bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Inferring intracellular signal transduction circuitry from molecular perturbation experiments

Authors:

Michelle L. Wynn[1,2], Megan Egbert[1], Nikita Consul[3], Jungsoo Chang[1], Zhi-Fen Wu[1], Sofia D. Meravjer[1], Santiago Schnell[2*]

[1]Department of Internal Medicine, Division of Hematology & Oncology and Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, MI, USA

[2]Department of Molecular & Integrative Physiology, Department of Computational Medicine & Bioinformatics, and Brehm Center for Diabetes Research, University of Michigan Medical School, Ann Arbor, Michigan, USA

[3]Columbia University College of Physicians & Surgeons, New York, New York, USA.

Corresponding author. Email: schnells@umich.edu.

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

The development of network inference methodologies that accurately predict connectivity in dysregulated pathways may enable the rational selection of patient therapies. Accurately inferring an intracellular network from data remains a very challenging problem in molecular systems biology. Living cells integrate extremely robust circuits that exhibit significant heterogeneity, but still respond to external stimuli in predictable ways. This phenomenon allows us to introduce a network inference methodology that integrates measurements of protein activation from perturbation experiments. The methodology relies on logic-based networks to provide a predictive approximation of the transfer of signals in a network. The approach presented was validated *in silico* with a set of test networks and applied to investigate the epidermal growth factor receptor signaling of a breast epithelial cell line, MFC10A. In our analysis, we predict the potential signaling circuitry most likely responsible for the experimental readouts of several proteins in the mitogen activated protein kinase and phosphatidylinositol-3 kinase pathways. The approach can also be used to identify additional necessary perturbation experiments to distinguish between a set of possible candidate networks.

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

It is increasingly likely that the large number of off-target effects associated with targeted therapeutics is due to the complex interactions and emergent behaviors intrinsic to dysregulated signaling pathways in cells. In order to rationally develop precise therapeutic avenues to target key dysregulated pathways in diseased cells, it is critical to develop methodologies that can help us understand how complex intracellular signaling pathways are wired as an integrated network within both normal and aberrant cells. While targeted molecular inhibitors hold great promise for controlling disease, in the clinic these drugs are often not as successful as their pre-clinical data indicated.

The pathways associated with epidermal growth factor (EGF) receptor signaling are of great clinical interest. EGF receptors can activate both the mitogen activated protein kinase (MAPK) pathway, which promotes motility, invasion, and angiogenic factor production, and the phosphatidylinositol-3 kinase (PI3K) pathway, which controls anchorage independent growth and modulates glucose metabolism (Plas and Thompson 2005; Robey and Hay 2009; van Golen et al. 2002; Zhang et al. 2003). Both the MAPK and PI3K pathways are highly dysregulated in cancer (Halilovic et al. 2010; McCubrey et al. 2006; Meier et al. 2005; Rommel et al. 1999; Won et al. 2012), neuroinflammation (Liu et al. 2014), obesity and type 2 diabetes (Bernal-Mizrachi et al. 2014; Kulkarni et al. 2012; Schultze et al. 2012), and developmental disorders (Hong et al. 2008; Nie and Chang 2007). Therapeutic targeting of the MAPK pathway has been less effective than hoped (McCormick 2011; Won et al. 2012), and several regulatory mechanisms have been suggested to explain the apparent cross-talk between the MAPK and PI3K pathways (Serra et al. 2011). Understanding precisely how these networks are wired in normal and diseased cells will have tremendous clinical significance.

Boolean logic can be used to simulate signal transfer in a molecular and biological network in a manner similar to signal transfer in a digital circuit. Boolean logic models of signaling networks approximate the sigmoidal regulation of a target molecule by an activator or inhibitor (Wynn et al. 2012). They can also be used to explicitly simulate enzyme activation and inhibition in a set of biochemical reactions as well as bistability and chemical hysteresis (Arkin and Ross 1994; Hjelmfelt and Ross 1995;

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Hjelmfelt et al. 1993). Logic-based models are powerful tools for studying complex systems, such as intracellular molecular networks, because they are qualitatively predictive and do not require parameter information or mechanistic details needed for more quantitatively precise kinetic methods (Albert et al. 2008; Glass and Kauffman 1973; Thomas 2006; Wynn et al. 2012). Dynamic signaling models based on logic networks have high predictive power only if the underlying logic network is accurately constructed. Of course, this is also true of highly quantitative kinetic-based differential equation models, which require knowledge of the underlying circuitry as well as a very large number of unknown parameters. While treating protein activation levels in binary terms of ***ON*** and ***OFF*** is a simplification of the complexity of molecular function within a cell, it is important to emphasize what ***ON*** and ***OFF*** values signify in a logic-based signaling model. Specifically, when a node is in an ***ON*** or ***OFF*** state, the molecule is assumed to be above or below, respectively, the threshold needed to induce an appreciable effect in the molecule or molecules it regulates (Wynn et al. 2012). This threshold assumption is similar to the continuous concentration approach used by Ross and colleagues (Arkin and Ross 1994; Hjelmfelt and Ross 1995; Hjelmfelt et al. 1993), where changes in concentrations are abrupt and switch from low to high concentration without an effective stable intermediate at steady-state.

A literature and/or curated database search is critical for developing an accurate logic network, but will never be sufficient to identify the underlying circuitry of an aberrant cell given the high number of genetic defects that may be present in such cells. Thus, we wondered if it was possible to infer the actual underlying network in a population of cells from experimental data routinely collected in the laboratory. To this end, we developed an approach to infer the signaling network circuitry inside a cell from experimentally measured western blot data using asynchronous Boolean logic simulation methods and a heuristic search with genetic algorithms. The objective of the approach is to identify the most likely underlying structure of a signaling network by minimizing the difference between simulated model output and experimental data collected from a series of molecular perturbation experiments. We applied the method to experimental data collected from a cell line and validated the general approach with a series of *in silico* tests, where the underlying network was known a priori.

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The ability to accurately infer an intracellular network from data remains a significant and difficult problem in molecular systems biology as well as a major obstacle to signaling based combination therapies. A variety of reverse engineering approaches have been developed (Crampin et al. 2004; Kholodenko et al. 2002; Stelniec-Klotz et al. 2012). When reverse engineering a gene, signaling, or metabolic network via Boolean approaches, it is necessary to identify a logical interaction between each molecule in the network, such that the observed experimental discretized data can be explained by the model. While algorithms exist to infer Boolean networks from experimental data, including REVEAL (Liang et al. 1998), the Akutsu algorithm (Akutsu et al. 1999), and ReBMM (Saeed et al. 2012), each differs in the approach used to identify a network. REVEAL uses an information criterion to reduce the network search space consistent with the observable data. The Akutsu algorithm uses an enumerating approach to find the networks consistent with observable data. ReBMM uses a sparse search algorithm based on a Bernoulli mixture model, which is used to determine the set of Boolean networks consistent with the observable data. These and similar algorithms were largely designed to infer gene regulatory logic networks from gene expression data and assume that Boolean networks are subject to deterministic synchronous updating.

In this paper, our objective is different: we are interested in the accurate inference of the underlying structure of a signaling network from experimental data of protein activation. We also rely on a stochastic and asynchronous logic method to simulate a population level response. We use a heuristic search with genetic algorithms to identify candidate logic networks consistent with observed data collected from a set of perturbation experiments. Our approach is applied to MCF10A cells and its validity is explored with a series of *in silico* network tests.

Materials and Methods

Here we provide a detailed description of the methods used as part of our network inference approach based on both *in vitro* and *in silico* perturbation experiments.

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Western blot analysis

MCF10A cells were serum starved overnight in a 1:1 mix of Ham's F12 and Dulbecco's modified Eagle medium (DMEM) with 1x Pen-Strep-Glutamine, 2.5 µg/mL fungizone, 5 µg/mL gentamycin, 10µg/mL insulin, 0.5µg/mL hydrocortisone, 0.02 µg/mL EGF from murine submaxillary gland (Sigma)**,** and 0.1µg/mL cholera toxin. In the morning, the media was changed to a 1:1 mix of Ham's F12 and DMEM containing no serum and no supplements, and cells were incubated for 60 min with or without 50 µM of LY294002 (Cell Signaling), a PI3K inhibitor, or 1 µM of PD0352901 (Cell Signaling), a MEK inhibitor, followed by treatment with or without 0.01 µg/mL EGF human recombinant protein (Millipore Cat. 324831) for 10 minutes. After 70 minutes, cells were harvested for protein extraction. Anti-bodies against phospho-AKT (S473 and T308), AKT, phospho-PTEN (S380), PTEN, phospho-ERK (T202/T204), ERK, phospho-Raptor (S792), Raptor, phospho-mTOR (S2448), mTOR, phospho-TSC2 (T1462) and TSC2 were obtained from Cell Signaling.

Logic Network Simulations

All logic network simulations used an asynchronous random order update scheme with *n* replicate networks. Typically, *n*=100, unless otherwise stated. Booleannet (Albert et al. 2008) was used for all asynchronous simulations. Because a random-order asynchronous update samples many different timescales (in contrast to the uniform timescales and deterministic outcomes of synchronous update methods), it is appropriate for simulating the heterogeneity in a population of cells (Chaves et al. 2005; Garg et al. 2008; Thomas 1973; Wynn et al. 2012).

Simulations ran for a pre-defined number of time steps (long enough for a logical steady state, also known as an attractor (Thomas and D'Ari 1990), to be reached). The steady state value of all nodes can be determined explicitly in a logic model if the initial values of all nodes in the network are known *a priori*. Because it will rarely be practical to measure all molecules in realistic biological networks, our simulation approach requires only knowledge of the initial activation state of nodes explicitly perturbed in the experiment simulated (e.g., constitutively ***ON*** or ***OFF*** proteins). All nodes not given an initial value at the start of a simulation are assigned random initial values in each of the

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

n replicate networks. The output of an asynchronous simulation (with *n* replicate networks) is a value between 0.0 and 1.0 for each node, indicating the probability a node is activated (***ON***) in the steady state. These probabilities are calculated directly from the average state of each node in *n* networks (Albert et al. 2008). As a direct consequence, steady state node probabilities greater than 0.0 but less than 1.0 indicate that, for the conditions tested, (a) the node is oscillating as part of a limit cycle attractor, (b) the node can be in one of two fixed point attractors, or (c) the node can be in both a limit cycle and a fixed point attractor.

Data Discretization

For the Western blots, we chose to use data from a normal cell line (MCF10A) that had unambiguous western blot signals (that is, it was clear upon visual inspection whether a node was ***ON*** or ***OFF***). Thus, it was not necessary to explicitly define a threshold when discretizing our experimental data (referred to as expected data). However, to apply a threshold parameter when discretizing western blot data where the binary state of a measured protein is not readily inferred, semi-quantitative densitometric methods are often used.

For the simulated data, we did not consider a value of 0.4 or 0.6 to be qualitatively different from 0.5 (the value expected when exactly half of the simulated networks for a given set of conditions were ***ON*** and exactly half were ***OFF***). Thus, this range served as the bin for the implicit threshold of 0.5. Because our simulated values were continuous probabilities that a node was ***ON*** or ***OFF***, we also decided to define a bin for 25% ***ON*** and 75% ***ON***. While in some ways these ranges are arbitrary, empirically we found that this helped our simulations converge to optimal solutions (when a “true” target network was known), suggesting that the approached helped smooth the stochastic nature of these probability values.

More details about the data discretization can be found in the Supplementary Materials of Wynn et al. (2014).

Predictive Scores

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Comparison of model readout with experimental data necessitated the discretization of experimental western blot data. It is important to note that because a phospho western blot is a qualitative assay dependent on many factors (including film exposure time), attempts at precise quantitation may be misleading and may not be necessary for the modeling approach used here. Values of ***ON*** or ***OFF*** were assigned to indicate the phosphorylation state of proteins in MCF10A cells for each condition tested. In general, faint or absent blots were assigned ***OFF*** values (0.0) and prominent blots were assigned ***ON*** values (1.0).

A predictive score (p-score) was developed to quantify how well simulated model output matched experimentally collected data (referred to as expected values). The p-score is calculated as:

$$\mathrm{p-score} = \sum_{i=1}^{N}\sum_{j=1}^{T} \left| Bin\big(SV(i,j)\big) - EV(i,j) \right| \left(\frac{1}{N \times T}\right)$$

where, *i* is the node identifier, *j* is the test condition identifier, *N* is the total number of nodes evaluated in the network, *T* is the total number of test conditions, *SV(i, j)* is the simulated value of the *i*-th node in the *j*-th test condition, *EV(i, j)* is the expected value of the *i*-th node in the *j*-th test condition. The function *Bin(x)*, partitions simulated values as follows (input → output):

1. $x = 0.00 \rightarrow 0.00$
2. $0.00 < x < 0.40 \rightarrow 0.25$
3. $0.40 \leq x \leq 0.60 \rightarrow 0.50$
4. $0.60 < x < 1.00 \rightarrow 0.75$
5. $x = 1.00 \rightarrow 1.00$

The *N* nodes evaluated in the p-score include only those nodes for which expected values were available. Consequently, *N* may be smaller than the total network node count. The scaled p-score will range from 0.0 (indicating a perfect match with experimental data) to 1.00 (indicating that the simulated output was wrong for all evaluated nodes in all test conditions).

Genetic Algorithm and Fitness Function

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

An implementation of a genetic algorithm (GA) was used to identify candidate logic networks from a large search space of possible logic networks. A GA is a type of heuristic search that is often used to find optimal solutions in very large search spaces that cannot reasonably be searched exhaustively. GAs leverage operations that mimic aspects of biological evolution to identify an optimal solution, such as fitness selection, reproduction, chromosomal crossover, and random mutation (Crampin et al. 2004; Deb 1999; Mitchell 1996). While GA operations may appear simple, the search performed is considered "highly nonlinear, massively multi-faceted, stochastic, and complex" (Deb 1999). Typically, the problem to be optimized is numerically encoded via vectors of numbers referred to as chromosomes. A fundamental component of a GA is a fitness function that assigns a numerical value based on the relative fitness of each chromosome. A GA begins with an initial population of chromosomes and proceeds by allowing the population to evolve over many generations.

We chose to use a GA because the search space in which we were looking for an optimal network configuration was often far too large to search exhaustively. In our case, the GA attempts to minimize the difference between experimental input data and simulated output of candidate logic networks under the same perturbation conditions used to collect the experimental data (**Figure 1**). The output of the fitness function used is a p-score. Our GA implementation does not stop when an optimal solution is found (i.e., a network with a p-score of 0.0) because we want the population to continue to evolve to other local optima, if they exist. The general steps followed in our GA implementation are:

1. In generation 0, randomly initialize a population of *p* chromosomes as a numeric vector, where each element of the vector corresponds to a node in the network, and each numeric value corresponds to a possible logic regulation rule for the node. Each unique chromosome represents a unique numeric encoding of a candidate logic network.
2. In each generation, evaluate the chromosomes in the current population by calculating their p-scores via a fitness function. The next generation's population is determined by cross-over via random tournament selection (Mitchell 1996) and mutation rates between 0.2 and 0.5.

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

3. Report all networks with an optimal p-score of 0.0.
4. Run until the number of generations set at the start of the simulation is reached.

For very large search spaces, GA were performed three times for at least 100 generations each to ensure that adequate coverage of the large search space was achieved. If there is uncertainty about the coverage of the search space, the number of generations can be increased in our approach.

Logic network search space generation

The space of all possible logic networks searched by the GA consists of all logic network combinations that can be made for the regulations possible for each node in a network. A script was used to enumerate all logically unique ways a single node could be regulated by all or a subset of nodes in a network (depending on whether any prior knowledge of the network is available). For *m* regulatory nodes, which represent the maximal set of nodes that can possibly regulate a target node, we assume each regulatory node appears no more than one time in a logic rule. Because order matters in a logical expression, we generate all permutations of the ways 1 to *m* nodes can by linked by logic operators as well as by parenthetic groupings. Finally, we scan all generated rules to check for logical equivalency and remove any redundant expressions (that is, any expression that generates a truth table identical to the truth table of an expression already included in the set). A unique identifier is assigned to each logically unique regulation rule possible for a target node. **Table 1** lists the number of unique logic regulations possible when a node is regulated by up to 6 other nodes in the network. **Table 2** and **Supplemental Table 1** show examples of logic regulations that make up two distinct logic network search spaces.

Computing environment

A custom python library called Netinf was written to support the network inference search and implementation of the GA. The library, which makes calls to Booleannet (Albert et al. 2008) for asynchronous network simulations, supports distributed computing using MPI for Python. All distributed simulations were run on the

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

University of Michigan's high performance computing cluster using Intel Nehalem/i7 Cores.

Results

The objective of our approach presented in this study is to accurately infer the underlying structure of a signaling network from experimental data. We first consider a hypothesized network of EGF based signaling in MCF10A cells. We then evaluate our method by testing its ability to identify a series of increasingly complex test networks.

EGFR signaling in MCF10A cells

A detailed literature search was performed to identify the important regulatory components of EGF induced signaling in the PI3K and MAPK pathways (**Table 3**). From this information, a canonical normal interaction network was constructed (**Figure 2A**). Interaction networks provide directional information about a set of regulatory events, but do not provide details about the hierarchy of regulations when more than one node regulates another node (Vidal et al. 2011; Wynn et al. 2012). For example, the interaction network indicates that **PIP3** is activated by **PI3K** and inhibited by **PTEN,** but does not indicate which effect is dominant when both molecules are present. The literature supporting each regulation in the interaction network was, therefore, carefully considered before deriving a set of hypothesized logic rules (**Table 4**), which make up the hypothesized logic network of normal EGF signaling in MCF10A cells (**Figure 2B**). In general, an ***AND NOT*** regulation means the inhibitor will dominate when both an activator and inhibitor are present, while an ***OR NOT*** regulation means that the activator will dominate when both an activator and inhibitor are present [see (Wynn et al. 2012) for a more detailed discussion].

The correctness of the hypothesized logic network model (**Figure 2B**) for normal-like cells was tested by comparing *in silico* model predictions to experimentally measured readouts of phosphorylated AKT (at S473 and T308) and ERK (at T202/T204) in the normal-like mammary epithelial MCF10A cell line. Six experimental conditions were performed *in vitro* (**Figure 2C**) and *in silico* (**Figure 2D**): (1) control, (2) EGF stimulation, (3) LY294002 (LY) alone, (4) LY followed by EGF stimulation, (5)

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

PD0325901 (PD) alone, and (6) PD followed by EGF stimulation. For *in vitro* immunoblot experiments, cells were serum starved overnight and then treated with or without PD, a direct inhibitor of MEK (Leyton et al. 2008), or LY, a direct inhibitor of PI3K (Qiang et al. 2004) for 1 h, followed by 10 min with or without EGF stimulation. For *in silico* experiments, the control condition was simulated by fixing the **EGFR** node to ***OFF*** for the duration of the simulation, while the presence of EGF in test conditions was simulated by fixing the **EGFR** node to ***ON*** for the duration of the simulation. The presence of LY or PD in test conditions was simulated by fixing **PI3K** or **MEK**, respectively, to ***OFF*** for the duration of the simulation. In all six experimental conditions tested, PTEN (which is known to be wild-type in MCF10A cells) was activated (**Figure 2C**). As a consequence, the **PTEN** node was fixed to ***ON*** in all simulations of MCF10A cells.

The hypothesized logic network (**Figure 2B**) correctly predicted the phosphorylation of **AKTS473**, **AKTT308**, and **ERK** in all six experimental conditions (**Figure 2C**), producing a p-score of 0.0, which indicates an exact qualitative match between experiment and simulation (see **Methods**). The hypothesized logic model does not include inhibitory regulation of **RAF** by **AKT** or **ERK** (Brummer et al. 2003; Ritt et al. 2010; Rommel et al. 1999; Zimmermann and Moelling 1999) (as indicated by dashed red lines in **Figure 2B**). We hypothesized that the negative regulation of **RAF** by **AKT** and **ERK**, which would be the only sources of **ERK** oscillations in our model, are “late effects” (as per the term used by Samaga et al. (2009) that will take place much later than our experimental timescale of 70 minutes. A longer experiment with multiple time points would be required to effectively measure possible oscillatory changes in ERK activation. Some experiments that report rapid growth factor induced oscillations in phosphorylated ERK, report changes in the relative ratio of phospho-ERK to total-ERK (Nakayama et al. 2008; Shankaran et al. 2009). While the ratio may be changing, a robust phospho ERK signal is often still observed. When an alternate version of the network that included negative regulation of **RAF** by **AKT** and **ERK** was tested, it produced a p-score of 0.08 (see **Supplemental Figure 1**).

These results suggest that the hypothesized model (**Figure 2B**) may represent the underlying circuitry driving EGF based signaling in the PI3K and MAPK pathways in

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

MCF10A cells, but do not definitively establish this to be the case. In an attempt to confirm or refute the proposed circuitry, we wondered how many other logical configurations of the nodes in the hypothesized logic network model could generate a p-score of 0.0. While an exhaustive search of all logic network configurations was not practical for a network of this size (**Table 1**), an optimized search via a GA (see **Methods**) found hundreds of alternate logic network configurations that can also generate a p-score of 0.0 based on the expected values of **AKTS473**, **AKTT308**, and **ERK** (**Figure 3C**). A subset of these networks can be discarded because the regulations involved do not represent what is expected in a normal model of the pathways considered. For example, all networks with a p-score of 0.0 that include **PTEN** activating (rather than inhibiting) **PIP3**'s role in the phosphorylation of **AKTT308** could be rejected. It may be imprudent to make this assumption, however, if we were attempting to uncover the signaling network inside a cancer or aberrant cell with unknown signaling adaptations.

To test if including experimental measurements for all or most nodes in the hypothesized network will narrow the list of candidate networks, we constructed a smaller 9-node hypothesized logic network (**Figure 3A**) where **PIP3** was the only node not directly measured or perturbed experimentally. While this version of the logic network also generated a p-score of 0.0 (**Figure 3B**), it is still not clear that the reduced network is an accurate representation of the underlying EGF based signaling circuitry of MCF10A cells. If we know which nodes *can* regulate other nodes in the network with a high degree of accuracy, but do not know precisely how these regulations occur, it is possible to reduce the complexity of the search space in which to look for other networks that may have a p-score of 0.0. First, **EGFR** and **PTEN** can be assumed to be input nodes (which means no other node in the network directly regulates them). This assumption is justified for PTEN because the phosphorylation of PTEN is unchanged in the 6 experimental conditions tested (**Figure 2C**). From the interaction network (**Figure 2A**), it can be assumed that **MEK** is regulated by **EGFR**, **ERK** is regulated by **MEK**, **PI3K** is regulated by **EGFR**, **PIP3** is regulated by **PI3K** and/or **PTEN**, and **AKT** is regulated by some combination of the phosphorylation of its T308 and S473 residues (which in turn can be assumed regulated by **PIP3** and **PI3K**, respectively).

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

If we assume only that the above regulations are possible, but assume nothing about the exact nature of the regulations, then a total of 36 logic regulation rules are possible. If the possibility that nodes other than **EGFR** and **PTEN** could be input nodes is included (in the case of **AKTS473** and **AKTT308** this would allow for the possibility of auto-phosphorylation), then 43 distinct logic regulation rules are possible (**Table 2**), which combine to form a search space of 41,067 distinct logic networks. p-scores were calculated after simulating all 41,067 logic networks, and 78 networks were found with a p-score of 0.0, leaving additional room for doubt about whether our hypothesized network (**Figure 3A**) can be considered correct.

Returning to the original hypothesized interaction network (**Figure 2A**), there are 3^{210} ($3^{n(n-1)}$, n = 15) possible interaction networks that can be constructed from the 15 nodes in this network and many more logic networks. If we make assumptions similar to those we made for the reduced network, there are a total of 275 possible logic rules that can regulate the 15 nodes, producing a very large search space of **2.4×10^{12}** logic networks (**Supplemental Table 1**). Obtaining additional expected values by performing additional perturbation experiments will likely improve the probability of finding the network that best reflects the network circuitry in the cell. Because laboratory experiments are resource intensive, it would be advantageous to be able to predict the next set of experiments that will best inform the network search. To this end, we designed a series of *in silico* experiments to test whether it is possible to infer the underlying signaling network from experimental data using a series of increasingly complex *in silico networks*. We also tested whether it is possible to reasonably predict the next best set of experiments to perform in the lab to facilitate discovery of the network that best represents the underlying signaling circuitry of a given cell type.

Validation of methodology in silico

We evaluated our network inference methodology with a series of test networks. We assume that between any two nodes in a logical circuit there can be, at most, two directional edges (depicted as a black dashed line in **Figure 4A**), where each directional edge can be in one of three regulatory states: (0) absent with no regulation, (1) inhibiting the target node, or (2) activating the target node. In a two-node network the

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

number of interaction networks and logic networks are equivalent (because there are no ***AND*** or ***OR*** operators possible). For $n > 2$, however, the number of logic networks grows faster than the number of possible interaction networks (**Table 1**).

The objective of the *in silico* tests performed was to recover the correct network from a large search space of networks (that is, to find a single network with a p-score of 0.0) using only the expected values generated by the correct network from a set of *in silico* perturbation experiments. The general approach followed was:

1. For all nodes in a test network known a priori, create a network search space by generating the set of all possible logic rules that may regulate each node (this set may be reduced by prior knowledge).
2. Identify a set of network perturbation experiments to perform *in silico*.
3. Run simulations of the test network under each perturbation condition to generate expected values.
4. Search for logic network configurations that can produce the same expected values as the target test network did in step 3, under identical perturbation conditions (i.e., p-score = 0.0). For large search spaces, the search was optimized using a GA, where the fitness function generates a p-score (see **Methods**) for each candidate network in the population of each generation.
5. If multiple networks with a p-score of 0.0 are found, attempt to reduce this number by predicting additional perturbation experiments that can differentiate between these networks. Repeat step 3.

Two-node test network

In a two-node network consisting of **A** and **B**, a total of 9 possible logic networks are possible. These networks can be uniquely identified in vector form as [$\mathbf{A}_i$, $\mathbf{B}_i$], where $\mathbf{A}_i$ and $\mathbf{B}_i$ correspond to one of the three logic rules possible for each node (**Figure 4A**). In this very simple example, the number of interaction and logic networks are equivalent (**Table 1**). For each of the 9 logic networks that can be formed from **A** and **B,** expected values were generated *in silico* with the following perturbations: (1) **A *ON***, (2) **A *OFF***, (3) **B *ON***, and (4) **B *OFF*** (**Supplemental Figure 2**). The *in silico* data were used as expected values analogous to how the *in vitro* western blot data was used as expected

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

values for the MCF10A cells (**Figure 2**). For each of the 9 sets of expected values, we ran a search that attempted to recover each of the 9 original networks. Because the search space is small, it was straightforward to perform an exhaustive search. As an example, when network [**1**, **2**] was the target search network (**Figure 5**), a p-score was generated for all 9 candidate networks. In all 9 searches, the correct network was the only network with a p-score of 0.0 and, thus, the correct network was always recovered based on the expected values generated by perturbing the network.

Three-node test network

The addition of a single node increases the size of the potential search space considerably. In a 3-node circuit consisting of **A**, **B**, and **C** (**Figure 4B**), there are 729 interactions and 2,197 logical networks possible (**Table 1;** and **Supplemental Table 2**). Expected values were generated *in silico* for 3 test networks: **[9, 1, 0]**, **[1, 3, 11]**, and **[3, 0, 10],** where each network is encoded as **[A_i, B_i, C_i]**, and *i* indicates a unique logic rule identifier possible for a given node (**Figure 4B**). Searches for the 3 three-node networks were run using both a GA and an exhaustive search (where all possible 2,197 logic networks were simulated). In all cases, a p-score was calculated for each candidate network tested. Any network with a p-score of 0.0 was reported as a possible network solution.

Three separate sets of *in silico* perturbation experiments (**Table 5**) were used to generate expected values for the three-node networks tested. The correct network was recovered from the expected values of all 3 sets of *in silico* perturbation tests for all 3 test network searches. Perturbation test sets I and II, however, produced non-unique solutions (with many other networks producing a p-score of 0.0). In contrast, perturbation test set III uniquely identified the correct network in all 3 target network searches (**Table 5**).

The three-node network searches performed via a GA were initialized with a population size of 100 and were allowed to run for 20 generations, requiring up to 2,000 simulations for each network search. While 2,000 is only slightly smaller than the 2,197 simulations needed to perform an exhaustive search, the number of simulations performed in 20 generations by the GA was always less than 2,000 because of

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

optimizations in the algorithm that prevent unnecessary repeated simulations of candidate networks maintained in multiple generations of the GA. In all cases, the GA was more efficient at recovering the target network (**Table 6**). Using perturbation test set III, the correct network was found on average in 6 or fewer generations (**Table 6)**. The longest GA search (for target network **[3, 0, 10]**) took an average of 41.7 minutes to complete 20 generations, while an exhaustive exploration of the full search space took an average of 98 minutes, under otherwise identical computing conditions.

Additional test networks

We performed similar tests using increasingly complex networks. Refer to the supplemental material for a detailed discussion about how we searched for a pre-defined Five and Ten Node network (**Supplemental Figure 3, Supplemental Figure 4**, respectively).

Unraveling the MCF10A network

Returning to the original hypothesized network of MCF10A signaling, we performed additional western blot experiments to obtain readouts representative of mammalian target of rapamycin complex 1 (mTORc1) signaling, which may play a fundamental role in the cross talk between the MAPK and PI3K signaling pathways (**Figure 2A-B**) (Foster and Fingar 2010; Hay 2005; Sarbassov et al. 2006; Sarbassov et al. 2005). Specifically, we measured the phosphorylation of raptor, mTOR and TSC2 under the same six experimental conditions used previously with MCF10A cells (**Figure 6A**). From these data, we concluded that mTORc1 is activated in all 6 experimental conditions, while the TSC1/2 dimer is inactive in all 6 experimental conditions. The activation of mTORc1 was inferred ***ON*** in all 6 conditions because a band was observed in all conditions for phospho-mTOR and phospho-raptor. While the band for phospho-mTOR was weakest in the control condition, we assume that its activity is above the threshold needed for functional activity. For simplicity, we also assumed that the other key molecular components of mTORc1 are present and in an active conformation. The activation of TSC1/2 was inferred ***OFF*** in all 6 conditions because a band representing phosphorylation to TSC2 at T1462 was observed in all 6 experimental conditions.

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Phosphorylation of TSC2 at T1462, which inactivates the molecule, is regulated by AKT. This residue was phosphorylated even in conditions where AKT was not activated, however. We therefore hypothesized that TSC2 is either constitutively deactivated in MCF10A cells or another key EGFR driven regulator of this residue is missing from our model.

When the expected values for **mTORc1** and **TSC12** activation were added to the expected values for **AKTS473**, **AKTT380**, and **ERK** (**Figure 6C**), the p-score of the original MCF10A hypothesized logic network (**Figure 2B**) increased from 0.00 to 0.10. The change in p-score reflects the difference between the simulated values of the hypothesized network and the expected value for **TSC12** activation in the control, LY, and PD conditions (**Figure 6C**). We used the GA to search for networks that are consistent with the MCF10A expected values in **Figure 6C.** First, we used the GA (generations=100, population=400) to search the informed search space described previously as consisting of 275 distinct logic rules and **2.4×10^{12}** distinct logic networks (**Supplemental Table 1**). While more than 1,000 networks with a p-score of 0.0 were found, they all included one of three regulation rules for TSC1/2, each of which violated our literature derived knowledge of AKT and/or ERK inhibition of TSC1/2 (Hahn-Windgassen et al. 2005; Hay 2005; Jiang and Liu 2008; Manning et al. 2002; Sabatini 2006). The three TSC1/2 rules contained in all networks found with a p-score of 0.0 were: (1) **$TSC12^{t+1} = AKT^{t}$**, (2) **$TSC12^{t+1} = AKT^{t}$ and not ERK^{t}**, and (3) **$TSC12^{t+1} = AKT^{t}$ and ERK^{t}**.

We next searched a larger search space of **5.7×10^{14}** possible networks where **EGFR** was added as a potential direct regulator of **mTORc1**, **S6K1**, and **TSC12**. The inclusion of **EGFR** as a possible regulator of these three nodes allowed us to test for the presence of molecules that regulate each of these nodes as a direct consequence of **EGFR** activation without including additional nodes in the network. In this search, more than 5000 networks with a p-score of 0.0 were found in the first GA run (suggesting there are likely tens of thousands of networks in this search space that produce steady state values that match the expected values). If we could measure the activation of every protein in the network under all six perturbation conditions, the number of networks consistent with our data would likely be far smaller than the very large set of

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

networks found with a p-score of 0.0. To attempt to reduce the set of candidate networks in this large search space without performing additional experiments, constraints were added to the GA for node regulations that we had high confidence in, based on the literature and our experiments. Specifically, the constraints required the following to be true:

1) ERK was only activated by **ERK^{t+1} = MEK^{t}**,
2) RAF was only activated by **RAF^{t+1} = $RAS\text{-}GTP^{t}$** ,
3) PIP3 was only activated by **$PIP3^{t+1}$ = $PI3K^{t}$ and not $PTEN^{t}$** .

In addition, rules that met any of the following criteria were removed from the search space because they violate what we expect in normal canonical signaling in these pathways (**Table 3**):

1) any potential regulation of **mTORc2** that included inhibition by **PI3K**.
2) any potential regulation of **AKT** that included inhibition by the phosphorylation of its T308 or S473 residues.
3) any potential regulation of **RAS-GTP** that included inhibition by **EGFR** or activation by **S6K1**.
4) any potential regulation of **PI3K** that included inhibition by **EGFR** or inhibition by **RAS-GTP**.

While the inclusion of these constraints reduced the effective search space by 5 orders of magnitude to 2.8 X 10^{9}, this still represents an extremely large search space. When we searched under these constraints, thousands of networks were again found with a p-score of 0.0. One of the networks found, however, was identical to the original hypothesized network of normal signaling (**Figure 2B**) in all but the **TSC12** regulation. The **TSC12** regulation in this network was **$TSC12^{t+1}$** = **$EGFR^{t}$ and (not AKT^{t} and not ERK^{t})**. The presence of this network is significant because it does not violate our literature knowledge of TSC1/2 inhibition by AKT and/or ERK, and it suggests that if the rest of our hypothesized network regulations are correct in **Figure 2B**, then an EGFR driven molecule other than AKT and ERK is also important for the regulation of TSC1/2 activation in MCF10A cells. Because of the closeness of this network to our original hypothesized network, we performed a perturbation analysis on this network (**Figure 6D**, **Supplemental Figure 5**). In the perturbation analysis, a single perturbation is made

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

to the logic network in the form of an edge deletion or an inverse regulation, and the resulting is p-score calculated. The reported perturbation response is calculated by dividing the p-score of the perturbed network by the maximum p-score found by all perturbations of a given type (e.g., node, logic operator, edge, or edge regulatory type). The most important regulations in the network were related to the regulation of mTORc1 and TSC1/2 as well as MEK activation by ERK and mTORc2 activation by PI3K.

Finally, we found that if additional constraints were added to the GA that violated literature knowledge of signaling interactions in this network, then we were ultimately able to reduce the number of networks with a p-score of 0.0 to a set that numbered in the hundreds rather than the thousands. A large number of these networks are the products of experimental uncertainty in some interactions in the network. For example, because T308 is never phosphorylated on AKT in MCF10A cells (almost certainly because of the presence of activated PTEN in these cells) (**Figure 2C**), without additional experiments that directly perturb this residue in some way, our methodology is not able to distinguish between the following two regulations for AKT activation:

$\mathbf{AKT^{t+1} = AKTS473^{t}}$

$\mathbf{AKT^{t+1} = AKTS473^{t}\ or\ AKTT308^{t}}$

Similarly, because PTEN was activated in all conditions in our experiments, the presence of active PI3K is logically unimportant (because inhibitory **PTEN** dominates activating **PI3K** in one of the rules), and thus the methodology cannot distinguish between these two regulations for PIP3:

$\mathbf{PIP3^{t+1} = not\ PTEN^{t}}$

$\mathbf{PIP3^{t+1} = PI3K^{t}\ and\ not\ PTEN^{t}}$

Moreover, the uncertainty in these regulations is reflected in the 0.0 perturbation response when the **AKTT308** to **AKT** or the **PI3K** to **PIP3** edge was dropped in the perturbation analysis (**Figure 6D**).

Discussion

When testing that a network inference approach is predictive, it is important to test it in a controlled a way. We inferred the EGF-driven PI3K and MAPK network circuity inside MCF10A mammary epithelial cells, and then evaluated the accuracy of

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

our predicted network with in silico tests against a target model network that was known a priori within live cells. In the case of MCF10A cells, we relied on phospho-western blot data generated from a series of perturbation experiments performed with and without EGF stimulation as the input (in the form of expected values) to the search algorithm.

Our results indicate that the method can be used to help infer the underlying signaling circuitry in a population of cells if sufficient experimental data are available. The approach developed can be readily applied to very large networks, but to perform a series of GA searches for a large network, access to high-performance computing facilities is recommended. A limitation of our approach is the implementation of a heuristic GA algorithm, which does not exhaustively search the network space. While synchronous simulation takes considerably less time to run than the asynchronous simulation approach we used, our testing revealed that synchronous simulations could not always find the expected network in a search space. For *in silico* tests, we generated expected values from known test networks and used these values as input to the search algorithm. The tests suggest that asynchronous logic simulations and a GA search can theoretically be used to identify an unknown network from the output of a set of network perturbations. Future development efforts are planned to further optimize the algorithm and ideally reduce the computing time necessary to converge to an optimal solution in large search spaces.

Our results also suggest that the more experimental perturbations performed, the more likely discovery of the source network is. We previously illustrated this point with a systematic perturbation analysis to infer the effects of Honokiol on the Notch signaling pathway in SW480 colon cancer cells (Wynn et al. 2014). Not surprisingly, we also observed that it was almost always necessary to obtain a readout of all non-input nodes in the network in order to converge on a single correct network in a search space. In a perturbation experiment, the state of all input nodes should be known *a priori.* Otherwise, it is not appropriate to treat the node as an input node. It is also clear that as the number of nodes in a network increases, the possible ways the nodes may be connected in a logic network becomes extremely large. Therefore, if attempting to infer the signaling network responsible for observed experimental readouts in a biological network, an exhaustive exploration of the search space of all possible logical circuits will

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

almost never be possible. To overcome this challenge, our results strongly suggest that a GA can provide a reasonable heuristic-based alternative to an exhaustive search. It should be emphasized, however, that even with the search optimization provided by the GA, to obtain results in less than a day for large search spaces (such as the ones used for the 10 node test network or the 15 node hypothesized MCF10A network), it was often necessary to distribute the simulations across up to 40 computing cores.

We relied on phospho-western blots for protein activation readouts. An alternative high-throughput approach for measuring protein activation is the use of multiplexed bead-based immunoassays, which consist of a set of antibodies bound to beads with individual fluorescent signatures, allowing detection of phosphorylation sites (Du et al. 2009; Krishhan et al. 2009). Because the multiplex system can detect several analytes in a single sample volume in far less time than equivalent data can be collected from traditional western blot experiments, this approach holds tremendous potential when used in tandem with protein network inference approaches. Saez-Rodriguez et al. (2009) used the continuous median fluorescence intensity readout of phospho bead-based immuno assays as input to another discrete Boolean based network inference approach. While their method differed from ours in a few important ways, the most significant difference is their use of synchronous Boolean simulations. It is worth noting that their inference approach also used a GA, providing additional support for the utility of heuristic based approaches for limiting the enormous network search spaces encountered in these types of problems.

Our approach recovered the target model network in all *in silico* tests performed. In some cases, it was necessary to perform additional *in silico* perturbation experiments in order to identify a single unique and correct network with a p-score of 0.00. In the case of the 5 node network, the target model network could only be distinguished in the informed search space. When we used no prior knowledge of the regulation of **B, C, D, or E** (the uninformed search space), a set of 52 networks with the same or nearly the same steady values under all possible perturbations combinations was found (**Supplemental Figure 3**). If this situation is encountered with a real network in the laboratory, we recommend to expand the initial hypothetical network with additional nodes, which are considered critical to the signaling mechanism.

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Similarly to Saez-Rodriguez et al. (2009), we did not identify one single logic network that matched our experimentally measured data. To do so would require performing many additional experimental perturbations as well as obtaining readouts of all or most activation states of the proteins in our hypothesized network, which was based on an extensive literature search. It is noteworthy that we did find a single network that matched the experimental data for all conditions tested, except for the logical regulation of TSC1/2. Our hypothesized network originally assumed that other regulators of TSC1/2 were unimportant relative to AKT and ERK. The final predicted network (**Figure 6D**) suggests that this is not the case and that another EGFR-regulated molecule may play an important role in TSC1/2 regulation. While this result stops short of inferring the most likely "target model" network inside MCF10A cells, it provides a tangible prediction for additional laboratory explorations related to the regulation of TSC1/2.

Of course, finding the "target model" network inside a living cell may not be fully attainable for any reverse engineering methodology and it may not even be possible, given the heterogeneity expected within a population of cells. In reality, what we are likely to find are mechanisms and models of a set of integrated pathways that have not yet been contradicted by experiment. This will be true of all modeling frameworks, not just Boolean based network models. Boolean logic, however, provides a predictive approximation of the transfer of signals in a circuit and has been used to build predictive models of the cell cycle (Davidich and Bornholdt 2008; Li et al. 2004; Ribba et al. 2006), gene regulatory networks (Albert and Othmer 2003; Chaves et al. 2005), circadian clocks (Akman et al. 2012), and signaling networks (Li et al. 2006; Zhang et al. 2008). Living cells behave as extremely robust integrated circuits that exhibit significant heterogeneity, but still respond to external stimuli in predictable ways. The deterministic cellular outcome, for example the induction of proliferative programs in response to growth factor stimuli or the induction of p53 oscillations in the presence of DNA damage, has been compared to digital biological computation (Lahav et al. 2004; Tamsir et al. 2011).

The approach presented here provides a means for inferring and/or validating a network hypothesis from experimental data. Our approach could be used to identify

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

novel adaptive resistance mechanisms of cancer cells to therapeutic drug targets. The reverse engineering methodology presented here can also be implemented to investigate the potential differences between normal and cancer cell lines. Ultimately a logic-based network model can provide only qualitative dynamics of the predicted order of signaling events but cannot, as it stands, provide precision in terms of timescales or the size of a response. Two-state Boolean models can predict that a molecule is activated but not the degree of activation or signal amplification. While more precise approaches with respect to time and signal amplitude are widely used in systems biology, the predictive power of these methods is substantially limited by their dependence on a large number of unknown parameters and kinetic mechanisms (Mourao et al. 2011; Srividhya et al. 2007; Wang et al. 2013; Wynn et al. 2012) that must be assumed prior to simulation. Thus, despite some limitations, when judged in the context of other methods, logic network models provide a predictive and robust methodology for modeling biochemical regulation without requiring prior knowledge of rate constants and reaction mechanisms.

Acknowledgements

SS and MLW acknowledge support from the James S. McDonnell Foundation's grant support for studying complex systems. This work was also partially supported by NIH R25 DK088752 (NC and SS), the University of Michigan Protein Folding Diseases Initiative (SS), the Breast Cancer Research Foundation (MLW, ME, ZFW, and SDM), and the Avon Foundation (ZFW and SDM).

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

References

Akman OE, Watterson S, Parton A, Binns N, Millar AJ, Ghazal P (2012) Digital clocks: simple Boolean models can quantitatively describe circadian systems J R Soc Interface 9:2365-2382. doi:10.1098/rsif.2012.0080

Akutsu T, Miyano S, Kuhara S (1999) Identification of genetic networks from a small number of gene expression patterns under the Boolean network model Pac Symp Biocomput 4:17-28. doi: 10.1142/9789814447300_0003

Albert I, Thakar J, Li S, Zhang R, Albert R (2008) Boolean network simulations for life scientists. Source Code Biol Med 3:16. doi:10.1186/1751-0473-3-16

Albert R, Othmer HG (2003) The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in Drosophila melanogaster. J Theor Biol 223:1-18. doi: 10.1016/S0022-5193(03)00035-3

Arkin A, Ross J (1994) Computational Functions in Biochemical Reaction. Networks Biophys J 67:560-578. doi: 10.1016/S0006-3495(94)80516-8

Bernal-Mizrachi E, Kulkarni RN, Scott DK, Mauvais-Jarvis F, Stewart AF, Garcia-Ocana A (2014) Human beta-cell proliferation and intracellular signaling Part 2: Still driving in the dark without a road map. Diabetes 63:819-831. doi:10.2337/db13-1146

Brummer T, Naegele H, Reth M, Misawa Y (2003) Identification of novel ERK-mediated feedback phosphorylation sites at the C-terminus of B-Raf. Oncogene 22:8823-8834 doi:10.1038/sj.onc.1207185

Chaves M, Albert R, Sontag ED (2005) Robustness and fragility of Boolean models for genetic regulatory networks. J Theor Biol 235:431-449. doi:10.1016/j.jtbi.2005.01.023

Crampin EJ, Schnell S, McSharry PE (2004) Mathematical and computational techniques to deduce complex biochemical reaction mechanisms. Prog Biophys Mol Biol 86:77-112. doi:10.1016/j.pbiomolbio.2004.04.002

Davidich MI, Bornholdt S (2008) Boolean network model predicts cell cycle sequence of fission yeast. PLoS ONE 3:e1672. doi:10.1371/journal.pone.0001672

Deb K (1999) An introduction to genetic algorithms. Sadhana 24:293-315. doi: 10.1007/BF02823145

Du J et al. (2009) Bead-based profiling of tyrosine kinase phosphorylation identifies SRC as a potential target for glioblastoma therapy. Nat Biotechnol 27:77-83. doi:10.1038/nbt.1513

Foster KG, Fingar DC (2010) Mammalian target of rapamycin (mTOR): Conducting the cellular signaling symphony. J Biol Chem 285:14071-14077. doi:10.1074/jbc.R109.094003

Garg A, Di Cara A, Xenarios I, Mendoza L, De Micheli G (2008) Synchronous versus asynchronous modeling of gene regulatory networks. Bioinformatics 24:1917-1925. doi:10.1093/bioinformatics/btn336

Glass L, Kauffman SA (1973) The logical analysis of continuous, non-linear biochemical control networks. J Theor Biol 39:103-129. doi: 10.1016/0022-5193(73)90208-7

Hahn-Windgassen A, Nogueira V, Chen C-C, Skeen JE, Sonenberg N, Hay N (2005) Akt activates the mammalian target of rapamycin by regulating cellular ATP level and AMPK activity. J Biol Chem 280:32081-32089. doi:10.1074/jbc.M502876200

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Halilovic E, She QB, Ye Q, Pagliarini R, Sellers WR, Solit DB, Rosen N (2010) PIK3CA mutation uncouples tumor growth and cyclin D1 regulation from MEK/ERK and mutant KRAS signaling. Cancer Res 70:6804-6814. doi:10.1158/0008-5472.CAN-10-0409

Hay N (2005) The Akt-mTOR tango and its relevance to cancer. Cancer Cell 8:179-183. doi:10.1016/j.ccr.2005.08.008

Hjelmfelt A, Ross J (1995) Implementation of logic functions and computations by chemical-kinetics. Physica D 84:180-193 doi:10.1016/0167-2789(95)00014-U

Hjelmfelt A, Schneider FW, Ross J (1993) Pattern recognition in coupled chemical kinetic systems. Science 260:335-337. doi:10.1126/science.260.5106.335

Hong CC, Kume T, Peterson RT (2008) Role of crosstalk between phosphatidylinositol 3-kinase and extracellular signal-regulated kinase/mitogen-activated protein kinase pathways in artery-vein specification. Circ Res 103:573-579. doi:10.1161/CIRCRESAHA.108.180745

Jiang B-H, Liu L-Z (2008) Role of mTOR in anticancer drug resistance: perspectives for improved drug treatment. Drug Resist Updat 11:63-76. doi:10.1016/j.drup.2008.03.001

Kholodenko BN, Kiyatkin A, Bruggeman FJ, Sontag E, Westerhoff HV, Hoek JB (2002) Untangling the wires: A strategy to trace functional interactions in signaling and gene networks. Proc Natl Acad Sci U S A 99:12841-12846. doi:10.1073/pnas.192442699

Krishhan VV, Khan IH, Luciw PA (2009) Multiplexed microbead immunoassays by flow cytometry for molecular profiling: Basic concepts and proteomics applications. Crit Rev Biotechnol 29:29-43. doi:10.1080/07388550802688847

Kulkarni RN, Mizrachi EB, Ocana AG, Stewart AF (2012) Human β-cell proliferation and intracellular signaling: driving in the dark without a road map. Diabetes 61:2205-2213. doi:10.2337/db12-0018

Lahav G, Rosenfeld N, Sigal A, Geva-Zatorsky N, Levine AJ, Elowitz MB, Alon U (2004) Dynamics of the p53-Mdm2 feedback loop in individual cells. Nat Genet 36:147-150. doi:10.1038/ng1293

Leyton J et al. (2008) Noninvasive imaging of cell proliferation following mitogenic extracellular kinase inhibition by PD0325901. Mol Cancer Ther 7:3112-3121. doi:10.1158/1535-7163.MCT-08-0264

Li F, Long T, Lu Y, Ouyang Q, Tang C (2004) The yeast cell-cycle network is robustly designed. Proc Natl Acad Sci U S A 101:4781-4786. doi:10.1073/pnas.0305937101

Li S, Assmann SM, Albert R (2006) Predicting essential components of signal transduction networks: A dynamic model of guard cell abscisic acid signaling. PLoS Biol 4:e312. doi:10.1371/journal.pbio.0040312

Liang S, Fuhrman S, Somogyi R (1998) Reveal, a general reverse engineering algorithm for inference of genetic network architectures. Pac Symp Biocomput:18-29

Liu X, Shah A, Gangwani MR, Silverstein PS, Fu M, Kumar A (2014) HIV-1 Nef induces CCL5 production in astrocytes through p38-MAPK and PI3K/Akt pathway and utilizes NF-kB, CEBP and AP-1 transcription factors. Sci Rep 4:4450. doi:10.1038/srep04450

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Manning BD, Tee AR, Logsdon MN, Blenis J, Cantley LC (2002) Identification of the tuberous sclerosis complex-2 tumor suppressor gene product tuberin as a target of the phosphoinositide 3-kinase/Akt pathway. Mol Cell 10:151-162. doi: 10.1016/S1097-2765(02)00568-3

McCormick F (2011) Mutant onco-proteins as drug targets: successes, failures, and future prospects. Curr Opin Genet Dev 21:29-33. doi:10.1016/j.gde.2010.12.002

McCubrey JA et al. (2006) Roles of the RAF/MEK/ERK and PI3K/PTEN/AKT pathways in malignant transformation and drug resistance. Adv Enzyme Regul 46:249-279. doi:10.1016/j.advenzreg.2006.01.004

Meier F et al. (2005) The RAS/RAF/MEK/ERK and PI3K/AKT signaling pathways present molecular targets for the effective treatment of advanced melanoma. Front Biosci 10:2986-3001.

Mitchell M (1996) An introduction to genetic algorithms. Complex adaptive systems. MIT Press, Cambridge, Mass.

Mourão MA, Srividhya J, McSharry PE, Crampin EJ, Schnell S (2011) A graphical user interface for a method to infer kinetics and network architecture (MIKANA). PLoS ONE 6:e27534. doi:10.1371/journal.pone.0027534

Nakayama K, Satoh T, Igari A, Kageyama R, Nishida E (2008) FGF induces oscillations of Hes1 expression and Ras/ERK activation. Curr Biol 18:R332-334. doi:10.1016/j.cub.2008.03.013

Nie S, Chang C (2007) PI3K and Erk MAPK mediate ErbB signaling in Xenopus gastrulation. Mech Dev 124:657-667. doi:10.1016/j.mod.2007.07.005

Plas DR, Thompson CB (2005) Akt-dependent transformation: there is more to growth than just surviving. Oncogene 24:7435-7442. doi:10.1038/sj.onc.1209097

Qiang YW, Yao L, Tosato G, Rudikoff S (2004) Insulin-like growth factor I induces migration and invasion of human multiple myeloma cells. Blood 103:301-308. doi:10.1182/blood-2003-06-2066

Ribba B, Colin T, Schnell S (2006) A multiscale mathematical model of cancer, and its use in analyzing irradiation therapies. Theor Biol Med Model 3:7. doi:10.1186/1742-4682-3-7

Ritt DA, Monson DM, Specht SI, Morrison DK (2010) Impact of feedback phosphorylation and Raf heterodimerization on normal and mutant B-Raf signaling. Mol Cell Biol 30:806-819. doi:10.1128/MCB.00569-09

Robey RB, Hay N (2009) Is Akt the "Warburg kinase"?-Akt-energy metabolism interactions and oncogenesis. Semin Cancer Biol 19:25-31. doi:10.1016/j.semcancer.2008.11.010

Rommel C et al. (1999) Differentiation stage-specific inhibition of the Raf-MEK-ERK pathway by Akt. Science 286:1738-1741. doi: 10.1126/science.286.5445.1738

Sabatini DM (2006) mTOR and cancer: Insights into a complex relationship. Nat Rev Cancer 6:729-734. doi:10.1038/nrc1974

Saeed M, Ijaz M, Javed K, Babri HA (2012) Reverse engineering Boolean networks: From Bernoulli mixture models to rule based systems. PLoS ONE 7:e51006. doi:10.1371/journal.pone.0051006

Saez-Rodriguez J, Alexopoulos LG, Epperlein J, Samaga R, Lauffenburger DA, Klamt S, Sorger PK (2009) Discrete logic modelling as a means to link protein signalling

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

networks with functional analysis of mammalian signal transduction. Mol Syst Biol 5:331. doi:10.1038/msb.2009.87

Samaga R, Saez-Rodriguez J, Alexopoulos LG, Sorger PK, Klamt S (2009) The logic of EGFR/ErbB signaling: theoretical properties and analysis of high-throughput data. PLoS Comput Biol 5:e1000438. doi:10.1371/journal.pcbi.1000438

Sarbassov DD et al. (2006) Prolonged rapamycin treatment inhibits mTORC2 assembly and Akt/PKB. Mol Cell 22:159-168. doi:10.1016/j.molcel.2006.03.029

Sarbassov DD, Guertin DA, Ali SM, Sabatini DM (2005) Phosphorylation and regulation of Akt/PKB by the rictor-mTOR complex. Science 307:1098-1101. doi:10.1126/science.1106148

Schultze SM, Hemmings BA, Niessen M, Tschopp O (2012) PI3K/AKT, MAPK and AMPK signalling: protein kinases in glucose homeostasis. Expert Rev Mol Med 14:e1. doi:10.1017/S1462399411002109

Serra V et al. (2011) PI3K inhibition results in enhanced HER signaling and acquired ERK dependency in HER2-overexpressing breast cancer. Oncogene 30:2547-2557. doi:10.1038/onc.2010.626

Shankaran H, Ippolito DL, Chrisler WB, Resat H, Bollinger N, Opresko LK, Wiley HS (2009) Rapid and sustained nuclear-cytoplasmic ERK oscillations induced by epidermal growth factor. Mol Syst Biol 5:332. doi:10.1038/msb.2009.90

Srividhya J, Crampin EJ, McSharry PE, Schnell S (2007) Reconstructing biochemical pathways from time course data. Proteomics 7:828-838. doi:10.1002/pmic.200600428

Stelniec-Klotz I et al. (2012) Reverse engineering a hierarchical regulatory network downstream of oncogenic KRAS. Mol Syst Biol 8:601. doi:10.1038/msb.2012.32

Tamsir A, Tabor JJ, Voigt CA (2011) Robust multicellular computing using genetically encoded NOR gates and chemical 'wires' Nature. 469:212-215. doi:10.1038/nature09565

Thomas R (1973) Boolean formalization of genetic control circuits. J Theor Biol 42:563-585. doi: 10.1016/0022-5193(73)90247-6

Thomas R (2006) Circular causality Syst Biol (Stevenage) 153:140-153. doi: 10.1049/ip-syb:20050101

Thomas R, D'Ari R (1990) Biological feedback. CRC Press, Boca Raton, Fla.

van Golen KL, Bao LW, Pan Q, Miller FR, Wu ZF, Merajver SD (2002) Mitogen activated protein kinase pathway is involved in RhoC GTPase induced motility, invasion and angiogenesis in inflammatory breast cancer. Clin Exp Metastasis 19:301-311. doi: 10.1023/A:1015518114931

Vidal M, Cusick ME, Barabasi AL (2011) Interactome networks and human disease. Cell 144:986-998. doi:10.1016/j.cell.2011.02.016

Wang YK, Hurley DG, Schnell S, Print C-G, Crampin EJ (2013) Integration of steady-state and temporal gene expression data for the inference of gene regulatory networks. PLoS ONE 8:e72103. doi:10.1371/journal.pone.0072103

Won JK, Yang HW, Shin SY, Lee JH, Heo WD, Cho KH (2012) The crossregulation between ERK and PI3K signaling pathways determines the tumoricidal efficacy of MEK inhibitor. J Mol Cell Biol 4:153-163. doi:10.1093/jmcb/mjs021

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Wynn ML, Consul N, Merajver SD, Schnell S (2012) Logic-based models in systems biology: a predictive and parameter-free network analysis method. Integr Biol (Camb) 4:1323-1337. doi:10.1039/c2ib20193c

Wynn ML, Consul N, Merajver SD, Schnell S (2014) Inferring the effects of Honokiol on the Notch signaling pathway in SW480 colon cancer cells. Cancer informatics 13:1-12. doi:10.4137/cin.s14060

Zhang G, He B, Weber GF (2003) Growth factor signaling induces metastasis genes in transformed cells: molecular connection between Akt kinase and osteopontin in breast cancer. Mol Cell Biol 23:6507-6519.

Zhang R et al. (2008) Network model of survival signaling in large granular lymphocyte leukemia. Proc Natl Acad Sci U S A 105:16308-16313. doi:10.1073/pnas.0806447105

Zimmermann S, Moelling K (1999) Phosphorylation and regulation of Raf by Akt (protein kinase B). Science 286:1741-1744. doi: 10.1126/science.286.5445.1741

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure Captions

Figure 1. High-level flow chart of the proposed methodology with genetic algorithm. Data from perturbation experiments are determined into an ON or OFF value and served as an input for the fitting. Prior knowledge on the network based on literature search, if any, defined the input initial search space. The genetic algorithm generates and evolves populations of networks based on the calculated fitness to the input experimental data. The algorithm evolves the output networks and calculates the fitness for the user-defined *n* generations.

Figure 2. Hypothesized network of epidermal growth factor receptor signaling in MCF10A cells. The hypothesized **(A)** interaction and **(B)** logic networks of normal-like EGF-driven signaling of the PI3K and MAPK pathways in MCF10A cells are presented. Logic network simulations included 15 nodes: **EGFR**, **PTEN**, **RAS-GTP**, **RAF**, **TSC12**, **PIP3**, **PI3K**, **S6K1**, **mTORc1**, **mTORc2**, **MEK**, **AKT**, **AKTT308**, **AKTS473**, **and ERK**. (**C**) Western blots of total and phosphorylated AKT, ERK, and PTEN in MCF10A cells in six experimental perturbation conditions are presented along with (**D**) the discretized form of these data as logical expected values. (**E**) The asynchronously simulated output of the hypothesized logic network for each perturbation condition as well as predicted steady states as logical values are also presented. The hypothesized logic network produced a p-score of 0.0.

Figure 3: Simplified form of the hypothesized network of epidermal growth factor receptor signaling in MCF10A cells. The hypothesized logic network was simplified to 9 nodes (**EGFR**, **PTEN**, **PIP3**, **PI3K**, **MEK**, **AKT**, **AKTT308**, **AKTS473**, **and ERK**). The steady-state output of this network is qualitatively identical to the larger hypothesized network (**Figure 2B**). The predicted steady state values (p-score = 0.0) and asynchronous simulation results under the six perturbation conditions are presented.

Figure 4. Simple logic networks. (A) Each black dashed edge between two nodes is a directional edge representing one of three regulation states: (0) no regulation, (1) inhibition, or (2) activation. For a two node network consisting of **A** and **B**, the logic

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

rules corresponding to each node are listed. A total of 9 logic networks can be constructed from these rules. Each logic network can be uniquely encoded as **[A_i, B_i]**, where **A_i** and **B_i** correspond to one of the three regulation rules possible for each node. For example, **[0, 0]** represents the case where no regulation exists between either node (i.e., they are both input nodes), **[1, 2]** represents the case where **A** is inhibited by **B** and **B** is activated by **A**, and **[0, 1]** represents the case where **A** is an input node and **B** is inhibited by **A**. **(B)** Similarly, for a three-node network consisting of **A**, **B**, and **C**, each logic network can be uniquely encoded as **[A_i, B_i, C_i,]**. The logic rules corresponding to each node are listed in **Supplemental Table 2**. A total of 729 interaction networks and 2,197 logic networks are possible for these three nodes

Figure 5. Results of the [1,2] network search. **(A)** Network **[1, 2]** is made of the following logic regulations: **A^{t+1} = not B^t** and **B^{t+1} = A^t**. **(B)** The expected values of network **[1, 2]** were generated based on the steady state values of nodes **A** and **B** under the following perturbation tests: (1) **A *ON***, (2) **A *OFF***, (3) **B *ON***, and (4) **B *OFF***. **(C)** In this search, the p-score of each of the 9 logic networks was calculated from the simulated output of each network and the expected values of network **[1, 2]**. As expected, in this example only network **[1, 2]** had a p-score of 0.0, indicating an exact match.

Figure 6. **Predicted network of epidermal growth factor receptor signaling in MCF10A cells.** (**A**) Western blot of total and phosphorylated Raptor, mTOR, and TSC2 in MCF10A cells in six experimental perturbation conditions are presented along with (**B**) the discretized form of these data as logical expected values. Expected values for the **AKTS473**, **AKTT308**, and **ERK** network nodes are also included for completeness. (**C**) Based on literature knowledge and a series of constrained genetic algorithm (GA) searches, we predict that the network in **Figure 2B** should also include an EGFR driven activator of TSC12. The predicted network generated a p-score of 0.0. (**D**) A perturbation analysis of the predicted network was performed. A perturbation response (the p-score after a single network perturbation divided by the maximum p-score produced by all perturbations of the same type) was calculated for each perturbation. The results of a node-based perturbation analysis (where a single node is set to

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

constitutively active and then constitutively inactive) is summarized. The boxes next to nodes list the perturbation response when the node was perturbed ***OFF*** (-) and ***ON*** (+). The table on the right summarizes the results of two different perturbation analyses: an edge-based perturbation analysis where one regulatory edge is removed and a regulatory-based perturbation analysis where each edge is inverted to its opposite regulatory type (e.g., activation to inhibition and inhibition to activation). Perturbations that had no effect on the network's p-score are highlighted in green (0.00), and perturbations that had the greatest effect on the network's p-score are highlighted in red (1.00).

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Tables

Table 1. The complexity of the network search space grows with the number of nodes in the network.

Nodes in network (*n*)	2	3	4	5	6
maximum number of regulator nodes per node (*n*-1)	1	2	3	4	5
interaction networks possible ($3^{n(n-1)}$)	3^2 = 9	3^6 = 729	3^{12} = 531,441	3^{20} = 3.5×10^9	3^{30} = 2.1×10^{14}
maximum unique logic regulation rule count for a single node (*r*)	3	13	95	1145	20,825
maximum logic network count (r^n)	3^2 = 9	13^3 = 2,197	95^4 = 1.8×10^{11}	$1{,}145^5$ = 2.0×10^{15}	$20{,}825^6$ = 8.2×10^{25}

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 2. All possible logic rules by node in the informed search space of the hypothesized PI3K and MAPK signaling network. The 43 rules combine to form an informed search space of 41,067 possible networks, which is the product of the regulation rule count of each node. All rule identifiers equal to 0 correspond to the case where the node is an input node (and no other node in the network regulates its activation state).

rule count	node rule identifier	node	regulation rule
1	0	AKT	$AKT^{t+1} = AKT^{t}$
2	1	AKT	AKT^{t+1} = not $AKTS473^{t}$
3	2	AKT	$AKT^{t+1} = AKTS473^{t}$
4	3	AKT	AKT^{t+1} = not $AKTT308^{t}$
5	4	AKT	$AKT^{t+1} = AKTT308^{t}$
6	5	AKT	AKT^{t+1} = not $AKTS473^{t}$ and not $AKTT308^{t}$
7	6	AKT	AKT^{t+1} = not $AKTS473^{t}$ or not $AKTT308^{t}$
8	7	AKT	AKT^{t+1} = not $AKTS473^{t}$ and $AKTT308^{t}$
9	8	AKT	AKT^{t+1} = not $AKTS473^{t}$ or $AKTT308^{t}$
10	9	AKT	$AKT^{t+1} = AKTS473^{t}$ and not $AKTT308^{t}$
11	10	AKT	$AKT^{t+1} = AKTS47^{t}$ or not $AKTT308^{t}$
12	11	AKT	$AKT^{t+1} = AKTS473^{t}$ and $AKTT308^{t}$
13	12	AKT	$AKT^{t+1} = AKTS473^{t}$ or $AKTT308^{t}$
14	0	AKTS473	$AKTS473^{t+1} = AKTS473^{t}$
15	1	AKTS473	$AKTS473^{t+1}$ = not $PI3K^{t}$
16	2	AKTS473	$AKTS473^{t+1} = PI3K^{t}$
17	0	AKTT308	$AKTT308^{t+1} = AKTT308^{t}$
18	1	AKTT308	$AKTT308^{t+1}$ = not $PIP3^{t}$
19	2	AKTT308	$AKTT308^{t+1} = PIP3^{t}$
20	0	EGFR	$EGFR^{t+1} = EGFR^{t}$
21	0	ERK	$ERK^{t+1} = ERK^{t}$
22	1	ERK	ERK^{t+1} = not MEK^{t}
23	2	ERK	$ERK^{t+1} = MEK^{t}$
24	0	MEK	$MEK^{t+1} = MEK^{t}$
25	1	MEK	MEK^{t+1} = not $EGFR^{t}$
26	2	MEK	$MEK^{t+1} = EGFR^{t}$
27	0	PI3K	$PI3K^{t+1} = PI3K^{t}$
28	1	PI3K	$PI3K^{t+1}$ = not $EGFR^{t}$
29	2	PI3K	$PI3K = EGFR^{t}$
30	0	PIP3	$PIP3^{t+1} = PIP3^{t}$
31	1	PIP3	$PIP3^{t+1}$ = not $PI3K^{t}$
32	2	PIP3	$PIP3^{t+1} = PI3K^{t}$
33	3	PIP3	$PIP3^{t+1}$ = not $PTEN^{t}$
34	4	PIP3	$PIP3^{t+1} = PTEN^{t}$

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

rule count	node rule identifier	node	regulation rule
35	5	PIP3	$PIP3^{t+1}$ = not $PI3K^{t}$ and not $PTEN^{t}$
36	6	PIP3	$PIP3^{t+1}$ = not $PI3K^{t}$ or not $PTEN^{t}$
37	7	PIP3	$PIP3^{t+1}$ = not $PI3K^{t}$ and $PTEN^{t}$
38	8	PIP3	$PIP3^{t+1}$ = not $PI3K^{t}$ or $PTEN^{t}$
39	9	PIP3	$PIP3^{t+1}$ = $PI3K^{t}$ and not $PTEN^{t}$
40	10	PIP3	$PIP3^{t+1}$ = $PI3K^{t}$ or not $PTEN^{t}$
41	11	PIP3	$PIP3^{t+1}$ = $PI3K^{t}$ and $PTEN^{t}$
42	12	PIP3	$PIP3^{t+1}$ = $PI3K^{t}$ or $PTEN^{t}$
43	0	PTEN	$PTEN^{t+1}$ = $PTEN^{t}$

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 3. Table of key interactions in the PI3K and MAPK signaling pathways. The (+) and (-) symbols indicate activation and inhibition, respectively, of a target molecule by a regulator molecule. Source references are available in the Appendix.

Regulator	Interaction	Target	Source(s)
EGFR	(+) recruits and phosphorylates GAB1, which phosphorylates PI3K via p85	PI3K	[1-4] [2, 5]
EGFR	(+) direct binding to Grb2/SOS, which catalyzes exchange of GDP for GTP on RAS	RAS-GTP	[6] [7, 8]
PTEN	(-) dephosphorylation	PIP3	[9]
RAS-GTP	(+) physical association with PI3K	PI3K	[10-12]
RAS-GTP	(+) membrane recruitment and phosphorylation	RAF	[12]
PI3K	(+) induces phosphorylation	PIP3	[10, 13, 14]
PI3K	(+) via ribosomes	mTORc2	[9, 15-17]
PIP3	(+) phosphorylation	AKT-T308	[9, 10, 13, 14]
mTORc2	(+) phosphorylation	AKT-S473	[9, 15, 16, 18]
RAF	(+) phosphorylation	MEK	[12]
MEK	(+) phosphorylation	Erk1/2	[12]
AKT-T308	(+) activation of AKT by phosphorylation on Thr308	AKT	[9, 19-21]
AKT-S473	(+) activation of AKT by phosphorylation on Ser473	AKT	[9, 19-21]
AKT	(-) serine phosphorylation on RAF-1, 14-3-3 binding	RAF	[22, 23]
AKT	(-) phosphorylation (Thr1462 & Ser939)	TSC12	[9, 16, 17, 24, 25]
AKT	(-) via phosphorylation of BAD	Apoptosis	[26]
AKT	(+) induce expression Cyclin D1, Cyclin D3, and CDK6, which promote proliferation	Proliferation	[27]
AKT	(+) induces expression GLUT, HK, and PFK2, which catalyzes the first committed step of glycolysis	Metabolism	[11, 28]
Erk1/2	(-) phosphorylation induces mobility shift	RAF	[29, 30]

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Regulator	Interaction	Target	Source(s)
Erk1/2	(-) phosphorylation (Thr1462)	TSC12	[17, 31, 32]
Erk1/2	(+) stabilization of Cyclin D1	Proliferation	[33]
TSC1/2	(-) inverse inhibitor (Ser2448)	mTORc1	[9, 16, 17, 28, 34]
mTORc1	(+) phosphorylation (Thr389, Ser371)	S6K1	[35]
S6K1	(-) via inhibiting phosphorylation (Ser636/639) of IRS1 which is physically associated with activation of PI3K	PI3K	[9, 11, 16, 17, 36] [16]
S6K1	(-) via inhibiting phosphorylation (Ser636/639) of IRS1 which activates GRB2/SOS, which catalyzes the GDP to GTP reaction on RAS	RAS-GTP	[9, 11, 16, 17, 36] [6, 7, 37, 38] [7, 8]

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 4. Logic function justifications for the hypothesized normal logic network of the PI3K and MAPK pathways in MCF10A cells. Source references are available in the Appendix.

Logic Rule	Rule Justification	Source(s)
$PTEN^{t+1} = PTEN^{t}$	PTEN is an input to the model. Phosphorylated PTEN is always present under the conditions tested in MCF10A cells	[39]
$EGFR^{t+1} = EGFR^{t}$	EGF is an input to the model, which stimulated its receptor, EGFR. The receptor is assumed to be activated only when EGF is added to MCF10A cells.	[40]
AKT^{t+1} = AKT-T308^{t} ***OR*** AKT-S473^{t}	While AKT requires phosphorylation at both AKT-T308 and AKT-S473 for full activation, we assume that partial activation is possible when only one is present. In support of this are data indicating that cells expressing PTEN were still able to activate AKT.	[20]
AKT-S473^{t+1} = mTORC2^{t}	There is evidence that mTORc2 is essential for full AKT activation (by directly phosphorylating AKT, likely due to PDK2 activity). We assume PDK2 is present in the system but do not include it in the model for simplicity.	[9, 19]
AKT-T308^{t+1} = PIP3^{t}	PIP3 recruits AKT to the cell membrane where AKT is activated by PDK1. We assume PDK1 is present in the system but do not include it in the model for simplicity.	[9, 20, 21]
RAF^{t+1} = RAS-GTP^{t} ***AND NOT*** AKT^{t} ***AND NOT*** Erk1/2RK^{t}	RAS-GTP directly activates RAF, and RAF must be phosphorylated and dephosphorylated on specific sites to remain active. Regardless of the activation state of RAF, AKT appears able to phosphorylate RAF, allowing the small molecule 14-3-3 to bind and reduce activation levels. Therefore, the rule is written as RAS-GTP 'and not AKT'. In addition, Erk1/2 is able to phosphorylate RAF on specific sites that induce a mobility shift in RAF, therefore reducing activation levels. The interaction between RAF and Erk1/2 appears independent of AKT, therefore the rule is written with 'and not Erk1/2'.	[22, 23] [12] [29, 30]
$Erk1/2^{t+1} = MEK^{t}$	ERK1/2 is directly activated by MEK in the MAPK pathway	[41]
$MEK^{t+1} = RAF^{t}$	RAF is upstream activator of MEK. The MEK inhibitor, PD0325901, is assumed to override activation of MEK by RAF phosphorylation in all conditions tested. The inhibitor is not modeled as a separate node in the network.	[41, 42] [43]
$mTORC1^{t+1}$ = S6K1^{t} ***OR NOT*** TSC12^{t}	S6K1 has a positive feedback, activating mTORc1 via h6a and h6c short isoforms. TSC1 and TSC2 form a heterodimer that inhibits the activity of Rheb. Rheb is a small GTPase required for activation of mTORc1. Thus, in the model we assume that mTORc1 is activated by S6K1 positive feedback or inactivated by TSC1/2 activity.	[44] [9]

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Logic Rule	Rule Justification	Source(s)
$mTORc2^{t+1}$ = $PI3K^{t}$	Evidence strongly suggests that PI3K is both necessary and sufficient for mTORc2 activation to occur via mTORc2 ribosomal binding. We assume that the ribosome concentration is above the regulatory threshold in the cells and therefore consider mTORc2 activation regulated by only PI3K.	[45, 46]
$PI3K^{t+1}$ = $EGFR^{t}$ ***OR*** $RAS\text{-}GTP^{t}$ ***OR NOT*** $S6K1^{t}$	This is a particularly complicated regulation. We simplify the network by using EGFR as a direct activator of PI3K instead of detailing the chain of activation from EGFR to GAB1, which normally activates PI3K by recruiting the p85 subunit. Erk1/2 negatively regulates PI3K by phosphorylating GAB1 and rendering it unable to recruit p85, but we assume this negative feedback is not activated within an hour and therefore do not include the feedback in the regulation of PI3K. In addition, S6K1 inhibits IRS1, which activates PI3K. Because insulin, an activator of IRS1, is not added in our experiments, we assume that IRS1 is only regulated by S6K1. Therefore, we simplify the interaction by writing S6K1 as an inhibitor of PI3K, independent of EGFR. Finally, RAS-GTP can also activate PI3K and its activation is not affected by EGFR or S6K1 levels. Finally, we assume the PI3K inhibitor, LY294002, overrides all activators of PI3K and results in complete inhibition of PI3K. The inhibitor is not modeled as a separate node in the network.	[1, 2, 5, 10, 16, 36, 47-52] [53]
$PIP3^{t+1}$ = $PI3K^{t}$ ***AND NOT*** $PTEN^{t}$	PI3K activates PIP3 at the cell membrane and PTEN dephosphorylates PIP3. We assume the inhibitor is dominant.	[9]
$RAS\text{-}GTP^{t+1}$ = $EGFR^{t}$ ***OR NOT*** $S6K1^{t}$	EGFR stimulates RAS-GTP via a GRB2/SOS complex, assumed to be present in all cells. Furthermore, S6K1 inhibits IRS1, which activates RAS-GTP. For simplicity, we leave IRS1 out of our network, as it is assumed to be only regulated by S6K1.	[1, 12] [36]
$S6K1^{t+1}$ = $mTORc1^{t}$	Phosphorylation of S6K1 is induced by mTORc1.	[54]
$TSC1/2^{t+1}$ = NOT AKT^{t} ***AND NOT*** $Erk1/2^{t}$	AKT inhibits TSC1/2 and Erk1/2 inhibits TSC1/2 through the activation of RSK1. Phosphorylation (Ser1798) of RSK1 inhibits TCS1/2 and results in increased mTORc1 expression. We assume that TSC1/2 is only ***ON*** when its activator is present and both its inhibitors are absent. While there are a number of negative regulators of the TSC1/2 dimer, our model only considers Erk1/2 inhibition and AKT inhibition of TSC1/2.	[32]

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 5. Perturbation tests used to search for three-node networks. Results are summarized for three network searches: **[9, 1, 0], [1, 3, 11], and [3, 0, 10]**. Networks are encoded as **[A_i, B_i, C_i]** where **i** indicates the regulation rule present in the network for each node (see **Supplemental Table 2** for the list of regulation rules by node).

		correct network found? / unique solution found?		
Test Set	**Perturbation tests in set**	**[9,1,0]**	**[1,3,11]**	**[3,0,10]**
I	1. **B *OFF***	Yes / No	Yes / No	Yes / No
II	1. **A *OFF* and B *ON*** 2. **A *OFF* and C *ON*** 3. **A *ON* and B *OFF*** 4. **A *ON* and C *OFF***	Yes / No	Yes / No	Yes / No
III	1. **A *ON*** 2. **A *OFF*** 3. **B *ON*** 4. **B *OFF*** 5. **C *ON*** 6. **C OFF**	Yes / Yes	Yes / Yes	Yes / Yes

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 6. Comparison of the time to find the correct network in three different three-node searches. Each search was run three times via genetic algorithm (GA). In addition, exhaustive evaluation of all 2,197 networks in the search space was performed three times. All simulations were distributed across 5 Nehalem core processors. GA tests were initialized with population size of 100 and ran for 20 generations. Values listed are the mean ± the standard deviation of 3 replicates of each network search. All searches used perturbation test set III, which is summarized in Table 5.

	Search for [9, 1, 0]	Search for [1, 3, 11]	Search for [3, 0, 10]
generation target network found in GA search (mean ± std)	**3.0 ± 2.0**	**4.3 ± 2.5**	**5.7 ± 4.0**
elapsed minutes to run 20 generations in the GA search (mean ± std)	**40.0 ± 12.5**	**36.3 ± 10.0**	**41.7 ± 9.2**
elapsed minutes to run exhaustive search (mean ± std)		**98.0 ± 26.9**	

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Appendix – References cited in Tables

1. Montagner, A., et al., *A novel role for Gab1 and SHP2 in epidermal growth factor-induced Ras activation.* Journal of Biological Chemistry, 2005. **280**(7): p. 5350-60.
2. Rodrigues, G.A., et al., *A novel positive feedback loop mediated by the docking protein Gab1 and phosphatidylinositol 3-kinase in epidermal growth factor receptor signaling.* Molecular and Cellular Biology, 2000. **20**(4): p. 1448-59.
3. Maroun, C.R., et al., *The Gab1 PH domain is required for localization of Gab1 at sites of cell-cell contact and epithelial morphogenesis downstream from the met receptor tyrosine kinase.* Molecular and Cellular Biology, 1999. **19**(3): p. 1784-99.
4. Cunnick, J.M., et al., *Phosphotyrosines 627 and 659 of Gab1 constitute a bisphosphoryl tyrosine-based activation motif (BTAM) conferring binding and activation of SHP2.* Journal of Biological Chemistry, 2001. **276**(26): p. 24380-7.
5. Zhang, S.Q., et al., *Receptor-specific regulation of phosphatidylinositol 3'-kinase activation by the protein tyrosine phosphatase Shp2.* Molecular and Cellular Biology, 2002. **22**(12): p. 4062-72.
6. Borisov, N., et al., *Systems-level interactions between insulin-EGF networks amplify mitogenic signaling.* Molecular Systems Biology, 2009. **5**: 256.
7. Ceresa, B.P. and J.E. Pessin, *Insulin regulation of the Ras activation/inactivation cycle.* Molecular and Cellular Biochemistry, 1998. **182**(1-2): p. 23-9.
8. Waters, S.B., et al., *Desensitization of Ras activation by a feedback disassociation of the SOS-Grb2 complex.* Journal of Biological Chemistry, 1995. **270**(36): p. 20883-6.
9. Hay, N., *The Akt-mTOR tango and its relevance to cancer.* Cancer Cell, 2005. **8**(3): p. 179-83.
10. Vojtek, A.B. and C.J. Der, *Increasing complexity of the Ras signaling pathway.* Journal of Biological Chemistry, 1998. **273**(32): p. 19925-8.
11. Robey, R.B. and N. Hay, *Is Akt the "Warburg kinase"?-Akt-energy metabolism interactions and oncogenesis.* Seminars in Cancer Biology, 2009. **19**(1): p. 25-31.
12. McCubrey, J.A., et al., *Roles of the RAF/MEK/ERK and PI3K/PTEN/AKT pathways in malignant transformation and drug resistance.* Advances in Enzyme Regulation, 2006. **46**(1): p. 249-79.
13. Wee, K.B., U. Surana, and B.D. Aguda, *Oscillations of the p53-Akt network: Implications on cell survival and death.* PLoS ONE, 2009. **4**(2): e4407.
14. Ciliberto, A., B. Novak, and J.J. Tyson, *Steady states and oscillations in the p53/Mdm2 network.* Cell Cycle, 2005. **4**(3): p. 488-93.
15. Facchinetti, V., et al., *The mammalian target of rapamycin complex 2 controls folding and stability of Akt and protein kinase C.* EMBO Journal, 2008. **27**(14): p. 1932-43.
16. Sabatini, D.M., *mTOR and cancer: insights into a complex relationship.* Nature Reviews Cancer, 2006. **6**(9): p. 729-34.

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

17. Jiang, B.-H. and L.-Z. Liu, *Role of mTOR in anticancer drug resistance: Perspectives for improved drug treatment.* Drug Resistance Updates, 2008. **11**(3): p. 63-76.
18. Ikenoue, T., et al., *Essential function of TORC2 in PKC and Akt turn motif phosphorylation, maturation and signalling.* EMBO Journal, 2008. **27**(14): p. 1919-31.
19. Sarbassov, D.D., et al., *Phosphorylation and regulation of Akt/PKB by the rictor-mTOR complex.* Science, 2005. **307**(5712): p. 1098-101.
20. Haas-Kogan, D., et al., *Protein kinase B (PKB/Akt) activity is elevated in glioblastoma cells due to mutation of the tumor suppressor PTEN/MMAC.* Current Biology, 1998. **8**(21): p. 1195-8.
21. Cantley, L.C. and B.G. Neel, *New insights into tumor suppression: PTEN suppresses tumor formation by restraining the phosphoinositide 3-kinase/AKT pathway.* Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(8): p. 4240-5.
22. Rommel, C., et al., *Differentiation stage-specific inhibition of the Raf-MEK-ERK pathway by Akt.* Science, 1999. **286**(5445): p. 1738-41.
23. Zimmermann, S. and K. Moelling, *Phosphorylation and regulation of Raf by Akt (protein kinase B).* Science, 1999. **286**(5445): p. 1741-4.
24. Hahn-Windgassen, A., et al., *Akt activates the mammalian target of rapamycin by regulating cellular ATP level and AMPK activity.* Journal of Biological Chemistry, 2005. **280**(37): p. 32081-9.
25. Manning, B.D., et al., *Identification of the tuberous sclerosis complex-2 tumor suppressor gene product tuberin as a target of the phosphoinositide 3-kinase/Akt pathway.* Molecular Cell, 2002. **10**(1): p. 151-62.
26. Datta, S.R., et al., *Akt phosphorylation of BAD couples survival signals to the cell-intrinsic death machinery.* Cell, 1997. **91**(2): p. 231-41.
27. Gharibi, B., M.S. Ghuman, and F.J. Hughes, *Akt- and Erk-mediated regulation of proliferation and differentiation during PDGFRβ-induced MSC self-renewal.* Journal of Cellular and Molecular Medicine, 2012. **16**(11): p. 2789-801.
28. Robey, R.B. and N. Hay, *Mitochondrial hexokinases, novel mediators of the antiapoptotic effects of growth factors and Akt.* Oncogene, 2006. **25**(34): p. 4683-96.
29. Brummer, T., et al., *Identification of novel ERK-mediated feedback phosphorylation sites at the C-terminus of B-Raf.* Oncogene, 2003. **22**(55): p. 8823-34.
30. Ritt, D.A., et al., *Impact of feedback phosphorylation and Raf heterodimerization on normal and mutant B-Raf signaling.* Molecular and Cellular Biology, 2010. **30**(3): p. 806-19.
31. Song, M.S., L. Salmena, and P.P. Pandolfi, *The functions and regulation of the PTEN tumour suppressor.* Nature Reviews Molecular and Cellular Biology, 2012. **13**(5): p. 283-96.
32. Roux, P.P., et al., *Tumor-promoting phorbol esters and activated Ras inactivate the tuberous sclerosis tumor suppressor complex via p90 ribosomal S6 kinase.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(37): p. 13489-94.

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

33. Halilovic, E., et al., *PIK3CA mutation uncouples tumor growth and cyclin D1 regulation from MEK/ERK and mutant KRAS signaling.* Cancer Research, 2010. **70**(17): p. 6804-14.
34. Foster, K.G. and D.C. Fingar, *Mammalian target of rapamycin (mTOR): Cconducting the cellular signaling symphony.* Journal of Biological Chemistry, 2010. **285**(19): p. 14071-7.
35. Magnuson, B., B. Ekim, and D.C. Fingar, *Regulation and function of ribosomal protein S6 kinase (S6K) within mTOR signalling networks.* Biochemical Journal, 2012. **441**(1): p. 1-21.
36. Ozes, O.N., et al., *A phosphatidylinositol 3-kinase/Akt/mTOR pathway mediates and PTEN antagonizes tumor necrosis factor inhibition of insulin signaling through insulin receptor substrate-1.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(8): p. 4640-5.
37. Skolnik, E.Y., et al., *The SH2/SH3 domain-containing protein GRB2 interacts with tyrosine-phosphorylated IRS1 and Shc: implications for insulin control of ras signalling.* EMBO Journal, 1993. **12**(5): p. 1929-36.
38. Noguchi, T., et al., *Role of SH-PTP2, a protein-tyrosine phosphatase with Src homology 2 domains, in insulin-stimulated Ras activation.* Molecular and Cellular Biology, 1994. **14**(10): p. 6674-82.
39. Hollestelle, A., et al., *Phosphatidylinositol-3-OH kinase or RAS pathway mutations in human breast cancer cell lines.* Molecular Cancer Research, 2007. **5**(2): p. 195-201.
40. Rao, G.S., S. Murray, and S.P. Ethier, *Radiosensitization of human breast cancer cells by a novel ErbB family receptor tyrosine kinase inhibitor.* International Journal of Radiation Oncology, 2000. **48**(5): p. 1519-28.
41. Rhee, S.H., et al., *MEK is a key modulator for TLR5-induced interleukin-8 and MIP3α gene expression in non-transformed human colonic epithelial cells.* Journal of Biological Chemistry, 2004. **279**(24): p. 25179-88.
42. Peyssonnaux, C. and A. Eychène, *The Raf/MEK/ERK pathway: new concepts of activation.* Biology of the Cell, 2001. **93**(1-2): p. 53-62.
43. Leyton, J., et al., *Noninvasive imaging of cell proliferation following mitogenic extracellular kinase inhibition by PD0325901.* Molecular Cancer Therapeutics, 2008. **7**(9): p. 3112-21.
44. Ben-Hur, V., et al., *S6K1 alternative splicing modulates its oncogenic activity and regulates mTORC1.* Cell Reports, 2013. **3**(1): p. 103-15.
45. Xie, X. and K.-L. Guan, *The ribosome and TORC2: Collaborators for cell growth.* Cell, 2011. **144**(5): p. 640-2.
46. Zinzalla, V., et al., *Activation of mTORC2 by association with the ribosome.* Cell, 2011. **144**(5): p. 757-68.
47. DeBerardinis, R.J., et al., *Brick by brick: metabolism and tumor cell growth.* Current Opinion in Genetics & Development, 2008. **18**(1): p. 54-61.
48. Majewski, N., et al., *Hexokinase-mitochondria interaction mediated by Akt is required to inhibit apoptosis in the presence or absence of Bax and Bak.* Molecular Cell, 2004. **16**(5): p. 819-30.
49. Lehr, S., et al., *Identification of major ERK-related phosphorylation sites in Gab1.* Biochemistry, 2004. **43**(38): p. 12133-40.

bioRxiv preprint doi: https://doi.org/10.1101/107730; this version posted April 9, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

50. Myers, M.G., Jr., et al., *The COOH-terminal tyrosine phosphorylation sites on IRS-1 bind SHP-2 and negatively regulate insulin signaling.* Journal of Biological Chemistry, 1998. **273**(41): p. 26908-14.
51. Downward, J., *Targeting RAS signalling pathways in cancer therapy.* Nature Reviews Cancer, 2003. **3**(1): p. 11-22.
52. Thimmaiah, K.N., et al., *Insulin-like growth factor I-mediated protection from rapamycin-induced apoptosis is independent of Ras-Erk1-Erk2 and phosphatidylinositol 3'-kinase-Akt signaling pathways.* Cancer Research, 2003. **63**(2): p. 364-74.
53. Qiang, Y.W., et al., *Insulin-like growth factor I induces migration and invasion of human multiple myeloma cells.* Blood, 2004. **103**(1): p. 301-8.
54. Feng, Z., et al., *The regulation of AMPK β1, TSC2, and PTEN expression by p53: Stress, cell and tissue specificity, and the role of these gene products in modulating the IGF-1-AKT-mTOR pathways.* Cancer Research, 2007. **67**(7): p. 3043-53.

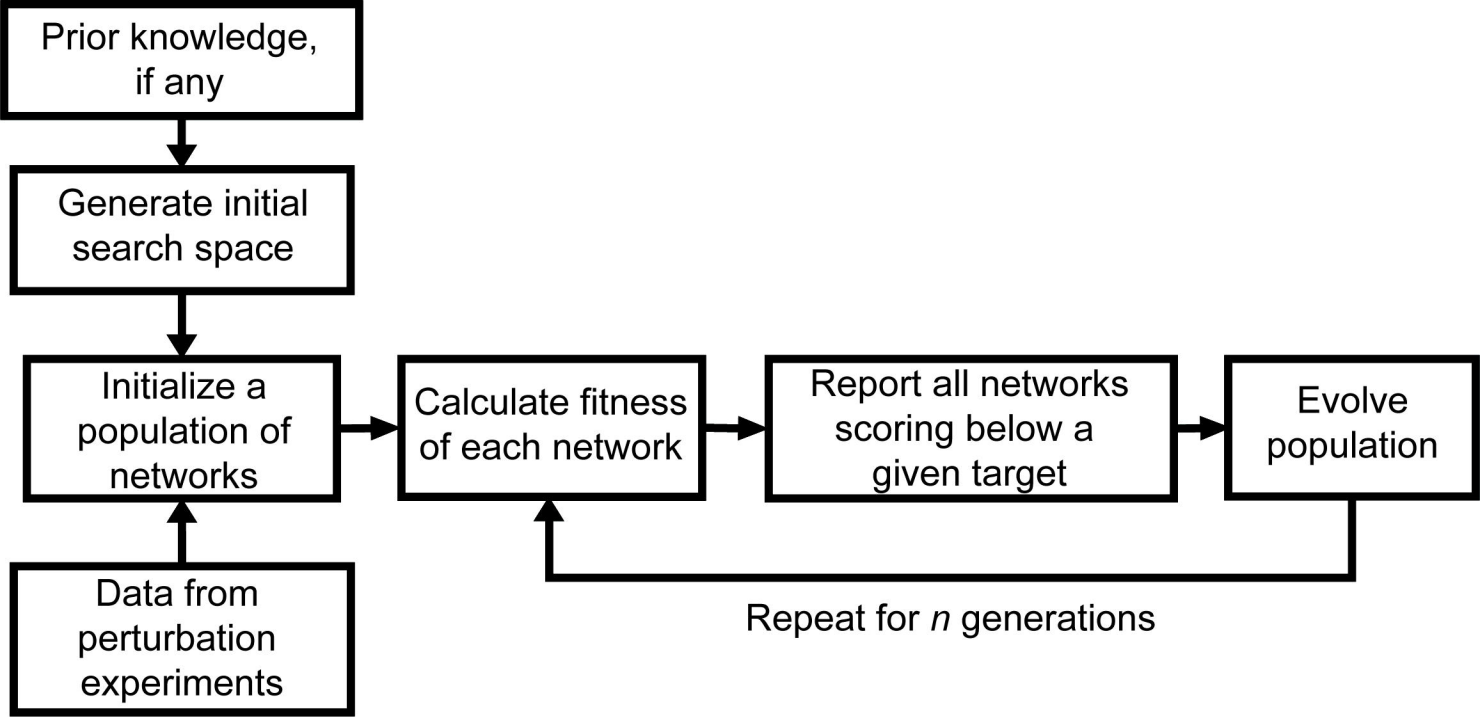

Prior knowledge, if any
Generate initial search space
Initialize a population of networks
Data from perturbation experiments
Calculate fitness of each network
Report all networks scoring below a given target
Evolve population
Repeat for *n* generations

Figure 1.

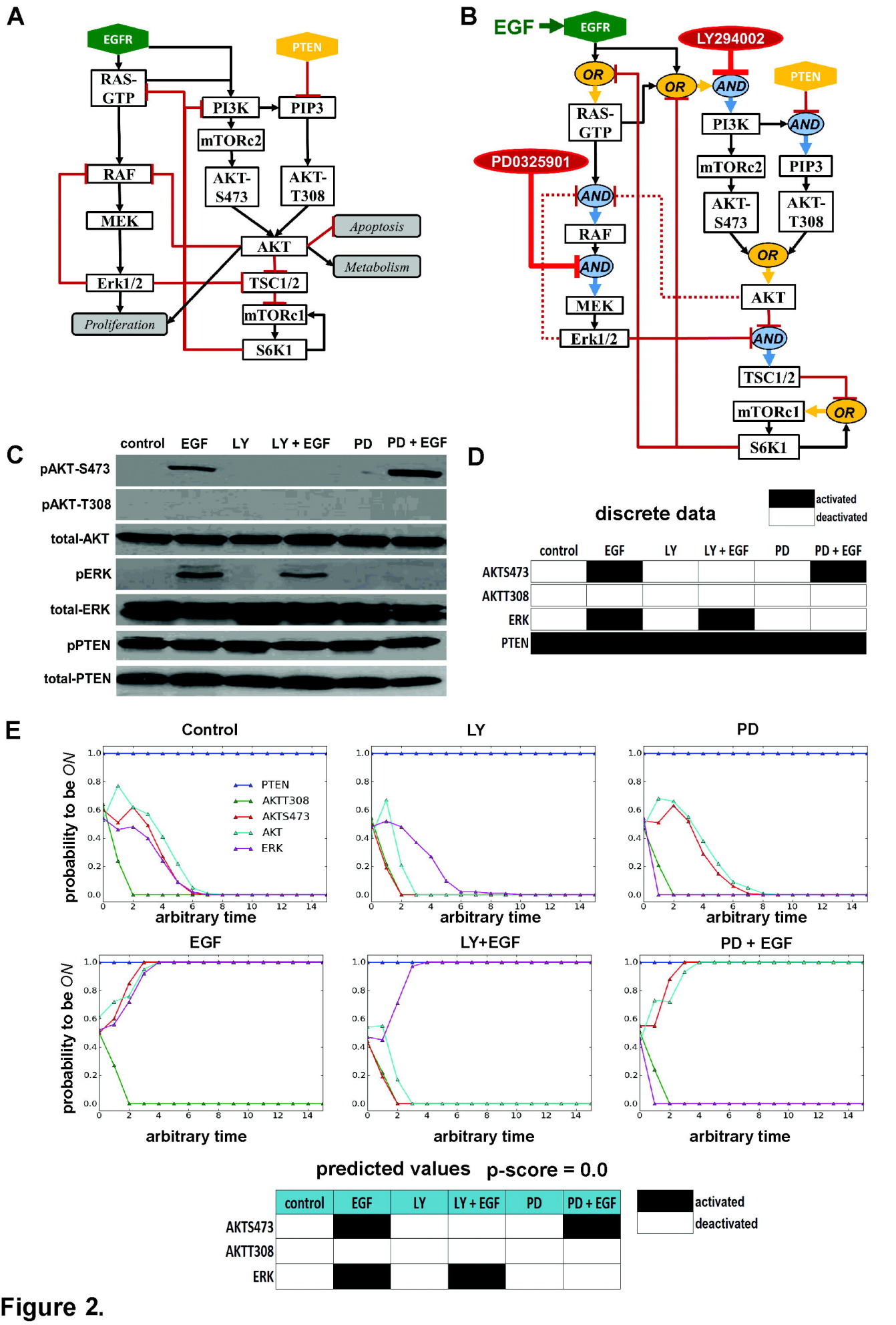

A
EGFR
PTEN
RAS-GTP
PI3K
PIP3
mTORc2
RAF
AKT-S473
AKT-T308
MEK
Apoptosis
AKT
Metabolism
Erk1/2
TSC1/2
Proliferation
mTORc1
S6K1
B
EGF
EGFR
LY294002
OR
OR
AND
PTEN
RAS-GTP
PI3K
AND
PD0325901
mTORc2
PIP3
AND
AKT-S473
AKT-T308
RAF
OR
AND
AKT
MEK
AND
Erk1/2
TSC1/2
mTORc1
OR
S6K1
C
control
EGF
LY
LY + EGF
PD
PD + EGF
pAKT-S473
pAKT-T308
total-AKT
pERK
total-ERK
pPTEN
total-PTEN
D
discrete data
activated
deactivated
control
EGF
LY
LY + EGF
PD
PD + EGF
AKTS473
AKTT308
ERK
PTEN
E
Control
LY
PD
EGF
LY+EGF
PD + EGF
probability to be ON
arbitrary time
PTEN
AKTT308
AKTS473
AKT
ERK
predicted values
p-score = 0.0
control
EGF
LY
LY + EGF
PD
PD + EGF
AKTS473
AKTT308
ERK
activated
deactivated

Figure 2.

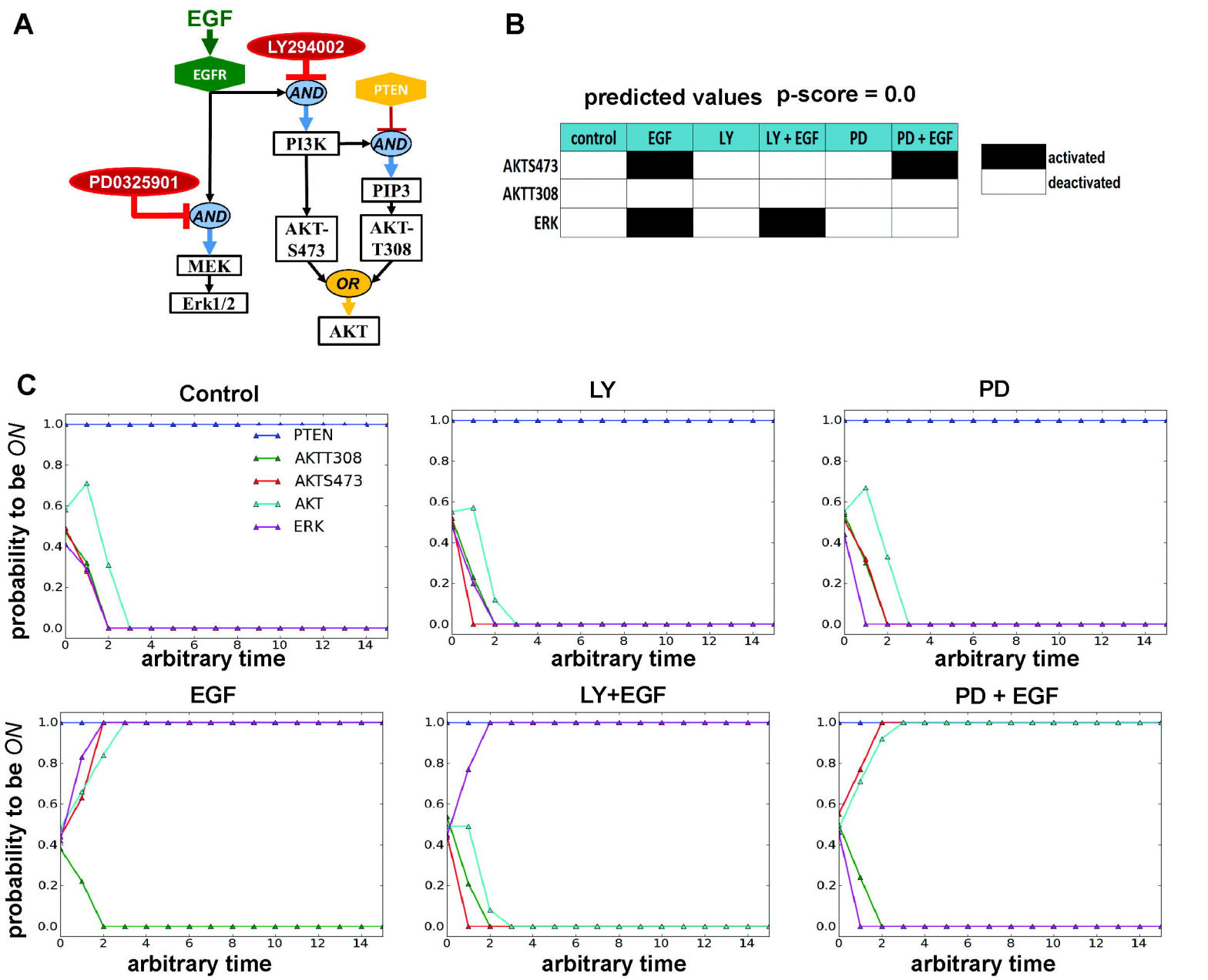

A
EGF
EGFR
LY294002
AND
PTEN
PI3K
AND
PIP3
PD0325901
AND
AKT-S473
AKT-T308
MEK
Erk1/2
OR
AKT
B
predicted values p-score = 0.0
control
EGF
LY
LY + EGF
PD
PD + EGF
AKTS473
AKTT308
ERK
activated
deactivated
C
Control
LY
PD
EGF
LY+EGF
PD + EGF
PTEN
AKTT308
AKTS473
AKT
ERK
probability to be ON
arbitrary time

Figure 3.

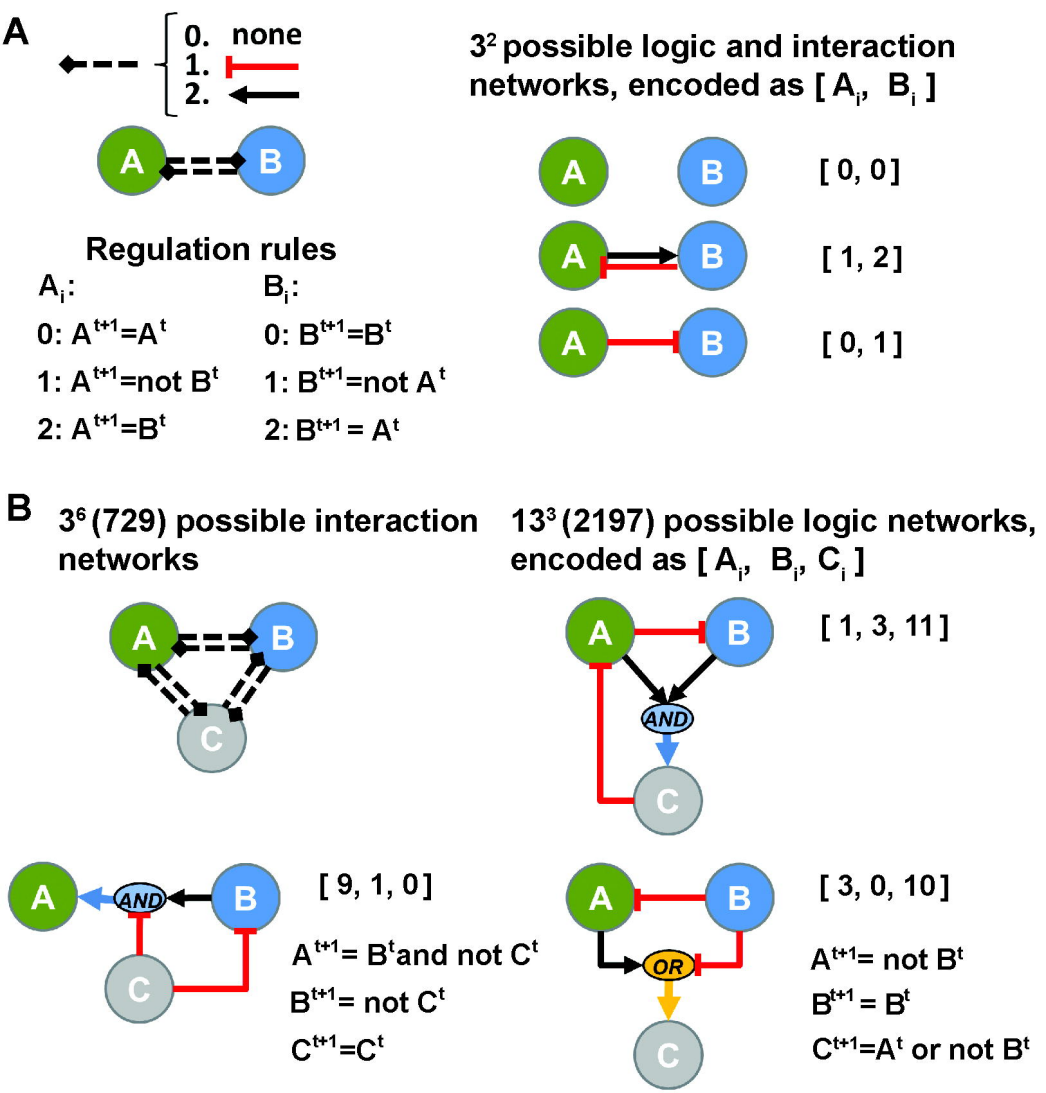
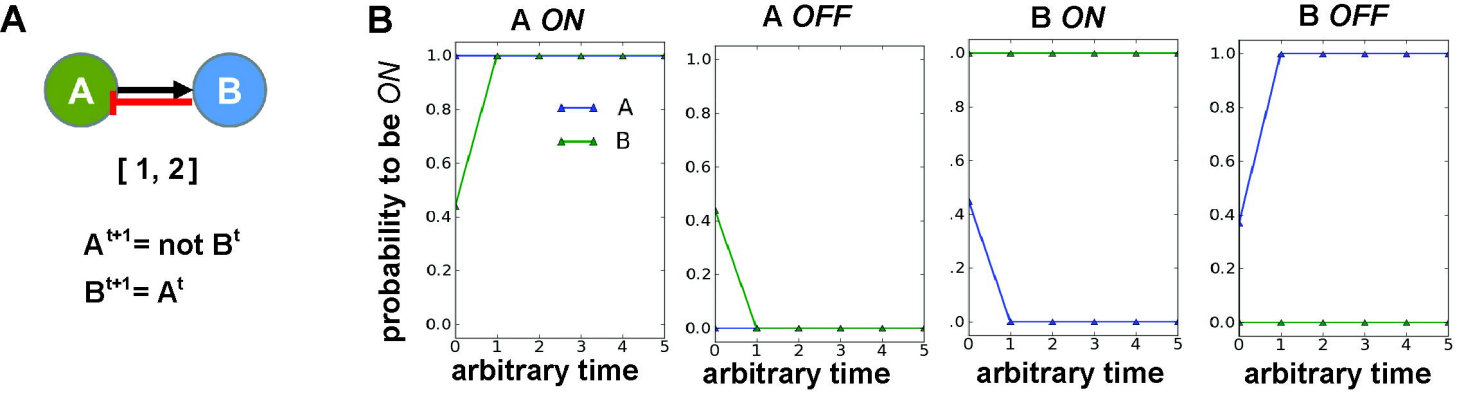
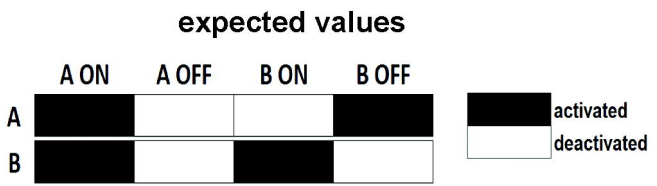

A
0. none
1.
2.
A
B
Regulation rules
A_i:
0: $A^{t+1}=A^t$
1: A^{t+1}=not B^t
2: $A^{t+1}=B^t$
B_i:
0: $B^{t+1}=B^t$
1: B^{t+1}=not A^t
2: $B^{t+1} = A^t$
3^2 possible logic and interaction networks, encoded as [A_i, B_i]
[0, 0]
[1, 2]
[0, 1]
B
3^6 (729) possible interaction networks
13^3 (2197) possible logic networks, encoded as [A_i, B_i, C_i]
C
[1, 3, 11]
AND
[9, 1, 0]
$A^{t+1}= B^t$and not C^t
B^{t+1}= not C^t
$C^{t+1}=C^t$
[3, 0, 10]
OR
A^{t+1}= not B^t
$B^{t+1} = B^t$
$C^{t+1}=A^t$ or not B^t

Figure 4.



C



D results of exhaustive search for [1, 2]

[0, 0]	0.50
[0, 1]	0.75
[0, 2]	0.25
[1, 0]	0.25
[1, 1]	0.50
[1, 2]	0.00
[2, 0]	0.75
[2, 1]	1.00
[2, 2]	0.50

Figure 5.

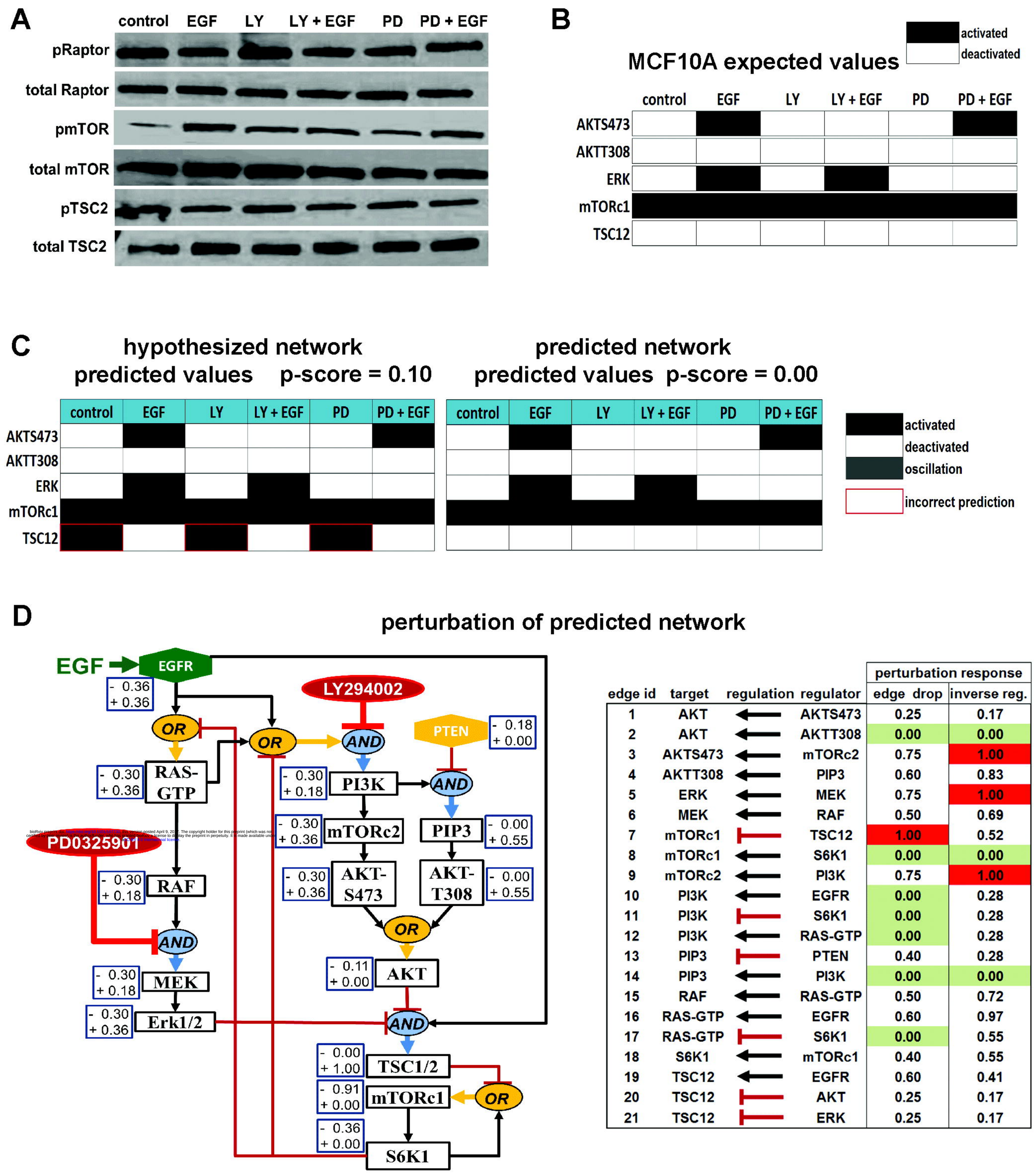

A
control
EGF
LY
LY + EGF
PD
PD + EGF
pRaptor
total Raptor
pmTOR
total mTOR
pTSC2
total TSC2
B
MCF10A expected values
activated
deactivated
AKTS473
AKTT308
ERK
mTORc1
TSC12
C
hypothesized network
predicted values
p-score = 0.10
predicted network
predicted values
p-score = 0.00
oscillation
incorrect prediction
D
perturbation of predicted network
EGF
EGFR
LY294002
PTEN
PD0325901
OR
AND
RAS-GTP
PI3K
mTORc2
PIP3
AKT-S473
AKT-T308
RAF
MEK
Erk1/2
AKT
TSC1/2
mTORc1
S6K1
- 0.36 + 0.36
- 0.18 + 0.00
- 0.30 + 0.36
- 0.30 + 0.18
- 0.00 + 0.55
- 0.11 + 0.00
- 0.00 + 1.00
- 0.91 + 0.00
perturbation response
edge id
target
regulation
regulator
edge drop
inverse reg.
1 AKT AKTS473 0.25 0.17
2 AKT AKTT308 0.00 0.00
3 AKTS473 mTORc2 0.75 1.00
4 AKTT308 PIP3 0.60 0.83
5 ERK MEK 0.75 1.00
6 MEK RAF 0.50 0.69
7 mTORc1 TSC12 1.00 0.52
8 mTORc1 S6K1 0.00 0.00
9 mTORc2 PI3K 0.75 1.00
10 PI3K EGFR 0.00 0.28
11 PI3K S6K1 0.00 0.28
12 PI3K RAS-GTP 0.00 0.28
13 PIP3 PTEN 0.40 0.28
14 PIP3 PI3K 0.00 0.00
15 RAF RAS-GTP 0.50 0.72
16 RAS-GTP EGFR 0.60 0.97
17 RAS-GTP S6K1 0.00 0.55
18 S6K1 mTORc1 0.40 0.55
19 TSC12 EGFR 0.60 0.41
20 TSC12 AKT 0.25 0.17
21 TSC12 ERK 0.25 0.17

Figure 6.