

SCIENTIFIC REPORTS



OPEN

Determining Relative Dynamic Stability of Cell States Using Boolean Network Model

Jae Il Joo¹, Joseph X. Zhou^{2,3}, Sui Huang² & Kwang-Hyun Cho¹ 

Cell state transition is at the core of biological processes in metazoan, which includes cell differentiation, epithelial-to-mesenchymal transition (EMT) and cell reprogramming. In these cases, it is important to understand the molecular mechanism of cellular stability and how the transitions happen between different cell states, which is controlled by a gene regulatory network (GRN) hard-wired in the genome. Here we use Boolean modeling of GRN to study the cell state transition of EMT and systematically compare four available methods to calculate the cellular stability of three cell states in EMT in both normal and genetically mutated cases. The results produced from four methods generally agree but do not totally agree with each other. We show that distribution of one-degree neighborhood of cell states, which are the nearest states by Hamming distance, causes the difference among the methods. From that, we propose a new method based on one-degree neighborhood, which is the simplest one and agrees with other methods to estimate the cellular stability in all scenarios of our EMT model. This new method will help the researchers in the field of cell differentiation and cell reprogramming to calculate cellular stability using Boolean model, and then rationally design their experimental protocols to manipulate the cell state transition.

Cell state transition is at the core of many cell biology processes in metazoan, such as cell differentiation, cell stress response, epithelial-to-mesenchymal transition (EMT) in development, but also in artificial cell type reprogramming, such as generation of iPSC (induced pluripotent stem cell) from differentiated cells or directed differentiation of multipotent cells, including stem cells and iPSC, into specialized cell types¹⁻⁴. Cell state transitions can be experimentally induced by ectopic control of the activity of key regulatory genes or by providing the appropriate environmental signals⁵⁻⁸. The conventional approach to identify these molecular levers to trigger the desired state transitions has been to use an educated guess based on known functions of key regulators or of relevant regulatory pathways and typically involves trial and error (such as high-throughput screening in the extreme case⁹). However, subsequent applications of the identified regulators, e.g. overexpressing or suppressing of a particular set of genes, and the use of empirical cytokine cocktails¹⁰, usually achieve low efficiency and necessitate selection of cells. The latter is further compromised by induction of undesired transitions, e.g. of stem cells into not only the intended lineage, but into “wrong” neighboring lineages¹¹⁻¹⁴. For example, generation of iPSC from differentiated cells (“reprogramming”) has an efficiency that is usually below 1% when using the classic Yamanaka protocol¹⁵, i.e. 1% of cells reach the desired destination state, the iPSC. Similarly, the efficiency of reprogramming pancreatic exocrine cells to beta cells using a set of three TFs is below 25%⁷. In reprogramming human embryonic stem cells human to retinal photoreceptor cells, less than 20% of cells at each step differentiated to desired cell types^{5,6}. The low efficiency of directed (de)differentiation is in line with the observation that the cells of the clonal cell population are heterogeneous¹⁶, suggesting the possibility that they may respond in distinct manners to the inducing signal applied.

The rationale for reprogramming protocols being derived from pathway diagrams involving “fate-determining” master regulators embodies linear causation and only allows for qualitative deterministic predictions of the fate outcome of manipulations. Such information on functions of regulatory factors and pathways cannot take into consideration the dynamics, including the ubiquitous stochastic fluctuations of gene activities¹⁷;

¹Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, 34141, Republic of Korea. ²Institute for Systems Biology, Seattle, WA, USA. ³Kavli Institute for Theoretical Physics, UC Santa Barbara, California, USA. Jae Il Joo and Joseph X. Zhou contributed equally to this work. Correspondence and requests for materials should be addressed to S.H. (email: shuang@systemsbiology.org) or K.-H.C. (email: ckh@kaist.ac.kr)

hence it does not allow prediction of a graded response, such as the submaximal efficiency ($\ll 100\%$) of desired state transitions. The pathway paradigm also cannot predict the pleiotropic and stochastic effects that lead to the diversification of cells into several lineages in response to the reprogramming perturbation. Here we propose a theoretical framework that goes beyond deterministic linear causation of regulatory pathways and takes into account entire gene regulatory network (GRN) and their dynamics. The latter refers to how GRN imposes constraints on the collective change of gene expression (since gene activities are coupled across all the genes) and determines trajectories of cell states, thereby affording robustness to cell states by constraining some of the random fluctuations, while still allowing cells to escape the constraints by the GRN. This systems biological approach is not only grounded on fundamental principles of the theory of dynamical systems but also offers guidance for the design of cell reprogramming protocols that can help understand the stochastic diversification of cell response to regulatory factor manipulations and can be used to improve efficiency of reprogramming into the desired state.

In brief, any observable cell state, such as a cell type, manifests a characteristic genome-wide gene expression pattern which allows the expressed genes to perform specific biological functions and to implement the characteristic phenotype of a cell. We use a vector $\mathbf{x} = (x_1, x_2, \dots, x_n)$ to represent the gene expression pattern consisting of gene expression values (“gene activities”) x_i of the n relevant genes associated with the state of a cell. Gene expression patterns arise because the individual gene activities are coordinated by the gene regulatory network. As a consequence of the gene regulatory interactions, each gene x_i can only change its value as a function of all the genes in the GRN or a particular subset; hence $\dot{x}_i = F_i(x_1, x_2, \dots, x_n)$, or for the entire system of the n genes in the network $\dot{\mathbf{x}} = \mathbf{F}(x_1, x_2, \dots, x_n)$ which is an n -dimensional system of ODEs (ordinary differential equations). An attractor \mathbf{x}^* , a stable fixed point or a limit cycle, corresponds to a distinct nominal cell state¹⁸, such as an embryonic stem cell, a liver cell, a stressed cell, etc., defined by its characteristic gene expression profile. In physics, \mathbf{x}^* is a non-equilibrium stable steady-state in the GRN as an open thermodynamic system.

The GRN of a genome produces a multitude of attractor states that have long been proposed to explain the discreteness and robustness of all the cell types of a metazoan organism¹⁹. The developmental relationship between these attractors is often depicted using Waddington’s pictorial epigenetic landscape, in which attractor states would correspond to the valleys²⁰. The epigenetic landscape, however, is more than a metaphor because a GRN’s dynamics can indeed be represented as a quasi-potential function $U(\mathbf{x})$ that can be theoretically constructed from the systems equations describing the GRN’s dynamics, $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}) + \boldsymbol{\eta}(\mathbf{t})$, which $\boldsymbol{\eta}(\mathbf{t})$ represents unbiased noise function (Fig. 1a)^{21–28}. In the quasi-potential landscape, each point represents a gene expression state $\mathbf{x} = (x_1, x_2, \dots, x_n)$ and the associated quasi-potential function $U(\mathbf{x})$; $U(\mathbf{x})$ informs about the system’s ability to undergo non-equilibrium state transitions even if $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}) + \boldsymbol{\eta}(\mathbf{t})$ is not a gradient system. In brief, from the Freidlin-Wentzell large deviation theory based on stochastic system²⁹, the “relative dynamic stability” can be defined by comparing probability of transitions between attractors which involve a noise-driven exit from an attractor to initiate a move into a neighboring attractor. Mathematically, when we take the limit of noise to zero, it has been shown that $U(\mathbf{x})$ of a deterministic system $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x})$ is related to transition probability of Freidlin-Wentzell stochastic system and represents the least action needed to work against the regulatory constraints imposed by the GRN (which produce the attractor states in the first place)²¹. This theoretical approach assumes Markov property and ergodicity, but the real biological system may not exactly satisfy Markov and ergodic assumptions, for instances, DNA methylation leads to memory effects (non-Markovian) and cells may not visit all possible gene expression states (nonergodic). However, the dynamical system $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x})$ is a first order approximation to explain the fundamental principle why the same genome can produce multicellular organism.

Since the construction of continuous models of biological system using $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x})$ requires the specific functional form of F_i for each gene i in the GRN and associated parameters, which are typically unknown, discrete Boolean network models are frequently used to model GRNs^{30–33}. Experimental work and known regulatory modalities can suggest an appropriate Boolean function in lieu of F_i , hence facilitating model construction^{34–36}. The problem is that a quasi-potential landscape $U(\mathbf{x})$ cannot be easily defined in the discrete model. Instead of the quasi-potential $U(\mathbf{x})$, several quantities have been developed to define the relative dynamic stability of a cell state in discrete Boolean network models^{30,31,33,37}. In the simplest formalism, the basin size of an attractor serves as an approximation of the relative dynamic stability of a given attractor, that is, the number of discrete states converging to that attractor (Fig. 1b)³³. Even though the basin size is defined in deterministic condition, a large basin may imply that perturbed states of an attractor \mathbf{x}^* by noise (e.g. “bit flip” of the value of a gene x_i from 0 to 1 or vice versa) would remain in the basin of the attractor. However, basin size is an inadequate measure for relative attractor stability since attractors with the equal size basin, depending on many other properties, may exhibit distinct resilience to random perturbations^{34,38,39}.

As a more appropriate relative stability index, the steady-state probability distribution function (SSP) has been proposed as an estimate of a quasi-potential in discrete Boolean network models³⁷. SSP is the probability density function $P(\mathbf{x})$ for the probability that in a stochastic system, at equilibrium (the steady state) a cell is observed to stay around the attractor. An inverse function of $P(\mathbf{x})$ can then be visualized as the depth of an attractor basin in a 3D landscape (Fig. 1c). However, the resulting landscape reflects steady-state properties; to establish a stability measure that is not limited to behaviors at equilibrium but quantifies the probability of exit from the attractor, the mean first passage time (MFPT) has been proposed to be used in a non-equilibrium state; MFPT is the average time required for a cell to exit from an attractor state so as to reach another attractor state, which considers all trajectories of cell state transitions that could be used in this process (Fig. 1d)³¹. From the MFPT, the stability difference (relative stability) and thus the preferred transition direction between two attractor states can be inferred. To generalize the relative stability from an attractor to an attracting basin, we can consider transitions between not only attractor states but also any “non-attractor” states (e.g. on transients) which may be located in separate attracting basins (Fig. 1e)³⁰. We defined the probability for such a generalized transition as basin transition rate (BTR). Although these various measures for dynamic stability had been developed for the Boolean model, to our

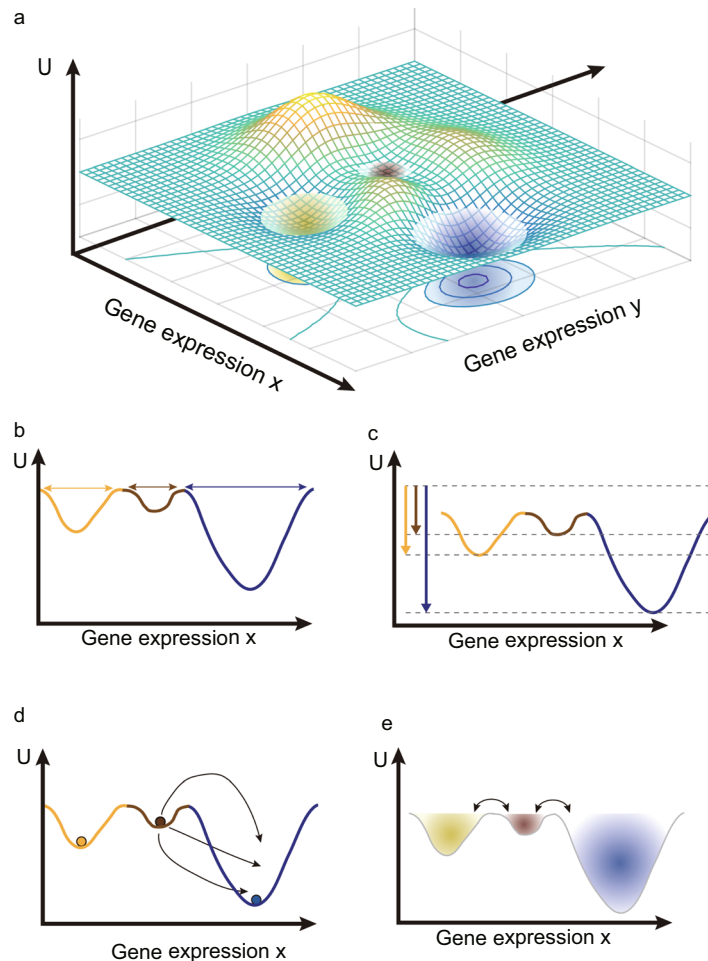


Figure 1. The scheme of quasi-potential landscape and relative dynamic stability. (a) A three-dimensional quasi-potential landscape. A point of XY plane is corresponding to a gene expression state of cell and elevations of each point $U(x, y)$ represent the quasi-potential. Three attractors are local minimums in the quasi-potential landscape. (b–e) are schematic descriptions of four different methods to calculate the relative dynamic stability of the attractors. For clarity, we used two-dimensional representation of the quasi-potential landscape. (b) The relative dynamic stability is defined by the basin size. (c) The dynamic stability is defined by the steady state probability distribution function. (d) The dynamic stability is defined by the mean first passage time. (e) The relative dynamic stability is defined by the basin transition rate.

best knowledge, these relative stability measures have not been systematically compared for a concrete biological question and have been rarely applied in the large-scale Boolean models due to their computational complexities.

In this paper, we constructed a simple discrete Boolean network model for the epithelial (E)-to-mesenchymal (M) transition (EMT), a biologically important cell state transition that is generic (not tissue specific) and plays a role in development, wound healing and regeneration of various processes and also has been implicated in cancer progression⁴⁰. Recently mathematical models using ODEs to model the underlying gene regulatory network that governs this circuit as well as experimental observations suggest that EMT is not simple binary switching between two stable states E and M, but may involve metastable intermediate states, referred to as hybrid states^{41–44}. The relative dynamic stability of three cell states (Epithelial, Mesenchymal and Epithelial/Mesenchymal Hybrid cells) can be generated by the GRN which also determines the transition probabilities. Using a Boolean network to model EMT, we systematically compared the above four measurements of relative dynamic stability (basin size, SSP, MFPT and BTR) and difference between them as well as the reasons for these differences. From that, we propose a new stability index for relative dynamic stability of attractor state. Our results could be useful in studies of cell state transitions, helping to choose the appropriate state stability quantity that is best suited for particular biological questions.

Results

Boolean network model for EMT. To compare the different measures for relative dynamic stability, we reconstructed a Boolean network model for a gene regulatory circuit that controls EMT and used the methods above to analyze the relative dynamic stability of various cell attractors in the model. EMT is a process by which epithelial cells lose their cell polarity and cell-cell adhesion and gain migration and invasion capabilities to become mesenchymal cells. This process has been implicated in tumor metastasis, the major cause of death of cancer. More recently, in studies of EMT in breast cancer, a third distinct cell state has been observed in addition to the epithelial (E) cell state, and the mesenchymal cell (M) state: an intermediate “hybrid” (H) cell state which displays both epithelial and mesenchymal features with regard to gene expression patterns^{41,42}. The existence of this additional state with promiscuous gene expression pattern has been postulated independently based on dynamical gene circuit models^{43,44}, and it is likely that additional distinct states are produced by the multi-stable dynamics of the underlying gene regulatory network. The three-state model provides an ideal benchmark system for our comparison of the four measures of dynamic stability because having more than two stable states offers a richer opportunity to order the attractors according to dynamic stability of and to examine its biological significance.

Establishing a Boolean network model requires that one constrains the gene regulatory interactions (network topology) and the gene expression profiles of each attractor to the experimental observations. The third piece of information is also needed: the Boolean logic function controlling expression of each gene is currently not readily observed and must be inferred (see below). We first surveyed the literature⁴³ and identified four key genes regulating EMT: *miR-34* ($\mu34$), *miR-200* ($\mu200$), *Snail* and *Zeb* (Fig. 2a). They form two mutually inhibitory toggle switches ($\mu34$ vs. *Snail* and $\mu200$ vs. *Zeb*). *Snail* and *Zeb* form a coherent feed-forward loop and a positive feed-back loop involving $\mu200$ and $\mu34$, respectively. One input node is added to four gene circuit to represent the external signals which induce EMT, thus obtaining a $n = 5$ nodes GRN. Before defining the Boolean logic function of each node, we first defined the Boolean state \mathbf{x}^* of the three attractors. $\mu200$ is an epithelial marker whereas *Zeb* is a mesenchymal marker, and we assumed that a cell in the hybrid state would express both markers of epithelial and mesenchymal cells. Thus, we defined the epithelial state, hybrid state and mesenchymal state as (Input, *Snail*, $\mu34$, *Zeb*, $\mu200$) = (0, 0, 1, 0, 1), (1, 1, 0, 1, 1) and (1, 1, 0, 1, 0) respectively (Fig. 2b).

We tried to identify Boolean functions of each gene based on the experimental observation and known biological mechanism of genes, but they failed in producing the defined three attractors. Hence, we generated all possible Boolean functions to identify the combination of Boolean functions associated with all the five nodes which produces the three defined attractors as point attractor states. As there are a large number of possible Boolean functions, to reduce the search space of Boolean functions, we used only a subset of Boolean function called nested canalizing Boolean functions (NCBFs, see Methods for details). The outputs of NCBFs can be determined by subsets of input arguments and so they exhibit stable dynamics which as usually observed in biological systems^{31,45–47}. Moreover, we added the proposed self-regulating link to the node $\mu200$ (Fig. 2a) which helped the mutual inhibitory circuit of *Zeb* and $\mu200$ to generate three stable attractor states⁴³. We were able to identify three Boolean models consisting of NCBFs which produced the three defined point attractors. However, two models included wrong NCBFs that $\mu34$ activated *Snail* and $\mu200$ could not inhibit *Zeb*, respectively, which were different from the known role of $\mu34$ and $\mu200$. Thus, we used the remained set of NCBFs to EMT Boolean model which agreed well with the known biology of genes (See details of the logic tables in Supplementary Table S1). Including the input node which sustains its initial value, our EMT Boolean model has 32 (2^5) states. These 32 states converge to three attractors which we defined as the epithelial, mesenchymal and hybrid attractor states, dividing the state space into three basins of attraction (Fig. 2c).

Validation of EMT Boolean network model with knock-out and overexpression experiments. Besides showing that the EMT Boolean model is valid in that it correctly represents the three stable cell states, we also validated our model against results from gene knock-out (KO) or over-expression (OE) experiments. We simulated the genetic mutations by setting the Boolean function of the manipulated node (mutated gene) in the network to return 0 (KO) or 1 (OE), irrespective of values of the other nodes. The genetic mutations caused new attractors to appear - we defined these novel attractor states using the values of $\mu200$ and *Zeb*: an epithelial-like attractor (EL, $\mu200 = 1$, *Zeb* = 0), a mesenchymal-like attractor (ML, $\mu200 = 0$, *Zeb* = 1), a hybrid-like attractor (HL, $\mu200 = 1$, *Zeb* = 1), and an undefined attractor (ND, $\mu200 = 0$, *Zeb* = 0).

Simulation results were compared with experimental observations (Table 1). Increased expression of *MIR200* family transcripts has been reported to induce mesenchymal-to-epithelial transition (MET, the reverse of EMT), and to suppress EMT, invade, and metastasize^{48,49}. Transfection of mesenchymal cells with *MIR200* family caused morphological change from mesenchymal to epithelial-like form with many cells aggregating together in groups⁵⁰. Our EMT Boolean network recapitulated this observation: the simulations showed that the mesenchymal attractor state disappeared if we set the value of the $\mu200$ node to 1. On the other hand, anti-*MIR200* treatment resulted in increased metastases *in vivo*, and caused an increase in the migration of cells and mesenchymal characteristics *in vitro*^{48,50}. This finding was reproduced in the simulation where setting to the $\mu200$ node value 0 resulted in a state space that had only mesenchymal and mesenchymal-like attractors. Similarly, when *MIR34a* was ectopically introduced into cells, *Snail1* protein levels and E-cadherin promoter activity decreased to baseline and mobility of the cell was decreased⁵¹; our model reproduced this observation because an epithelial point attractor and a cyclic attractor with repeating non-mesenchymal state was generated by setting the *MIR34* node to the value 1. Conversely, re-expression of *SNAIL* resulted in the motile response of *MIR34a*-expressing cell⁵¹ which agrees with our corresponding simulation of up-regulating *Snail*: the resulting state space contained mesenchymal and hybrid attractors but no epithelial attractor. Lastly, simulation of over-expression of *Zeb* resulted in the removal of the epithelial attractor, which is consistent with the observed association of the induction of EMT in breast cancer cells with increased expression of *ZEB1/ZEB2*⁴⁸. These comparisons with experimental findings

Node	UP/ DOWN	Experiment	Simulation results		Ref.
			Epithelial	Mesenchymal	
$\mu 200$	UP	Reversal of EMT	Yes	No	48
$\mu 200$	DOWN	EMT	No	Yes	48
<i>Zeb</i>	UP	Induction of EMT	No	Yes	48
$\mu 200$	UP	Abrogate EMT	Yes	No	49
$\mu 200$	UP	MET	Yes	No	50
$\mu 200$	DOWN	EMT	No	Yes	50
$\mu 34$	UP	MET	Yes	No	51
<i>Snail</i>	UP	EMT	No	Yes	51

Table 1. Cross validation of EMT model from knock-out or over-expression experiments. KO and OE simulation results of EMT Boolean model are validated by experiment results. Simulation results column shows existence of epithelial attractor and mesenchymal attractor.

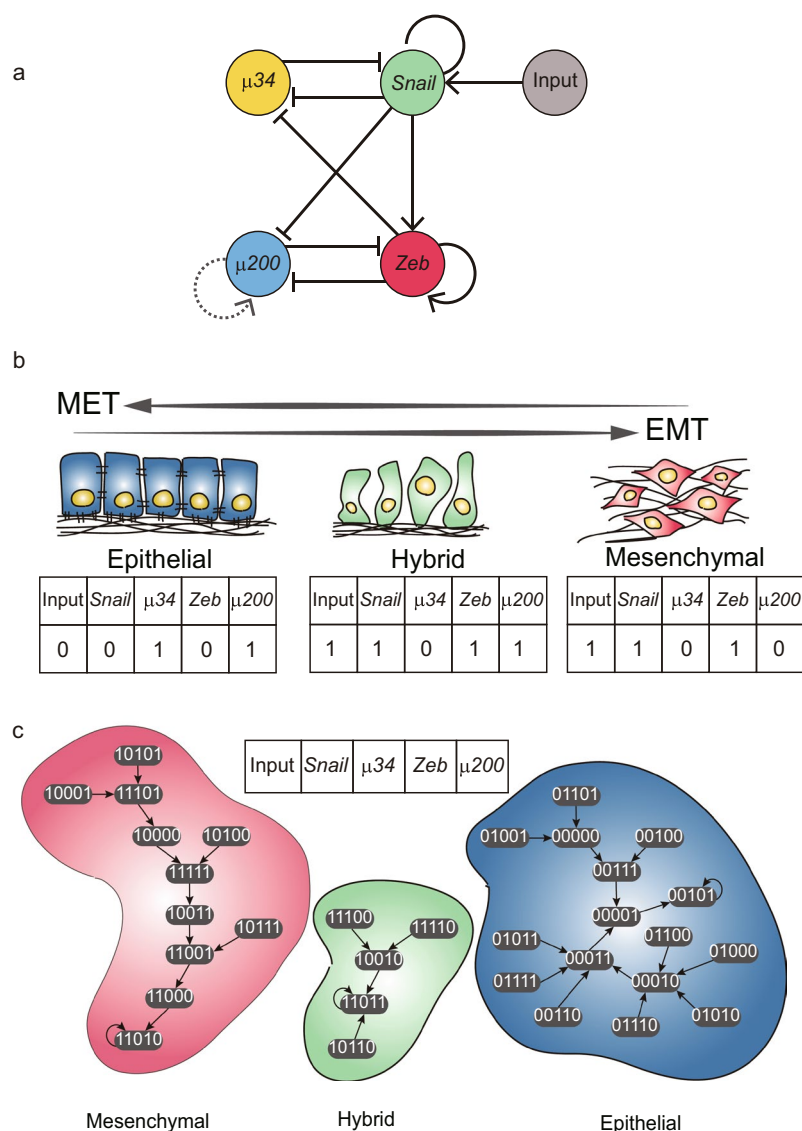


Figure 2. The EMT Boolean network model and three cell attractors and attracting basins. (a) A network structure for the EMT model. The sharp arrows represent the activating regulations while the blunt arrows represent the repressing regulations between nodes. A dashed link is for a proposed regulation in this model. (b) The definitions of three cell attractors in Boolean state. (c) Transition map of the EMT Boolean model. Each ellipse represents one Boolean state of the model. Without internal and external perturbation, state changes of the EMT model follow arrows due to the Boolean functions. There are three cell attractors and each basin of them is distinguished by different colors: blue for epithelial cells, red for mesenchymal cells and green for hybrid cells.

demonstrated that our EMT Boolean network model not only correctly generated the three cell attractor states that produce biological characteristic gene expression patterns but also recapitulated the effects of genetic mutations of the gene regulatory network that governs EMT.

Computing the relative dynamic stability of the three cell attractors in EMT using the four known methods. Using the validated EMT Boolean model, we calculated the relative dynamic stability of cell attractors based on the four methods described above: (i) the attractor basin size, (ii) the steady state probability distribution function (SSP), (iii) the mean first passage time (MFPT) and (iv) basin transition rate matrix (BTR). The purpose is to compare the different methods to measure the stability of the attractors, i.e. their ability to determine the relative stability. For calculating SSP, MFPT and BTR, we constructed a stochastic model, which is identical to the deterministic EMT Boolean model biologically, by introducing biological gene expression noise $\eta = 0.01$ to the deterministic model (for details, see State transition matrix and steady state probability density function in Methods). The relative attractor stability has a functional (measurable) consequence in that it affects the preferred directionality of attractor transitions (EMT, MET) and hence can be benchmarked against experiments. This will allow us to determine which of these four methods best matches experimental results, which in turn will help to design better measures for the relative dynamic stability.

The attractor of the epithelial cell state has 16 network states (50%) in its basin (Fig. 3a). The mesenchymal attractor has a basin of 11 states (34%) whereas the hybrid attractor has a basin of only 5 states (16%). Using the attracting basin size as a measure, we can rank the attractors for their relative dynamic stability: the epithelial cell is the most stable one. For the relative dynamic stability calculated using SSP, only the attractor states have meaningful values whereas other states have almost zero values. The epithelial, mesenchymal, and hybrid attractors have the steady-state probabilities P_{ss} as 0.437, 0.272, and 0.168 respectively (Fig. 3b). The steady state probability distribution P_{ss} shows that the epithelial attractor is the most stable state whereas the hybrid attractor is the least stable one. This is consistent with the experimental characterization of the hybrid cell state as a metastable and stem-cell like intermediate state⁴².

Rather than measuring the relative dynamic stability using a quantity for each attractor separately, such as basin size and steady state probability, the other two methods, MFPT and BTR, consider pairwise relationships to define the stability. As shown in Fig. 3c, the MFPT values for the transitions between epithelial and hybrid attractors are, for the two directions, 182.53 (Epithelial to Hybrid) and 107.53 (Hybrid to Epithelial). Thus, the epithelial attractor is more stable (since cells stays on average longer in it) than the hybrid one. Similarly, since the MFPT from the mesenchymal attractor to the hybrid attractor is larger than the one of the backward transition, 76.93 and 50.06, respectively – the mesenchymal attractor is more stable than the hybrid attractor. Lastly, the epithelial attractor is more stable than the mesenchymal attractor because the transition from the epithelial attractor to the mesenchymal attractor takes longer than the transition in the opposite direction (118.14 and 107.79, respectively). Importantly, this triangle of net preferred direction is transitive, allowing us to rank the three attractors: The epithelial cell is the most stable while the hybrid cell is the least stable based on MFPT, which agrees with the previous measurements.

Finally, we examined the basin transition rates. From the basin transition matrix for the BTR values, we can compare two properties of basin transition: self-transition rate and relative transition rate (Fig. 3d). The epithelial, mesenchymal, and hybrid basin have the self-transition rates of 0.990, 0.978, and 0.963 respectively. Between the epithelial basin and mesenchymal basin, the transition from the epithelial basin to the mesenchymal basin is the preferred one. The transition from the epithelial basin to the hybrid basin is also preferred to the backward transition. Lastly, transition from the mesenchymal basin to the hybrid basin is less probable than the transition from the hybrid basin to the mesenchymal basin. To compare the relative stability associated with each basin, we need to consider the relative transition rates. This allowed us to compare the relative stability: The epithelial basin is the most stable while hybrid basin is the least stable among three basins (Fig. 3d).

In conclusion, all four methods for calculating the relative dynamic stability were in agreement with each other in ranking the three cell attractors according to their relative stability. These results are supported by experimental observations: the hybrid cell state has stem cell-like property and tends to spontaneously differentiates to epithelial and mesenchymal cell states, suggesting that it is the least stable^{42,52}. Mesenchymal cells tend to transform into the epithelial state spontaneously - which supports the simulation result that the epithelial cell is the most stable cell type⁵³.

A new stability index based on 1-degree neighborhood graph. Next, we compared the four measurements of relative dynamic stability for the cell attractors in the EMT Boolean network model with respect to genetic mutations. We knocked-out or overexpressed each gene and measured again the relative dynamic stability. Interestingly, differences appeared between the four methods (Supplementary Figs S1 and S2 and Supplementary Table S2). For instance, in the $\mu34$ knock-out (KO) simulation, the hybrid and the mesenchymal cell states were equivalently stable if basin size and BTR were used as a measure of relative stability, whereas the mesenchymal cell state was more stable than the hybrid cell state if SSP and MFPT were used (Supplementary Fig. S1c,h,l and p).

To understand the cause of this discrepancy, we considered *one-degree neighbor states* (1-degree neighbor) as shown in Fig. 4a–c. A 1-degree neighbor of a given attractor state is the state whose n-bit state vector has one-bit value difference (Hamming distance of 1) from that of the attractor state. The 1-degree neighbor may reside in the same basin or in another basin of attraction. Because transitions from an attractor state to its 1-degree neighbor are the most probable transitions (except the transition to itself), the distribution of these 1-degree neighbors of each attractor in the various basins can predict the major trajectories of attractor transitions driven by random noise. When we use basin size as a measure of relative dynamic stability, we usually think that the perturbed states from an attractor are uniformly distributed and a larger basin would contain more the perturbed states. However, we found that the distribution of 1-degree neighbors of each attractor was independent of its basin

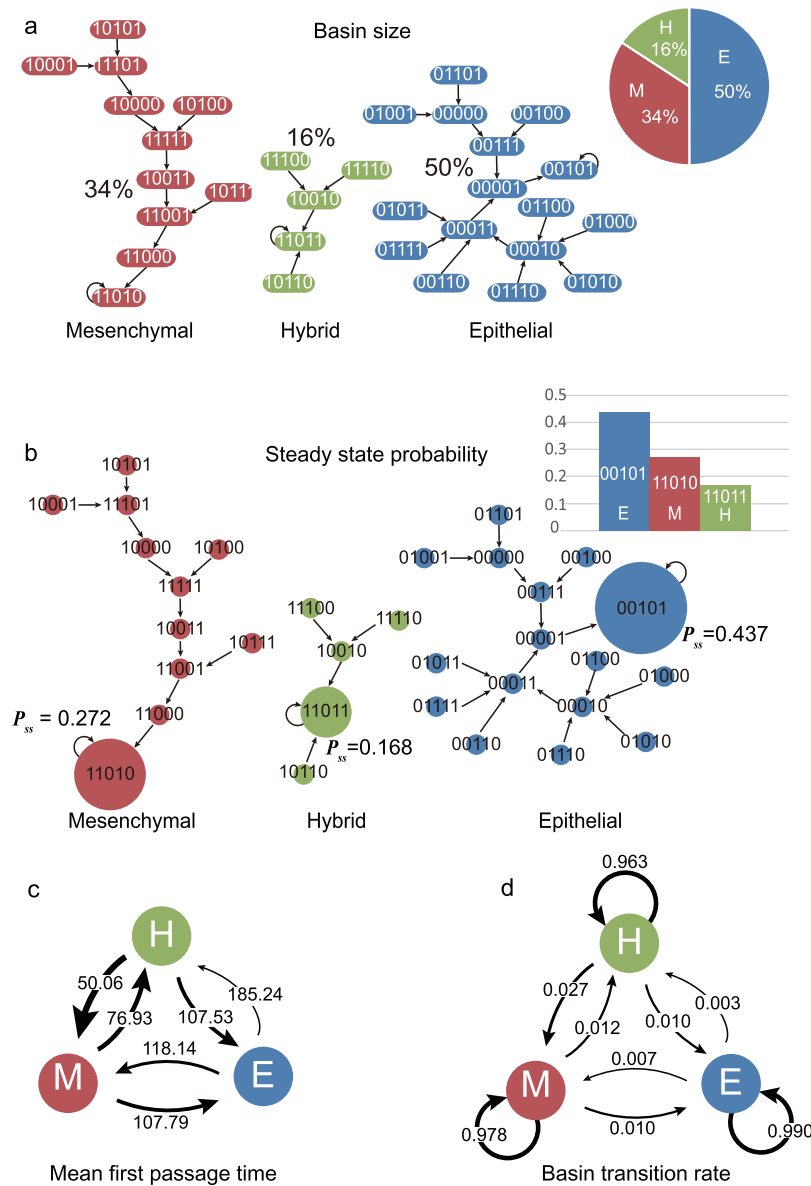


Figure 3. The relative dynamic stability of the attractors in the EMT Boolean model computed by four different methods. **(a)** The basin size of three attracting basins. The basin size is a count of states (ellipses) having the same color with a cell attracting basin. Basin size of each attractor is represented by a ratio to the number of all possible states. **(b)** Steady state probability distribution functions P_{ss} of three cell attractor states. The size of circle represents the steady state probability of each state. Non-attractor states have very small probabilities. **(c)** Mean first passage time between three cell attractor states. Width of an arrow is thicker as MFPT is shorter because the shorter is MFPT, the easier is the transition. **(d)** Basin transition rates between three cell attractor states. Contrary to MFPT, width of each arrow is thicker as BTR is higher.

size. Moreover, the relative dynamic stability analysis using BTR on the $\mu34$ KO model did not agree with our transition trajectory analysis based on the 1-degree neighbors. For instance, BTRs between mesenchymal and hybrid states were equivalent, however, the distribution of 1-degree neighbors of these two states were not same; four 1-degree neighbors of the hybrid state were in the mesenchymal basin (Fig. 4c) whereas only three 1-degree neighbors of mesenchymal state were in the hybrid basin (Fig. 4b). Interestingly, MFPTs between mesenchymal and hybrid states agreed with the biased distribution of 1-degree neighbors of two cell states. The relationship between SSP and 1-degree neighbor distribution is more complex than other measurements because SSP is the result of the equilibrium state at longer time scale whereas the 1-degree neighbor distribution only predicts the results of the dynamics in the short time scale. However, we can infer that SSP of an attractor is high when many 1-degree neighbors of all the attractors locate in the basin of that respective attractor.

From this, we propose a new stability index S_A for relative dynamic stability of an attractor A in the Boolean network model as follows:

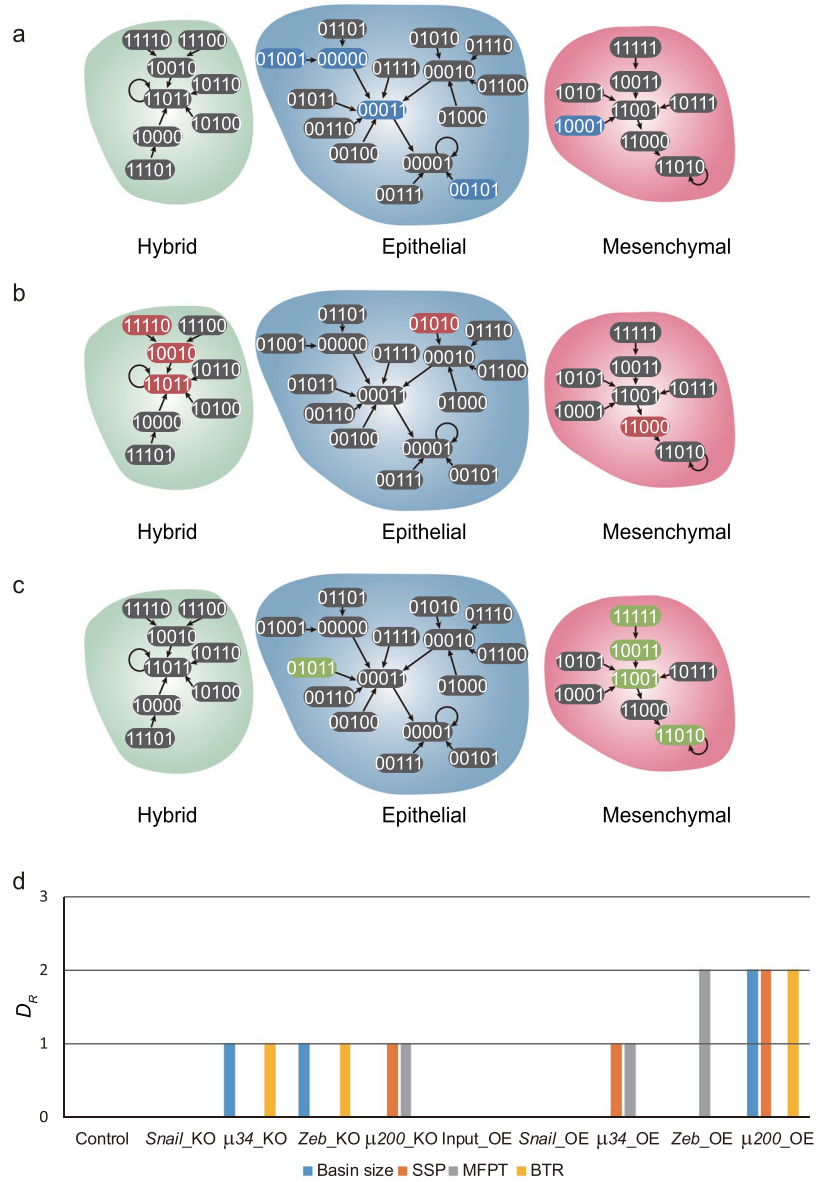


Figure 4. The 1-degree neighbor distribution and the stability index for EMT Boolean network models with various mutations. **(a–c)** The 1-degree neighbor distribution graphs for three cell attractors of the $\mu34$ KO model. 1-degree neighbors of each attractor A state are highlighted with the same color as the respective attractors A . **(a)** 1-degree neighbors of the epithelial cell attractor state (blue), **(b)** 1-degree neighbors of the mesenchymal cell attractor state (red), **(c)** 1-degree neighbors of the hybrid cell (green). **(d)** Differences between the stability index S_A and other measures of relative dynamic stability. The value 0 of D_R indicates that the stability index produces the same result with other measure R .

$$S_A = \sum_i O_i^A - \sum_{j=A} O_A^j \quad (1)$$

where O_i^j is defined as a ratio of 1-degree neighbors of attractor i in the basin of attractor j to total 1-degree neighbors of attractor i (Supplementary Table S3). The larger the stability index S_A , the more stable is an attractor state in a Boolean network model. To verify the relation between the stability index and relative dynamic stability, we compared the new stability index with other stability measurements (Fig. 4d and see Material and Method for details). In the cases of $\mu34$ knock-out and Zeb knock-out networks, basin size and BTR produced different rankings of attractor stability than by SSP and MFPT (Supplementary Figs S1 and S2). By contrast, the stability index S_A reproduced the rankings obtained by using SSP and MFPT (Fig. 4d). However, in the cases of $\mu200$ knock-out and $\mu34$ over-expression networks, the stability index S_A failed to produce the same stability ranking as SSP and MFPT. Since the differences of attractors in the $\mu200$ and $\mu34$ mutated networks were very small (Supplementary Figs S1j,n and S2h,m), we propose that the biased distribution graph of 2-degree neighbors or higher-degree

neighbors will explain these slightly different results. In the *Zeb* over-expression network, using MFPT resulted in ranking of attractor stability that was different from other measurements (Fig. 4d). In the $\mu 200$ over-expression network, our stability index S_s agreed with MFPT but differs from the methods based on basin size, SSP and BTR. Overall, our simple stability index agreed with the most of other methods to calculate the relative dynamic stability of Boolean models of various mutated networks.

From the 1-degree neighbor graph, we can identify the major trajectories of transitions between cell states. A major trajectory consists of two stages: (i) an escaping transition from a starting attractor to one of its 1-neighbor in the basin of end attractor due to noisy fluctuation of a gene expression, and (ii) following sequential transitions toward the end attractor actuated by Boolean functions. The epithelial cell state had one major trajectory toward the mesenchymal state but no major trajectory toward the hybrid state (Fig. 5a). Both mesenchymal and hybrid states had one major trajectory toward epithelial state (Fig. 5b and d); the mesenchymal state had three major trajectories to the hybrid state and the hybrid state could transit to mesenchymal state along four major trajectories (Fig. 5b and c). Especially, one of the trajectories from the hybrid to mesenchymal state started by the activation of $\mu 34$ (third position in the state vector). If we prevented the activation of $\mu 34$ in hybrid cell state, it would block one major trajectory, but the other trajectories would remain, so the differentiation of hybrid state would not be blocked (Fig. 5d). This observation may confirm the widely held notion that manipulation of gene expression of a single gene (node) to achieve a cell attractor state transition is not sufficient³.

Discussion

In this work, we constructed a Boolean network model for the EMT process and analyzed the relative dynamic stability of attractor states using various quantities. Four relative dynamic stability measures found in the literature were applied to examine our EMT Boolean network model. They agreed with each other in the wild-type networks but displayed different results in the cases with the genetic mutations that permanently altered the activity of individual genes. From this observation, we propose a new index for the relative dynamic stability of Boolean network by analyzing the graph of 1-degree neighbors of the attractors.

The four stability measures are based on different assumptions which naturally can lead to different results. First, basin size assumes fully deterministic dynamics whereas the other measures assume stochastic dynamics. Second, basin size and BTR assume that a basin represents a biological cell state but SSP and MFPT assumes that the cell state is only the attractor state no matter whether it is a limit cycle or a fixed point attractor. In the case of limit cycles, we used the sum of SSP of each state in a cyclic attractor to compare with SSP of other attractor, but it is still required to be further studied whether SSP of each state in cyclic attractor or the sum of them is meaningful. Finally, SSP assumes that the probability at the attractor state is the most important whereas MFPT assumes that relative difference of attractor transitions is the most significant. These assumptions contribute to the differences in the simulation results of mutations of a Boolean network. Hence, people should choose the appropriate method among four methods according to the purpose and model of the analysis.

When we consider a large-scale Boolean network model, there are additional problems with the currently available methods to calculate relative dynamic stability of attractors. Because computational complexity increases exponentially with the size of the network, it will be difficult to enumerate all the states to determine the basin size, and to calculate steady state probability function, mean first passage time or the transition rates between attractor states. However, our proposed stability index is simple to calculate which enable us to apply it to large-scale Boolean network models once the attractor states are identified. In large-scale networks, the noise level η is important to the stability index. A probability of transition to 1-degree neighbors is $N \times \eta(1 - \eta)^{N-1}$ but the transition probability to 2-degree neighbors is $\frac{N(N-1)}{2} \times \eta^2(1 - \eta)^{N-2}$ for a network of size N . Therefore, when $\eta N \ll 1 - \eta$, transition probability to 2-degree neighbors is much lower than the transition probability to 1-degree neighbors, and our dynamic stability index defined by only 1-degree neighbors can be reliable. In other words, when the dynamic stability indexes are different among attractors, 2-degree neighbors can be ignored. However, when the stability indexes of two attractors are identical, 2-degree neighbors can make a slight difference of the relative stability of them (Fig. 4d).

We added an auto-regulation link on $\mu 200$ even though there is no identified auto-regulation of $\mu 200$ (Fig. 2a). However, the added auto-regulation is matched to the largest parameter of synthesis rate of $\mu 200$ in ODE model⁴³. Interestingly, the added auto-regulation cannot cancel inhibitory effect of *Zeb* and *Snail* when one of them is activated and another is inactivated, but the auto-regulation of $\mu 200$ can buffer the inhibitory effect of *Zeb* and *Snail* when both are activated (Table 1). This is not general but seems to be required to construct tertiary stable toggle switch in Boolean model.

We used the synchronous updating Boolean model in this research. However, there are many researches using asynchronous updating Boolean models or multi-value logical model^{54–59}. These researchers have identified attractor states but not analyzed relative stability of attractors due to lack of stability measure in their models⁶⁰. Even though intrinsic and extrinsic noise is not defined in asynchronous model generally, 1-degree neighbors can be defined as in synchronous model and hence the suggested stability index may be relevant in asynchronous model. Moreover, if we define 1-degree neighbors considering multi-value of node, the stability index can be applied in multi-value logical model as well.

Genetic mutations and epigenetic modifications modify the Boolean functions and hence the landscape of gene regulatory network. However, as in Fig. 5, transient perturbations cannot modify the Boolean function and the landscape, but they can repress or promote some trajectories of cell state transition; it is possible that undesired cell state transition can occur through unaffected trajectories. Moreover, if the desired cell state is less stable than other cell states, then undesired cell state transition will occur by intrinsic and extrinsic noise.

In Boolean network model, the stable motifs associated to an attractor should be controlled to reach the attractor from any initial states⁵⁸. Among 10 cases with a single mutation, only one case (Input_KO) resulted in a single attractor of state space (Figs S1 and S2). From that, we can predict that most of induced cell state transitions

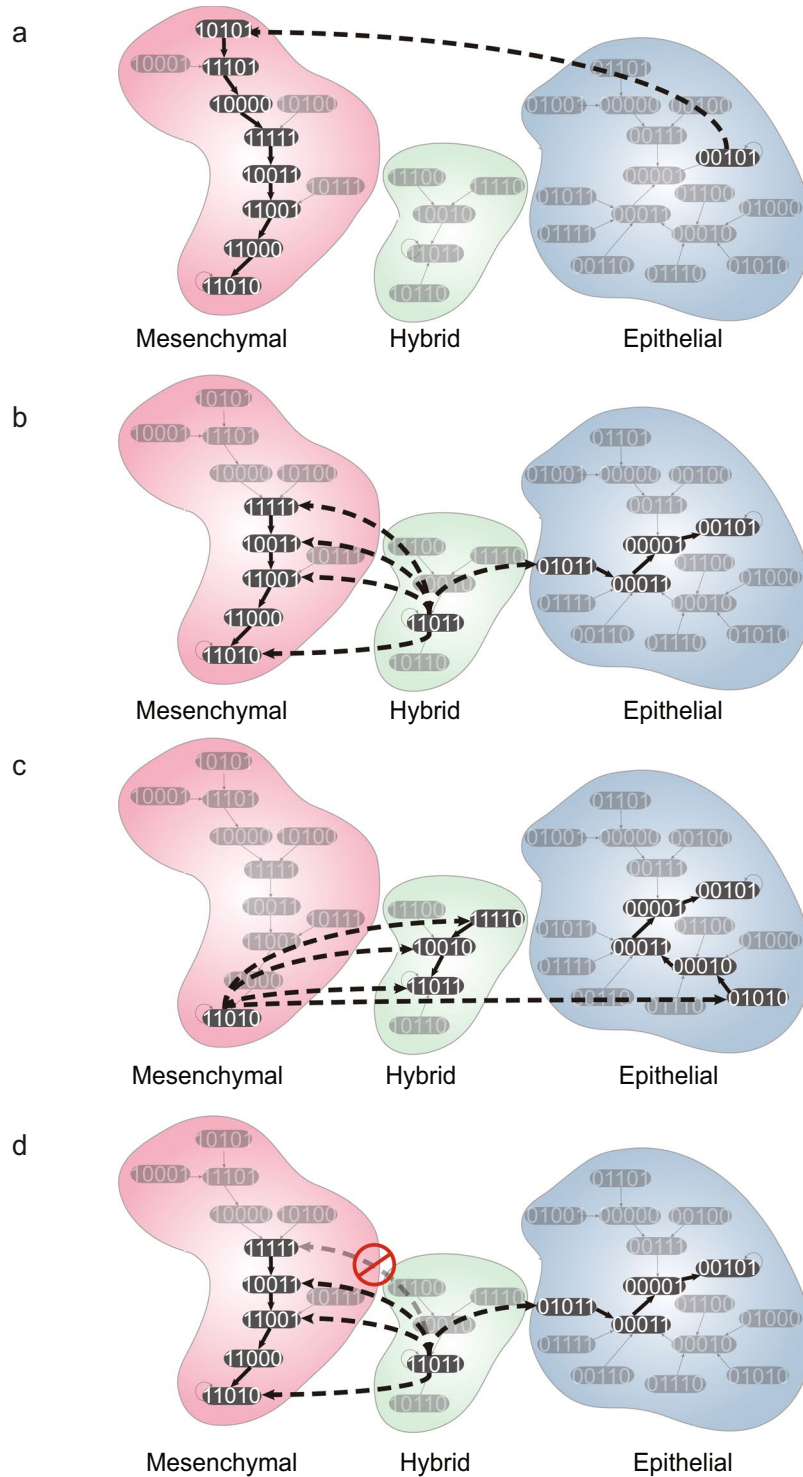


Figure 5. The Analysis of major transition trajectories between the three cell attractors of EMT Boolean model. The major transition trajectories from (a) epithelial cells, (b) hybrid cells and (c) mesenchymal cells to other cell state attractors. The dashed bold arrows represent the escaping transition driven by noise and the solid bold arrows represent the following sequential transitions due to execution of Boolean functions. (d) This example demonstrates that inhibition of one single node μ_{34} (third position in the state vector) is not sufficient to block the transition from hybrid to mesenchymal attractor.

may result in multiple attractors including undesired ones and hence the efficiencies of them is usually low. To improve the efficiency, we need to make the desired attractor as a unique attractor of the state space⁶¹ or as the most stable attractor relatively. Therefore, computing the relative dynamic stability in Boolean networks model can be relevant in the study of the cell state transition, such as cell differentiation, cell reprogramming, cancer metastasis and drug resistance.

Methods

Boolean network model and attracting basin size. The Boolean network model, among the historically oldest discrete model for dynamical systems, was firstly proposed by S. A. Kauffman¹⁹. Boolean networks represent the dynamics of an influence (regulatory) network in which the nodes influence the activity values of each other via the edges. Nodes can take the state being off or on, taking the Boolean values (0 or 1); edges represent the regulatory interactions (influences) between two nodes and represent the modality of the influence (e.g. activation or inhibition). The latter is formalized by the Boolean function f_i associated with each node i which integrates the values of its input nodes to update its value. Thus, to represent the states of n nodes of a Boolean network, a state vector $\mathbf{x}(t)$ at specific time t is defined as: $\mathbf{x}(t) = (x_1(t), \dots, x_i(t), \dots, x_n(t))$, n is the number of nodes of a network. Because the value of node is either 0 or 1, $\mathbf{x}(t)$ is represented as a n -bit binary vector. The value of node i is updated at each discrete time step by a Boolean function $f_i(\mathbf{x}_{j \in I_i})$; I_i is the set of upstream regulators that serve as input of node i , $\mathbf{x}_{j \in I_i}$ is a state vector of all regulating nodes. Using the Boolean functions, we can calculate a state vector $\mathbf{x}(t+1)$ which is the unique state of $\mathbf{x}(t)$:

$$\mathbf{x}(t+1) = \mathbf{F}(\mathbf{x}(t)) = (f_1(\mathbf{x}_{j \in I_1}), \dots, f_n(\mathbf{x}_{j \in I_n})) \quad (2)$$

\mathbf{F} is a state transition function of the entire Boolean network. Note that here values of each node are updated simultaneously, i.e. we employ the synchronous updating policy. Because each node can have two possible values, the state space contains 2^n possible vector states.

When an initial state $\mathbf{x}(0)$ of the network reaches a steady state $\mathbf{x}(t)$, i.e. $\mathbf{x}(t) = \mathbf{x}(t+L)$, if $L = 1$, we call the state $\mathbf{x}(t)$ as a point attractor; if $L > 1$, the states $\mathbf{x}(t), \mathbf{x}(t+1), \dots, \mathbf{x}(t+L-1)$ represents a limit cycle. A Boolean network can have multiple point attractors and limit cycles - every point in the state space converges to one of them. The basin of an attractor is the set of states $\mathbf{x}(0)$ that will converge to the same attractor. Therefore, the basin size of an attractor is the number of states in the basin of an attractor.

State transition matrix and steady state probability density function. Boolean networks can be used to model stochastic gene expression noise⁶², in which, according to some scheme the values of each node randomly (spontaneously) flip to the alternative state. In such models the network can undergo spontaneous transitions stochastically between two attractors due to the intrinsic noise in the system. Since the transition probability between any two states can be calculated from the driving force $\mathbf{F}(\mathbf{x}(t))$, we can use a discrete Markov model to trace the stochastic transitions of the whole system at any time. The transition matrix T in the Markov model can be obtained from the Boolean functions $\mathbf{F}(\mathbf{x}(t))$ in a deterministic manner: an entry T_{lm} is 1 only if a state l is the successor state of a state m according to $\mathbf{F}(\mathbf{x}(t))$, otherwise, $T_{lm} = 0$. Since gene expression dynamics is stochastic, with this implementation it is possible that state transitions sometimes do not follow $\mathbf{F}(\mathbf{x}(t))$. To represent this stochastic process, let η to be a probability in which noise changes one bit in the n -bit binary vector, then the stochastic transition matrix P_{lm} according to noise is as follows⁶³:

$$P_{lm} = \begin{cases} C_{H(l,m)}^n \eta^{H(l,m)} \cdot (1 - \eta)^{n-H(l,m)} & (l \neq m) \\ 0 & (l = m) \end{cases} \quad (3)$$

$H(l, m)$ is the Hamming distance between two states l and m . The stochastic transition matrix T_{lm}^* is defined as follows:

$$T_{lm}^* = (1 - \eta)^n \cdot T_{lm} + P_{lm} \quad (4)$$

Then T^* contains the state transition probability between network states which is driven by both the deterministic force $\mathbf{F}(\mathbf{x}(t))$ and the noise η . We define a 2^n -dimensional vector $\mathbf{P}(t)$ as the probability distribution function of all states in the network at specific time t , and we can get $\mathbf{P}(t+1)$ at the next time step by using T^* :

$$\mathbf{P}(t+1) = T^* \cdot \mathbf{P}(t) \quad (5)$$

After long enough time, the probability distribution function will reach a steady state \mathbf{P}_{ss} :

$$\mathbf{P}(t+1) = \mathbf{P}(t) = \mathbf{P}_{ss} \quad (6)$$

where \mathbf{P}_{ss} is called the steady state probability distribution function of a Boolean network.

Mean first passage time (MFPT). The Mean First Passage Time (MFPT) M_{ij} is the expected time steps for the occurrence of the first passage from a state j to another state i in a stochastic process⁶⁴. Let f_{ij}^n be the probability that first passage time from j to i is n time steps. Then, we can define the transition probability p_{ij}^n from an initial state j to a final state i in n time steps as follows:

$$p_{ij}^n = f_{ij}^n + f_{ij}^{n-1} \times p_{jj}^1 + f_{ij}^{n-2} \times p_{jj}^2 + \dots + f_{ij}^1 \times p_{jj}^{n-1} \quad (7)$$

Finally, we have the recursive definition of f_{ij}^n :

$$f_{ij}^n = p_{ij}^n - \sum_{k=1}^{n-1} f_{ij}^{n-k} \times p_{jj}^k \quad (8)$$

and MFPT M_{ij} is a sum of time step n times its probability f_{ij}^n . By using the fundamental matrix $Z = (I - T^* + W)^{-1}$, where I is an identity matrix and each column of W is P_{ss} , we can get the MFPT matrix:

$$M_{ij} = \frac{Z_{ii} - Z_{ij}}{P_{ss}^i} \quad (9)$$

where P_{ss}^i is the i th element of P_{ss} ⁶⁵.

Basin transition rate matrix. Another method for characterizing the dynamics of a Boolean network is to define a cell type as a basin of attraction and calculate the transition rate between the basins by using Monte Carlo simulation³⁰. With this idea, we obtained the basin transition matrix B by using state transition matrix T^* . An entry B_{UV} is a probability of basin transition from basin V to basin U :

$$B_{UV} = \sum_{i \in U} \sum_{j \in V} T_{ij}^* / \text{size}(V). \quad (10)$$

Nested Canalizing Boolean Function. Since the actual Boolean functions of each node of a Boolean network is difficult to determine, we may need to guess certain structures of them. From the limited biological experiment data, we know that a knock-out or knock-down of a key regulating gene usually decreases the expression of the regulated gene significantly. Thus, we used a subset of Boolean function called nested canalizing Boolean functions (NCBF) which considers this kind of regulatory logic. NCBF satisfy the constrain that there is a permutation of inputs $(\sigma_1, \dots, \sigma_n)$ such that:

$$\begin{aligned} f(x_{\sigma_1} = a_{\sigma_1}, \dots, x_{\sigma_n}) &= b_{\sigma_1}; \\ f(x_{\sigma_1} \neq a_{\sigma_1}, x_{\sigma_2} = a_{\sigma_2}, \dots, x_{\sigma_n}) &= b_{\sigma_2}; \\ &\dots; \\ f(x_{\sigma_1} \neq a_{\sigma_1}, x_{\sigma_2} \neq a_{\sigma_2}, \dots, x_{\sigma_n} = a_{\sigma_n}) &= b_{\sigma_n} \end{aligned} \quad (11)$$

where b_{σ_i} is the analyzed Boolean value of input x_{σ_i} . For example, there is a Boolean function $f(x_1, x_2, x_3) = x_1 \text{ AND } x_2 \text{ AND NOT } x_3$, this is a NCBF because there is a permutation $[1, 3, 2]$ of inputs $(1, 2, 3)$ such that $f(x_1 = 0, x_3, x_2) = 0$; $f(x_1 \neq 0, x_3 \neq 1, x_2) = 0$; $f(x_1 \neq 0, x_3 \neq 1, x_2 = 1) = 1$ (for this function, there are more permutations of inputs). However, $f(x_1, x_2, x_3) = (x_1 \text{ AND } x_2) \text{ OR } (\text{NOT } x_1 \text{ AND } x_3)$ is not a NCBF because there is no such permutation of inputs $(1, 2, 3)$.

Comparisons between the stability index and other relative stability measures. To compare the stability index with other relative stability measures, we used rankings of attractor states based on the relative dynamic stability:

$$D_R = \sum_{A_i} (r_S(A_i) - r_R(A_i))^2 \quad (12)$$

where $r_S(A_i)$ is a ranking of an attractor A_i by using the proposed stability index and $r_R(A_i)$ is the ranking by another relative stability measure R . Hence, D_R shows a difference between the stability index and a relative stability measure R .

Data availability. All data generated or analysed during this study are included in this published article (and its Supplementary Information files).

References

1. Buganim, Y., Faddah, D. A. & Jaenisch, R. Mechanisms and models of somatic cell reprogramming. *Nat Rev Genet* **14**, 427–439, <https://doi.org/10.1038/nrg3473> (2013).
2. Lamouille, S., Xu, J. & Derynck, R. Molecular mechanisms of epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol* **15**, 178–196, <https://doi.org/10.1038/nrm3758> (2014).
3. Ladewig, J., Koch, P. & Brustle, O. Leveling Waddington: the emergence of direct programming and the loss of cell fate hierarchies. *Nat Rev Mol Cell Biol* **14**, 225–236 (2013).
4. Brzezinski, J. A. & Reh, T. A. Photoreceptor cell fate specification in vertebrates. *Development* **142**, 3263–3273, <https://doi.org/10.1242/dev.127043> (2015).
5. Osakada, F. *et al.* Toward the generation of rod and cone photoreceptors from mouse, monkey and human embryonic stem cells. *Nat Biotechnol* **26**, 215–224, <https://doi.org/10.1038/nbt1384> (2008).
6. Klimanskaya, I. *et al.* Derivation and comparative assessment of retinal pigment epithelium from human embryonic stem cells using transcriptomics. *Cloning Stem Cells* **6**, 217–245, <https://doi.org/10.1089/1536230042323420> (2004).
7. Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J. & Melton, D. A. *In vivo* reprogramming of adult pancreatic exocrine cells to beta-cells. *Nature* **455**, 627–U630, <https://doi.org/10.1038/nature07314> (2008).
8. Kehat, I. *et al.* Human embryonic stem cells can differentiate into myocytes with structural and functional properties of cardiomyocytes. *J Clin Invest* **108**, 407–414, <https://doi.org/10.1172/Jci12131> (2001).
9. Paull, D. *et al.* Automated, high-throughput derivation, characterization and differentiation of induced pluripotent stem cells. *Nat Methods* **12**, 885–892, <https://doi.org/10.1038/nmeth.3507> (2015).
10. Ebrahimi, B. Reprogramming barriers and enhancers: strategies to enhance the efficiency and kinetics of induced pluripotency. *Cell Regen (Lond)* **4**, 10, <https://doi.org/10.1186/s13619-015-0024-9> (2015).
11. Hanna, J. H., Saha, K. & Jaenisch, R. Pluripotency and Cellular Reprogramming: Facts, Hypotheses, Unresolved Issues. *Cell* **143**, 508–525, <https://doi.org/10.1016/j.cell.2010.10.008> (2010).

12. Bargaje, R. *et al.* Cell population structure prior to bifurcation predicts efficiency of directed differentiation in human induced pluripotent cells. *Proc Natl Acad Sci USA* **114**, 2271–2276, <https://doi.org/10.1073/pnas.1621412114> (2017).
13. Mojtahedi, M. *et al.* Cell Fate Decision as High-Dimensional Critical State Transition. *Plos Biol* **14**, e2000640, <https://doi.org/10.1371/journal.pbio.2000640> (2016).
14. Ogawa, M. Hemopoietic stem cells: stochastic differentiation and humoral control of proliferation. *Environ Health Perspect* **80**, 199–207 (1989).
15. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676, <https://doi.org/10.1016/j.cell.2006.07.024> (2006).
16. Chang, H. H., Hemberg, M., Barahona, M., Ingber, D. E. & Huang, S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* **453**, 544–547, <https://doi.org/10.1038/nature06965> (2008).
17. Huang, S. Reprogramming cell fates: reconciling rarity with robustness. *Bioessays* **31**, 546–560, <https://doi.org/10.1002/bies.200800189> (2009).
18. Huang, S., Eichler, G., Bar-Yam, Y. & Ingber, D. E. Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys Rev Lett* **94**, <https://doi.org/10.1103/PhysRevLett.94.128701> (2005).
19. Kauffman, S. A. Metabolic stability and epigenesis in randomly constructed genetic nets. *J Theor Biol* **22**, 437–467 (1969).
20. Waddington, C. H. *The strategy of the genes; a discussion of some aspects of theoretical biology* (Allen & Unwin, 1957).
21. Zhou, J. X., Aliyu, M. D. S., Aurell, E. & Huang, S. Quasi-potential landscape in complex multi-stable systems. *J R Soc Interface* **9**, 3539–3553, <https://doi.org/10.1098/rsif.2012.0434> (2012).
22. Ao, P. Potential in stochastic differential equations: novel construction. *J Phys a-Math Gen* **37**, L25–L30, <https://doi.org/10.1088/0305-4470/37/3/L01> (2004).
23. Yin, L. & Ao, P. Existence and construction of dynamical potential in nonequilibrium processes without detailed balance. *J Phys a-Math Gen* **39**, 8593–8601, <https://doi.org/10.1088/0305-4470/39/27/003> (2006).
24. Kwon, C., Ao, P. & Thouless, D. J. Structure of stochastic dynamics near fixed points. *P Natl Acad Sci USA* **102**, 13029–13033, <https://doi.org/10.1073/pnas.0506347102> (2005).
25. Ao, P., Kwon, C. & Qian, H. On the existence of potential landscape in the evolution of complex systems. *Complexity* **12**, 19–27, <https://doi.org/10.1002/cplx.20171> (2007).
26. Wang, J., Xu, L. & Wang, E. K. Potential landscape and flux framework of nonequilibrium networks: Robustness, dissipation, and coherence of biochemical oscillations. *P Natl Acad Sci USA* **105**, 12271–12276, <https://doi.org/10.1073/pnas.0800579105> (2008).
27. Wang, J., Xu, L., Wang, E. K. & Huang, S. The Potential Landscape of Genetic Circuits Imposes the Arrow of Time in Stem Cell Differentiation. *Biophys J* **99**, 29–39, <https://doi.org/10.1016/j.bpj.2010.03.058> (2010).
28. Wang, J., Zhang, K., Xu, L. & Wang, E. Quantifying the Waddington landscape and biological paths for development and differentiation. *P Natl Acad Sci USA* **108**, 8257–8262, <https://doi.org/10.1073/pnas.1017017108> (2011).
29. Freidlin, M. I., Wentzell, A. D. & Wentzell, A. D. *Random perturbations of dynamical systems* (Springer-Verlag, 1984).
30. Alvarez-Buylla, E. R. *et al.* Floral Morphogenesis: Stochastic Explorations of a Gene Network Epigenetic Landscape. *Plos One* **3**, <https://doi.org/10.1371/journal.pone.0003626> (2008).
31. Zhou, J. X., Samal, A., d'Herouel, A. F., Price, N. D. & Huang, S. Relative stability of network states in Boolean network models of gene regulation in development. *Biosystems* **142**, 15–24, <https://doi.org/10.1016/j.biosystems.2016.03.002> (2016).
32. Li, F., Long, T., Lu, Y., Ouyang, Q. & Tang, C. The yeast cell-cycle network is robustly designed. *Proc Natl Acad Sci USA* **101**, 4781–4786, <https://doi.org/10.1073/pnas.0305937101> (2004).
33. Choi, M., Shi, J., Jung, S. H., Chen, X. & Cho, K. H. Attractor landscape analysis reveals feedback loops in the p53 network that control the cellular response to DNA damage. *Sci Signal* **5**, ra83, <https://doi.org/10.1126/scisignal.2003363> (2012).
34. Wang, R. S., Saadatpour, A. & Albert, R. Boolean modeling in systems biology: an overview of methodology and applications. *Phys Biol* **9**, 055001, <https://doi.org/10.1088/1478-3975/9/5/055001> (2012).
35. Atias, N., Gershenson, M., Labazin, K. & Sharan, R. Experimental design schemes for learning Boolean network models. *Bioinformatics* **30**, i445–452, <https://doi.org/10.1093/bioinformatics/btu451> (2014).
36. Sharan, R. & Karp, R. M. Reconstructing Boolean models of signaling. *J Comput Biol* **20**, 249–257, <https://doi.org/10.1089/cmb.2012.0241> (2013).
37. Han, B. & Wang, J. Quantifying robustness and dissipation cost of yeast cell cycle network: The funneled energy landscape perspectives. *Biophys J* **92**, 3755–3763, <https://doi.org/10.1529/biophysj.106.094821> (2007).
38. Albert, I., Thakar, J., Li, S., Zhang, R. & Albert, R. Boolean network simulations for life scientists. *Source Code Biol Med* **3**, 16, <https://doi.org/10.1186/1751-0473-3-16> (2008).
39. Huang, S. Gene expression profiling, genetic networks, and cellular states: an integrating concept for tumorigenesis and drug discovery. *J Mol Med (Berl)* **77**, 469–480 (1999).
40. Thiery, J. P., Acloque, H., Huang, R. Y. & Nieto, M. A. Epithelial-mesenchymal transitions in development and disease. *Cell* **139**, 871–890, <https://doi.org/10.1016/j.cell.2009.11.007> (2009).
41. Yu, M. *et al.* Circulating Breast Tumor Cells Exhibit Dynamic Changes in Epithelial and Mesenchymal Composition. *Science* **339**, 580–584, <https://doi.org/10.1126/science.1228522> (2013).
42. Grosse-Wilde, A. *et al.* Stemness of the hybrid Epithelial/Mesenchymal State in Breast Cancer and Its Association with Poor Survival. *Plos One* **10**, e0126522, <https://doi.org/10.1371/journal.pone.0126522> (2015).
43. Lu, M., Jolly, M. K., Levine, H., Onuchic, J. N. & Ben-Jacob, E. MicroRNA-based regulation of epithelial-hybrid-mesenchymal fate determination. *Proc Natl Acad Sci USA* **110**, 18144–18149, <https://doi.org/10.1073/pnas.1318192110> (2013).
44. Zhang, J. Y. *et al.* TGF-beta-induced epithelial-to-mesenchymal transition proceeds through stepwise activation of multiple feedback loops. *Science Signaling* **7**, <https://doi.org/10.1126/scisignal.2005304> (2014).
45. Kauffman, S., Peterson, C., Samuelsson, B. & Troein, C. Random Boolean network models and the yeast transcriptional network. *P Natl Acad Sci USA* **100**, 14796–14799, <https://doi.org/10.1073/pnas.2036429100> (2003).
46. Murrugarra, D. & Dimitrova, E. S. Molecular network control through boolean canalization. *EURASIP J Bioinform Syst Biol* **2015**, 9, <https://doi.org/10.1186/s13637-015-0029-2> (2015).
47. Kauffman, S., Peterson, C., Samuelsson, B. & Troein, C. Genetic networks with canalizing Boolean rules are always stable. *Proc Natl Acad Sci USA* **101**, 17102–17107, <https://doi.org/10.1073/pnas.0407783101> (2004).
48. Ahmad, A. *et al.* Phosphoglucose isomerase/autocrine motility factor mediates epithelial-mesenchymal transition regulated by miR-200 in breast cancer cells. *Cancer Res* **71**, 3400–3409, <https://doi.org/10.1158/0008-5472.CAN-10-0965> (2011).
49. Gibbons, D. L. *et al.* Contextual extracellular cues promote tumor cell EMT and metastasis by regulating miR-200 family expression. *Genes Dev* **23**, 2140–2151, <https://doi.org/10.1101/gad.1820209> (2009).
50. Gregory, P. A. *et al.* The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nat Cell Biol* **10**, 593–601, <https://doi.org/10.1038/ncb1722> (2008).
51. Kim, N. H. *et al.* A p53/miRNA-34 axis regulates Snail1-dependent cancer cell epithelial-mesenchymal transition. *J Cell Biol* **195**, 417–433, <https://doi.org/10.1083/jcb.201103097> (2011).
52. Mani, S. A. *et al.* The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **133**, 704–715, <https://doi.org/10.1016/j.cell.2008.03.027> (2008).
53. Scheel, C. *et al.* Paracrine and autocrine signals induce and maintain mesenchymal and stem cell states in the breast. *Cell* **145**, 926–940, <https://doi.org/10.1016/j.cell.2011.04.029> (2011).

54. Saadatpour, A., Albert, I. & Albert, R. Attractor analysis of asynchronous Boolean models of signal transduction networks. *J Theor Biol* **266**, 641–656, <https://doi.org/10.1016/j.jtbi.2010.07.022> (2010).
55. Kerkhofs, J., Roberts, S. J., Luyten, F. P., Van Oosterwyck, H. & Geris, L. Relating the chondrocyte gene network to growth plate morphology: from genes to phenotype. *Plos One* **7**, e34729, <https://doi.org/10.1371/journal.pone.0034729> (2012).
56. Steinway, S. N. *et al.* Network modeling of TGFbeta signaling in hepatocellular carcinoma epithelial-to-mesenchymal transition reveals joint sonic hedgehog and Wnt pathway activation. *Cancer Res* **74**, 5963–5977, <https://doi.org/10.1158/0008-5472.CAN-14-0225> (2014).
57. Steinway, S. N. *et al.* Combinatorial interventions inhibit TGFbeta-driven epithelial-to-mesenchymal transition and support hybrid cellular phenotypes. *NPJ Syst Biol Appl* **1**, 15014, <https://doi.org/10.1038/npsbsa.2015.14> (2015).
58. Zanudo, J. G. & Albert, R. Cell fate reprogramming by control of intracellular network dynamics. *PLoS Comput Biol* **11**, e1004193, <https://doi.org/10.1371/journal.pcbi.1004193> (2015).
59. Grieco, L. *et al.* Integrative modelling of the influence of MAPK network on cancer cell fate decision. *PLoS Comput Biol* **9**, e1003286, <https://doi.org/10.1371/journal.pcbi.1003286> (2013).
60. Mendes, N. D., Monteiro, P. T., Carneiro, J., Remy, E. & Chaouiya, C. Quantification of reachable attractors in asynchronous discrete dynamics. *Preprint at*, <https://arxiv.org/abs/1411.3539> (2014).
61. Kim, J., Park, S. M. & Cho, K. H. Discovery of a kernel for controlling biomolecular regulatory networks. *Sci Rep* **3**, 2223, <https://doi.org/10.1038/srep02223> (2013).
62. Shmulevich, I., Dougherty, E. R. & Zhang, W. Gene perturbation and intervention in probabilistic Boolean networks. *Bioinformatics* **18**, 1319–1331 (2002).
63. Shmulevich, I., Gluhovsky, I., Hashimoto, R. F., Dougherty, E. R. & Zhang, W. Steady-state analysis of genetic regulatory networks modelled by probabilistic Boolean networks. *Comp Funct Genom* **4**, 601–608, <https://doi.org/10.1002/cfg.342> (2003).
64. Kemeny, J. G. & Snell, J. L. *Finite markov chains* (Van Nostrand, 1960).
65. Grinstead, C. M., Snell, J. L. & Snell, J. L. *Introduction to probability* 2nd rev. edn, (American Mathematical Society, 1997).

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grants funded by the Korea Government, the Ministry of Science and ICT (2017R1A2A1A17069642, 2015M3A9A7067220, and 2013M3A9A7046303) and by the KAIST Future Systems Healthcare Project from the Ministry of Science, ICT & Future Planning, the KUSTAR-KAIST Institute, Korea under the R&D program supervised by the KAIST, and the grant of the Korean Health Technology R&D Project, Ministry of Health & Welfare, Republic of Korea (HI13C2162). This work was partially supported by National Institute of General Medical Sciences (NIGMS) Grant R01GM109964 and NIGMS National Centers for Systems Biology Grant 2P50GM076547-06A1. This research was also partially supported by the National Science Foundation Grant PHY11-25915. The content is solely the responsibility of the authors. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

K.-H.C. designed the project and supervised the study. J.I.J. and J.X.Z. conceived the idea and the theoretical analysis. J.I.J. curated data and performed data analysis. J.I.J. and J.X.Z. drafted the manuscript, K.-H.C. and S.H. edited and wrote the paper.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-30544-0>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018