Uncovering the role of mutations in Epithelial-to-Mesenchymal transition through computational analysis of the underlying gene regulatory network

Jorge Enrique Narváez-Chávez[1], Elena R. Álvarez-Buylla[2], and Juan Carlos Martínez-García[3]

[1] Oracle Global Mexico, CDMX, México.
jorge.narvaez@oracle.com
[2] Instituto de Ecología & Center of Complexity Sciences, UNAM, México.
eabuylla@gmail.com
[3] Automatic Control Department, Cinvestav-IPN, CDMX, México.
juancarlos.martinez@cinvestav.mx

Abstract

We illustrate here a systems-based computational analysis technique intended for uncovering dynamic properties of gene regulatory networks described in discrete Boolean terms. This through the use of the algebraic Semi-Tensor Product (STP), and with the purpose of exploring the regulatory consequences of genetic mutations in the modification of cell phenotypes. The proposed technique derives from the state-based reachability analysis of dynamic systems via the design of open-loop perturbative control schemes. We choose as a case-of-study the discrete Boolean network that describes in qualitative formal terms the transcriptional gene regulatory network that underlies the Epithelial-to-Mesenchymal transition in the context of epithelial cancer. We are particularly interested in qualitatively understanding the systems level consequences of mutations in specific genes that regulate the phenotypic transition. More specifically, we are interested in bringing to light potential preventive therapeutic interventions that promote the Mesenchymal-to-Epithelial transition.

1 Introduction

The increasing availability of experimental data from the exploration of cancer diseases forces the use of computational analysis techniques. This is not only with the purpose of satisfying the natural scientific curiosity, but above all with the priority purpose of providing information, in a fast, efficient and reliable manner, designed to promote the development of preventive therapeutic methods in the clinical setting. This is one of the major concerns of medical systems biology [1]. In this context, following a bottom-up approach, one of the mathematical modeling tools that has been remarkably successful is the one that uses discrete Boolean networks both to

describe in qualitative systems-based terms gene regulatory processes, and to uncover the basic generic principles that govern them [3, 14]. This has been particularly important in the case of cancer [12]. Indeed, the fundamental role played by Epithelial-to-Mesenchymal Transition (EMT) in the on set and progression of epithelial cancer has been made clear [6], and the gene regulatory network that underlies the cellular dynamics of EMT has been characterized. This is thanks to the use of systems-based mathematical modeling techniques based on experimental data [7]. In methodological terms, the construction of the discrete Boolean model of a given gene regulatory network asks for the integration of experimental data about the regulatory interactions of the involved biomolecular agents, described in terms of logic rules. Seen in the terms of the state-space theory, any gene regulatory network is in fact a dynamic system whose state vector, *i.e.* the vector descriptive variable, is defined by the set of genes that constitute it. At each instant, the state vector reports the activity status of each one of the genes. In discrete Boolean terms, this state of activity is represented in a binary way: if the gene is expressed in the cell of interest, its state of activity is represented by the number 1 and by 0 otherwise (*i.e.* the "true" and the "false" logic values). From a very general point of view, the system can stay in the same state it is in or change to a different state. When a system reaches a state that does not change in time, this state is known as an *attractor* state, and the regulatory logic of the system will cause it to stay there. Some systems known as *multistable systems* can have several attractor states. This means that the way their elements are related allows them to exist in different combinations that make the state of the system remain unchanged (we are not concerned here by cyclic attractors). As pointed out in [7], the core gene regulatory network that underlies EMT is in fact a multi-stable dynamical system whose state vector has 9 elements, and it has only tree fixed point attractors. These configurations characterize the phenotypes of the cells that are involved in the transition process: *epithelial cells*, *senescent epithelial cells*, and *stem-like mesenchymal cells* (with a tumorigenic potential). Adding stochasticity to this Boolean description, in order to take into consideration the stochastic nature of transcriptional regulation, and following the methodology exposed in [5], it has been established in [7] the temporal sequence and global order of cell-fate attainment pattern under the stochastic Boolean gene regulatory network model during epithelial carcinogenesis. The most probable temporal order of attractor attainment for a population of cells that initially show and epithelial phenotype correspond to the expected transition from an epithelial to a senescent to a mesenchymal stem-like phenotype. In this article we shall remain within a deterministic theoretical framework, keeping in mind that the dynamics of transcriptional gene regulatory processes is essentially stochastic.

Taking as a point of departure what is exposed in [7], and in order to explore the role of mutations in the Epithelial-to-Mesenchymal transition, we apply here a reachability analysis to the underlying core gene regulatory network. Our methodology is depicted in schematic terms in Figure 1) and is based on the transformation of the corresponding qualitative discrete Boolean description to an equivalente algebraic dynamic system via the so-called Semi-Tensor Product (STP) approach [3], as exposed in [2]. Then, the algebraic transformation allow us to find (when possible) perturbative control laws that promote transitions from initial given attractors to different final attractors. These control laws are then biologically interpreted in terms of gene gain or loss of function mutations. It is important for us to mention that we are not affirming here that the control strategy that we use can be implemented experimentally. This only shows us in qualitative terms the feasibility of inducing desired phenotypic transitions.
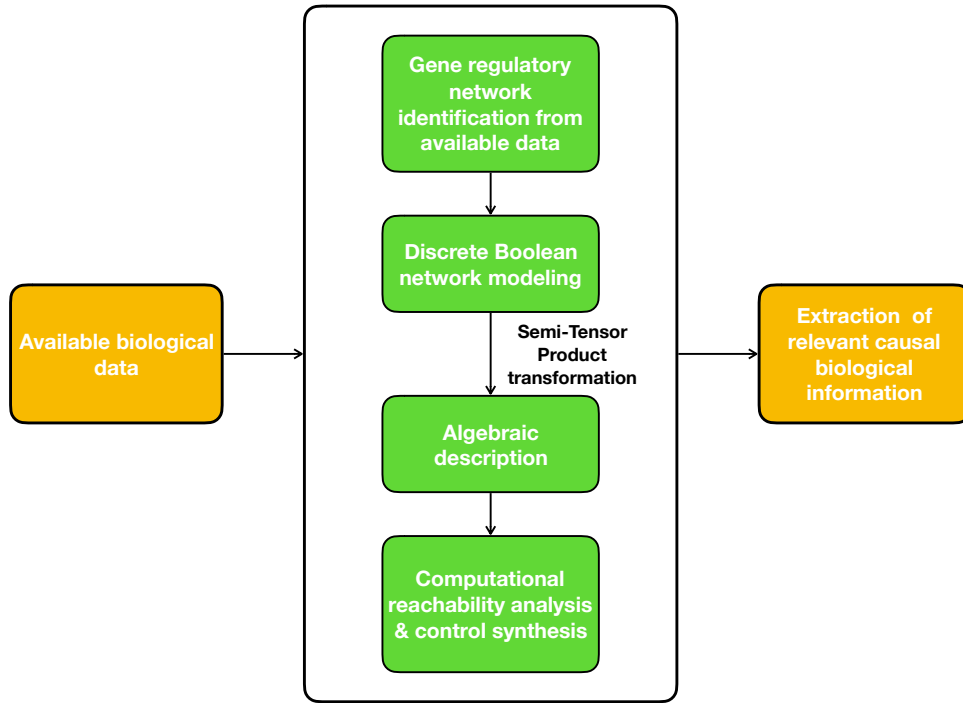
Figure 1: **Methodological framework**. From biological data we characterize the logical rules that underlie gene regulatory processes. The resulting networks are then described by means of Boolean networks, which are later transformed to discrete linear dynamical systems by means of the Semi-Tensor Product. It is then necessary to carry out the computational analysis of reachability. This in order to elucidate the role played by specific genes in the dynamics of the network. Control laws are then designed to explore the dynamics of possible transitions between the stable state configurations. Finally, relevant biological information is extracted.

2   Methods

Lets us first recall here in sketchy way the algebraic tools that we use to uncover the systems level consequences of mutations (loss or gain of function) in specific genes that regulate EMT in the context of epithelial cancer. We shall first recall how a synchronous Boolean network can be described as a linear discrete-time dynamic systems. Then, we shall show how attractor transitions can be attained via the design of control laws based on a reachability analysis.

2.1   A synchronous Boolean network (BN) as a linear discrete-time dynamic system

A synchronous Boolean network (BN) is a discrete-time dynamic system of n Boolean variables (in the case of a transcriptional gene regulatory network the integer n corresponds to the number of genes in the network). Its activity state (or *state of expression*) is given by a vector of Boolean variables $[x_1(t) \quad x_2(t) \quad \ldots \quad x_n(t)]^T$, and changes, in discrete fixed-steps as $x_i(t +$

1) $= f_i(x_1, x_2, \ldots, x_n)$, where $f_i : \mathcal{D}^n \rightarrow \mathcal{D}$ is the regulatory Boolean function related to the node x_i (a logical variable), which is built in agreement to the combinatorial action of its regulators; $\mathcal{D}$ stands for the set of logical values $[1, 0]$ (1 corresponds to "true" and 0 to "false"). In experimentally grounded Boolean gene regulatory network models, a particular case of Boolean networks, the set of f_i, $i = 1, 2, \ldots, n$, expresses the relationship between the genes that share regulatory interactions, involved in the process of interest and derived from experimental data. These functions consists of logical rules. The state vector of the network has 2^n different configurations. The Semi-Tensor Product approach (STP), which is based on the matrix expression of logic, is used to rewrite the set of Boolean functions into algebraic operators [3]. The Boolean network is thus converted into a conventional discrete-time linear system $x(t + 1) = Lx(t)$, where: $x(t) = \ltimes_{i=1}^n x_i(t) := x_1(t) \ltimes x_2(t) \ltimes \cdots \ltimes x_n(t)$ is the STP of the Boolean variables ($\ltimes$ denotes the left-semi tensor product[1]). $x_i \in \Delta_2$ represents the vector form of the i-th state ($\Delta_k := \left\{ \delta_k^1, \delta_k^2, \ldots, \delta_k^k \right\}$, with δ_2^i denoting the i-column of the identity matrix I_2); $L := M_1 \prod_{j=2}^n \left[\left(I_{2^n} \otimes M_j \right) \Phi_n \right]$ is called the *network transition matrix* of the BN (see the details in [3]), with M_i, $i = 1, \ldots, n$, standing for the structure matrix such that $x_i(t + 1) = M_i x(t)$ and $\Phi_n := \prod_{i=1}^n I_{2^{i-1}} \otimes \left[\left(I_2 \otimes W_{[2,2^{j-i}]} \right) M_r \right]$ is the group power-reducing matrix, $W_{[2,2^{j-i}]} = \left[I_{2^{j-i}} \otimes \delta_2^1, \ldots, I_{2^{j-i}} \otimes \delta_2^2 \right]$ is a swap matrix, and M_r is the power-reducing matrix.

2.2   The Boolean control network

In order to modify the dynamics of the BN, a set of Boolean inputs $u_j(t)$ are connected in some specific nodes. Thus, we have the Boolean control network given by $x_i(t + 1) = \bar{f}_i(x_1, x_2, \ldots, x_n, u_1, u_2, \ldots, u_m)$, where $u_1, u_2, \ldots, u_m$, are (logical) control inputs. We apply the STP to convert this network into a bilinear system. So:

$$x(t + 1) = \bar{L}u(t)x(t), \tag{1}$$

with $u(t) := \ltimes_{i=1}^m u_i$. Taking into consideration the case where the control is *a free Boolean sequence*, the Boolean control network is said to be controlled from x_0 to x_d at the s-th step if we can find a control sequence $u(t)$, $t = 0, 1, \ldots, s - 1$, such that the initial state $\ltimes_{i=1}^n x_i(0) = x_0$ can be driven to the destination state $\ltimes_{i=1}^n x_i(s) = x_d$. This control law exists if and only if $x_d \in \text{Col} \left\{ \tilde{L}^s x_0 \right\}$, where $\text{Col} \left\{ \tilde{L}^s x_0 \right\}$ stands for the set of columns of $\tilde{L}^s x_0$, and $\tilde{L}^s$ results from $\tilde{L} = \bar{L} W_{[2^n, 2^m]}$ (which allows the Boolean control network to be expressed as $x(t + 1) = \tilde{L} x(t) u(t)$) and the repetition that yields $x(s) = \tilde{L}^s x(0) u(0) u(1) \cdots u(s - 1)$. Moreover, if x_d is equal to the k-th column of $\tilde{L}^s x_0$, then the controls should be $u(0) u(1) \cdots u(s - 1) = \delta_{2^{ms}}^k$.

2.3   Attaining attractor transitions

In order to find the specific control law that drive the system from a given initial state to a given final state, we use the so-called controlability matrix of the system. Once the Boolean control network is given in its algebraic form (1), the controllability matrix is computed by:

$$C = \sum_{s=1}^{2^{m+n}} \sum_{i=1}^{2^m} \text{Blk}_i \left(\bar{L}^{(s)} \right) \in \mathcal{B}_{2^n \times 2^n}, \tag{2}$$

---

[1]If T is a *np*-dimensional row vector and X is a p-dimensional column vector, we split T into p equal blocks, say $T^1, \ldots, T^p$, which are $1 \times n$ matrices, then the left semi-tensor product of T and X is defined as $T \ltimes X := \sum_{i=1}^p T^i x_i$.

where $\bar{L}^{(s)}$ means we use the Boolean algebra in the product of matrices. $\text{Blk}_i\left(\bar{L}^{(s)}\right)$ and $\mathcal{B}_{2^n \times 2^n}$ stand for the i-th $n \times n$ block of matrix $\bar{L}^{(s)}$ and the set of $2^n \times 2^n$ Boolean matrices, respectively. The entries of C, *i.e.* c_{ij}, indicate whether the i-state is reachable from the j-state, under a set of admissible Boolean inputs. Then (Theorem 20 in [4] and Corollary 16.3 in [3]): (i) $x_d = \delta^i_{2^n}$ is reachable from $x_0 = \delta^j_{2^n}$ if and only if $c_{ij} > 0$; (ii) the system is controllable at $x_0 = \delta^j_{2^n}$ if and only if $\text{Col}_j(C) > 0$; (iii) the system is controllable if and only if $C > 0$. Since we are interested in transitions between attractors due to the manipulation of specific genes, we choose one of the attractors as the initial state and we determine which of the other attractors we can reach by applying a control sequence acting on a specific gene (the addition of the control input to the chosen gene can be done by considering it as a regulatory interaction between the gene and a control node, chosen in logical terms either considering the logical conjunction or the logical disjunction). In each case, we computed the control sequence considering for two given attractors each one of the genes. That is, we test the reachability for each and every one of the genes and in each case we determine the shortest control sequence that allows reaching the target attractor, if it belongs to the set of states reachable from the initial attractor. This methodology is summarized as follows:

- *Step 1.* Initialize $k = 1$. Then, given the Boolean gene regulatory network in its algebraic form, recover the experimentally observed attractors.

- *Step 2.* Add a control input u on the k-th node of the BN. If $k = n$, go to Step 5; otherwise, go to the next step.

- *Step 3.* Convert the control Boolean network into its algebraic form, and then compute the k-th controllability matrix C_k.

- *Step 4.* Set $k = k + 1$ and go back to Step 2.

- *Step 5.* Save $C_q k$, $k = 1, 2, \ldots, n$. Stop.

After a straightforward computation, the set of C_k, $k = 1, 2, \ldots, n$ contains all the *available state trajectories* among pairs of states, when the k-th gene is perturbed, if and only if $(c_{ij})_k > 0$. Furthermore, those trajectories can be restricted to a subset associated to the *attractor transitions*. In what follows we briefly expose what concerns the gene regulatory network that underlies EMT in the context of epithelial cancer.

3   Computational experiments

We present in this section the computer-based experiments that implement the reachability analysis. These experiments are intended to uncover the role of mutations in the dynamics of the core gene regulatory network that underlies EMT in the context of epithelial cancer.

The core regulatory network that underlies EMT in the context of epithelial cancer, includes 9 nodes: Snai2, ESE2, p16, E2F, Cyclin, TELasa, NFkB, Rb, and p53. The specific functional identity of these genes as well as the details of construction of the regulatory logic rules from available experimental data are exposed in [7]. We simplified the original set of logic rules to facilitate the encoding of the resulting discrete Boolean network in algebraic terms. The simplified set of logic rules is shown in Table 1. Once the three attractors of the network have been determined (corresponding to the epithelial phenotype, the senescent epithelial phenotype and the stem-like mesenchymal phenotype, as shown in Figure 2), a pair

Attractors with 1 state(s)

Figure 2: **Attractors of the core gene regulatory network that underllies EMT.** This figure shows the fixed-point attractors of the gene regulatory network that underlies EMT in the context of epithelial cancer (in the form of columns, from left to right, respectively, showing the activity states of the 9 genes involved). Which is to say: the epithelial phenotype, the senescent epithelial phenotype, and the stem-like mesenchymal phenotype. These attractors were obtained via the Boolnet R toolbox [10]. In the case of the attractor corresponding to the epithelial cell phenotype, the basin of attraction includes 92 states, the one corresponding to the stem-like mesenchymal phenotype has a basin of 288 states, and the one corresponding to the senescent epithelial phenotype has a basin of 132 states.

of attractors is chosen and the previously exposed methodology is applied. In order to explore all the possible transitions a control entry was added gene by gene and for each addition the reachability analysis was performed, as well as the consequent construction of the perturbative control action that promotes the transition towards the destination attractor in case it is feasible. The details of the application of the methodology are found in [11], as well as the computer code developed in Matlab$^{\text{TM}}$, which is based on the tools provided in [3].

Node	Simplified logic rule
Snai2	$(Snai2 \wedge NFkB) \vee \neg ESE2$
ESE2	$\neg Snai2 \vee (ESE2 \wedge \neg NFkB)$
p16	$(E2F \wedge p16) \vee (p16 \wedge \neg Snai2) \vee (p53 \wedge \neg Snai2) \vee (p53 \wedge \neg TELasa)$
E2f	$\neg p53 \wedge \neg Snai2 \wedge \neg Rb$
Cyclin	$(\neg Snai2 \wedge \neg p16) \vee (\neg ESE2 \wedge \neg p16 \wedge NFkB)$
TELasa	$\neg ESE2$
NFkB	$Snai2 \vee ESE2 \vee p16 \vee NFkB$
Rb	$p53 \vee (p16 \wedge Cyclin)$
p53	$(\neg Snai2 \wedge p16 \wedge \neg TELasa) \vee (\neg Snai2 \wedge \neg TELasa \wedge \neg NFkB)$

Table 1: **Simplified logic rules of the core gene regulatory network underlying EMT**. In this table the symbols $\neg$, $\wedge$, and $\vee$, stand for logical negation, logical disjunction, and logical conjunction, respectively.
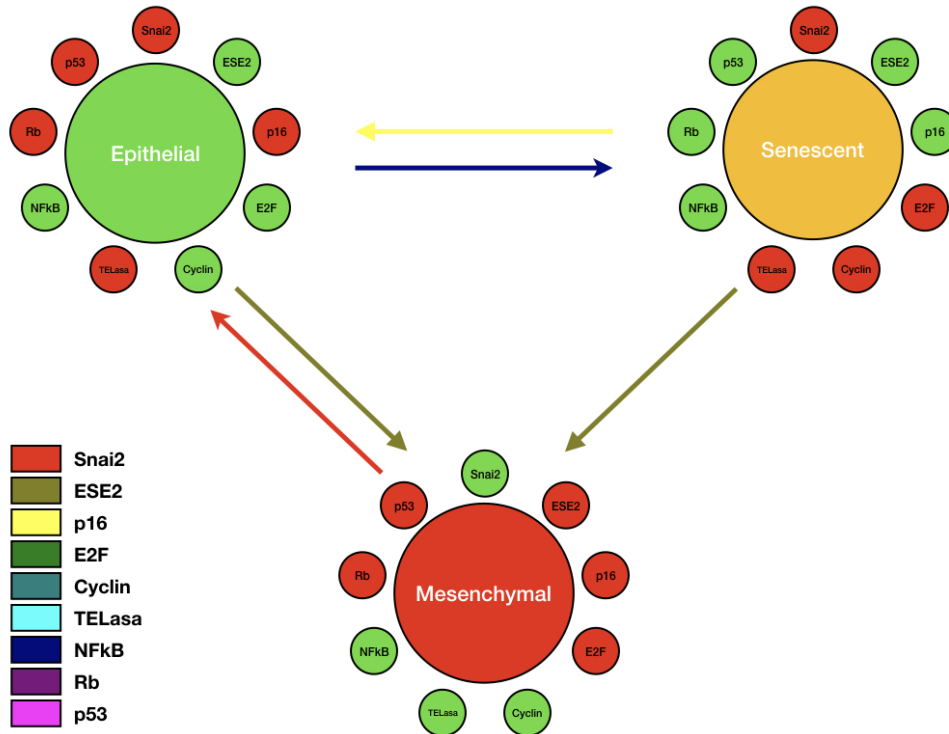
Figure 3: **Promotion of phenotypic transitions.** This figure shows the attractors of the discrete Boolean networks that describes the gene regulatory network that underlies EMT, which correspond to the gene expression identities of the epithelial cells, senescent epithelial cells, and the stem-like mesenchymal cells. The 9 genes that make up the network are represented by the small labeled circles that are around the circle that represents the cell phenotype. The activity status is indicated in green if it corresponds to an expressed gene and in red if it is not. The arrows represent transitions promoted by the manipulation of the activity state of a specific gene via the action of a free sequence controller applied through logical conjunction. The manipulated gene is shown on the left of the image by means of a small box. The color of this corresponds to that of the promoted transition. To simplify our exposition we only depict the transitions that results from the manipulation of p16, NFkB, ESE2, and Snai2.

4   Results

Note that in our methodology the addition of an input with a constant value equal to 0, this is corresponding to the logical value "false", to a given gene that is active, through a regulation action defined by the logical conjunction, simulates the loss-of-function mutation of the gene. As for the gain-of-function mutation of a given gen, this can be simulated by a constant regulatory action corresponding to logical disjunction acting on the gene in the inactive state at the given attractor. Due to its interest in terms of the prevention of epithelial cancer, we have been interested in this study particularly in the transition from the stem-like mesenchymal cell phenotype to the epithelial cell phenotype (this because the stem-like mesenchymal cell phenotype has a tumorigenic potential). In Figure 3 we show what results from our study in this
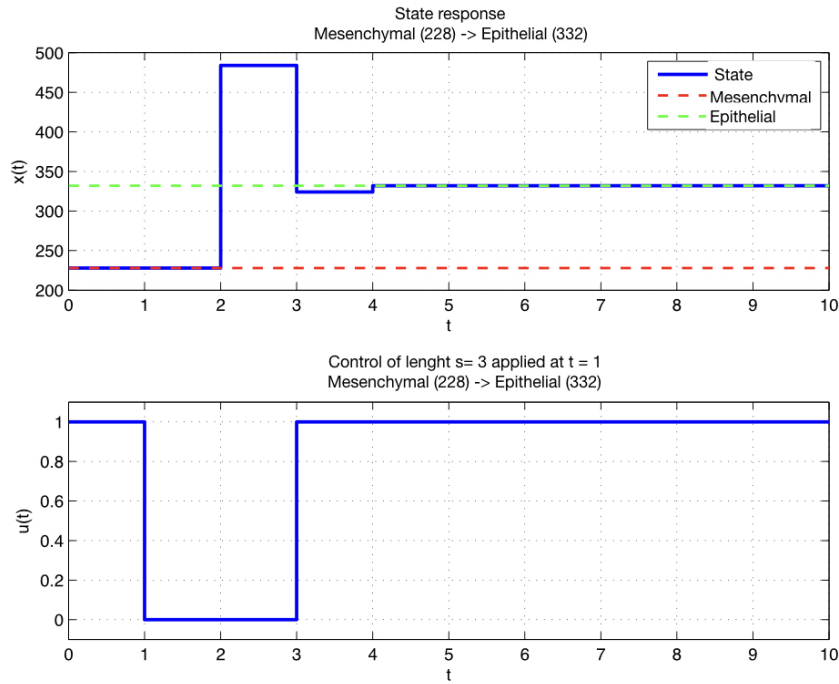
Figure 4: **Control sequence designed to promote a mesenchymal-to-epithelial transition.** This figure shows the free sequence control action exerted on the Snai2 gene to promote the transition to the epithelial phenotype from the stem-like mesenchymal phenotype. The stem-like mesenchymal phenotype is represented by the integer in base ten corresponding to the binary representation of the corresponding attractor, this is the number 228. The epithelial phenotype is represented by the number 332. The graphic at the top shows the transition, while the one below shows the control action. As can be seen, a loss of function is applied to Snai2 at instant 1, which translates into the abandonment of the mesenchymal phenotype at instant 2. The state of loss of function of Snai2 is maintained for two instants and its recovery does not prevent it from the phenotypic state now corresponds to the epithelial one.

case. As can be seen, it is possible to promote the transition from the undesirable mesenchymal phenotype to the epithelial phenotype by ensuring the loss-of-function mutation of the Snai2 gene. Figure 4 shows a control sequence that promotes the desired transition.

5   Discusion

As shown in Figure 2, the basin of attraction of the stem-like mesenchymal phenotype includes 288 states. This corresponds to the 56.5% of the total possible phenotypic configurations, *i.e.* $2^9 = 512$. This means that, in stochastic fluctuation terms, it is very difficult for there to be a spontaneous transition from the stem-like mesenchymal phenotype to the most desirable epithelial phenotype or to the senescent epithelial phenotype. However, as previously shown, the transition can be in fact induced by a loss-of-function mutation of the Snai2 gene in the attractor corresponding to the stem-like mesenchymal cell phenotype (see 4). Experimental

evidence has been found on the possible induction of a Mesenschymal-to-Epithelial transition that attenuates the malignancy of cancer [13]. This experiment was carried out through the indirect repressive manipulation of Snai2 through the introduction of reprogramming factors in squamous cell carcinoma (SCC) cells (the indirect repressive manipulation do not involve direct mutation of the gene). This shows how relatively simple qualitative studies based on computer-based analysis, such as the one presented in this paper, can motivate the exploration of new therapeutic strategies. In terms of the control strategy used in our study to promote transitions, between given attractors (stable cell phenotypes), and that supports our reachability analysis, it is important to mention that it corresponds to the class of perturbative control actions. The reachability analysis could also be performed using non-perturbative control approaches, *e.g.* via the feedback vertex set [16], stable motifs [15], through algebraic methods [9], and also via canalization-based control [8] to identify portier genes.

6   Conclusions

To conclude, since the transcriptional core gene regulatory network that underlies EMT is not isolated from the rest of the cell's transcriptional regulatory machinery, the class of reachability studies, such as the one we show here, can be used to identify nodes in the network that act to vehicle regulation processes supported by the interconnection of the network with other transcriptional regulatory networks, such as those underlying the inflammatory response and the cell-cycle, as well as metabolic processes and of various kinds. We must point out that our methodology is essentially qualitative and as such complementary to other approaches. These include bottom-up approaches that use perturbation theory to describe regulatory processes in differential terms), as well as top-down genomic approaches. The main advantage that characterizes methods like the one we present is that it allows us to explore hypotheses around phenotypic plasticity, without burdening ourselves with the parametric complexity that characterizes standard quantitative approaches.

Acknowledgements

References

[1] Elena R Álvarez-Buylla, Juan Carlos Martínez-García, José Dávila-Velderrain, Elisa Domínguez-Hüttinger, and Mariana Esther Martínez-Sánchez. *Modeling Methods for Medical Systems Biology: Regulatory Dynamics Underlying the Emergence of Disease Processes*, volume 1069. Springer, 2018.

[2] José E Chairez-Veloz, José Luis Caldu-Primo, José Dávila-Velderrain, Alberto Soria-López, Elena R Álvarez-Buylla, and Juan Carlos Martínez-García. Controllability analysis of the core gene regulatory network underlying epithelial-to-mesenchymal transition in the context of epithelial cancer. In *International Conference on Systems Biology (ICSB 2018), Lyon, France, October,28-November 1*, 2018.

[3] Daizhan Cheng, Hongsheng Qi, and Zhiqiang Li. *Analysis and Control of Boolean Networks. A Semi-tensor Product Approach.* Communications and Control Engineering. Springer, 2011.

[4] Daizhan Cheng, Hongsheng Qi, and Yin Zhao. On boolean control networks — an algebraic approach. *IFAC Proceedings Volumes*, 44(1):8366–8377, 2011. 18th IFAC World Congress.

[5] José Davila-Velderrain, Juan Carlos Martinez-Garcia, and Elena R. Alvarez-Buylla. Modeling the epigenetic attractors landscape: toward a post-genomic mechanistic understanding of development. *Frontiers in Genetics*, 6:160–160, 2015.

[6] Raghu Kalluri and Robert A. Winberd. The basics of epithelial-mesenchymal transition. *Journal of Clinical Investigation*, 119(6), June 2009.

[7] Luis F. Méndez-López, Jose Davila-Velderrain, Elisa Domínguez-Hüttinger, Christian Enríquez-Olguín, Juan Carlos Martínez-García, and Elena R. Alvarez-Buylla. Gene regulatory network underlying the immortalization of epithelial cells. *BMC Systems Biology*, 11(1):24, 02 2017.

[8] David Murrugarra and Elena S. Dimitrova. Molecular network control through boolean canalization. eurasip j. *Bioinformatics Syst. Biol*, 9, 2015.

[9] David Murrugarra, Alan Veliz-Cuba, Boris Aguilar, and Reinhard Laubenbacher. Identification of control targets in boolean molecular network models via computational algebra. *BMC systems biology*, 10(1):1–11, 2016.

[10] Christoph Müssel, Martin Hopfensitz, and Hans A. Kestler. Boolnet – an r package for generation, reconstruction and analysis of boolean networks. *Bioinformatics*, 26(10):1378–1380, 2010.

[11] Jorge Narváez-Chávez. Controllability analysis of the gene regulatory network of the epithelial-mesenchymal transition in the development of epithelial cancer (in spanish). Master's thesis, Centro de Investigación y de Estudios Avanzados del I.P.N. Departamento de Control Automático, México (https://repositorio.cinvestav.mx/handle/cinvestav/2192), 2018.

[12] Sriganesh Srihari, Venkatesh Raman, Hon Wai Leong, and Mark A. Ragan. Evolution and controllability of cancer networks: A boolean perspective. *EEE/ACM Trans Comput Biol Bioinform*, 2014.

[13] Mikiro Takaishi, Masahito Tarutani, Junji Takeda, and Shigetoshi Sano. Mesenchymal to epithelial transition induced by reprogramming factors attenuates the malignancy of cancer cells. *PLOS ONE*, 11(6):e0156904, 06 2016.

[14] Jing Yang and Robert A. Weinberg. Epithelial-mesenchymal transition: At the crossroads of development and tumor metastasis. *Developmental Cell*, 14(6):818–829, 06 2008.

[15] Jorge GT Zañudo and Réka Albert. Cell fate reprogramming by control of intracellular network dynamics. *PLoS computational biology*, 11(4):e1004193, 2015.

[16] Jorge GT Zañudo, Gang Yang, and Réka Albert. Structure-based control of complex networks with nonlinear dynamics. *Proceedings of the National Academy of Sciences*, 114(28):7234–7239, 2017.