anti-correlated with EWS-FLI1.

With the correlated and anti-correlated mRNAs two datasets of qualitative observations were built:

(A) *EWS-FLI1 inhibition*. This dataset represented the qualitative global change in time of the mRNAs when EWS-FLI1 was inhibited (Day0-Day11). It was built by classifying as '−' the correlated mRNAs, and as '+' the anti-correlated mRNAs. It consisted of 54 $\{+, -\}$ variations of the mRNAs present in the EWS-FLI1 signaling network.

(B) *EWS-FLI1 reactivation*. This dataset represented the qualitative global change in time of the mRNAs when the EWS-FLI1 oncogene was reactivated (Day11-Day17). It was obtained by changing the signs of dataset A to their opposite value.

3.1.3 Native regulatory network – Generic model

In an independent work, an annotated gene regulatory and signaling model involving 130 genes, including EWS-FLI1, was designed from the genes that responded to EWS-FLI1 as follows. Using information from TRANSPATH and literature, interactions were selected to describe signal pathways that regulate key functions involved in tumor progression (cell cycle phase transitions, apoptosis, and cell migration) [18].

From this native model we designed an influence graph. In order to capture the effect of post-translational regulations, each node of the native model was divided into two molecular species: *mRNA nodes*, and *protein nodes*. Influences between nodes were set according to the nature of each interaction (transcriptional or post-translational) provided by the annotation of the native model. The resulting model contained 287 nodes and 644 edges.

From this influence graph we built a first system of constraints, called *Generic model*, by imposing the generic function GEN_T (Eq. (4)) over each node T of the network.

3.1.4 Refined model

In order to constrain more the behavior of the system, we designed a new influence graph based on the native regulatory network described in the previous subsection. This new graph had the following characteristics:

- The phosphorylated proteins were modeled according to their change in activity after the phosphorylation. That is, if A is a protein that becomes active when phosphorylated by B, we added in our influence graph the interaction '$B -> A^{act}$ +'. In the opposite case, we added the interaction '$B -> A^{act}$ –'. The native regulatory network modeled the phosphorylations as state-change transitions, while the influence graph of the Refined model models them as positive or negative influences over the protein activity.
- The variation of 36 network products was modeled using an specific (non-generic) qualitative function, designed taking into account the specificity of post-translational interactions. The four qualitative functions introduced into this model were detailed in Eqs. (5), (6), (7), (8).
- The nodes that were neither phosphorylated proteins nor receiving a specific qualitative function were modeled using the generic function.

The model of constraints associated to the new influence graph was called *Refined model*. Notice that instead of the generic function GEN, that is always satisfied, the new functions that we introduced in the Refined model are designed according to the biological literature on the products implied in the network that we are currently considering. Some rules may be quite generic and used in other contexts, but this deserves a specific mathematical study which is not our purpose presently. The Refined model of the EWS-FLI1 network had 296 nodes (36 received a non-generic qualitative function), and 430 $\{+,-\}$ edges. The nodes in this network were mRNA nodes and *active* protein nodes.

3.2 Results

We applied the process described in Section 2.1.3 to test the consistency between the Generic and the Refined model with respect to observed variations during inhibition and reactivation of EWS-FLI1 (Section 3.1.2). Both models were found to be consistent with both datasets.

When confronted with the variations observed during *inhibition* of EWS-FLI1, 37 nodes were predicted by the Generic model, while 55 were predicted by the Refined model. With the Generic model 2 mRNA nodes, 2 protein-complexes, and 33 protein activities were predicted. With the Refined model 6 mRNA nodes, 13 protein-complexes, and 36 protein activities were predicted.

TABLE 2
New predictions obtained using the Refined model. The third column specifies which nodes were modeled using a specific qualitative function.

Predicted node			Refined Model Prediction Prob.	Specific Function
CCNA2_CDC2_act	→	–	0.00601	X
CCNA1_CDC2_act	→	–	0.00582	X
CCNB1_CDC2_act	→	–	0.00547	X
CCND3_CDK4_act	→	–	0.00463	X
CCND3_CDK6_act	→	–	0.00463	X
CCNA1_CDK2_act	→	–	0.00457	X
CCNA2_CDK2_act	→	–	0.00447	X
E2F2	→	–	0.00408	X
CCND2_CDK4_act	→	–	0.00277	X
CCND2_CDK6_act	→	–	0.00277	X
E2F3	→	–	0.00248	X
CCNB1_CDC2	→	–	0.00248	X
E2F5	→	–	0.00232	X
CCND1_CDK6_act	→	–	0.00222	X
CCND1_CDK4_act	→	–	0.00222	X
mRNA_TP73	→	–	0.00006	
mRNA_APAF1	→	–	0.00006	
mRNA_CDKN2A	→	–	0.00003	
mRNA_RB1	→	–	0	

The list of new predictions obtained with the Refined model using gene variations observed after EWS-FLI1 inhibition is shown in Table 2. All predictions from the Generic model but one (RBL1) were also predicted from the Refined model. The RBL1 protein was modeled in the Generic model without taking into account the influences that phosphorylate it, thus its change was easy to predict as it received only one influence coming from its mRNA node. In the Refined model, however, the *active* protein RBL1 received also phosphorylation influences; from the values of these influences it was not possible to predict a definite change for RBL1.

As expected, designing logical rules to describe the effects of post-translational interactions significantly increased the number of predictions (+ 44.7%). Additionally, most of the new predictions target protein or protein-complex nodes. We will analyze them in the pathways perturbed by EWS-FLI1 inhibition, then reactivation in Section 4.

Observed variations during the inhibition of EWS-FLI1 produced a dataset of 54 observations on mRNA nodes (denoted by $S_+ \cup S_-$). For both models, the Generic and the Refined, 1000 random datasets were generated.

From our analysis we obtained that the Generic model had a $p_{consistent}$ of 0.52 (524 random consistent datasets over 1000) whereas the Refined model had a $p_{consistent}$ of 0.31 (312 random consistent datasets). This confirms that the generic functions used in the Generic model do not constrain enough gene variations. Furthermore, adding logical rules only on 36 nodes (12% of the total) highly reduces the range of random observations compatible with the topology of the model. In other words, the Refined model is more specific than the Generic model.

Finally, we computed the prediction probabilities, that is, the chance of obtaining the same predictions when the models are confronted with consistent random datasets (see Table 2). The prediction probability (P_{pred}) was 0.31 on average for the predictions of the Generic model. It decreased to 0.21 for the predictions of the Refined model. Moreover, 34% of the predictions from the Refined model had a P_{pred} inferior to 0.05, while only 3% of the predictions from the Generic model had a P_{pred} inferior to 0.05. The nodes with low prediction probability express that their $\{+,-\}$ value is not easy to predict by chance, being specific to the EWS-FLI1 signals. In Table 3 we show comparative values between both types of models: generic and refined.

TABLE 3
Comparative results obtained using the Refined and the Generic models

	Ref. model	Gen. model
Nodes receiving non-generic rules	36	0
Predictions	55	37
Consistency test significance	0.31	0.52
Prediction probability (on average)	0.2	0.3
Stronger predictions ($P_{pred} < 0.05$)	30%	5%

In addition, refining a model increases the robustness with respect to noise effects in the dataset of observations: incorrect observations have very few effects over the predictions, although they may influence the consistency test – which is a good point, since this allows us to detect errors in the dataset (see details on our supplementary Webpage [23]).

Altogether, designing logical rules allows significantly reducing the range of observations compatible with the model. Also, it generates non-trivial predictions with low prediction probability. The prediction probabilities are in the trivial case reduced to the probability of choosing a specific mRNA by chance. This is the case of all the predictions obtained with the Generic model; by adding logical rules in the model, we modify the distribution of their prediction probabilities. The Prediction probabilities for all the predicted nodes are available on our supplementary Webpage [23].

4 Studying the correlation between the cell cycle S-phase progression and the EWS-FLI1 activation

The logical rules added to the model increased the number of $\{+,-\}$ predictions on the network nodes. This opens the way to investigate which signaling pathways can explain an observed phenotype. In Section 2.3 we proposed two methods that given a network $\mathcal{N}$ and a dataset μ of observations, obtained a subset of interactions of $\mathcal{N}$ and a subset of observations of μ that minimally explained a known fact. In our case we were interested on understanding:

- the arrest of cell cycle when EWS-FLI1 is inactivated, and
- the restarting of cell cycle when EWS-FLI1 is reactivated.

On that account we confronted the *Refined model* with the significant $\{+,-\}$ variations of the network nodes issued from EWS-FLI1 inactivation, then reactivation. The output of these methods, shown as a graph, represented a cascade of interactions originated by a small set of genes. The combined behavior of these genes explained the known variation of the cell cycle progression.

4.1 Phenotype I: the cell cycle S progression decreases when EWS-FLI1 is inhibited

The cell cycle S-phase progression node ('ccS') receives nine influences in the EWS-FLI1 signaling network. We may shortly describe them as follows:

- Three influences are issued from the competition between proteins E2F1,2,3 and RB1. RB1 is a member of the pocket protein family. When it is active, it sequestrates E2F1,2,3 and thus prevents the E2F1,2,3 normally transcription of genes important to the S-phase progression [25].
- One influence is triggered by the CCNE-CDK2 active protein-complex [24].
- Three influences are coming from E2F6,7,8, which have been reported to inactivate the transcription of genes responsible for the S-phase progression [30].
- Two influences are triggered from the complex formed by proteins E2F4,5 with pocket proteins RBL2,1 respectively [25], [24].

Since the priority order of the 'ccS' regulators is not established, the 'ccS' node was modeled using the generic function as GEN_{ccS}. However, experiments show that 'ccS' is inhibited when EWS-FLI1 is. In addition, 54 mRNA nodes were measured to change significantly due to EWS-FLI1 inhibition.

Our objective was to explain an important effect of the EWS-FLI1 inhibition: the decreasing in cell cycle S-phase transition. In symbolic language, this can be viewed as extracting the sub-graph linking observed mRNA nodes with the 'ccS' node coded as '–'. For this purpose we used the Algorithm 1, which was presented in Section 2.3. The computation time of this analysis was of 86 s[2]. The results of this analysis showed that three of the nine influences that 'ccS' receives were '–', three were '+' (thus contrary to the 'ccS' observation), and three were not fixed, that is, they may be '+' or '–' without representing a contradiction between the model and the dataset.

In Fig. 2 we describe the 3 inhibited pathways when EWS-FLI1 is inhibited. Notice that there is not a direct connection between EWS-FLI1 and the 'ccS' inhibition. Hence, the propagation of EWS-FLI1 inhibition can be tracked down. Its effect on the cell cycle S transition is mediated via the negative variation of three nodes of the network: $E2F2^{mRNA}$, $E2F3^{mRNA}$, and $CCND3^{mRNA}$.

2. All CPU times indicated in this paper were obtained on a Linux PC equipped with Intel Core2 2.16GHZ processor and 2GB of main memory.

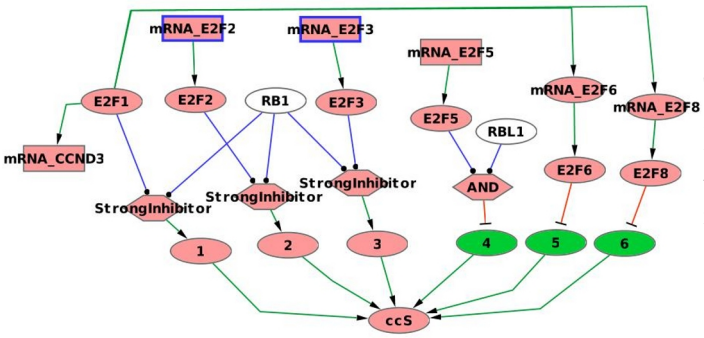

Fig. 2. EWS-FLI1 network sub-graph describing six influences over the cell cycle node ('ccS'). Arrows ending in '–>' or '–|' refer to activations or inhibitions, green/red nodes represent up-/down-regulated products, and octagonal shaped nodes are those modeled using non-generic rules. The depicted 'ccS' influences (nodes 1-6) were predicted during the EWS-FLI1 inhibition; the observations in the dataset causing these fixed influences were tracked down (squared shaped nodes). Of them, two did not have predecessors in the graph (blue bordered nodes).

Interestingly, the first two nodes do not receive other influences. The three influences that contradict the sense of variation of the S-phase progression are also depicted in Fig. 2. One of them comes from the protein-complex formed by E2F5 and RBL1, which is reviewed to inactivate the S-phase progression in [25]. In this review they also illustrate a case were E2F5 -/- Mefs mutant cells show cell cycle arrest, even after the E2F5-RBL1 complex is not formed. We arrived to a similar conclusion since the influence coming from E2F5-RBL1 complex does not impact the arrest of the cell cycle progression. The other two influences come from E2F6 and E2F8. Literature is still elusive on the role of those genes. In the absence of further knowledge, these results suggest that in the case of A673 cell lines those influences are not enough to promote the cell cycle progression.

The conclusion of this first analysis is that E2F2 and E2F3 transcriptions may be regulated directly or indirectly by EWS-FLI1. This regulation may be enough to stop A673 cells in G1 phase.

4.2 Phenotype II: the cell cycle S progression activates when EWS-FLI1 reactivates

We showed that the specific rules introduced into the model were able to explain the 'ccS' inactivation when EWS-FLI1 was inhibited. However, when applying Algorithm 1 to the second dataset (EWS-FLI1 reactivation), the cell cycle progression ('ccS': '+') was not explained by any observation in the dataset. In fact, when creating the temporal layer of nodes that represent the nine influences arriving to the 'ccS' node, two nodes (influences) were predicted to '–'. The rest of the temporal nodes over 'ccS' were not predicted.

To overcome this problem, we applied Algorithm 2 (see Section 2.3) to the EWS-FLI1 reactivation dataset. The computation time was of 103 s and the outputted sub-graph is shown in Fig. 3. The new nodes predicted to have a $\{+,-\}$ variation in order to explain the '+' variation of the 'ccS' node are shown in Fig. 3 colored in dark-green/red.

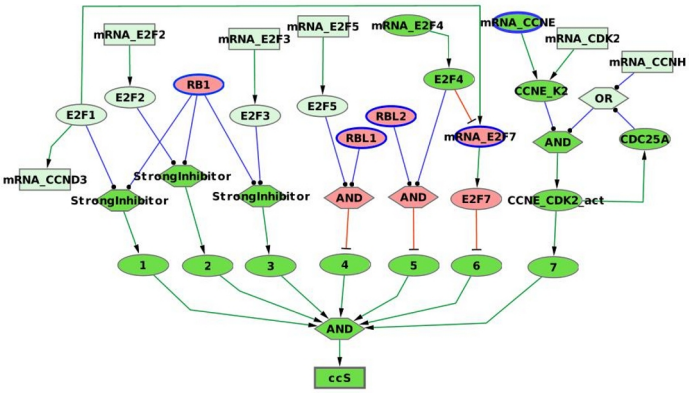

Fig. 3. Predicted $\{+,-\}$ variations (green/red colors) in order to provide 7 positive influences over the 'ccS' node. Arrows ending in '–>' or '–|' refer to activations or inhibitions, octagonal shaped nodes are those modeled using non-generic rules. The impact from the dataset of observations (square bordered nodes) on the pathways controlling the 'ccS' node is given by the light-green colored nodes. This impact cannot predict as '+' any of the 7 influences that the 'ccS' receives. By adding the AND_{ccS} function into the model, new predictions (red/dark-green nodes), consistent with the dataset, were generated. These new predictions justified the '+' change of the 'ccS' node. A blue bordered node was enough to explain at least one positive influence over the 'ccS' (nodes 1-7 fixed to '+').

We extracted a subset of these new predictions. Each node in this subset was enough to explain at least one positive influence over 'ccS'. These nodes were: (RB1: –), (RBL1,2: –), ($E2F7^{mRNA}$: –), and ($CCNE^{mRNA}$: +). As a further step, we applied Algorithm 1 to these nodes to check whether its $\{+,-\}$ predicted value could be explained by an observed value of its predecessors. These results are shown in Table 4.

From this last analysis we conclude that the pocket family proteins (RB1, RBL1, RBL2) should be inactivated post-translationally in order to explain the 'ccS' node as activated (+). The reason is that the mRNA of RBL1 is observed as '+' in the dataset, while the mRNAs of the other pocket proteins are predicted to be up-regulated. This means that during the EWS-FLI1 reactivation the mRNA of the pocket proteins increases. However, the active value of the pocket family proteins should be '–' in order to explain the cell cycle S-phase progression. Thus, we conclude that these proteins are inactivated due to post-translational phosphorylations coming from any of the products listed in the column "Origin E" of

TABLE 4
Initially unobserved/unpredicted products, that if fixed to a specific sign (in parenthesis), explain at least one positive influence over 'ccS'. The upper 'e' in the name of a product specifies its mRNA variation. **NR** is the number of regulators of the product, **E vs. O** shows how many regulators explain the sign of the product (E) and how many the opposite sign (O), and **Origin E/O** is the set of mRNA observations in the dataset that explain/contradict the sign of the product. When only one origin is found, we list in addition the name of the regulator and its effect over the product.

Product	NR	E vs. O	Origin E	Origin O
RB1 (–)	11	7 vs 0	CCND1,2,3 (+) PRKCB (+) CDK4 (+) CCNH (+)	
RBL1,2 (–)	10	6 vs 0	CCND1,2,3 (+) CDK4 (+) CCNH (+)	
$E2F7^{e}$ (–)	2	1 vs 1	E2F4 (+) E2F4 -|	CCND3 (+) E2F1 ->
$CCNE^{e}$ (+)	7	5 vs 1	E2F2,3 (+) CCND2,3 (+) EWS-FLI1 (+)	CCND3 (+) E2F8 -|

Table 4.

Another interesting result is the influence coming from $CCNE^{mRNA}$, which has 7 regulators, 5 of them providing an explanation for a '**+**' change of it, including EWS-FLI1. Only one of its regulators, E2F8, contradicts this change. In Fig. 2 E2F8 also contradicted the 'ccS' expected behavior. Finally, $E2F7^{mRNA}$ illustrates a clear example of priority order. In order to provide a positive influence over the 'ccS' node, $E2F7^{mRNA}$ needs to be repressed by E2F4. The second influence $E2F7^{mRNA}$ receives, coming from E2F1, should be absent or not strong enough.

5 Conclusion

In this paper we conceived and tested new qualitative functions to model the effect of complex regulatory influences on steady state changes in a regulatory network. Using logical functions built with the $\vee$ and $\wedge$ operators, we modeled the global response to a stress of various macromolecular interactions such as competitions, sequestration and releasing, and complex inactivation-reactivation. We show, in comparison with a model that does not implement such functions, that the prediction number is significantly increased – both on real experimental data and on random generated datasets. Experimentalists can have more confidence in those predictions: using logical functions decreases significantly the probability to obtain them from random datasets. Moreover, our predictions concern to active proteins, difficult to be widely measured by high-throughput methodologies.

We proposed a methodology that exploits those logical functions to study the impact of targeted manipulations on key nodes in the network. Biologically, these *in silico* manipulations can simulate the effect of a manipulation on target genes – using inhibitors (drugs or sh-RNA) or enhancers (*e.g.* strong promoters). Doing this, we assume that the globally observed variations on the network nodes remain stable when one manipulates specific molecules. With the proposed methodology one can work on relatively large networks (hundreds of products), for which one cannot manage to reason in an intuitive way.

Our method generates new hypotheses that suggest new experiments. First, on the experimental dataset, we were able to make predictions on the behavior of specific nodes such as E2F5, IER3, or the PDGF pathway (see our supplementary Webpage [23]). These predictions should be confirmed experimentally. Second, the *in silico* manipulation of nodes influencing the cell cycle transition to S-phase suggests that E2F2,3 are direct or indirect targets of EWS-FLI1; also, a list of variations of the network nodes is predicted to be at the origin of the cell cycle reactivation. Our analysis points out different network nodes important in the cell cycle S-phase transition, which could be interesting targets to biologists in order to better understand the correlation between EWS-FLI1 and the cell cycle.

In the long term, we aim at improving the model by adding new interactions. It may be possible that these new interactions obey to classes of rules that were not modeled in our approach, but that can be approached similarly to post-translational interactions. It is the case of post-transcriptional interactions such as alternative splicing. An important issue will be to discover new Boolean functions that could be automatically extracted from biological networks. From this perspective, this work contributes to automatically confront biological knowledge on molecular interactions with experimental data.

Acknowledgement

The authors would like to thank M. Le Borgne for his invaluable work in implementing the BioQualiV2 Python library. The authors would also like to thank the two following groups at Institute Curie, for providing data and biological insights: *Genetics and biology of pediatric tumors and sporadic breast cancers* (leaded by O. Delattre) and *Bioinformatics, biostatistics, epidemiology and computational systems biology of cancer* (leaded by E. Barillot). The work of the first author was funded by CONICYT (Comisión Nacional de Investigación Científica y Tecnológica), Chile. This work was supported by ANR-06-BYOS-0004 SITCON project.

References

[1] S. A. Kauffman, C. Peterson, B. Samuelsson, and C. Troein, "Metabolic stability and epigenesis in randomly constructed genetic nets," *Journal of Theoretical Biology*, vol. 22, pp. 437–467, 1969.

[2] R. Thomas, "Regulatory networks seen as asynchronous automata : a logical description," *J. Theor. Biol.*, vol. 153, pp. 1–23, 1991.

[3] S. Bornholdt, "Boolean network models of cellular regulation: prospects and limitations," *J R Soc Interface*, vol. 5 Suppl 1, pp. 85–94, Aug 2008.

[4] S. Kauffman, *The origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, 1993.

[5] S. Kauffman, C. Peterson, B. Samuelsson, and C. Troein, "Random boolean network models and the yeast transcriptional network," *PNAS*, vol. 100, no. 25, pp. 14 796–14 799, 2003.

[6] R. M. Gutierrez-Rios, D. A. Rosenblueth, J. A. Loza, A. M. Huerta, J. D. Glasner, F. R. Blattner, and J. Collado-Vides, "Regulatory network of Escherichia coli: consistency between literature knowledge and microarray profiles," *Genome Res*, vol. 13, no. 11, pp. 2435–2443, 2003.

[7] M. J. Herrgard, B.-S. Lee, V. Portnoy, and B. O. Palsson, "Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in Saccharomyces cerevisiae." *Genome Res*, vol. 16, no. 5, pp. 627–35, 2006.

[8] M. W. Covert, N. Xiao, T. Chen, and J. Karr, "Integrating metabolic, transcriptional regulatory and signal transduction models in Escherichia coli," *Bioinformatics*, vol. 18, pp. 2044–50, 2008.

[9] J. Baumbach and L. Apeltsin, "Linking cytoscape and the corynebacterial reference database coryneregnet," *BMC Genomics*, vol. 9, p. 184, 2008.

[10] T. Ideker, V. Thorsson, J. A. Ranish, R. Christmas, J. Buhler, J. K. Eng, R. Bumgarner, D. R. Goodlett, R. Aebersold, and L. Hood, "Integrated Genomic and Proteomic Analyses of a Systematically Perturbed Metabolic Network," *Science*, vol. 292, no. 5518, pp. 929–934, 2001. [Online]. Available: http://www.sciencemag.org/cgi/content/abstract/292/5518/929

[11] O. Radulescu, S. Lagarrigue, A. Siegel, P. Veber, and M. L. Borgne, "Topology and static response of interaction networks in molecular biology." *J R Soc Interface*, vol. 3, no. 6, pp. 185–96, 2006.

[12] A. Siegel, O. Radulescu, M. L. Borgne, P. Veber, J. Ouy, and S. Lagarrigue, "Qualitative analysis of the relation between DNA microarray data and behavioral models of regulation networks." *Biosystems*, vol. 84, no. 2, pp. 153–74, 2006.

[13] P. Veber, C. Guziolowski, M. Le Borgne, O. Radulescu, and A. Siegel, "Inferring the role of transcription factors in regulatory networks," *BMC Bioinformatics*, vol. 9, p. 228, 2008.

[14] B. DasGupta, G. A. Enciso, E. Sontag, and Y. Zhang, "Algorithmic and complexity results for decompositions of biological networks into monotone subsystems," *BioSystems*, vol. 90, pp. 161–178, 2007.

[15] C. Guziolowski, J. Gruel, O. Radulescu, and A. Siegel, "Curating a large-scale regulatory network by evaluating its consistency with expression datasets," in *Computational Intelligence Methods for Bioinformatics and Biostatistics - Selected revised papers*, ser. Lecture Notes in Computer Science, vol. 5488. Salerno, Italy: Springer-Verlag, 2009, pp. 144–155.

[16] K. M. Mani, C. Lefebvre, K. Wang, W. K. Lim, K. Basso, R. Dalla-Favera, and A. Califano, "A systems biology approach to prediction of oncogenes and molecular perturbation targets in B-cell lymphomas," *Mol. Syst. Biol.*, vol. 4, p. 169, 2008.

[17] T. Baumuratova, D. Surdez, B. Delyon, G. Stoll, O. Delattre, O. Radulescu, and A. Siegel, "Localizing active transcriptional and post-translational influences in gene regulatory networks," *submitted*.

[18] G. Stoll, A. Zinovyev, E. Novikov, L. Martignetti, E. Barillot, F. Tirode, K. Laud, N. Guillon, D. Surdez, O. Delattre, O. Radulescu, T. Baumuratova, S. Blachon, M. L. Borgne, P. Veber, C. Guziolowski, and A. Siegel, "Systems biology sitcon project for studying ewing sarcoma," in *International Conference on Systems Biology*, Goteborg, Sweden, 2008.

[19] R. Gutiérrez, L. Lejay, A. Dean, F. Chiaromonte, D. Shasha, and G. Coruzzi, "Qualitative network models and genome-wide expression data define carbon/nitrogen-responsive molecular machines in Arabidopsis," *Genome Biology*, vol. 8, no. 1, p. R7, 2007.

[20] J. Dormoy, "Controlling qualitative resolution," in *7th National Conference on Artificial Intelligence, AAAI'88*, Saint Paul, Min., USA, 1988.

[21] P. Veber, M. L. Borgne, A. Siegel, and O. Radulescu, "Complex qualitative models in biology: A new approach," *Complexus*, vol. 2, pp. 3–4, 2004/2005.

[22] M. L. Borgne, "Solving losely coupled constraints," http://hal.inria.fr/docs/00/39/70/98/PDF/RR-6958.pdf, INRIA, Tech. Rep. RR-6958, 2009.

[23] "Supplementary webpage," www.irisa.fr/symbiose/projects/signalingNets/logicalRules.html.

[24] A. Obaya and J. Sedivy, "Regulation of cyclin-cdk activity in mammalian cells," *Cell Mol Life Sci.*, vol. 59, no. 1, pp. 126–42, 2002.

[25] J. DeGregory, "The genetics of the e2f family of transcription factors: shared functions and unique roles," *Biochim Biophys Acta.*, vol. 1602, no. 2, pp. 131–50, 2002.

[26] I. Perez-Roger, S. Kim, B. Griffiths, A. Sewing, and H. Land, "Cyclins d1 and d2 mediate myc-induced proliferation via sequestration of p27(kip1) and p21(cip1)," *EMBO J.*, vol. 18, no. 19, pp. 5310–20, 1999.

[27] Y. Moreau, S. Aerts, B. Moor, B. Strooper, and M. Dabrowski, "Comparison and meta-analysis of microarray data: from the bench to the computer desk," *TRENDS in Genetics*, vol. 19, no. 10, pp. 570–577, 2003.

[28] S. Kazmi, Y. Kim, B. Pei, N. Ravi, D. Rowe, H. Wang, A. Wong, and D. Shin, "Meta Analysis of Microarray Data Using Gene Regulation Pathways," in *Proceedings of the 2008 IEEE International Conference on Bioinformatics and Biomedicine*. IEEE Computer Society, 2008, pp. 37–42.

[29] O. Delattre, J. Zucman, B. Plougastel, C. Desmaze, T. Melot, M. Peter, H. Kovar, I. Joubert, P. de Jong, G. Rouleau, *et al.*, "Gene fusion with an ets DNA-binding domain caused by chromosome translocation in human tumours," *Nature*, vol. 359, pp. 162–165, 1992.

[30] J. DeGregori and D. Johnson, "Distinct and overlapping roles for e2f family members in transcription, proliferation and apoptosis," *Curr Mol Med*, vol. 6, no. 7, pp. 739–48, 2006.