

Supporting information for: Quantifying the landscape and kinetic paths for epithelial-mesenchymal transition from a core circuit

Chunhe Li^{1,2,3,*}, Tian Hong^{1,2}, Qing Nie^{1,2,*}

1 Department of Mathematics,

2 Center for Complex Biological Systems, University of California, Irvine, Irvine, CA 92697, USA

3 Shanghai Center for Mathematical Sciences, Fudan University, Shanghai 200433, China

* E-mail: chunhe.lee@gmail.com, qnie@math.uci.edu

Population-level Cell State Transition Model

From the MFPT we acquired (Fig. 5 in the main text), we estimated the transition rate (inverse of the MFPT) between different attractors (cell types). We further constructed a Markov cell state transition model, to calculate the population evolutions for cells. With n_1, n_2, n_3 denoting the number of cells for three different cell types, we wrote down the state transition model as three ODEs:

$$\begin{aligned}\frac{dn_1}{dt} &= v_1 - (k_{12} + k_{13}) * n_1 + k_{21} * n_2 + k_{31} * n_3 \\ \frac{dn_2}{dt} &= v_2 - (k_{21} + k_{23}) * n_2 + k_{12} * n_1 + k_{32} * n_3 \\ \frac{dn_3}{dt} &= v_3 - (k_{31} + k_{32}) * n_3 + k_{13} * n_1 + k_{23} * n_2\end{aligned}\tag{1}$$

which determine the temporal evolution of the cell population starting from an initial condition. Here, k_{ij} represents the transition rate from state i to state j (the proportion of cells that switch from state i to state j per unit time), which is calculated from: $k_{ij} = 1/MFPT_{ij}$. v_i indicates the proliferation rate of cells for cell state i . For the case of no population selection pressure, $v_i = 0$ for all i . The proliferation rate v_i can be change to be larger than 0 to consider cells having a selective advantage. In this work, we only explored the case for no population selection pressure ($v_i = 0$ for all i). By solving the above ODEs, we obtained the steady state proportions of cells for three cell types, at different SNAIL level (Fig. 5D). The population-level cell state transition model confers a way to explore the temporal evolution as well as the steady state distribution of cell populations, and to make comparisons with experiments directly. It can be harnessed to investigate the effects of different treatments on the cell fate determination, e.g. to simulate the growth or extinction of cancer cell populations under different interventions (making perturbations to certain genes or microRNA).

The landscape and the population model predict that as the SNAIL signal is strengthened, the steady state proportion for E state cells and P state cells decreases and the steady proportion for M state cells increases. This prediction is consistent with the landscape results (Fig. 3), and can be tested experimentally.

Hamilton-Jacobian Approach for Path Integral

From the path integral formalism, we can evaluate the weights of the kinetic paths. The most probable trajectory can be obtained when the transition action $S(x)$ is minimized directly. The Lagrangian can be written as [1–4]:

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

and the generalized momentum can be written out as: $\mathbf{P}(\mathbf{x}) = \frac{\partial L}{\partial \dot{\mathbf{x}}} = \frac{1}{2D}(\dot{\mathbf{x}} - \mathbf{F}(\mathbf{x}))$. In the dynamic systems, the Hamiltonian of the system has the form:

$$H(\mathbf{x}) = -L(\mathbf{x}) + \mathbf{P}(\mathbf{x}) \cdot \dot{\mathbf{x}} = E_{eff} \quad (3)$$

From the above equation, we have $\frac{1}{4D}\dot{\mathbf{x}}^2 - V(\mathbf{x}) = E_{eff}$ and $|\dot{\mathbf{x}}| = \sqrt{4D(E_{eff} + V(\mathbf{x}))}$. After substituting Eq. S3 into the action, we obtain $S(\mathbf{x}) = \int (\mathbf{P}(\mathbf{x}) \cdot \dot{\mathbf{x}} - H(\mathbf{x}))dt$. We can find that the action characterizing the weights of the paths depends on the values of the Hamiltonian. Specific values of the Hamiltonian are corresponding to specific values of the final time T . For a fixed Hamiltonian, a corresponding optimal path exits when the action $S(\mathbf{x})$ is minimized.

According to the least action principle, if the Hamiltonian of the system is constant, the variation of the action, for given initial and final coordinates and initial and final time, is zero. Giving a variation of the final time T and leaving the initial and the final coordinates fixed, we have $\delta S = -H\delta t$. For a constant Hamiltonian, $\delta S = -E\delta t$. We define $S_0 = \int \mathbf{P}(\mathbf{x}) \cdot \dot{\mathbf{x}}dt$, since $S(\mathbf{x}) = \int (\mathbf{P}(\mathbf{x}) \cdot \dot{\mathbf{x}} - H(\mathbf{x}))$. We find $\delta S_0 = 0$. Thus, the action S_0 is minimized in terms of all the paths that satisfy the constant Hamiltonian and passing through the final point at any instant.

When it goes to the multidimensional questions, the action depends not only the initial and final coordinates but also on the initial and final time. In the HJ framework, we can transform the formulations into a different representation in x space: $S_0 = S_{HJ}(\mathbf{x}) = \int \sum_i \frac{1}{2D}(\dot{\mathbf{x}}_i - \mathbf{F}_i)dx_i = \int \sum_i p_i(\mathbf{x})dx_i$. Here p_i is the associated momentum. Now the action only depends on the initial and final coordinates. This action can be further simplified and is equivalent to a line integral along a particular one dimensional path l so that $S_{HJ}(\mathbf{x}) = \int \sum_i p_i(\mathbf{x})dx_i = \int p_l dl$ where $p_l = \sqrt{(E_{eff} + V(\mathbf{x}))/D} - \frac{1}{2D}F_l$. Therefore, the formulism is switched from the time-dependent to the Hamiltonian-dependent (HJ) description [1-4]. The dominant path connection given initial and final states is obtained by minimizing the action in the HJ representation $S_{HJ} = \int_{x_i}^{x_f} (\sqrt{(E_{eff} + V(\mathbf{x}))/D} - \frac{1}{2D}F_l)dl$, where dl is an infinitesimal displacement along the path trajectory. E_{eff} is a free parameter that determines the total time elapsed during the transition.

In this work, for simplification we chose $E_{eff} = -V_{min}(x)$, which is the effective potential by minimizing $V(\mathbf{x})$, and corresponding to the longest kinetic time. Eventually, the optimal paths were obtained by minimizing the discrete target function:

$$S_{HJ} = \sum_n^{N-1} \left(\sqrt{(E_{eff} + V(n))/D} - \frac{1}{2D}F_l(n) \right) \Delta l_{n,n+1} + \lambda P \quad (4)$$

where

$$\begin{aligned} P &= \sum_i^{N-1} (\Delta l_{i,i+1} - \langle \Delta l \rangle)^2 \\ (\Delta l)_{n,n+1}^2 &= \sum_i (\mathbf{x}_i(n+1) - \mathbf{x}_i(n))^2 \\ F_l(n) &= \sum_i \mathbf{F}_i(\mathbf{x}(n))(\mathbf{x}_i(n+1) - \mathbf{x}_i(n)) / \Delta l_{n,n+1} \\ V(n) &= \sum_i \left(\frac{1}{4D} \mathbf{F}^2(\mathbf{x}_i) + \frac{1}{2} \sum_j \frac{\partial \mathbf{F}_j(\mathbf{x}_i)}{\partial \mathbf{x}_j} \right) \end{aligned} \quad (5)$$

Here, $\Delta l_{n,n+1}$ is the Euclidean measure of the n th elementary path step, and P is a penalty function keeping all the length elements close to their average, which becomes irrelevant in the continuum limit. The minimization of the discrete HJ effective action was performed by applying a simulated annealing algorithm or the conjugate gradient algorithm. In this work, we chose the discrete steps n as 20.

Relationship between Diffusion Equation and Master Equation

The Diffusion equation (Fokker-Planck Equation) is the probability evolution equation for continuous variables, whereas the Master equation describes the probability evolution for discrete variables.

A Master equation describing the probability evolution can be represented in discrete state space: $\frac{\partial P(\mathbf{x}, t | \mathbf{x}_0, t_0)}{\partial t} = \sum_{j=1}^M [a_j(\mathbf{x} - \nu_j)P(\mathbf{x} - \nu_j, t | \mathbf{x}_0, t_0) - a_j(\mathbf{x})P(\mathbf{x}, t | \mathbf{x}_0, t_0)]$. Here P is the probability at molecular number vector \mathbf{x} and time t , given initial time t_0 and molecular number \mathbf{x}_0 . M is the number of reactions, a_j is the reaction rates and ν is the state vector for molecular number changes caused by one reaction. When the molecular number is sufficiently large, the master equation is equivalent to the Langevin equation or diffusion equation (Fokker-Planck Equation) in continuous representation. Here the variables are continuous concentrations, rather than discrete molecular number. The Langevin equation describing the motion of the system can be written as: $\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}) + \mathbf{G} \cdot \Gamma$, where \mathbf{G} is a tensor representing the spatial (concentration) dependent part of the noise $G_{ij}(\mathbf{x}) = v_{ji} \sqrt{a_j(\mathbf{x})}$ ($i=1,2,\dots,N, j=1,2,\dots,M$) [5] and Γ is a vector representing gaussian white noise corresponding to the time dependent part of the noise, which is defined as $\langle \Gamma_j(t) \rangle = 0$ and $\langle \Gamma_i(t) \Gamma_j(t') \rangle = 2\delta_{ij} \delta(t - t')$ ($\delta_{ij} = 1$ for $i = j$, and $\delta_{ij} = 0$ for $i \neq j$).

The Fokker-Planck or diffusion equation characterizes the continuous description of the intrinsic statistical fluctuations [5]:

$$\frac{\partial P(\mathbf{x}, t)}{\partial t} = - \sum_{i=1}^N \frac{\partial}{\partial x_i} F_i(\mathbf{x}) P(\mathbf{x}, t) + \frac{1}{2} \sum_{i=1, i'=1}^N \frac{\partial^2}{\partial x_i \partial x_{i'}} D_{ii'}(\mathbf{x}) P(\mathbf{x}, t)$$

Here \mathbf{x} stands for the set $\{x_i\} (i = 1, 2, 3, \dots, N)$, N is the number of species, and M is the number of reactions. F , D and G are separately defined as: $F_i(\mathbf{x}) = \sum_{j=1}^M v_{ji} a_j(\mathbf{x}) (i = 1, 2, \dots, N)$, $D_{ii'}(\mathbf{x}) = \sum_{j=1}^M v_{ji} v_{ji'} a_j(\mathbf{x}) (i, i' = 1, 2, \dots, N)$, and $G_{ij}(\mathbf{x}) = v_{ji} \sqrt{a_j(\mathbf{x})}$ ($i=1,2,\dots,N, j=1,2,\dots,M$) [5]

Therefore, a Master equation for the discrete description of dynamical system can be transformed to a corresponding probability diffusion equation for the continuous description when the molecular number is sufficiently large.

References

1. Faccioli P, Sega M, Pederiva F, Orland H (2006) Dominant pathways in protein folding. *Phys Rev Lett* 97: 1-4.
2. Wang J, Zhang K, Lu H, Wang E (2005) Quantifying kinetic paths of protein folding. *Biophysical journal* 89: 1612–1620.
3. Wang J, Zhang K, Wang EK (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: 125103:1-13.
4. Wang J, Zhang K, Xu L, Wang EK (2011) Quantifying the waddington landscape and biological paths for development and differentiation. *Proc Natl Acad Sci USA* 108: 8257-8262.
5. Gillespie DT (2000) The chemical langevin equation. *The Journal of Chemical Physics* 113: 297–306.