

Research Article

A Boolean Network Approach to Estrogen Transcriptional Regulation

Guillermo de Anda-Jáuregui,¹ Jesús Espinal-Enríquez,^{1,2}
Santiago Sandoval-Motta,^{1,2,3} and Enrique Hernández-Lemus^{1,2} 

¹Computational Genomics Division, National Institute of Genomic Medicine, 14610 México City, Mexico

²Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, 04510 México City, Mexico

³Research Chairs Program, National Council on Science and Technology (Conacyt), 03940 México City, Mexico

Correspondence should be addressed to Enrique Hernández-Lemus; ehernandez@inmegen.gob.mx

Received 1 March 2019; Accepted 6 May 2019; Published 22 May 2019

Academic Editor: Yongtang Shi

Copyright © 2019 Guillermo de Anda-Jáuregui et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Gene expression governs important biological processes such as the cell's growth cycle and its response to environmental signals. Alterations of this complex network of transcriptional interactions often lead to unstable expression states and disease. Estrogen is a sex hormone known for its roles in cell proliferation. Its expression has been involved in several physiological functions such as regulating the menstrual and reproduction cycles in women. Altered expression states where estrogen levels are atypically high have been associated with an increased incidence of breast, ovarian, and cervix cancer. To better understand the implications of deregulation of the estrogen and estrogen receptor regulatory networks, in this work we generated a dynamical model of gene regulation of the estrogen receptor transcription network based on known regulatory interactions. By using an adaptation to classical Boolean Networks dynamics we identified proliferative and antiproliferative gene expression states of the network and also to identify key players that promote these altered states when perturbed. We also modeled how pairwise gene alterations may contribute to shifts between these two proliferative states and found that the coordinated subexpression of E2F1 and SMAD4 is the most important combination in terms of promoting proliferative states in the network.

1. Introduction

1.1. Estrogen Regulation. Estrogen is the primary human female sex hormone, responsible for the development of the female reproductive system and the emergence of secondary sexual characteristics. Estrogens are primarily synthesized in the ovaries but can also be produced in the adrenals and testes. From here, these molecules are distributed to target organs and tissues [1, 2].

To regulate biological processes, estrogen binds to estrogen receptor (ER) proteins. Estrogen may readily diffuse through the cell membrane to the cytosol and bind to ERs. ERs belong to either a nuclear or a membrane class. The nuclear class is comprised of two ERs: alpha and beta. ER-alpha is encoded by the ESR1 gene, located in Chromosome 6(6q25.1-q25.2), meanwhile ER-beta is encoded by the ESR2 gene located in Chromosome 14 (14q23.2-q23.3) [3].

The expression of estrogen receptors is regulated through complex genetic and epigenetic control mechanisms. The most extensively described regulatory mechanisms involve the activity of transcription factor (TF) proteins [4]. These TFs bind to regulatory regions in the genome, inducing changes (either positive or negative) in the basal transcription rate of the genes, ultimately affecting the concentration of the encoded protein.

Estrogen-bound ERs can induce changes in the biological state of a cell through two different mechanisms: a nongenomic and a genomic mechanism. The nongenomic mechanism involves signal transduction through secondary messengers via the Estrogen Signaling pathway and related crosstalk pathways [5]. The genomic mechanism of ERs involves its activity as a transcription factor.

Estrogen-bound ERs translocate to the nucleus where they bind to Estrogen Response Elements located in the

promoter region of estrogen-regulated genes. This binding can either increase or decrease their respective gene transcription rate [6]. We may consider that these two phenomena are not disconnected. ER-regulated genes may be involved in the regulation of other TFs which in turn may regulate ER expression, forming a complex gene regulatory network.

Both experimental and computational studies have been performed to map this gene regulatory network in different biological contexts. The description of the gene regulatory mechanisms governing the expression of ERs, as well as the genes whose expression is regulated through ERs, is still an open problem; however, current knowledge of this regulatory network is enough for us to implement models of the gene regulatory dynamics involving estrogen and estrogen receptor genes.

1.2. Boolean Models of Gene Regulation. Boolean networks have been increasingly used as models for simulating the dynamics of gene regulatory networks. This description treats genes (nodes in the network) as binary variables so genes can only be in two possible expression states: active or inactive. The dynamical behavior of these networks has been proved to give insights into the dynamics of real cellular systems. Boolean networks have been used to model the response of the immune system in response to a respiratory infection [7]. They have served to identify potential therapeutic targets in the blood cancer T cell large granular lymphocyte [8].

Of particular interest in these models are the attractors of the dynamical system, defined as the set of recurrent states that the networks reach after some time [9]. The dynamics of the network are simulated by determining the state of each element of the network at a posterior time, given the current state of its regulators. The order by which each state of each element is updated can be either synchronous or asynchronous (see Appendix), and thus the system can be either Markovian or purely deterministic.

Since Boolean networks are discrete dynamical systems with finite support (there are exactly 2^N possible states on a Boolean network with N nodes), the evolution of the system will produce recurrent states. The trajectories will fall into one of a set of steady states or cycles called attractors. The collection of attractors along with the respective configurations that lead to them is called the attractor landscape.

The determination of the set of attractor states and the convergence dynamics leading to those attractors constitutes the solution to the Boolean network dynamics problem. It is important to note that these attractors are often consistent with what is observed experimentally in the gene expression patterns of specific phenotypes or cell types [9–11].

In this paper we present a Boolean model for the estrogen and estrogen receptor gene regulatory network for which we analyze its dynamics under different configurations of presence / absence of nodes and determine which configurations lead to a proclivity to proliferative and antiproliferative states.

2. Analysis

2.1. Construction of the Estrogen Receptor Regulatory Network. To study the gene regulatory phenomenon surrounding the estrogen receptor using the Boolean formalism, a suitable network is required. The reconstruction of such regulatory network is not trivial and will often represent a fragment of the whole set of gene regulatory interactions that could be involved in real organisms. However, taking into account only the first-order regulators often represents a good first approach model to understand how the network behaves under different circumstances [8, 12].

The problem may be approached using much formalism, including coexpression and information-theoretical methods [13], literature mining strategies [14], and even reconstruction using the Boolean formalism itself [15]. Such approaches may become quite computationally expensive, and the reconstructed network itself must be validated against known bibliographic information. With these considerations in mind, we decided to construct a background network for our Boolean dynamics based on well-documented, curated information of transcriptional regulations available in databases.

We constructed this network of gene regulation centered on estrogen receptors α and β (ESR1 and ESR2). We identify regulatory interactions previously described and validated in the current biomedical literature. For this and in order to increase the reliability of the constructed network, we made use of two complementary databases of regulatory interactions: RegNetwork and IPA,

Our first source of regulatory interactions, RegNetwork [16] is a database of transcriptional regulatory relationships in human. This database comprises curated regulations from various other databases (BioGrid, FANTOM, HPRD, IntAct, JASPAR, TRANSFAC, TRED, among others.) that integrate the current understanding of regulatory interactions derived from experimental and theoretical data, for instance by resorting to gene expression datasets for known gene perturbations, and ChIP-seq assays. The database assigns different confidence levels to the listed interactions based on the information provided by each of its sources. Furthermore, it makes a distinction between predicted and experimentally validated regulatory interactions.

Our second source of regulatory interactions is the Ingenuity Pathway Analysis (IPA) platform (IPA ©, QIAGEN Redwood City). A commercial software package that relies on a proprietary Ingenuity Knowledge Base [17] which contains a causal network derived from experimental observations as well as records from manually curated biomedical literature.

Using these two sources, we constructed our *estrogen receptor regulatory network* following these steps:

- (1) Extract first neighbors of ESR1 and ESR2 from RegNetwork; consider only interactions with experimental evidence and high confidence.
- (2) Extract interactions between first neighbors of ESR1 and ESR2, considering only interactions with experimental evidence and high confidence. This will be Network 1.

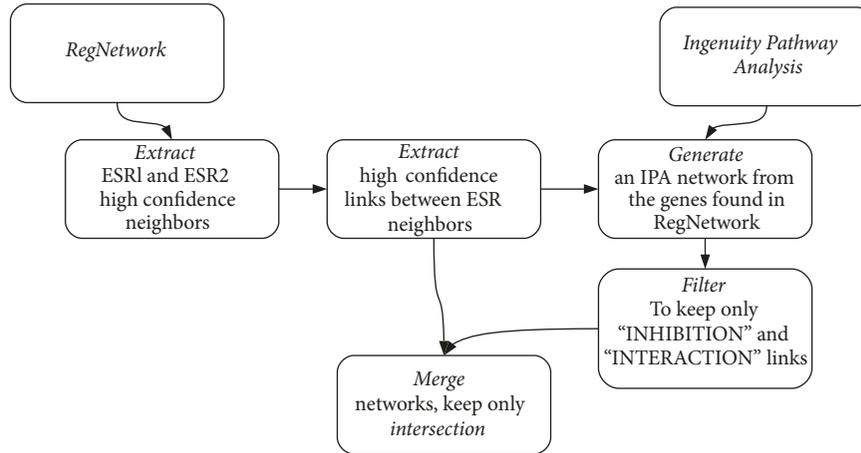


FIGURE 1: Pipeline followed in this work.

- (3) Filter out nodes in Network 1 with $out - degree = 0$ iteratively, until there are no nodes with $out - degree = 0$. This will be Network 2.
- (4) Take the nodes in Network 2, and use them as the input of an IPA analysis to generate networks based on interactions described in IPA. We merge all IPA networks and then remove nodes that were not present in Network 2 to generate Network 3.
- (5) We filter Network 3 by keeping only those links that are described by IPA as being either "INHIBITION" or "ACTIVATION," and we remove all nodes with $out - degree = 0$ iteratively until there are no nodes with $out - degree = 0$. This generates Network 4.
- (6) We generate a network with the intersection of nodes and edges of network 2 and network 4 and iteratively remove nodes with $out - degree = 0$. This generates the final *estrogen receptor regulatory network* to be used in this work (Figure 1).

It is known that the general inference of this kind of networks is a computationally expensive problem (NP-hard). By considering a limited search space as we just sketched, we are able to directly delve into the phenomenological implications of the studied system. This is so since fixing the network in compliance to established experimental facts and known biological information by resorting to causal inference has proven to be an effective way to proceed, in particular in cases like the estrogen pathway, for which a large number of experimental evidence sources are available.

2.2. Boolean Rules and Dynamics. Boolean rules are the set of logical constraints that a node may have in a Boolean network. This rule depends on the state of regulators of the node, and the activatory/inhibitory nature of those regulators. Let us explain one dynamic rule of this signaling pathway. The estrogen receptor 1, ESR1, is a well-known gene that encodes a protein which participates in processes such as DNA binding, activation of transcription, or sexual development. This protein (among other functions) binds to the androgen receptor

TABLE 1: ESR1 truth table.

AR(t)	BRCA1(t)	STAT5A(t)	ESR1($t + 1$)
0	0	0	0
0	0	1	0
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	0
1	1	0	1
1	1	1	0

(AR), which in turn regulates negatively the expression of ESR1. STAT5A is another transcription factor that promotes ESR1 activity. At the same time, BRCA1, a crucial gene in DNA damage response, regulates positively the activity of ESR1.

ESR1 has two positive regulators (AR and STAT5A) and one repressor (BRCA1). The dynamical state of ESR1 at time $t + 1$, will depend on the state of AR, BRCA1, and STAT5A at time t . The logical rule of ESR1 may be written as in Table 1

In this case, the single or combined action of BRCA and STAT5A will activate the estrogen receptor ($ESR1 = 1$) only if AR is not present; otherwise, ESR1 will acquire a value of 0. Regarding the combined regulation of ESR1, as in the rest of this dynamical system, the negative regulators exert a stronger influence than the positive ones. This differentiated influence of negative regulators has been applied in other biological Boolean networks before [9–11], since in some cases, it is not possible to know experimentally the effect of combined action of more than one gene/protein over a determined molecule. It has been shown that the differentiated action of negative and positive regulators is often in agreement with the biological system.

2.3. Boolean Model and Attractor Landscape Analysis. Our simplified estrogen receptor regulatory network consists of 14 nodes (Figure 2). In this dynamic model, each node can

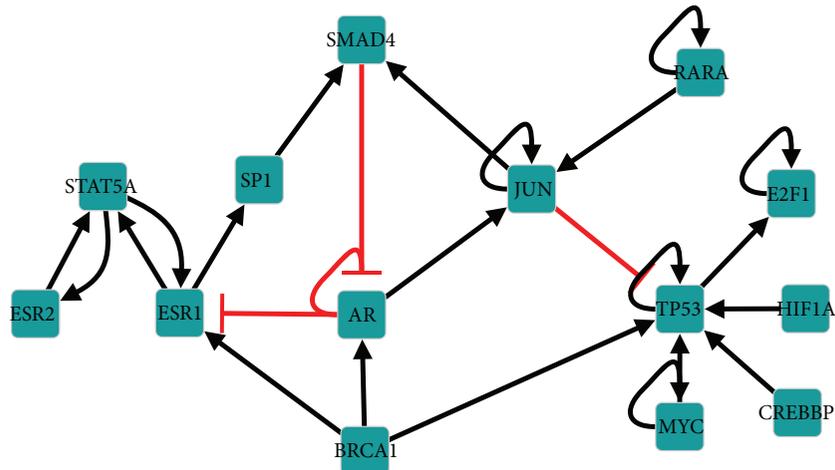


FIGURE 2: The estrogen receptor regulatory network. It was constructed from curated transcriptional interactions found in RegNetwork and IPA. Gene nodes include ESR1 and ESR2, the estrogen receptors. It has 25 interactions representing transcriptional regulation among these genes. Inhibitory regulation is represented in red; activatory regulation is depicted in black.

acquire a set of discrete values that correspond to its possible expression levels. Due to the lack of experimental data on the kinetic constants for each interaction of the network, we constructed a model that focuses on the functional state of expression of each component, rather than on their exact concentrations. These levels of expression are modeled through discrete variables that take a finite number of values. Since all the elements of the network are considered binary, our network has fourteen binary nodes giving a total of $\Omega = 2^{14}$ possible dynamical states for the network.

At every time step of the dynamics, the expression level of all components of the network is updated simultaneously according to

$$\sigma_n(t+1) = F_n(\sigma_{1n}(t), \sigma_{2n}(t), \dots, \sigma_{kn}(t)) \quad (1)$$

where $\sigma_n(t)$ represents the state of the n^{th} element of the network at time t . Here $\{\sigma_{1n}, \sigma_{2n}, \dots, \sigma_{kn}\}$ are the k_n regulators of σ_n and $F_n(\cdot)$ is a discrete function, stated as a logical rule, that explicitly states the corresponding expression level of σ_n given the current expression levels of its regulators. $F_n(\cdot)$ is constructed according to experimental evidence regarding the regulatory interactions (activator or inhibitor) for each node (see methods for references). All the functions $F_n(\cdot)$ for the estrogen network are listed in Supplementary Information 1.

Since there is a finite number of possible dynamical configurations for the entire network (Ω), starting the dynamics from any of these configurations and successively iterating (1) for each node will make the network traverse through a series of states until a periodic pattern of activity is reached. This periodic set of states is known as an attractor.

Several attractors might exist for a given network, and several initial configurations may lead to the same attractor. For a fixed set of logical functions $F_n(\cdot)$, the particular attractor the network falls into depends entirely on the initial condition the network starts from. Each attractor has a basin of attraction defined by all of the initial conditions that lead to

that particular attractor. These attractors can be thought of as stable patterns of activity of real biological systems as has been shown previously [8, 9]. In this case, the attractors reached will represent the proliferative or antiproliferative state in cells under the transcription signals triggered by estrogen.

Despite these parallelisms, a direct comparison of an attractor to real expression levels might not be so straightforward. Cyclic attractors (attractors composed by several states) are very common whereas experimental gene expression is often presented as a single value. Additionally, gene expression measurements are commonly taken from cell populations, which makes the final measurement an average of single measurements taken over a time window. For this reason we have used a modification of the classic Boolean Network approach where we average the gene expression in the whole attractor basin to end up with a single level of expression for each gene. This approach has been used previously to accurately simulate the behavior of small gene regulatory networks [12]

Here the state of each element of the network is represented by its average expression over a time lapse. The length of the window (L) where values will be averaged corresponds to the length of the attractor reached. For instance, let us assume a cyclic attractor consisting in four network states is reached. For simplicity let us assume the states are

State (1) $[1]_1[0]_2[0]_3[0]_4[0]_5[0]_6[0]_7[0]_8[0]_9[0]_{10}[0]_{11}$
 $[0]_{12}[0]_{13}[0]_{14}$,

State (2) $[0]_1[0]_2[0]_3[0]_4[0]_5[0]_6[0]_7[0]_8[0]_9[0]_{10}[0]_{11}$
 $[0]_{12}[0]_{13}[0]_{14}$,

State (3) $[1]_1[1]_2[0]_3[0]_4[0]_5[0]_6[0]_7[0]_8[0]_9[0]_{10}[0]_{11}$
 $[0]_{12}[0]_{13}[0]_{14}$,

State (4) $[0]_1[1]_2[0]_3[0]_4[0]_5[0]_6[0]_7[0]_8[0]_9[0]_{10}[0]_{11}$
 $[0]_{12}[0]_{13}[0]_{14}$.

Now, we want the averaged state of $[node]_1$, which is represented by the first digit of the network state

([1], [0], [1], [0]). Then the corresponding expression level for $node_1$ for this attractor will be $(1 + 0 + 1 + 0)/4 = 0.5$ where 4 is the size of the attractor.

Since a network can have more than one attractor, we will end up with an expression level for each of these attractors. In order to account for each of their basins of attraction and end up with a single value for the expression level of a gene, we have incorporated a weighted average using the entire set of attractors (N) for each network. We define the average expression level of σ_n as

$$\sigma_n = \sum_{a=1}^N \omega_a \left(\frac{\sum_{\tau}^{L_a} \sigma_n(\tau)}{L_a} \right) \quad (2)$$

where N represents all the different attractors, so the external sum is carried out over all the existing attractors numbered 1 to N. The parameter ω_a is the fraction of initial conditions that lead to attractor “a” over the total possible conditions, that is, the size of the basin of the a_{th} attractor over Ω . The internal sum is carried out over all the $\sigma_n(\tau)$ states of the a_{th} attractor, of size L_a . This means that the final level of expression of the gene σ_n will be the sum of the averaged expression levels of σ_n in each attractor, with each attractor average weighted by its corresponding basin of attraction.

This modification, apart from allowing an easier comparison between the model and experimental data, resembles the way in which experimental data is gathered for gene expression, where traditionally measurements of the level of expression represent the population average, as cells in the population may be at different stages of a stable pattern of gene expression.

It is important to note that mutations in our model (e.g., a deletion/malfunction of a gene) are represented by keeping the value of the deleted node equal to zero throughout all the dynamics. In the case of gene overexpression, we kept the value of the overexpressed gene equal to its maximum state over the whole simulation.

2.4. Perturbation Analysis. To simulate altered physiological states, we explored the dynamics of the network if the state of genes is perturbed.

- (i) We simulated the *overexpression* of a gene by setting it to an *ON* state at the beginning of the simulation and keeping it that way throughout the simulation regardless of the state of its regulators.
- (ii) We simulated the *knockout* of a gene by setting it to an *OFF* state at the beginning of the simulation and keeping it that way throughout the simulation regardless of the state of its regulators.

The simulation of the overexpression of a gene may represent either an increase in the activity or concentration of said gene in a phenotype and may also be used as a model of the activity of an agonist drug. Similarly, the simulation of the knockout of a gene may also be used as a model of the activity of an antagonist drug.

We analyzed the full set of single *overexpression* and *knockout* perturbations for all genes in the network. We

also analyzed the full set of 2-hit perturbations for all genes in the network, including all *overexpression/overexpression*, *overexpression/knockout*, and *knockout/knockout* pairs. This has been conducted successfully in other biological systems by members of our group, such as the calcium-dependent signaling pathway of the spermatozoa of the sea urchin *S. purpuratus*, in its searching for the egg [10].

2.5. Proliferation Index. We used the Boolean dynamics to quantify biological features in a particular phenotype. In this work, we focused on identifying whether a given phenotype may tend to be proliferative or antiproliferative. To do this, we considered whether the expression of each gene may be involved in processes that are proliferative or antiproliferative, beyond their role as regulators in the network.

We constructed a *Proliferation Index* (PI), in which we consider, for a given phenotype, the average state of each gene throughout the attractor landscapes associated to the phenotype.

$$PI = \frac{\sum \langle P \rangle - \sum \langle AP \rangle}{\sum \langle P \rangle + \sum \langle AP \rangle} \quad (3)$$

Where *P* are proliferative genes and *AP* are anti-proliferative genes and $\langle \cdot \rangle$ are the appropriate ensemble averages. In order to assess whether a gene was considered *proliferative* or *anti-proliferative*, we performed a systematic analysis of the literature using a combination of Pubmed <https://www.ncbi.nlm.nih.gov/pubmed/>, the Gene database [18], and the Genetics Home Reference <https://ghr.nlm.nih.gov/>. In Supplementary Information 2 we provide the bibliographic evidence used to assign a *proliferative* or *anti-proliferative* value to each gene in the network. Our index should not be confused with the proliferative index (or growth fraction), which is used in the clinical setting.

It is worth to mention that the Proliferation Index (PI) defined here is the result of averaging the state value of nodes during the attractor period. The PI is a measure that integrates the attractor landscape in terms of the proliferative/antiproliferative phenotype.

3. Results

3.1. The Estrogen Transcriptional Network. By following our construction methodology, we are able to recover an *estrogen receptor regulatory network* composed of 14 nodes and 25 directed and signed interactions. Four of these interactions are inhibitory, while the rest correspond to activation. A visualization of this network may be found in Figure 2.

The network dynamics of this network is depicted in Figure 3, where each dynamical state is represented as a colored point, and the transition between two consecutive states is represented as a straight line. As previously mentioned, attractors of the network dynamics may be punctual or cyclic. In the figure we observe both cases.

3.2. Effects of Perturbations on Proliferation Based on Network Dynamics. Through the use of well-curated biological knowledge along with Boolean network dynamics, we

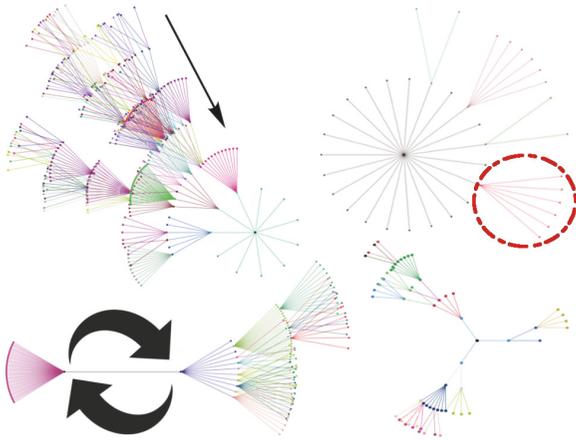


FIGURE 3: Excerpt of the attractor landscape. Fan-like representation of a set of 4 attractors of the attractor landscape of the estrogen/estrogen receptor network. Each point represents a state of the network. Connections represent temporal succession between states, with the outward points representing preceding states (see black arrow). Same colored fans represent a collection of states that lead to the same future state of the network (e.g., red dashed circle). A cyclic attractor of size = 2 is also represented (see blue and purple circular arrows). The length of the connections is inconsequential as all time steps between states are fixed.

developed a model that may elucidate the contribution of gene perturbations to an observable phenotypic trait. We focused on the proliferative state that is achieved through gene perturbation. This could reflect the changes in cell growth observed in diseases such as cancer, but it also can be used to model the effects that an external perturbation (such as a pharmacological intervention) may have in the phenotype.

3.2.1. The Effects of Single Perturbations on the Proliferative Phenotype. In Figure 4 we present the result of the perturbation of single genes in terms of the *Proliferation Index (PI)*, compared to the PI value for the wild-type (*WT*) phenotype.

A total of 28 perturbations were performed, which may be seen in Figure 4(a). Overall, 15 of these perturbations induce PI value higher than the one for the *WT* phenotype ($PI = -0.0761$), whereas 13 lead to a reduction of the PI value with respect to the *WT*. The maximum PI value is achieved through the knockout of E2F1 ($PI = 0.3029$), while the minimum PI value is achieved through the knockout of STAT5A ($PI = -0.5600$).

We may observe that the effects of gene overexpression and knockout are different in terms of the PI. In Figures 4(b) and 4(c) we may observe the PI for overexpressions and knockouts separately. In the case of overexpressions (Figure 4(b)), the PI values are less spread, ranging from -0.3670 to 0.0335 with 9 perturbations increasing the PI with respect to the *WT* and 5 decreasing it. In the case of knockouts (Figure 4(c)), these cover a broader range including the aforementioned overall maximum (E2F1 knockout) and minimum

(STAT5A knockout); 4 perturbations increase the PI with respect to *WT*, and 10 decrease it.

In Figure 4(d) we present the genes in the network in a scatterplot, where the x-axis represents PI when the gene is knocked-out, and the y-axis represents PI when the gene is overexpressed. We trace four quadrants with respect to PI for the *WT* phenotype. We may observe that all four antiproliferative genes are placed in the lower, right quadrant, indicating that their knockouts lead to more proliferation, while their overexpression leads to less proliferation.

3.2.2. Effects of Two-Hit Perturbations on the Proliferative Phenotype. In Figure 5 we present the results of the simultaneous perturbation of two genes of the Estrogen Receptor Regulatory Network in terms of the Proliferative Index as heatmaps. In Figure 5(a), we present the result of the simultaneous overexpression of two genes. In Figure 5(b), we show the effect of the combined overexpression of a gene (shown in the rows of the heatmap) and the knockdown of another gene (shown in the columns of the heatmap). Finally, in Figure 5(c), we show the effect of double gene knockouts.

For each type of two-hit perturbation, we may find a maximum and minimum PI value. In the case of the double overexpression, the minimum PI value is achieved with the overexpression of TP53 and AR ($PI = -0.5313$), while the maximum PI value is achieved with the overexpression of ESR1 and AR ($PI = 0.1876$). For overexpression/knockout combinations, the minimum PI value results from the overexpression of TP53 and the knockout of STAT5A ($PI = -0.8871$), while the maximum PI value comes from overexpressing ESR1 and knocking out SMAD4 ($PI = 0.4719$). In the case of double knockouts, knocking out both STAT5A and JUN lead to the minimum PI value ($PI = -1.0196$) while the double knockout of SMAD4 and E2F1 generates the maximum PI value ($PI = 0.8033$).

Through the double perturbation of genes, it is possible to reach more extreme changes in PI than by targeting a single gene alone. For instance, the lowest PI value obtained ($PI = -1.0196$, from the double knockout of STAT5A and JUN) is much lower than the lowest PI obtained from a single gene perturbation ($PI = -0.5600$ from the single knockout of STAT5A). Similarly, the highest PI value obtained ($PI = 0.8033$ from the double knockout of SMAD4 and E2F1) is higher than the highest PI value obtained from single perturbations ($PI = 0.3029$ from the knockout of E2F1). Importantly, and similar to what was observed in single perturbations, the most extreme changes in PI come from knockout perturbations.

4. Discussion

We have shown that with the Boolean approach it is possible to perturb the dynamical state of the estrogen transcriptional network and observe single or multitarget perturbations. As it is expected, the effect of altering one or more elements in the network dynamics will be different in terms of the Proliferation Index.

In the upper, left quadrant of the scatterplot in Figure 4(d), representing single gene perturbations, we

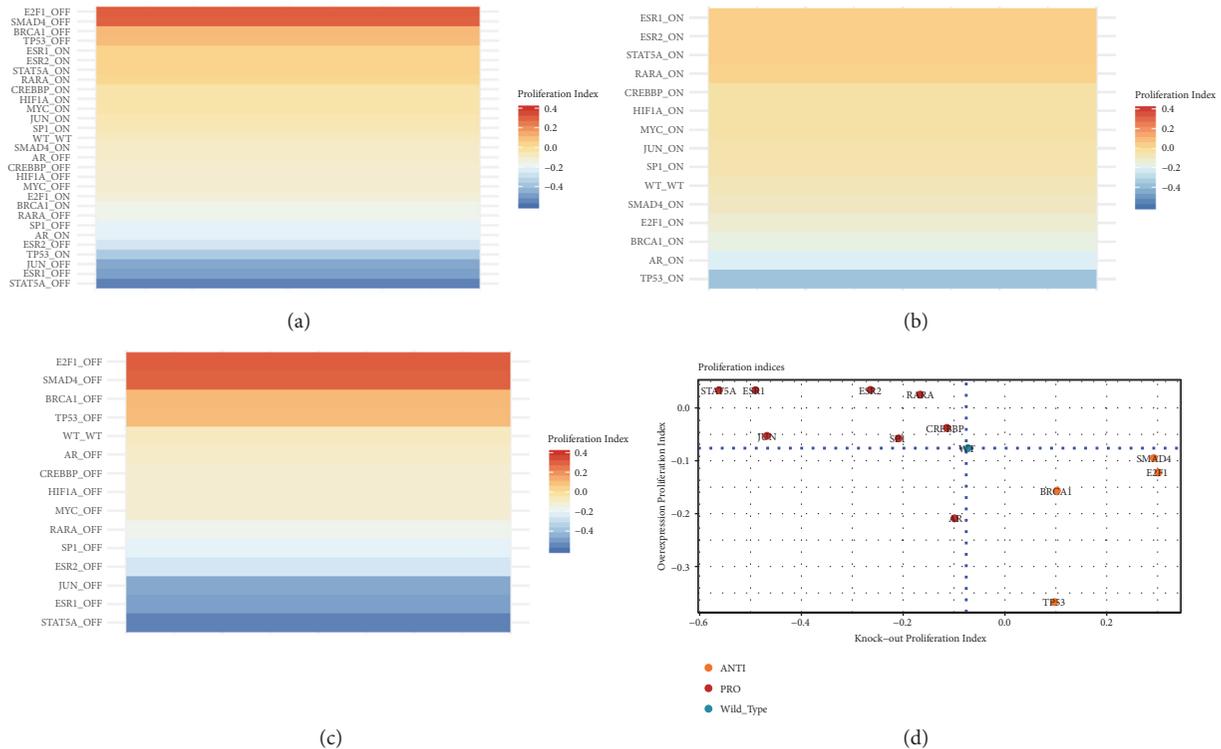


FIGURE 4: Proliferation indexes for single gene perturbations. Each column shows the PI values after a perturbation. (a) shows the set of overexpression and knockouts, (b,c) represent overexpression and knockout separately. (d) is a scatterplot showing the PI values of all genes after overexpressing and knocking out the genes.

may find proliferative genes (STAT5A, ESR2, MYC, JUN, etc.), meaning the overexpression of these genes lead to more proliferation, while their knockout leads to less proliferation. An interesting finding is the curious case of AR. This is the only gene in the network that is located in the lower left quadrant, indicating that both its overexpression and knockout lead to a decrease of the proliferative index with respect to the wild type.

The E2F1 gene is a well known tumor suppressor gene. It participates in both control of cell cycle and cell death processes. It has been observed experimentally that lower expression values of E2F1 gene are frequent in malignant tumors in breast cancer [19]. As in our network dynamics, the highest PI value was obtained by knocking out E2F1 gene, which is in agreement with the experimental results. Analogously, as STAT5A, being one of the main activators of ESR1 and ESR2, its inhibition decreases substantially PI .

By observing Figure 5, representing two-hit perturbations, it is evident that each type of perturbation generates different clustering patterns. It may be seen that in the case of double overexpressions, we observe a more homogeneous distribution of the PI values. In the case of the overexpression/knockout combinations, the PI patterns tend to be more dominated by the knockout genes (as the pattern observed is of vertical stripes). Finally, in the case of double knockouts, we may find well defined clusters that are related to the double knockout of antiproliferative genes, proliferative genes, or the combination of a proliferative and antiproliferative gene.

It is worth noting that the lowest PI value results from the concerted action of the overexpression of a keystone tumor suppressor (TP53) and the concomitant knockout of a proproliferative gene (STAT5A), whose single knockout leads to the lowest individual PI value. The aforementioned results may have important implications in different scenarios, such as cancer, where drug combinations may have deep impact in clinical outcomes.

Interestingly, the resulting network, after causal inference, contains Master Regulators such as P53, E2F, SMAD4, STAT5A, AR, ESR1, MYC, FOS, or JUN. It is well known that these Master Regulators determine the cell phenotype in health and disease and its deregulation may have profound implications, in cases such as cancer [20].

PI was constructed acknowledging the pro- and antiproliferative activity of said regulators. The relevance of having Master Regulators in our network is that the “fine tuning” of them would imply the switch to a proliferative state or a cell cycle arrest.

The Boolean approach used here has several advantages, such as the fast and direct set of results that are obtained by a relatively simple model. It is not necessary to know the reaction rates or other biological parameters that are often difficult to obtain experimentally. Perturbation analysis is also easy to obtain and interpret. Another advantage is the possibility to perturb more than one molecule *in silico* and analyze results in terms of transient times, attractor landscapes, or basins of attraction.

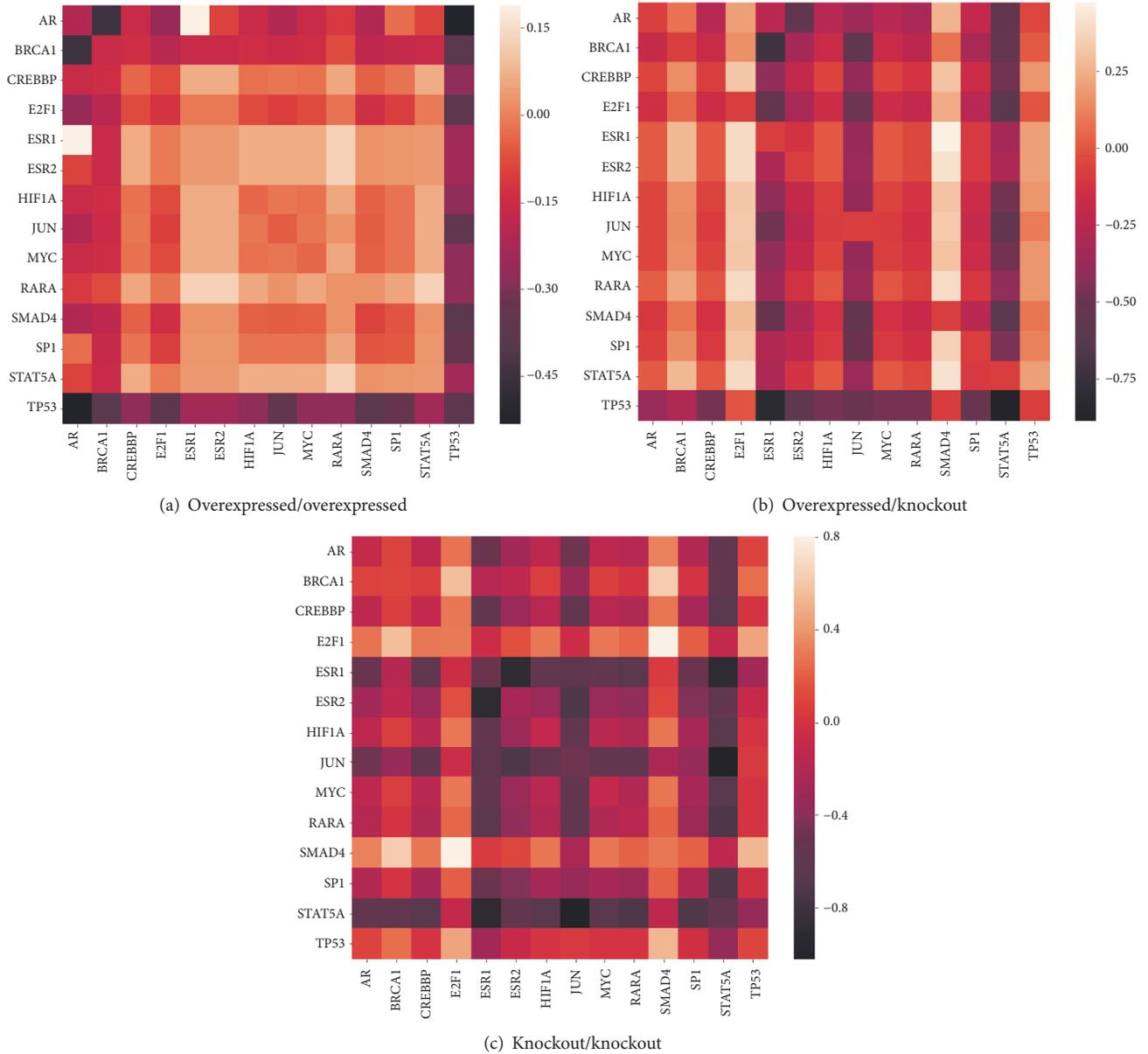


FIGURE 5: Proliferation indexes for two-hit perturbations. The three heatmaps show the PI value for perturbation of a couple of genes in the network.

However, these kinds of models also present some issues that must be taken into account to have a better interpretation: the model only uses two discrete states, losing the fine-tuning of studying the system as a continuous model. Time evolution is also discrete, but it is widely known that biological molecules have a particular time for reaction. Despite the fact that the Boolean model uses a discrete time evolution, this does not significantly differ from an attractor landscape obtained by a nonsynchronous update evolution dynamics.

All of these caveats obviously have influence on the interpretation, but after a careful construction of the dynamical rules of the network, the results of the Boolean dynamics are a good generator of hypotheses and may be used as a first step in the searching for experimental corroborations.

5. Conclusions

In this work we have demonstrated *in silico* that altering the dynamical state of key biomolecules of the proliferative estrogen regulatory networks is possible to shift the dynamical state from a proproliferative towards an antiproliferative one and vice versa. The Proliferation Index presented here, despite being similar to other indexes used in cancer-related Boolean networks [21], provides elements of analysis and suggests possible experimental approaches in terms of altering the estrogen-dependent cell proliferation.

These kinds of approaches may be useful to test the usage of different drugs with a known or unknown effects and evaluate the final outcome searching for a more personalized medicine.

Appendix

Boolean Networks as Dynamical Systems

This brief appendix provides some definitions of dynamical systems and Boolean networks included for the sake of completeness. The study of the dynamical evolution of networked systems has been gaining importance and recognition in the physics/applied mathematics/complex systems/computer sciences literature. This is so since a wide variety of non-trivial phenomena has been characterized as arising of the dynamic evolution of interdependent agents. Features like cooperation, spreading, and synchronization dynamics on networks have been characterized. For instance, the work of Wang and coworkers [22] presents an application of novel centrality measures to account for modified diffusion (spreading) on complex networks while information sharing and cooperation have been characterized in the works of the Chengyi group [23, 24].

Boolean networks, in particular, are a class of (deterministic or stochastic) sequential dynamical systems. Boolean networks usually consist on a (finite) set of Boolean logic variables governed by a set of finitary functions of the form $\mathcal{F} : B^k \rightarrow B$, where $B = \{0, 1\}$ is a binary logic or *Boolean domain* (e.g., an algebra of logical truth values) (it is possible to build Boolean domains with more than two logical states. The formalism extension to these cases is straightforward.) and k is the *arity* (number of arguments or Cartesian product dimension), a nonnegative integer. A Boolean function is thus a propositional formula in k variables which takes a series of inputs from a subset of the Boolean variables and as an output produces the state of the corresponding variable. The set of Boolean functions determines the connectivity of the set of variables that become the nodes of a network, whose topology is given by the combination of Boolean functions for all the variables [25, 26].

For the Boolean network dynamics, the state of the network at a given time $t + 1$ is determined via the evaluation of each of the variables' function on the state of the network at a previous time t . This may be done on a *synchronous* (all nodes' states updated at once) or *asynchronous* (hierarchical updating given the position of a given node in the network) way. Depending upon updating procedures, the system's dynamics may be Markovian or non-Markovian (often finite-Markovian) [26, 27].

Given the fact that Boolean networks are discrete dynamical systems with finite support (there are exactly 2^N possible states on a classical—i.e., 2-state—Boolean network with N nodes), the evolution of the system will produce recurrent states. The trajectories will fall into one of a set of steady states or cycles called *attractors*. The set of attractors of a dynamical system is called the *attractor landscape*. The determination of the set of attractor states and the convergence dynamics leading to those attractors constitutes the solution to the Boolean network dynamics problem [27].

The Boolean networks studied here belong to a class of deterministic dynamical systems. Such systems may be represented by a set of differential equations describing the dynamical evolution in phase space. Deterministic Boolean

networks may also be represented as a discrete dynamical system (a map) that when iterated reproduces the full dynamics of the network including the set of attractors. This was the way we proceeded here. Since iterated maps and differential equations are two equivalent representations of the evolution of a dynamical system [28], our approach does not lose any generality.

Data Availability

All relevant data has been included in the supplementary materials.

Conflicts of Interest

The authors have no conflict of interest to declare.

Authors' Contributions

Guillermo de Anda-Jáuregui, Jesús Espinal-Enríquez, and Santiago Sandoval-Motta contributed equally to this work.

Acknowledgments

The research leading to these results has received funding from Consejo Nacional de Ciencia y Tecnología (grant number 285544/2016, Ciencia Básica, and 2115/2017, Fronteras de la Ciencia (Jesús Espinal-Enríquez)), as well as federal funding from the National Institute of Genomic Medicine (Enrique Hernández-Lemus). Enrique Hernández-Lemus also acknowledges support from the 2016 Marcos Moshinsky Research Chair in the Physical Sciences. Jesús Espinal-Enríquez acknowledges support from Fundación Miguel Alemán in Health Research. Santiago Sandoval-Motta acknowledges support from the program Cátedras CONACYT. The funders had no role in the design of this research.

Supplementary Materials

Supplementary 1. Supplementary Material 1. Regulatory functions of the estrogen transcriptional networks. Each file contains the regulatory function for all those genes in the network, including the regulatory genes, as well as the discrete value of the target gene after taking into account the value of its regulators.

Supplementary 2. Supplementary Material 2. Bibliographic evidence associated with the proliferative and antiproliferative nature of the genes in the network.

References

- [1] W. E. Stumpf, "Nuclear concentration of 3H-estradiol in target tissues. Dry-mount autoradiography of vagina, oviduct, ovary, testis, mammary tumor, liver and adrenal," *Endocrinology*, vol. 85, no. 1, pp. 31–37, 1969.
- [2] J. Cui, Y. Shen, and R. Li, "Estrogen synthesis and signaling pathways during aging: from periphery to brain," *Trends in Molecular Medicine*, vol. 19, no. 3, pp. 197–209, 2013.

- [3] F. Pedoutour, B. J. Quade, S. Weremowicz, P. Dal Cin, S. Ali, and C. C. Morton, "Localization and expression of the human estrogen receptor beta gene in uterine leiomyomata," *Genes, Chromosomes and Cancer*, vol. 23, no. 4, pp. 361–366, 1998.
- [4] L. Giacinti, P. P. Claudio, M. Lopez, and A. Giordano, "Epigenetic information and estrogen receptor alpha expression in breast cancer," *The Oncologist*, vol. 11, no. 1, pp. 1–8, 2006.
- [5] G. De Anda-Jáuregui, R. A. Mejía-Pedroza, J. Espinal-Enríquez, and E. Hernández-Lemus, "Crosstalk events in the estrogen signaling pathway may affect tamoxifen efficacy in breast cancer molecular subtypes," *Computational Biology and Chemistry*, vol. 59, pp. 42–54, 2015.
- [6] P. Ascenzi, A. Bocedi, and M. Marino, "Structure-function relationship of estrogen receptor α and β : Impact on human health," *Molecular Aspects of Medicine*, vol. 27, no. 4, pp. 299–402, 2006.
- [7] J. Thakar, M. Pilione, G. Kirimanjeswara, E. T. Harvill, and R. Albert, "Modeling systems-level regulation of host immune responses," *PLoS Computational Biology*, vol. 3, no. 6, Article ID e109, 2007.
- [8] A. Saadatpour, R.-S. Wang, A. Liao et al., "Dynamical and structural analysis of a t cell survival network identifies novel candidate therapeutic targets for large granular lymphocyte leukemia," *PLoS Computational Biology*, vol. 7, no. 11, Article ID e1002267, 2011.
- [9] J. Espinal, M. Aldana, A. Guerrero, C. Wood, A. Darszon, and G. Martínez-Mekler, "Discrete dynamics model for the speract-activated Ca²⁺ signaling network relevant to sperm motility," *PLoS ONE*, vol. 6, no. 8, Article ID e22619, 2011.
- [10] J. Espinal-Enríquez, A. Darszon, A. Guerrero, and G. Martínez-Mekler, "In Silico determination of the effect of multi-target drugs on calcium dynamics signaling network underlying sea urchin spermatozoa motility," *PLoS ONE*, vol. 9, no. 8, Article ID e104451, 2014.
- [11] J. Espinal-Enríquez, D. A. Priego-Espinosa, A. Darszon, C. Beltrán, and G. Martínez-Mekler, "Network model predicts that CatSper is the main Ca²⁺ channel in the regulation of sea urchin sperm motility," *Scientific Reports*, vol. 7, no. 1, article no. 4236, 2017.
- [12] S. Pérez-Landero, S. Sandoval-Motta, C. Martínez-Anaya et al., "Complex regulation of Hsf1-Skn7 activities by the catalytic subunits of PKA in *Saccharomyces cerevisiae*: Experimental and computational evidences," *BMC Systems Biology*, vol. 9, no. 1, article no. 42, 2015.
- [13] S. Barbosa, B. Niebel, S. Wolf, K. Mauch, and R. Takors, "A guide to gene regulatory network inference for obtaining predictive solutions: Underlying assumptions and fundamental biological and data constraints," *BioSystems*, vol. 174, pp. 37–48, 2018.
- [14] B. A. McGregor, S. Eid, A. E. Rumora et al., "Conserved transcriptional signatures in human and murine diabetic peripheral neuropathy," *Scientific Reports*, vol. 8, no. 1, 2018.
- [15] S. Barman and Y.-K. Kwon, "A novel mutual information-based Boolean network inference method from time-series gene expression data," *PLoS ONE*, vol. 12, no. 2, Article ID e0171097, 2017.
- [16] Z.-P. Liu, C. Wu, H. Miao, and H. Wu, "RegNetwork: An integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse," *Database*, vol. 2015, pp. 1–12, 2015.
- [17] A. Krämer, J. Green, J. Pollard, and S. Tugendreich, "Causal analysis approaches in ingenuity pathway analysis," *Bioinformatics*, vol. 30, no. 4, pp. 523–530, 2014.
- [18] N. A. O'Leary, M. W. Wright, J. R. Brister et al., "Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation," *Nucleic Acids Research*, vol. 44, no. 1, pp. D733–D745, 2016.
- [19] D. Worku, F. Jouhra, G. W. Jiang, N. Patani, R. F. Newbold, and K. Mokbel, "Evidence of a tumour suppressive function of E2F1 gene in human breast cancer," *Anticancer Research*, vol. 28, no. 4 B, pp. 2135–2139, 2008.
- [20] H. Tovar, R. García-Herrera, J. Espinal-Enríquez, and E. Hernández-Lemus, "Transcriptional master regulator analysis in breast cancer genetic networks," *Computational Biology and Chemistry*, vol. 59, pp. 67–77, 2015.
- [21] J. Espinal-Enríquez, R. A. Mejía-Pedroza, and E. Hernández-Lemus, "A Boolean network model for invasive thyroid carcinoma," in *Proceedings of the Artificial Life Conference 2016*, pp. 570–577, Cancun, Mexico, July 2016.
- [22] J. Wang, C. Li, and C. Xia, "Improved centrality indicators to characterize the nodal spreading capability in complex networks," *Applied Mathematics and Computation*, vol. 334, pp. 388–400, 2018.
- [23] C. Xia, X. Li, Z. Wang, and M. Perc, "Doubly effects of information sharing on interdependent network reciprocity," *New Journal of Physics*, vol. 20, no. 7, Article ID 075005, 2018.
- [24] C. Chen, Y. Hu, and L. Li, "NRP1 is targeted by miR-130a and miR-130b, and is associated with multidrug resistance in epithelial ovarian cancer based on integrated gene network analysis," *Molecular Medicine Reports*, vol. 13, no. 1, pp. 188–196, 2016.
- [25] M. Leone, A. Pagnani, G. Parisi, and O. Zagordi, "Finite size corrections to random Boolean networks," *Journal of Statistical Mechanics: Theory and Experiment*, no. 12, Article ID P12012, 2006.
- [26] B. Derrida and Y. Pomeau, "Random networks of automata: A simple annealed approximation," *EPL (Europhysics Letters)*, vol. 1, no. 2, pp. 45–49, 1986.
- [27] U. Bastolla and G. Parisi, "The modular structure of Kauffman networks," *Physica D: Nonlinear Phenomena*, vol. 115, no. 3–4, pp. 219–233, 1998.
- [28] M. W. Hirsch, R. L. Devaney, and S. Smale, *Differential Equations, Dynamical Systems, and Linear Algebra*, vol. 6, Academic Press, New York, NY, USA, 1974.

