

## ARTICLE OPEN

## A landscape view on the interplay between EMT and cancer metastasis

Chunhe Li<sup>1,2</sup> and Gabor Balazsi<sup>3,4</sup>

The epithelial–mesenchymal transition (EMT) is a basic developmental process that converts epithelial cells to mesenchymal cells. Although EMT might promote cancer metastasis, the molecular mechanisms for it remain to be fully clarified. To address this issue, we constructed an EMT–metastasis gene regulatory network model and quantified the potential landscape of cancer metastasis-promoting system computationally. We identified four steady-state attractors on the landscape, which separately characterize anti-metastatic (A), metastatic (M), and two other intermediate (I1 and I2) cell states. The tetrastable landscape and the existence of intermediate states are consistent with recent single-cell measurements. We identified one of the two intermediate states I1 as the EMT state. From a MAP approach, we found that for metastatic progression cells need to first undergo EMT (enter the I1 state), and then become metastatic (switch from the I1 state to the M state). Specifically, for metastatic progression, EMT genes (such as ZEB) should be activated before metastasis genes (such as BACH1). This suggests that temporal order is important for the activation of cellular programs in biological systems, and provides a possible mechanism of EMT-promoting cancer metastasis. To identify possible therapeutic targets from this landscape view, we performed sensitivity analysis for individual molecular factors, and identified optimal interventions for landscape control. We found that minimizing transition actions more effectively identifies optimal combinations of targets that induce transitions between attractors than single-factor sensitivity analysis. Overall, the landscape view not only suggests that intermediate states increase plasticity during cell fate decisions, providing a possible source for tumor heterogeneity that is critically important in metastatic progress, but also provides a way to identify therapeutic targets for preventing cancer progression.

*npj Systems Biology and Applications* (2018)4:34; doi:10.1038/s41540-018-0068-x

## INTRODUCTION

Cancer metastasis, as the most fatal stage of cancer, accounts for over 90% of cancer deaths.<sup>1</sup> The epithelial–mesenchymal transition (EMT) plays critical roles in embryonic development and might contribute to cancer metastasis.<sup>2–5</sup> Many classical EMT marker genes have been connected to cancer metastasis.<sup>6</sup> However, it remains elusive how to elucidate the mechanistic connections between EMT and cancer metastasis quantitatively.

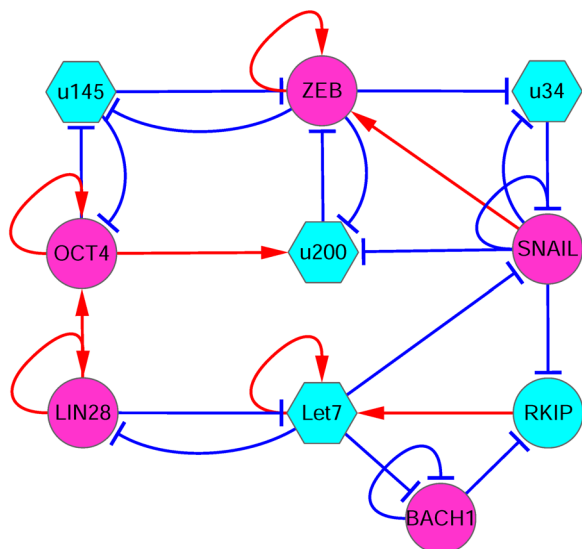
Mathematical modeling has been useful for studying EMT<sup>7,8</sup> and metastasis<sup>9</sup> starting from gene regulatory networks. Yet, in biological systems it is important to consider the effects of the stochasticity, and it remains challenging to study the global properties of the dynamical systems and the gene regulatory networks.<sup>10–14</sup> For example, the classic Waddington landscape<sup>15</sup> has been proposed as a metaphor to explain cellular development and differentiation. Recently, the epigenetic landscape for the biological networks has been constructed from different approaches,<sup>16–23</sup> and employed to investigate the stochastic dynamics of embryonic development and cancer.<sup>13,16,24–34</sup> In the landscape picture, different cell types are described as the basins of attraction on a potential surface, and cell fate determination is viewed as a ball rolling from one basin to another on the landscape surface by crossing certain potential barriers. The barrier heights separating the attractors or basins quantify the

degrees of difficulty for cells to switch from one cell type to the other. Of note, although many papers have been published using landscape approaches for modeling networks of genes and living cells, the theoretical foundation of this approach has only become available very recently, i.e., Waddington's epigenetic landscape metaphor now has a rigorous mathematical and chemical kinetic foundation.<sup>35–37</sup> These recent studies show that the landscapes of biochemical reaction networks emerge in the mathematical limit  $N \rightarrow \infty$ .<sup>35–37</sup>

In this work, we aim to discover the relationship between EMT and metastasis, and study the mechanism of EMT-promoting metastasis from a gene regulatory network perspective. To do so, we first construct an EMT–metastasis regulatory network by merging an EMT gene network and a metastasis network obtained by literature mining, and then develop a computational model corresponding to this joint system. Using this model, we find four steady states (attractors), two of which represent the anti-metastatic (A) and metastatic (M) states, while the other two represent intermediate states (I1 and I2). The intermediate state I1 is similar to the EMT state, with higher expression of the EMT marker genes (such as ZEB). This tetrastable landscape of metastasis is corroborated by recent single-cell experimental data. More interestingly, we find that the cancer metastasis process is a stepwise process, by first stepping into an intermediate state I1 (or

<sup>1</sup>Shanghai Center for Mathematical Sciences, Fudan University, Shanghai, China; <sup>2</sup>Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China; <sup>3</sup>The Louis and Beatrice Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, New York, USA and <sup>4</sup>Department of Biomedical Engineering, Stony Brook University, Stony Brook, New York, USA  
Correspondence: Chunhe Li (chunheli@fudan.edu.cn)

Received: 23 January 2018 Revised: 2 July 2018 Accepted: 4 July 2018  
Published online: 23 August 2018



**Fig. 1** The diagram for the core circuit of EMT-metastasis network including 10 gene nodes and 26 regulation links (8 activations and 18 repressions). Red arrows represent activation and blue bars represent repression. Magenta nodes represent pro-metastatic genes, and cyan nodes represent anti-metastatic genes. Circle nodes represent proteins and hexagonal nodes represent microRNAs. u34: miR34, u200: miR200, u145: miR145

EMT state), and then switching to the metastatic state. We propose that the existence of intermediate states increases the plasticity of cancer cells, offering a possible source for tumor heterogeneity. Moreover, we find that the temporal order for the activation of EMT marker genes and metastasis marker genes is critical for metastatic progression. First, the EMT marker gene ZEB is activated, which then promotes the activation of metastatic genes. This suggests a mechanism for EMT-promoting metastasis.

To identify effective anti-cancer strategies from our computational models, we performed single-factor sensitivity analysis to unravel how each link influences the landscape topography. We used the transition action as a measure to quantify the difficulty of transition between attractors, since a smaller transition action, corresponding to a smaller barrier height, means that it is easier for the system to make the corresponding transition. Specifically, we changed each activation or inhibition constant (characterizing individual regulation strength) individually to study their influences on the transition actions between the A attractor and the M attractor. We identified some optimal combinations of targets from minimizing transition actions from the metastatic (M) state to the anti-metastatic (A) state. We show quantitatively that the strategies from our optimization method are more effective than those from the single-parameter sensitivity analysis. This supports the view that advanced cancer is a network disease rather than a disease arising from single-gene defects. Therefore, an effective therapeutic strategy should be targeting multiple networks with different functions (e.g., related to different hallmarks of cancers) in appropriate order. Our results suggest that intermediate states play critical roles in cancer progression and offer a possible explanation for tumor heterogeneity. We also provide an approach to identify the optimal combinations of therapeutic targets for preventing cancer progression.

## RESULTS

### Mathematical model for the EMT-metastasis network

Based on previous computational works on EMT and cancer metastasis,<sup>9,38</sup> we first constructed an EMT-metastasis gene regulatory network by literature mining (Fig. 1), which includes

10 nodes (genes) and 26 links (regulations). Based on the network structure, we generated the ordinary differential equations (ODEs) describing the time evolution of relative expression levels for each of the 10 genes. To incorporate biological knowledge to the model, we developed these equations based on the following:

1. The mathematical equations for the let7-RKIP-BACH1 circuit originate from the model in ref.<sup>9</sup> and include biological details such as the interaction between let7 and BACH1. Hill coefficients for the regulations among let7, RKIP, and BACH1 were likewise taken from ref.<sup>9</sup>.
2. For the EMT and stemness gene regulatory circuit, our models are based on other recent studies.<sup>38,39</sup> The numbers of binding sites for the interaction between microRNAs and proteins constrain the choice of Hill coefficients (see Table S1 in SI).
3. For finding the other parameter, a major biological constraint is to obtain multistability because in a cancer metastasis system there should be at least two cell states, i.e., the metastatic cell state and the anti-metastatic cell state.

We used Hill functions to describe the activation and inhibition regulations among different genes.<sup>16,21,24</sup> The ODEs that govern the time evolution of variables (relative expression level of different genes) are in Eq. (1):

$$F_i = \sum_{j=1}^N \frac{A_{ji} X_j^n}{S^n + X_j^n} + \sum_{j=1}^N \frac{B_{ji} S^n}{S^n + X_j^n} - kX_i \quad (1)$$

and the ODEs for let7 ( $X_7$ ) and BACH1 ( $X_{10}$ ) are<sup>9</sup>:

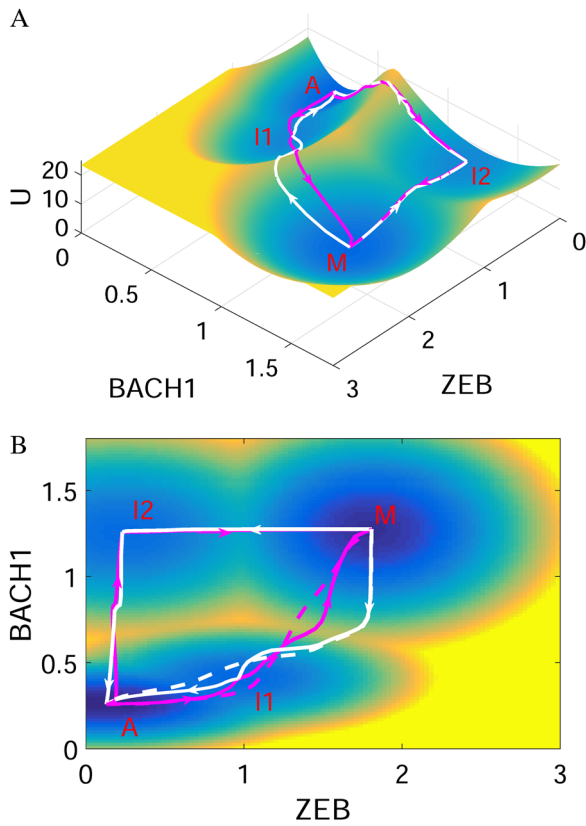
$$F_7 = \frac{sa2X_7^{ns}}{S^{ns} + X_7^{ns}} + \frac{G_L X_8^{nr}}{S^{nr} + X_8^{nr}} + \frac{bS^n}{S^n + X_9^n} - k_{LB} X_7 X_{10} - kX_7 \quad (2)$$

$$F_{10} = G_0 B + \frac{G_B S_B^{nb}}{S_B^{nb} + X_{10}^{nb}} - k_{LB} X_7 X_{10} - kX_{10}. \quad (3)$$

Here,  $F_i (i = 1, 2, \dots, 10)$  represents the driving force for the time evolution of expression levels of 10 genes. There are totally  $N = 10$  equations describing the time evolution of 10 variables.  $S$  represents the threshold of the sigmoidal function, and  $n$  is the Hill coefficient, which determines the steepness of the sigmoidal function.<sup>17,24</sup> Here, the default parameter values for the Hill function are specified as:  $S = 0.5$ ,  $n = 4$ .  $A$  and  $B$  are the interaction matrices for activation and inhibition interactions.  $A_{ji} = a$  when gene  $j$  activates gene  $i$ , otherwise  $A_{ji} = 0$ .  $B_{ji} = b$  when gene  $j$  inhibits gene  $i$ , otherwise  $B_{ji} = 0$  (see Table S2 for the interaction matrix  $M$ ). In addition,  $a$  is the activation constant,  $b$  is the inhibition constant, and  $k$  is the degradation rate for different genes (see the Supporting Information for the descriptions of parameters, and Table S1 for the values of parameters). In Eq. (1), the first term represents the activation effect of gene  $j$  on gene  $i$  (this term represents the self-activation when  $i = j$ ), while the second term represents the repression effect from gene  $j$  to gene  $i$  (this term represents the self-repression when  $i = j$ ). Finally, the last term represents the degradation of gene products (proteins, RNAs).

### Tetrastable landscape for cancer metastasis

The probabilistic evolution for a stochastic dynamical system is governed by the diffusion equations. For a high-dimensional gene regulatory system such as the one here, it is difficult to solve the diffusion equations directly. Following the self-consistent approximation approach (see Methods), we determined the steady-state probability distribution and then mapped out the potential landscape for the EMT-metastasis system. Because it is difficult to visualize the landscape in a ten-dimensional space, we selected two variables as the coordinates and projected the ten-



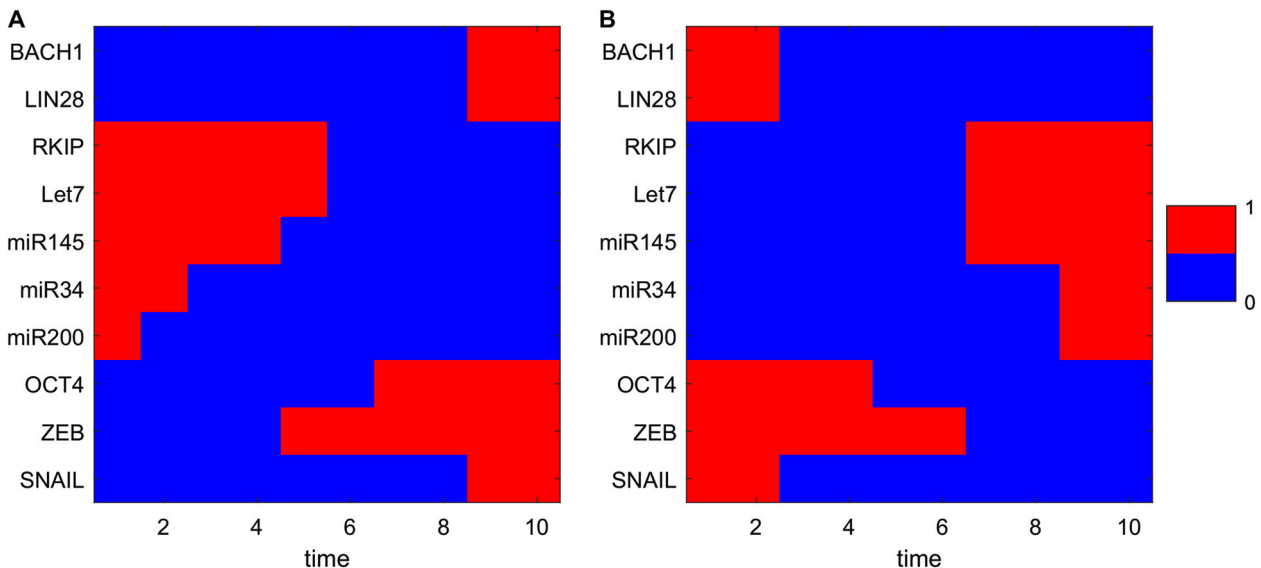
**Fig. 2** The landscape and corresponding minimum action paths (MAPs) for the EMT-metastasis network are shown in three-dimensional **(a)** and two-dimensional figures **(b)**. Magenta solid lines represent the MAP from A state (anti-metastatic cell state) to I1 states, and to M state (metastatic cell state), and the white solid lines represent the MAP from M to I1, and to A state. The dashed lines represent the direct MAP from A to M and from M to A states, respectively. A: anti-metastatic state, M: metastatic state, I1, I2: intermediate state. Here, ZEB and BACH1 are selected as the two coordinates

dimensional landscape into this two-dimensional space, by integrating out the other eight gene variables. We chose the two key variables “ZEB” and “BACH1” as the two coordinates for the landscape, since ZEB is a major EMT marker gene, and BACH1 is a major metastasis marker and regulator gene. We need to stress that our major conclusions do not depend on the specific choice of the coordinates (see Figs. S1–S4 for landscapes with other gene pairs as coordinates) because we also calculated the transition actions between different attractors, based on the ten-dimensional space of gene expressions. We found four stable cell states emerging on the landscape for the EMT-metastasis system (Fig. 2). The landscape surface is characterized by different colors, where the blue region represents lower potential or higher probability, and the red region represents higher potential or lower probability. The four basins of attraction on the landscape represent four different cell states characterized by different gene expression patterns in the ten-dimensional state space. These states separately correspond to M state (metastatic state, high ZEB/high BACH1 expression), A state (anti-metastatic state, low ZEB/low BACH1 expression), and two intermediate states (I1 and I2, intermediate ZEB and BACH1 expression).

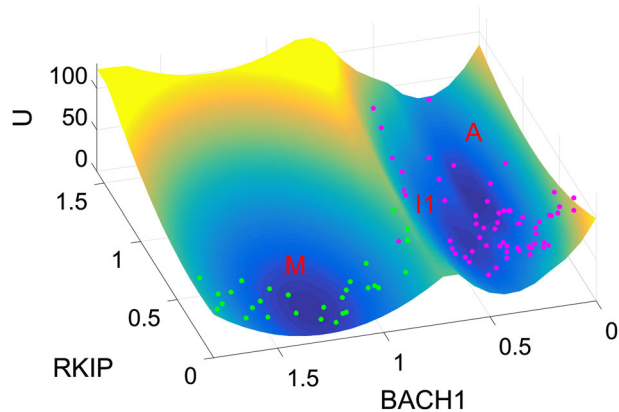
Of note, besides for the anti-metastatic state (A state) and metastatic state (M state), the two other intermediate states (I1 and I2) appear on the landscape. We propose that the I1 state corresponds to the EMT/partial EMT or some similar state,<sup>7,8</sup> since it has higher expression of EMT marker genes (such as ZEB and SNAIL), but lower expression of metastatic marker genes (such as BACH1).

#### Metastasis is a stepwise process

To study the transitions among individual cell types, we calculated kinetic transition paths by minimizing the transition actions between attractors,<sup>31,40</sup> obtaining minimum action paths (MAPs). The MAPs for different transitions are shown on the landscape in Fig. 2. The magenta MAP from the A state to the M state corresponds to the pro-metastatic process, while the white MAP from the M state to the A state corresponds to anti-metastatic process. The lines represent the MAPs, with the arrows denoting the directions of the transitions. The MAP for pro-metastatic process and the MAP for anti-metastatic process are irreversible,



**Fig. 3** Transition path from A state to M state **(a)** corresponding to the metastatic progression) and from M state to A state **(b)** corresponding to the anti-metastatic progress) in terms of expression levels of 10 different genes. The relative gene expressions are discretized to 0 or 1. 1 represents that the corresponding genes are in the on (activated) state and 0 represents that the corresponding genes are in the off (repressed) state. X-axis shows the 10 time points along the transition path



**Fig. 4** Landscape comparisons with single-cell experimental data (MDA-MB-231 cells). RKIP and BACH1 are chosen as the two coordinates. Each point represents a normalized gene expression value for one cell from single-cell experiments.<sup>9</sup> Green points, untreated cells; magenta points, cells after shBACH1 treatment (BACH1 knockdown). M: metastatic state, A: anti-metastatic state, I1: intermediate state

since the forward and reverse kinetic paths are not identical. This irreversibility of kinetic transition paths is caused by the non-gradient force, or curl flux.<sup>14,20,41</sup>

Since there are two intermediate states between A and M, in principle there should be two routes from the A to the M attractor, i.e., passing through the I1 state or passing through the I2 state. However, by calculating the MAP from the A to the M state, we found that the MAP from A to M and the MAP from M to A (dashed lines in Fig. 2b) are more similar to the case of passing through the I1 state, compared to that passing through I2 state (Figs. 2, 3). This demonstrates that for the transition process from A to M, i.e., metastatic progression, the cells experience a staged transition process. Specifically, cells first rise along the ZEB axis reaching the I1 state (EMT state), and then move up along the BACH1 axis, which possibly corresponds to a process in which cells acquire increased ability of invasion. This suggests that the EMT state might serve as a premetastatic state, promoting the progression toward metastasis. For the reverse process, i.e., the transition from metastatic M state to anti-metastatic A state, cells first decrease the expression of BACH1 reaching the I1 state and then decrease ZEB expression, finally reaching the A state. This indicates that a better anti-metastasis strategy would be first targeting BACH1 inducing the transition from M to I1, rather than first targeting ZEB governing the EMT process.

To investigate the metastatic process for multiple gene expressions, we visualized the ten-dimensional MAP from the A to the M state by discretizing the expressions of the 10 genes. From Fig. 3, we found that for the pro-metastatic process, three microRNAs (miR145, miR34, miR200) need to be downregulated first, and then the EMT marker gene ZEB must be activated, leading to the activation of the metastasis gene BACH1. These results indicate the importance of microRNAs in preventing cells from obtaining metastatic ability. Additionally, ZEB is activated before BACH1 (consistent with the 2D landscape picture in Fig. 2). This suggests that the order of switching on or switching off for different genes is critical and cells need to undergo EMT before transforming into metastasis cells.

Our landscape picture is also supported by the recent single-cell experiments, considering the expression of BACH1 and RKIP (Fig. 4),<sup>9</sup> mapped to the landscape with BACH1 and RKIP as the two coordinates. Green points (each point represents a single cell) correspond to untreated breast cancer cells, while magenta points correspond to breast cancer cells after shBACH1 treatment (BACH1 knockdown). We find that the untreated metastatic cells are

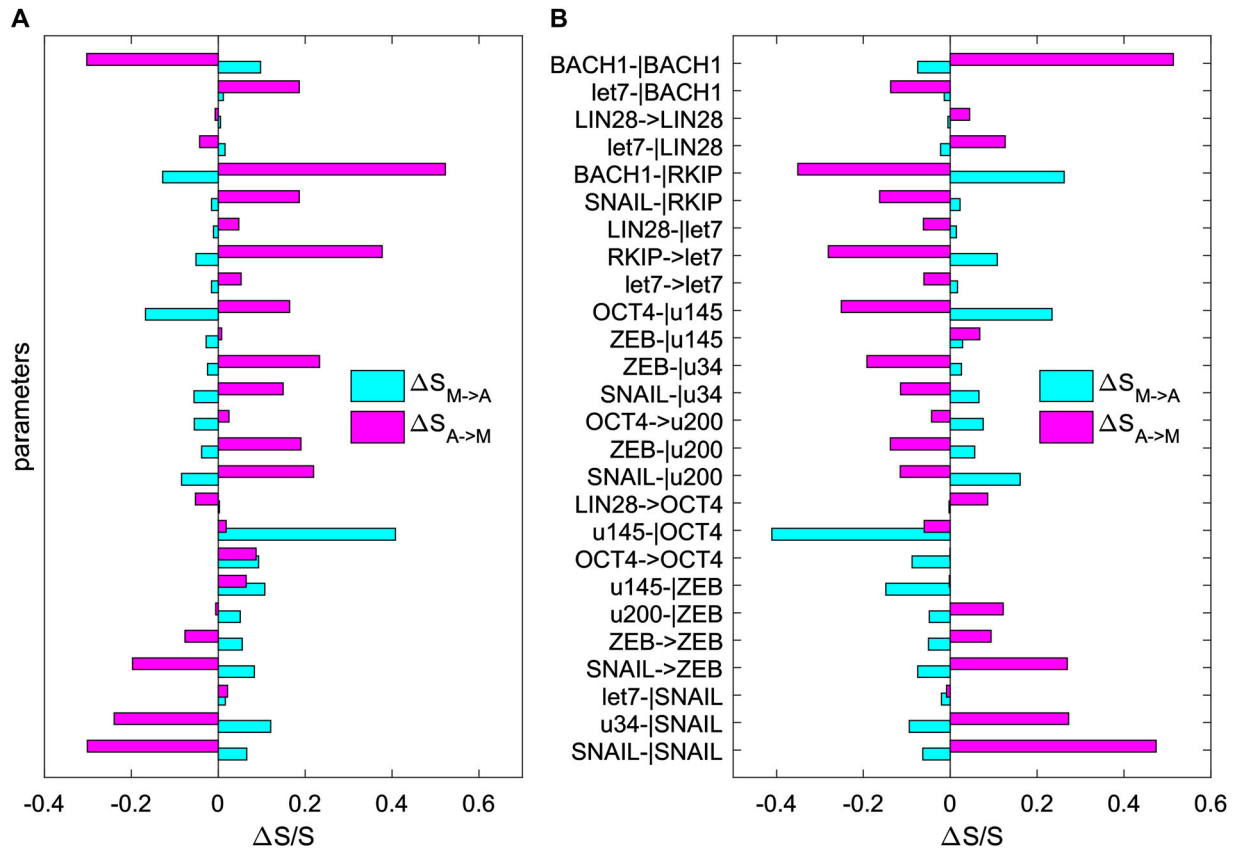
located around the metastasis (M) basin, whereas after BACH1 knockdown treatment, cells relocate around the A and I1 basins. This means that BACH1 knockdown lowers the energy barrier between the M basin and I1 basin, promoting the transition from the M state to the I1/A state. There are also some cells around the saddle point (green points), which reflects the heterogeneity of cancer cells and knockdown efficiency. Indeed, the single-cell data indicates that not all cells move to the A state. Yet, both A state cells and I1 state cells are separated from the M state, suggesting efficient treatment, because both A state cells and I1 state cells lose their metastatic ability (losing their expression of BACH1). Therefore, displaying these single-cell data on the landscape supports the existence of the intermediate state in the metastatic progression. It also indicates that a possible strategy for the metastasis prevention is to push back cells into the intermediate I1 state (or EMT state). This also means that an effective way of metastasis prevention would be to target an EMT cell before it enters metastatic state, since the I1 state or EMT state is much easier to convert into the anti-metastatic A state compared with the M state (Figs. 2, 4).

#### Sensitivity analysis for individual links

Traditionally, cancer has been studied as a disease arising from individual genetic defects (a disease from individual mutation). Recently, however, it has been proposed that cancer is a network diseases,<sup>42–44</sup> which can be understood as attractors in the state space of gene regulatory networks.<sup>42,45–47</sup> Therefore, it is important to ask how to alter normal (or anti-metastatic) and cancer (or metastatic) attractors to stabilize normal attractors and destabilize cancer attractors by targeting certain genes or regulations between genes. In the gene network models, this corresponds to the task of identifying the critical parameters that determine the metastatic to anti-metastatic state transitions. One natural way to accomplish this is to perform a single-factor sensitivity analysis for barrier heights between the M attractor and the A attractor, or the transition actions from M attractor to A attractor. This is preferable since a multiple-factor sensitivity analysis has very large computational cost.

Here we used two quantities to measure the difficulty of transitions between attractors. To quantify the stability of the system from the landscape topography, we defined the potential barrier height as the potential difference between a local minimum and the corresponding saddle point. A larger barrier means that it is harder for the system to surmount the barrier and switch from one attractor to another. Therefore, the barrier height provides a quantitative measure for the relative stability of each attractor and measures the difficulty of transitions among attractors in the system. Additionally, we also used transition actions to quantify the difficulty of transitions between attractors, since a smaller transition action, corresponding to a larger potential barrier, means an easier transition between attractors. In principle, the transition action should be correlated with the potential barrier height. However, we obtained the barrier height from the self-consistent and Gaussian approximation, so the transition action provides a more precise description for the barrier-crossing process.

Our sensitivity analysis is based on calculating the change of the transition actions between attractors after changing each parameters. Specifically, we change each activation or repression constant (representing individual regulation strength) individually to see how the transition actions between the M attractor and the A attractor are affected. In this way, we can identify those critical elements (here we focus on the links between genes) that govern the M to A transitions, as shown in Fig. 5. We find that some paramount links (to change the transition actions between M and A attractor significantly) include (see Table S3 for link ID): the self-repression of BACH1 (link 26), the repression of BACH1 on RKIP



**Fig. 5** Sensitivity analysis for the 26 key parameters (regulatory strengths among different genes, including 8 activation constants and 18 inhibition constants) on the transition action ( $S_{M \rightarrow A}$  and  $S_{A \rightarrow M}$ ). Y-axis represents the 26 parameters. X-axis represents the percentage of the change of the transition action (S) relative to S with default parameters. Here,  $S_{M \rightarrow A}$  represents the transition action from attractor M to attractor A (cyan bars), and  $S_{A \rightarrow M}$  represents the transition action from attractor A to attractor M (magenta bars). **a** Each parameter is increased by 20%, individually. **b** Each parameter is decreased by 20%, individually. u145: miR145, u34: miR34, u200: miR200

**Table 1.** Metastasis-suppressing interventions identified from optimization of transition actions for EMT-metastasis model

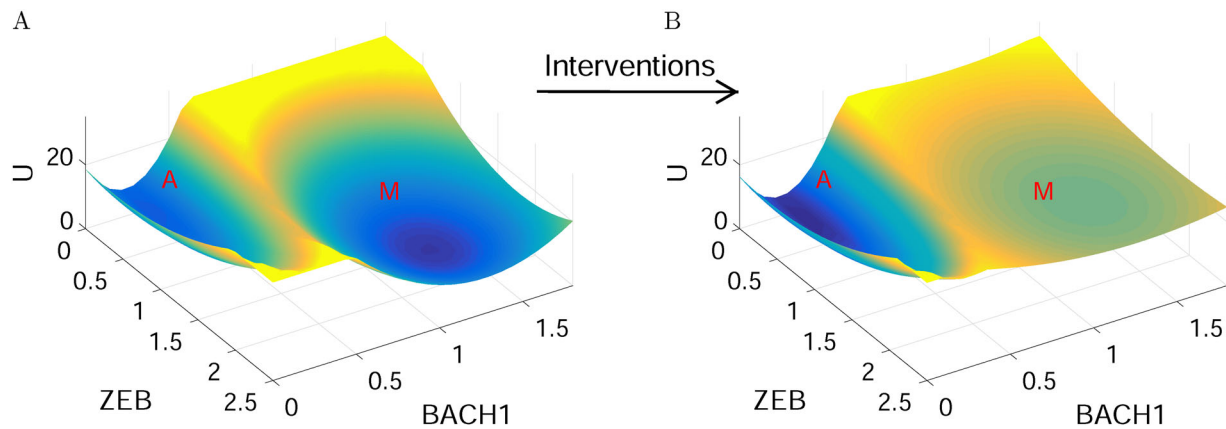
Target ID	Direction	Biological meaning of intervention	Sensitivity	Supporting references
22	↓	Decrease repression of BACH1 on RKIP	27.9%	Prediction
17	↓	Decrease repression of OCT4 on miR145	7.0%	50
26	↑	Increase self-repression of BACH1	7.0%	9,49
19	↓	Decrease activation of RKIP on Let7	6.7%	51
25	↑	Increase repression of Let7 on BACH1	6.1%	9,49
2	↑	Increase repression of miR34 on SNAIL	6.0%	4,48
1	↑	Increase self-repression of SNAIL	5.8%	48
15	↓	Decrease repression of ZEB on miR34	4.5%	Prediction
13	↑	Increase activation of OCT4 on miR200	4.3%	Prediction
20	↓	Decrease repression of LIN28 on Let7	3.2%	51

Each regulation is represented by a target ID (Table S3). Direction represents the direction of changes in regulation strengths for the corresponding interventions. Sensitivity is defined as how much each parameter (regulation strength) change accounts for the total change of parameters, averaged across the six parameter sets (representing the heterogeneity of tumor samples). Here, the targets are sorted by sensitivity and the top 10 targets are showed. The top 10 parameter changes account for about the 79% of the total parameter change. The column of “Supporting references” shows the consistent experimental evidences for the corresponding interventions

(link 22), the activation of RKIP on let7 (link 19), the repression of OCT4 on miR145 (link 17), the repression of miR145 on OCT4 (link 9), the activation of SNAIL on ZEB (link 4), the repression of miR34 on SNAIL (link 2), the self-repression of SNAIL (link 1). These key regulations are sorted according to their sensitivity (defined as the

relative change in transition actions caused by each parameter change). The top 10 sensitive regulations are shown in Table S4.

The predictions from this sensitivity analysis agree well with experimental data. For example, experiments show that Snail plays a critical role in tumor growth and metastasis of ovarian carcinoma through regulation of MMP activity,<sup>48</sup> ZEB and Snail are



**Fig. 6** An illustration for the landscapes of cancer metastasis network before and after interventions. Before interventions, the M state is deep and stable; after interventions, the M state becomes very shallow and cells are more inclined to stay in the A state. A: anti-metastatic state, M: metastatic state

upregulated in metastatic cells,<sup>4</sup> and the transcription factor BACH1 has been identified as a regulator of metastasis-associated genes, which are critical for bone metastasis formation.<sup>9,49</sup> Additionally, the interaction between the microRNA miR145 and the stem cell pluripotency gene OCT4 might play important roles in tumor growth and invasion,<sup>50</sup> and the microRNA let7 is proposed to inhibit cell motility by regulating the genes in the actin cytoskeleton pathway in breast cancer.<sup>51</sup> The consistency between our model predictions and previous experiments supports our models and landscape approaches.

Landscape control identifies optimal combinations of therapeutic targets

The single-factor sensitivity can be used to discover the effects of individual genes on the landscape or the transition actions. However, the combined influence from multiple parameters is difficult to identify in this way because it requires multiple-factor sensitivity analysis on parameters, with a large computational cost. Based on transition actions optimization,<sup>31,52,53</sup> we can predict optimal therapeutic combinations for targeting multiple genes or regulations among them for controlling the landscape. Therefore, we aimed to identify the optimal combinations of targets (regulations) for inducing M to A state transitions by optimizing (minimizing) the transition action from the M to the A attractor.

The top 10 targets from optimization are shown in Table 1. The comparison between Table 1 and Fig. 5 (or Table S4) indicates good consistency, in addition to some differences. For example, considering the fact that experimentally it might be only possible to tune a few genes (e.g., three) simultaneously, we found that the top three targets from the MAP optimization approach (Table 1, link 22, 17, 26) are not identical to the top three targets from the single-factor sensitivity analysis (Table S4, link 9, 17, 22). Yet, the both strategies suggest concordantly that an efficient treatment should target multiple circuits with different functions simultaneously. For example, link 22 (repression of BACH1 on RKIP) and link 17 (repression of OCT4 on miR145) separately characterize the metastasis circuit and the EMT circuit. Nonetheless, the optimization results suggest a better strategy (by adding a third interventions), namely, to target link 26 rather than link 9 for a larger sensitivity. This is probably because link 9 shares some of its function with link 17 since they are both characterizing the mutual repression between OCT4 and miR145. Therefore, a better strategy is to target another key metastasis gene BACH1 (link 26) rather than the repression of miR145 on OCT4 (link 9). This result is reasonable, and offers a quantitative illustration about why the MAP optimization method is useful to identify successful

interventions, which would be difficult to identify from single-factor sensitivity analysis.

Finally, we can infer how the landscape of the metastasis network looks before and after interventions suggested from above procedures (Fig. 6). Before any interventions, the M state is deep and stable; after interventions, the M state becomes very shallow and cells are more inclined to remain in the A state, suggesting a successful intervention for cancer metastasis treatment. Similar results could also provide an explanation for the relapse of cancers after treatment. A potential reason could be that many drugs for tumors are only pushing cells away from cancer attractors (changing the populations of tumor cells), while leaving the cancer metastasis network landscape unaltered. This again underlines that cancer cells are characterized by the attractors determined by cancer networks, and an effective way for cancer treatment should not be only killing tumor cells (i.e., not changing the landscape topography), but rather changing the landscape of cancer networks (making the cancer or metastatic attractors less stable as illustrated in Fig. 6).<sup>28</sup> This could be implemented by targeting multiple genes or regulations simultaneously as suggested in this work.

## DISCUSSION

Understanding the mechanisms underlying metastatic progression is critical to developing new therapeutic strategies for cancer treatment. Accumulating evidences show that stochastic fluctuations play important roles in cancer progression.<sup>54</sup> EMT is considered to be vital to the cancer metastatic process. However, the molecular mechanisms for how EMT promotes metastasis are not fully understood. In this work, based on the stochastic dynamics theory of gene networks we revealed the EMT-metastasis landscape through constructing an EMT-metastasis gene regulatory network. We identified four stable cell states characterized by the attractors on the landscape, in agreement with recent single-cell experimental data. Besides for the anti-metastatic attractor and metastatic attractor, we found two intermediate states (I1 and I2), one of which is very similar in gene expression pattern to an EMT cell state. This implies that metastatic progression is a stepwise process and there are two possible routes for the A to M transition, i.e., going through the I1 or going through the I2 state. However, our results based on MAPs between attractors indicate that the MAP for the A to M transition is closer to the path from A to I1, and then to M, i.e., the metastatic progression prefers going through the I1 (EMT) state rather than the I2 state.

These landscape results have a number of implications for cancer progression. First, cancer metastasis is a stepwise process. For the metastatic progression, the first step for cells is to cross the barrier between A and I1 and enter the I1 or EMT state (meaning that cells finish the EMT transition). The second step is crossing the barrier between I1 and M state to enter the metastatic state. The staged process for metastatic progression or the existence of intermediate states increases the plasticity of cell fate determinations. It also provides a possible source for tumor heterogeneity. Second, the intermediate state I1 provides a possible therapeutic anti-cancer target because it would be much easier to induce metastatic cells switching to the I1 state than switching to the A state. Third, our results indicate that the temporal order is critical for the metastatic progression. EMT circuit (e.g., the EMT marker gene ZEB) has to be activated before the metastatic marker gene BACH1 is activated (Fig. 2). The alternative case would be first activating BACH1 and then activating the EMT circuit (A to I2 and to M), but our transition action results suggest that this route has larger cost (action) and is not preferred by metastatic cells.

Our landscape results agree with recent single-cell data, which disperse within the A, I1, and M basins, as well as around the saddle points between basins, reflecting the heterogeneity of tumor cell populations. Recently, single-cell approaches have been developed to extract gene expression signatures for different cell populations.<sup>55,56</sup> In principle, one can integrate single-cell data considering more genes involved in EMT, metastasis, as well as other hallmarks of cancer. It would be interesting to combine gene network modeling and such single-cell data analysis, which is still very challenging. This work provides a starting point and a possible way toward that goal.

In the development of novel cancer therapeutic approaches, a major challenge in the field is, can we shift the goal from inducing tumor cell death (i.e., removing cells from particular “basins”) toward a strategy that seeks to alter the landscape topography to prevent relapse. Given the inevitability of cells eventually returning to steady-state proportions following a perturbation or drug treatment, this challenge is significant and important. In this work, we performed some initial attempts to tackle this challenge. We propose some optimal therapeutic targets for cancer treatment, focusing mostly on the regulatory interactions between different genes. We show that the optimal combinations of targets from minimizing transition actions are more effective than the strategy from single-factor sensitivity analysis. These results provide some new insights into potential therapeutic applications by targeting multiple genes. The successful application of these interventions will depend on whether physicians can actually tune these regulations in practice. Recent work indicates that one can tune the strengths of gene regulatory interactions through a synthetic biology approach.<sup>57</sup> With the further development of experimental techniques, we believe these predictions on optimal interventions of anti-cancer targets can eventually be tested from experimentally, potentially providing the guidance for designing the new anti-cancer strategies.

In this work, for simplicity we only consider the interplay between EMT and cancer metastasis. As suggested by Weinberg et al.,<sup>43,44</sup> cancer is a complex disease determined by many hallmarks of cancer. It would be really interesting to incorporate the other hallmarks of cancer, such as cancer immunity, cancer metabolism, etc., into our current metastatic network, and rediscover the landscape of cancer.<sup>27</sup> One limitation of the current work is that we only used single-cell data on two gene products (proteins) for comparisons with our computational model. In principle, one could collect single-cell data involving more EMT-cancer metastasis gene products at the RNA level by single-cell sequencing or at the protein level by mass cytometry. However, these technologies are still not routinely available or affordable to most laboratories, and hence it remains challenging to obtain and combine single-cell data with mathematical modeling. Another

limitation of this work is that the current models are single cell based. The cell–cell communications and interactions in populations of cells have been suggested to play critical roles in all kinds of systems such as neuron synchronization in circadian rhythms<sup>58</sup> and different forms of cancers.<sup>59,60</sup> Future work could incorporate cell–cell interactions on top of the current gene regulatory network models, where we could study the influence of other factors (e.g., the growth factor secretion and the proliferation rates for different cell types) on the landscape. This could better reflect the underlying biology and allow us to obtain more insights into intricate mechanisms for EMT and cancer metastasis.

## METHODS

### Self-consistent mean field approximation

The time evolution a dynamical system is determined by a probabilistic diffusion equation. Given the system state  $P(X_1, X_2, \dots, X_N, t)$ , where  $X_1, X_2, \dots, X_N$  represent the concentration of molecules or gene expression levels, we expect to have N-dimensional partial differential equation, which are hard to solve because of the very large state space of the system. Following a self-consistent mean field approach,<sup>20,24,41,61,62</sup> we can split the probability into the products of the individual probabilities:  $P(X_1, X_2, \dots, X_N, t) \sim \prod_i^n P_i(X_i, t)$  and solve the probability self-consistently. In this way, we effectively reduce the dimensionality of the system from  $M^N$  to  $MN$ , and therefore the computation of the problem becomes tractable. In some sense, our approximation is similar to the Poisson product result, which was shown to be exact for deficiency zero reaction networks.<sup>63</sup>

However, for the multidimensional system, it is still challenging to solve the diffusion equations directly. We start from the moment equations and assume specific probability distribution based on physical argument, i.e., we assume some specific connections between moments. In principle, once we know all moments, we can obtain the probability distributions. In this work, we assume Gaussian distribution as an approximation, which means we need to calculate two moments, the mean and the variance.

When the diffusion coefficient  $D$  is small, the moment equations can be approximated to<sup>64,65</sup>:

$$\dot{\bar{\mathbf{x}}}(t) = F[\bar{\mathbf{x}}(t)] \quad (4)$$

$$\dot{\boldsymbol{\sigma}}(t) = \boldsymbol{\sigma}(t)\mathbf{A}^T(t) + \mathbf{A}(t)\boldsymbol{\sigma}(t) + 2D[\bar{\mathbf{x}}(t)]. \quad (5)$$

Here,  $\mathbf{x}$ ,  $\boldsymbol{\sigma}(t)$ , and  $\mathbf{A}(t)$  are vectors and tensors, and  $\mathbf{A}^T(t)$  is the transpose of  $\mathbf{A}(t)$ . The elements of matrix  $A$  are specified as:  $A_{ij} = \frac{\partial F_i[\mathbf{x}(t)]}{\partial x_j(t)}$ . In this work we model intracellular biochemistry based on stochastic chemical kinetics, since an arbitrary noise structure would not be very realistic. Of note, one fundamental insight from the recent theory is that the landscape is sensitively dependent on the noise structure, e.g., the  $A$  matrix.<sup>35</sup> Also, the approach we used here is neither van Kampen  $\Omega$ -expansion because our drift term  $F$  is not linearized, nor the Kramers–Moyal expansion, which is questionable for computing barrier crossing.<sup>66</sup> It is worth pointing out that a “Large Deviation Theory” of the general stochastic chemical kinetic system has been derived recently by Ge and Qian.<sup>35</sup>

Based on these equations, we can solve  $\mathbf{x}(t)$  and  $\boldsymbol{\sigma}(t)$ . Here, we only consider the diagonal elements of  $\boldsymbol{\sigma}(t)$  from the mean field approximation. Therefore, the evolution of probability distribution for each variable can be acquired from the Gaussian approximation:

$$P(x, t) = \frac{1}{\sqrt{2\pi\sigma(t)}} \exp - \frac{[x - \bar{x}(t)]^2}{2\sigma(t)}. \quad (6)$$

Here,  $\bar{\mathbf{x}}(t)$  and  $\boldsymbol{\sigma}(t)$  are the solutions of Eqs. (2) and (3). The probability distribution obtained above corresponds to one steady state or basin of attraction. If the system has multiple steady states, then there should be several probability distributions localized at each basin with different variances. Therefore, the total probability is the weighted sum of all these probability distributions. The weighting factors ( $w_1, w_2$ ) are the sizes of the basins, representing the relative sizes of different basin of attraction. For example, for a bistable system, the probability distribution takes the form:  $P(x, t) = w_1 P^a(x) + w_2 P^b(x)$ , here  $w_1 + w_2 = 1$ . We determine the weights  $w_i$  by giving a large number of random initial conditions to the ODEs to be solved, and then collect the statistics from all of these different solutions.

For example, for a bistable system, if 10% of the random initial conditions lead to the first steady state, and 90% of the random initial conditions lead to the second steady state, then the weight  $w_1$  for the first basin is 0.1 and  $w_2$  for the second basin is 0.9.

From the mean field approximation, we can extend this formulation to the multidimensional case by assuming that the total probability is the produce of each individual probability for each variable. Finally, we can construct the potential landscape by:  $U(x) = -\ln P_{ss}(x)$ , with  $P_{ss}$  representing the steady-state probability distribution.

To determine how the stable states correspond to different cell states, we first obtain the steady-state expression for different genes from the deterministic ODEs, Eqs. (1–3), and then identify the biological cell states with these steady states, naming them according to their relative gene expression. Based on the cancer biology literature, the genes and miRNAs in Fig. 1 can be classified as either pro-metastatic (including ZEB, SNAIL, OCT4, LIN28, and BACH1), or anti-metastatic (including RKIP, Let7, miR200, miR34, and miR145). Taking the tetrastable landscape (Fig. 2) as an example, we have listed the steady-state gene expression values for 10 different genes for the four stable steady states, individually (Table S5). In this ten-dimensional gene expression state space, the A state (second column in Table S5) has lower relative expression values for all five pro-metastatic genes and high relative expression for all five anti-metastatic genes. Thus, we define this state as the A (anti-metastatic) state. On the contrary, the M state (last column in Table S5) has higher relative expression values for all five pro-metastatic genes and lower relative expression for all five anti-metastatic genes. Therefore, we define this state as the M (metastatic) state. In the middle, the I1 state has higher relative expression for EMT marker genes (such as ZEB and SNAIL) and lower relative expression for metastasis marker gene (such as BACH1), so we defined I1 as an EMT-like state. Similarly, we identified another intermediate state I2.

To validate the self-consistent approximation method, previously we have compared the results from the self-consistent approximation and Langevin dynamics for a simple two-dimensional model.<sup>24</sup> There we showed that for a simple “mutual repression with self-activation” model, the landscapes from the self-consistent approximation preserve similar global properties (number of attractors, relative stability of the basins of attraction) as the Langevin dynamics method.<sup>24</sup>

### MAPs and optimization of transition actions

Within the cell, there exists intrinsic noise arising from statistical fluctuations of the finite number of molecules, and external noise originating from highly dynamical and inhomogeneous environments, which can be critical to the dynamics of the system.<sup>67–69</sup> Because the number of cancer cells and immune cells are usually huge, we consider the external noise as the major noise source of the cancer-immunity interplay system. A dynamical system in the fluctuating environments can be described by:  $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}) + \zeta$ . Here,  $\mathbf{x} = (x_1(t), x_2(t), \dots, x_{10}(t))$  represents the vector of the relative gene expressions.  $\mathbf{F}(\mathbf{x})$  is the vector for the driving force of chemical reaction.  $\zeta$  is Gaussian noise term whose autocorrelation function is  $\langle \zeta_i(\mathbf{x}, t) \zeta_j(\mathbf{x}, 0) \rangle = 2D\delta(t)$ , and  $D$  is diffusion coefficient matrix.

Following the approaches<sup>26,52,53,70,71</sup> based on the Wentzell–Freidlin theory,<sup>72</sup> the most probable transition path from one attractor  $i$  at time 0 to attractor  $j$  at time  $T$ ,  $\phi_{ij}^*(t)$ ,  $t \in [0, T]$ , can be acquired by minimizing the action functional over all possible paths:

$$S_T[\phi_{ij}] = \frac{1}{2} \int_0^T |\dot{\phi}_{ij} - \mathbf{F}(\phi_{ij})|^2 dt. \quad (7)$$

This optimal path is called MAP. We calculated MAPs numerically by applying minimum action methods used in refs. <sup>52,53</sup>. In our case,  $T$  is set to be 20 and we verified that larger values of  $T$  would not improve accuracy of simulations significantly.

To predict the optimal combination of therapeutic targets, we applied the approach of optimizing transition actions to the cancer metastasis model.<sup>52</sup> Here, the purpose is to predict therapeutic targets (26 parameters characterizing the regulatory strengths among genes) that can induce the transition from M state to A state. Specifically, the optimization process is to minimize the difference between the transition action from M to A and the transition action from A to M, ( $\Delta S = S_{M \rightarrow A} - S_{A \rightarrow M}$ ), by tuning each of the 26 parameters. We used the Adaptive Minimum Action Method<sup>53</sup> to find the minimum of the transition actions, and the matlab fuction “fmincon” was used to perform the minimizations.

### Data availability

The data and matlab codes that support the findings of this study are available from the corresponding author upon reasonable request.

### ACKNOWLEDGEMENTS

C.L. is supported by the National Natural Science Foundation of China (11771098), Natural Science Foundation of Shanghai, China (17ZR1444500), and Thousand Youth Talents Program. G.B. was supported by NIGMS grants 1R35GM122561 (MIRA) and 1R01GM106027-01, and the Laufer Center for Physical and Quantitative Biology.

### AUTHOR CONTRIBUTIONS

C.L. conceived the research, performed the research, analyzed the results, and wrote the manuscript. G.B. analyzed the results and wrote the manuscript.

### ADDITIONAL INFORMATION

**Supplementary information** accompanies the paper on the *npj Systems Biology and Applications* website (<https://doi.org/10.1038/s41540-018-0068-x>).

**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.

### REFERENCES

- Brabletz, T., Lyden, D., Steeg, P. S. & Werb, Z. Roadblocks to translational advances on metastasis research. *Nat. Med.* **19**, 1104–1109 (2013).
- Nieto, M. A. The ins and outs of the epithelial to mesenchymal transition in health and disease. *Annu. Rev. Cell Dev. Biol.* **27**, 347–376 (2011).
- Thiery, J. P., Acloque, H., Huang, R. Y. & Nieto, M. A. Epithelial-mesenchymal transitions in development and disease. *Cell* **139**, 871–890 (2009).
- Nakaya, Y. & Sheng, G. Emt in developmental morphogenesis. *Cancer Lett.* **341**, 9–15 (2013).
- Jia, D., Jolly, M. K., Kulkarni, P. & Levine, H. Phenotypic plasticity and cell fate decisions in cancer: insights from dynamical systems theory. *Cancers* **9**, 70 (2017).
- Heerboth, S., et al. Emt and tumor metastasis. *Clin. Transl. Med.* **4**, 6 (2015).
- Lu, M., Jolly, H., Levine, H., Onuchic, J. & Ben-Jacob, E. MicroRNA-based regulation of epithelial-hybrid-mesenchymal fate determination. *Proc. Natl Acad. Sci. USA* **110**, 18144–18149 (2013).
- Zhang, J., et al. Tgf-b-induced epithelial-to-mesenchymal transition proceeds through stepwise activation of multiple feedback loops. *Sci. Signal* **7**, ra91 (2014).
- Lee, J., et al. Network of mutually repressive metastasis regulators can promote cell heterogeneity and metastatic transitions. *Proc. Natl Acad. Sci. USA* **111**, E364–E373 (2014).
- Ferrell, J. E. Bistability, bifurcations, and waddington's epigenetic landscape. *Curr. Biol.* **22**, R458–R466 (2012).
- Frauenfelder, H., Sliagar, S. G. & Wolynes, P. G. The energy landscapes and motions of proteins. *Science* **254**, 1598–1603 (1991).
- Qian, H. Cooperativity in cellular biochemical processes: noise-enhanced sensitivity, fluctuating enzyme, bistability with nonlinear feedback, and other mechanisms for sigmoidal responses. *Annu. Rev. Biophys.* **41**, 179–204 (2012).
- Wang, J. Landscape and flux theory of non-equilibrium dynamical systems with application to biology. *Adv. Phys.* **64**, 1–137 (2015).
- Wang, J., Xu, L. & Wang, E. K. Potential landscape and flux framework of non-equilibrium networks: robustness, dissipation and coherence of biochemical oscillations. *Proc. Natl Acad. Sci. USA* **105**, 12271–12276 (2008).
- Waddington, C. H. *The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology*. (Allen and Unwin, London, 1957).
- Wang, J., Zhang, K., Xu, L. & Wang, E. K. Quantifying the waddington landscape and biological paths for development and differentiation. *Proc. Natl Acad. Sci. USA* **108**, 8257–8262 (2011).
- 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 (2010).
- Liao, C. & Lu, T. A minimal transcriptional controlling network of regulatory T cell development. *J. Phys. Chem. B* **117**, 12995–13004 (2013).
- Lv, C., Li, X., Li, F. & Li, T. Energy landscape reveals that the budding yeast cell cycle is a robust and adaptive multi-stage process. *PLoS Comput. Biol.* **11**, e1004156 (2015).



20. Li, C. & Wang, J. Landscape and flux reveal a new global view and physical quantification of mammalian cell cycle. *Proc. Natl Acad. Sci. USA* **111**, 14130–14135 (2014).
21. Lu, M., Onuchic, J. & Ben-Jacob, E. Construction of an effective landscape for multistate genetic switches. *Phys. Rev. Lett.* **113**, 078102 (2014).
22. Ge, H. & Qian, H. Landscapes of non-gradient dynamics without detailed balance: stable limit cycles and multiple attractors. *Chaos* **22**, 023140 (2012).
23. Feng, H., Han, B. & Wang, J. Adiabatic and non-adiabatic non-equilibrium stochastic dynamics of single regulating genes. *J. Phys. Chem. B* **115**, 1254–1261 (2011).
24. Li, C. & Wang, J. Quantifying cell fate decisions for differentiation and reprogramming of a human stem cell network: Landscape and biological paths. *PLoS Comput. Biol.* **9**, e1003165 (2013).
25. Li, C. & Wang, J. Quantifying waddington landscapes and paths of non-adiabatic cell fate decisions for differentiation, reprogramming and transdifferentiation. *J. R. Soc. Interface* **10**, 20130787 (2013).
26. Chen, C., et al. Mathematical models of the transitions between endocrine therapy responsive and resistant states in breast cancer. *J. R. Soc. Interface* **96**, 20140206 (2014).
27. Li, C. & Wang, J. Quantifying the underlying landscape and paths of cancer. *J. R. Soc. Interface* **10**, 20140774 (2014).
28. Li, C. & Wang, J. Quantifying the landscape for development and cancer from a core cancer stem cell circuit. *Cancer Res.* **75**, 2607–2618 (2015).
29. Huang, S. Genetic and non-genetic instability in tumor progression: link between the fitness landscape and the epigenetic landscape of cancer cells. *Cancer Metastasis Rev.* **32**, 423–448 (2013).
30. Li, C., Hong, T. & Nie, Q. Quantifying the landscape and kinetic paths for epithelial–mesenchymal transition from a core circuit. *Phys. Chem. Chem. Phys.* **18**, 17949–17956 (2016).
31. Li, C. Identifying the optimal anticancer targets from the landscape of a cancer–immunity interaction network. *Phys. Chem. Chem. Phys.* **19**, 7642–7651 (2017).
32. Xu, L., Zhang, K. & Wang, J. Exploring the mechanisms of differentiation, dedifferentiation, reprogramming and transdifferentiation. *PLoS ONE* **9**, e105216 (2014).
33. Ao, P., Galas, D., Hood, L. & Zhu, X. Cancer as robust intrinsic state of endogenous molecular–cellular network shaped by evolution. *Med. Hypotheses* **70**, 678–684 (2008).
34. Li, S., Zhu, X., Liu, B., Wang, G. & Ao, P. Endogenous molecular network reveals two mechanisms of heterogeneity within gastric cancer. *Oncotarget* **6**, 13607 (2015).
35. Ge, H. & Qian, H. Mesoscopic kinetic basis of macroscopic chemical thermodynamics: a mathematical theory. *Phys. Rev. E* **94**, 052150 (2016).
36. Huang, S., Li, F., Zhou, J. X. & Qian, H. Processes on the emergent landscapes of biochemical reaction networks and heterogeneous cell population dynamics: differentiation in living matters. *J. R. Soc. Interface* **14**, 20170097 (2017).
37. Ge, H. & Qian, H. Mathematical formalism of nonequilibrium thermodynamics for nonlinear chemical reaction systems with general rate law. *J. Stat. Phys.* **166**, 190–209 (2017).
38. Lu, M., et al. Toward decoding the principles of cancer metastasis circuits. *Cancer Res.* **74**, 4574–4587 (2014).
39. Jolly, M. K., et al. Towards elucidating the connection between epithelial–mesenchymal transitions and stemness. *J. R. Soc. Interface* **11**, 20140962 (2014).
40. Wang, J., Zhang, K. & Wang, E. K. Kinetic paths, time scale, and underlying landscapes: a path integral framework to study global natures of nonequilibrium systems and networks. *J. Chem. Phys.* **133**, 1–13 (2010).
41. Wang, J., Li, C. & Wang, E. K. Potential and flux landscapes quantify the stability and robustness of budding yeast cell cycle network. *Proc. Natl Acad. Sci. USA* **107**, 8195–8200 (2010).
42. Kauffman, S. Differentiation of malignant to benign cells. *J. Theor. Biol.* **31**, 429–451 (1971).
43. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
44. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).
45. Huang, S., Ernberg, I. & Kauffman, S. Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. *Semin. Cell Dev. Biol.* **20**, 869–876 (2009).
46. Creixell, P., Schoof, E. M., Eler, J. T. & Linding, R. Navigating cancer network attractors for tumorspecific therapy. *Nat. Biotechnol.* **30**, 842–848 (2012).
47. Davies, P. & Lineweaver, C. Cancer tumors as metazoa 1.0: tapping genes of ancient ancestors. *Phys. Biol.* **8**, 015001 (2011).
48. Jin, H., et al. Snail is critical for tumor growth and metastasis of ovarian carcinoma. *Int. J. Cancer* **126**, 2102–2111 (2010).
49. Liang, Y., et al. Transcriptional network analysis identifies bach1 as a master regulator of breast cancer bone metastasis. *J. Biol. Chem.* **287**, 33533–33544 (2012).
50. Sachdeva, M. & Mo, Y. Y. mir-145-mediated suppression of cell growth, invasion and metastasis. *Am. J. Transl. Res.* **2**, 170–180 (2010).
51. Hu, X., et al. The heterochronic microRNA let-7 inhibits cell motility by regulating the genes in the actin cytoskeleton pathway in breast cancer. *Mol. Cancer Res.* **11**, 240–250 (2013).
52. Wells, D. K., Kath, W. L. & Motter, A. E. Control of stochastic and induced switching in biophysical networks. *Phys. Rev. X* **5**, 031036 (2015).
53. Zhou, X., et al. Adaptive minimum action method for the study of rare events. *J. Chem. Phys.* **128**, 104111 (2008).
54. Marusyk, A., Almendro, V. & Polyak, K. Intra-tumour heterogeneity: a looking glass for cancer? *Nat. Rev. Cancer* **12**, 323–334 (2012).
55. Lawson, D. A., et al. Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. *Nature* **526**, 131–135 (2015).
56. Petropoulos, S., et al. Single-cell RNA-seq reveals lineage and x chromosome dynamics in human preimplantation embryos. *Cell* **165**, 1012–1026 (2016).
57. Wu, F., Su, R. Q., Lai, Y. C. & Wang, X. Engineering of a synthetic quadrastable gene network to approach waddington landscape and cell fate determination. *eLife* **6**, e23702 (2017).
58. Kalsbeek, A., Merrow, M., Roenneberg, T. & Foster, R. Suprachiasmatic nucleus: cellular clocks and networks. *Neurobiol. Circadian Timing* **199**, 129 (2012).
59. Tetta, C., Ghigo, E., Silengo, L., Deregibus, M. C. & Camussi, G. Extracellular vesicles as an emerging mechanism of cell-to-cell communication. *Endocrine* **44**, 11–19 (2013).
60. Skog, J., et al. Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. *Nat. Cell Biol.* **10**, 1470–1476 (2008).
61. Sasai, M. & Wolynes, P. Stochastic gene expression as a many-body problem. *Proc. Natl Acad. Sci. USA* **100**, 2374–2379 (2003).
62. Zhang, B. & Wolynes, P. G. Stem cell differentiation as a many-body problem. *Proc. Natl Acad. Sci. USA* **111**, 10185–10190 (2014).
63. Anderson, D. F., Craciun, G. & Kurtz, T. G. Product-form stationary distributions for deficiency zero chemical reaction networks. *Bull. Math. Biol.* **72**, 1947–1970 (2010).
64. Hu, G. *Stochastic Forces and Nonlinear Systems* (Shanghai Scientific and Technological Education Press, Shanghai, 1994).
65. Van Kampen, N. G. *Stochastic Processes in Chemistry and Physics* (North Holland, Amsterdam, 1992).
66. Vellela, M. & Qian, H. Stochastic dynamics and non-equilibrium thermodynamics of a bistable chemical system: the schogl model revisited. *J. R. Soc. Interface* **6**, 925–940 (2009).
67. Swain, P. S., Elowitz, M. B. & Siggia, E. D. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl Acad. Sci. USA* **99**, 12795–12800 (2002).
68. Kaern, M., Elston, T. C., Blake, W. J. & Collins, J. J. Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.* **6**, 451–464 (2005).
69. Thattai, M. & Van, O. A. Intrinsic noise in gene regulatory networks. *Proc. Natl Acad. Sci. USA* **98**, 8614–8619 (2001).
70. Weinan, E., Ren, W. & Vanden-Eijnden, E. Minimum action method for the study of rare events. *Commun. Pure Appl. Math.* **57**, 637–656 (2004).
71. Heymann, M. & Vanden-Eijnden, E. The geometric minimum action method: a least action principle on the space of curves. *Commun. Pure Appl. Math.* **61**, 1052–1117 (2008).
72. Freidlin, M. & Weber, M. Random perturbations of dynamical systems and diffusion processes with conservation laws. *Probab. Theory Relat. Fields* **128**, 441–466 (2004).



**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/>.