PCCP

PAPER

View Article Online
View Journal | View Issue



Cite this: *Phys. Chem. Chem. Phys.*, 2016, **18**, 17949

# Quantifying the landscape and kinetic paths for epithelial–mesenchymal transition from a core circuit†

Chunhe Li,*[abc] Tian Hong[ab] and Qing Nie*[ab]

Received 11th May 2016,
Accepted 6th June 2016

DOI: 10.1039/c6cp03174a

www.rsc.org/pccp

Epithelial–mesenchymal transition (EMT), as a crucial process in embryonic development and cancer metastasis, has been investigated extensively. However, how to quantify the global stability and transition dynamics for EMT under fluctuations remains to be elucidated. Starting from a core EMT genetic circuit composed of three key proteins or microRNAs (microRNA-200, ZEB and SNAIL), we uncovered the potential landscape for the EMT process. Three attractors emerge from the landscape, which correspond to epithelial, mesenchymal and partial EMT states respectively. Based on the landscape, we analyzed two important quantities of the EMT system: the barrier heights between different basins of attraction that describe the degree of difficulty for EMT or backward transition, and the mean first passage time (MFPT) that characterizes the kinetic transition rate. These quantities can be harnessed as measurements for the stability of cell types and the degree of difficulty of transitions between different cell types. We also calculated the minimum action paths (MAPs) by path integral approaches. The MAP delineates the transition processes between different cell types quantitatively. We propose two different EMT processes: a direct EMT from E to P, and a step-wise EMT going through an intermediate state, depending on different extracellular environments. The landscape and kinetic paths we acquired offer a new physical and quantitative way for understanding the mechanisms of EMT processes, and indicate the possible roles for the intermediate states.

## 1 Introduction

Epithelial–mesenchymal transition (EMT) plays crucial roles in embryonic development and cancer metastasis.[1,2] Through EMT, the phenotypes of cells undergo many changes, such as losing cell–cell adhesion and cell polarity, and acquiring migratory and invasive properties.[3] Cells that have undergone EMT sometimes revert back to the epithelial state. This mesenchymal–epithelial transition (MET) enables the migratory cells to settle and proliferate at a distant organ. The dynamic and reversible behaviors of EMT are controlled by gene regulatory networks that consist of some key molecules and their interactions in epithelial and mesenchymal cells.[4] EMT networks have been studied using mathematical models,[5–7] as a ternary switch circuit with a noise-buffering integrator, or two cascading coupled bistable switches. Interestingly, the networks enable the formation of a hybrid epithelial–mesenchymal state, which is stable under certain environmental conditions. It has been proposed that the hybrid EMT phenotype is associated with invasiveness of cancer cells.[5,6]

In cells, there are intrinsic fluctuations due to limited number of molecules and external fluctuations due to inhomogeneous environments.[8–10] Noise has been shown to play critical roles in cell fate decision processes during developmental patterning.[11–13] However, it is unclear what roles noise play in the EMT networks, and it remains challenging to elucidate the global properties of the EMT system, such as the global stability of different states and associated dynamics under fluctuations. The potential landscape theory might provide a route for addressing these issues. Waddington[14] proposed a landscape picture for development and differentiation of cells, as a metaphor. Recently the Waddington landscapes for development as well as for cancer have been quantified.[15–22] From the landscape theory, different phenotypes can be depicted as the basins of attraction on a potential surface, and the cell fate decision process is viewed as a ball rolling down from one basin to another on the landscape. Functional states correspond to high probability or low potential states, and non-functional states correspond to the low probability or high potential states. The barrier heights between the attractors or basins quantify

[a] *Department of Mathematics, University of California, Irvine, Irvine, CA, USA. E-mail: chunhe.lee@gmail.com, qnie@math.uci.edu; Tel: +16315385236*
[b] *Center for Complex Biological Systems, University of California, Irvine, Irvine, CA, USA*
[c] *Shanghai Center for Mathematical Sciences, Fudan University, Shanghai, China*
† Electronic supplementary information (ESI) available. See DOI: 10.1039/c6cp03174a

the degrees of difficulty for the cells to switch from one cell type to another. In addition, an open question of the cell fate decision is whether the noise (including extrinsic noise and intrinsic noise) drives the transitions on a fixed potential landscape or, alternatively, the noise acts in concert with an exogenously altered potential landscape. This can be addressed by investigating how the landscape shape changes as the key parameters in the network (*e.g.* the synthesis or degradation rate of genes, or regulatory strength among genes) and the noise level are varied. The landscape theory as a metaphor has been used as a pictorial illustration for EMT transitions previously.[6,23] However, a more rigorous and quantitative study of EMT networks using the landscape theory is needed to gain more insights into the EMT system.

It is understood that the cell fate determination among three phenotypes (epithelial, partial EMT, mesenchymal cells) is governed by a core regulatory circuit consisting of two modules with four core components. There are two transcription factors SNAIL and ZEB and two microRNAs (miRs) miR-34 and miR-200. The circuit is mostly determined by two cross-inhibition feedback loops, miR-34/SNAIL circuit and miR-200/ZEB circuit. For simplicity, we only consider one couple of mutual repression, *i.e.* the regulation between the protein ZEB and the microRNA miR-200, in this study (Fig. 1) SNAIL is treated as an input for the network, which activates the ZEB transcription factor. The two transcription factors SNAIL and ZEB promote the expression of some mesenchymal marker genes, such as N-cadherin and vimentin, and repress the expression of epithelial marker genes, such as E-cadherin.[24] Therefore, this core circuit is expected to capture the major regulation and dynamics of EMT.

In this work, we aim to uncover the potential landscape for epithelial–mesenchymal transition. We start from a core EMT circuit (miR-200/ZEB) with 3 key genes or microRNAs. We uncovered the underlying potential landscape for the EMT process. Three attractors emerge on the landscape, and characterize the epithelial, mesenchymal, and partial EMT states (intermediate state), individually. We quantified the transition processes between these cell types by identifying the minimum action paths (MAP) among these three attractors. The forward transition MAP (from E to M) and backward transition MAP (from M to E) are not identical, indicating the irreversibility caused by the non-zero probabilistic flux. We found that the potential barriers for basins are correlated with the kinetic transition rates among these cell types, quantified by the first passage time (MFPT). We also construct a population-level state transition model, which can be harnessed to compare with experiments directly. Our landscape approach also indicates possible explanation for the existence of intermediate states. We propose two different EMT processes. In the condition with a strong signal (*e.g.* TGFβ signal or environment pressure), the system might go from the E state to the M state directly. In the condition with a weak signal, the system might make an indirect transition, from E to P, and finally to the M state. This provides a way to increase the plasticity of cell fate decision.

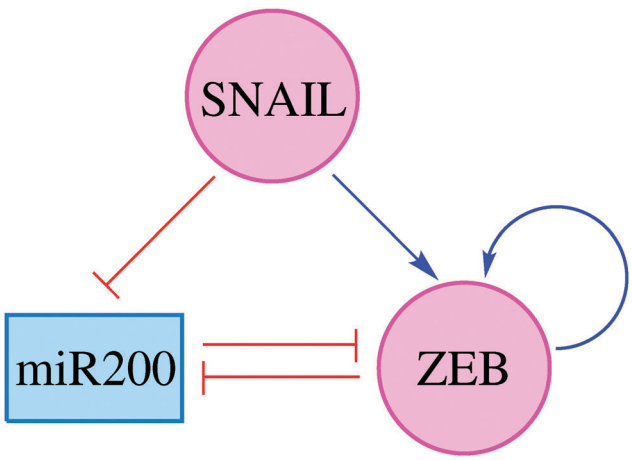


Fig. 1 The diagram for the core EMT circuit including 3 nodes (proteins or microRNAs). Blue arrows represent transcriptional activation and red short bars represent transcriptional or translational repression. Rectangular node represents microRNA, and circle nodes represent proteins.

## 2 Results

### 2.1 Potential landscape and minimum action path (MAP) for the EMT circuit

The probability evolution for a stochastic dynamical system can be captured by the diffusion equations. For a 2-dimensional system, the diffusion equation has the form:[25,26]

$$\frac{\partial P(x_1,x_2,t)}{\partial t} = -\frac{\partial}{\partial x_1}[F_1(x_1,x_2)P] - \frac{\partial}{\partial x_2}[F_2(x_1,x_2)P] + D\left(\frac{\partial^2 P}{\partial {x_1}^2} + \frac{\partial^2 P}{\partial {x_2}^2}\right) \quad (1)$$

Here $P(x_1, x_2, t)$ is the probability evolution function, $x_1$ and $x_2$ represent the concentration of the micro-RNA miR-200 and the protein ZEB, respectively, and $F_1$ and $F_2$ are the driving force for variables $x_1$ and $x_2$ ($F_1 = \frac{dx_1}{dt}$ and $F_2 = \frac{dx_2}{dt}$). The forms of $F_1$ and $F_2$ are shown in eqn (3) (see the Methods section). $D$ is diffusion coefficient matrix. In this work, we assume that $D$ is homogeneous and constant (the diagonal elements $D_{11} = D_{22} = D$ and the non-diagonal elements $D_{12} = D_{21} = 0$). In the current work, we focus on the effects of external noises (environment changes) on the EMT system, and apply the additive noise in our model (the diagonal elements of the diffusion matrix are constant). We have applied the landscape theory to different systems under external noises[16–18,22] and under intrinsic noises[27] (multiplicative noise is used), separately. In mammalian cells, the number of protein molecules is usually abundant. Therefore we expect that the source of fluctuations is mostly from the external noise rather than from the internal noise.

By solving the diffusion equations for the long time limit, we obtained the steady state probabilistic distribution of the EMT network. We take the zero flux boundary condition, *i.e.* $n\vec{j}_{ss} = 0$, where $n$ indicates the unit normal vector of boundary. This corresponds to the conservation of total probability. We used COMSOL Multiphysics (version 4.3) to solve the diffusion

 

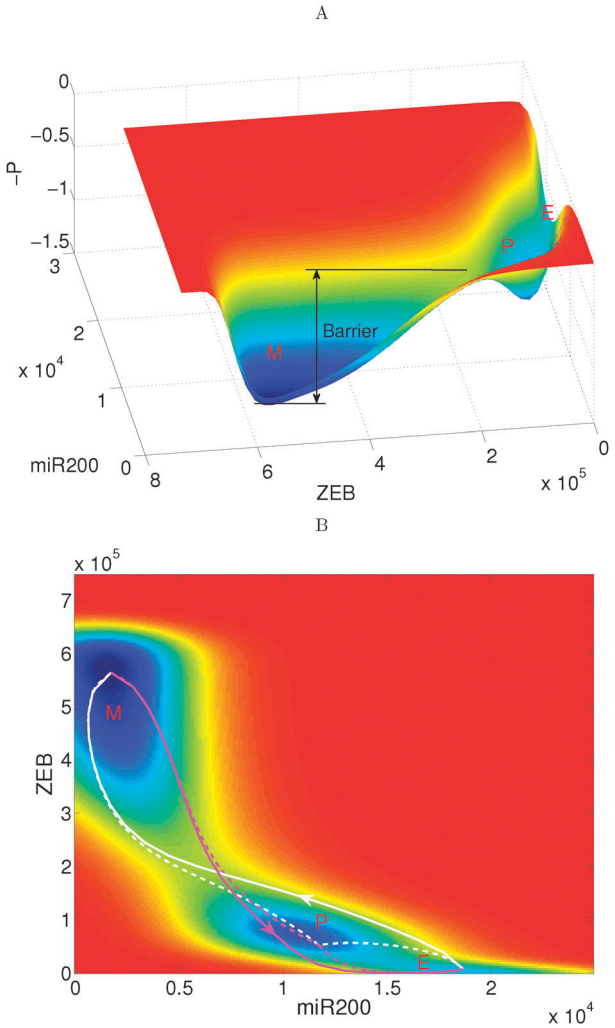


Fig. 2 The landscape and the corresponding kinetic paths for the EMT network are shown in 3-dimensional (A) and 2-dimensional figures (B). White lines represent the MAP for EMT, and the purple lines represent the MAP for MET. The dashed lines represent the MAP between P and E states as well as between P and M states. The parameters are set as: Snail = 200 K molecules (thousand molecules). E: epithelial state, M: mesenchymal state, P: partial EMT state.

equations.[28] In this way, we mapped out the potential landscape for the EMT system (Fig. 2) by $U = -\ln(P_{ss})$.[16–19,29–31] Here, $P_{ss}$ represents the probability distribution of the steady state, and $U$ is the dimensionless potential.

Fig. 2 displays a tristable landscape for the EMT system. The blue region represents lower potential or higher probability, and the red region represents higher potential or lower probability. Here, the three basins of attraction represent three cellular phenotypes (E for epithelial, M for mesenchymal, and P for the partial EMT state) respectively. From the landscape point of view, the transition between different cellular types can be understood as a ball rolling from one basin to the other by surmounting certain barriers (saddle points).

We calculated the dominant kinetic transition paths between different attractors by minimizing the transition action (see methods to acquire the dominant transition path), which are also called minimum action paths (MAPs). The MAPs for different transitions are shown on the landscape (Fig. 2B). The white (from the E state to the M state, EMT) and magenta (from the M state to the E state, MET) lines represent the MAPs with arrows denoting the directions of the transitions. We noticed that the MAPs do not go through the corresponding saddle points exactly. This is because the dynamics of non-equilibrium systems is not only determined by the landscape gradient but also the curl flux.[29,32,33]

Both EMT and MET processes have been observed experimentally by manipulating the TGF-β level.[34,35] The MAP for EMT and the MAP for MET are irreversible, reflected by the fact that forward and reverse paths are not identical (Fig. 2B, compare the white and magenta solid lines). The irreversibility for EMT paths has also been found by other work.[6] This irreversibility of MAPs is a consequence of the non-gradient force, curl flux.[29,33] The MAP from E to M shows that the EMT does not necessarily go through the P state (Fig. 2B, white solid line), which might provide an explanation why experimentally the partial EMT states cannot always be easily observed. In addition, we observed that both EMT and MET processes are initiated by rapid downregulation of the genes that are highly expressed in the original state (*e.g.*, the MET path, represented by the magenta solid line, starts with rapid downregulation of ZEB). This prediction in terms dynamics of key molecules can be tested experimentally.

The dashed lines in Fig. 2B show the kinetic paths between the intermediate P state and the E state as well as between the P state and the M state. The MAPs acquired between E, P, and M states show that the path from E, through P, to M (white dashed line) is not the minimum action path for the EMT. In other words, the minimum action path for EMT (white solid line) does not go through the intermediate P state. One natural question following this is why cells sometimes choose an EMT transition path with a higher cost, or larger action, by switching to P and then to M rather than going to M directly and how cells manage to achieve it. One possibility is that this is one way for cells to increase plasticity. A step-wise EMT from E, to P, then to the M state, should correspond to the case in which the system does not have strong enough signals to make complete EMT transition (*e.g.*, from weak TGF-β signal or limited environment pressure). In this case, the system may first enter the intermediate P state to stay for a while due to the limited driving signals. Depending on the extracellular or metabolic conditions, the system will decide whether it will further switch to M to finish the EMT or will return back to the E state (purple dash line from P to E). Therefore, the intermediate partial EMT state increases the plasticity for EMT, and ensures the robustness of EMT dynamics.

To answer the question how cells manage to follow the route going through P, rather than the direct path (least action path), we calculated the value of least actions for the transitions among E, P, and M states (choice of parameters are corresponding to Fig. 2), shown in Table 1. We found that although the total transition action for the staged EMT ($S_{E\to P} + S_{P\to M}$ = 2.36 + 6.79 = 9.15) is larger than the transition action of the direct EMT ($S_{E\to M}$ = 8.73), the first step of the staged EMT has

Table 1 Least actions for the transitions among E, P, and M states. The transition actions are represented by a 3 × 3 matrix $S$, with $S_{ij}$ representing the least action from state $i$ to state $j$. N denotes not applicable

| $S_{ij}$ | E state | P state | M state |
|---|---|---|---|
| E state | N | 2.36 ($S_{E\to P}$) | 8.73 ($S_{E\to M}$) |
| P state | 2.48 ($S_{P\to E}$) | N | 6.79 ($S_{P\to M}$) |
| M state | 8.75 ($S_{M\to E}$) | 6.34 ($S_{M\to P}$) | N |

smaller action than the direct EMT ($S_{E\to P} = 2.36 < S_{E\to M} = 8.73$). Therefore E state cells actually have very high probability to go to the P state because the transition action from E to P is much smaller than the transition action from E to M.

We found that for the reverse direction (MET), the difference in MAPs between the direct and staged transitions is overall less than the corresponding difference for the forward direction (EMT) (Fig. 2B). This is supported by the calculation of transition actions (Table 1). The difference of the transition actions between direct and staged MET is actually very small, which can be quantified by the relative change of actions $((S_{M\to P} + S_{P\to E} - S_{M\to E})/S_{M\to E} = 0.8\%)$. In contrast, the relative change of actions for the direct EMT and the staged EMT is $(S_{E\to P} + S_{P\to M} - S_{E\to M})/S_{E\to M} = 4.8\%$. The similarity in the paths of direct and staged MET suggests that the cells do not need to commit to one of these paths at the early stage of transition, and they can decide whether they need to revert to P or E even after the initiation of differentiation. In the situation of EMT, the cells need to decide whether they need to commit to P or M at the beginning of the transition because the decision would be difficult to change later. This indicates that the partial EMT has to be regulated more precisely, because errors at the early stage may lead to dramatic consequences later on.

### 2.2 Mean first passage time (MFPT) and potential barrier

For the current simplified EMT circuit, the SNAIL level represents the major input signal. This is because SNAIL is the immediate downstream signaling molecule of some external stimulations, including TGF-beta.[5,6] So, we explored the effects of the SNAIL level on the landscape for EMT, as shown in Fig. 3. As the SNAIL level increases, the landscape switches gradually from a stable E state to a P and M coexisting state, to tristability, and finally to a stable M state. A strong SNAIL level promotes the transition of cell types from epithelial inclined cells to mesenchymal inclined cells, which is indicated by the change in landscape topography.

We also explored the effects of the noise level (characterized by the diffusion coefficient $D$) on the landscape. As the noise level increases (Fig. 4), the landscape becomes more and more widely distributed, and the attractors become less stable. When the noise is very large, the E state disappears, and only the P state and the M state exist. This suggests a possibility that noise can work in concert with the altered potential landscape and induce EMT transitions.

The stability of the dynamical system is relevant to the escape time from the basins of attraction. For the probabilistic description of the network using diffusion equations, the mean first passage time (MFPT) $\tau(\mathbf{x})$ starting from a state $\mathbf{x}$ is determined

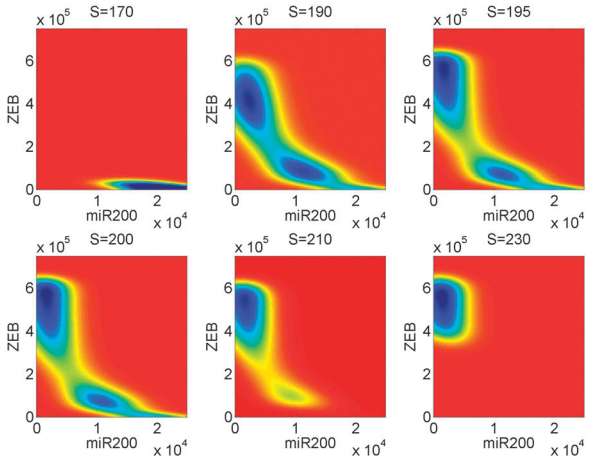


Fig. 3 The landscape at different Snail levels. The unit for S (SNAIL level) is K molecules (thousand molecules). The blue region represents higher probability, and the red region represents lower probability.

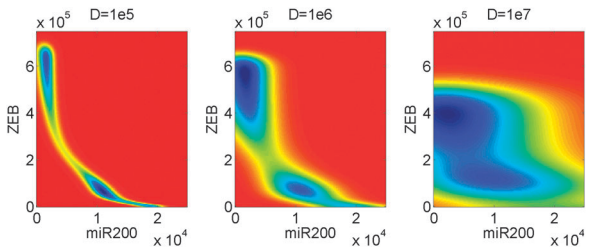


Fig. 4 The landscape at different diffusion coefficients $D$, quantifying different noise levels. The blue region represents higher probability, and the red region represents lower probability.

by[19,26] $\mathbf{F}\cdot\nabla\tau + D\nabla\tau\cdot\mathbf{D}\cdot\nabla\tau = -1$. It reflects the average time for a transition from an initial state to a given final state. In the final state an absorbing boundary condition ($\tau = 0$) is taken, and for the outer boundary a reflecting boundary condition $n\cdot\nabla\tau = 0$ is taken.

By solving the above partial differential equation numerically, we calculated the MFPT from one attractor to another (Fig. 5A and B). As the level of SNAIL increases, the MFPT for EMT decreases gradually and the MFPT for MET increases gradually (Fig. 5A). This behavior arises because SNAIL promotes EMT through activating the ZEB transcription factor as well as inhibits micro-RNA miR-200, therefore making the transition time of EMT shorter. Fig. 5B shows that as the SNAIL level increases, the MFPT from E to P increases and the MFPT from P to M decreases (Fig. 5B). This indicates that SNAIL promotes EMT mostly by affecting the second half stage (the stage from P to M).

To quantify the stability of the system based on the landscape topography, we define the barrier height as the potential difference between the local minimum and the saddle point. The barrier height can also be defined as the potential difference between the local minimum and the global maximum along the MAP by considering the effects of the probabilistic flux.[32] A larger barrier means that it is harder for the system to cross

 

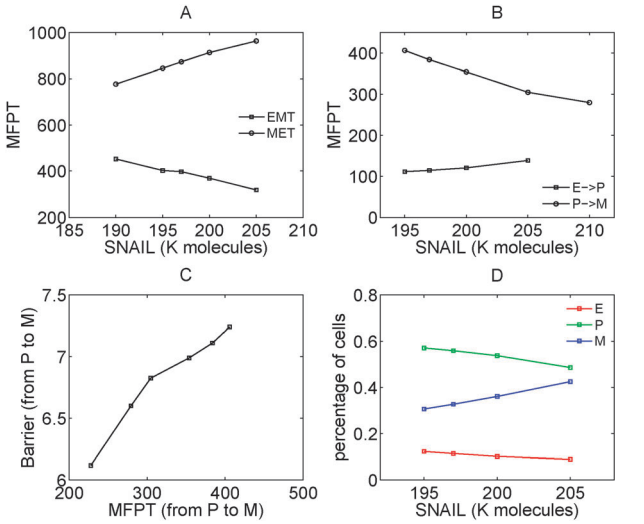


Fig. 5 (A and B) MFPT for EMT and MET transitions at different Snail levels. (C) Barrier *versus* MFPT from the P state to the M state. (D) Cell percentage for three phenotypes at steady states from the state transition model at different Snail levels. K molecules: thousand molecules.

the barrier and switch from one basin to another. Therefore, the barrier height provides a quantitative measure for the global stability for the EMT system. To investigate the relationship between MFPT and barrier height, we focus on a specific transition process, *i.e.* the transition from the partial EMT state to the M state. The good correlation between MFPT and barrier height (Fig. 5C) is reasonable because the larger barrier makes it harder for the system to escape from one basin, and therefore leads to a larger switching time. This implicates that the potential barrier from landscape topography and the MFPT provide quantitative measures for the global stability of the EMT system.

Based on the landscape and MFPT, our model predicts that the larger SNAIL level leads to less transition time from the E state to the M state (Fig. 5A). This is partially supported by the experiments indicating that SNAIL promotes EMT by activating the TGFβ signal.[36] Our results also predict that the switching time from E to P increases while the switching time from P to M decreases as the SNAIL level increases (Fig. 5B). This suggests that SNAIL promotes EMT mostly by speeding up the transition from the P (partial EMT) state to the M state. This prediction can be tested by future experiments.

From the MFPT we acquired, we estimated the transition rate (inverse of the MFPT) between different attractors (cell types). We further constructed a Markov cell state transition model (ESI†) to calculate the population evolutions for cells. The landscape and the population model predict that as the SNAIL signal is strengthened, the steady state proportion for E state cells and P state cells decreases and the steady proportion for M state cells increases (Fig. 5D). This prediction is consistent with the landscape results (Fig. 3), and can be tested experimentally. We have summarized the predictions from our model and some possible validations from experiments in Table 2.

Table 2 Model predictions and experimental supports

| Model predictions | Supporting references |
|---|---|
| Forward and reverse paths for EMT are not identical (Fig. 2B) | Ref. 6 |
| EMT starts with rapid downregulation of miR-200 (Fig. 2B) | Prediction |
| MET starts with rapid downregulation of ZEB (Fig. 2B) | Prediction |
| Increasing the SNAIL level leads to faster EMT (Fig. 5A) | Ref. 34 |
| Increasing the SNAIL level leads to slower transition from E to P and faster transition from P to M (Fig. 5B) | Prediction |
| As the SNAIL level is increased, E state and P state cell population decreases and M state cell population increases (Fig. 5D) | Prediction |

# 3 Discussion

The epithelial–mesenchymal transition play critical roles in embryonic development and cancer metastasis.[2] A core circuit for EMT (miR-200 and ZEB cross-inhibition feedback loop) has been investigated both experimentally and theoretically. However, learning the global properties for the EMT system is still challenging, particularly using deterministic approaches.

In this work, based on a simplified miR-200/ZEB model considering the microRNA regulation in detail, we uncovered the potential landscape for this system. We introduced the stochastic description for the network, and solved the probability evolution diffusion equations. The landscape for EMT displays three basins of attraction, which corresponds to three different phenotypes: epithelial (E), mesenchymal (M), and partial EMT states (P). In the process of EMT, some intermediate state cell types have been observed including partial epithelium and partial mesenchyme.[2] The partial EMT in adult epithelial tissues is also seen as the sign of tumor initiation.[37,38] In our landscape picture, the P basin on the landscape corresponds to these intermediate (partial EMT) states, and the transition from E to P or from M to P represents the partial EMT transition. We quantified the MAP as the kinetic paths for the transitions between these three attractors. The irreversible nature of MAPs embodies the effects of the non-equilibrium curl flux.[29,33,39]

We also calculated the switching time between different cell types, characterized by the mean first passage time (MFPT). The SNAIL promotes EMT mainly by influencing the transition from P to M, *i.e.* the increase of the SNAIL level speeds up the switching from P to M. To quantify the landscape change, we define the potential barrier. The barrier has a positive correlation with the MFPT, indicating that it can serve as a quantitative measure for the relative stability of different basins in the EMT system.

By estimating the transition rate among different cell types from MFPT, we expanded our cellular level model to a state transition model in the level of cell population. The population model provides a way to make comparisons with experiments directly. It would also be interesting to incorporate the effects of selection pressure (different proliferation rates for different cell types) to the population model and see the influence of selection pressure on the EMT process.

The partial EMT or intermediate states have been suggested to play important roles in various biological processes, for example, during wound healing, cells at the border of the wound experience a partial EMT process.[1] These intermediate cell types move as maintaining loose contacts rather than migrating as individual cells. Our landscape picture provides some hints on the roles of intermediate states (P state here). The transition for EMT (from E to M) is usually driven by signals (*e.g.* TGFβ signal or environment pressure), since it needs to cross different potential barriers on the landscape. However, if in some case the system does not have a strong enough driving signal, the intermediate state might be valuable. For example, the system can first switch from the E to P state and stay there for a while (because at this time the signal is not strong enough to drive the system going to M directly), and make further switch from the P to M state or going back to the E state at other times, depending on the environments. Therefore, the intermediate states increase the plasticity of cell fate determination, and increase the robustness of EMT dynamics.

Recent evidence suggests that cells undergoing EMT gain stem cell-like properties, thus giving rise to cancer stem cells (CSCs).[40–42] This suggests that ZEB, an EMT marker gene, also promotes the formation of CSCs. CSCs are highly tumorigenic cell types and are suggested to play important roles in oncogenesis due to their abilities of initiating tumors and driving metastasis.[43,44] It has been hypothesized that the frequent recurrence of cancer after chemotherapy or radiotherapy is due to the CSCs not being removed in these treatments.[45] In our landscape picture, the M attractor, marked by the activation of the ZEB gene, can be close to some CSC attractors in some higher-dimensional gene expression state spaces. When the cancer cells or normal cells undergo an EMT, they not only initiate metastasis, but also promote the formation of CSCs. Therefore, our results also suggest that the ZEB and relevant microRNA (miR-200) may serve as potential anti-cancer targets by blocking the EMT process or forming CSCs.

The simplified network used in this study is the minimum motif that governs a tristable EMT system. Our investigation on the transitions driven by both noise and external control of the landscape provided new physical insights for understanding quantitatively the mechanisms of EMT, which is not discussed in the previous studies focusing on the deterministic models.[5–7] We expect that most of our conclusions are valid for more complete EMT networks. However, more complex networks[6,23,46] involve more molecular details, which may influence the dynamics of the EMT system. For example, a recent work suggests that the mutual repression between OVOL2 and ZEB are critical to the multi-step EMT transitions.[23] Understanding the mechanisms of those multi-step transitions in the larger networks using landscape theory will be the topic of future studies. We anticipate that by exploring the landscape and paths of these networks, one can obtain more information on the mechanisms of the EMT process as well as cancer metastasis. This will facilitate the development of new strategies for cancer prevention or treatment. Our method can also be applied to other genetic circuits, to study their stochastic dynamics and kinetic transition paths.

# 4 Methods

## 4.1 Model of the EMT network

EMT modeling has been performed both deterministically and stochastically in some recent studies.[5,6] Lu *et al.* developed a theoretical framework to explore the micro-RNA translation-transcription regulation dynamics.[5] It is believed that the three phenotypes are determined by a core regulatory unit consisting of four core components, two transcription factors SNAIL and ZEB and two microRNAs (miRs) miR-34 and miR-200. The circuit is mostly determined by two mutually repressed feedback loops. For simplicity, we only consider one couple of mutual repressed genes, ZEB and miR-200 (Fig. 1). SNAIL is treated as an input for the network, which promotes the activation of ZEB. The two transcription factors SNAIL and ZEB promote the expression of some mesenchymal marker genes, such as N-cadherin and vimentin, and repress the expression of epithelial marker genes, such as E-cadherin.[24]

Starting from the topology of the network, we can write down the ordinary differential equations (ODEs) that govern the time evolution of variables (gene expression level or protein concentration). With the Hill function describing the activation or repression regulations, the ODEs have the following form:

$$\begin{aligned}\dot{\mu 200} &= g1 \times HS(ZEB, \lambda_{Z1}) \times HS(S, \lambda_{S1}) - mZEB \times Y\mu(\mu 200) \\ &\quad - k1 \times \mu 200 \\ \dot{mZEB} &= g2 \times HS(ZEB, \lambda_{Z2}) \times HS(S, \lambda_{S2}) - mZEB \times Ym(\mu 200) \\ &\quad - k2 \times mZEB \\ \dot{ZEB} &= g3 \times mZEB \times L(\mu 200) - k3 \times ZEB \end{aligned} \quad (2)$$

Here, μ200 denotes miR-200, mZEB denotes mRNA of ZEB, and ZEB represents the ZEB protein. In addition, g1 and g2 are the basal synthesis rate for miR200 and mZEB respectively, k1, k2 and k3 are the constant degradation rates of miR200, mZEB and ZEB respectively, and g3 is the translation rate of protein ZEB for each mRNA in the absence of microRNAs. The translation rate $L(\mu)$ is defined as $L(\mu) = \sum_{i=0}^{n} l_i C_n^i M_n^i(\mu)$, the mRNA active degradation rate $Ym(\mu)$ is defined as $Ym(\mu) = \sum_{i=0}^{n} rm_i C_n^i M_n^i(\mu)$, and the miR-200 active degradation rate is defined as $Y\mu(\mu) = \sum_{i=0}^{n} ru_i C_n^i M_n^i(\mu)$. $l_i$ is the individual translation rate, $rm_i$ and $ru_i$ are the individual active degradation rates for mRNA and microRNA, where $M_n^i(\mu) = (\mu/\mu 0)^i/(1 + \mu/\mu 0)^n$.[5,47] Here HS is the shifted Hill function, defined as $HS(B, \lambda) = H^-(B) + \lambda \times H^+(B)$, $H^-(B) = 1/(1 + (B/B_0)^{nB})$, $H^+(B) = 1 - H^-(B)$. $\lambda$ is the fold change from the basal synthesis rate due to protein ZEB. $\lambda > 1$ represents activations and $\lambda < 1$ represents repressions. nB is the number of binding sites for different regulations. The SNAIL level S serves as the external input signal for the EMT circuit. We followed the previous work for

 

Table 3 Parameter values for EMT model. k: thousand molecules

| | | Hour$^{-1}$ | | Hour$^{-1}$ | | Molecules | |
|---|---|---|---|---|---|---|---|
| nBZ1 | 3 | $\lambda_{Z1}$ | 0.1 | k1 | 0.05 | S01 | 180k |
| nBZ2 | 2 | $\lambda_{Z2}$ | 7.5 | k2 | 0.5 | S02 | 180k |
| n | 6 | $\lambda_{S1}$ | 0.1 | k3 | 0.1 | S | 200k |
| nBS1 | 2 | $\lambda_{S2}$ | 10 | g3 | 0.1k | μ0 | 10k |
| nBS2 | 2 | g1 | 2.1k | g2 | 11 | | |

the parameter selection (Tables 3 and 4), and the number of the binding sites for miR-200 regulations to ZEB mRNA was chosen to be $n = 6$.[5]

The degradation rate of mRNA has been shown to be much faster than that of proteins and microRNAs,[48] so we assume that mRNA can reach quick equilibrium. By setting $F_2 = \mathrm{d(mZEB)}/\mathrm{d}t = 0$, we can obtain mZEB = g2 × HS(ZEB, $\lambda_{Z2}$) × HS($S$, $\lambda_{S2}$)/(Ym(μ200) + k2), and further reduce the 3-variable equations to 2-variable equations for μ200 and ZEB:

$$\dot{\mu 200} = \mathrm{g1} \times \mathrm{HS(ZEB}, \lambda_{Z1}) \times \mathrm{HS}(S, \lambda_{S1}) - \mathrm{mZEB} \times \mathrm{Y}\mu(\mu 200) - \mathrm{k1} \times \mu 200$$

$$\dot{\mathrm{ZEB}} = \mathrm{g3} \times \mathrm{g2} \times \mathrm{HS(ZEB}, \lambda_{Z2}) \times \mathrm{HS}(S, \lambda_{S2})/(\mathrm{Ym}(\mu 200) + \mathrm{k2}) \times L(\mu 200) - \mathrm{k3} \times \mathrm{ZEB} \quad (3)$$

Writing the equations governing the dynamics of the EMT system in the vector form as $\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{F}(\mathbf{x}) + \eta(t)$, where the $F(x)$ is the driving force for the system. For the stochastic system, the system is determined by probability distributions. The probability distribution follows master equations for the discrete case and diffusion equations for the continuous case. In continuous representation, the diffusion equations governing probability evolution can be written as $\partial P(\mathbf{x},t)/\partial t + \nabla\cdot\mathbf{J}(\mathbf{x},\mathbf{t}) = 0$, where the probability flux $\boldsymbol{J}$ takes the form $\boldsymbol{J}(\mathbf{x},t) = \mathbf{F}(\mathbf{x})P(\mathbf{x},t) - D\nabla P(\mathbf{x},t)$.

Biological systems are usually open systems exchanging energy and materials with environments, and form a non-quilibrium steady state (NESS). One of the distingushed characteristics of a NESS is the nonzero flux in the system.[49] So the driving force $F$ can be decomposed to two terms ($\mathbf{F} = -D\cdot\nabla U + \mathbf{J}_{ss}/P_{ss}$), the potential gradient $U$ and the probability flux $\mathbf{J}$ with divergence-free nature.

### 4.2 Kinetic path from path integral

In the cells, there exist intrinsic noise from statistical fluctuations of the finite number of molecules, and external noise from highly dynamical and inhomogeneous environments. Both of them can be significant to the dynamics of the system.[8–10]

Table 4 Parameter values for miR binding rates in the EMT model

| nB (number of miR binding sites) | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $l_i$ | 1.0 | 0.6 | 0.3 | 0.1 | 0.05 | 0.05 | 0.05 |
| $rm_i$ (hour$^{-1}$) | | 0.04 | 0.2 | 1.0 | 1.0 | 1.0 | 1.0 |
| $ru_i$ (hour$^{-1}$) | | 0.005 | 0.05 | 0.5 | 0.5 | 0.5 | 0.5 |

Therefore, one needs to study the cellular network dynamics under fluctuating conditions in order to model the cellular inner and outer environments realistically. A network of chemical reactions in fluctuating environments can be addressed by $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}) + \zeta$, where $\mathbf{x} = (x_1(t), x_2(t), \ldots, x_6(t))$ represents the vector of protein concentrations or gene expression levels. $\mathbf{F}(\mathbf{x})$ is the vector for the driving force of chemical reactions or gene regulations. $\zeta$ is the Gaussian noise term satisfied with: $\langle\zeta_i(\mathbf{x},t)\rangle = 0$ and $\langle\zeta_i(\mathbf{x},t)\zeta_i(\mathbf{x},t')\rangle = 2\mathrm{D}_{ij}\delta_{ij}\delta(t-t')$ ($\delta_{ij} = 1$ for $i = j$, and $\delta_{ij} = 0$ for $i \neq j$), where $\delta(t)$ is the Dirac delta function and $D$ is the diffusion coefficient matrix.

The dynamics for the probability of starting from an initial configuration $\mathbf{x}_{\mathrm{initial}}$ at $t = 0$ and ending at a final configuration $\mathbf{x}_{\mathrm{final}}$ at time $t$, in terms of the Onsager–Machlup functional, can be formulated[15,50] as

$$P(\mathbf{x}_{\mathrm{final}}, t, \mathbf{x}_{\mathrm{initial}}, 0) = \int \mathbf{D}\mathbf{x} \exp\left[-\int \mathrm{d}t \left(\frac{1}{2}\nabla\cdot\mathbf{F}(\mathbf{x}) + \frac{1}{4}(\mathrm{d}\mathbf{x}/\mathrm{d}t - \mathbf{F}(\mathbf{x}))\cdot\frac{1}{\mathbf{D}(\mathbf{x})}\cdot(\mathrm{d}\mathbf{x}/\mathrm{d}t - \mathbf{F}(\mathbf{x}))\right)\right] = \int \mathbf{D}\mathbf{x}\exp[-S(\mathbf{x})] = \int \mathbf{D}\mathbf{x}\exp\left[-\int L(\mathbf{x}(t))\mathrm{d}t\right].$$

Here, $\mathbf{D}(\mathbf{x})$ is the diffusion coefficient matrix. The integral over $\mathbf{D}\mathbf{x}$ denotes the sum over all possible paths from the state $\mathbf{x}_{\mathrm{initial}}$ at time $t = 0$ to the state $\mathbf{x}_{\mathrm{final}}$ at time $t$. The exponent factor gives the weight of each path. Thus, the probability of network dynamics from the initial state $\mathbf{x}_{\mathrm{initial}}$ to the final state $\mathbf{x}_{\mathrm{final}}$ is equal to the sum of all possible paths with different weights. $S(\mathbf{x})$ is the transition action and $L(\mathbf{x}(\mathrm{t}))$ is the Lagrangian or the weight for each path.

The path integrals can be approximated with a set of dominant paths, since each path is exponentially weighted, and the other subleading path contributions are often small and can be neglected. So, the dominant path with the optimal weight can be acquired through minimization of the action or Lagrangian. In our case, we identify the dominant paths as the biological paths for EMT.

## Acknowledgements

C.L. thanks Prof. Jin Wang for valuable comments on the manuscript. C.L. thanks Feng Zhang for helpful discussions. The study was partially supported by NSF grants DMS 1161621, and DMS 1562176, and NIH grants P50-GM076516, R01-GM107264, and R01NS095355.

## References

1 J. P. Thiery, H. Acloque, R. Y. Huang and M. A. Nieto, *Cell*, 2009, **139**, 871–890.
2 Y. Nakaya and G. Sheng, *Cancer Lett.*, 2013, **341**, 9–15.
3 M. A. Nieto, *Annu. Rev. Cell Dev. Biol.*, 2011, **27**, 347–376.
4 M. A. Nieto, *Science*, 2013, **342**, 1234850.
5 M. Lu, H. Jolly, H. Levine, J. Onuchic and E. Ben-Jacob, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 18144–18149.

6 J. Zhang, X. Tian, H. Zhang, Y. Teng, R. Li, F. Bai, S. Elankumaran and J. Xing, *Sci. Signaling*, 2014, **7**, ra91.
7 X. Tian, H. Zhang and J. Xing, *Biophys. J.*, 2013, **105**, 1079–1089.
8 P. S. Swain, M. B. Elowitz and E. D. Siggia, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 12795–12800.
9 M. Kaern, T. C. Elston, W. J. Blake and J. J. Collins, *Nat. Rev. Genet.*, 2005, **6**, 451–464.
10 M. Thattai and O. A. Van, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 8614–8619.
11 L. Zhang, K. Radtke, L. Zheng, A. Q. Cai, T. F. Schilling and Q. Nie, *Mol. Syst. Biol.*, 2012, **8**, 613.
12 L. Wang, J. Xin and Q. Nie, *PLoS Comput. Biol.*, 2010, **6**, e1000764.
13 T. Hong, E. S. Fung, L. Zhang, G. Huynh, E. S. Monuki and Q. Nie, *J. R. Soc., Interface*, 2015, **12**, 20150258.
14 C. H. Waddington, *The strategy of the genes: a discussion of some aspects of theoretical biology*, Allen and Unwin, London, 1957, p. 290.
15 J. Wang, K. Zhang, L. Xu and E. K. Wang, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 8257–8262.
16 C. Li and J. Wang, *PLoS Comput. Biol.*, 2013, **9**, e1003165.
17 J. Wang, L. Xu, E. K. Wang and S. Huang, *Biophys. J.*, 2010, **99**, 29–39.
18 C. Li and J. Wang, *J. R. Soc., Interface*, 2014, **10**, 20140774.
19 F. Zhang, L. Xu, K. Zhang, E. K. Wang and J. Wang, *J. Chem. Phys.*, 2012, **137**, 065102.
20 C. Chen, W. Baumann, J. Xing, L. Xu, R. Clarke and J. Tyson, *J. R. Soc., Interface*, 2014, **96**, 20140206.
21 J. T. Margaret, B. K. Chu, M. Roy and E. L. Read, *Biophys. J.*, 2015, **109**, 1746–1757.
22 C. Li and J. Wang, *Cancer Res.*, 2015, **75**, 2607–2618.
23 T. Hong, K. Watanabe, C. H. Ta, A. Villarreal-Ponce, Q. Nie and X. Dai, *PLoS Comput. Biol.*, 2015, **11**, e1004569.
24 H. Peinado, D. Olmeda and A. Cano, *Nat. Rev. Cancer*, 2007, **7**, 415–428.
25 G. Hu, *Stochastic Forces and Nonlinear Systems*, Shanghai Scientific and Technological Education Press, Shanghai, 1994, pp. 68–74.
26 N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry*, North Holland, Amsterdam, 3rd edn, 2007.
27 C. Li, J. Wang and E. K. Wang, *Chem. Phys. Lett.*, 2011, **505**, 75–80.
28 *Introduction to COMSOL Multiphysics, version 4.3*, COMSOL Inc., Burlington, MA, 2012.
29 J. Wang, C. Li and E. K. Wang, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 8195–8200.
30 C. Li and J. Wang, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 14130–14135.
31 C. Liao and T. Lu, *J. Phys. Chem. B*, 2013, **117**, 12995–13004.
32 H. Feng, K. Zhang and J. Wang, *Chem. Sci.*, 2014, **5**, 3761–3769.
33 J. Wang, L. Xu and E. K. Wang, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 12271–12276.
34 P. A. Gregory, C. P. Bracken, E. Smith, A. G. Bert, J. A. Wright, S. Roslan, M. Morris, L. Wyatt, G. Farshid and Y.-Y. Lim, *Mol. Biol. Cell*, 2011, **22**, 1686–1698.
35 A. Gal, T. Sjoblom, L. Fedorova, S. Imreh, H. Beug and A. Moustakas, *Oncogene*, 2008, **27**, 1218–1230.
36 D. Medici, E. D. Hay and B. R. Olsen, *Mol. Biol. Cell*, 2008, **19**, 4875–4887.
37 L. Huang and S. K. Muthuswamy, *Curr. Opin. Genet. Dev.*, 2010, **20**, 41–50.
38 A. Grosse-Wilde, A. F. dHerouel, E. McIntosh, G. Ertaylan, A. Skupin, R. E. Kuestner, A. del Sol, K.-A. Walters and S. Huang, *PLoS One*, 2015, **10**, e0126522.
39 M. Turcotte, *Quant. Biol.*, 2016, **4**, 69–83.
40 S. A. Mani, W. Guo, M.-J. Liao, E. N. Eaton, A. Ayyanan, A. Y. Zhou, M. Brooks, F. Reinhard, C. C. Zhang and M. Shipitsin, *Cell*, 2008, **133**, 704–715.
41 A. Singh and J. Settleman, *Oncogene*, 2010, **29**, 4741–4751.
42 X. Liu, S. Johnson, S. Liu, D. Kanojia, W. Yue, U. P. Singh, Q. Wang, Q. Wang, Q. Nie and H. Chen, *Sci. Rep.*, 2013, **3**, 2473.
43 K. Sampieri and R. Fodde, *Seminars in cancer biology*, 2012, pp. 187–193.
44 T. Reya, S. J. Morrison, M. F. Clarke and I. L. Weissman, *Nature*, 2001, **414**, 105–111.
45 L. L. C. Marotta and K. Polyak, *Curr. Opin. Genet. Dev.*, 2009, **19**, 44–50.
46 M. K. Jolly, D. Jia, M. Boareto, S. A. Mani, K. J. Pienta, E. Ben-Jacob and H. Levine, *Oncotarget*, 2015, **6**, 25161–25174.
47 M. Lu, M. K. Jolly, R. Gomoto, B. Huang, J. Onuchic and E. Ben-Jacob, *J. Phys. Chem. B*, 2013, **117**, 13164–13174.
48 A. Lipshtat, A. Loinger, N. Q. Balaban and O. Biham, *Phys. Rev. Lett.*, 2006, **96**, 188101.
49 H. Qian, *J. Phys. Chem. B*, 2006, **110**, 15063–15074.
50 J. Wang, K. Zhang, H. Lu and E. Wang, *Biophys. J.*, 2005, **89**, 1612–1620.

 