Title
Landscape and kinetic path quantify critical transitions in epithelial-mesenchymal transition

Journal
Biophysical Journal, 120(20)

ISSN
0006-3495

Authors
Lang, Jintong
Nie, Qing
Li, Chunhe

Publication Date
2021-10-01

DOI
10.1016/j.bpj.2021.08.043

Peer reviewed

Biophysical *Journal*
Article

Biophysical Society

Landscape and kinetic path quantify critical transitions in epithelial-mesenchymal transition

Jintong Lang,[1,2] Qing Nie,[3] and Chunhe Li[1,2,4,*]
[1]Institute of Science and Technology for Brain-Inspired Intelligence and [2]Shanghai Center for Mathematical Sciences, Fudan University, Shanghai, China; [3]Department of Mathematics, University of California, Irvine, Irvine, California; and [4]School of Mathematical Sciences, Fudan University, Shanghai, China

ABSTRACT Epithelial-mesenchymal transition (EMT), a basic developmental process that might promote cancer metastasis, has been studied from various perspectives. Recently, the early warning theory has been used to anticipate critical transitions in EMT from mathematical modeling. However, the underlying mechanisms of EMT involving complex molecular networks remain to be clarified. Especially, how to quantify the global stability and stochastic transition dynamics of EMT and what the underlying mechanism for early warning theory in EMT is remain to be fully clarified. To address these issues, we constructed a comprehensive gene regulatory network model for EMT and quantified the corresponding potential landscape. The landscape for EMT displays multiple stable attractors, which correspond to E, M, and some other intermediate states. Based on the path-integral approach, we identified the most probable transition paths of EMT, which are supported by experimental data. Correspondingly, the results of transition actions demonstrated that intermediate states can accelerate EMT, consistent with recent studies. By integrating the landscape and path with early warning concept, we identified the potential barrier height from the landscape as a global and more accurate measure for early warning signals to predict critical transitions in EMT. The landscape results also provide an intuitive and quantitative explanation for the early warning theory. Overall, the landscape and path results advance our mechanistic understanding of dynamical transitions and roles of intermediate states in EMT, and the potential barrier height provides a new, to our knowledge, measure for critical transitions and quantitative explanations for the early warning theory.

SIGNIFICANCE Epithelial-mesenchymal transitions (EMTs) play critical roles in cancer metastasis and development. Recent studies have proposed that early warning signals can be used to predict critical transitions in EMT. However, the mechanism of early warning related to cell fate decisions and stochastic transition dynamics in EMT has yet to be elucidated. Here, we constructed the potential landscape of EMT based on a comprehensive gene regulatory network model and identified multiple attractor cell states, including epithelial, mesenchymal, and intermediate states. The transition path results highlight that intermediate states can accelerate EMT. Importantly, the potential barrier height provides a new, to our knowledge, measure for critical transitions and quantitative explanations for the mechanism of the early warning theory.

INTRODUCTION

Epithelial-mesenchymal transition (EMT) has been suggested to play an important role in the normal embryonic development of the metazoan body, differentiation into tissues and organs, and wound healing (1,2). In disease, EMT contributes to organ fibrosis (3) and the formation of metastasis in cancer (4). However, it remains elusive how to elucidate the underlying mechanism and transition dynamics of EMT quantitatively.

Mathematical models have been widely used to study the dynamics of EMT (5–19). Tian et al. established models showing that the negative feedback loop between SNAIL1 and miR-34 and the one between ZEB1 and miR-200 play important roles in EMT (17). Hong et al. extended the previous model and proposed the critical involvement of Ovol2 and its associated molecular regulations (12). Huang et al. established a more inclusive ordinary differential equation (ODE) model with 18 genes to simulate EMT (11). Steinway et al. identified network perturbations that suppress TGF-β-driven EMT, with the goal of suppressing invasive properties of cancer cells by Boolean modeling with 65 genes (20). Although different models have been proposed to study EMT based on deterministic or stochastic models, the global

Submitted May 10, 2021, and accepted for publication August 30, 2021.

*Correspondence: chunheli@fudan.edu.cn

Editor: Kevin Janes.

https://doi.org/10.1016/j.bpj.2021.08.043

stability and stochastic transition dynamics of EMT, e.g., the roles of intermediate states, have not been fully clarified.

The classic Waddington landscape has been proposed as a metaphor for the development and differentiation of cells (21). Recently, the Waddington epigenetic landscapes for biological networks have been quantified from various approaches (22–27) and used to study the stochastic dynamics of development (22,28,29) and cancerization (13,30). An advantage of the landscape-transition path approach is that it can address the problem of global stability and therefore provides a global and quantitative measure for how a perturbation (gene or regulation changes) will influence the transition (cell fate decisions). In this work, we aim to employ landscape theory to study the stochastic dynamics of EMT. Firstly, we constructed a gene regulatory network for EMT by collecting evidence of gene interactions based on previous works (9–12,16), which includes 16 nodes (genes) and 53 links (regulations). Then, we constructed the ODE models based on the network structure. By screening the parameter space, we identified the parameter regions for multistable states. The multistability and intermediate E/M states have been suggested both from theoretical and experimental explorations (9,10,16,17,31). Therefore, we focus on the parameter regions that can generate more than two stable states, which means that in addition to the epithelial state (E state) and the mesenchymal state (M state), there will also be intermediate states between them in various cases. As a typical example, we map out a landscape with six stable attractors or cell states, which correspond to the E state, M state, and four intermediate states.

Based on the transition action theory, we calculated the kinetic transition paths of EMT, which include the dynamical information of the genes switching on or off. The transition paths predicted by our model are supported by experimental data. We performed single-factor sensitivity analysis to uncover how each link influences the cell state transition and identified effective strategies for regulating target genes to inhibit EMT. Previous works (18,19) emphasized that intermediate states play critical roles on EMT process. Goetz et al. inferred that intermediate states may accelerate EMT in some cases from a simplified state transition model (32). However, the evidence for this conclusion from a molecular network perspective is still lacking. To test whether this conclusion holds from the EMT gene network perspective, we calculated the distribution of transition actions from E to M state for a wide range of parameter regions and found that the appearance of intermediate state will make the transition action decreasing (corresponding to faster transition) in certain scenarios. These results provide a theoretical basis for intermediate state accelerating EMT, from the perspective of molecular network dynamics.

Recently, predicting tipping points has been a key topic because it may be important to prevent "critical transitions," e.g., in cases of ecosystems, the climate, and the financial market (33–35), as well as in many cases related to disease (36–40). Different indicators based on time-series data have been suggested as early warning signals (EWSs) to predict tipping points, including variance, autocorrelation function, coefficient of variation, etc. (41–48). Recently, Sarkar et al. identified the variance as the EWS for EMT based on a simplified EMT regulatory network model (49). Specifically, they proposed the "critical slowing down" of the system as it approaches the tipping point as the key property for EWSs, which are linked to the variance in the EMT case. Here, we intend to explore how this theory works in a more comprehensive EMT network model. We performed corresponding analysis related with EWSs from our EMT regulatory network. We identified the potential barrier height from the landscape as a new, to our knowledge, EWS and demonstrated that it performs better than other traditional indicators including variance, autocorrelation, etc., in terms of predicting the critical phase transition. In fact, the "critical slowing down" of the system in transition period can be well explained by the decrease of barrier height from landscape topography. This provides a possible physical explanation for the well-known EWS theory. Overall, our work provides a holistic view for the mechanistic understanding of EMT transition dynamics and suggests new roles of intermediate states. The landscape approach provides a new, to our knowledge, tool to predict critical cellular transitions in EMT and metastasis progression.

MATERIALS AND METHODS

Self-consistent mean field approximation

The time evolution of a dynamical system is determined by a probabilistic diffusion equation (Fokker-Planck equation). Given the system state $P(x_1, x_2, \ldots, x_N, t)$, with $x_1, x_2, \ldots, x_N$ representing the levels of components (e.g., the gene expression level), we will have a N-dimensional partial differential equation. In our network, there are 16 genes, so $N = 16$ here. It is difficult to solve a high-dimensional diffusion equation because of the huge state space of the system. Following a self-consistent mean field approach (24,28,50,51), we can split the probability into the products of probabilities of individual ones, $P(x_1, x_2, \ldots, x_N, t) \sim \prod_i^n P_i(x_i, t)$, and solve the probability self-consistently. In this way, we effectively reduce the dimensionality of the system from M^N to MN (M is the dimension for each gene expression level), and the computation of above problem becomes feasible.

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 this work, we assume Gaussian distribution as an approximation, which means we need to calculate two moments, the mean and the variance. When the elements in diffusion coefficient matrix D is small, the moment equations can be approximated to (52,53)

$$\dot{\overline{\mathbf{x}}}(t) = \mathbf{F}[\overline{\mathbf{x}}(\mathbf{t})] \tag{1}$$

and

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

Here, $\mathbf{x}$, $\boldsymbol{\sigma}(t)$, and $\mathbf{A}(t)$ are vectors and tensors and $\mathbf{A}^{\mathbf{T}}(t)$ is the transpose of $\mathbf{A}(t)$. The elements of matrix A are specified as $A_{ij} = \frac{\partial F_i[X(t)]}{\partial x_j(t)}$. $\mathbf{D}[\overline{\mathbf{x}}(\mathbf{t})]$ is the

diffusion matrix. For the external noise only cases, $\mathbf{D}[\overline{\mathbf{x}}(\mathbf{t})]$) is not dependent on $\overline{\mathbf{x}}(\mathbf{t})$, i.e., $\mathbf{D} = diag(d, d, \ldots, d)$. For the intrinsic noise cases, $\mathbf{D}\overline{\mathbf{x}}(\mathbf{t})]$) is dependent on $\overline{\mathbf{x}}(\mathbf{t})$ (see Supporting materials and methods). Based on these equations, we can solve $\overline{\mathbf{x}}(t)$ 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:

$$P(x,t) = \frac{1}{\sqrt{2\pi\sigma(t)}} e^{-\frac{[x-\overline{x}(t)]^2}{2\sigma(t)}}. \tag{3}$$

Here, $\overline{\mathbf{x}}(t)$ and $\boldsymbol{\sigma}(t)$ are the solutions of Eqs. 1 and 2. From the mean field approximation, we can extend this formulation to the multidimensional case by assuming that the total probability is the product of each individual probability for each variable: $P(x_1, x_2, \ldots, x_N, t) \sim \prod_{i=1}^{N} P(x_i, t)$, namely

$$P_j(x_1, x_2, \ldots, x_N, t) = \frac{1}{(2\pi)^{(N/2)} \prod_{i=1}^{N} \sqrt{\sigma_i(t)}} e^{-\sum_{i=1}^{N} \frac{[x_i - \overline{x}_i(t)]^2}{2\sigma_i(t)}}. \tag{4}$$

The probability distribution obtained above corresponds to one steady state or basin of attraction. If the system has multiple stable states, there will be several probability distributions localized at each basin with different variances. Therefore, the total probability is the weighted sum of all these probability distributions:

$$P(x_1, x_2, \ldots, x_N, t) = \sum_{j=1}^{N_{ss}} P_j(x_1, x_2, \ldots, x_N, t) \cdot w_j, \tag{5}$$

where N_{ss} denotes the total number of the stable states and j denotes the corresponding stable state.

The weighting factors (w_j) characterize the relative sizes of different basin of attraction. We determine the weights w_j by giving a large number of random initial conditions to the ODEs to be solved and collecting the statistics from all of these different solutions. Finally, we can construct the potential landscape by $U(x) = -lnP_{ss}(x)$ (22,24), with P_{ss} representing the steady-state probability distribution (in practice, we let t be very large to obtain the steady state). To visualize the landscape, we chose two of the transcriptional factors and microRNAs to display a two-dimensional landscape.

State transition path from path-integral approach

After we obtain the landscape, we calculate the transition path from one stable state to another one. Based on the path-integral approach (22,54), we have

$$\begin{aligned} P_t(\mathbf{x}_{final}, t, \mathbf{x}_{initial}, 0) &= \int \mathbf{x} exp\Big[- \int dt \Big(\frac{1}{2} \nabla \cdot \mathbf{F}(\mathbf{x}) \\ &\quad + \frac{1}{4} (d\mathbf{x}/dt - \mathbf{F}(\mathbf{x})) \cdot \frac{1}{\mathbf{D}(\mathbf{x})} \cdot \\ (d\mathbf{x}/dt - \mathbf{F}(\mathbf{x}))) &= \int D\mathbf{x} exp[-S(\mathbf{x}(t))] \\ &= \int D\mathbf{x} exp\Big[- \int L(\mathbf{x}(t)) dt \Big], \end{aligned} \tag{6}$$

where P_t represents the transition probability, $S(\mathbf{x}(t))$ is the action, and $L(\mathbf{x}(t))$ is the Lagrangian.

To calculate the most probable transition path from one stable state to another one, we need to minimize the action S to maximize the transition probability. Here, the Lagrangian is written as (22,54)

$$L(\mathbf{x}) = \frac{1}{4D} \dot{\mathbf{x}}^2 + V(\mathbf{x}) - \frac{1}{2D} \mathbf{F}(\mathbf{x}) \cdot \dot{\mathbf{x}}, \tag{7}$$

where $V(\mathbf{x}) = \frac{1}{4D}\mathbf{F}^2 + \frac{1}{2} \nabla \cdot \mathbf{F}(\mathbf{x})$.

So, we can write the generalized momentum and Hamiltonian

$$\mathbf{P_m}(\mathbf{x}) = \frac{\partial L}{\partial \dot{\mathbf{x}}} = \frac{1}{2D} (\dot{\mathbf{x}} - \mathbf{F}(\mathbf{x})) \tag{8}$$

and

$$H(\mathbf{x}) = -L(\mathbf{x}) + \mathbf{P_m}(\mathbf{x}) \cdot \dot{\mathbf{x}} = E_{eff}. \tag{9}$$

Here, we chose $E_{eff} = -V_{min}$, with V_{min} being the minimum of effective potential. In this case, the path connects two stable states, so V will reach its minimum when x is the most stable state among multiple stable states.

Then, we substitute Eq. 9 into the action and obtain $S(\mathbf{x}) = \int (\mathbf{P_m}(\mathbf{x}) \cdot \dot{\mathbf{x}} - H(\mathbf{x}))dt$. To calculate the action of the path, we need to transform the formulations into a different representation in x space and discretize the integral. The target function can be written as

$$S = \sum_{n=1}^{N_{tp}-1} \Big(\sqrt{(E_{eff} + V(n))/D} \Big) - \frac{1}{2D} F_l(n)) \Delta l_{n,n+1} + \mathrm{l} P_1, \tag{10}$$

where N_{tp} is the total number of points on the transition path and P_λ is a penalty function keeping all the length elements close to their average.

$$\begin{aligned} P_\lambda &= \sum_{n=1}^{N_{tp}-1} (\Delta l_{n,n+1} - \langle \Delta l \rangle)^2 \\ \Delta l_{n,n+1} &= \sqrt{\sum_{i=1}^{N} (\mathbf{x}_i(n+1) - \mathbf{x}_i(n))^2} \\ F_l(n) &= \sum_{i=1}^{N} \mathbf{F}_i(\mathbf{x}(n))(\mathbf{x}_i(n+1) - \mathbf{x}_i(n))/\Delta l_{n,n+1} \\ V(n) &= \sum_{i=1}^{N} \Big(\frac{1}{4D} \mathrm{F}^2(\mathbf{x}_i) + \frac{1}{2} \sum_{j=1}^{N} \frac{\partial \mathbf{F}_j(\mathbf{x}_i)}{\partial x_j} \Big), \end{aligned} \tag{11}$$

where N represents the number of genes (here $N = 16$) and i is the index for different dimensions. In this way, we can calculate the transition action of any path. Finally, we can obtain the most probable transition path by minimizing the transition action S. Here, we used simulated annealing algorithm for optimization.

Data availability statement

A MATLAB (The MathWorks, Natick, MA) implementation of the landscape and kinetic path of EMT has been deposited at GitHub (https://github.com/chunhelilab/EMT).

RESULTS

Gene regulatory network model of EMT

EMT is a complex and heterogenous biological process controlled by underlying gene regulatory networks. The regulatory networks of EMT have been studied from previous work based on ODE models (9–11,16,17). However, these previous studies usually focus on small circuits for EMT. To reveal deeper molecular mechanisms, we aim to construct a relatively complete gene regulatory network for EMT. Based on the EMT network of Huang et al. (11), we added an important transcriptional factor Ovol2 (12), an important microRNA miR-145 (13,14), and merged three microRNAs with similar functions (miR-200a, miR-200b, and miR-200c) and obtained a gene regulatory network for EMT including 16 components. We checked the experimental evidence of all gene regulations in this network by literature mining (Table S1). In this network, CDH1 and VIM are treated as the E marker and M marker, respectively.

Our EMT model includes 16 representative gene or microRNA components. Here, x_i ($i = 1, 2, \ldots, 16$) stands for the expression levels of TGF-β, ZEB1, ZEB2, SNAI1, SNAI2, TWIST1, FOXC2, GSC, TCF3, VIM, miR-145, miR-141, miR-200, miR-34a, Ovol2, and CDH1, respectively. The circuit consists of 12 transcriptional factors and four microRNAs. The network diagram is shown in Fig. 1.

Based on the gene regulatory network structure, we wrote down the ODEs to describe the time evolution of relative expression levels for each of the 16 genes or microRNAs. Here we used Hill functions to describe the activation and inhibition regulations among different genes (22,25,28). The ODEs take the form

$$F_i = \dot{x}_i = \frac{dx_i}{dt} = \sum_{j\in A_i} \frac{a_{ji}x_j^n}{s_{ji}^n + x_j^n} + \sum_{j\in B_i} \frac{b_{ji}s_{ji}^n}{s_{ji}^n + x_j^n} - k_i x_i. \quad (12)$$

Here, F_i represents the driving force for the time evolution of the activity of the ith gene ($i = 1, 2, \ldots, 16$). k_i is the degradation rate. A_i denotes the aggregate of genes that activate the ith gene in the network, and B_i denotes the aggregate of genes that inhibit the ith gene in the network. a_{ji} is the activation constant from the jth to the ith gene, and b_{ji} is for the inhibition. s_{ji} is the threshold level for the regulation from x_j to x_i, and n is the Hill coefficient of the regulation (see Supporting materials and methods for detailed descriptions of parameter setting).

Landscape with multiple intermediate states for EMT and mesenchymal-epithelial transition

Recent studies have identified intermediate cell states both at the single-cell level and population levels across different cancer types (31,55,56). To study the multistable properties and roles of intermediate states, a key property for the model is being able to generate multistable states. After screening the parameters (see Supporting materials and methods for details of parameter-searching approaches and Tables S2–S7 for results), we found that intermediate states commonly exist in a wide range of parameter space. There could be a monostable state or two to six states coexisting depending on parameter choice, and five or six states coexisting are very common (Tables S2–S7). Here, we focus on the multistable state cases. For example, we identified a parameter region leading to a six-stable-state landscape, including four

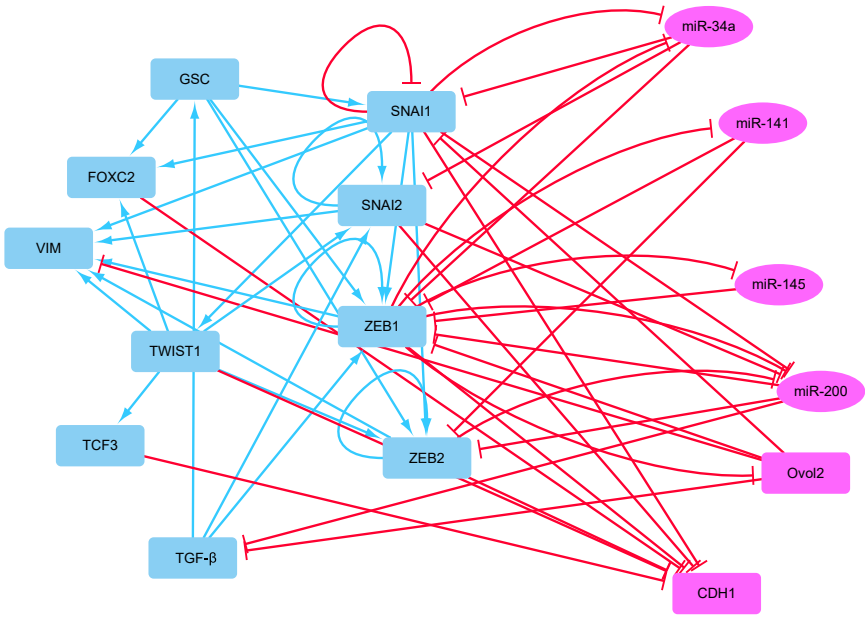

FIGURE 1 The diagram of EMT gene regulatory network. The network consists of 12 transcriptional factors (*rectangle nodes*), four microRNAs (*ellipse nodes*), and 53 regulatory links. Blue arrows represent activation, and red bars represent repression. The six genes represented by the pink nodes (miR-145, miR-141, miR-200, miR-34a, Ovol2, and CDH1) are E markers, and the other genes represented by the blue nodes are M markers. To see the figure in color, go online

intermediate states (Table S8). The parameter set of this case is listed in Table S9.

The probabilistic evolution of a stochastic dynamical system is determined by the diffusion equation. To study the stochastic dynamics of a gene regulatory network, one way is to solve the probability evolution from the diffusion equations. However, the diffusion equation is hard to solve directly for a high-dimensional system, e.g., the EMT network here. Here, we use a self-consistent approximation approach (see Materials and methods) to quantify the steady-state probability distribution (P_{ss}) (24,30,50,51). Then, we can obtain the landscape by $U = -ln(P_{ss})$ (22,24,26,28). The landscape here is a high-dimensional curved surface, which is hard for visualization. So, we chose the key variables VIM, ZEB1, and CDH1 as the coordinates and projected the landscape to two-dimensional space, as VIM and ZEB1 are major M marker genes and CDH1 is a major E marker gene (Figs. 2 and S1).

To illustrate the typical landscape with multiple intermediate states, here we used a case with six stable states coexisting as an example. On the landscape, the basins or attractors (*blue region*) represent stable states or phenotypes (Fig. 2; Table S8). It follows that the closer to the stable states, the lower the potential energy and the more likely the cell is to stay. The six basins of attraction on the landscape represent six different cell states characterized by different gene expression patterns in the 16-dimensional state space (corresponding to the six stable states in Table S8). These states correspond to the E state (high CDH1 and low VIM/ZEB1 expression), M state (high VIM/ZEB1 and low CDH1 expression), three intermediate states close to the E state (IE1, IE2, and IE3, intermediate expression close to the E state), and one intermediate state close to M state (IM, intermediate expression close to the M state), respectively. Fig. S2 shows the landscape projected onto other dimensions. The number of stable states and the property of intermediate states do not change, whereas in certain coordinates, some of the stable states might overlap. We also showed the tetrastable (Fig. S1, *A and B*) and tristable landscapes (Fig. S1, *C and D*) as parameter values vary, which shows that multistable states exist in a wide range of parameter space. In the following, we will take the six-stable-state landscape as an example to study the transition properties of EMT and the roles of intermediate states.

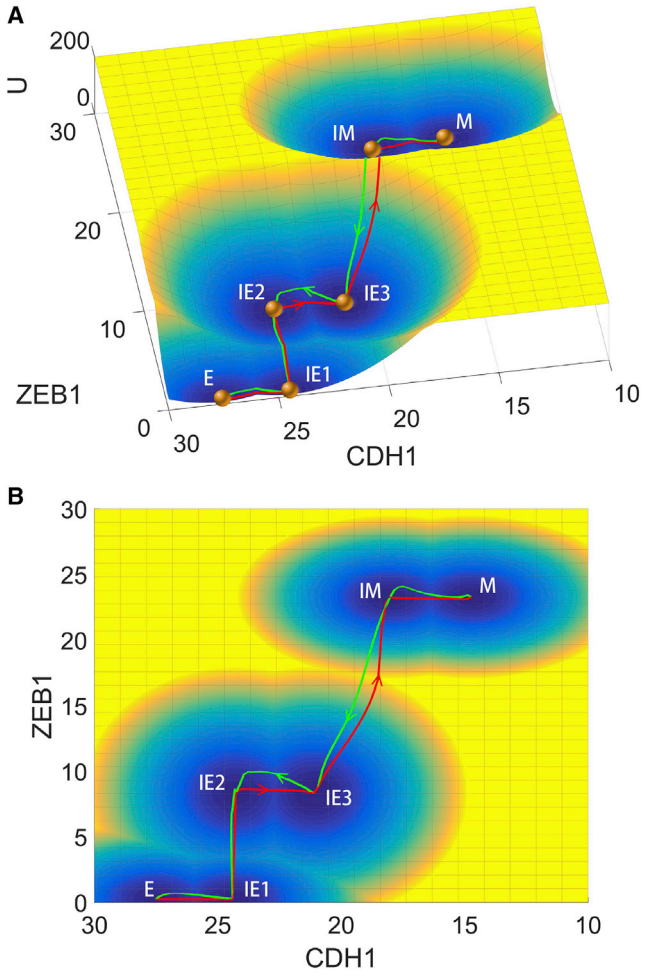

FIGURE 2 The landscape and transition paths between different attractor states in the CDH1/ZEB plane shown in three-dimensional (*A*) and two-dimensional (*B*) figures. The landscape surface is characterized by different colors, with the blue region representing lower potential or higher probability and the yellow region representing higher potential or lower probability. The red curve represents the transition path of EMT, and the green one represents the transition path of MET. The six stable states (E, IE1, IE2, IE3, IM, and M) correspond to the cell states in Table S8. E, epithelial state; IE1/2/3, intermediate epithelial state 1/2/3; IM, intermediate mesenchymal state; M, mesenchymal state. Here, $a = 2$, $\tilde{a} = 7.9$, $b = 4$, $s_a = 5.8$, and $s_b = 0.21$. To see the figure in color, go online

Kinetic transition paths for EMT are supported by experimental data

Transition path theories have been used to quantify the most probable path for the transition between cell attractors based on large deviation theory (57,58). However, one limitation for the transition path from large deviation theory is that it assumes zero noise limitation (59). To overcome this limitation, here we resort to a path-integral approach to calculate the transition paths between stable states so that we are able to study the effects of noise on these transition paths. Based on the path-integral approach (see Materials and methods for details) (22), we calculated the 16-dimensional transition path or minimal action paths (MAPs) between neighboring stable states. The MAPs are shown in Fig. 2, *A and B* on the landscape. The red MAP from the E state to the M state corresponds to the EMT process, and the green MAP from the M state to the E state corresponds to the mesenchymal-epithelial transition (MET) process. Here, we minimize the transition action by the method of simulated annealing. Fig. S3 displays the stability and convergence of the simulated annealing method, which supports the reliability of

the results for MAPs. The results of landscape and transition paths indicate that the cell fate transition process for EMT needs to go through the intermediate states (Figs. 2 and S1).

To support the transition path results from our models, we obtain some experimental data (TGF-β-treated lung adenocarcinoma cells (GSE17708)) (60), which yield a trajectory moving from the E to the M state, and make comparisons between the paths from experiments and the paths from our model. The comparison results for the transition paths are shown in Fig. 3 *A*. Here, the experimental data are obtained from two repeated experiments under the same conditions (60). We also visualized the 16-dimensional MAP from the E to M state by showing the expressions of the 16 genes changing with time, respectively. In current data sets, there are no data for all the microRNAs. Most data of GSC are zero, which is consistent with the fact that the expression of GSC is very small (Table S8). In addition, the data of TWIST1 have large errors (it has very intensive oscillation, and there is a big difference between the two groups of experimental data). So, we only show the comparison of paths for 10 components here. We found that the transition paths from models are essentially in agreement with the two experimental data groups (Fig. 3 *A*).

To quantify how the transition paths from modeling agree with experimental trajectories, we discretize the paths to binary form (Fig. 3 *B*). In this way, we can quantify the similarity between the paths from models and the paths from experiments. Here, we define the similarity for two matrices as the ratio between the number of same elements in two matrices and the number of matrix elements. Here, each element of a matrix represents the on or off state for corresponding genes (Fig. 3 *B*). The similarity degree between modeling and experiment 1 is 80.00%, and the one between modeling and experiment 2 is 76.67%, whereas the similarity degree between experiment 1 and experiment 2 is 83.33%. The similarity between experiments and models is similar to the one between the two experiments, which indicates that the MAP from modeling is supported by experiments.

To probe the roles of network structure on generating stable states and corresponding transition paths, we perturbed the network structure to recalculate the MAP of EMT and compared these MAPs with the experimental data as we did for the wild-type network (Fig. S4). Specifically, we modified n_r regulatory links in the network, for which we picked n_r as 10, 20, …, 200 (there are 16 genes in total,

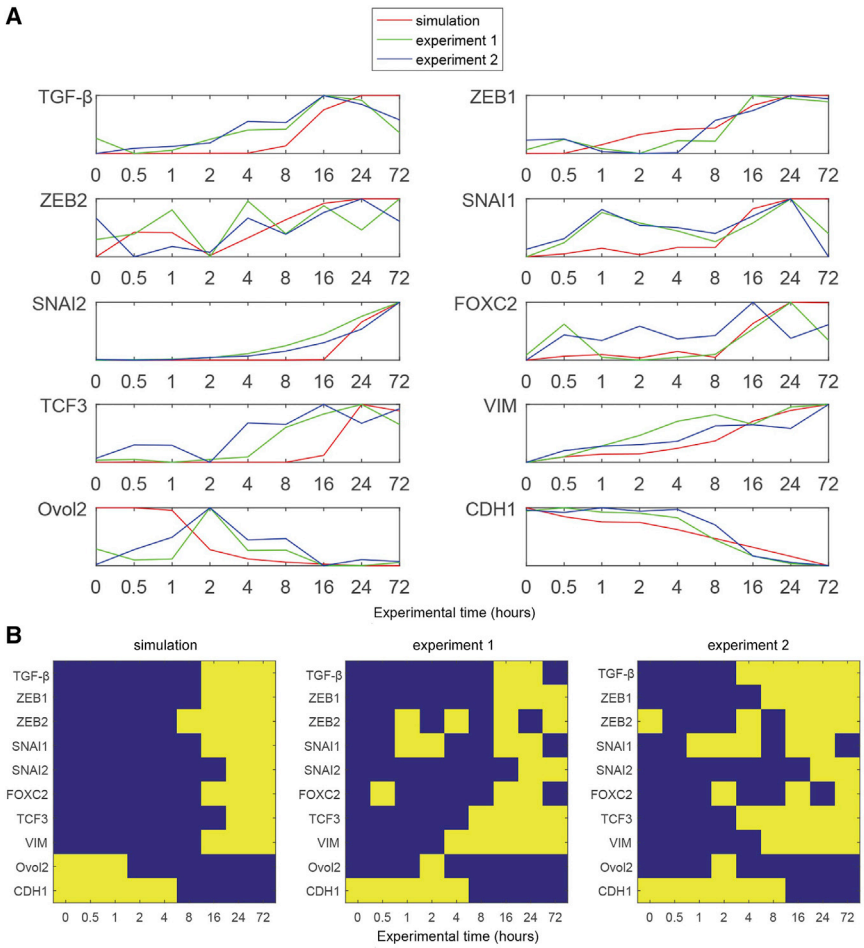

FIGURE 3 (*A*) Comparison for the transition path of EMT between the model and the experimental data in terms of activity of 10 different genes. The gene expressions on the paths are normalized for comparison. The experimental paths are based on the data from TGF-β-treated lung adenocarcinoma cells (GSE17708) (60). (*B*) Corresponding comparisons between models and experiments after the gene expressions are discretized to 0 or 1. Here, 1 represents that the expression of gene is high and 0 that the expression of gene is low. The *X* axis shows the nine time points along the transition path. To see the figure in color, go online

so the total possible edge number is $16 \times 16 = 256$). For each n_r, we simulated 30 different mutated networks and calculated the average and variance of their similarity degree with experimental data. Here, we perturbed the network so that each "mutant" network either has an edge removed or has an edge with reversed sign (from activation to inhibition or from inhibition to activation). We define the similarity degree as 0 when the mutated network only generates a monostable state, as the experimental data show two stable states, E and M. As shown in Fig. S4, *A and B*, with the increase of the mutated edge number n_r, the average value of similarity degree decreases, and it is much lower than the one for the wild-type network (Fig. 3). For example, when the mutated edge number is equal to or larger than 20, there is very low correlation (smaller than 0.15) for the transition path results between models and both experimental data sets (Fig. S4, *A and B*). These results support the importance of network topology on generating multistable E/M states and provide further support for our gene regulatory network model of EMT.

In addition, Fig. S4 *C* indicates that the average number of stable states decreases with the increase of n_r, which supports that multistability of EMT network is determined by the topology of the regulatory network. It has been shown in many regulatory networks, including those involved in EMT (61), that positive feedback loops enable multistability. Once several regulations are modified and such loops are broken, there is a high probability that the system will lose its multistability, consistent with recent experiments showing that breaking positive feedback loops will compromise multistable behavior (62).

An advantage of the path-integral approach compared with the methods based on large deviation theory is that we can study the effects of noise level on the transition paths. For the external noises cases (characterizing environment changes) (63,64), we explored the effects of noise level d (see Materials and methods) on the transition paths. As shown in Fig. S5, larger d leads to larger difference between the two paths (path from E to M and path from M to E). This is because larger d means the system has stronger ability of diffusion, which makes the system not have to follow the path with the lowest potential energy, and the effect of flux tends to be more obvious (24,65).

Intermediate states accelerate EMT

A critical question in EMT is what the role of intermediate states is. With an intermediate state, the cells can make changes on the expression of a small part of genes as the first step for a partial transition and stay in an intermediate state. The intermediate state possesses a certain level of stability (as an attractor), which will reduce the possibility of the cells returning to the E state. From the view of energy landscape, more intermediate states mean lower barrier height between the stable states, which may facilitate the transition. However, too many intermediate states make the energy level rise and fall repeatedly, which increases the difficulty of the transition. Therefore, the roles of multiple intermediate states might be a combinatorial effect of above two factors in terms of promoting or hindering EMT. Interestingly, recent work proposed that the intermediate states can accelerate EMT based on a state transition model (32). However, whether this conclusion holds from the molecular network level has yet to be clarified. So, we aim at studying this issue from a stochastic dynamics perspective based on our gene network model of EMT.

To uncover the roles of intermediate states, we studied the relationship between the transition action and the number of intermediate states (N_{IM}) as parameters vary. To avoid a significant change of landscape as parameters vary, we first performed analysis when parameters change in a small range (Fig. 4). In this parameter region, there are relative smaller changes in landscapes (e.g., similar E and M states are preserved as parameters vary), so we can focus on the role of intermediate states on state transitions. Here, the five parameters a, $\tilde{a}$, b, s_a, and s_b are selected independently and randomly ($a = 0.5$, $\tilde{a} = 8$, $b = 2$, $s_a = \sim 5.1$–5.7, and $s_b = \sim 1.2$–1.6). This parameter area is taken from the area of Table S7 that can generate two to six stable states (namely $\sim$0–4 intermediate states). We screened more than 2000 sets of parameters to calculate the transition actions and classified the results of transition actions by N_{IM} (Fig. 4, *A–E*). For each parameter set, we calculated the transition action from the E state to the M state based on the path-integral approach, which can also be used to measure the transition rate of EMT (9). Smaller transition action means faster transition. Similarly, we studied two other small parameter regions (Figs. S6 and S7), which are taken from the areas of Tables S3 and S5 that can generate two to six stable states. We found that for all three cases (three different small parameter ranges, corresponding to Figs. 4, S6, and S7, individually), there are certain optimal number of intermediate states (transition actions have some minima or transition rates have some maxima as the number of intermediate states increases). However, the optimal number of intermediate states is different under different parameter regions. The transition action reaches a minimum when there is one (Fig. S7), two (Fig. 4), or three (Fig. S6) intermediate states and reaches a maximum when there are four intermediate states (Figs. 4, S6, and S7). These results support that when parameter changes in relative smaller regions, the intermediate states promote the transitions from E to M, although the optimal number of intermediate states may depend on specific conditions.

Furthermore, to study how larger changes of landscapes affect the relationship between the transition action and the number of intermediate states, we screened more than 10,000 sets of parameters (in larger parameter ranges) to calculate the transition actions and classified the results of

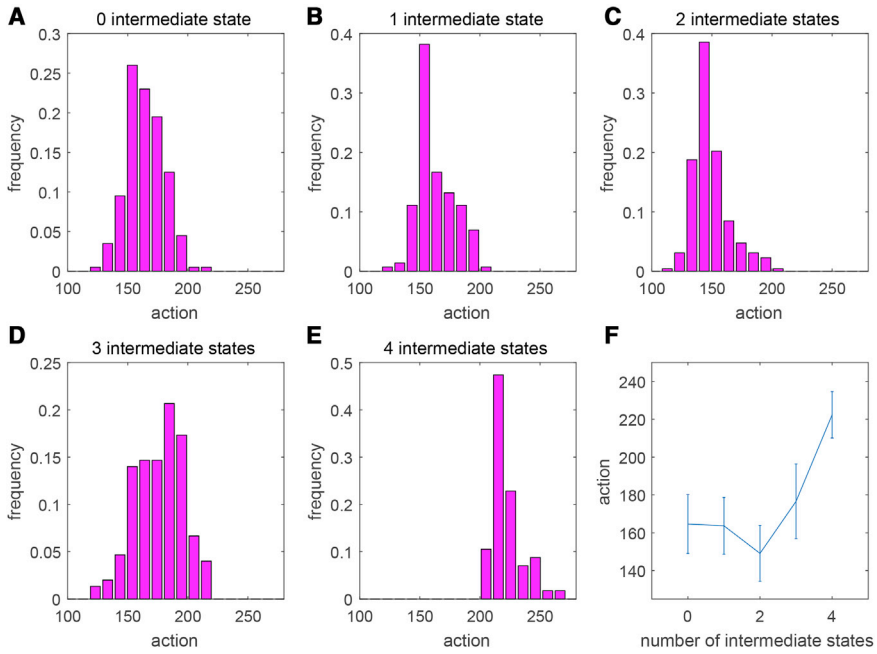

FIGURE 4 (*A–E*) The distribution for transition action of EMT for different numbers of intermediate states in a small parameter region ($a = 1$, $\vec{a} = 8$, $b = 2$, $s_a = \sim 5.1–5.7$, and $s_b = \sim 1.2–1.6$). Here, 2000 random parameter sets are sampled to calculate the transition action from the E state to the M state. (*F*) Mean transition action and corresponding variance (error bars) under different numbers of intermediate states (N_{IM}). To see the figure in color, go online

transition actions by the number of intermediate states (Fig. S8). This parameter area contains all the parameter areas of Figs. 4, S6, and S7. We found that the variances of transition actions are very large for the same N_{IM}, which indicates that even if N_{IM} is the same, transition paths may be quite different under larger parameter changes. This indicates that the roles of intermediate states on promoting EMT can be complicated and depend on specific conditions. Other factors, e.g., the interaction with the local microenvironment could also play a role.

These results suggest that some intermediate states may accelerate the EMT process, but too many intermediate states may hamper EMT. A possible explanation is that the occurrence of intermediate states increases the plasticity of cell fate transitions, and the stepwise transition lowers the total transition actions from E to M. However, there is also another outcome for the increase of number of intermediate states, which is that the system needs to cross more barriers to finish the transition from E to M. As the number of intermediate states increases, when the latter factor becomes dominant, it leads to the transition action increasing again and the EMT being hampered. However, the number of intermediate states with faster EMT may vary for different models of EMT depending on specific conditions. Our results based on the dynamical molecular network model suggest that the intermediate states may have a complex role. Certain number of intermediate states may promote EMT, which is consistent with recent findings (32). But we also propose that when the number of intermediate states is over a certain number, the EMT may be hampered (32). We expect that this prediction can be tested by further experimental study.

Barrier height from landscape serves as an EWS for critical transitions in EMT

Recently, the theory of EWSs have been applied to predict critical transitions in EMT. Some indicators for EWSs have been suggested, including variance and autocorrelation functions. In fact, in ecological and climate systems, many indicators related with system dynamics have been suggested as EWSs to predict critical transitions (41,42). Based on a simplified EMT model, Sarkar et al. proposed that there exists EWSs for EMT (49). It remains unclarified whether this conclusion holds in a more realistic EMT network including more molecular details. With our EMT regulatory network, we aim to study this issue.

We first generated stochastic time series of gene expressions by Langevin stochastic simulations (Supporting materials and methods). With parameter a (activation strength for regulations in EMT network) characterizing the signal of EMT activation, the trajectory is generated with time-varying signal a, which starts from 0 at day 0 and then increases up to 8 at day 20 (the rate of a changing (da/dt) is 0.016 $\mu M/h^2$). The network structure in Fig. 1 shows that all activation regulations are among M genes, so increasing a will promote transitions from the E state to the M state. Here, we considered the effects of internal noise in modeling (see Supporting materials and methods) because the level of system state (concentrations) will affect the variance of trajectories. Fig. S9 shows the stochastic trajectories of all 16 genes. We see there are three obvious abrupt changes in the trajectory of Fig. S9, so we divide the trajectory into four segments corresponding to four stages of EMT, i.e., E $\rightarrow$ IE, IE $\rightarrow$ IM, IM $\leftrightarrow$ M, and M state (Fig. 5). This is because the landscape

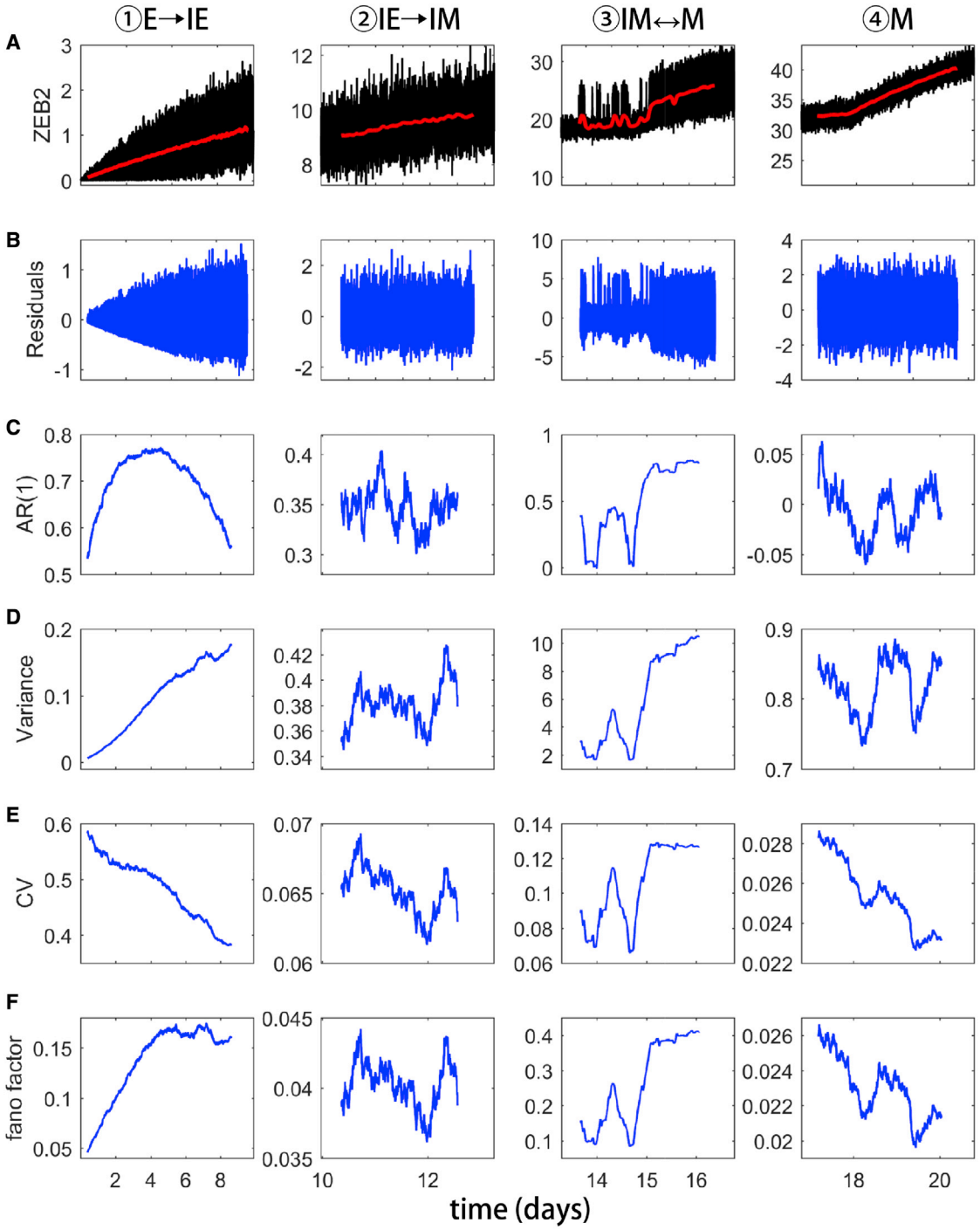

FIGURE 5 (*A*) The trajectory of ZEB2 for four different transition stages in EMT. (*B*) Residual time series after applying a Gaussian filter (red curve in figure *A* is the result after filtering). (*C–F*) Four indicators of EWSs for EMT calculated from the filtered time series after using a rolling window of 10% of the data length. Four indicators are AR (1) (*C*), variance (*D*), coefficient of variance (*E*), and Fano factor (*F*). Four columns from left to right correspond to four different transition stages (1, days ∼1–9; 2, days ∼10–12; 3, days ∼13–16; 4, days ∼17–20). To see the figure in color, go online

will change with the parameter a, leading to cells staying in different stable states at different time. The exact transition time points are day 9, day 12, and day 16 (Fig. 5).

We notice that the third segment of trajectory oscillates back and forth between the two stable states (the M state and the IM state). A question is whether this oscillation behavior is caused by the sudden increase of noise. To study this possibility, we compared the magnitude of internal noise (proportional to $\sqrt{\rho_i^+ + \rho_i^-}$) in the chemical Langevin equation (Eq. S6) with the bandwidth of the trajectory

(width of trajectory, Fig. S10). As shown in Fig. S10, the trajectory has three obvious abrupt changes when the parameter a changes uniformly that divide the trajectory into four segments. We see that the bandwidth has a sudden increase in the third segment of the trajectory, which characterizes the oscillation behavior in this stage. If this is caused by the sudden increase of noise, we will anticipate a similar increase of noise in the third segment of the trajectory. However, the change of noise as a increases is similar to the change of trajectory and different with the change of bandwidth. This means that the oscillation behavior in the third stage is not completely determined by the noise increase. It should be caused by the decrease of barrier height between stable states.

Then, we consider a time-series segment from the E state to the M state (Fig. 5 *A*) based on the trajectory in Fig. S9 to see possible EWSs. We filter possible nonstationarities of the trajectory by Gaussian filter, obtain the remaining residues (Fig. 5 *B*) for EWS analysis (66), and calculate four indicators as EWSs for EMT with a rolling window having a length of 10% (Fig. 5, *C–F*). From left to right, we show the properties of trajectory in different time periods. These four segments correspond to the four key transition stages (also see Fig. S9). We also obtained similar results for MET (Fig. S11), including stochastic trajectory of MET (Fig. S11 *A*), remaining residues (Fig. S11 *B*), and four indicators as EWSs for MET with a rolling window having a length of 10% (Fig. S11, *C–F*).

Here, we take ZEB2 as an example; the analyses of other genes are shown in Figs. S12 (EMT) and S13 (MET). Of note, we found that not all genes have EWSs. For example, we see that the variances of ZEB1, ZEB2, SNAI2, TWIST1, FOXC2, TCF3, and VIM increase significantly before critical points (Fig. S9 1), but not for TGF-β, SNAI1, GSC, and all six E genes (Fig. S9 2).

Previous study suggested that variance and temporal lag-1 autocorrelation (AR (1)) of the expression of some genes will increase before critical transitions (49). From our EMT network model, we focus on the four indicators for the first three stages (Fig. 5, *first three columns*) because the last stage corresponds to a single M state with no phase transition. We see that the AR (1) and variance of gene expression increase in EMT for four stages, individually (Fig. 5, *C and D*, the *first three columns*), but there is no such behavior for MET (Fig. S11, *C and D*, the *first three columns*). It seems that the variance and AR (1) are not accurate enough as EWSs in EMT and MET.

One possible reason for why the variance is not working well as the EWS here is because it does not involve the effects of the mean of gene expression, as a possible reason for increasing variance is actually the increasing magnitude of gene expressions. Based on Fig. S9, only M genes show obvious EWSs. So, we only focus on the trajectories of the M genes. From Eqs. S4–S6, we can see clearly that the intrinsic noise level (proportional to $\sqrt{\rho_i^+ + \rho_i^-}$) increases with x_j ($j \in A_i$) (first term in Eq. S4). Here, the ith gene and jth gene are both M genes (Fig. 1 shows that activation regulations only exist between M genes). From Fig. S9, during EMT the expressions of all M genes increase. So, the increase of the mean of gene expressions (M genes) contributes to the increase of noise, or the increase of variance of the trajectory. In other words, the increase of variance may be due to two factors, one related to the phase change and another one related to the change of the magnitude of gene expression. So, we also studied two other indicators: the coefficient of variance (CV) and the Fano factor (the ratio between variance and mean), which both involve the mean and the variance. With the CV or Fano factor, we can focus on the former factor (phase change) because our purpose is to identify the early warning signals related with critical transitions. We found that both the CV and Fano factor have better performance in terms of predicting critical transitions marked by their increase before critical points for EMT and MET (Fig. 5, *E and F*, the *first three columns*).

Specifically, the Fano factor increases before the cell transits to another stable state (Figs. 5 *F 1, 2, and 3* and S11 *F* 1), and it decreases when the cell tends to be stable (Figs. 5 *F 4* and S11 *F* 4). This is clear at the beginning and ending phases of the transition (stages 1 and 4 for both EMT and MET). However, this phenomenon is not that obvious in the transition between intermediate states (Figs. 5 *F 2* and S11 *F 2 and 3*). We suppose that this is because a segment of the trajectory contains oscillations between the two stable states (Figs. 5 3 and S11 2). These results demonstrate that the Fano factor is a better indicator as an EWS in EMT and MET, compared with the other three indicators studied here.

We further studied the mechanism of EWSs based on landscape theory. In Fig. 6 *A*, we show the landscapes under different parameter a with a typical stochastic trajectory (landscape for intrinsic noise). Clearly, the first, second, and fourth segments of trajectory are all for transitions from one stable state to the other (E to IE, IE to IM, and IM to M state). For the third segment, the trajectory oscillates between the IM state and the M state. These results imply that EWSs are related to the change of the stable states. Here, the variance of the stable states was calculated from stochastic simulations with intrinsic noise (see Supporting materials and methods). As is shown in Fig. 6 *B*, for all three stages (stages 1–3, representing the transition from one stable state to another) before the critical point, the Fano factor (*red line*) of the former stable state gradually increases. The trend of the Fano factor is similar to that in Fig. 5, which indicates that the theoretical trend of the Fano factor is consistent with the simulation results. We also acquired a similar conclusion in MET (Fig. S14).

Importantly, we identified the barrier height from the potential landscape as a new, to our knowledge, EWS and

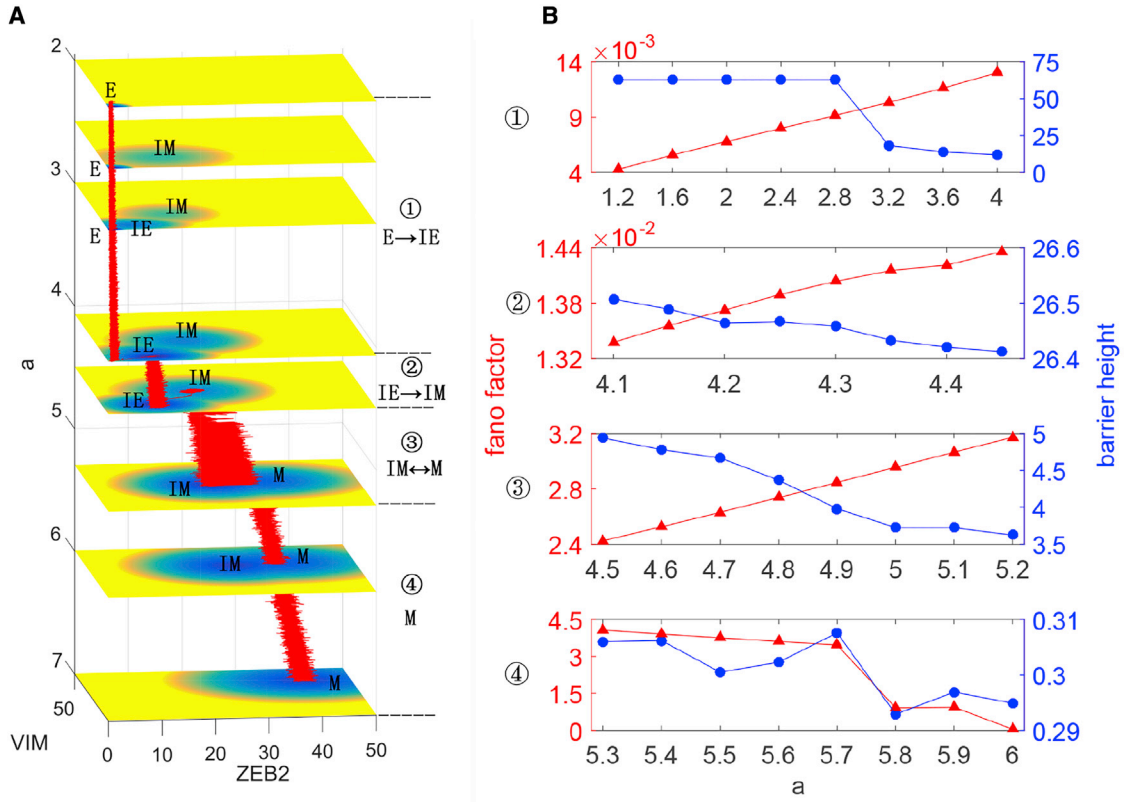

FIGURE 6 (*A*) A three-dimensional landscape of EMT under intrinsic noise with a typical stochastic trajectory from the E state to the M state. Different layers of two-dimensional landscape are for different values of parameter *a*. The red trajectory is a typical trajectory from E to M state from stochastic simulations. The red arrows represent the direction of the transitions. The parameter *a*, characterizing the activation signal of EMT, varies from 2 to 7 from top to down. The whole transition is divided into four stages with 1 $a = \sim$1.2–4.1, 2 $a = \sim$4.1–4.5, 3 $a = \sim$4.5–5.3, and 4 $a = \sim$5.3–6.0 (corresponding to the four stages in Fig. 5). (*B*) The Fano factor (σ^2/x, *red line*; here σ is the standard deviation of ZEB2 and x is the expression of ZEB2) of the E state (stage 1), IE state (stage 2), IM state (stage 3), and M state (stage 4) and the barrier height (*blue line*) from the E state to the IM state (stage 1, $a = \sim$1.2–3.0), from the E state to the IE state (stage 1, $a = \sim$3.0–4.1), from the IE state to the IM state (stage 2), from the IM state to the M state (stage 3), and from the M state to the IM state (stage 4). To see the figure in color, go online

found that the barrier height provides a better indicator for critical transitions than other indicators studied here, such as variance and AR (1). Here, the barrier height is defined as the potential difference between the local minimum (stable state) and neighboring saddle point for corresponding transitions. We found that the barrier height between stable states is decreasing for each stage (Fig. 6 *B*, *blue line*), which makes cells more likely to move to adjacent states. There is a significant decrease of barrier height at about $a = 3$ in the first stage as the landscape changes from the bistable to the tristable state. Because of the appearance of the IE state, the barrier height between the E state and the adjacent stable state is significantly reduced. That is the major reason for the upcoming transition from the E to the IE state. In the third stage of EMT, the magnitude of the barrier height between the IM state and the M state is very small, which explains why the trajectory oscillates in these two stable states (Fig. 6, stage 3). In the first half of this stage, the barrier is slightly larger, so the trajectory does not always oscillate between two stable states. For the second half of this stage, the trajectories oscillate more frequently as the barrier height becomes even smaller (Fig. 6 *A*, stage 3). For the fourth stage of the transitions, the magnitude of the barrier height becomes very small, indicating a stable M state. From a global point of view, the barrier height from the E state to the M state keeps decreasing, which indicates that the difficulty of the transition is decreasing. We also acquired a similar conclusion for MET (Fig. S14). These results explain why the barrier height is more accurate for predicting critical transitions as it characterizes the underlying transition dynamics of stable state switching.

These results also demonstrate that the global stability, rather than local stability, is more important to determine the system behavior. If we analyzed some local indicator for stability of the system, e.g., the real part of the dominant eigenvalue of Jacobian matrix of $F(x)$ at an individual stable state, we found that the local stability does not give consistent results for predicting critical transition for all four stages as the barrier height does (Fig. S15). To sum up, the results of landscape reveal the underlying and physical mechanism for the emergence of EWSs, and the change of

barrier height, as a new, to our knowledge, EWS, offers a quantitative explanation for the appearance of phase transition and the change of feasibility of transition between different stable states.

Gene classification based on intermediate states

During the EMT, different genes have different expression patterns. Here, we choose 200 cases (parameter sets) with more than three stable states to observe the rough paths for each gene against parameter perturbations (Fig. 7). We connect all adjacent stable states as the rough paths to interrogate the properties of different genes. Fig. 7 *A* shows the paths of all 16 genes, which are classified to three classes, and Fig. 7 *B* shows three typical average transition paths for three types of genes, individually.

Based on the results for parameter perturbations, the 16 genes in the network can be divided into three categories: 1) SNAI2 has an irregular path, probably because of its self-activation. Although ZEB1 and ZEB2 have self-activation, too, their changing trends are different from SNAI2. 2) The changes of transition paths for ZEB2, VIM, miR-200, and CDH1 are smoother, as indicated by the color in Fig. 7 *A*, probably because these genes are affected by more other genes in the network. As shown in Fig. 1, the second type of gene is activated or inhibited by at least four genes, whereas other genes are affected by at most three genes (except ZEB1 and SNAI2). According to Eq. 12, it is reasonable to consider that genes regulated by more other genes may change more smoothly because of the averaging effects from other genes. 3) Different from above two cases, the expression of the other 10 genes displays a sigmoidal shape of transition path, and they keep the highest expression (*yellow*) or lowest expression (*blue*) most of the time during the transition process. This indicates that these genes will not change their expression much at the beginning and the end of EMT, and their expression will change considerably between intermediate states. Furthermore, the expression of M genes in this class (TGF-β, ZEB1, SNAI1, TWIST1, FOXC2, and TCF3) increases and the expression of E genes in this class (miR-145, miR-141, miR-34a, and Ovol2) decreases almost at the same time because their patterns in Fig. 7 *A* are similar.

In summary, we classify the genes based on their changing trend during EMT. These results may help us to design regulatory strategies: compared with type 1 and type 2 genes, the type 3 genes seem more sensitive to perturbations because they have a very short transition time in the middle. So, it may be less effective to target them at the beginning or the end of EMT. Because EMT has been suggested to promote cancer metastasis (14), a better anti-metastasis strategy could be targeting them in the middle of EMT, e.g., targeting them in the intermediate state stage. This is consistent with recent work suggesting that intermediate states are critical to cancer metastasis (67,68). We also calculated the location of intermediate states in different conditions (Fig. S16). For

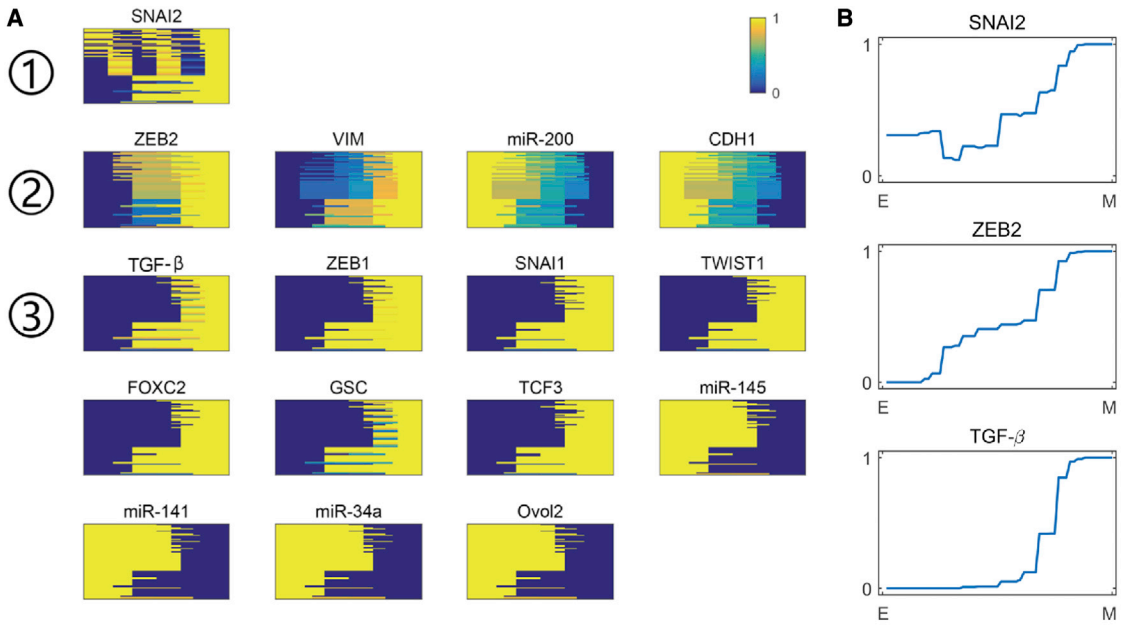

FIGURE 7 (*A*) Classification of the genes through their changing trends during EMT. The *X* axis represents the virtual time from E to intermediate, and to M state. The *Y* axis represents 100 different samples. Each line represents a sample. The yellow region represents higher expression, and the blue region represents lower expression. The 16 genes are divided into three types. (*B*) Average of the changing curves of each type of genes. The normalized expressions of these three genes during EMT are shown. The *X* axis, from left to right, represents the virtual time from E to intermediate state and to M state. The *Y* axis represents the normalized gene expression. The first type of gene (SNAI2) has obvious oscillations. The second type of gene (ZEB2) changes gradually. The third type of gene exhibits sigmoidal change during the transition. To see the figure in color, go online

most cases, the stable states exist on both a positive axis and negative axis, which indicates that if there are more than two intermediate states, most likely the IE state and IM state will exist at the same time, reflecting the heterogeneity for the intermediate states in EMT.

Identify influential factors for EMT and MET from global sensitivity analysis

To regulate EMT, a key issue is how to identify critical factors for EMT and MET. In the gene regulatory network model, it means identifying the critical parameters which determine the EMT. For example, it is important to ask how to stabilize the E state and destabilize the M state by targeting certain genes or regulations between genes. One natural way to accomplish these two missions is to perform a global sensitivity analysis for the transition actions from E attractor to M attractor (or from M attractor to E attractor). Sensitivity analysis can show the influence of different parameters on the system, and transition actions can measure the possibility of transition from one state to another for a cell.

For simplicity, we constrain the system in the bistable region (the E state and the M state coexisting, no intermediate state). From the definition of transition action, the cell needs more energy to overcome the barrier to transform to the other state when the transition action is larger. The sensitivity analysis here is based on calculating the change of the transition action, individually for E to M and M to E, after changing the model parameters one by one. To inhibit EMT, we tend to make the transition action of EMT become larger, namely the cell needs more energy to transform to the M state, after changing the parameters. Here, we choose each factor of $K = \{k_i\}$ (the degradation rate of genes) to perform the sensitivity analysis. The result is shown in Fig. 8.

The simulation results show that the regulation of gene expression has different effects on EMT and MET. Increasing k_E (the innate degradation rate of an E gene, Fig. 8 *A*, the *top six lines*) does not necessarily make it easier for cells to transform to M state. On the other hand, decreasing k_M does not necessarily make it easier for cells to transform to the E state. Thus, we focus on the genes and regulations with significant effects on state transitions next.

As shown in Fig. 8, the top five sensitive regulations are TCF3, ZEB1, ZEB2, Ovol2, and SNAI2. In previous work (5–14,16,69,70), we know SNAI1, SNAI2, ZEB1, ZEB2, and TCF3 are the most essential regulators of EMT. In this work, we found that TCF3 is the strongest single-factor activator of EMT in some cases, which is partially supported by some experimental work (71). Interestingly, less expression of TCF3 (larger degradation rate, Fig. 8 *A*) will not influence the transition much, but more expression of TCF3 (Fig. 8 *B*) will greatly increase the action required for EMT. It is anticipated that more experiments will focus on TCF3 to test its roles in regulating EMT in the future. Meanwhile, the ZEB family (including ZEB1 and ZEB2) also plays a key role of the transition, as suggested by previous experiments (71,72). Once the expression of ZEB1 decreases (Fig. 8 *A*), EMT will become harder, whereas MET will become easier. On the contrary, if the expression of ZEB1 increases (Fig. 8 *B*), EMT will be easier, whereas MET will be harder. Further, Ovol2 also plays an important role in EMT (12). If the expression of Ovol2 decreases (Fig. 8 *A*), $S_{E\to M}$ will decreases a lot, which means the EMT will become easier.

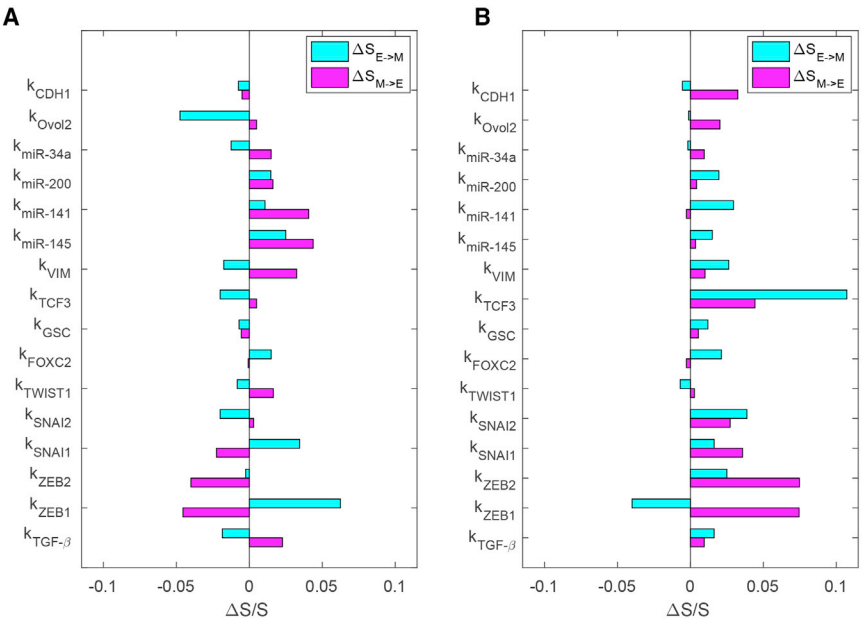

FIGURE 8 Sensitivity analysis for parameter $K = \{k_i\}$, the degradation rate of each gene, on the transition action ($S_{E\to M}$ and $S_{M\to E}$). The Y axis represents k_i of 16 genes, individually. The X axis represents the percentage of the change of the transition action (ΔS) relative to action with default parameters (S), namely the sensitivity of the genes ($\Delta S/S$). Here, $S_{E\to M}$ denotes the transition action from the E state to the M state, and $S_{M\to E}$ denotes the transition action from the M state to the E state. (*A*) Each k_i is increased by 5% individually. (*B*) Each k_i is decreased by 5% individually. Here, we fix the perturbation level to be 5% to avoid phase transition of the model. Other parameters are set as $a = 1$, $\tilde{a} = 3$, $b = 3$, $s_a = 3$, and $s_b = 0.5$. To see the figure in color, go online

The EMT system also shows strong sensitivity to SNAI2. When the expression of SNAI2 increases (Fig. 8 *B*), it becomes more difficult for both EMT and MET, possibly because of the role of SNAIL2 on stabilizing an intermediate EMT phenotype (73). Similarly, TCF3, ZEB2, SNAIL1, and TGF-β also show this property. From the network topology, ZEB2 and SNAI2 both have self-activation. As suggested by previous work, strong self-activation will promote more stable intermediate states (74). This indicates that ZEB2 and SNAI2 might make EMT and MET both harder by promoting stable intermediate states. For microRNAs (miRNAs), miR-145 and miR-141 show strong sensitivities from our model. If their expression decreases, MET will be remarkably harder (Fig. 8 *A*), which is supported by the experiments (75,76).

DISCUSSION

EMT is an important phenotypic switching (1–4,77). Mathematical modeling has been applied to study the dynamical properties for EMT. However, the molecular mechanisms of EMT have not been fully understood. In this work, we constructed a relatively inclusive gene regulatory network for EMT including 16 genes. By screening the parameters, we modeled the EMT in different conditions and showed that intermediate states exist in a wide range of parameter regions, which is supported by experiments (31). Based on the stochastic dynamics theory of gene networks, we revealed the landscape of EMT, which quantifies the global stability and stochastic transition dynamics of EMT. We identified multistable cell states characterized by the attractors on the landscape. We calculated the kinetic transition path of EMT based on the path-integral approach to identify the order of gene switching in EMT and studied the effects of noise on the transition path (22). Larger noise, leading to larger probability flux (24,65), makes the system deviate from the path with the lowest potential energy. We demonstrated that the 16-dimensional transition paths of EMT from the model are supported by experimental data.

In the study of EMT, an interesting question is what the roles of intermediate states are. Recent work proposed a role of intermediate state for accelerating EMT from a simplified state transition model (32). Whether this conclusion holds from a molecular network perspective remains unsolved. Here, by screening multiple parameter values, we calculated the distribution of transition actions for EMT to study the relationship between the number of intermediate states (N_{IM}) and the rate of the transition. Roughly speaking, as N_{IM} increases, the mean transition action first decreases, and after a certain N_{IM}, the mean transition action begins to increase. That means intermediate states in certain condition can accelerate EMT, which agrees with a recent study (32). As the number of intermediate states N_{IM} increases, the energy landscape becomes more complex. Besides the number of intermediate states, the depth of basins for intermediate states would also influence the rate of transition, and the rate changing effect becomes limited for large numbers of intermediates (78). However, the optimal number of intermediate states is different under different parameter regions, which suggests that the role of intermediate states interacting with the local microenvironment would make a difference to the transition path. These predictions can be verified by further experiments.

In gene regulatory network modeling, a critical issue is the selection of parameters. In this work, we performed the parameter selection by first setting the parameters randomly and independently in reasonable regions and then searching for the multistability. So, in some sense we considered the possible collinearity among these parameters to ensure that the dynamical system generates multistability. In other words, we actually selected the parameters as an ensemble (79).

The EWS theory has been applied to EMT study to anticipate the critical transition points recently (49). Here, we studied a few indicators for critical transitions from our EMT model. We verified the existence of EWSs in EMT through our complex molecular network model and showed that some genes, including ZEB1, ZEB2, SNAI2, TWIST1, FOXC2, TCF3, and VIM, have obvious EWSs at multiple stages, whereas the other genes do not display such behavior obviously. Interestingly, we demonstrated that the Fano factor could be a better indicator as an EWS compared with traditionally used ones, although it does not perform as well as the potential barrier height. We proposed that the potential barrier height based on the landscape topography provides a new, to our knowledge, indicator for predicting critical transitions, which performs the best compared with other indicators including the Fano factor, the CV, variance, etc. This is reasonable because the barrier height is a global quantity based on the potential landscape, whereas other indicators here are all local measures for certain variables. So, the potential barrier height can characterize the dynamical behavior of the system more accurately. The landscape not only provides a new, to our knowledge, indicator for EWS with better accuracy but also offers an explanation for the underlying mechanism of the well-known EWS theory. One limitation of barrier height as an EWS is that the current approach for obtaining the barrier height depends on the quantified landscape, which requires a dynamical model as a prior and is not a completely data-driven approach. This could be further improved by absorbing the data-driven landscape construction approaches (80,81).

The concept of barrier height based on landscape theory has been proposed to predict transitions in a multistable system (9–11,13–15,22–30,54). It is understandable that barrier height from potential landscape may be related to the early warning theory of critical transitions. However, to our

knowledge the underlying physical mechanism for EWSs remains to be clarified quantitatively. For example, different indicators based on time-series data have been suggested as EWSs to predict critical transitions, including variance, autocorrelation function, coefficient of variation, etc. (41–48). Recently, Sarker et al. proposed some indicators for EWSs of EMT (49). However, the underlying physical mechanism for why these EWSs can work remains unclear. In this work, based on landscape theory, we propose barrier height as a new global measure for EWSs, and provide a quantitative explanation for the mechanism of EWS. An interesting question is what the difference between EMT and other multistable systems is. From current work, the EMT does not show obvious differences with other multistable systems because the multistable property mostly comes from the network structure with positive feedback loops. A possible difference between EMT and other multistable systems is the common existence of intermediate states in EMT, as the partial EMT states or hybrid E/M states have been extensively explored both in modeling and experimental studies (10,16,17,67). Another possible difference between EMT and other multistable systems is that we found that intermediate states promote E to M transitions. Whether this conclusion holds for other multistable system warrants further explorations.

Through classifying the genes by their trend of transitions in each step, we propose their unique regulation strategy for each type of gene. The strategy may provide a route for regulating EMT more efficiently. By global sensitivity analysis of each gene on model parameters, we identified some genes playing critical roles on regulating EMT or MET. For example, besides the SNAIL family and the ZEB family, our models predict that TCF3 has a significant regulatory influence on EMT.

In this work, we quantified the global stability and stochastic transition dynamics of EMT by landscape and transition path theory. We mapped out a landscape with multiple intermediate states for EMT and proposed possible roles of intermediate states for accelerating EMT. We provide a physical and quantitative explanation for the mechanism of the well-known EWS theory and propose the potential barrier height as a new, to our knowledge, indicator for critical transitions. Our approaches can be applied to other dynamical systems or gene regulatory systems.

SUPPORTING MATERIAL

Supporting material can be found online at https://doi.org/10.1016/j.bpj.2021.08.043.

AUTHOR CONTRIBUTIONS

C.L. designed research. J.L. and C.L. performed research. J.L., Q.N., and C.L. analyzed data. J.L., Q.N., and C.L. wrote the manuscript.

ACKNOWLEDGMENTS

C.L. thanks Professor Luonan Chen for valuable discussions.

C.L. is supported by the National Key R&D Program of China (2019YFA0709502) and the National Natural Science Foundation of China (11771098). Q.N. is partly supported by a National Science Foundation grant DMS1763272 and a grant from the Simons Foundation (594598).

REFERENCES

1. Thiery, J. P. 2002. Epithelial-mesenchymal transitions in tumour progression. *Nat. Rev. Cancer.* 2:442–454.
2. Kalluri, R., and R. A. Weinberg. 2009. The basics of epithelial-mesenchymal transition. *J. Clin. Invest.* 119:1420–1428.
3. Rastaldi, M. P., F. Ferrario, …, G. D'Amico. 2002. Epithelial-mesenchymal transition of tubular epithelial cells in human renal biopsies. *Kidney Int.* 62:137–146.
4. Grünert, S., M. Jechlinger, and H. Beug. 2003. Diverse cellular and molecular mechanisms contribute to epithelial plasticity and metastasis. *Nat. Rev. Mol. Cell Biol.* 4:657–665.
5. Steinway, S. N., J. G. Zañudo, …, R. Albert. 2014. Network modeling of TGFβ signaling in hepatocellular carcinoma epithelial-to-mesenchymal transition reveals joint sonic hedgehog and Wnt pathway activation. *Cancer Res.* 74:5963–5977.
6. Cohen, D. P. A., L. Martignetti, …, L. Calzone. 2015. Mathematical modelling of molecular pathways enabling tumour cell invasion and migration. *PLoS Comput. Biol.* 11:e1004571.
7. Khan, F. M., S. Marquardt, …, B. M. Pützer. 2017. Unraveling a tumor type-specific regulatory core underlying E2F1-mediated epithelial-mesenchymal transition to predict receptor protein signatures. *Nat. Commun.* 8:198.
8. Méndez-López, L. F., J. Davila-Velderrain, …, E. R. Alvarez-Buylla. 2017. Gene regulatory network underlying the immortalization of epithelial cells. *BMC Syst. Biol.* 11:24.
9. Li, C., T. Hong, and Q. Nie. 2016. Quantifying the landscape and kinetic paths for epithelial-mesenchymal transition from a core circuit. *Phys. Chem. Chem. Phys.* 18:17949–17956.
10. Lu, M., M. K. Jolly, …, E. Ben-Jacob. 2013. MicroRNA-based regulation of epithelial-hybrid-mesenchymal fate determination. *Proc. Natl. Acad. Sci. USA.* 110:18144–18149.
11. Huang, B., M. Lu, …, J. N. Onuchic. 2017. Interrogating the topological robustness of gene regulatory circuits by randomization. *PLoS Comput. Biol.* 13:e1005456.
12. Hong, T., K. Watanabe, …, X. Dai. 2015. An Ovol2-Zeb1 mutual inhibitory circuit governs bidirectional and multi-step transition between epithelial and mesenchymal states. *PLoS Comput. Biol.* 11:e1004569.
13. Li, C., and J. Wang. 2015. Quantifying the landscape for development and cancer from a core cancer stem cell circuit. *Cancer Res.* 75:2607–2618.
14. Li, C., and G. Balazsi. 2018. A landscape view on the interplay between EMT and cancer metastasis. *npj Syst. Biol. Appl.* 4:34.
15. Kang, X., and C. Li. 2021. A dimension reduction approach for energy landscape: identifying intermediate states in metabolism-EMT network. *Adv. Sci. (Weinh.).* 8:2003133.
16. Zhang, J., X. J. Tian, …, J. Xing. 2014. TGF-β-induced epithelial-to-mesenchymal transition proceeds through stepwise activation of multiple feedback loops. *Sci. Signal.* 7:ra91.
17. Tian, X. J., H. Zhang, and J. Xing. 2013. Coupled reversible and irreversible bistable switches underlying TGFβ-induced epithelial to mesenchymal transition. *Biophys. J.* 105:1079–1089.
18. MacLean, A. L., T. Hong, and Q. Nie. 2018. Exploring intermediate cell states through the lens of single cells. *Curr. Opin. Syst. Biol.* 9:32–41.

19. Sha, Y., D. Haensel, …, Q. Nie. 2019. Intermediate cell states in epithelial-to-mesenchymal transition. *Phys. Biol.* 16:021001.
20. Steinway, S. N., J. G. T. Zañudo, …, R. Albert. 2015. Combinatorial interventions inhibit TGFβ-driven epithelial-to-mesenchymal transition and support hybrid cellular phenotypes. *NPJ Syst. Biol. Appl.* 1:15014.
21. Waddington, C. H. 1957. The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology. Allen and Unwin, London.
22. Wang, J., K. Zhang, …, E. Wang. 2011. Quantifying the Waddington landscape and biological paths for development and differentiation. *Proc. Natl. Acad. Sci. USA.* 108:8257–8262.
23. Lv, C., X. Li, …, T. Li. 2015. Energy landscape reveals that the budding yeast cell cycle is a robust and adaptive multi-stage process. *PLoS Comput. Biol.* 11:e1004156.
24. Li, C., and J. Wang. 2014. Landscape and flux reveal a new global view and physical quantification of mammalian cell cycle. *Proc. Natl. Acad. Sci. USA.* 111:14130–14135.
25. Lu, M., J. Onuchic, and E. Ben-Jacob. 2014. Construction of an effective landscape for multistate genetic switches. *Phys. Rev. Lett.* 113:078102.
26. Ge, H., and H. Qian. 2012. Landscapes of non-gradient dynamics without detailed balance: stable limit cycles and multiple attractors. *Chaos.* 22:023140.
27. Ge, H., and H. Qian. 2016. Mesoscopic kinetic basis of macroscopic chemical thermodynamics: a mathematical theory. *Phys. Rev. E.* 94:052150.
28. Li, C., and J. Wang. 2013. Quantifying cell fate decisions for differentiation and reprogramming of a human stem cell network: landscape and biological paths. *PLoS Comput. Biol.* 9:e1003165.
29. Su, H., G. Wang, …, X. Zhu. 2017. Decoding early myelopoiesis from dynamics of core endogenous network. *Sci. China Life Sci.* 60:627–646.
30. Li, C. 2017. Identifying the optimal anticancer targets from the landscape of a cancer-immunity interaction network. *Phys. Chem. Chem. Phys.* 19:7642–7651.
31. Pastushenko, I., A. Brisebarre, …, C. Blanpain. 2018. Identification of the tumour transition states occurring during EMT. *Nature.* 556:463–468.
32. Goetz, H., J. R. Melendez-Alvarez, …, X. J. Tian. 2020. A plausible accelerating function of intermediate states in cancer metastasis. *PLoS Comput. Biol.* 16:e1007682.
33. May, R. M., S. A. Levin, and G. Sugihara. 2008. Complex systems: ecology for bankers. *Nature.* 451:893–895.
34. Scheffer, M., S. R. Carpenter, …, J. Vandermeer. 2012. Anticipating critical transitions. *Science.* 338:344–348.
35. Lenton, T. 2011. Early warning of climate tipping points. *Nature Clim Change.* 1:201–209.
36. Trefois, C., P. M. Antony, …, R. Balling. 2015. Critical transitions in chronic disease: transferring concepts from ecology to systems medicine. *Curr. Opin. Biotechnol.* 34:48–55.
37. Korolev, K. S., J. B. Xavier, and J. Gore. 2014. Turning ecology and evolution against cancer. *Nat. Rev. Cancer.* 14:371–380.
38. Scheffer, M., J. Elizabeth Bolhuis, …, M. G. M. Olde Rikkert. 2018. Quantifying resilience of humans and other animals. *Proc. Natl. Acad. Sci. USA.* 115:11883–11890.
39. Chen, L., R. Liu, …, K. Aihara. 2012. Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers. *Sci. Rep.* 2:342.
40. Jiang, Z., L. Lu, …, L. Chen. 2020. SMAD7 and SERPINE1 as novel dynamic network biomarkers detect and regulate the tipping point of TGF-beta induced EMT. *Sci. Bull. (Beijing).* 65:842–853.
41. Dai, L., D. Vorselen, …, J. Gore. 2012. Generic indicators for loss of resilience before a tipping point leading to population collapse. *Science.* 336:1175–1177.
42. Veraart, A. J., E. J. Faassen, …, M. Scheffer. 2011. Recovery rates reflect distance to a tipping point in a living system. *Nature.* 481:357–359.
43. Carpenter, S. R., W. A. Brock, …, M. L. Pace. 2008. Leading indicators of trophic cascades. *Ecol. Lett.* 11:128–138.
44. Guttal, V., and C. Jayaprakash. 2008. Changing skewness: an early warning signal of regime shifts in ecosystems. *Ecol. Lett.* 11:450–460.
45. Seekell, D. A., S. R. Carpenter, and M. L. Pace. 2011. Conditional heteroscedasticity as a leading indicator of ecological regime shifts. *Am. Nat.* 178:442–451.
46. Carpenter, S. R., J. J. Cole, …, B. Weidel. 2011. Early warnings of regime shifts: a whole-ecosystem experiment. *Science.* 332:1079–1082.
47. Boettiger, C., and A. Hastings. 2012. Quantifying limits to detection of early warning for critical transitions. *J. R. Soc. Interface.* 9:2527–2539.
48. Tirabassi, G., J. Viebahn, …, S. C. Dekker. 2014. Interaction network based early-warning indicators of vegetation transitions. *Ecol. Complex.* 19:148–157.
49. Sarkar, S., S. K. Sinha, …, P. S. Dutta. 2019. Anticipating critical transitions in epithelial-hybrid-mesenchymal cell-fate determination. *Proc. Natl. Acad. Sci. USA.* 116:26343–26352.
50. Zhang, B., and P. G. Wolynes. 2014. Stem cell differentiation as a many-body problem. *Proc. Natl. Acad. Sci. USA.* 111:10185–10190.
51. Sasai, M., and P. G. Wolynes. 2003. Stochastic gene expression as a many-body problem. *Proc. Natl. Acad. Sci. USA.* 100:2374–2379.
52. Hu, G. 1994. Stochastic Forces and Nonlinear Systems. Shanghai Scientific and Technological Education Publishing House, Shanghai, China.
53. Van Kampen, N. G. 2007. Stochastic Processes in Chemistry and Physics, Third Edition. North Holland, Amsterdam.
54. Wang, J., K. Zhang, and E. Wang. 2010. Kinetic paths, time scale, and underlying landscapes: a path integral framework to study global natures of nonequilibrium systems and networks. *J. Chem. Phys.* 133:125103.
55. Jolly, M. K., D. Jia, …, H. Levine. 2015. Coupling the modules of EMT and stemness: a tunable 'stemness window' model. *Oncotarget.* 6:25161–25174.
56. Sha, Y., S. Wang, …, Q. Nie. 2021. Inference of intercellular communications and multilayer gene-regulations of epithelial-mesenchymal transition from single-cell transcriptomic data. *Front. Genet.* 11:604585.
57. Wells, D. K., W. L. Kath, and A. E. Motter. 2015. Control of stochastic and induced switching in biophysical networks. *Phys. Rev. X.* 5:031036.
58. Zhou, X., W. Ren, and W. E. 2008. Adaptive minimum action method for the study of rare events. *J. Chem. Phys.* 128:104111.
59. Freidlin, M., and M. Weber. 2004. Random perturbations of dynamical systems and diffusion processes with conservation laws. *Probab. Theory Relat. Fields.* 128:441–466.
60. Abnaof, K., N. Mallela, …, H. Fröhlich. 2014. TGF-β stimulation in human and murine cells reveals commonly affected biological processes and pathways at transcription level. *BMC Syst. Biol.* 8:55.
61. Hari, K., B. Sabuwala, …, M. K. Jolly. 2020. Identifying inhibitors of epithelial-mesenchymal plasticity using a network topology-based approach. *NPJ Syst. Biol. Appl.* 6:15.
62. Celià-Terrassa, T., C. Bastian, …, Y. Kang. 2018. Hysteresis control of epithelial-mesenchymal transition dynamics conveys a distinct program with enhanced metastatic ability. *Nat. Commun.* 9:5005.
63. Swain, P. S., M. B. Elowitz, and E. D. Siggia. 2002. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. USA.* 99:12795–12800.
64. Kaern, M., T. C. Elston, …, J. J. Collins. 2005. Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.* 6:451–464.
65. Wang, J., L. Xu, and E. Wang. 2008. Potential landscape and flux framework of nonequilibrium networks: robustness, dissipation, and

coherence of biochemical oscillations. *Proc. Natl. Acad. Sci. USA.* 105:12271–12276.

66. Dakos, V., S. R. Carpenter, …, M. Scheffer. 2012. Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data. *PLoS One.* 7:e41010.
67. Jolly, M. K., M. Boareto, …, H. Levine. 2015. Implications of the hybrid epithelial/mesenchymal phenotype in metastasis. *Front. Oncol.* 5:155.
68. Tripathi, S., H. Levine, and M. K. Jolly. 2020. The physics of cellular decision making during epithelial-mesenchymal transition. *Annu. Rev. Biophys.* 49:1–18.
69. Udyavar, A. R., D. J. Wooten, …, V. Quaranta. 2017. Novel hybrid phenotype revealed in small cell lung cancer by a transcription factor network model that can explain tumor heterogeneity. *Cancer Res.* 77:1063–1074.
70. Wooten, D. J., and V. Quaranta. 2017. Mathematical models of cell phenotype regulation and reprogramming: make cancer cells sensitive again! *Biochim. Biophys. Acta Rev. Cancer.* 1867:167–175.
71. Cuevas, E. P., P. Eraso, …, F. Portillo. 2017. LOXL2 drives epithelial-mesenchymal transition via activation of IRE1-XBP1 signalling pathway. *Sci. Rep.* 7:44988.
72. Aigner, K., B. Dampier, …, A. Eger. 2007. The transcription factor ZEB1 (deltaEF1) promotes tumour cell dedifferentiation by repressing master regulators of epithelial polarity. *Oncogene.* 26:6979–6988.
73. Subbalakshmi, A. R., S. Sahoo, …, M. K. Jolly. 2021. A computational systems biology approach identifies SLUG as a mediator of partial epithelial-mesenchymal transition (EMT). *Cells Tissues Organs* 1–14, Published online February 10, 2021.
74. Li, C., L. Zhang, and Q. Nie. 2018. Landscape reveals critical network structures for sharpening gene expression boundaries. *BMC Syst. Biol.* 12:67.
75. Sachdeva, M., and Y. Y. Mo. 2010. miR-145-mediated suppression of cell growth, invasion and metastasis. *Am. J. Transl. Res.* 2:170–180.
76. Huang, Y., J. Tong, …, Z. Chen. 2015. miR-141 regulates TGF-β1-induced epithelial-mesenchymal transition through repression of HIPK2 expression in renal tubular epithelial cells. *Int. J. Mol. Med.* 35:311–318.
77. Font-Clos, F., S. Zapperi, and C. A. M. La Porta. 2018. Topography of epithelial-mesenchymal plasticity. *Proc. Natl. Acad. Sci. USA.* 115:5902–5907.
78. Wagner, C., and T. Kiefhaber. 1999. Intermediates can accelerate protein folding. *Proc. Natl. Acad. Sci. USA.* 96:6716–6721.
79. Gutenkunst, R. N., J. J. Waterfall, …, J. P. Sethna. 2007. Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput. Biol.* 3:1871–1878.
80. Chen, Z., S. An, …, L. Wan. 2019. DensityPath: an algorithm to visualize and reconstruct cell state-transition path on density landscape for single-cell RNA sequencing data. *Bioinformatics.* 35:2593–2601.
81. Guo, J., and J. Zheng. 2017. HopLand: single-cell pseudotime recovery using continuous Hopfield network-based modeling of Waddington's epigenetic landscape. *Bioinformatics.* 33:i102–i109.