

## RESEARCH ARTICLE

## A plausible accelerating function of intermediate states in cancer metastasis

Hanah Goetz<sup>1</sup>, Juan R. Melendez-Alvarez<sup>1</sup>, Luonan Chen<sup>2,3</sup>, Xiao-Jun Tian<sup>1\*</sup>

**1** School of Biological and Health Systems Engineering, Arizona State University, Tempe, Arizona, United States of America, **2** Key Laboratory of Systems Biology, Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, Shanghai, China, **3** Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, China

✉ These authors contributed equally to this work.

\* [Xiaojun.Tian@asu.edu](mailto:Xiaojun.Tian@asu.edu)



## Abstract

Epithelial-to-mesenchymal transition (EMT) is a fundamental cellular process and plays an essential role in development, tissue regeneration, and cancer metastasis. Interestingly, EMT is not a binary process but instead proceeds with multiple partial intermediate states. However, the functions of these intermediate states are not fully understood. Here, we focus on a general question about how the number of partial EMT states affects cell transformation. First, by fitting a hidden Markov model of EMT with experimental data, we propose a statistical mechanism for EMT in which many unobservable microstates may exist within one of the observable macrostates. Furthermore, we find that increasing the number of intermediate states can accelerate the EMT process and that adding parallel paths or transition layers may accelerate the process even further. Last, a stabilized intermediate state traps cells in one partial EMT state. This work advances our understanding of the dynamics and functions of EMT plasticity during cancer metastasis.

## OPEN ACCESS

**Citation:** Goetz H, Melendez-Alvarez JR, Chen L, Tian X-J (2020) A plausible accelerating function of intermediate states in cancer metastasis. *PLoS Comput Biol* 16(3): e1007682. <https://doi.org/10.1371/journal.pcbi.1007682>

**Editor:** Qing Nie, University of California Irvine, UNITED STATES

**Received:** October 20, 2019

**Accepted:** January 24, 2020

**Published:** March 10, 2020

**Copyright:** © 2020 Goetz et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the manuscript and its Supporting Information files.

**Funding:** This project was supported by the ASU School of Biological and Health Systems Engineering and NSF grant (EF-1921412) (to X-JT) ([https://www.nsf.gov/awardsearch/showAward?AWD\\_ID=1921412](https://www.nsf.gov/awardsearch/showAward?AWD_ID=1921412)). HG and JRM-A were also supported by the Arizona State University Dean's Fellowship. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Author summary

Epithelial-mesenchymal transition (EMT) is a basic biological process, in which epithelial cells undergo multiple biochemical changes, lose cell-cell junctions and polarization, and become a mesenchymal phenotype with migratory and invasive properties. Recent studies have illustrated the existence and importance of the partial EMT states. It has become increasingly apparent that the EMT has strong differentiation plasticity. This plasticity is heavily implicated in cancer cell invasion and metastasis. However, it is still unclear how the number of intermediate states changes the EMT process. Here, we use a hidden Markov model to describe the EMT process. By fitting with the experimental data, we find that unobservable microstates exist within the observable macrostates: epithelial, partial EMT, and mesenchymal. Additionally, we find that increasing the number of states between the start and end of EMT or including alternative transition avenues via parallel paths or transition layers can accelerate the EMT process. This study suggests a non-trivial function of the EMT plasticity during cancer metastasis.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

Epithelial-to-mesenchymal transition (EMT) is a fundamental cellular process in which polarized epithelial cells lose various cell-cell junctions and adhesion and gain migratory and invasive properties to become mesenchymal cells [1, 2]. EMT is very important in embryonic development, tumorigenesis, metastasis, tumor stemness, and therapy resistance [3, 4]. Remarkably, EMT is not a binary process but instead proceeds with multiple partial intermediate states, collectively known as partial or hybrid EMT states [3, 5–11]. The partial EMT state retains some characteristics of epithelium but also shows features of mesenchymal cells [12–14]. One partial EMT state was predicted through mathematical modeling of the EMT core regulatory network and was verified with quantitative experiments by our previous works [5, 6]. Thereafter, many different partial EMT states were proposed [8, 9, 15–17]. More and more experimental data shows a different number of partial EMT states in various cancer cell lines [18–23]. Recently, several partial EMT phenotypes were found during cancer metastasis *in vivo* in a skin cancer mouse model [24, 25] and prostate cancer [26]. While many partial EMT states have been found, their functions are still not fully understood during cancer metastasis [4, 27–29].

Currently, the function of partial EMT states has been studied in the context of coupling with other cellular processes. For example, acquisition of stem-like properties dictates its coupling with cancer stemness [11, 30–34], circulating tumor cells (CTCs) [35, 36], and drug resistance [37]. Thus, the partial EMT cells hold the highest metastatic potential. Instead of full EMT, partial EMT is found to be critical for renal fibrosis [38–40]. There are many potential couplings of partial EMT and other biological processes, such as cell cycle [40], renal fibrosis [41] and metabolisms [42]. The cross-talk among the regulators of EMT and other cellular processes guides the coupling mechanism and full functional characteristics of partial EMT states.

While it is important to investigate specific functions of each partial EMT state in the context of additional cellular processes, an interesting question about the general function is whether the number of partial EMT states affect the transition itself. Given that different cell lines may have different partial EMT states, epithelial cells in different systems may also undergo different steps to become mesenchymal cells. Here, we use a general hidden Markov model to describe cell fate transitions during the EMT process. Our analysis makes several non-intuitive predictions. First, several microstates exist within one macrostate, which makes EMT a non-Markov process. Second, increasing the number of intermediate states can accelerate the EMT process. Third, transition in parallel and layer modes can further accelerate the process. Lastly, the existence of a stabilized intermediate state traps cells within their current phenotype. We conclude with discussions on how the dynamics and functions of EMT plasticity are controlled by the number of states during cancer metastasis.

## Materials and methods

### Transition rate and energy barrier

In this work, we assume that the total energy barrier one epithelial cell has to cross to transition to the mesenchymal state is  $\Delta E$ . If there is no intermediate state, then following the Arrhenius equation, the transition rate of EMT is  $k = k_0 e^{-\Delta E/k_{BT}}$ , where the value of  $k_0$  is a pre-exponential factor, which determines the time scale of the system. For simplicity, we define the energy unit as  $k_{BT}$ , thus we can simplify the equation to  $k = k_0 e^{-\Delta E}$ . Suppose there are  $N_{int}$  intermediate states between the epithelial and mesenchymal states and that the energy is divided into  $N_{int} + 1$  steps; the transition rate for step  $i$  is  $k_i = k_0 e^{-\delta E_i}$  where  $\delta E_i$  is the energy barrier for this step and  $\sum \delta E_i = \Delta E$ . We considered two scenarios:

1. The energy is divided evenly into all the steps, and thus the transition rate for each step is same:  $k_i = k_0 e^{-\Delta E / (N_{int} + 1)}$ .
2. There is one stabilized intermediate state where the energy barrier is  $\delta E_s = m \delta E_i$ . We considered case one:  $\delta E_s + \sum_{i \neq s} \delta E_i = \Delta E$ . The transition rate is  $k_i = k_0 e^{-\frac{\Delta E}{(N_{int} + m)}}$  for the transition from the regular intermediate states, and  $k_s = k_0 e^{-\frac{\Delta E}{(N_{int} + m)}}$  for the transition from the stabilized intermediate state, or case two: the transition rate is  $k_i = k_0 e^{-\frac{\Delta E}{(N_{int} + 1)}}$  for the transition from the regular intermediate states (same as scenario 1), and  $k_s = k_0 e^{-\frac{\Delta E}{(N_{int} + 1)}}$  for the transition from the stabilized intermediate state.

### Estimation of $k_0$ and the energy barrier $\Delta E$

The total energy barrier from epithelial to mesenchymal,  $\Delta E$  and  $k_0$ , were estimated by the following equations:

$$k_{fit} = k_0 e^{-\Delta E / (N_{fit} + 1)}$$

where  $N_{fit} = 9$  is the number of intermediate states and the transition rate  $k_{fit} = 3.4261$  is from the best fitting of the experimental data. For each  $\Delta E$ , we can calculate  $k_0$  and the transition rates for all cases.

### Parameter fitting

**Based on one intermediate state and an irreversible EMT process.** First, we considered one intermediate state in the model to fit the parameters. We assumed EMT to be an irreversible process under high dose of inducer TGF- $\beta$ .  $\vec{p}(t) = [p_E(t), p_P(t), p_M(t)]$  is the vector that represents the probability of a cell belonging in each state during the process of EMT. Thus, the equation governing the dynamics of  $\vec{p}$  is

$$\frac{d}{dt} \vec{p} = \begin{bmatrix} -k_1 & 0 & 0 \\ k_1 & -k_2 & 0 \\ 0 & k_2 & 0 \end{bmatrix} \vec{p}$$

where  $k_1$  is the transition rate from the epithelial state to the pEMT state, and  $k_2$  is the transition rate from the pEMT state to the mesenchymal state.

The equations are solved and the solution is:

$$p_E(t) = e^{-k_1 t},$$

$$p_P(t) = \frac{k_1}{k_2 - k_1} e^{-k_1 t} - \frac{k_1}{k_2 - k_1} e^{-k_2 t},$$

$$p_M(t) = 1 - \frac{k_2}{k_2 - k_1} e^{-k_1 t} + \frac{k_1}{k_2 - k_1} e^{-k_2 t}.$$

**Based on a flexible number of intermediate states and an irreversible EMT process.**

The above model does not accurately describe the experimental data. Thus, we extended the model by considering  $N_{int} + 1$  microstates; this is based on an assumption that each

macrostate, including the epithelial state and pEMT state, consist of multiple microstates. Suppose that EMT is an irreversible process where the cells transition independently from one microstate to the next at the same rate  $k$ . Based on the assumption that the  $\Delta E$  is evenly divided into these steps, the following equations can be used to represent the dynamics of the probability of the microstates:

$$\frac{d}{dt}\vec{p} = \begin{bmatrix} -k & 0 & 0 & \dots & 0 & 0 & 0 \\ k & -k & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & k & -k & 0 \\ 0 & 0 & 0 & \dots & 0 & k & 0 \end{bmatrix} \vec{p}.$$

The solution of this equation is:

$$p_n(t) = \frac{1}{(n-1)!} e^{-kt} k^{n-1} t^{n-1}, (1 \leq n \leq N_{int} + 1),$$

$$p_M(t) = 1 - \sum_{n=1}^{N_{int}+1} p_n(t).$$

**Based on a flexible number of intermediate states and a reversible EMT process.** EMT has some probability to be reversible even under high dose of inducer TGF- $\beta$ , which may result from the noise in the cell or other stochasticity. We also considered a case where the EMT process is reversible with the cells transitioning independently between one microstate and the next at the forward rate  $k_1$  and backward rate  $k_{-1}$ . The following equations can be used to represent the dynamics of the probability of the microstates:

$$\frac{d}{dt}\vec{p} = \begin{bmatrix} -k_1 & k_{-1} & 0 & \dots & 0 & 0 & 0 \\ k_1 & -k_{-1} - k_1 & k_{-1} & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & k_1 & -k_1 - k_{-1} & k_{-1} \\ 0 & 0 & 0 & \dots & 0 & k_1 & -k_{-1} \end{bmatrix} \vec{p}.$$

**First arrival time**

The first arrival time (FAT) distribution is used to quantify the time cells take to be transitioned into the mesenchymal state or stabilized state in various scenarios. The FAT distribution is  $f_M(t) = \frac{d}{dt}p_M(t)$  with normalization ( $\int_0^\infty f_M dt = 1$ ).

**For the one-path scenario.** When there is a total of  $N_{int}$  intermediate states, the probability of a cell being in the  $n$ -th intermediate state is given by:

$$p_n(t) = \frac{1}{(n-1)!} e^{-kt} k^{n-1} t^{n-1}, (1 \leq n \leq N_{int} + 1),$$

$$f_M(t) = \frac{1}{N_{int}!} e^{-kt} k^{N_{int}+1} t^{N_{int}}.$$

The corresponding mean first arrival time (MFAT) to the mesenchymal state is

$$\langle f_M \rangle = \int_0^\infty (f_M \cdot t) dt = \frac{N_{int} + 1}{k} = \frac{N_{int} + 1}{k_0 e^{-\Delta E/(N_{int}+1)}}$$

which indicates that the MFAT non-linearly depends on  $N_{int}$ .

**For the parallel path scenario.** The probability of a cell in the  $n_i$ -th intermediate state of the  $i$ -th path is given by:

$$p_{n_i}(t) = \frac{k_i^{n_i}}{(k_i - k_A)^{n_i}} \cdot \left( e^{-k_A t} - e^{-k_i t} \sum_{j=0}^{n_i-1} \left( \frac{((k_i - k_A)t)^j}{j!} \right) \right).$$

The FAT distribution is given by:

$$f_M(t) = \frac{dp_M}{dt} = \sum_{i=1}^{N_{pth}} k_i p_{N_i},$$

where  $N_{pth}$  is the number of paths and  $N_i$  is the number of intermediate states in the  $i$ -th path.

The MFAT to the mesenchymal state is

$$\sum \frac{k_i + N_i \cdot k_A}{k_A^2},$$

where  $k_A = \sum k_i$  and  $k_i$  is transition rate in the  $i$ -th path.

**For the layered scenario.** The probability of a cell in the  $i$ -th layer is given by:

$$p_{L_i}(t) = N_{L1} \cdot \left( \frac{e^{-N_{L1}kt}}{(1 - N_{L1})^i} - e^{-kt} \sum_{j=1}^i \frac{(kt)^{i-j}}{(i-j)!(1 - N_{L1})^j} \right),$$

and the FAT distribution is given by:

$$f_M(t) = N_{L1} \cdot k \cdot \left( \frac{e^{-N_{L1}kt}}{(1 - N_{L1})^{N_{Ly}}} - e^{-kt} \sum_{j=1}^{N_{Ly}} \frac{(kt)^{N_{Ly}-j}}{(N_{Ly}-j)!(1 - N_{L1})^j} \right),$$

and the MFAT to the mesenchymal state is given by:

$$\frac{N_{Ly} + 1/N_{L1}}{k},$$

where  $N_{L1}$  is the number of states in the first layer,  $N_{Ly}$  is the number of layers, and  $k$  is the transition rate from one intermediate state to the next (the same for all steps in the layered case).

**The cases with one stabilized state**

For any case that has one stabilized state at  $S$ , such that  $1 < S < N_{int} + 1$ , the probability of a cell in each state goes as follows:

$$\begin{aligned}
 P_E(t) &= e^{-k_1 t} \\
 P_1 &= k_1 t e^{-k_1 t} \\
 &\vdots \\
 P_{S-1} &= \frac{k_1^{S-1}}{(S-1)!} t^{S-1} e^{-k_1 t} \\
 P_S &= \frac{k_1^S}{(k_1 - k_2)^S} e^{-k_2 t} - \sum_{i=1}^S \frac{k_1^S t^{i-1}}{(k_1 - k_2)^{S+1-i} (i-1)!} e^{-k_1 t} \\
 P_{S+1} &= \frac{k_1^{S+1} k_2}{(k_1 - k_2)^{S+1}} e^{-k_2 t} - \sum_{i=1}^{S+1} \frac{k_1^{S+1} k_2 t^{i-1}}{(k_1 - k_2)^{S+2-i} (i-1)!} e^{-k_1 t} \\
 &\vdots \\
 P_N &= \frac{k_1^N k_2}{(k_1 - k_2)^N} e^{-k_2 t} - \sum_{i=1}^N \frac{k_1^N k_2 t^{i-1}}{(k_1 - k_2)^{N+1-i} (i-1)!} e^{-k_1 t} \\
 P_M(t) &= 1 - \left( \sum_{i=1}^N P_{P_i}(t) \right) - P_E(t).
 \end{aligned}$$

Here,  $k_1$  denotes the transition rate from the non-stabilized state and  $k_2$  the transition from the stabilized intermediate state. In order for the state to be stabilized as described, it must be true that  $k_2 < k_1$ .

**The MFAT to the stabilized state.** The MFAT to the stabilized state is

$$\frac{S + 1}{k_1}.$$

Thus, the acceleration effect of increasing the number of intermediate states or adding paths or layers also applies to the stabilized state.

**Mean dwelling time.** The mean dwelling time distribution is used to quantify the time cells spend within the stabilized intermediate state for various cases. The ODE of the probability of cells in the stabilized intermediate state is

$$P'_S = k_1 P_{S-1} - k_2 P_S$$

where  $k_1 P_{S-1}$  represents the transition into the stabilized state and  $k_2 P_S$  represents the transition out. The mean dwelling time ( $D_S(t)$ ) is defined as the average time that cells spend within this intermediate state, which only depends on  $k_2$ .  $D_S(t)$  is as follows:

$$D_S(t) = \frac{\int_0^\infty t e^{-k_2 t}}{\int_0^\infty e^{-k_2 t}} dt = \frac{1}{k_2}.$$

**Stochastic model**

To simulate the stochasticity of the cell state transitions during EMT at the single-cell level, we also developed a stochastic model using the Gillespie algorithm.

1. Define the system according to the cell state transitions during EMT, using the following matrices:  $S$  is the cell state the transition comes from,  $P$  is the cell state that the transition is to, and vector  $K$  denotes the transition rates. For example, for the system with only one path and  $N_{int}$  intermediate states,  $S$  and  $P$  have  $N_{int} + 2$  columns and  $N_{int} + 1$  rows while  $K$  has one column and  $N_{int} + 1$  rows, with the pattern:

$$S = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix}, \quad P = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix}, \quad K = \begin{bmatrix} k \\ k \\ \vdots \\ k \end{bmatrix}.$$

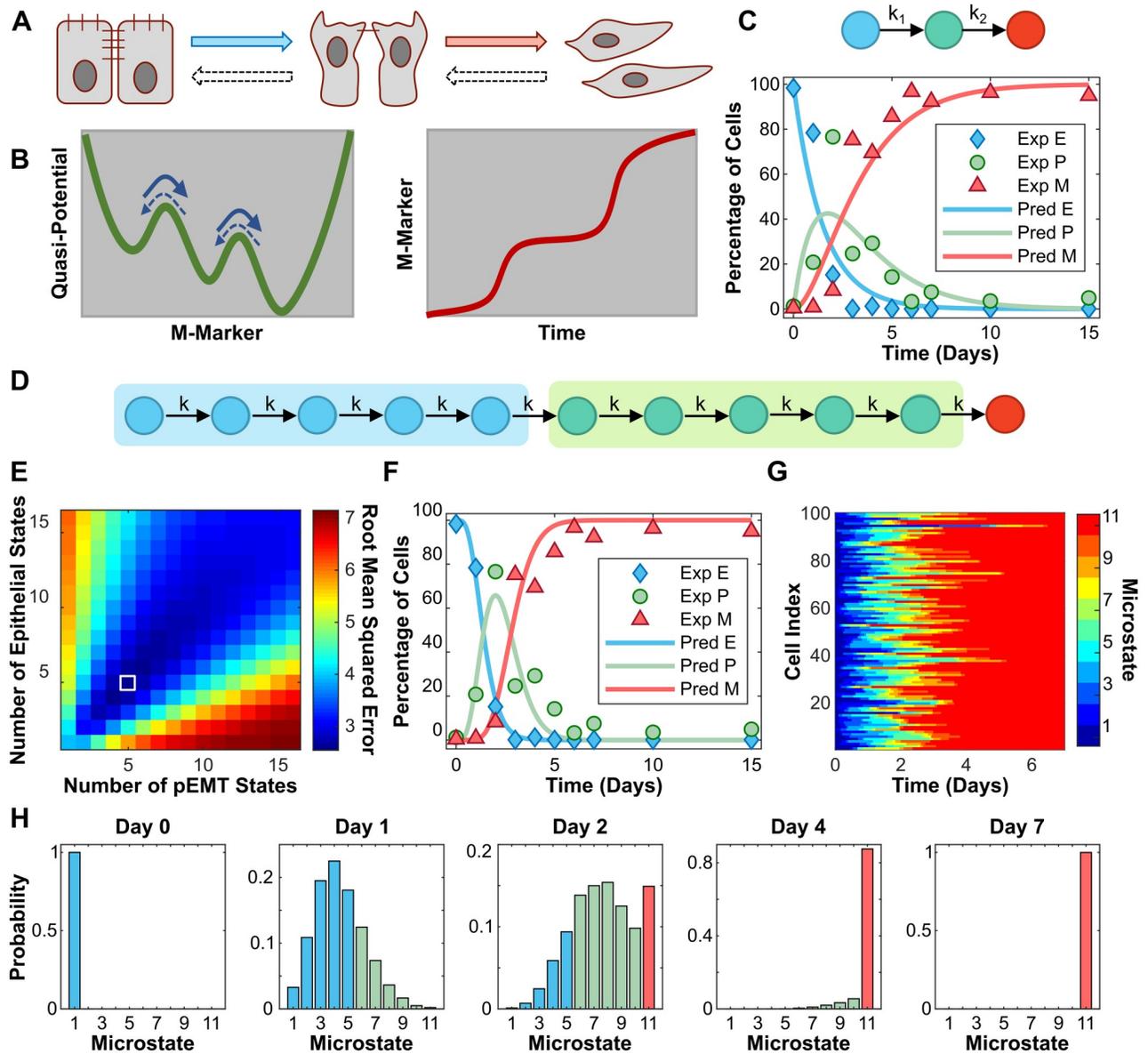
2. Initialize the system with time  $t_0$  and cell state  $x_0 = [1; 0; \dots; 0]$  as one column vector.
3. Determine the rate of each cell transition:  $a = Sx_0K$ , and  $a_0 = \sum a$ .
4. Generate two uniformly distributed random numbers,  $rn_1$  and  $rn_2$ , between  $[0, 1]$ .
5. Determine the time it takes for the cell state transition:  $dt = \ln(1/rn_1)/a_0$ .
6. Determine the step in which the cell state transition occurs,  $r$ , which satisfies:
 
$$\sum_{i=1}^{i=r-1} a(i) < rn_2 \cdot a_0 < \sum_{i=1}^{i=r} a(i).$$
7. Update the system:  $x_0 = x_0 - S(r, :) + P(r, :)$  and  $t = t + dt$ .
8. Repeat step 3–7 until  $t$  is more than the maximum time,  $T_{max}$ .

## Results

### EMT is a non-Markov process

Consistent with theoretical prediction [5], three steady states were found in TGF- $\beta$  induced EMT in the MCF10A cell line with quantitative measurements of E-cadherin and Vimentin at the single cell level [6]. That is, EMT progresses through three functional cell states in this cell line: from the initial epithelial state to the partial EMT state, where cells lose some of the cell-cell adhesion, and then to the full mesenchymal state (Fig 1A). As shown in Fig 1B (left), this system can be represented with a metaphorical three-well landscape in one dimension along the mesenchymal marker (M-Marker) axis. Each well represents a stable or metastable cell phenotype during the EMT process and the transitions among them are indicated with arrows. With high concentration of TGF- $\beta$ , the cell will be driven from the epithelial state (the first well) to the mesenchymal state (last well) through the partial EMT state (middle well), indicated by solid arrows. It is noted that the energy barrier for the forward transition is much smaller than the energy barrier for the reverse transition of each step at high concentration of TGF- $\beta$ . Thus, the forward transition rate (Fig 1B, solid arrows) is much larger than the reverse transition rate (Fig 1B, dash arrows). The dynamics of the M-Marker (measured by the Vimentin protein) at the single cell level shows two steps with cells staying in the intermediate state temporarily (Fig 1B, right) [5]. The underlying mechanism is that during EMT the phenotype changes of one cell are controlled by complex genetic regulatory networks and stepwise activation of multiple feedback loops, as demonstrated by Ref. [5, 6].

Due to the cell-cell variability, the cells do not synchronize the transition through these states but make the transitions in a stochastic manner. However, the fraction of cells in each



**Fig 1. EMT is a non-Markov process.** (A) Cell phenotype transitions among epithelial, partial EMT, and mesenchymal states during EMT. (B) Metaphorical three-well landscape in one dimension along the mesenchymal marker (M-Marker) axis shows the cell phenotype transition during EMT. Full arrows show the order of EMT while dashed arrows show reverse process, mesenchymal-to-epithelial transition (MET). (C) Three-state Markov model for EMT (top). Best fit of the three-state model to previous experimental data from Ref. [6] on temporal changes of the percentage of cells in three states during EMT with  $k_1 = 0.6657$ ,  $k_2 = 0.4908$ , and  $RMSE = 13.9$  (bottom). (D) An extended model with  $N_E$  epithelial microstates (blue circles) and  $N_P$  partial EMT microstates (green circles). (E) The fitting score, root mean squared error ( $RMSE$ ), with the extended model in the space of  $N_E$  and  $N_P$ . The best fit is at  $N_E = 5$  and  $N_P = 5$  (white square,  $RMSE = 2.55$ ,  $k = 3.4261$ ). (F) The dynamics of cells in three macrostates during EMT with the best-fitted extended model overlaid with experimental data. (G) Stochastic simulation shows the state transition of 100 simulated cells over time. (H) Probability of cells in each microstate at several time points. 10,000 simulated cells with a stochastic model (see [Materials and methods](#) for details) are sampled to calculate the probability.

<https://doi.org/10.1371/journal.pcbi.1007682.g001>

state shows deterministic dynamics, as seen in the quantified experimental data (Fig 1C) from Ref. [6]. To better understand EMT at the population level, we first build a simple model by assuming that cells are initially held in the epithelial state and will be driven by a high dose of TGF- $\beta$ , as used in the experiments, to the partial EMT state before ending in the mesenchymal

state, all with constant average rates ( $k_1$  and  $k_2$ ) (Fig 1C, top). We fit this simple model with the experimental data and find that the model does not describe the data with strong accuracy (Fig 1C, bottom). An inaccurate fitting was also found using a two-state model without any partial EMT states (S1 Fig, panel A). With three-state or two-state models, the transition from the partial EMT state to the mesenchymal state shows a quicker dynamic than the observed data. Thus, three-state or two-state models are not a good representation of the EMT process at the population level, and the cells still pass through other intermediate states temporarily before reaching the final mesenchymal state, which is the only attractor at high concentration of TGF- $\beta$ .

To find out the underlying mechanism for the observed dynamics, we extend the simple model by adding several microstates for each cell macrostate (Fig 1D) by assuming that some unmeasured variables can further distinguish the epithelial state and partial EMT state into several microstates, with the epithelial and partial EMT states being the macrostates used in describing the overall EMT process. While we do not know the actual number of the epithelial and partial EMT microstates ( $N_E$  and  $N_P$ , respectively), a fitting of the extended model is done with all the combinations of  $N_E$  and  $N_P$  (Fig 1E) by assuming an equal transition rate from one state to the following state ( $k$ ). According to the fitting score (root mean squared error), the best fit is located when  $N_E = 5$  and  $N_P = 5$  (white square in Fig 1E). The best fitting displayed in Fig 1F shows the optimal conditions in which our model reproduces the EMT dynamics at the population level with better accuracy than that of the assumption of only one microstate in both the epithelial and partial EMT states. It is noted that the three macrostates used in our description are defined by E-Cadherin and Vimentin single-cell data in Ref. [6], not including other makers. In general, the macrostates are defined by the corresponding steady states observed in the experiment with the defined markers. This implies that the number of macrostates observed may differ depending on the markers used. Four macrostates were defined based on additional makers such as *Ovol2* [9]. Here, we limited the fitting of the Markov model to the temporal dynamics of the three macrostates as defined in Ref. [6]. Currently, there is no experimental evidence for the existence of nine intermediate states in this cell line, suggesting the existence of microstates during EMT. These results suggest that EMT is a non-Markov process with many hidden epithelial microstates and partial EMT microstates, which could further be distinguished by measuring other variables in the system.

This conclusion is further confirmed with a stochastic model to demonstrate the evolution of cell transitions in these microstates (see [Materials and methods](#) for more details). Fig 1G shows 100 typical simulated stochastic trajectories of cells to complete the EMT process. With all cells beginning in the epithelial state, prominent cell-to-cell variation is shown in the time it takes a cell to convert to different microstates, indicating unsynchronized transition at the single cell level. At the population level, the statistic results with 10,000 simulated cells show the dynamics of the distribution of cells in the microstates (Fig 1H). Few reach the mesenchymal state (red bar) by day two. However, most cells will reach the final EMT state by day four. This is also consistent with the experimental data [6] (S1 Fig, panel B). Furthermore, fitting the experimental data with a model by considering the reversible state transitions suggests more microstates (S2 Fig, panel A). Taken together, our analysis based on fitting a general Markov model with the experimental data suggests that EMT is a non-Markov process, where several microstates exist within the epithelial and partial EMT states.

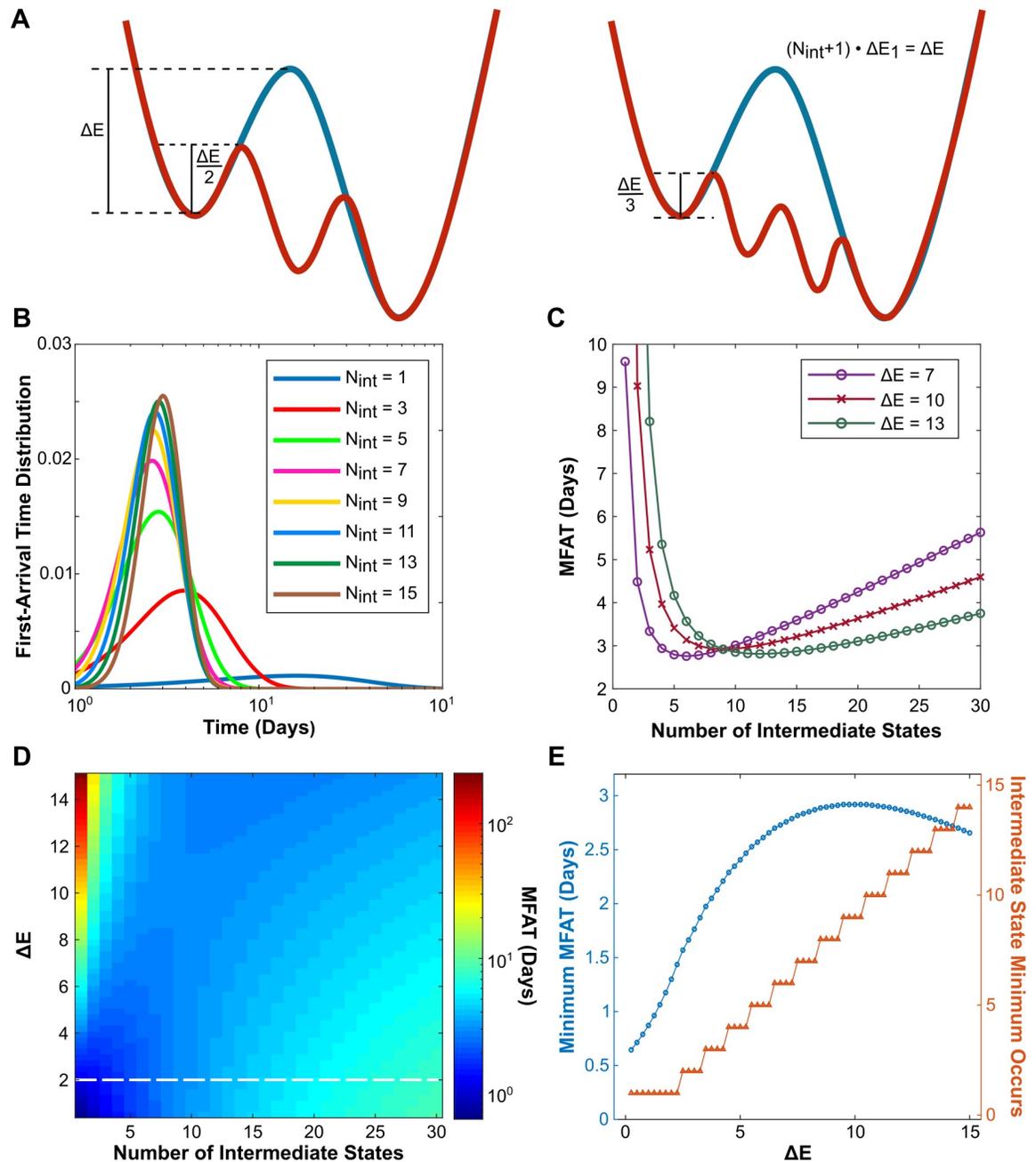
### Increasing the number of intermediate states can accelerate EMT

Our analysis with a Markov model for EMT at the population level based on the experimental data of MCF10A cells suggests that multiple microstates exist in each macrostate during the

EMT process. However, the number of microstates and macrostates can be cell-type specific. Different numbers of observable steady states have been found during EMT and cancer metastasis [9, 15–23], which leads us to wonder how the number of micro- and macro-states affects the EMT dynamics and cancer metastasis. For this reason, we now consider all states between the initial epithelial and final mesenchymal as “intermediate” states in order to generalize our results and allow for application into cancer cell lines. To compare the systems with different numbers of intermediate states, we made one assumption: the total energy barrier for the full transition from epithelial (E) to mesenchymal (M) is the same despite the total number of states needed to complete EMT. As shown in Fig 2A (left), if the energy barrier from E to M is  $\Delta E$  without any intermediate states (blue curve), adding one additional state makes the transition two steps and the barrier becomes  $\Delta E/2$  (red curve) for each step. Similarly, the barrier is  $\Delta E/3$  for each step if two intermediate states are considered (Fig 2A, right, red curve). This assumption comes from the hypothesized monotonical energy gradient based on epigenetic changes between different EMT states [43] and existence of ‘checkpoints’ in the EMT continuum [44]. This assumption is also consistent with the cascading bistable switches mechanism in which EMT proceeds through step-wise activation of multiple feedback loops [5, 6, 9, 45]. The cells need to change the profiles of gene expression to make the full EMT transition. Under this assumption with one intermediate state, the cells can make changes on the expression of some genes as the first step for a partial transition, and then make changes on the rest of the genes for the second step to complete the transition. The profiles of the genes that control the eukaryotic cell phenotypes are usually sophisticated by positive feedback loops to avoid undesired random phenotype switching from noises or short pulse stimulus. These feedback loops can be mutual inhibitions between two groups, one of them controls cell phenotype I while the other group controls cell phenotype II. Thus, the cell fate transitions through EMT intermediate states are analogous to crossing a series of energy barriers controlled by these positive feedback loops.

First, we systematically study how the EMT process changes with fixed  $\Delta E$  and the number of intermediate states,  $N_{int}$ . The distributions of the first-arrival time (FAT), defined as the time that cells take to arrive to the mesenchymal state, is used to measure how fast the EMT process is completed. As shown in Fig 2B, with increase of  $N_{int}$ , the FAT distribution first shifts to the left and then to the right, becoming increasingly more narrow. To see this affect more clearly, we further calculate the mean first arrival time (MFAT) and find a non-monotonic dependence of MFAT on  $N_{int}$  (Fig 2C). As  $N_{int}$  increases, the MFAT first shows a rapid decrease. However, after a certain  $N_{int}$ , the MFAT begins to slowly increase again. These data suggest that the partial intermediate states have a potential function of accelerating the EMT process. Furthermore, we analyze how  $\Delta E$  affects the EMT process. As shown in Fig 2C, the MFAT displays the same non-monotonic pattern for three different  $\Delta E$  values, although the position of the minimum varies. At these minimum points, the EMT process is completed the fastest. This interesting phenomena is further confirmed by the analysis of the MFAT in the space of  $\Delta E$  and  $N_{int}$  (Fig 2D). The same trend of MFAT can be observed for all  $\Delta E > 2$  (threshold shown by the dash line). This result suggests that the EMT process is not a continuous process but proceeds with a finite number of discrete microstates.

For a  $\Delta E$  value lower than the threshold, MFAT constantly increases with  $N_{int}$  (Fig 2D). The distribution of the first-arrival time for one case with small  $\Delta E$  is shown in S3 Fig. One can see that the transition time is already very small in this case and that increasing intermediate states barely increases the transition rate for each step, thus causing more time to complete full EMT. This is generally not the case for the EMT process in a real system, as the minimal MFAT is less than one day, but it demonstrates the underlying mechanism of the non-monotonic dependence of MFAT on  $N_{int}$ . That is, increasing  $N_{int}$  after the minimum



**Fig 2. Increasing the number of intermediate states can accelerate the EMT process.** (A) The assumptions that the total energy barrier ( $\Delta E$ ) for the full transition from epithelial to mesenchymal is fixed to compare the effects of the number of intermediate states ( $N_{int}$ ) on the dynamics of EMT. The energy barrier is  $\Delta E/(N_{int} + 1)$  for each step for the system with  $N_{int}$  intermediate states. (B) The distribution of the first arrival time to the mesenchymal state with different  $N_{int}$  and fixed  $\Delta E = 10$ . (C) The dependence of the mean first arrival time (MFAT) on  $N_{int}$  for three different energy barrier values as indicated. (D) MFAT in the space of  $\Delta E$  and  $N_{int}$ . A dashed white line represents the threshold above which MFAT shows a non-monotonic dependence on  $N_{int}$ . (E) The dependence of the minimum MFAT (blue) on  $\Delta E$  alongside the intermediate state that this minimum occurs (red).

<https://doi.org/10.1371/journal.pcbi.1007682.g002>

does not help to increase the transition rate for each step, it just simply increases the number of steps. Thus, there is a trade-off to accelerate EMT through increasing the number of intermediate states. The dependence of the minimum MFAT alongside the intermediate state in which it occurs ( $N_{int}^{min}$ ) is shown in Fig 2E. With increase of  $\Delta E$ , both the minimal MFAT and

$N_{int}^{min}$  increase, although a small decline of minimal MFAT is observed for large  $\Delta E$ . Taken together, a potential function of multiple partial EMT states is to accelerate EMT, and the number of intermediate states necessary to achieve fastest EMT increases with the total energy barrier.

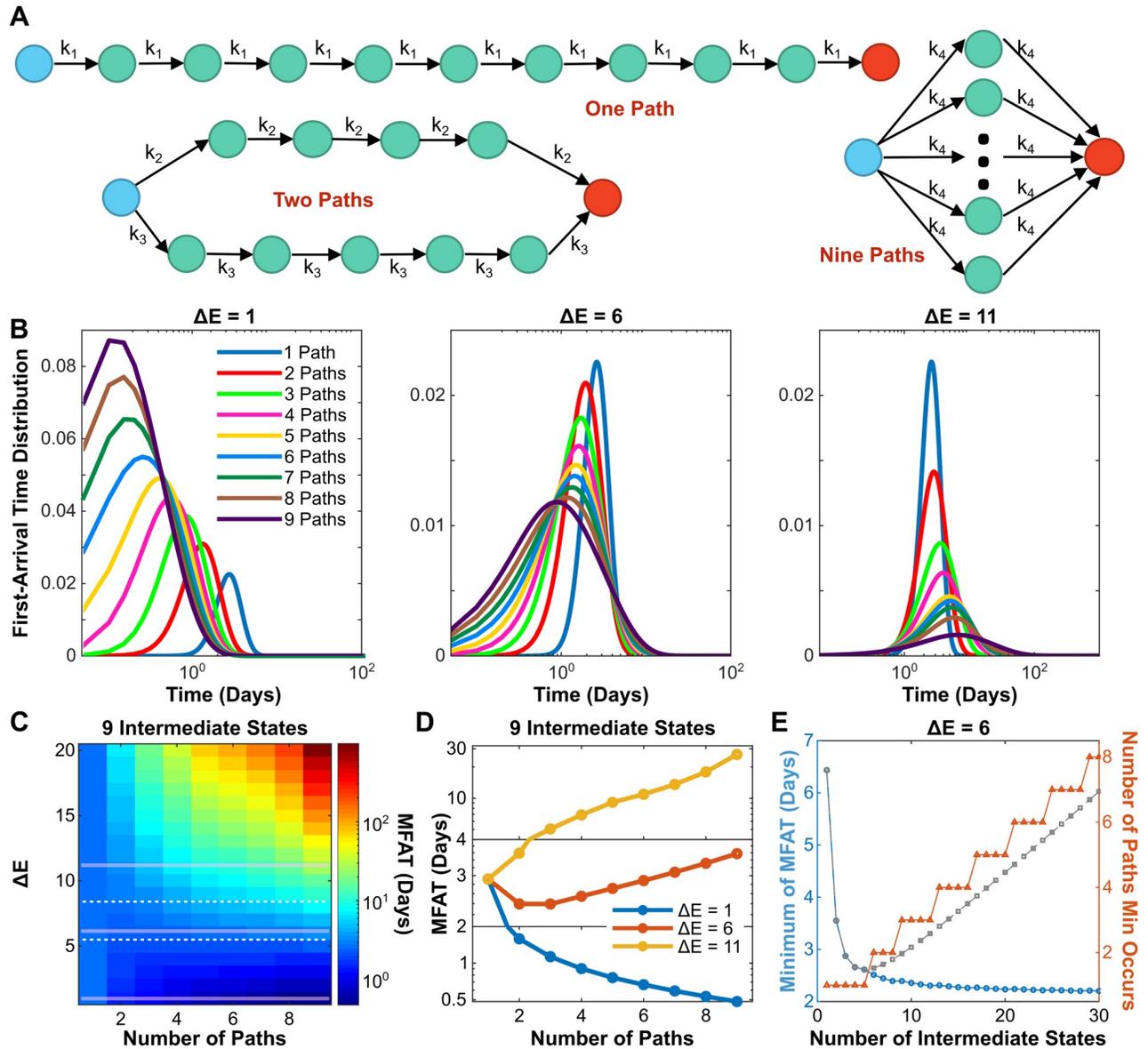
### Adding parallel paths or transition layers further accelerates the EMT process

We have discussed the scenario with only one path in which epithelial cells proceed through a step-wise transition of intermediate states to become mesenchymal cells. Recently, parallel EMT paths were reported [46, 47]. Here, we consider multiple paths in parallel (Fig 3A) and study how the number of paths ( $N_{pth}$ ) affects EMT. For  $N_{int}$  intermediate states, there are  $N_{int}$  possible cases ( $N_{pth} = 1 \sim N_{int}$ ) based on the one simplification that each path contains close to the same number of intermediate states. For example, if there are nine intermediate states and two paths, four states would belong to one path and five to the other (Fig 3A).

To demonstrate the effect of multiple paths, we first study the system with nine intermediate states ( $N_{int} = 9$ , the best fit from Fig 1E). The FAT distribution to the mesenchymal state is shown in Fig 3B under various  $N_{pth}$  and three levels of energy barrier,  $\Delta E$ . For small  $\Delta E$  (left panel), the FAT distribution shifts to the left as  $N_{pth}$  increases, implying that more paths create a faster transition. However, for large  $\Delta E$  (right panel), the FAT distribution reflects the opposite, implying that more paths create a slower transition. These also suggests a non-monotonic dependence of FAT on  $N_{pth}$  for moderate  $\Delta E$ . As shown in Fig 3B (middle panel), the FAT distribution peak shifts to the left but the distribution width becomes larger with increase of  $N_{pth}$ . This suggests that the monotonic/non-monotonic relationship between the FAT and  $N_{pth}$  is dependent on the energy barrier.

To further confirm this conclusion, we systematically study how the MFAT of the mesenchymal state changes in the space of  $\Delta E$  and  $N_{pth}$  (Fig 3C). The MFAT dependence on  $N_{pth}$  indeed shows three trends according to the value of  $\Delta E$ . For small (below the bottom white dashed line) or large (above the top white dashed line) values of  $\Delta E$ , the dependence is monotonic, either decreasing or increasing respectively as exemplified in Fig 3D (blue and yellow curves). On the other hand, for a bounded range of  $\Delta E$  values between the two white dashed lines (Fig 3C), it is non-monotonic as exemplified in Fig 3D (red curve). The non-monotonic dependence of MFAT on  $N_{int}$  can change depending on  $N_{pth}$  (S4 Fig). For  $N_{pth} = 1$ , MFAT shows a non-monotonic dependence on  $N_{int}$ , the same as Fig 2C. However, the position of the minimum shifts to the right as  $N_{pth}$  is increased to 2 ~ 3. The non-monotonicity is then lost with a continuous increase of  $N_{pth}$ . The minimum of MFAT decreases monotonically with the increase of  $N_{int}$ . In order to maintain this dynamics, the number of parallel paths must also increase with  $N_{int}$  (Fig 3E). That is, increasing  $N_{pth}$  makes the dependence of MFAT on  $N_{int}$  more monotonically decreasing. These data suggest that increasing both  $N_{int}$  and  $N_{pth}$  together accelerates the EMT process.

We have now discussed the scenario in which there is only one path and the scenario of multiple paths in parallel and found that increasing the number of the intermediate states or parallel paths could accelerate the EMT process. Here, another scenario is considered by assuming that the epithelial cell has to pass a fixed number of intermediate states to become mesenchymal. In other words, the epithelial cell must pass  $N_{ly}$  of layers to become mesenchymal (Fig 4A). This is similar to the parallel paths scenario but has one difference; it is possible that two paths converge at one intermediate state in the layered scenario. The difference between two topologies is exemplified in S5 Fig, panel A. One can see that a single extra transition step exists in the layered case compared with the parallel case. In the layered topology, the

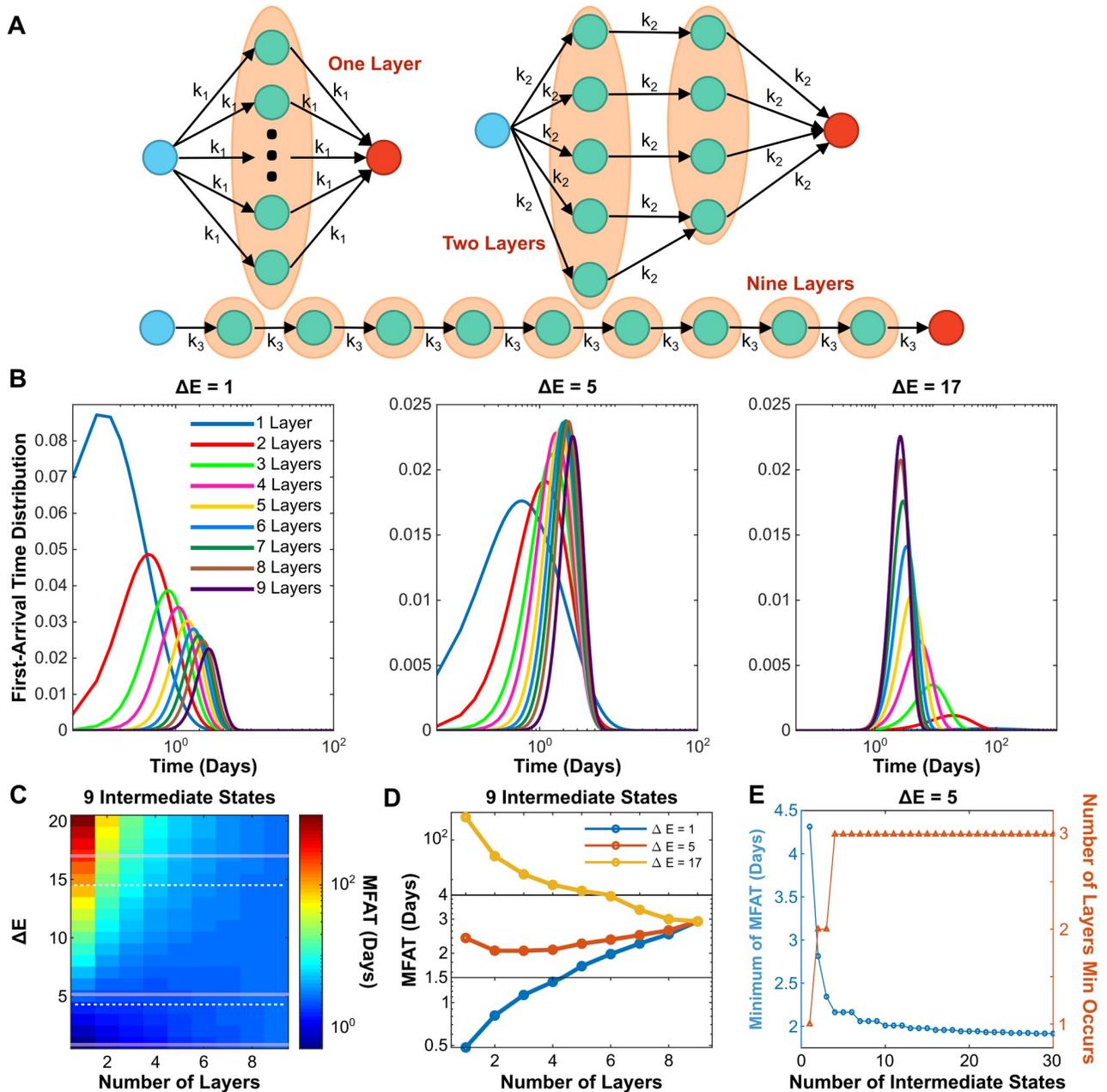


**Fig 3. Adding parallel paths can further accelerate the EMT process.** (A) Diagram of cell phenotype transition through nine intermediate states with multiple parallel paths. (B) Dependence of FAT distribution on  $N_{pth}$  with three different  $\Delta E$  values and  $N_{int} = 9$ . (C) Phase diagrams showing MFAT subjected to  $N_{pth}$  and  $\Delta E$  with  $N_{int} = 9$ . Dashed lines represent two thresholds for monotonic and non-monotonic dependence of MFAT on  $N_{pth}$ . (D) Three typical examples show the monotonic and non-monotonic dependence of MFAT on  $N_{pth}$  according to  $\Delta E$  value. (E) The dependence of minimal MFAT (blue curve) with a comparison to one parallel path (grey dashed line), and the  $N_{pth}$  the minimum occurs on (red) for  $\Delta E = 6$ .

<https://doi.org/10.1371/journal.pcbi.1007682.g003>

transition rates in all steps are the same, which depends on the number of layers  $N_{ly}$ . The MFAT to the mesenchymal state as a function of  $\Delta E$  is not equivalent for the two cases (S5 Fig, panel B). Interestingly, the two curves cross at one specific  $\Delta E$ . For  $\Delta E$  values below this cross-point, EMT is completed faster in the parallel topology. However, for higher  $\Delta E$  values, the parallel topology completes EMT at a slower rate than the layered topology.

Similar to the parallel path scenario, the first-arrival time distribution depends on  $\Delta E$  (Fig 4B). If  $\Delta E$  is small, the FAT distribution shifts to the right with increase of  $N_{ly}$  (Fig 4B, left panel), while it shifts to the left as  $\Delta E$  becomes large (Fig 4B, right panel). The FAT shows a



**Fig 4. Adding transition layers can further accelerate the EMT process.** (A) Diagram of cell state transition through nine intermediate states with multiple transition layers. (B) Dependence of FAT distribution on the number of layer  $N_{ly}$  with three different  $\Delta E$  and  $N_{int} = 9$ . (C) Phase diagrams showing MFAT subjected to  $N_{ly}$  and  $\Delta E$  with  $N_{int} = 9$ . Dashed lines represent two thresholds for monotonic and non-monotonic dependence of MFAT on  $N_{ly}$ . (D) Three typical examples show monotonic and non-monotonic dependence of MFAT on  $N_{ly}$ . (E) The dependence of minimum MFAT (blue curve) and the  $N_{ly}$  the minimum occurs on (red) for  $\Delta E = 5$ .

<https://doi.org/10.1371/journal.pcbi.1007682.g004>

non-monotonic dependence on  $N_{ly}$  with the moderate  $\Delta E$  (Fig 4B, middle panel). Fig 4C shows the MFAT in the space of  $N_{ly}$  and  $\Delta E$  for the system with nine intermediate states. In general, a higher energy barrier gives a slower mean time arrival regardless of the number of layers. However, the trend of the MFAT versus  $N_{ly}$  depends on  $\Delta E$ . MFAT increases with  $N_{ly}$  for small  $\Delta E$  (below the bottom white dashed line), but decreases for large  $\Delta E$  (above the top

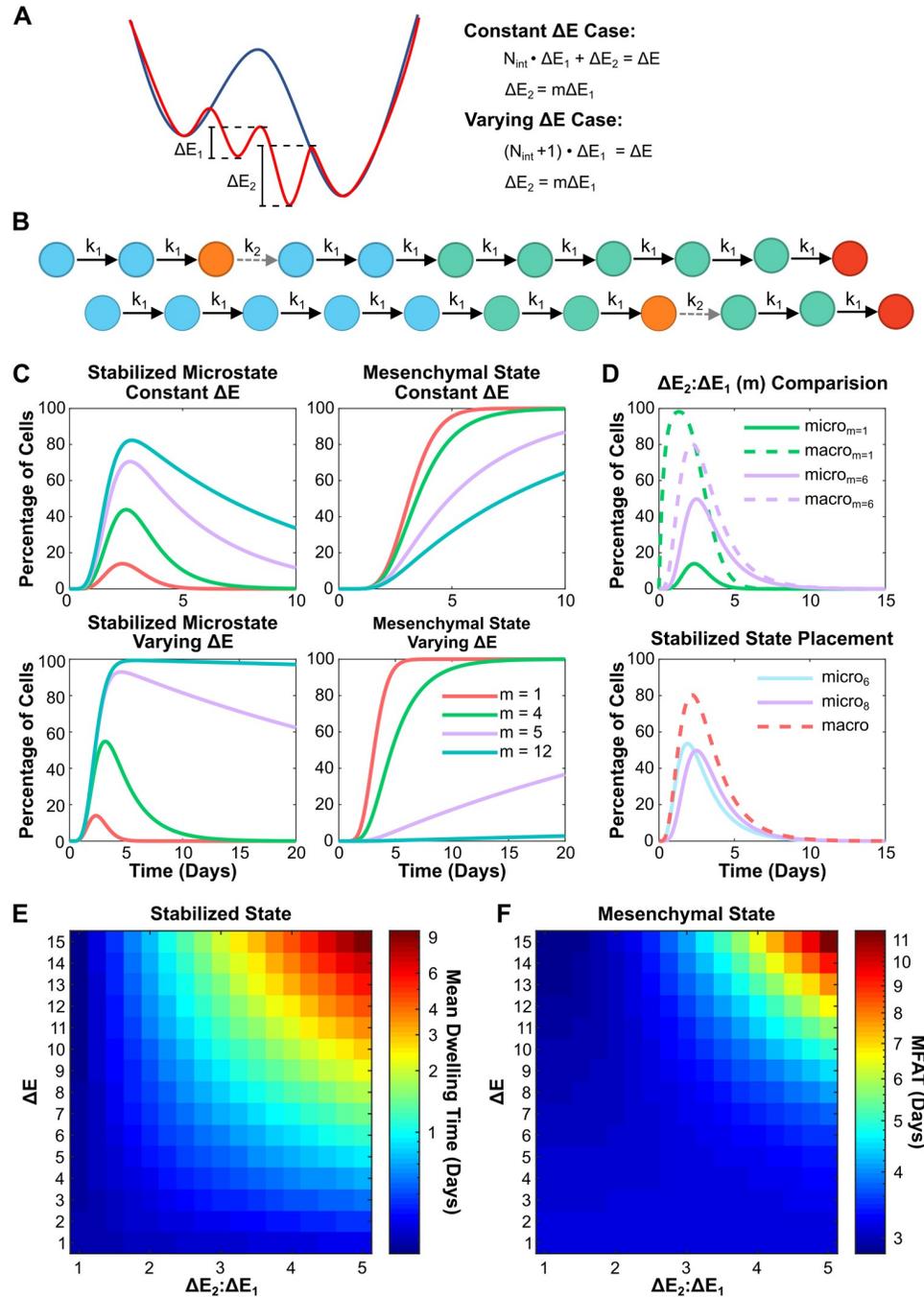
white dashed line) and shows a non-monotonic dependence for moderate  $\Delta E$  (between the two white dashed lines), as exemplified by the three cases in Fig 4D. Similar to the parallel path scenario, the minimum of MFAT decreases monotonically with  $N_{int}$  as long as more layers are provided for the system with large number of  $N_{int}$  (Fig 4E). This is also confirmed with various values of  $\Delta E$  (Fig 4E and S6 Fig).

Taken together, adding more parallel paths or transition layers can further accelerate the EMT process. In both scenarios, a non-monotonic dependence of the mean first arrival time on the number of paths or layers is found. The MFAT increases with the number of intermediate states after the minimal if the system only has one path, but can further decrease with multiple paths or transition layers. Thus, a combination of increasing the number of intermediate states and parallel paths (or transition layers) can always accelerate the EMT process.

### Stabilized intermediate state traps a cell within its current phenotype

We have discussed the scenarios in which the same energy barrier is considered for each step during the EMT process. However, it is very possible that one of the intermediate states is stabilized by specific regulators for some cancer cells [48], making the transition from this step more difficult. While mesenchymal cells contribute to cancer metastasis in many cases, the partial EMT states were found to be more aggressive [30, 49]. Here, we consider this scenario where one of the intermediate states is more stable than the others; we call this state the “stabilized” state, which is related to certain states within the more aggressive forms of cancer cells. As shown in Fig 5A, our metaphoric landscape now has one well that is deeper than the others. In this stabilized state, more energy is needed for the transition to the next state. That is, the energy barrier for the transition at this step ( $\Delta E_2$ ) is larger than the energy barrier for other steps ( $\Delta E_1$ ). Two cases are considered here: 1)  $\Delta E_2$  increases and  $\Delta E_1$  decreases in order to maintain the overall energy barrier  $\Delta E$  (Fig 5A, “constant  $\Delta E$  case”), and 2)  $\Delta E_2$  increases as  $\Delta E_1$  remains the same to change the overall energy barrier (Fig 5A, “varying  $\Delta E$  case”). Fig 5B represents the diagram of the EMT process with one stabilized state, which includes two transition rates,  $k_1$  as the transition rate from the non-stabilized states, and  $k_2$  as the transition from the stabilized state.

Similar to the mechanism for the accelerated transition to the mesenchymal state in Figs 2–4, increasing the number of intermediate states before the stabilized state and adding layers or paths should also accelerate the transition to the stabilized state, given that the formula of the MFAT to the stabilized state is similar to the one to the mesenchymal state (see [Materials and methods](#)). To further understand the impact of a stabilized state on EMT, the distribution of cells at the stabilized state and the mesenchymal state is analyzed with different  $\Delta E_2$ :  $\Delta E_1$  ratios. As the ratio increases, the dwelling time inside the stabilized state grows and it takes more time to arrive to the mesenchymal state for both the constant and varying  $\Delta E$  cases (Fig 5C). That is, the greater the ratio, the more time the cancer cells can stay in the aggressive stabilized state for cancer metastasis. It also implies that cells stay longer in the macrostate where the stabilized intermediate state belongs (Fig 5D, top panel) and that the placement of the stabilized intermediate state within this macrostate does not affect the macrostate behavior (Fig 5D, bottom panel). The dependence of the mean dwelling time within the stabilized state and the MFAT is further analyzed. As shown in Fig 5E and 5F, with increase of either the ratio  $\Delta E_2$ :  $\Delta E_1$  or the overall energy barrier, the mean dwelling time within the stabilized state and the MFAT increase. The same dependence is found for the varying  $\Delta E$  case (S7 Fig, panels A-B) but the scale of changes significantly increases. Another difference between two cases is the reverse dependence of MFAT on  $\Delta E_2$ :  $\Delta E_1$  as



**Fig 5. A stabilized intermediate state traps a cell within its current phenotype.** (A) The metaphorical landscape of EMT with one stabilized state, which has a deeper well ( $\Delta E_2$ ) than the others ( $\Delta E_1$ ). Two cases were considered here: (1) the constant  $\Delta E$  case in which  $\Delta E_2$  increases and  $\Delta E_1$  decreases but overall energy barrier  $\Delta E$  remains the same, and (2) the varying  $\Delta E$  case in which only  $\Delta E_2$  increases and  $\Delta E_1$  remains the same. (B) Example of a stabilized state occurring within the EMT process. Orange represents the stabilized intermediate state and the gray dashed arrow represents the transition from this stabilized state ( $k_2$ ). (C) The distribution of cells at the stabilized state and the mesenchymal state under different energy barrier ratio,  $m$  ( $= \Delta E_2:\Delta E_1$ ). (D) Top: The distribution of the cell at the stabilized microstate (solid lines, the 6th microstate) or at the corresponding macrostate (dash lines, including the 6-10th microstates) with  $m = 1$  (green, non-stabilized) or  $m = 6$  (purple, stabilized). Bottom: The distribution of the cell at the stabilized microstate (solid lines) or at the corresponding macrostate (red dashed line) with the stabilized state in 6th intermediate state (blue) or the 8th intermediate state (purple). (E-F) Phase diagrams showing mean dwelling time at the stabilized state (E) and MFAT to mesenchymal (F) subjected energy barrier ratio and the total energy barrier  $\Delta E$ , in the constant  $\Delta E$  case.

<https://doi.org/10.1371/journal.pcbi.1007682.g005>

shown in [S7 Fig](#) panels C-D when  $\Delta E$  is small. Taken together, a greater overall energy barrier or greater intermediate state energy barrier ratio will cause for a longer stay in the aggressive stabilized state as well as a slower transition into the mesenchymal state; this may contribute to cancer metastasis.

## Discussion

Multiple partial EMT states have been proposed and verified to exist in the EMT process, and the number of states involved depends on the cell line. Here, we focused on the dynamics and functions of EMT plasticity and provided microstate and macrostate concepts for EMT. Based on the parameter fitting of a Markov model of EMT with the experimental data, we proposed a statistical mechanism for EMT in which many unobservable microstates exist within the observable macrostates. That is, EMT is a process with many microstates, which can be mapped into different observable macrostates. The microstates may encompass other dimensions of the cell that are coupled to the EMT process, such as the cell cycle, stemness, and metabolism. How the number of microstates is determined is not clear but is believed to largely depend on the cell line, which may be revealed with analysis of single-cell time-course data [50].

Here, we performed a systematical analysis on how the number of intermediate states changes cell transformation. We considered several scenarios including a single transition path, parallel transition paths, and layered transitions. We found that in the single transition path scenario, increasing the number of intermediate states can accelerate EMT, but too many intermediate states can also potentially decelerate EMT, thus showing non-monotonic behavior. However, adding parallel paths or transition layers can further accelerate the EMT process, especially for a system with a large number of intermediate states.

Our results suggest that the number of intermediate EMT states may function as an indicator for the malignancy of the cancer. That is, the more intermediate EMT states in one cancer type, the better the chance it will metastasize. A complementary feature of multiple intermediate phenotypes was reported to stabilize the cancer stem cell population [51]. Thus, the malignancy level can be quantified by the number of potential intermediate EMT states. It is estimated that metastasis causes 90% of deaths related to cancer [49]. If we can stop or slow down the transition, we have a great chance of treating cancer. Our results suggest one potential strategy targeting on the EMT spectrum, i.e., reducing the number of the potential partial EMT states.

Lastly, the existence of a stabilized state within the EMT process causes a slower transition. It is reported that cells in the partial EMT state contribute more to cancer metastasis than the mesenchymal state [52]. Further more, some partial EMT states can be stabilized by the cancer cells [53]. Our simulation shows that stabilizing one intermediate state will trap the cells there for a much longer time period, slowing down full EMT. Multiple intermediate states emerge due to various interlinked positive feedback loops formed in the regulatory network of EMT, which has been demonstrated in many previous works that combine mathematical modeling and quantitative experiments [5–11, 45, 54, 60]. It also has been shown that removal of counteracting determinants traps cells in the rare myeloid transition state [55]. Similarly, if we can tune the strength of the feedback loops that are responsible for the generation of these partial EMT states, the partial EMT states will destabilize and even disband. This will make the first step of metastasis mediated by EMT or partial EMT more difficult and thus give us more time to treat the cancer.

Here in this work, we did not consider the molecular mechanism of the intermediate states but focused on the how the number of the intermediate states affect the EMT process. Further

work is needed to verify the prediction derived here and the detailed molecular mechanism needs to be determined for the potential therapeutic targets. Our results suggest that a small perturbation of the signaling pathway could change the EMT process significantly and thus EMT-related diseases, such as cancer metastasis and renal fibrosis. Our previous work on renal fibrosis suggests that knockout of the EMT genes might not be the optimal treatment design for acute kidney injury and long-term fibrosis [41]. Thus, future works can search control strategies to slow down or accelerate EMT with dynamic perturbation of the signaling pathway toward optimal treatment of cancer or fibrosis.

The accelerating function of intermediate states on EMT is analogous to the phenomena in protein folding [56]. It is becoming increasingly apparent that many other cellular processes have multiple intermediate states instead of just one binary or continuous process. For example, stem cell differentiation consists of many intermediate states [57–59]. Thus, it will be exciting to analyze one biological process in different perspectives, including statistical mechanics, mathematical modeling, and single-cell analysis.

## Supporting information

**S1 Fig. Effects from change in the number of states during EMT.** (A) Best fit of the two-state model to previous experimental data from Ref. [6] on temporal changes of the percentage of cells in two states during EMT with  $k = 0.2855$  and  $RMSE = 16.67$ . (B) The dynamics of the EMT process with the stochastic model overlaid with experimental data from Ref. [6]. The simulation data is sampled with 10,000 cells from the stochastic model.

(TIFF)

**S2 Fig. Fitting the experimental data with a model that considers reversible cell state transitions suggests more microstates during EMT.** (A) EMT progression through a continuum of reversible intermediate states. (B) The fitting score, root mean squared error ( $RMSE$ ), for the model with reversibility at the space of  $N_E$  and  $N_P$  gives the best fit at  $N_E = 9$  and  $N_P = 8$  (white square,  $RMSE = 0.0064$ ,  $k_1 = 7.1258$ , and  $k_{-1} = 0.6694$ ). (C) The dynamics of the EMT process with the best-fitted model of reversible intermediate states overlaid with experimental data from Ref. [6].

(TIFF)

**S3 Fig. FAT distribution for different number of states during EMT.** The FAT distribution to the mesenchymal state with different numbers of intermediate states and  $\Delta E = 0.1$ .

(TIFF)

**S4 Fig. Adding parallel paths changes the dependence of MFAT to the mesenchymal state on  $N_{int}$ .** MFAT as a function of  $N_{int}$  under various  $N_{pth}$  with  $\Delta E = 6$ .

(TIFF)

**S5 Fig. Comparison of the path-dependent and layer-dependent topologies.** (A) Diagram of cell phenotype transition through fifteen intermediate states with four parallel paths (left) and four transition layers (right). (B) The dependence of MFAT on  $\Delta E$  under three similar path- and layer-dependent topologies.

(TIFF)

**S6 Fig. Adding transition layers changes the dependence of MFAT to the mesenchymal state on  $N_{int}$ .** Minimum MFAT (A) and the corresponding number of layers (B) in the space of number of  $N_{int}$  and  $\Delta E$ .

(TIFF)

**S7 Fig. Comparison of two mechanisms of stabilizing one intermediate state on EMT dynamics.** (A-B) Phase diagram of the mean dwelling time of the stabilized state (A) and the MFAT to the mesenchymal state (B) on the energy barrier ratio,  $\Delta E_2: \Delta E_1$ , and total energy barrier,  $\Delta E$ , for the varying  $\Delta E$  case. (C-D) Phase diagram of the MFAT to the mesenchymal state on energy barrier ratio  $\Delta E_2: \Delta E_1$  and total energy barrier  $\Delta E (\leq 1)$  in the constant  $\Delta E$  case (C) and the varying  $\Delta E$  case (D).  
(TIFF)

## Author Contributions

**Conceptualization:** Luonan Chen, Xiao-Jun Tian.

**Data curation:** Hanah Goetz, Juan R. Melendez-Alvarez, Xiao-Jun Tian.

**Formal analysis:** Hanah Goetz, Juan R. Melendez-Alvarez, Xiao-Jun Tian.

**Funding acquisition:** Xiao-Jun Tian.

**Investigation:** Hanah Goetz, Juan R. Melendez-Alvarez, Xiao-Jun Tian.

**Methodology:** Hanah Goetz, Juan R. Melendez-Alvarez, Xiao-Jun Tian.

**Project administration:** Xiao-Jun Tian.

**Resources:** Xiao-Jun Tian.

**Software:** Xiao-Jun Tian.

**Supervision:** Xiao-Jun Tian.

**Validation:** Xiao-Jun Tian.

**Visualization:** Hanah Goetz, Juan R. Melendez-Alvarez, Xiao-Jun Tian.

**Writing – original draft:** Hanah Goetz, Xiao-Jun Tian.

**Writing – review & editing:** Hanah Goetz, Juan R. Melendez-Alvarez, Xiao-Jun Tian.

## References

1. Thiery JP, Acloque H, Huang RYJ, Nieto MA. Epithelial-Mesenchymal Transitions in Development and Disease. *Cell*. 2009; 139(5):871–890. <https://doi.org/10.1016/j.cell.2009.11.007> PMID: 19945376
2. Nieto MA. The ins and outs of the epithelial to mesenchymal transition in health and disease. *Annu Rev Cell Dev Biol*. 2011; 27:347–76. <https://doi.org/10.1146/annurev-cellbio-092910-154036> PMID: 21740232
3. Nieto MA, Huang RYJ, Jackson RA, Thiery JP. EMT: 2016. *Cell*. 2016; 166(1):21–45. <https://doi.org/10.1016/j.cell.2016.06.028> PMID: 27368099
4. Brabletz T, Kalluri R, Nieto MA, Weinberg RA. EMT in cancer. *Nat Rev Cancer*. 2018; 18(2):128–134. <https://doi.org/10.1038/nrc.2017.118> PMID: 29326430
5. Tian XJ, Zhang H, Xing J. Coupled reversible and irreversible bistable switches underlying TGF $\beta$ -induced epithelial to mesenchymal transition. *Biophys J*. 2013; 105:1079–89. <https://doi.org/10.1016/j.bpj.2013.07.011> PMID: 23972859
6. Zhang J, Tian XJ, Zhang H, Teng Y, Li R, Bai F, et al. TGF $\beta$ -induced epithelial-to-mesenchymal transition proceeds through stepwise activation of multiple feedback loops. *Sci Signal*. 2014; 7:ra91. <https://doi.org/10.1126/scisignal.2005304> PMID: 25270257
7. Huang RYJ, Wong MK, Tan TZ, Kuay KT, Ng AHC, Chung VY, et al. An EMT spectrum defines an anoikis-resistant and spheroidogenic intermediate mesenchymal state that is sensitive to e-cadherin restoration by a src-kinase inhibitor saracatinib (AZD0530). *Cell Death & Disease*. 2013; 4(11):e915–e915. <https://doi.org/10.1038/cddis.2013.442>

8. Lu M, Jolly MK, Levine H, Onuchic JN, Ben-Jacob E. MicroRNA-based regulation of epithelial-hybrid-mesenchymal fate determination. *Proceedings of the National Academy of Sciences*. 2013; 110(45):18144–18149. <https://doi.org/10.1073/pnas.1318192110>
9. Hong T, Watanabe K, Ta CH, Villarreal-Ponce A, Nie Q, Dai X. An Ovol2-Zeb1 Mutual Inhibitory Circuit Governs Bidirectional and Multi-step Transition between Epithelial and Mesenchymal States. *PLOS Computational Biology*. 2015; 11(11):e1004569. <https://doi.org/10.1371/journal.pcbi.1004569> PMID: 26554584
10. Abell AN, Jordan NV, Huang W, Prat A, Midland AA, Johnson NL, et al. MAP3K4/CBP-Regulated H2B Acetylation Controls Epithelial-Mesenchymal Transition in Trophoblast Stem Cells. *Cell Stem Cell*. 2011; 8(5):525–537. <https://doi.org/10.1016/j.stem.2011.03.008> PMID: 21549327
11. Jordan NV, Johnson GL, Abell AN. Tracking the intermediate stages of epithelial-mesenchymal transition in epithelial stem cells and cancer. *Cell Cycle*. 2011; 10(17):2865–2873. <https://doi.org/10.4161/cc.10.17.17188> PMID: 21862874
12. Kalluri R, Weinberg R. The basics of epithelial-mesenchymal transition. *J Clin Invest*. 2009; 119:1420–8. <https://doi.org/10.1172/JCI39104> PMID: 19487818
13. Futterman MA, García AJ, Zamir EA. Evidence for partial epithelial-to-mesenchymal transition (pEMT) and recruitment of motile blastoderm edge cells during avian epiboly. *Developmental Dynamics*. 2011; 240(6):1502–1511. <https://doi.org/10.1002/dvdy.22607> PMID: 21412939
14. Leroy P, Mostov KE. Slug Is Required for Cell Survival during Partial Epithelial-Mesenchymal Transition of HGF-induced Tubulogenesis. *Molecular Biology of the Cell*. 2007; 18(5):1943–1952. <https://doi.org/10.1091/mbc.E06-09-0823> PMID: 17344479
15. Steinway S, Zañudo J, Ding W, Rountree C, Feith D, Loughran TJ, et al. Network modeling of TGF $\beta$  signaling in hepatocellular carcinoma epithelial-to-mesenchymal transition reveals joint sonic hedgehog and Wnt pathway activation. *Cancer Res*. 2014; 74(21):5963–77. <https://doi.org/10.1158/0008-5472.CAN-14-0225> PMID: 25189528
16. Steinway S, Zañudo J, Michel P, Feith D, Loughran T, Albert R. Combinatorial interventions inhibit TGF $\beta$ -driven epithelial-to-mesenchymal transition and support hybrid cellular phenotypes. *NPJ Syst Biol Appl*. 2015; 1:15014. <https://doi.org/10.1038/npijsba.2015.14> PMID: 28725463
17. Joo JI, Zhou JX, Huang S, Cho KH. Determining Relative Dynamic Stability of Cell States Using Boolean Network Model. *Sci Rep*. 2018; 8(1):12077. <https://doi.org/10.1038/s41598-018-30544-0> PMID: 30104572
18. Schliekelman M, Taguchi A, Zhu J, Dai X, Rodriguez J, Celiktas M, et al. Molecular portraits of epithelial, mesenchymal, and hybrid States in lung adenocarcinoma and their relevance to survival. *Cancer Res*. 2015; 75:1789–800. <https://doi.org/10.1158/0008-5472.CAN-14-2535> PMID: 25744723
19. Tan TZ, Miow QH, Miki Y, Noda T, Mori S, Huang RY, et al. Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients. *EMBO Mol Med*. 2014; 6(10):1279–93. <https://doi.org/10.15252/emmm.201404208> PMID: 25214461
20. Tan T, Miow Q, Huang R, Wong M, Ye J, Lau J, et al. Functional genomics identifies five distinct molecular subtypes with clinical relevance and pathways for growth control in epithelial ovarian cancer. *EMBO Mol Med*. 2013; 5:1051–66. <https://doi.org/10.1002/emmm.201201823> PMID: 23666744
21. Huang R, Wong M, Tan T, Kuay K, Ng A, Chung V, et al. An EMT spectrum defines an anoikis-resistant and spheroidogenic intermediate mesenchymal state that is sensitive to e-cadherin restoration by a src-kinase inhibitor, saracatinib (AZD0530). *Cell Death Dis*. 2013; 4:e915. <https://doi.org/10.1038/cddis.2013.442> PMID: 24201814
22. Font-Clos F, Zapperi S, La Porta CAM. Topography of epithelial–mesenchymal plasticity. *Proceedings of the National Academy of Sciences*. 2018; 115(23):5902–5907. <https://doi.org/10.1073/pnas.1722609115>
23. Karacosta LG, Anchang B, Ignatiadis N, Kimmey SC, Benson JA, Shrager JB, et al. Mapping Lung Cancer Epithelial-Mesenchymal Transition States and Trajectories with Single-Cell Resolution. *bioRxiv*. 2019.
24. Pastushenko I, Brisebarre A, Sifrim A, Fioramonti M, Revenco T, Boumahdi S, et al. Identification of the tumour transition states occurring during EMT. *Nature*. 2018; 556:463–468. <https://doi.org/10.1038/s41586-018-0040-3> PMID: 29670281
25. Pastushenko I, Blanpain C. EMT Transition States during Tumor Progression and Metastasis. *Trends in Cell Biology*. 2019; 29(3):212–226. <https://doi.org/10.1016/j.tcb.2018.12.001> PMID: 30594349
26. Stylianou N, Lehman ML, Wang C, Fard AT, Rockstroh A, Fazli L, et al. A molecular portrait of epithelial–mesenchymal plasticity in prostate cancer associated with clinical outcome. *Oncogene*. 2019; 38(7):913–934. <https://doi.org/10.1038/s41388-018-0488-5> PMID: 30194451
27. Chaffer CL, San Juan BP, Lim E, Weinberg RA. EMT, cell plasticity and metastasis. *Cancer Metastasis Rev*. 2016; 35(4):645–654. <https://doi.org/10.1007/s10555-016-9648-7> PMID: 27878502

28. Aiello NM, Maddipati R, Norgard RJ, Balli D, Li J, Yuan S, et al. EMT Subtype Influences Epithelial Plasticity and Mode of Cell Migration. *Dev Cell*. 2018; 45(6):681–695 e4. <https://doi.org/10.1016/j.devcel.2018.05.027> PMID: 29920274
29. Ye X, Weinberg RA. Epithelial–Mesenchymal Plasticity: A Central Regulator of Cancer Progression. *Trends in Cell Biology*. 2015; 25(11):675–686. <https://doi.org/10.1016/j.tcb.2015.07.012> PMID: 26437589
30. Mani SA, Guo W, Liao MJ, Eaton EN, Ayyanan A, Zhou AY, et al. The Epithelial-Mesenchymal Transition Generates Cells with Properties of Stem Cells. *Cell*. 2008; 133(4):704–715. <https://doi.org/10.1016/j.cell.2008.03.027> PMID: 18485877
31. Schmidt JM, Panzilius E, Bartsch HS, Irmeler M, Beckers J, Kari V, et al. Stem-cell-like properties and epithelial plasticity arise as stable traits after transient Twist1 activation. *Cell Rep*. 2015; 10(2):131–9. <https://doi.org/10.1016/j.celrep.2014.12.032> PMID: 25578726
32. Grosse-Wilde A, d'Hérouël AF, McIntosh E, Ertaylan G, Skupin A, Kuestner RE, et al. Stemness of the hybrid Epithelial/Mesenchymal State in Breast Cancer and Its Association with Poor Survival. *PLOS ONE*. 2015; 10(5):e0126522. <https://doi.org/10.1371/journal.pone.0126522> PMID: 26020648
33. Jolly MK, Jia D, Boareto M, Mani SA, Pienta KJ, Ben-Jacob E, et al. Coupling the modules of EMT and stemness: A tunable ‘stemness window’ model. *Oncotarget*. 2015; 6(28).
34. Biddle A, Gammon L, Liang X, Costea DE, Mackenzie IC. Phenotypic Plasticity Determines Cancer Stem Cell Therapeutic Resistance in Oral Squamous Cell Carcinoma. *EBioMedicine*. 2016; 4:138–45. <https://doi.org/10.1016/j.ebiom.2016.01.007> PMID: 26981578
35. Yu M, Bardia A, Wittner B, Stott S, Smas M, Ting D, et al. Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. *Science*. 2013; 339:580–4. <https://doi.org/10.1126/science.1228522> PMID: 23372014
36. Cheung K, Padmanaban V, Silvestri V, Schipper K, Cohen J, Fairchild A, et al. Polyclonal breast cancer metastases arise from collective dissemination of keratin 14-expressing tumor cell clusters. *Proc Natl Acad Sci U S A*. 2016; 113:E854–63. <https://doi.org/10.1073/pnas.1508541113> PMID: 26831077
37. Shibue T, Weinberg R. EMT, CSCs, and drug resistance: the mechanistic link and clinical implications. *Nat Rev Clin Oncol*. 2017; 14:611–629. <https://doi.org/10.1038/nrclinonc.2017.44> PMID: 28397828
38. Lovisa S, LeBleu V, Tampe B, Sugimoto H, Vadnagara K, Carstens J, et al. Epithelial-to-mesenchymal transition induces cell cycle arrest and parenchymal damage in renal fibrosis. *Nat Med*. 2015; 21:998–1009. <https://doi.org/10.1038/nm.3902> PMID: 26236991
39. Grande M, Sánchez-Laorden B, López-Blau C, De FC, Boutet A, Arévalo M, et al. Snail1-induced partial epithelial-to-mesenchymal transition drives renal fibrosis in mice and can be targeted to reverse established disease. *Nat Med*. 2015; 21:989–97. <https://doi.org/10.1038/nm.3901> PMID: 26236989
40. Xing J, Tian XJ. Investigating epithelial-to-mesenchymal transition with integrated computational and experimental approaches. *Phys Biol*. 2019; 16(3):031001. <https://doi.org/10.1088/1478-3975/ab0032> PMID: 30665206
41. Tian XJ, Zhou D, Fu H, Zhang R, Wang X, Huang S, et al. Sequential Wnt Agonist then Antagonist Treatment Accelerates Tissue Repair and Minimizes Fibrosis. *arXiv*. 2018.
42. Jia D, Lu M, Jung KH, Park JH, Yu L, Onuchic JN, et al. Elucidating cancer metabolic plasticity by coupling gene regulation with metabolic pathways. *Proceedings of the National Academy of Sciences*. 2019; 116(9):3909–3918. <https://doi.org/10.1073/pnas.1816391116>
43. Tam WL, Weinberg RA. The epigenetics of epithelial-mesenchymal plasticity in cancer. *Nature Medicine*. 2013; 19(11):1438–1449. <https://doi.org/10.1038/nm.3336> PMID: 24202396
44. McFaline-Figueroa JL, Hill AJ, Qiu X, Jackson D, Shendure J, Trapnell C. A pooled single-cell genetic screen identifies regulatory checkpoints in the continuum of the epithelial-to-mesenchymal transition. *Nature Genetics*. 2019; 51(9):1389–1398. <https://doi.org/10.1038/s41588-019-0489-5> PMID: 31477929
45. Celia-Terrassa T, Bastian C, Liu DD, Eil B, Aiello NM, Wei Y, et al. Hysteresis control of epithelial-mesenchymal transition dynamics conveys a distinct program with enhanced metastatic ability. *Nat Commun*. 2018; 9(1):5005. <https://doi.org/10.1038/s41467-018-07538-7> PMID: 30479345
46. Panchy N, Azeredo-Tseng C, Luo M, Randall N, Hong T. Integrative Transcriptomic Analysis Reveals a Multiphasic Epithelial–Mesenchymal Spectrum in Cancer and Non-tumorigenic Cells. *Frontiers in Oncology*. 2020; 9. <https://doi.org/10.3389/fonc.2019.01479> PMID: 32038999
47. Wang W, Xing J. Analyses of Multi-dimensional Single Cell Trajectories Quantify Transition Paths Between Nonequilibrium Steady States. *bioRxiv*. 2020; p. 2020.01.27.920371.
48. Biswas K, Jolly M, Ghosh A. Stability and mean residence times for hybrid epithelial/mesenchymal phenotype. *Phys Biol*. 2019; 16:025003. <https://doi.org/10.1088/1478-3975/aaf7b7> PMID: 30537698

49. Chaffer CL, Weinberg RA. A Perspective on Cancer Cell Metastasis. *Science*. 2011; 331(6024):1559–1564. <https://doi.org/10.1126/science.1203543> PMID: 21436443
50. MacLean AL, Hong T, Nie Q. Exploring intermediate cell states through the lens of single cells. *Current Opinion in Systems Biology*. 2018; 9:32–41. <https://doi.org/10.1016/j.coisb.2018.02.009> PMID: 30450444
51. Ta CH, Nie Q, Hong T. Controlling Stochasticity in Epithelial-Mesenchymal Transition Through Multiple Intermediate Cellular States. *Discrete Continuous Dyn Syst Ser B*. 2016; 21(7):2275–2291. <https://doi.org/10.3934/dcdsb.2016047> PMID: 29497351
52. Jolly M, Boareto M, Huang B, Jia D, Lu M, Ben-Jacob E, et al. Implications of the Hybrid Epithelial/Mesenchymal Phenotype in Metastasis. *Front Oncol*. 2015; 5:155. <https://doi.org/10.3389/fonc.2015.00155> PMID: 26258068
53. Jolly M, Tripathi S, Jia D, Mooney S, Celiktas M, Hanash S, et al. Stability of the hybrid epithelial/mesenchymal phenotype. *Oncotarget*. 2016; 7:27067–84. <https://doi.org/10.18632/oncotarget.8166> PMID: 27008704
54. Tian XJ, Zhang H, Zhang J, Xing J. Reciprocal regulation between mRNA and microRNA enables a bistable switch that directs cell fate decisions. *FEBS Letters*. 2016; 590(19):3443–3455. <https://doi.org/10.1002/1873-3468.12379> PMID: 27578189
55. Olsson A, Venkatasubramanian M, Chaudhri VK, Aronow BJ, Salomonis N, Singh H, et al. Single-cell analysis of mixed-lineage states leading to a binary cell fate choice. *Nature*. 2016; 537(7622):698–702. <https://doi.org/10.1038/nature19348> PMID: 27580035
56. Wagner C, Kiefhaber T. Intermediates can accelerate protein folding. *Proceedings of the National Academy of Sciences*. 1999; 96(12):6716–6721. <https://doi.org/10.1073/pnas.96.12.6716>
57. Lo NA, de JSA, Khoueiry R, Cho D, Ferlazzo G, Perini I, et al. PDGFR $\alpha$ <sup>+</sup> Cells in Embryonic Stem Cell Cultures Represent the In Vitro Equivalent of the Pre-implantation Primitve Endoderm Precursors. *Stem Cell Reports*. 2017; 8:318–333. <https://doi.org/10.1016/j.stemcr.2016.12.010>
58. Hayashi K, de Sousa Lopes SMC, Tang F, Lao K, Surani MA. Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. *Cell Stem Cell*. 2008; 3(4):391–401. <https://doi.org/10.1016/j.stem.2008.07.027> PMID: 18940731
59. Stumpf PS, Smith RCG, Lenz M, Schuppert A, Müller FJ, Babbie A, et al. Stem Cell Differentiation as a Non-Markov Stochastic Process. *Cell Systems*. 2017; 5(3):268–282.e7. <https://doi.org/10.1016/j.cels.2017.08.009> PMID: 28957659
60. Liu Y, Xue M, Du S, Feng W, Zhang K, Zhang L et al. Competitive endogenous RNA is an intrinsic component of EMT regulatory circuits and modulates EMT. *Nat Commun*. 2019; 10(1):1637. <https://doi.org/10.1038/s41467-019-09649-1> PMID: 30967542