OPTIMIZING THERAPEUTIC TARGETS FOR BREAST CANCER USING BOOLEAN NETWORK MODELS

Domenico Sgariglia[1], Flavia Raquel Gonçalves Carneiro[2,3,4], Luis Alfredo Vidal de Carvalho[5], Carlos Eduardo Pedreira[1], Nicolas Carels[6], Fabricio Alves Barbosa da Silva[7]

[1]Engenharia de Sistemas e Computação, COPPE-UFRJ, Rio de Janeiro, Brazil, [2]Center of Technological Development in Health (CDTS), FIOCRUZ, Rio de Janeiro, Brazil, [3]Laboratório Interdisciplinar de Pesquisas Médicas Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro, Brazil, [4]Program of Immunology and Tumor Biology, Brazilian National Cancer Institute(INCA), Rio de Janeiro 20231050, Brazil, [5]Faculdade de Medicina, UFRJ, Rio de Janeiro, Brazil, [6]Platform of Biological System Modeling, Center of Technological Development in Health (CDTS), FIOCRUZ, Rio de Janeiro, Brazil, [7]Scientific Computing Program (PROCC), FIOCRUZ, Rio de Janeiro, Brazil

Keywords: Boolean networks, systems biology of cancer, gene regulatory network analysis, epigenetic landscape attractors, apoptosis

ABSTRACT

Since cancer is a genetic disease, studying gene regulatory networks related to this pathology offers the possibility of obtaining helpful information for therapeutic purposes. However, the complexity expressed by the interconnections between network components grows exponentially with the number of genes in the system. In this report, the Boolean logic's use for regulating the existing relationships between network components has allowed us to simplify the modeling capable of producing attractors representing the cell phenotypes from breast cancer RNA-seq data. From this point of view, a therapeutic objective is to induce the cell, through appropriate interventions, to leave the current cancer attractor to migrate toward another attractor physiologically different from cancer. In this report, we developed a computational method that identifies network nodes whose inhibition achieves the therapeutic objective of transitioning from one tumor attractor to another associated with apoptosis based on transcriptomic data from cell lines. We used previously published *in vitro* experiments, in which the knockdown of a few proteins led a breast cancer cell line to death, to validate the model taking into account genes related to apoptosis in cancer cells. The method proposed in this manuscript uses heterogeneous data integration, structural network analysis, and biological knowledge to identify potential targets of interest for breast cancer on a personalized medicine approach.

INTRODUCTION

Cancer is a disease mainly characterized by uncontrolled cell growth. The characteristic of this deregulation is to modify the cellular homeostasis-inducing features called "hallmarks of cancer" common to all tumors [1]. Most human body tissues can be affected by this pathology, with approximately 85% of cancers occurring in epithelial cells called carcinomas. When affecting glandular tissue, they are called adenocarcinomas. Breast cancer falls into the adenocarcinoma category. In terms of incidence, it represents the neoplastic pathology that most affect women. Globally, the incidence of this disease was, in 2020, approximately 2,261,419 new cases and 684,996 deaths worldwide [2].

Furthermore, treating this pathology frequently produces deleterious side effects for patients. For example, one study found that patients suffered from 38 negative symptoms grouped into 5 clusters resulting from undergoing chemotherapy [3]. Thus, it is essential to develop new intervention strategies capable of improving therapies and limiting their adverse collateral effects. With this aim, we propose a Boolean modeling approach for breast cancer. Cancer is a genome disease with multifactorial consequences [4]. Cancer cells' DNA presents several alterations due to the oncogenic process. Examples are mutations at single base pairs, indels, and epigenetic marks.

Because the hallmarks of cancer are related to several cell pathways involving many genes, cancer induces network modifications. Furthermore, these pathways are often not independent but intertwined in forming a single process represented by tumor dynamic and progression [5-7]. Alterations in the epigenetic landscape and gene regulation network [8] offer the opportunity of modeling cancer attractors [9]. This report extends a previous analysis [10] of attractors identified in a gene regulatory network based on breast cancer data. More specifically, by integrating the Boolean network with a new set of genes related to apoptosis, the new attractors found as a result of target inactivation allowed our gene model to evolve toward a cell death phenotype as seen in the bench experiment *in vitro* [11].

In this paper, we described an algorithm to optimize the choice of network elements capable of inducing trajectories among attractors in the epigenetic landscape. We also defined an indicator that allowed us to measure the effect on the network when inducing a trajectory from a malignant to an apoptosis state by direct intervention on its vertices. We added a set of genes representative of the apoptosis process to the gene regulatory network involved in breast malignancy. The state of activation or inhibition of each gene of this group allowed us to assess whether network perturbations successfully forwarded its phenotype from a malignant to an apoptosis state. We calibrated the network based on (i) the gene expression level typical of the malignant attractor and (ii) the gene to be inhibited for inducing a malignant cell line into apoptosis, as known from an *in vitro* experiment. The result obtained in the laboratory [11], in which a breast cancer cell line was induced to death by silencing five genes, was used to verify the system dynamics according to its perturbation by permanently forcing to zero (silencing) specific targets.

We verified that the network structure obtained from the interactome was compatible with forwarding the malignant attractor toward apoptosis by forcing to zero the same network vertices as those silenced in the *in vitro* experiment. The ability of our model to reproduce the conditions favorable to malignant cell death obtained *in vitro* allowed us to optimize the target choice to forward the system dynamics from the malignant attractor to the apoptosis one. We performed this optimization through specific analysis techniques of the network structure, which allowed us to

identify the vertices whose inhibition could emulate and optimize the results obtained in the *in vitro* experiment.

MATERIALS AND METHODS

The different steps that have allowed the realization of this research are indicatively summarized in the following figure:

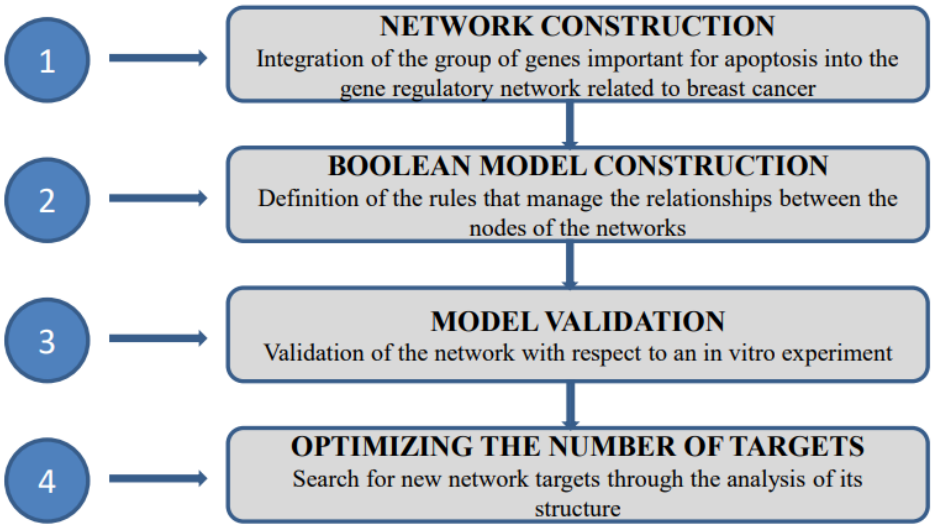

Fig. 1: Steps for Boolean network construction and dynamic simulation.

1 NETWORK CONSTRUCTION

We used a gene regulation network implemented in a previous report as a starting point [10]. The choice of network nodes was based on breast cancer RNA-seq data. The selected genes were associated with their respective hallmarks in the MSigDB repository. The two critical hallmarks of cancer, i.e., "UNLIMITED REPLICATIVE POTENTIAL" and "EVASION OF CELL DEATH", were peculiarly well represented. We added 28 new genes to the initial network of 103 genes [10]. Twenty-five of the newly introduced nodes are apoptosis-related genes, inducing or inhibiting this mechanism. In addition to these 25 new vertices, two others already present in the previous study also play an essential role in cellular apoptosis as part of the apoptotic cascade. We will call these 27 genes *apoptosis-related genes*. This update was necessary to model the transition from the malignant state to that of apoptosis induced by network perturbation through specific vertice inactivation. Protein-protein interactions were obtained from the IntAct interactome (IntAct database, version updated in December 2017, accessed on January 11, 2018, at ftp://ftp.ebi.ac.uk/pub/databases/intact/current/psimitab/intact-micluster.txt) to set the network connection between genes. The direction of the connections and their nature (activation or inhibition) were identified by reference to the Metacore database [12]. To validate the network for a biological situation tested *in vitro*, we have set up our system in such a way as to reproduce the result of Tilli

et al. [11], who obtained the death of a cancer cell line (MDA-MB-231) by inhibiting five genes through RNA interference.

For this purpose, we downloaded the RNA-seq data of two cell lines (MCF10A and MDA-MB-231) from the Gene Expression Omnibus (GEO) repository (https://www.ncbi.nlm.nih.gov/gds/). MDA-MB-231 is a malignant cell line from triple-negative breast cancer, while MCF10A is a non-tumoral cell line that we considered as the control in this report. For MCF10A, we downloaded the following RNA-seq: SRR2149928, SRR2149929, SRR2149930, SRR2870783, SRR2872995 and for MDA-MB-231, the RNA-seq: ERR493677, ERR493680 corresponding to the *body* portion, and ERR493678, ERR493679 corresponding to the *protrusion* portion of the cells, respectively. According to the GEO repository, cells were grown on a 3-micrometer pore polycarbonate transwell filter. The ones that crossed the filter and allowed to form protrusions through the pores for 2 hrs. After this time, the cells were washed, and each side of the filter was lysed for RNA preparation. This protocol fractionated the cells into protrusion and cell-body fractions. Below, we binarized the up-regulated genes for each MDA-MB-231 RNA-seq of the *body* and *protrusion* with every MCF10A RNA-seq to produce variability in the experiment and analyze system robustness.

We used the degree distribution of vertices to verify the convergence of our network model with scale-free networks, which are typical of cell signaling pathways [19]. To assess the structure of our network, we compared its degree distribution with random graphs [27], Watts and Strogatz small-world networks [28], and scale-free networks [19], which were generated from the same number of vertices. To compare these distributions, we used the complementary cumulative distribution function (CCDF) defined by equation 1, which gives the probability F of a vertex having a given degree of connectivity or higher.

$$\boldsymbol{F(k)} = \sum_{\boldsymbol{k'=k}}^{\infty} \quad \boldsymbol{P(k')} \tag{1}$$

2 BOOLEAN MODEL CONSTRUCTION

Having implemented the directed graph of our model, we defined the conditions that allowed its dynamic modeling by describing the transfer functions required for the system's evolution in discrete time steps. The goal was to direct the dynamic evolution of the elements of the network in such a way as to consistently reproduce the conditions detected in the *in vitro* experiment of Tilli et al. [11].

Inferring the rules of interaction of a network in a realistic way is one of the most difficult challenges in systems biology. To tackle this task, we used boolean nested canalizing functions [13], through which the variables positioned in a specific order dominate the function. We adopted the hierarchization of transfer functions (i.e., by using nested canalizing functions) because it approximates the biological system behavior [14,15]. In addition, many network nodes presented a considerable number of inputs, and the connection degree of its vertices followed a power law. In this situation, nested canalizing functions provide more stability to the system [16], essential for controlling the noise typical of biological systems. The use of nested canalizing functions and the necessity to tune the model conformity with the biological facts let us manually set the transfer rules for each gene [16,17]. Harris et al. [18] showed that most of the updating rules of the genes belonged to the class of canalizing functions. A Boolean function is canalizing if only one input can determine the output. If this input does not play the canalizing role, the other inputs are considered

to perform this function. According to these considerations, we define the number of Boolean function inputs given in Equation 2.

$$\boldsymbol{f(x) = f(x_1, \dots., x_z)} \quad (\mathbf{2})$$

Considering a set of Boolean variables $\boldsymbol{x = \{x_1, x_2, \dots, x_z\}}$, the input was defined essential (see equation 3) if the condition below was satisfied:

$$\boldsymbol{f(x_1, \dots, x_{i-1}, 0, x_{i+1}, \dots, x_z) \neq f(x_1, \dots, x_{i-1}, 1, x_{i+1}, \dots, x_z)} \quad \mathbf{(3)}$$

The essential inputs were defined as *canalizing* if there were values $\boldsymbol{a, b} \in \{\mathbf{0}, \mathbf{1}\}$ that satisfied equation 4:

$$\boldsymbol{f(x_1, \dots, x_{i-1}, a, x_{i+1}, \dots, x_z) = b} \quad (\mathbf{4})$$

for all remaining combinations of variables $\boldsymbol{x - \{x_i\}}$, where x is a canalizing input value and x_i is a canalized value.

A function with $\boldsymbol{z}$ essential input is defined as nested if it is z-times canalized with $\boldsymbol{x_1, \dots, x_z}$ canalizing inputs and $\boldsymbol{a_1, \dots, a_z}$ canalizing values, to which correspond the canalized value $\boldsymbol{b_1, \dots, b_z}$.

Nested canalizing functions are an extension of the canalizing function formalism [13], in which the order of the inputs is considered to assign the canalizing role.

Canalizing functions can be nested if it is possible to set them with $\boldsymbol{z}$ inputs and $\boldsymbol{z - 1}$ Boolean operators AND ($\wedge$) or OR ($\vee$) with a priority proceeding from left to right. Defining ($* \in \{\wedge, \vee\}$), we have equation 5:

$$\boldsymbol{f_i(x_1, x_2, \dots, x_z) = x_{i1} * \left(x_{i2} * \left(\dots (x_{iz-1} * x_{iz+1})\right)\right)} \quad (\mathbf{5})$$

In the implementation phase of this report, each input of the function was coupled to a single logical operator, which can be an $\wedge$ (and) or an $\vee$ (or). In light of these rules, the transfer functions of the network have been implemented according to Equation 6

$$\boldsymbol{node_c = (a_1 \vee a_2 \vee \quad \dots \vee a_n) \wedge \neg\left(b_1 \wedge \neg\left(b_2 \wedge \neg\left(\dots \wedge \neg(b_n)\right)\right)\right)} \quad (\mathbf{6})$$

where $\boldsymbol{node_c}$ is the nth node of the network, $\boldsymbol{a_i}$, $\boldsymbol{i = [1, n]}$, $\boldsymbol{n}$ represents the number of nodes with activation function on $\boldsymbol{node_c}$, and $\boldsymbol{b_j, j = [1, m]}$, $\boldsymbol{m}$ is the number of nodes with inhibition function on $\boldsymbol{node_c}$. To create suitable conditions for the network to reproduce the results obtained in the *in vitro* experiment [11], we applied some changes to the general scheme of the nested canalizing functions illustrated above in the nodes representing the *TP53*, *HIF1A*, *RELA*, *NFKB1*, *HDAC1*, *STAT3*, *BCL2*, *CASP3*, and *BRCA1* genes as shown in equations 7 to 11:

$$\boldsymbol{node_{TP53,HIF1A,CASP3} = (a_1 \vee a_2 \vee, \dots, \vee a_n) \wedge (\neg b_1) \wedge (\neg b_2 \vee, \dots, \vee \neg b_m)} \quad (7)$$

$$\boldsymbol{node_{RELA}} = (\boldsymbol{a_1}) \wedge (\boldsymbol{a_2} \vee, \dots, \vee \boldsymbol{a_n}) \wedge (\neg \boldsymbol{b_1}) \wedge (\neg \boldsymbol{b_2} \vee \ , \dots, \vee \neg \boldsymbol{b_m}) \tag{8}$$

$$\boldsymbol{node_{NFKB1}} = (\boldsymbol{a_1}) \wedge \left(\boldsymbol{a_2} \vee \ , \dots, \boldsymbol{a_n}\right) \wedge (\neg \boldsymbol{b_1} \vee \neg \boldsymbol{b_2} \vee \ , \dots, \vee \neg \boldsymbol{b_m}) \tag{9}$$

$$\boldsymbol{node_{HDAC1}} = (\boldsymbol{a_1}) \wedge (\boldsymbol{a_2} \vee \ , \dots, \vee \boldsymbol{a_n}) \tag{10}$$

$$\boldsymbol{node_{STAT3,BCL2,BRCA1}} = \left(\boldsymbol{a_1} \vee \boldsymbol{a_2} \vee_{,} \dots, \vee \boldsymbol{a_n}\right) \wedge (\neg \boldsymbol{b_1} \vee \neg \boldsymbol{b_2} \vee_{,\dots,} \vee \neg \boldsymbol{b_n}) \tag{11}$$

Where some input variables with activation or inhibition roles on $node_c$ ceased to be independent of the other elements of the function and assume a cumulative role for the final result, which the other nodes cannot replace.

Once the constitutive elements of the gene regulation network were defined, and its structure analyzed, the next step was to establish the type of mathematical formalism to be used for the dynamic analysis of the system. The choice was to use a directed graph model based on Boolean logic. Boolean network modeling is among the simplest methods for dynamic modeling, but at the same time, with the characteristics of being reliable in providing insights into the dynamics of a system.

Considering a Boolean variable $\boldsymbol{B}$, it takes the value True (1) or False (0) according to whether a given gene is up-regulated or not in the RNA-seq of MDA-MB-231 (malignant) compared to MCF10A (control) cell lines. Thus, for the $\boldsymbol{n}$ vertices of our network, we have equation 12.

$$\boldsymbol{X} = \{\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}, \dots., \boldsymbol{x_n}\}, \ \ \boldsymbol{x_i} \in \boldsymbol{B} \tag{12}$$

If time is described as a discrete scalar value, we represent the states of the network as a vector whose components are its vertices (equation 13)

$$\vec{\boldsymbol{x}} = \left(\boldsymbol{x_1}(\boldsymbol{t}), \dots\dots, \boldsymbol{x_n}(\boldsymbol{t})\right) \tag{13}$$

and system state space trajectories depend on the Boolean functions of the $\boldsymbol{n}$ vertices of the network (equation 14).

$$\boldsymbol{x_i}(\boldsymbol{t}+\boldsymbol{1}) = \boldsymbol{f_i}\left(\vec{\boldsymbol{x}}(\boldsymbol{t})\right), \boldsymbol{f_i}: \boldsymbol{B^n} \rightarrow \boldsymbol{B} \tag{14}$$

In this report, we adopted a simultaneous update of the network vertices, called synchronous update mode. An asynchronous update turns out to be more realistic from a biological point of view. Still, the update mode should not be essential, considering the greater computational and conceptual convenience of synchronous updates and the improved system stability by using nested canalized transfer functions [21]. The synchronous update occurs in consecutive temporal progress states (equation 15).

$$\boldsymbol{T}\left(\vec{\boldsymbol{x_i}}^{\boldsymbol{t}}, \vec{\boldsymbol{x_i}}^{\boldsymbol{t+1}}\right) = \boldsymbol{T_1}\left(\boldsymbol{x_1^t}, \boldsymbol{x_1^{t+1}}\right) \wedge \dots \wedge \boldsymbol{T_n}\left(\boldsymbol{x_n^t}, \boldsymbol{x_n^{t+1}}\right) \tag{15}$$

3 MODEL VALIDATION

First, we compared MDA-MD-231 RNA-seq samples (two *body* and two *protrusion*) to every MCF10A one. For this, we applied the process of RPKM for RNA-seq normalization as described in Pires et al. [20] to the read counts of the twenty paired RNA-seq samples. Then, we subtracted each normalized value of MCF10A RNA-seq sample from that of each MDA-MB-231 data and applied a log transformation according to equation 12 for positive values (malignant up-regulated genes) using the pipeline described by Pires et al. [20].

$$y = x * Log_2(x+1) \tag{16}$$

This pipeline classifies genes as up-regulated when the log transform of differential expression is greater than a critical value. This critical value was obtained by fitting a Gaussian curve with a 95% degree of confidence on data for a p-value of 0.025 using a Python script (see Pires et al. 2021). For each of the twenty MBA-MD-231 vs. MCF10A comparisons, all genes classified as up-regulated according to this process were set to "1" and the other to "0" (**Supplementary Table S1**).

Input : V nodes of the Boolean network; N cell lines with body-fraction MDA-MD 231 RNA-Seq values ($B_{n,v}$); M cell lines with RNA-Seq values MDA-MD 231 of protrusion fraction ($P_{m,v}$); C cell lines of MCF10A type RNA-Seq values ($CO_{c,v}$);

Output : Boolean values of RNA-Seq

1 ***for*** *all* $n \in N$ ***do***
2 ***for*** *all* $v \in V$ ***do***
3 $B_{n,v} \leftarrow RPKM_{upper} = \frac{RC_g * 10^9}{L*(RC_{pc} - (\delta * RC_{pc}))}$
4 ***for*** *all* $m \in M$ ***do***
5 ***for*** *all* $v \in V$ ***do***
6 $P_{m,v} \leftarrow RPKM_{upper} = \frac{RC_g * 10^9}{L*(RC_{pc} - (\delta * RC_{pc}))}$
7 ***for*** *all* $c \in C$ ***do***
8 ***for*** *all* $v \in V$ ***do***
9 $CO_{c,v} \leftarrow RPKM_{upper} = \frac{RC_g * 10^9}{L*(RC_{pc} - (\delta * RC_{pc}))}$
10
11 ***for*** *all* $v \in V$ ***do***
12 ***for*** *all* $n,m \in (N, M)$ ***do***
13 ***for*** *all* $c \in C$ ***do***
14 $gene_diff_{v,n,c} \leftarrow B_{n,v} - CO_{c,v}$
15 $gene_diff_{v,m,c} \leftarrow P_{m,v} - CO_{c,v}$
16 $G_{v,n,c} \leftarrow gene_diff_{v,n,c} + lg_2(gene_{diff_{v,n,c}} + 1)$
17 $G_{v,m,c} \leftarrow gene_diff_{v,m,c} + lg_2(gene_{diff_{v,m,c}} + 1)$

Block 1
Apply the process of RPKM for RNA-seq normalization for all genes in the network. MDA-MD 231 body (lines 1-3), MDA-MD 231 protrusion (lines 4-6), MCF10A (lines 7-9).

Block 2
Subtract each normalized value of MCF10A RNA-seq from that of each MDA-MB-231 (lines 11-15), both for body (line 14) and protrusion (line 15). A log transform is applied (body type line 16, Protrusion line 17)

18

19

20 *P-value* ← *0.025*

21 ***for*** *all n,m ∈ (N, M)* ***do***

22 ***for*** *all c ∈ C* ***do***

23 $\mu_{n,c} = mean_{n,c}$

24 $\mu_{m,c} = mean_{m,c}$

25 $\sigma_{n,c} = standard_deviation_{n,c}$

26 $\sigma_{m,c} = standard_deviation_{m,c}$

27 $PDF_{n,c} = \frac{1}{\sigma_{n,c}\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu_{n,c}}{\sigma})^2}$

28 $PDF_{m,c} = \frac{1}{\sigma_{m,c}\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu_{m,c}}{\sigma})^2}$

29 $CDF_{n,c} = P(X \leq x),\ \boldsymbol{for}\ all\ x_{n,c}$

30 $CDF_{m,c} = P(X \leq x),\ \boldsymbol{for}\ all\ x_{m,c}$

31 $critical_value_{n,c} = inverse.CDF_{n,c}(1 - 0.025)$

32 $critical_value_{m,c} = inverse.CDF_{m,c}(1 - 0.025)$

33

34 ***for*** *all v ∈ V* ***do***

35 ***for*** *all n,m ∈ (N, M)* ***do***

36 ***for*** *all c ∈ C* ***do***

37 ***if*** $G_{v,n,c} \geq critical_value_{n,c,V}$

38 $G_{v,n,c} \leftarrow 1$

39 ***else:***

40 $G_{v,n,c} \leftarrow 0$

41 ***if*** $G_{v,m,c} \geq critical_value_{m,c,V}$

42 $G_{v,m,c} \leftarrow 1$

43 ***else:***

44 $G_{v,n,c} \leftarrow 0$

Block 3

For each of the groups formed by all the nodes of the network obtained in BLOCK 2, we established the threshold (line 20), for both cellular lines

For body and protrusion lines we obtained the average (lines 23-24), standard deviation (lines 25-26), normal distribution (lines 27-28), cumulative distribution function (lines 29-30), critical threshold value through the inverse of cumulative distribution function (lines 31-32).

Block 4

Each of the values detected (lines 34-36) from the difference of RNA-seq values (lines 16-17) is compared with the specific threshold of a given group (lines 37-41). If this value is greater than the threshold is assigned 1 (lines 38-42), otherwise it's assigned the Boolean value 0 (lines 40-44).

Fig. 2: Procedure used for binarization of RNA-seq values. BLOCK 1: normalization of RNA-seq values. BLOCK 2: subtraction of each normalized value of MCF10A RNA-seq from each value of MDA-MB-231data and log transformation application. BLOCK 3: determination of the threshold value for which to attribute a specific Boolean value. BLOCK 4: attribution of a Boolean value based on the critical value.

Using the BooleanNet library [22] applied to the binary values of each MDA-MB-231 RNA-seq normalized data paired with those of every MCF10A sample on the corresponding genes of our network (**Supplementary Table S2**), we looked for the attractors produced by the dynamic evolution of the network. The participation of the 27 apoptosis-related genes in the attractors of this initial setting served as a reference to assess the effect of subsequent network modifications (Fig. 2)

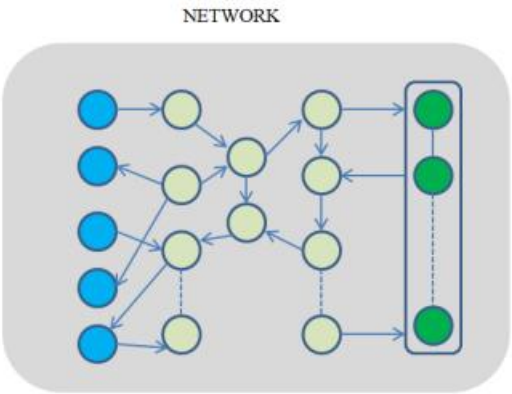

Fig. 3: Structure of the gene regulation network under study. The blue color indicates the five target genes, *CSNK2B*, *HSP90AB1*, *TK1*, *VIM*, *YWHAB*, which were inhibited in the *in vitro* experiment of Tilli et al. [11]. The dark green nodes represent the apoptosis-related genes. The nodes in light green exemplify the rest of the network genes. There is no vertex inhibition in this network representation.

Then, we simulated the conditions of the *in vitro experiment* performed by Tilli et al. [11], which induced the death of the MDA-MB-231 cancer cell line by transiently inhibiting *TK1*, *VIM*, *YWHAB*, *CSNK2B*, and *HSP90AB1* genes through siRNA interference. Below, for clarity, we call these five targets *bench targets*. To mimic this *in vitro* experiment, we constantly inhibited the vertices corresponding to these bench targets in the dynamic evolution of the network, as shown in Fig. 3.

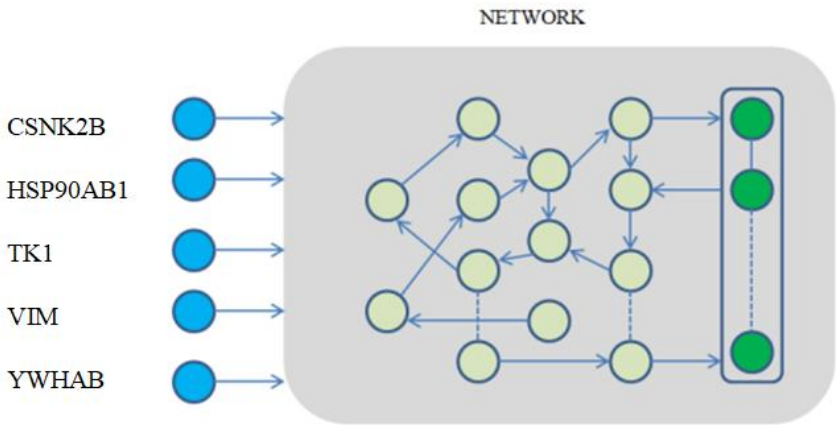

Fig. 4: The five bench targets (blue nodes) were set to zero (inhibition) for network dynamics simulation.

In Fig. 3, we compared the activation or inhibition of each of the 27 apoptotic genes and their participation in the attractors following the inhibition of the five bench targets of the original network of Fig. 1. This comparison enabled us to assess the functional correspondence of our model with the *in vitro* experiment of Tilli et al. [11]. In addition, during the network simulation, *TP53*

was constantly inhibited because this gene is not a tumor suppressor in MDA-MB-231 due to mutations [42].

4 OPTIMIZING THE NUMBER OF TARGETS

Given the death induction of MDA-MB-231 *in vitro* by *CSNK2B*, *HSP90AB1*, *TK1*, *VIM*, *YWHAB* silencing [11] and the knowledge of the primary 27 apoptosis-related genes, we propose a method to identify genes that can lead cancer cells to programmed cell death.

Since the goal was to maximize the number of the 27 genes involved in the apoptosis attractor in the sense that the genes that induce apoptosis are "turned on" and genes that inhibit it are "turned off", we looked for the best target genes involved in this process, given the network structure. Following this reasoning, we investigated the network modularity based on the connection level of their vertex members. We identified the modular structure using the Clauset-Newman-More greedy modularity maximization algorithm [23], considering the network as not directed. Modularity was computed as described in Equation 17

$$\boldsymbol{Q} = \sum_{\boldsymbol{c}\in\boldsymbol{C}}\left[\frac{\boldsymbol{L_c}}{|\boldsymbol{E}|} - \Upsilon\left(\frac{\boldsymbol{K_c}}{2|\boldsymbol{E}|}\right)^{\boldsymbol{2}}\right] \tag{17}$$

Where c corresponds to communities, L_c is the number of intra-community links, Υ is the parameter of the resolution, and k_c is the sum of degrees within the community.

In its initial stage, this algorithm creates a partition where each node is its cluster. Then, it iteratively joins the pairs of clusters to increase modularity. A negative value is assigned to the initial modularity matching a singleton, which becomes larger and larger until it reaches a positive peak in correspondence with the best solution found by the algorithm. Eventually, the value reverts to zero when all nodes are in the same community. Finally, the algorithm finds the partition corresponding to the peak value in a backward process. The algorithm was implemented through the Python library NetworkX [24] . Having identified the communities in the network, we looked if the 27 apoptosis-related genes were clustered or scattered among these detected modules, possibly forming a single community in the network according to the criteria imposed by the algorithm. This approach has already been adopted in [25], where drug targets for cancer were mapped on a modularized network to detect the presence of modules on which the therapeutic action was concentrated.

Because of the structural clusterization shown by the apoptosis-related genes, we looked for nodes that would bridge these apoptosis-related clusters with the five bench targets. Finding these bridge nodes might minimize the path of target genes capable of reproducing the results of the bench experiment. For this purpose, we applied the Dijkstra algorithm through Python's NetworkX [24] to look for the shortest path between the five bench targets (*CSNK2B*, *HSP90AB1*, *TK1*, *VIM*, *YWHAB*) and each of the apoptosis-related genes.

A similar approach was adopted by George et al. [26], in which intermediate genes along the shortest path were identified as therapeutic targets and ranked according to the number of shortest paths in which they participated.

Since we know from the *in vitro* experiment that inhibiting the five bench targets led MDA-MB-231 to cell death, we used this knowledge to find alternative vertices to these five nodes whose

inhibition would activate the greatest number of genes of the apoptosis group. The simplest way to perform this task was to look for the smallest set of vertices with the common property of participating in at least one shortest path where each of the five genes, *CSNK2B, HSP90AB1, TK1, VIM, YWHAB* were considered as starting nodes. Each apoptosis-related gene was considered the final node (Figure 4).

Input: *List of in vitro experiment target genes G; list of apoptosis gene group A; list of network nodes V; adjacent list of the network T.*

Output: *New_target_list.*

```
1  for all g ∈ G do
2      for all a ∈ A do
3  C_ga = T.shortest_path_list_from_g_to_a
4          for all v ∈ V do
5              if v = g or v = a
6                  C_ga ← C_ga {v}
7  for all v ∈ V do
8      for all g ∈ G do
9          for all a ∈ A do
10             if v ∈ C_ga
11                 {...}L_g ← v
12 for all v ∈ V do
13     for all g ∈ G do
14             if v ∈ L_g
15                 x_v ← v + 1
16 New_target_list ← v ∈ max{...} x_v
```

Block 1

All the shortest paths from the in vitro experiment target gene $\boldsymbol{g}$ to each gene of the list of apoptosis group $\boldsymbol{a}$ are detected, obtaining a number of shortest paths equal to $\boldsymbol{g} \times \boldsymbol{a}$ (lines 1-3). In each of them the starting gene and the target gene are eliminated (lines 4-6).

Block 2

For each node of the network $\boldsymbol{v}$, its presence is verified in the shortest paths detected in block 1 (lines 7-9). If the node is present in at least one of the paths existing between a gene $\boldsymbol{g}$ and all the genes of group $\boldsymbol{a}$ (lines 10), it is inserted in the specific list relating to the specific target gene $\boldsymbol{g}$ (lines 11).

Block 3

For each node of the network $\boldsymbol{v}$ (line 12), its presence is verified in each of the detected lists (lines 13-14) previously established (***Block 2***, line 11). For each list in which the $\boldsymbol{v}$ gene is present, the value of $\boldsymbol{v}$ is incremented by one (line 15). The nodes of network $\boldsymbol{v}$ with the highest $\boldsymbol{x_v}$ values are assigned to the final list "***New_target_list***" (line 16).

Fig.5: Pseudocode of target vertex determination in shortest paths. Lines 1-6 (block 1): Detection of the shortest paths via Python networkX library between the five bench targets and the apoptosis-related genes. Lines 7-11 (block 2): Creation of a list for each bench target gene containing the nodes detected by the algorithm on every shortest path. Line 12-17 (block 3): Insertion in the list of new targets of the genes present in the largest number of lists of the previous step. The asymptotic complexity of the algorithm is $O(\max\{C_{1,2}(G * A * V), C_3(V * G)\})$.

The procedure indicated in Figure 4 allowed us to find new target genes capable of acting on the configuration of the apoptosis-related group. By "configuration", we mean the number of activated

apoptosis-related genes that may induce apoptosis and the consistency of this number across the RNA-seq comparisons of this study.

To assess the effect of inactivating the new target vertices identified through the shortest path strategy on the configuration of the apoptosis-related group, we looked for the activation or inhibition state of the apoptosis-related genes for the twenty comparisons analyzed (ten for *body*, and ten for *protrusion*, corresponding to the two fractions of MDA-MB-231). This procedure allowed us to compare the effect of inhibiting the new targets within the shortest paths to that obtained by inhibiting the original five bench targets.

RESULTS

1 GENE REGULATORY NETWORK

The breast cancer regulatory network in this study comprises the genes used in a previous report for calculating Boolean attractors [10] into which 25 new genes were integrated (Fig. 5). These 25 genes, in addition to 2 genes already present in the previous network, play an essential role in the process of cellular apoptosis.

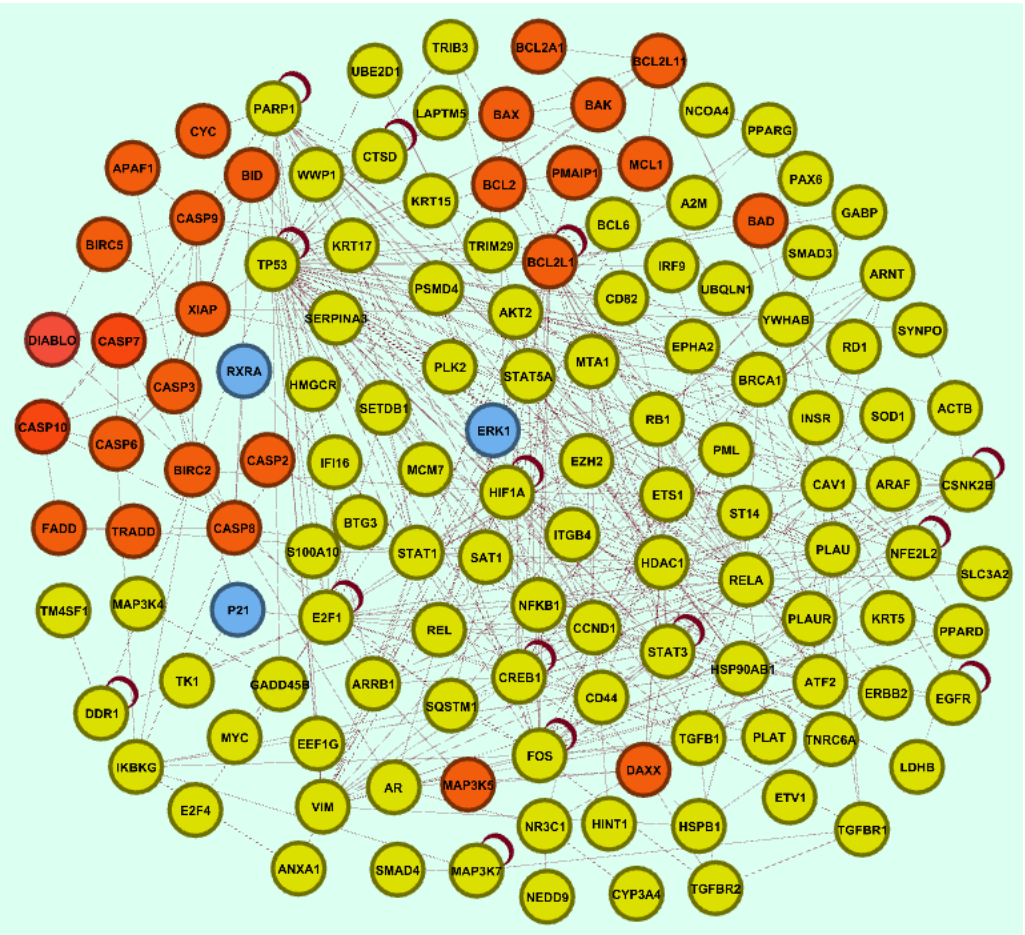

Fig 6: Breast cancer gene regulatory network used in this report. Twenty-five red nodes and the three blue nodes, not part of the apoptosis group, are the new vertices added to the network used in the previous report (yellow nodes). In the group of twenty-seven red genes related to apoptosis, two (*BAX* and *BCL2L1*) were already present in our earlier work network.

The network shown in Figure 5 is composed of **131** nodes and **494** edges. Boolean transfer functions determine the system's dynamics, in which the different nodes can play an activating or inhibiting role on the other nodes of which they are inputs (**Supplementary Table S3**). We deduced the nature of these interactions through the recourse to appropriate software [12], through which it was also possible to integrate the genes of the apoptosis-related group with the pre-existing network.

2 STRUCTURAL ANALYSIS OF THE NETWORK

We identified the structural characteristics of the network through comparison with the canonical network types. The comparison, based on degree distribution, was performed with three types of networks well defined in the literature, the Erdos-Renvy [27], the Watts-Strogatz [28], and the Barabasi-Albert networks [19].

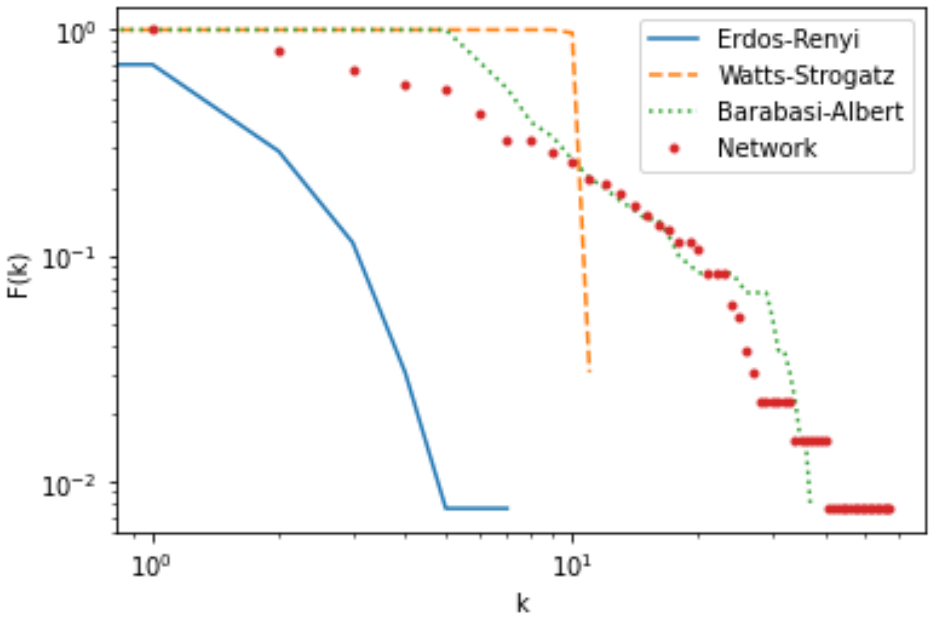

Fig. 7: Convergence of our model with three canonical networks. Log-log plot of Erdos and Renvi (blue line), Watts and Strogatz (orange line), and scale-free networks (green line) networks compared to the experimental network of this study (red).

From the plot of Fig. 6, representing a complementary cumulative distribution function, we verified that the network used in this report was very similar to the network model representing a power-law degree distribution. As Albert et al. [19] indicated, we can find this type of distribution in many real networks, including those describing complex biological systems such as the network used for our analysis.

3 ATTRACTOR ANALYSIS

When analyzing the gene regulation network as illustrated in Fig. 2, i.e., without any vertex inhibition, we observed that the percentage of genes that satisfies the suitable conditions for the transition of the malignant state to that of apoptosis was 29.6% in *body* and 14.8% in *protrusion* samples, respectively. The pro-apoptosis role of just two genes of the Bcl-2 family suggests a configuration unsuitable for the cell apoptosis process, which was supported by the fact that *CASP3, 6, 7* (fundamental for the intrinsic apoptosis pathway) were off. In addition, *XIAP* and *DIABLO*, which are inhibitors of *CASP9* and the IAP family (endogenous proteins that block caspases), respectively, did not meet the necessary conditions for activating an apoptosis process (Fig. 7).

BODY		*PROTRUSION*	
	1-10		1-10
APAF1	F	APAF1	F
BAD	T	BAD	T
BAK	F	BAK	F
BAX	F	BAX	F
BCL2	T	BCL2	T
BCL2A1	T	BCL2A1	T
BCL2L1	F	BCL2L1	F
BCL2L11	F	BCL2L11	F
BID	F	BID	F
BIRC2	F	BIRC2	X
BIRC5	F	BIRC5	F
CASP10	T	CASP10	T
CASP2	T	CASP2	X
CASP3	F	CASP3	F
CASP6	F	CASP6	F
CASP7	F	CASP7	F
CASP8	T	CASP8	X
CASP9	F	CASP9	F
CYC	F	CYC	F
DIABLO	F	DIABLO	F
FADD	T	FADD	X
MP3K5	F	MP3K5	F
MCL1	T	MCL1	T
PMAIP1	F	PMAIP1	F
TRADD	F	TRADD	F
XIAP	T	XIAP	T
DAXX	F	DAXX	F

Fig. 8: Configuration of the apoptosis-related genes in the attractors of the twenty comparisons of two MDA-MB-231 sample types, *body*, and *protrusion*. The seven genes indicated in red favor a pro-apoptotic mechanism if the gene is inhibited. On the other hand, the 20 genes displayed in black favor apoptosis if genes are activated (**Supplementary Table 1**). The *DAXX* gene is an exception, for which we did not find any characterization of its pro-apoptosis state in the literature. T (for True or activation), F (for False or inhibition), and X (for a continuous alternation between T and F) represent the boolean value of the 27 apoptosis-related genes within the detected attractors.

T, F, and X background colors indicate if there are matches between the detected state and the ideal one of the corresponding genes for apoptosis induction. The green background means correspondence, while the orange one indicates a divergence.

To find attractors that mimic the results described in Tilli et al. [11], we inhibited the action of the five bench targets, which led the cell line MDA-MB-231 to death in vitro. The percentages of genes in the apoptosis attractors that assume the state of inhibition or activation supporting cell death were 74% and 55% in the comparisons 1 to 6 and 7 to 10 , respectively, for *body* samples. For *protrusion* samples, the genes in the apoptosis phenotype state were found in 41% (comparisons 1 to 6) and 74% (7 to 10) (Fig. 8). There is a quantitative and qualitative difference when comparing these results with those shown in Fig. 7 (obtained without any inactivation of the network genes). Quantitatively, the percentages of genes in which the attractors coincided with the apoptosis state are significantly higher, indicating that the model was somewhat representative of the in vitro experiment. Qualitatively, the apoptotic state of Bcl-2 and Caspase families, as well as *DIABLO* and *XIAP*, corroborates the expected bench experiment in the *body* 1-6 and *protrusion* 7-10 comparisons and partially observed in the *body* 7-10 and *protrusion* 1-6 ones because of a difference in the RNA-seq profiles of the MCF10A cell lines used in the comparisons.

BODY			***PROTRUSION***		
	1-6	**7-10**		**1-6**	**7-10**
APAF1	**F**	**F**	**APAF1**	**F**	**F**
BAD	**F**	**T**	**BAD**	**F**	**F**
BAK	**T**	**F**	**BAK**	**T**	**T**
BAX	**T**	**F**	**BAX**	**T**	**T**
BCL2	**F**	**T**	**BCL2**	**F**	**F**
BCL2A1	**F**	**T**	**BCL2A1**	**F**	**F**
BCL2L1	**F**	**F**	**BCL2L1**	**F**	**F**
BCL2L11	**T**	**F**	**BCL2L11**	**T**	**T**
BID	**F**	**F**	**BID**	**F**	**F**
BIRC2	**F**	**F**	**BIRC2**	**X**	**F**
BIRC5	**F**	**F**	**BIRC5**	**X**	**F**
CASP10	**T**	**T**	**CASP10**	**T**	**T**
CASP2	**T**	**T**	**CASP2**	**X**	**T**
CASP3	**T**	**T**	**CASP3**	**X**	**T**
CASP6	**T**	**T**	**CASP6**	**T**	**T**
CASP7	**T**	**T**	**CASP7**	**X**	**T**
CASP8	**T**	**T**	**CASP8**	**X**	**T**
CASP9	**T**	**T**	**CASP9**	**F**	**T**
CYC	**F**	**F**	**CYC**	**F**	**F**
DIABLO	**T**	**T**	**DIABLO**	**X**	**T**
FADD	**T**	**T**	**FADD**	**X**	**T**
MP3K5	**F**	**F**	**MP3K5**	**F**	**F**
MCL1	**F**	**T**	**MCL1**	**F**	**F**
PMAIP1	**T**	**T**	**PMAIP1**	**T**	**T**
TRADD	**F**	**F**	**TRADD**	**F**	**F**
XIAP	**F**	**F**	**XIAP**	**F**	**F**
DAXX	**F**	**F**	**DAXX**	**F**	**F**

Fig. 9: Attractors obtained by inhibiting the five bench targets: *CSNK2B*, *HSP90AB1*, *TK1*, *VIM*, and *YWHAB*. Columns *BODY* 1-6 and 7-10 refer to the ten *body* samples, while columns *PROTRUSION* 1-6 and 7-10 refer to the ten *protrusion* ones.

4 NETWORK MODULARITY ANALYSIS

The investigation of the network's modularity highlighted that the apoptosis-related genes tended to cluster in two groups with well-defined functional characteristics, as illustrated in Fig. 9.

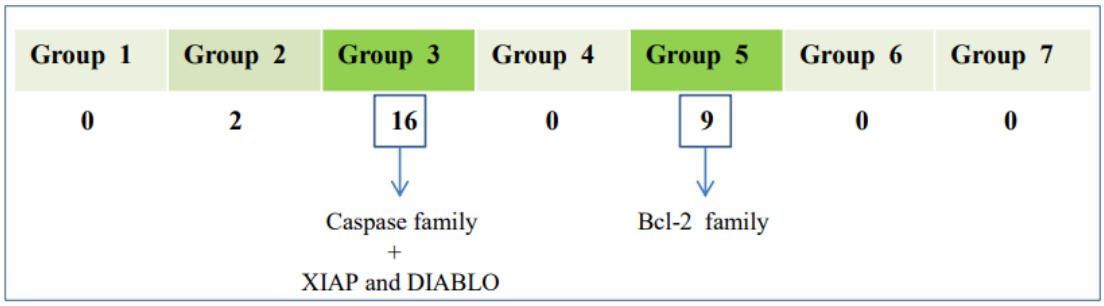

Fig. 10: Network modularization process through the Clauset-Newman-More algorithm with the number of components of the apoptosis group in each of the seven modules. Groups 3 and 5, highlighted in green, indicate the groups belonging to the same modules relevant to the cell apoptosis process (mainly Caspases and Bcl-2 families)[29].

Groups 3 and 5 contain 92.5% of the apoptosis-related genes. It is worth noting that the entire Caspases plus *XIAP* and *DIABLO* group is clustered in group 3, while the whole Bcl-2 family is clustered in group 5. This modular distribution (**Supplementary Table S4**) showed that the apoptosis-related genes cluster together. This characteristic is essential in the methodological setting adopted in this report because it intuitively allows using a relatively limited number of target vertices to activate them.

5 SHORTEST PATH EVALUATION

Searching for the shortest path connecting the five bench targets with the apoptosis-related genes allowed us to identify three genes, *HIF1A*, *XIAP*, and *BCL2*, present in at least one of these shortest paths. Given that *XIAP* and *BCL2* belong to the apoptosis group that we used as an indicator of the state of the network and that they are also the final nodes of the shortest paths, we did not assess the effects of their inhibition on the apoptosis attractor. However, since only *HIF1A* did not belong to the apoptosis group, *XIAP*, and *BCL2* were replaced by their input nodes that were *STAT5A* and *BRCA1*, respectively (Fig. 10).

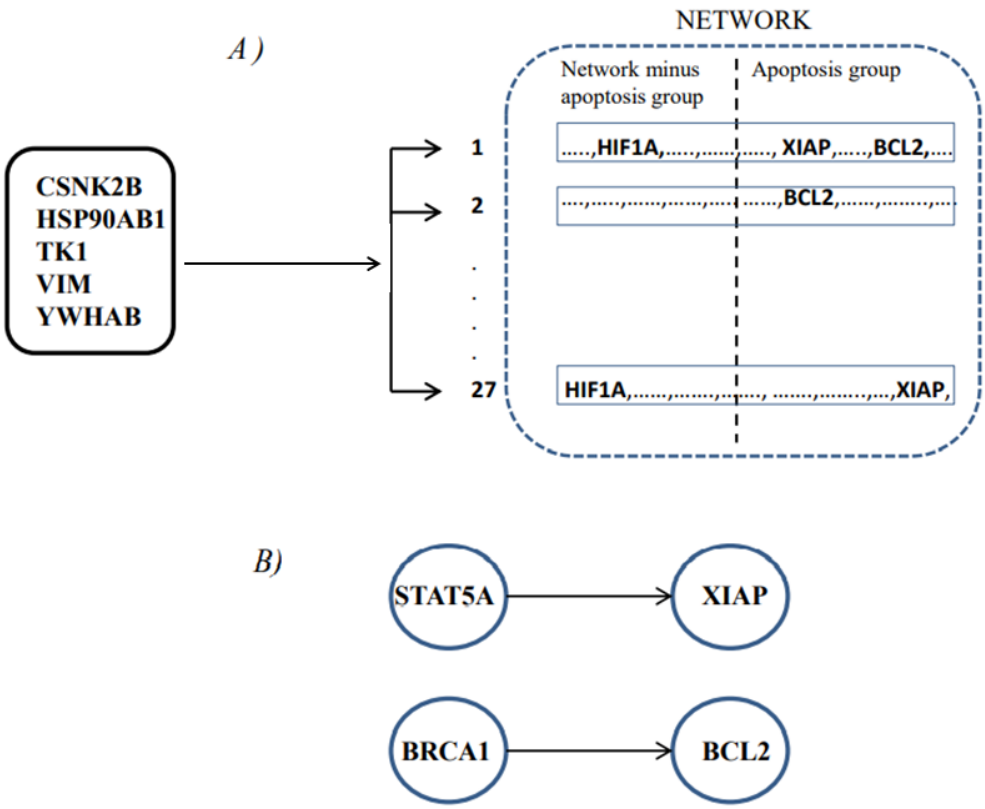

Fig. 11: Process of new target identification by the shortest search. **Panel A:** *HIF1A*, *XIAP*, and *BCL2* are present in at least one of the 27 shortest paths. From these three genes, only *HIF1A* did not belong to the apoptosis-related group**. Panel B:** *STA5A* and *BRCA1* represent the only input genes of *XIAP* and *BCL2*. Thus, *HIF1A*, *STAT5A*, and *BRCA1* were the optimized target nodes detected within the shortest paths between the five bench targets and the 27 apoptosis-related genes. These three vertices were excellent candidates to complete the new set of optimized target genes that trigger the network within the apoptosis state.

6 OPTIMIZING THE NUMBER OF TARGETS

The vertices with the greatest centrality between the five bench targets and the 27 apoptosis-related genes provided the opportunity to test the effect of inhibiting the new targets on the apoptosis attractor and to propose a new approach to choosing therapeutic targets. *STAT5A*, *BRCA1*, and *HIF1A* were vertices with high centrality detected through the application of the algorithm described in Fig. 4. Therefore, we used these three genes as targets for inhibition instead of *CSNK2B*, *HSP90AB1*, *TK1*, *VIM*, and *YWHAB*. The three-gene combination activated the attractor of apoptosis in all sample comparisons. The inhibition of *HIF1A*, *STAT5A*, and *BRCA1* in place of the five bench targets produced the results illustrated in Fig. 11.

BODY		*PROTRUSION*	
	1-10		1-10
APAF1	F	APAF1	F
BAD	F	BAD	F
BAK	T	BAK	T
BAX	T	BAX	T
BCL2	F	BCL2	F
BCL2A1	F	BCL2A1	F
BCL2L1	F	BCL2L1	F
BCL2L11	T	BCL2L11	T
BID	T	BID	T
BIRC2	F	BIRC2	F
BIRC5	F	BIRC5	F
CASP10	T	CASP10	T
CASP2	T	CASP2	T
CASP3	T	CASP3	T
CASP6	T	CASP6	T
CASP7	T	CASP7	T
CASP8	T	CASP8	T
CASP9	T	CASP9	T
CYC	F	CYC	F
DIABLO	T	DIABLO	T
FADD	T	FADD	T
MP3K5	F	MP3K5	F
MCL1	F	MCL1	F
PMAIP1	T	PMAIP1	T
TRADD	F	TRADD	F
XIAP	F	XIAP	F
DAXX	F	DAXX	F

Fig. 12: Activation or inhibition status detected in the 27 genes constituting the apoptosis-related group in the attractors of 10 *body* (1-10) and ten *protrusion* (1-10) sample comparisons by inhibiting *HIF1A*, *STAT5A*, and *BRCA1* instead of *CSNK2B*, *HSP90AB1*, *TK1*, *VIM* and *YWHAB*.

We showed in Fig.11 the simultaneous induction to cell apoptosis for *Bcl-2* and *Caspase* gene families in every combination of RNA-seq from *body* and *protrusion* types. It is also important to note that *BID* is activated in this simulation. The activation of *BID*, a pro-apoptosis member of the *Bcl-2* family, is responsible for activating downstream *Caspases* through the direct activation of *BAX* and *BAK*. This activation is absent in the results obtained by simulating the apoptosis attractor through the five bench genes (Fig. 8). The same consideration must be made for *DIABLO*, which was not constantly expressed in the six RNA-seq protrusion combinations from 1 to 6 (Fig. 8)
Therefore, we analyzed the induction of the network into the apoptosis attractor through different strategies of gene inhibition. By inhibiting the five bench targets, we obtained a configuration suitable for apoptosis in a considerable portion of the genes belonging to this group (green color in the graph) (Fig. 12). The proportion of genes activated in the apoptosis-related group was significantly greater (74.1%) when compared to no-bench target inhibition (first column). When *HIF1A*, *STAT5A*, and *BRCA1* were chosen as targets for inhibition, the proportion of genes activated in the apoptosis group was even larger (77.8%) and independent of the fraction of MDA-MB-231 or from the MCF10A RNA-seq used to find the up-regulated genes in MDA-MB-231 (Fig. 12, right column).

The results above suggest the transition from a malignant basin of attraction specific to *body* and *protrusion* samples to a single basin of attraction representing a generalized state of cellular apoptosis due to *HIF1A*, *STAT5A*, and *BRCA1* inhibition (Fig. 13).

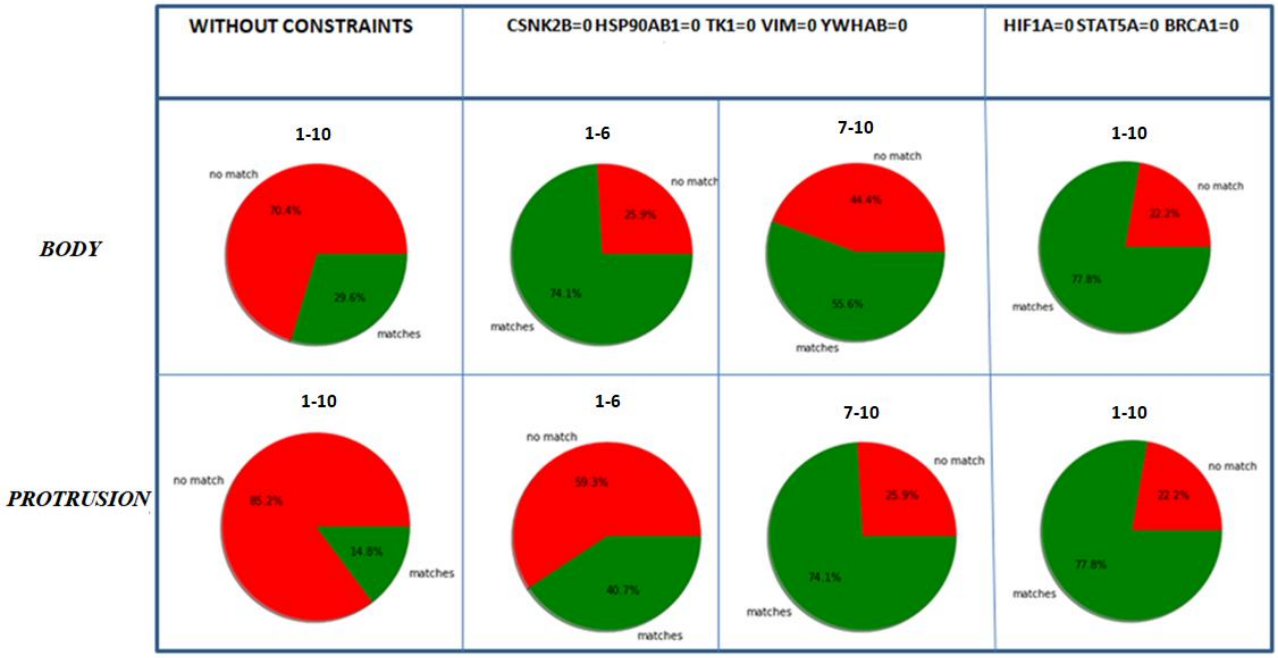

Fig. 13: The results detected on the *body* and *protrusion* fractions of MDA-MB-231 used in this report, respectively. The columns identify which network genes were kept silenced in the dynamic simulation of the model. The green color, as opposed to the red one, indicates the percentage of the genes of the apoptosis group presenting a stable configuration of activation or inhibition associated with the attractor of the apoptosis phenotype.

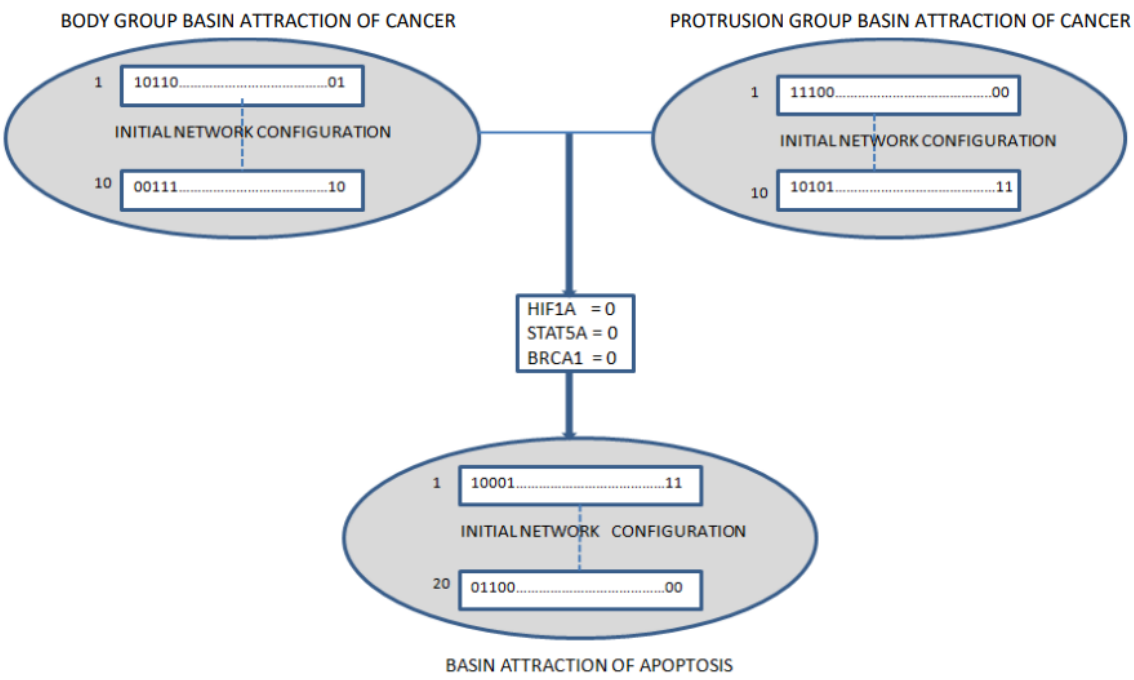

Fig. 14: Boolean description of the transition from two basins of attraction representing the malignant cellular state of *body* and *protrusion* to a single basin of attraction of the cellular apoptosis state, obtained by inhibiting *HIF1A*, *STAT5A*, and *BRCA1*.

DISCUSSION

Cancer is a complex process influenced by several factors whose molecular approach is no longer sufficient to understand the biological mechanisms that underlie its evolution. For this reason, the trend is to increasingly integrate molecular data with dynamic properties of the biological networks simulated through computing and mathematical modeling [30].

The use of quantitative or qualitative modeling depends on the type of data available. The need for kinetic parameters makes modeling through ordinary differential equations complex and only applicable to gene regulatory networks of limited size [31]. Qualitative modeling using Boolean networks represents a valid alternative, enabling a relatively simple dynamic simulation of complex biological systems [32] and helpful in searching regulatory interactions in protein expression [33] and therapeutic intervention strategies [34].

In addition to these documented characteristics, we provided functional compatibility checks for the Boolean model implemented in this report with the results obtained in an *in vitro* experiment [11] whereby silencing five genes by siRNA. As a result, the malignant MBA-MB-231 cell line was forced into apoptosis. The data we produced clearly show how our system can generate functionally compatible results with those obtained from the *in vitro* test by acting through inhibition on the same genes. Therefore we reproduced the evolution of an actual biological system into the Boolean dynamics of the gene regulatory network implemented for our research.

In this report, using a Boolean network representing a set of up-regulated genes in breast cancer has allowed us to detect attractors corresponding to specific cellular phenotypes and verify the Boolean network's compatibility with an existing biological system through the comparison with an *in vitro* experiment [11]. The assignment of the Boolean values to the network nodes was carried out following the algorithm illustrated in Fig. 2. Through this procedure, Booleanization of RNA-seq values was performed based on the gene expression differences between malignant and non-malignant cell lines. Considering the difference in gene expression between different cell lines, it also allowed us to obtain an interesting number of comparisons on which to perform the dynamic simulation of our model. In fact, instead of considering only one control for the four malignant samples (two body and two protrusion) on which to Booleanize the RNA-seq values and perform the search for attractors, we used with five samples of the non-malignant cell line MCF10A in order to derive the gene expression differences. This strategy produced at the combinatorial level a number of comparisons equal to twenty, allowing the possibility of applying the procedure used in this report on numerically more significant network configurations.

Considering cancer phenotype as basins of attraction in the epigenetic landscape [35], this report aimed to cause the transition from a basin of attraction of malignant type to that of apoptosis [36] through the perturbation of a subset of genes belonging to the network. For this purpose, we developed an algorithm (Fig. 5) that optimizes the choice of the network elements to produce a transition from one specific phenotype to another.

The approach of exploring the relationship existing between the results of an *in vitro* experiment, the insertion of a specific group of genes for apoptosis into the system, and the investigation of the

network structure through the analysis of shortest paths between the five bench targets and the apoptosis-related group, represents an innovative strategy for clinical applications to increase patient benefit in personalized approaches of cancer therapies.

The group of apoptosis-related genes inserted into the network was a reference to assess the induction of cell death attractors by vertex inhibition. Fig. 13 schematically shows the results obtained in this report by adopting this methodology. By inhibiting the three genes *HIF1A*, *STAT5A*, and *BRCA1*, we observed the transition of the system dynamics from the malignant basins of attraction to that of cell apoptosis in all samples analyzed . The comparison of this result (Fig. 12), with that obtained by reproducing the in vitro experiment (Fig. 9), shows the robustness of the data obtained by applying the procedure of Fig. 5. The inhibition of the five genes (*CSNK2B, HSP90AB1, TK1, VIM, YWHAB*) described in Tilli et al. [11] promotes a configuration of the 23 apoptosis-related genes favorable to apoptosis. However, this configuration is not quantitatively uniform in all comparisons. For example, lines from 1 to 6 were more favorable to apoptosis in the body sample while more favorable from 7 to 10 in the protrusion sample (Fig. 9). Instead, the inhibition of the three genes (*HIF1A, STAT5A, BRCA1*) detected through the procedure described in Fig. 5 resulted in a profile of activated apoptosis-related genes more reproducible over all the comparisons (Fig. 12). Despite the difference in attractors between body and protrusion samples, the induction of apoptosis by inhibiting HIF1A, STAT5A and BRCA1 makes the method robust.

As shown in Fig. 11, the target genes detected by the procedure described in Fig. 5 should be *HIF1A*, *XIAP*, and *BCL2* considering that our challenge was to identify target genes able to activate or inhibit the 27 apoptosis-related genes, whatever their configuration. Thus we decided to avoid indicating *XIAP* and *BCL2* as targets. One option was to replace them with their input nodes (those not belonging to the apoptosis group), *STAT5A*, and *BRCA1*. Thus, it is not excluded that the combined inhibition action on the *HIF1A*, *XIAP*, and *BCL2* genes may have a relevant function in the induction of cancer cell apoptosis.

For each of the *HIF1A*, *STAT5A*, and *BRCA1* genes, the relationship between their inhibition and apoptosis induction in the MDA-MB-231 cell line has already been identified in several reports. Inhibition of *HIF1A* expression by siRNA suppresses cell growth of MDA-MB-231 and increases apoptosis [37]. There is a correlation between the inhibition of *STAT5A* and a decrease in metastasis and growth of breast cancer tumor cells [38,39]. In addition, the knockdown of *STAT5A* restores cellular sensitivity to TRAIL-induced apoptosis [40]. Finally, the RNAi-mediated silencing of *BRCA1* increased the percentage of apoptotic cells in synergy with the transfection of miR-342 [41]. Moreover, *BRCA1*-depleted MDA-MD-231 cells showed increased susceptibility to proteasome inhibitors [43].

The results obtained in this report depended on tuning the transfer functions (eq. 6-10) so that the model could match the *in vitro* experiment [11]. Nevertheless, the correspondence between the result obtained *in vitro* and by computational simulation suggested sufficient model representativity for further optimization and applications in therapeutic contexts. Therefore, it is possible to integrate specific experimental results with computational hypotheses formulated to address cancer-related therapeutic problems.

The results obtained in this report come from the processing of data extracted from specific biological samples. The ever-increasing availability of this type of information related to particular pathological situations of cancer shows the adaptability of our model to different configurations of the same pathology to be addressed, placing the method elaborated in this report in the context of personalized medicine.

CONCLUSION

By tackling the study of the complex dynamics of a gene regulatory network related to breast cancer through the detected attractors, we have computationally demonstrated a method capable of directing cell fate by acting directly on some system elements.
By combining the knowledge derived from direct laboratory experience, comprehensive bibliographic research related to each system element, and applying structural analyses to the network, a modeling strategy was developed that may provide potentially helpful therapeutic indications.
As further developments of the proposed methodology, it is possible to imagine therapeutic contexts in which clinical and laboratory information can be integrated with heterogeneous biological networks to tackle a specific pathological situation. In this perspective, we can place the methodology developed in this report in the context of a personalized therapeutic approach.

DATA AVAILABILITY STATEMENT

Source code and supplementary data used in this study are available at:
https://github.com/Domenico321/therapeutic-optimization

REFERENCE

1. Hanahan D. Hallmarks of Cancer: New Dimensions. Cancer Discov. 2022 Jan;12(1):31-46, https://doi.org/10.1158/2159-8290.CD-21-1059.
2. Sung, Hyuna, et al. "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries." *CA: a cancer journal for clinicians* 71.3 (2021): 209-249.
3. Chan CWH, Law BMH, So WKW, Chow KM, Waye MMY. Novel Strategies on Personalized Medicine for Breast Cancer Treatment: An Update. Int J Mol Sci. 2017 Nov 15;18(11):2423, https:// doi.org/10.3390/ijms18112423.
4. Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. Nat Med. 2004 Aug;10(8):789-99, https://doi.org/10.1038/nm1087.
5. Yuan R, Zhu X, Wang G, Li S, Ao P. Cancer as robust intrinsic state shaped by evolution: a key issues review. Rep Prog Phys. 2017 Apr;80(4):042701,https://doi.org/10.1088/1361-6633/aa538e.
6. Barillot, Emmanuel, et al. *Computational systems biology of cancer*. Boca Raton, FL: CRC Press, 2013. pp. 6-7
7. Feng Y, Spezia M, Huang S, Yuan C, Zeng Z, Zhang L, Ji X, Liu W, Huang B, Luo W, Liu B, Lei Y, Du S, Vuppalapati A, Luu HH, Haydon RC, He TC, Ren G. Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis. Genes Dis. 2018 May 12;5(2):77-106. https://doi.org/10.1016/j.gendis.2018.05.001.

8. Waddington, C.(1957). The Strategy of The Genes: A discussion of some aspects of theoretical biology. London: George Allen and Unwin, 262
9. Huang S, Ernberg I, Kauffman S. Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. Semin Cell Dev Biol. 2009 Sep;20(7):869-76,https://doi.org/10.1016/j.semcdb.2009.07.003.
10. Sgariglia D, Conforte AJ, Pedreira CE, Vidal de Carvalho LA, Carneiro FRG, Carels N, Silva FABD. Data-Driven Modeling of Breast Cancer Tumors Using Boolean Networks. Front Big Data. 2021 Oct 20;4:656395. https://doi/10.3389/fdata.2021.656395.
11. Tilli TM, Carels N, Tuszynski JA, Pasdar M. Validation of a network-based strategy for the optimization of combinatorial target selection in breast cancer therapy: siRNA knockdown of network targets in MDA-MB-231 cells as an in vitro model for inhibition of tumor development. Oncotarget. 2016 Sep 27;7(39):63189 63203. https://doi.org/10.18632/oncotarget.11055.
12. Ekins S, Nikolsky Y, Bugrim A, Kirillov E, Nikolskaya T. Pathway mapping tools for analysis of high content data. Methods Mol Biol. 2007;356:319-50. https://doi.org/10.1385/1-59745-217-3:319.
13. Nikolajewa S, Friedel M, Wilhelm T. Boolean networks with biologically relevant rules show ordered behavior. Biosystems. 2007 Jul-Aug;90(1):40-7. https://doi.org/ 10.1016/j.biosystems.2006.06.007.
14. Kauffman, Stuart A. *The origins of order: Self-organization and selection in evolution*. Oxford University Press, USA, 1993.pp. 447-449
15. Shmulevich I, Gluhovsky I, Hashimoto RF, Dougherty ER, Zhang W. Steady-state analysis of genetic regulatory networks modelled by probabilistic boolean networks. Comp Funct Genomics. 2003;4(6):601-8. https://doi.org/10.1002/cfg.342.
16. Kauffman S, Peterson C, Samuelsson B, Troein C. Genetic networks with canalyzing Boolean rules are always stable. Proc Natl Acad Sci U S A. 2004 December 7;101(49):17102-7. https://doi.org/10.1073/pnas.0407783101.
17. Szallasi Z, Liang S. Modeling the normal and neoplastic cell cycle with "realistic Boolean genetic networks": their application for understanding carcinogenesis and assessing therapeutic strategies. Pac Symp Biocomput. 1998:66-76.
18. Harris, Stephen E., et al. "A model of transcriptional regulatory networks based on biases in the observed regulation rules." *Complexity* 7.4 (2002): 23-40. https://doi.org/10.1002/cplx.10022.
19. Albert R. Scale-free networks in cell biology. J Cell Sci. 2005 November 1;118(Pt 21):4947-4957. https://doi.org/10.1242/jcs.02714.
20. Pires JG, da Silva GF, Weyssow T, Conforte AJ, Pagnoncelli D, da Silva FAB, Carels N. Galaxy and MEAN Stack to Create a User-Friendly Workflow for the Rational Optimization of Cancer Chemotherapy. Front Genet. 2021 February 18;12:624259. https://doi.org/ 10.3389/fgene.2021.624259.
21. Kauffman S, Peterson C, Samuelsson B, Troein C. Random Boolean network models and the yeast transcriptional network. Proc Natl Acad Sci U S A. 2003 December 9;100(25):14796-9. https://doi.org/10.1073/pnas.2036429100.
22. Albert I, Thakar J, Li S, Zhang R, Albert R. Boolean network simulations for life scientists. Source Code Biol Med. 2008 November 14;3:16. https://doi.org/10.1186/1751-0473-3-16.

23. Clauset A, Newman ME, Moore C. Finding community structure in very large networks. Phys Rev E Stat Nonlin Soft Matter Phys. 2004 Dec;70(6 Pt 2):066111. https://doi.org/10.1103/PhysRevE.70.066111.
24. Hagberg, Aric, Pieter Swart, and Daniel S Chult. *Exploring network structure, dynamics, and function using NetworkX*. No. LA-UR-08-05495; LA-UR-08-5495. Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008.
25. Takeshi Hase, Kaito Kikuchi, Samik Ghosh, Hiroaki Kitano. "Identification of drug-target modules in the human protein–protein interaction network." *Artificial Life and Robotics* 19.4 (2014): 406-413.https:doi.org/10.1007/s10015-014-0178-5.
26. George RA, Liu JY, Feng LL, Bryson-Richardson RJ, Fatkin D, Wouters MA. Analysis of protein sequence and interaction data for candidate disease gene prediction. Nucleic Acids Res. 2006;34(19):e130. https://doi.org/10.1093/nar/gkl707.
27. P. Erdós and A. Rényi, "On the evolution of random graph I", *Publicationes Mthematicae Debrecen* 6:290(1959).
28. Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. Nature. 1998 Jun 4;393(6684):440-2. https://doi.org/10.1038/30918.
29. Pecorino, Lauren. *Molecular biology of cancer: mechanisms, targets, and therapeutics*. Oxford university press, 2012. pp. 149-160
30. Kitano H. Systems biology: a brief overview. Science. 2002 Mar 1;295(5560):1662-4. https://doi.org/10.1126/science.1069492.
31. Karlebach G, Shamir R. Modelling and analysis of gene regulatory networks. Nat Rev Mol Cell Biol. 2008 Oct;9(10):770-80. https://doi.org/10.1038/nrm2503.
32. Schwab JD, Kühlwein SD, Ikonomi N, Kühl M, Kestler HA. Concepts in Boolean network modeling: What do they all mean? Comput Struct Biotechnol J. 2020 March 10;18:571-582. https://doi.org/10.1016/j.csbj.2020.03.001.
33. Dahlhaus M, Burkovski A, Hertwig F, Mussel C, Volland R, Fischer M, Debatin KM, Kestler HA, Beltinger C. Boolean modeling identifies Greatwall/MASTL as an important regulator in the AURKA network of neuroblastoma. Cancer Lett. 2016 Feb 1;371(1):79-89. https://doi.org/10.1016/j.canlet.2015.11.025.
34. Bloomingdale P, Nguyen VA, Niu J, Mager DE. Boolean network modeling in systems pharmacology. J Pharmacokinet Pharmacodyn. 2018 Feb;45(1):159-180. https://doi.org/10.1007/s10928-017-9567-4.
35. Crespo I, Perumal TM, Jurkowski W, del Sol A. Detecting cellular reprogramming determinants by differential stability analysis of gene regulatory networks. BMC Syst Biol. 2013 December 19;7:140. https://doi.org/10.1186/1752-0509-7-140.
36. Shi Y, Chang M, Wang F, Ouyang X, Jia Y, DU H. Role and mechanism of hypoxia-inducible factor-1 in cell growth and apoptosis of breast cancer cell line MDA-MB-231. Oncol Lett. 2010 Jul;1(4):657-662. https://doi.org/10.3892/ol_00000115.
37. Zeng Y, Min L, Han Y, Meng L, Liu C, Xie Y, Dong B, Wang L, Jiang B, Xu H, Zhuang Q, Zhao C, Qu L, Shou C. Inhibition of STAT5a by Naa10p contributes to decreased breast cancer metastasis. Carcinogenesis. 2014 Oct;35(10):2244-53. https://doi.org/10.1093/carcin/bgu132.

38. Medler TR, Craig JM, Fiorillo AA, Feeney YB, Harrell JC, Clevenger CV. HDAC6 Deacetylates HMGN2 to Regulate Stat5a Activity and Breast Cancer Growth. Mol Cancer Res. 2016 Oct;14(10):994-1008. https://doi.org/10.1158/1541-7786.MCR-16-0109.
39. Yoshida T, Zhang Y, Rivera Rosado LA, Zhang B. Repeated treatment with subtoxic doses of TRAIL induces resistance to apoptosis through its death receptors in MDA-MB-231 breast cancer cells. Mol Cancer Res. 2009 Nov;7(11):1835-44. https://doi.org/10.1158/1541-7786.MCR-09-0244.
40. Crippa E, Folini M, Pennati M, Zaffaroni N, Pierotti MA, Gariboldi M. miR-342 overexpression results in a synthetic lethal phenotype in BRCA1-mutant HCC1937 breast cancer cells. Oncotarget. 2016 Apr 5;7(14):18594-604. https://doi.org/10.18632/oncotarget.7617.
41. Hui L, Zheng Y, Yan Y, Bargonetti J, Foster DA. Mutant p53 in MDA-MB-231 breast cancer cells is stabilized by elevated phospholipase D activity and contributes to survival signals generated by phospholipase D. Oncogene. 2006 November 23;25(55):7305-10. https://doi.org/10.1038/sj.onc.1209735.
42. Gu Y, Bouwman P, Greco D, Saarela J, Yadav B, Jonkers J, Kuznetsov SG. Suppression of BRCA1 sensitizes cells to proteasome inhibitors. Cell Death Dis. 2014 December 18;5(12):e1580. https://doi.org/10.1038/cddis.2014.537.