BIOINFORMATICS ORIGINAL PAPER

Vol. 27 no. 4 2011, pages 548–555
doi:10.1093/bioinformatics/btq703

Systems biology

Advance Access publication December 30, 2010

Cancer therapy design based on pathway logic

Ritwik Layek[1], Aniruddha Datta[1], Michael Bittner[2] and Edward R. Dougherty[1,2,*]

[1]Department of Electrical and Computer Engineering, Texas A & M University, College Station, TX 77843-3128 and [2]Translational Genomics Research Institute, Phoenix, AZ 85004, USA

Associate Editor: Trey Ideker

ABSTRACT

Motivation: Cancer encompasses various diseases associated with loss of cell cycle control, leading to uncontrolled cell proliferation and/or reduced apoptosis. Cancer is usually caused by malfunction(s) in the cellular signaling pathways. Malfunctions occur in different ways and at different locations in a pathway. Consequently, therapy design should first identify the location and type of malfunction to arrive at a suitable drug combination.

Results: We consider the growth factor (GF) signaling pathways, widely studied in the context of cancer. Interactions between different pathway components are modeled using Boolean logic gates. All possible single malfunctions in the resulting circuit are enumerated and responses of the different malfunctioning circuits to a 'test' input are used to group the malfunctions into classes. Effects of different drugs, targeting different parts of the Boolean circuit, are taken into account in deciding drug efficacy, thereby mapping each malfunction to an appropriate set of drugs.

Contact: edward@ece.tamu.edu

Received on August 24, 2010; revised on October 30. 2010; accepted on December 15, 2010

1 INTRODUCTION

In eukaryotic multicellular organisms, life is sustained by a systematic coordination between different cells and all extracellular signals. Each cell has its own functionality and its future is determined by various intrinsic and extrinsic biological signals. For instance, a cell's proliferation, differentiation or induction of apoptosis are determined by a number of different signals. From the time of a cell's birth (by division of its parent cell), the cell's state is tightly controlled by different biological regulations. Cell signaling is a form of communication between different cells. These signals can be chemical or electrical impulses. Communication via electrical impulses is typically associated with nerve cells (neurons), which are attached to each other and the action potential transmits from neuron to neuron. For general somatic cells, proteins are usually the signaling molecules used for communication. The interactions between the different signaling molecules are multivariate in nature and hence difficult to study. As a result, historically biologists have focussed on studying the marginal interaction between the signaling molecules, leading to what is called *pathway* information. Although pathway knowledge cannot provide the complete multivariate picture of the overall cellular signal transduction, it is clear that one has to have a mechanism for incorporating this prior information into any signal transduction model that one develops. A procedure to do precisely that was recently developed in Layek *et al.* (2011). In that paper, tools from digital system design were used to generate Boolean networks consistent with given pathway information. Furthermore, it was shown using a specific biological example that a network designed using that approach could replicate relevant experimentally observed behavior from the published literature.

In this article, our goal is to go a few steps further. Here, we are not content with just producing a Boolean network model from given pathway information. Instead our objective is to utilize such a model to (i) enumerate all the possible fault scenarios; (ii) use the response of the model to a test input to determine which fault or class of faults has occurred; and (iii) finally use this information to prescribe an appropriate therapeutic action. To keep the discussion biologically focussed, we will consider the specific case of growth factor (GF) signaling pathways.

The article is organized as follows. In Section 2, pathways and networks are defined in a formal way. In Section 3, cancer is modeled as faults in the underlying signaling network. In Section 4, drug therapies are modeled as interventions to alter aberrant network behavior emanating from a fault. Section 5 gives a biological example showing the power of our methodology. Specifically, fault classification and intervention results for our example are presented. Finally, Section 6 contains some concluding remarks.

2 FROM PATHWAYS TO NETWORKS

From a systems viewpoint, the behavior of a living cell is analogous to that of a multi-input multi-output (MIMO) feedback system. Although the actual protein concentrations in the cell are continuous variables, there are at least three reasons why a discrete type of modeling would be preferred. First, although the continuous model may dictate the exact dynamics, using the current technology it is impossible to reliably measure the concentration of each protein inside the cell in real time. Second, many of the genes/proteins inside the cell exhibit ON/OFF switch-like behavior, which is more readily accommodated using quantization within the digital domain (Jacob and Monod, 1961; Loriferne, 1982). Third, the discrete-time systems are easier to analyze, model and control in real time in comparison to continuous-time systems (Kuo, 1977). In the next subsection, we introduce Boolean networks (BNs) that constitute a popular framework for discrete-time, discrete-space modeling of biological systems including genetic regulatory networks. In addition, we formally define signaling pathways.

2.1 BNs and pathways

A BN (Glass and Kauffman, 1973; Kauffman, 1969, 1993), $B=(V,F)$, on n genes/proteins is defined by a set of nodes

*To whom correspondence should be addressed.

(genes/proteins) $V=\{x_1,\ldots,x_n\}$, $x_i\in\{0,1\}$, $i=1,\ldots,n$, and a list $F=(f_1,\ldots,f_n)$, of Boolean functions, $f_i:\{0,1\}^{n+m}\rightarrow\{0,1\}$, $i=1,\ldots,n$, $m\geq 0$ is the number of external inputs e.g. GFs, stresses, metabolites, etc. Each node x_i represents the state/expression of the gene i, where $x_i=0$ means that gene/protein i is OFF (unexpressed or inactive according to their biological significance) and $x_i=1$ means that gene/protein i is ON (expressed or active). The function f_i is called the *predictor function* for gene/protein i. Updating the states of all genes/proteins in B is done synchronously at every time step according to their predictor functions. If the predictor functions are known, the dynamics of the BN will solely depend on the set of input variables. The dynamic behavior of the BN is quite different in the presence of different external inputs. For m different external inputs, we can model the dynamical system as 2^m different closed BNs, each one of which we may define as a 'context'. The switching between contexts here occurs in response to changes in the activity status of external input variables and is, therefore, deterministic. A stochastic view of context switching has been adopted in several earlier papers, for instance Pal *et al.* (2005); Xiao and Dougherty (2006).

In this article, the starting point is the theoretical construction of context sensitive BNs or input–output BNs from the known biological knowledge of signal transduction pathways. Toward this end, we define the term *pathway segment* $A\overset{t:a,b}{\longrightarrow}B$ to mean that if gene/protein A assumes the value $a\in\{0,1\}$ then gene/protein B transitions to $b\in\{0,1\}$ in no more than t subsequent time steps. A *pathway* is defined to be a sequence of pathway segments of the form $A\overset{t_1:a,b_1}{\longrightarrow}B\overset{t_2:b_2,c}{\longrightarrow}C$. A systematic general procedure for generating a family of BNs consistent with a set of given signaling pathways, or minor variations thereof, is presented in Layek *et al.* (2011). This procedure makes use of standard techniques from digital systems design such as Karnaugh maps (Karnaugh, 1953). A network generated by such a procedure could form the starting point for the exposition of the results of this article, although the specific biological example considered here does not require the use of this machinery.

3 MODELING CANCER AS FAULTS IN THE SIGNALING NETWORK

In molecular biology, the marginal behavior of the normal cell is described using signaling pathways. BNs represent a paradigm that can be used to incorporate this information to model the overall dynamic behavior of the cell, consistent with the pathway knowledge. However, the translational motivation behind this type of dynamical modeling is to facilitate corrective intervention when the cell behaves abnormally. Cancer is actually a disease of several faults in the network. A 'fault' is defined by any structural error of the physical system, such that the dynamics become aberrant. For example, the accumulation of point mutations in the genomic DNA may cause the signaling pathways to behave erratically leading to proliferation. On the other hand, sometimes the fault may not be in the genetic code of a particular protein, but rather it is in the protein synthesis factory ribosome, or in some control mechanism of alternative splicing. The fault could also be in the chromosomal spindle resulting in unequal splitting of the chromosomal DNA between the two daughter cells during cell division. Any of these different kinds of errors could cause structural changes in the regulatory network, thereby changing its dynamics and steady-state behavior. In this section, we try to model different types of biological errors within the BN (digital electronics) framework. In a BN, the faults can be broadly divided into two types.

- Stuck-at fault: A stuck-at fault means that a point in the network circuitry is stuck to a particular value. As a result, the incoming information is no longer communicated beyond the faulty point; instead, only the stuck-at value is passed on to the outgoing port. Clearly, stuck-at faults can commonly be of two types: 'stuck-at-1' faults and 'stuck-at-0' faults with obvious interpretations. We next present an example to show that modeling via stuck-at faults makes biological sense.

  In the mitogen-activated protein kinase (MAPK) pathways, an important signaling protein kinase is the Ras protein. Ras is phosphorylated by many upstream proteins (by GF-mediated pathways). Once activated, Ras activates downstream proteins that have transcriptional control on cyclin D1 and hence cell cycle progression. However, the inherent enzymatic GTPase activity of Ras hydrolyzes the active Ras–GTP complex into the inactive Ras–GDP complex, so that Ras activity ceases after some delay. However, if due to some mutations in the Ras gene, the GTPase activity of the Ras protein is lost, the once activated Ras protein will be constitutively active and will signal the downstream transcription causing proliferation and cancer (Weinberg, 2006). This constitutive activation of Ras can be modeled as a 'stuck-at-1' fault in the Ras node of the BN model of the cell signal transduction. Indeed, the 'stuck-at' fault is a very common one in cancer biology. One of the earliest findings of a very prevalent mutation in cancer was the identification of the Ras oncogene family members, HRAS, KRAS and NRAS. These genes play a critical role in the signaling that drives proliferation. KRAS genes constitutively activated by mutations are found at the very high rate of 17–25% in human cancers (Kranenburg, 2005).
- Bridging fault: As the name suggests, a bridging fault refers to the disruption of old interconnections and incorporation of new interconnections in the network. Bridging faults also make biological sense. The molecular signal transduction relies on the sequences and 3D conformations of the molecules involved. So, any variation in the sequence and 3D conformation of a molecule (mainly protein) will alter its functionality. As a result, many pathways involving that molecule will become inactive while the altered molecule may open up new ones. Without any loss of generality, this kind of aberrant behavior could be modeled as a bridging fault in the BN.

  Indeed, the 'bridging' fault is also a common occurrence in human cancers. A wide variety of tumor types carry chromosomal translocations, where parts of different chromosomes have been joined together. The first such event to be associated with a specific cancer is the Philadelphia chromosome, a translocation joining chromosomes 9 and 22 (Nowell and Hungerford, 1960) and fusing the BCR and ABL genes. The event makes the action of the Abl kinase constitutive in its stimulation of proliferation and inhibition of DNA repair and, if this happens in early blood cell progenitors in the bone marrow, can cause chronic myelogenous leukemia. A variety of drugs that inhibit this kinase activity can produce remission of the disease.

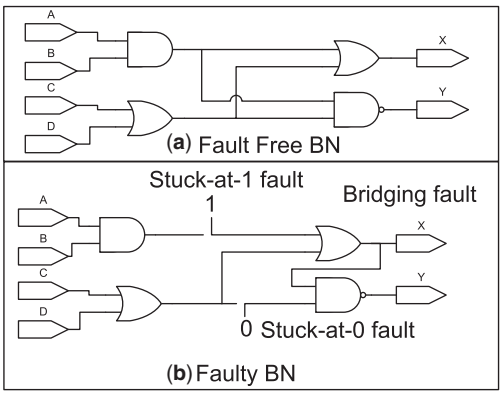


Fig. 1. Stuck-at faults and bridging faults in a digital circuit.

Stuck-at faults and bridging faults are illustrated in Figure 1, where a fault-free BN is shown in Figure 1a while the corresponding faulty network is shown in Figure 1b.

Based on the preceding discussion, it is clear that cancer can be broadly modeled as multiple stuck-at and bridging faults in the BNs corresponding to the normal signaling pathways. In Abramovici *et al.* (1990), extensive theoretical work on digital system testing and fault modeling is presented, which the engineers have been successfully using for digital circuit testing for quite some time now. One of the goals of this article is to use a similar approach for the prediction of fault locations in cancerous networks and the design of intervention policies to compensate for the effect of these faults. For the sake of simplicity, we will focus only on single stuck-at faults. The more general case of cancer modeling involving multiple stuck-at and bridging faults will be taken up in future publications.

3.1 Test inputs and fault detectability

In this article, we will primarily focus on the non-feedback input–output modules of biological systems. Consider the BN of Figure 1a that has four inputs and two outputs as shown. Now suppose that the only possible fault in this network is the stuck-at-1 fault shown in Figure 1b. Following Abramovici *et al.* (1990), for a combinatorial circuit (i.e. non-feedback BN) N, let $Z(x)$ denote the output vector for the input vector x. The presence of a fault f transforms N into N_f with output function $Z_f(x)$ for the same input vector x. We say that a test vector t detects the fault f iff $Z_f(t) \neq Z(t)$. Clearly, for the stuck-at-1 fault in Figure 1b, the test input vector $ABCD = 0000$ can detect the fault because, $Z(0000) = 01$ while $Z_f(0000) = 11$. However, the test input vector 1111 cannot detect the fault since $Z(1111) = Z_f(1111) = 10$. These ideas about fault detectability will be applied to a biological example in Section 5.4.

4 MODELING DRUG INTERVENTION

In a cancerous network, identification of the fault locations is only a part of the task. The major challenge lies in finding the best possible drug or drug combinations with which to intervene. From a theoretical perspective, we can consider the non-cancerous and cancerous (faulty) networks as two different boolean networks. In general, it will be impossible to make a cancerous network revert to the original non-cancerous one using any sort of drug intervention, because the mutations leading to cancer are usually irreversible.

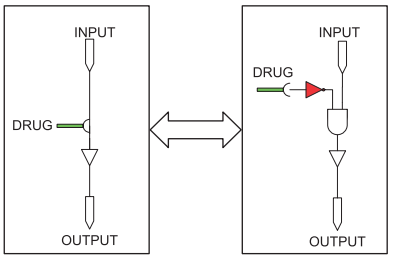


Fig. 2. Drug intervention modeling.

Instead, what the best drug combination could do is to nullify some of the deadly effects (like constitutive cell division) of the cancerous faulty system and try to kill the cell by inducing apoptosis.

The following modeling of drug intervention is inspired by the biological effect of the drug on the pathways. A drug goes into the cell to bind a particular kinase to deactivate its phosphorylating capability. This means that the drug can cut the effect of that particular kinase on molecules further downstream. Hence, the drug can be modeled as an inverted input to an 'AND' gate at the target point of the BN. This schematic modeling of drug intervention is shown in Figure 2.

In this article, our goal is not to derive the mathematical expression for the optimal drug intervention policy, since most mathematically derived policies may be difficult, or impossible, to biochemically implement. Instead, our objective is to model known and well-tested cancer drugs separately and then to find the best suboptimal combination of drugs for a particular cancerous network. The method is described in detail in Section 5.5, where it is applied to a biological example.

5 BIOLOGICAL EXAMPLE: GFS AND CELLULAR SIGNAL TRANSDUCTION

In multicellular eukaryotic organisms, the cell numbers are very tightly controlled, and cells divide to form more cells only when they receive signals from other cells directing them to do so. The external signals that stimulate a cell to divide are usually called *GFs* or *mitogens*. Normally, these are protein or steroid ligands. The external signal directing a cell to divide is usually communicated to the cell division machinery inside the cell through a transmembrane protein called a *GF receptor*. These transmembrane proteins contain the amino acid tyrosine and activate the cell division machinery inside the cell by phosphorylating some key proteins; hence, they are also sometimes referred to as *receptor tyrosine kinases*. Each growth factor binds to its membrane-bound receptor with great specificity and when that happens, an intracellular signaling cascade occurs that can result in enhanced cell proliferation, enhanced protein synthesis or inhibition of apoptosis. In this article, we will focus on the signaling pathways associated with a number of growth factors. One of the reasons for this choice is that these signaling pathways have not only been widely studied in the context of cancer but also different cancer drugs, known to affect different parts of the pathways, are currently available.

Before presenting a detailed schematic diagram of the components involved in these pathways and their interactions, it is appropriate

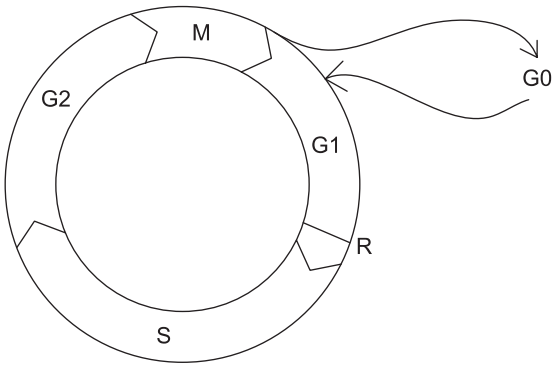


Fig. 3. The eukaryotic cell cycle: G_0 (quiscence), G_1 (gap 1), R (restriction point), S (synthesis), G_2 (gap 2) and M (mitosis).

to first briefly review the eukaryotic cell cycle and point out how malfunctions in the associated control system can lead to cancer.

5.1 Cell cycle control, DNA mutation and cancer

In a multicellular organism, cell growth and proliferation are tightly controlled by the cell cycle control system. The typical eukaryotic cell cycle has four phases called G_1(Gap 1), S(Synthesis), G_2(Gap 2) and M(Mitosis) as shown in Figure 3. The resting phase G_0 is a phase where the cell has made a decision (in the G_1 phase) to temporarily withdraw from the cell cycle. The G_0 and G_1 phases are in equilibrium with each other so that a resting cell in the G_0 phase can readily reenter the cell cycle, if the external conditions require additional cells to be produced.

In the G_1 phase, the cell processes all the extracellular signals (through different pathways) and decides whether to go back to G_0 or proceed toward the S or DNA synthesis phase. The R checkpoint (see Fig. 3) is very critical in the cell cycle regulation. Once the cell goes past the R checkpoint, the progression of the cell cycle no longer depends on the mitogens (the growth factors or the inputs of the transduction pathways). Cyclin–cyclin-dependent kinase (CDK) complexes play major roles in the regulation of the cell cycle dynamics. The growth factor activation of the receptor tyrosine kinases results in rapid accumulation of Cyclin D1. Similarly in normal cells, removal of growth factors results in rapid decline in the Cyclin D1 level. This Cyclin D(1 or 2) and CDK4/6 complexes carry the cell past the R checkpoint. Beyond this checkpoint, although there are mechanisms to check for correct DNA replication and proper apportioning of the chromosomes between the two daughter cells, there are no more decisions made between whether to remain in quiescence or to proceed to proliferation. Thus, after the R checkpoint, the cell cycle is more or less automated and independent of the extracellular inputs. In normal cells, if there is no mitogen during the $G_0 \leftrightarrows G_1$ transition, the cell will not enter the S phase. However, in cancerous cells, the proto-oncogenes can get mutated to become oncogenes. The translated oncoproteins have 3D conformations, which are quite different from that of the corresponding normal protein and can behave differently. For instance, if Ras proto-oncogene mutates to Ras oncogene, the encoded Ras oncoprotein can become constitutively active and start perpetually signaling to the downstream proteins. In that case, even if there is no mitogenic signaling from the outside, the cell will be stimulated to divide. Similarly, mutation in pro-apoptotic genes

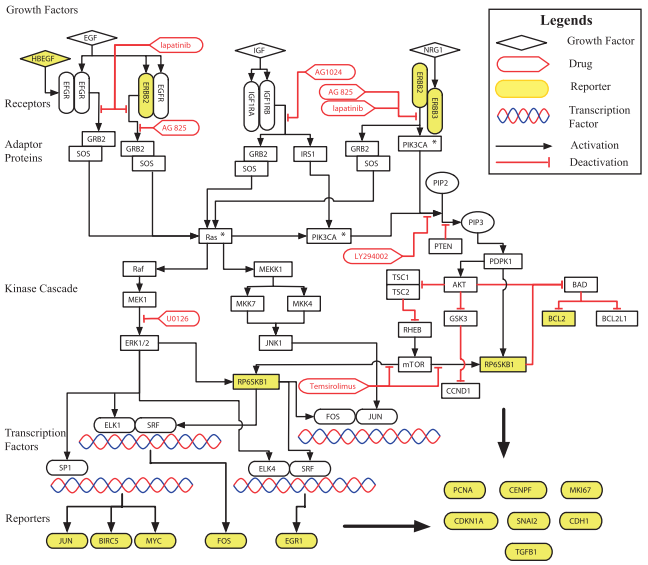


Fig. 4. A schematic diagram of the GF signaling pathways (the yellow color is used for the reporter proteins, which will be measured in future experiments).

can stop apoptosis resulting in tumorigenesis. Since almost all the genes/proteins along the important proliferation/apoptosis pathways are prone to mutation, the number of possibilities for mutation leading to cancer is quite large.

5.2 GF pathways and their boolean circuit modeling

The particular set of signaling pathways that we will focus on in this article are the so-called GF pathways. Our goal is to model these signaling pathways as an input–output Boolean circuit and to use the latter for (i) enumerating the different fault (or malfunction) possibilities, (ii) carrying out fault classification and (iii) designing the appropriate corrective action (or therapy). Such modeling must necessarily be preceded by a biological understanding of the different components of this pathway and their interactions. Figure 4 is a schematic diagram showing the different components of this pathway and their interactions, as currently understood by biologists, for example the Kegg collection of pathways (http://www.genome.jp/kegg/pathway.html) and the NIH BioCarta collection of pathways (http://cgap.nci.nih.gov/Pathways/BioCarta_Pathways). The input nodes in the diagram are the GFs (shown in the rhombuses in Fig. 4). The external signals corresponding to the GFs are transmitted through the kinase cascades and finally activate the appropriate transcription factors. The black and red lines in the diagram indicate relationships which are known to be activating and inhibitory, respectively. Figure 4 also shows six different cancer drugs (red boxes) and the points in the pathway where they are believed to intervene. These signal transduction pathways in Figure 4 constitute a module in a larger tightly controlled network of cell growth, cell division and metabolism.

Believing these pathways to be true, we can develop another level of abstraction by modeling using BNs. For most of the pathways, the modeling is trivial. Figure 5 describes the procedure quite clearly. This modeling approach is quite intuitive and logical and can be applied to the pathways in Figure 4 to arrive at Figure 6.

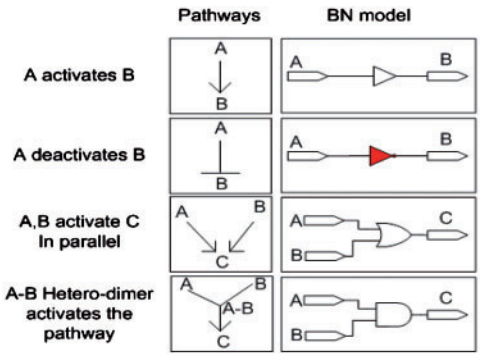


Fig. 5. Going from simple pathways to boolean logic circuits.

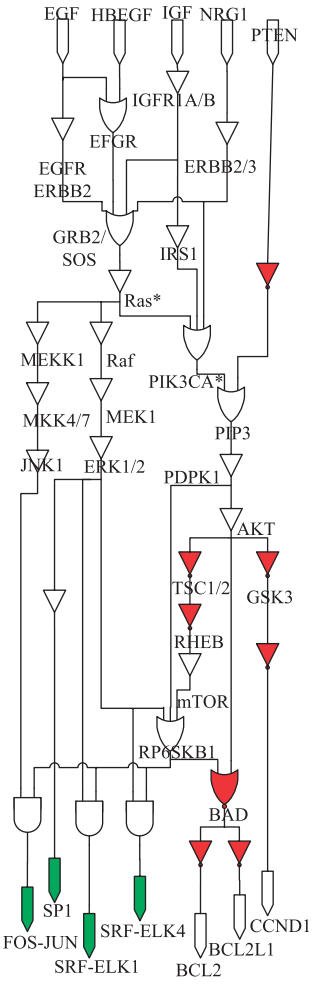


Fig. 6. An input–output BN model of the signal transduction pathways shown in Figure 4.

In the Boolean circuit shown in Figure 6, the seven outputs of interest, shown at the bottom of the figure, are transcription factors (marked in green) and the activation status of some key proteins (not colored). As we will see, such a Boolean circuit model can play an important role in understanding the proliferation versus quiescence decision for a cell. Here, it is appropriate to point out that in those cases where a detailed biological interaction diagram such as Figure 4 is not supplied by biologists, one can use the approach developed in Layek *et al.* (2011) to develop a Boolean network starting from isolated pathway information.

5.3 Modeling faults and therapeutic interventions using the Boolean circuit

Any mutation of any gene or post-transcriptional modification of the corresponding protein can constitutively turn 'ON' or 'OFF' that particular protein. This fits in precisely within the stuck-at fault paradigm considered in Section 3. For the sake of simplicity, in our GF pathways case study, we will consider only single faults of the stuck-at type. In addition, we will only consider the stuck-at faults, which can lead to cancer. For the Boolean circuit shown in Figure 6, the possible locations for the different stuck-at errors, which can induce proliferation and stop apoptosis, are shown in Figure 7a. The numbers are color coded to distinguish between the 'stuck-at-1' and 'stuck-at-0' faults. Specifically, the black numerals refer to the stuck-at-1 faults while the red numerals refer to the stuck-at-0 faults.

As discussed in Section 4, a drug targets particular enzymes along the pathways and cuts off the connectivity of that enzyme to the downstream proteins. This connection cleavage can be achieved via various mechanisms. For instance, the drug may have the capability to bind a target protein and inhibit it from undergoing phosphorylation. For our case study, we consider six potent cancer drugs. Our objective here is not to study their detailed mechanisms of action. Instead, we are interested in using the knowledge from biologists to mark in their intervention locations and corresponding activities on the Boolean circuit of Figure 6. This leads to the effects shown in Figure 7b. Such pictorial representation of the drug activity information is useful.

For instance, let us consider the drug '*lapatinib*' which is known to work on EGFR, ERBB2, EFGR or ERBB3 by inhibiting the signaling capabilities of these receptor tyrosine kinases. From Figure 7b, one can conclude that the drug 'lapatinib' will likely be responsive for the treatment of cancers caused by mutations in the receptor tyrosine kinases although it will probably be ineffective against cancers caused by mutations in the Ras protein, which lies further downstream. Two central objectives of this article are as follows: (i) to use the information contained in Figure 7a to group the numbered faults into different classes; and (ii) to use the information in Figure 7b to predict which set of drugs/drug combinations would be most effective against a particular fault. These objectives are pursued in the next two subsections.

5.4 Fault analysis and classification

From Figure 7a, we see that there are 24 possible fault locations. Alternatively, we could have arrived at the fault locations based on our biological understanding. As already indicated, in this article we will be confining ourselves to the analysis of single faults only. So, for our purposes, the fault can be any one of the 24 faults in the figure. Carrying forward the discussion from Section 3, we use f_i^1 to denote the fault at the i-th location. Then the sample space for the single fault modeling can be defined as $F^1 = \{f_1^1, f_2^1, f_3^1, \ldots, f_{24}^1\}$. Here the superscript 1 refers to the fact that we are considering only single faults. Now if $f_i^1 \in F^1$ occurs, an input vector t detects the fault iff the output vector Z in the faultless system differs

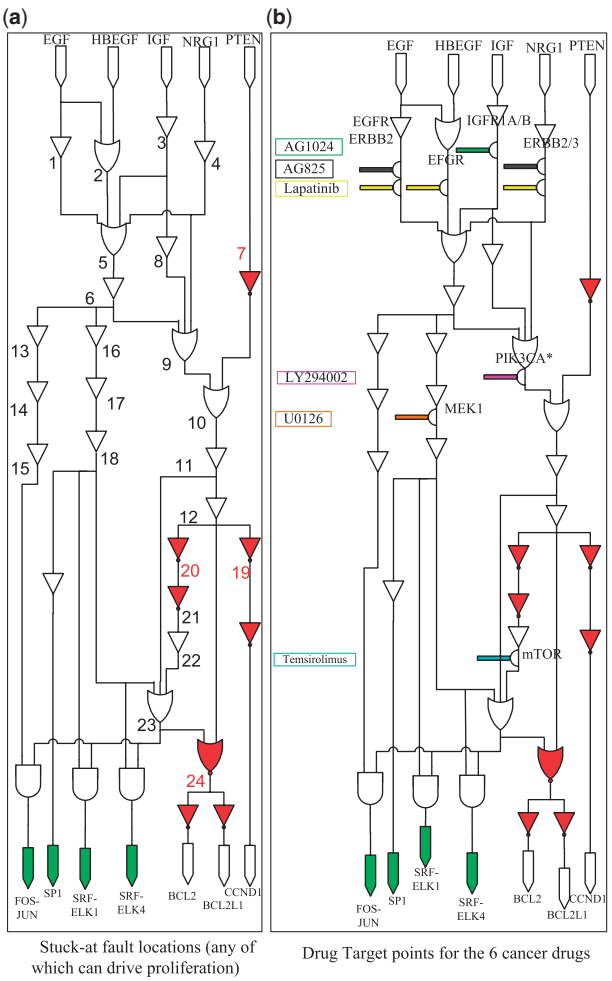
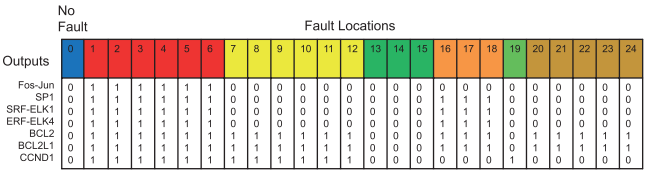


Fig. 7. Possible fault locations and drug intervention locations: (**a**) proliferative stuck-at fault locations (black numerals are stuck-at-1 faults and red numerals are stuck-at-0 faults) and (**b**) intervention locations for the available cancer drugs.

from the output vector Z_{f_1} in the faulty system. Mathematically, $Z(t) \neq Z_{f_1}(t)$. If we cannot find such an input t, we say the fault is undetectable. In the circuit shown in Figure 7a, the only input vector that can detect any $f_i^1 \in F^1$ for this particular network is $V = 00001$, which is achieved with $EGF = 0$, $HBEGF = 0$, $IGF = 0$, $NRG1 = 0$ and $PTEN = 1$. This is due to the fact that for any other binary input V, all the outputs are equal to 1, regardless of whether a stuck-at fault is present or not. This result is not at all surprising. Indeed, when there is no GF outside the cellular membrane and also the tumor suppressor protein *PTEN* is active, we expect to see all the proliferative transcription factors and anti-apoptotic factors deactivated or turned 'OFF'. However, if there are faults (mutations) in the signal transduction pathways, we could see proliferation even in the absence of active input signals (mitogens).

5.4.1 Single fault simulation: In this subsection, computer analysis for the single fault model of the circuit in Figure 6 is presented. The single fault model of the Boolean circuit is shown in Figure 7a. The input vector is



(a) Single Fault Simulation

Output	Equivalent Fault Groups
1111111	1,2,3,4,5,6
0000111	7,8,9,10,11,12
0000000	0(No Fault),13,14,15
0111110	16,1h7,18
0000001	19
0000110	20,21,22,23,24

(b) Equivalent Faults for Input = 00001

Fig. 8. Single fault simulation: (**a**) output simulation in presence of all single stuck-at faults for input V = 00001 and (**b**) equivalent single stuck-at fault groups for input V = 00001.

$V = [EGF, HBEGF, IGF, NRG1, PTEN]$. Each input can take binary values. For this simulation, we take $V = 00001$. The output vector is $Z = [FOS-JUN, SP1, SRF-ELK1, SRF-ELK4, BCL2, BCL2L1, CCND1]$. For the fault-free circuit, we get the output $Z(00001) = 0000000$. Now for the 24 different faults which may induce cancer in the given circuit, the outputs are tabulated in Figure 8a.

5.4.2 Fault classification: From the outputs shown in Figure 8a, we can classify the faults into different groups of equivalent faults. Faults which generate the same output vectors for a particular test input vector are called *'equivalent faults'* with respect to that input test vector. The information in Figure 8a leads us to sets of equivalent faults for the test input vector $V = 00001$. The equivalent fault groups along with their corresponding outputs are shown in Figure 8b.

From Figure 8b, it is clear that any fault in the locations 13, 14, 15 cannot be detected from the output since the corresponding output is the same as that for the fault-free case. Hence, this class of faults is said to be 'undetectable'. It is true that 'undetectable faults' cannot be compensated for based on observations of the output. Assuming that the outputs are true indicators of the processes being monitored, there is no reason why we should be concerned with faults that do not manifest themselves in the outputs. Hence, this is not a major concern especially if we are only interested in the behavior of the outputs.

5.5 Simulation results for drug intervention

Since we have only the six available drugs, we define a drug vector of length 6 as follows. If a particular drug is applied it is assigned the value 1, otherwise it is assigned the value 0. Consequently, the drug vector space has cardinality $2^6 = 64$. The simulation is carried out for all of the possible faults, taken one at a time, and for each of the 64 different drug vectors, and the corresponding outputs are computed. The drug vector is defined by $[lapatinib, AG825, AG1024, U0126, LY294002, Temsirolimus]$.

5.5.1 Continuous real mapping of the output vector: To avoid introducing any possible ambiguity about the origin of the proliferative signaling, we take the same input vector (00001) that

we have previously used for the fault analysis. In the no fault case, with the drug vector 000000 we get the output 0000000 which is certainly non-proliferative. However, in the presence of faults, the outputs will be different. The objective of this simulation is to determine the best possible drug sequence which can nullify the effect of the fault, i.e. produce an output close to 0000000 or away from the proliferative output 11 11 111. We note that although all the output vectors are represented as binary numbers, assigning the usual binary weights to the digits here does not make any biological sense. In other words, 11 11 111 here does not really mean 127 or 0000111 does not really mean 7. Consequently, we need to determine some transformation which will map these $128 = 2^7$ output vectors to a continuous real number scale in a biologically meaningful way. One way to do this is to proceed as follows.

If we examine the components of the output vector, we see that out of the seven components, four are transcription factors, which express (turn ON) the important genes leading to proliferation. The remaining three components capture the activation status of some key proteins in the cytoplasm. So, these two groups of outputs have different biological significance and should be encoded separately. A possible mathematical transformation on the output vectors is described next.

The output vector is

$OUTPUT = [FOS-JUN, SP1, SRF-ELK1, SRF-ELK4, BCL2, BCL2L1, CCND1]$.

Now suppose we take the number of active transcription factors as the first variable (F) and the number of active remaining outputs as the second variable (S). The mathematical transformation makes use of these two variables as described in Equation (1) below:

$$\begin{aligned} \text{Output} &= [a,b,c,d,e,f,g] \\ F &= a+b+c+d \\ S &= e+f+g \\ P &= F \times S \\ S &= F+S \\ \psi(\text{Output}) &= \alpha P + (1-\alpha)S, \end{aligned} \tag{1}$$

where $\alpha \in (0,1)$ is a design parameter. The above encoding scheme counts the number of active transcription factors and the number of active key proteins and combines these two counts via a non-linear many-to-one map, the idea being to quantify the degree of abnormal behavior , e.g. proliferation in the absence of GFs, etc. With α chosen as 0.5, the function ψ's values over the full sweep on the drug vectors and faults are shown in Figure 9. Here, the fault numbers and drug vectors are listed along the horizontal and vertical directions, respectively. The results are color coded for easier visualization, and the color codes used are tabulated on the right side in Figure 9.

5.5.2 Interpretation of the result: From the output tables and the color codes, we see that the color green corresponds to non-proliferation while the color red corresponds to a high chance of proliferation even in the absence of mitogenic signals. So, the best drug vector will be the one which can drive the largest number of faulty circuits toward non-proliferative (green) outputs. For example, the drug vector 000110 drives all of the faults 1–6 to green and most of the remaining boxes along that row away from red. So, the drug combination of U0126 and LY294002 will likely

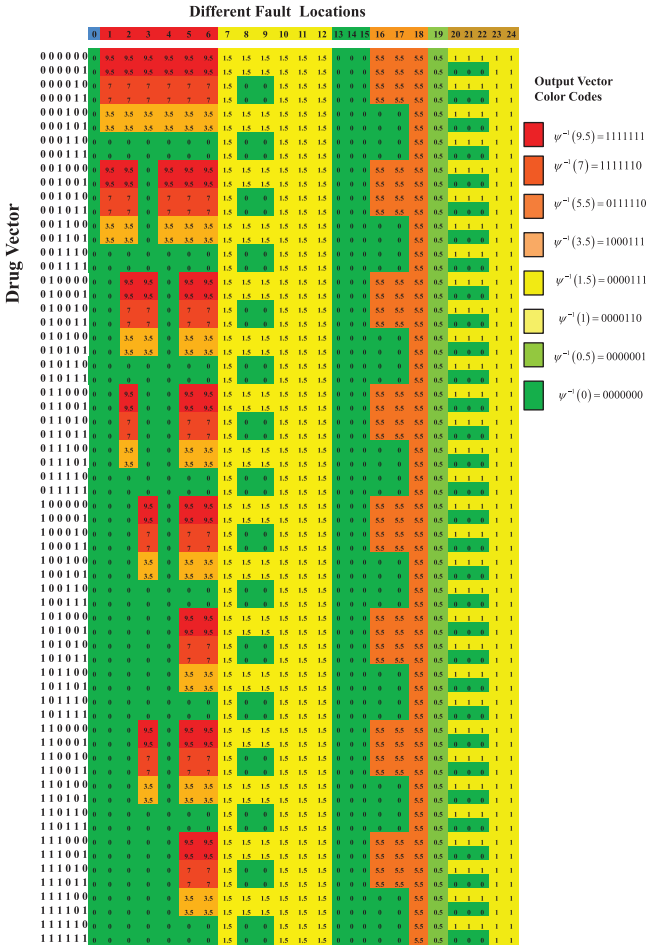


Fig. 9. Drug vector response in the presence of a single fault: (Left) output responses of the circuit for all drug vectors in presence of all single stuck-at faults and (Right) the map between the color codes and the output vectors (red means extreme proliferation while green means non-proliferation).

be effective in producing a non-proliferative output. Another point to note is that there can be faults (like fault 18 in Fig. 9) whose output cannot be altered using any drug sequence. This is not at all surprising and is consistent with the pathway information that we have. Indeed, the fault location 18 is at the $ERK1/ERK2$ protein and there is no available drug in our list downstream of that protein. Consequently, no drug in this particular case study would be able to block the effect of a mutated $ERK1/ERK2$ protein.

6 CONCLUDING REMARKS

In this article, we have presented a new approach for designing cancer therapies based on available pathway information and the manner in which drugs target specific pathway connections. Relevant pathway information is first used to produce BNs whose state transitions are consistent with the given pathway information or minor variations of it. The BN is then realized as a digital circuit which is used to (i) enumerate all the possible fault scenarios; (ii) classify the faults into different classes based on their responses to a particular test input; and (iii) prescribe an appropriate course

of therapeutic action, tailored to the fault or set of faults that has occurred. To keep the discussion focused on practical translational science, the entire presentation has been carried out specific to the GF signaling pathways. These pathways are widely studied in the context of cancer and also have a number of associated drugs known to target them at different points. Because the entire procedure is embedded in classical circuit theory, it can be implemented using slight variations of existing electrical engineering software.

Funding: This work was supported in part by the National Science Foundation under (Grant ECCS-0701531) and in part by the Qatar National Research Fund (a member of the Qatar Foundation) under (Grant NPRP08-148-3-051). The statements made herein are solely the responsibility of the authors.

Conflict of Interest: none declared.

REFERENCES

Abramovici,M. *et al.* (1990) *Digital Systems Testing and Testable Design*, 1st edn. IEEE Press, The IEEE, Inc, New York.

Glass,L. and Kauffman,S. (1973) The logical analysis of continuous, nonlinear biochemical control networks. *J. Theor. Biol.*, **39**, 103–129.

Jacob,F. and Monod,J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.*, **3**, 318–356.

Karnaugh,M. (1953) The map method for synthesis of combinational logic circuits. *Trans. Amer. Inst. Elect. Eng.*, **72**, 593–599.

Kauffman,S. (1969) Metabolic stability and epigenesis in randomly constructed genetic nets. *Theor. Biol.*, **22**, 437–467.

Kauffman,S. (1993) *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, New York.

Kranenburg,O. (2005) The kras oncogene: past, present, and future. *Biochim. Biophys. Acta*, **1756**, 81–82.

Kuo,B.C. (1977) *Digital Control Systems*. 1st edn. SRL Pub. Co.,Champaign, IL.

Layek,R. *et al.* (2011) From biological pathways to regulatory networks. *Mol. BioSyst.*, [Epub ahead of print, doi:10.1039/C0MB00263A, The Royal Society of Chemistry].

Loriferne,B. (1982) *Analog-Digital and Digital-Analog Conversion*. 1st edn. Heyden, Philadelphia, PA.

Nowell,P. and Hungerford,D. (1960) A minute chromosome in human chronic granulocytic leukemia. *Science*, **132**, 1497.

Pal,R.*et al.* (2005) Intervention in context-sensitive probabilistic boolean networks. *Bioinformatics*, **21**, 1211–1218.

Weinberg,R.A. (2006) *The Biology of Cancer*. 1st edn. Garland Science, Princeton.

Xiao,Y. and Dougherty,E.R. (2006) Optimizing consistency-based design of context-sensitive gene regulatory networks. *IEEE Trans. Circuits Syst. I*, **53**, 2431–2437.