RESEARCH ARTICLE

[1]Research Institute of Intelligent Complex Systems, Fudan University, Shanghai 200433, China; [2]International Research Center for Neurointelligence, The University of Tokyo Institutes for Advanced Study, The University of Tokyo, Tokyo 113-0033, Japan; [3]Key Laboratory of Mathematics and Its Applications (LMAM) and School of Mathematical Sciences, Peking University, Beijing 100871, China; [4]Key Laboratory of Systems Biology, Shanghai Institute of Biochemistry and Cell Biology, Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences, Shanghai 200031, China; [5]Key Laboratory of Systems Health Science of Zhejiang Province, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Hangzhou 310024, China; [6]School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China and [7]Guangdong Institute of Intelligence Science and Technology, Zhuhai 519031, China

*Corresponding authors. E-mails: kaihara@g.ecc.utokyo.ac.jp; tieli@pku.edu.cn; lnchen@sibs.ac.cn

INFORMATION SCIENCE

Energy landscape decomposition for cell differentiation with proliferation effect

Jifan Shi [1,2], Kazuyuki Aihara[2,*], Tiejun Li[3,*] and Luonan Chen [4,5,6,7,*]

ABSTRACT

Complex interactions between genes determine the development and differentiation of cells. We establish a landscape theory for cell differentiation with proliferation effect, in which the developmental process is modeled as a stochastic dynamical system with a birth-death term. We find that two different energy landscapes, denoted U and V, collectively contribute to the establishment of non-equilibrium steady differentiation. The potential U is known as the energy landscape leading to the steady distribution, whose metastable states stand for cell types, while V indicates the differentiation direction from pluripotent to differentiated cells. This interpretation of cell differentiation is different from the previous landscape theory without the proliferation effect. We propose feasible numerical methods and a mean-field approximation for constructing landscapes U and V. Successful applications to typical biological models demonstrate the energy landscape decomposition's validity and reveal biological insights into the considered processes.

Keywords: energy landscape, cell differentiation, stochastic systems

INTRODUCTION

Cell differentiation makes the biological world rich and colorful. Modeling and understanding the dynamics of gene regulatory networks (GRNs) are essential to explore the underlying mechanism of cell differentiation [1]. Since Waddington proposed his seminal metaphor of the epigenetic landscape [2,3], differentiation has been intuitively described as a ball rolling down a surface, i.e. the energy landscape.

There are mainly two approaches to construct the energy landscape for biosystems. One is the data-based approach, which tries to identify cluster/cell types, differentiation trajectories, pseudotime and cell pluripotency from the experimental data, especially single-cell RNA sequencing (scRNA-seq) data [4–8]. Various algorithms have been proposed from the theory of graph [9,10], entropy [6,11,12] and dynamical systems [13–16]. Among these data-based methods, the population balance analysis (PBA) [13–15] and landscape of differentiation dynamics (LDD) [16] especially involve cell proliferation and death rates, which generalize the hypothesis of cell differentiation from equilibrium to non-equilibrium steady processes. The other approach is the model-based approach, which tries to build

dynamic equations from the GRN, analyze the dynamical behavior of the system and then construct energy landscapes using numerical simulations. Wang *et al.* proposed a practical framework for constructing energy landscapes for biosystems without cell birth and death effects [17–19]. They defined the landscape as $U(\boldsymbol{x}) = -\epsilon \log P_{ss}(\boldsymbol{x})$, where $\boldsymbol{x}$ is the gene expression vector, $P_{ss}(\boldsymbol{x})$ is the steady probability density function (PDF) of the system and ϵ is related to the amplitude of small intrinsic noise. By changing parameter values to control the biological process, the landscape pattern varies, which gives an intuitive description of cell-type locations and transition probability between clusters. This framework has been widely used in modeling the budding yeast cell cycle [17], human stem cell fate [19] and *Caenorhabditis elegans* ageing [20]. Without the cell proliferation effect, however, the direction of differentiation is usually not intrinsic in Wang's landscape but controlled by manually setting parameters. Beyond the topics above, there are also many other interesting works related to the energy landscape theory [21–27].

In this study, we follow the model-based approach but consider the birth and death rates

(BDRs) of cells in differentiation, which is inspired by the PBA and LDD. We show that there are two important energy landscapes to describe differentiation dynamics, which are denoted $U(x)$ and $V(x)$ and are also computable using the model. The metastable states in $U(x)$ represent cell types, and the value of $V(x)$ implicates pluripotency. The negative gradient of $V(x)$ shows the direction of differentiation. Taking BDRs into account is essential for constructing the pluripotency landscape $V(x)$. Here, we explain our theory of the energy landscape decomposition (ELD) in cell differentiation with proliferation effect. Numerical algorithms for constructing energy landscapes and the mean-field approximation (MFA) in high-dimensional cases are also proposed. We use three examples to show the application of the ELD, where $U(x)$ and $V(x)$ intuitively explain the processes of cell differentiation.

RESULTS

Modeling cell differentiation at the population level

Motivated by Weinreb *et al.* [13] and Briggs *et al.* [15], we modeled cell development through a stochastic dynamical system with the birth-death term. We denote gene expression levels using $x \in \mathbb{R}^n$. During cell differentiation, the cell population is quantified using a PDF $p(x, t)$, whose evolution follows a generalized form of the well-known Fokker-Planck equation (FPE), i.e.

$$\frac{\partial p(x, t)}{\partial t} = -\nabla \cdot (b(x)p(x, t)) + \epsilon \Delta p(x, t) + R(x)p(x, t), \qquad (1)$$

with an initial PDF $p(x, t = 0)$ at $t = 0$, where $b(x)$ represents the biological interactions between genes, ϵ is a parameter standing for the noise amplitude and $R(x)$ stands for the net BDR of cells at x. In a practical application, $b(x)$ is usually modeled using either activated or inhibited Hill functions, according to the considered GRN. Here $R(x) > 0$ indicates the proliferation of cells, while $R(x) < 0$ indicates the death of cells. When $R(x) \equiv 0$, (1) reduces to the standard FPE without any cell proliferation effect. To ensure a non-explosive and non-degenerative steady PDF $p_{ss}(x)$ for (1), we require the following condition as a constraint for $R(x)$:

$$\int R(x)p_{ss}(x) \, dx = 0. \qquad (2)$$

It is necessary to have a non-trivial $R(x)$ to characterize the population balance of the cell birth and death in the steady state.

Modeling cell differentiation at the single-cell level

The population-level dynamics in (1) can also be viewed as the probabilistic interpretation at the single-cell level. We consider a cell ω with gene expression $X_t(\omega)$ starting from Y_0 at $t = 0$ and propose a weighted stochastic dynamics in the Itô sense as

$$dX_t(\omega) = b(X_t(\omega)) \, dt + \sqrt{2\epsilon} \, dW_t(\omega), \qquad (3a)$$

$$X_t|_{t=0} = Y_0(\omega), \qquad (3b)$$

$$d\rho_t(\omega) = R(X_t(\omega))\rho_t(\omega) \, dt, \qquad (3c)$$

$$\rho_t|_{t=0} = 1, \qquad (3d)$$

where Y_0 is distributed according to $p_0(x)$, W_t is a standard Brownian motion with independent components and $\rho_t(\omega)$ stands for a time-varying weight for cell ω. The connection between (1) and (3) is represented as

$$p(x, t) = \mathbb{E}\{\rho_t(\omega)\delta(x - X_t(\omega))\}, \qquad (4)$$

where δ is the Dirac delta function and the expectation is taken over all the possible trajectories ω (see the online supplementary material for details). The weight $\rho_t(\omega)$ increases at t when $R(X_t(\omega)) > 0$, which corresponds to cell proliferation, while it decreases at t when $R(X_t(\omega)) < 0$, which corresponds to cell death. If $R(x) \equiv 0$, the weight $\rho_t(\omega)$ is a constant for every cell and the system reduces to the case without any cell proliferation, which is discussed by Wang *et al.* [17,18].

Energy landscape decomposition

Based on the above dynamical modeling of cell development, we focus on the construction of the landscapes using the model-based approach, i.e. we assume that $b(x)$, ϵ, and $R(x)$ are known a priori.

We denote by $P_U(x)$ the steady PDF with the known BDR $R(x)$ and by $P_0(x)$ the steady PDF with $R(x) \equiv 0$, for (1). Note that the notation $P_0(x)$ is different from the initial distribution $p(x, t = 0)$. Then, two energy landscapes are constructed as

$$U(x) = -\epsilon \log P_U(x), \qquad (5)$$

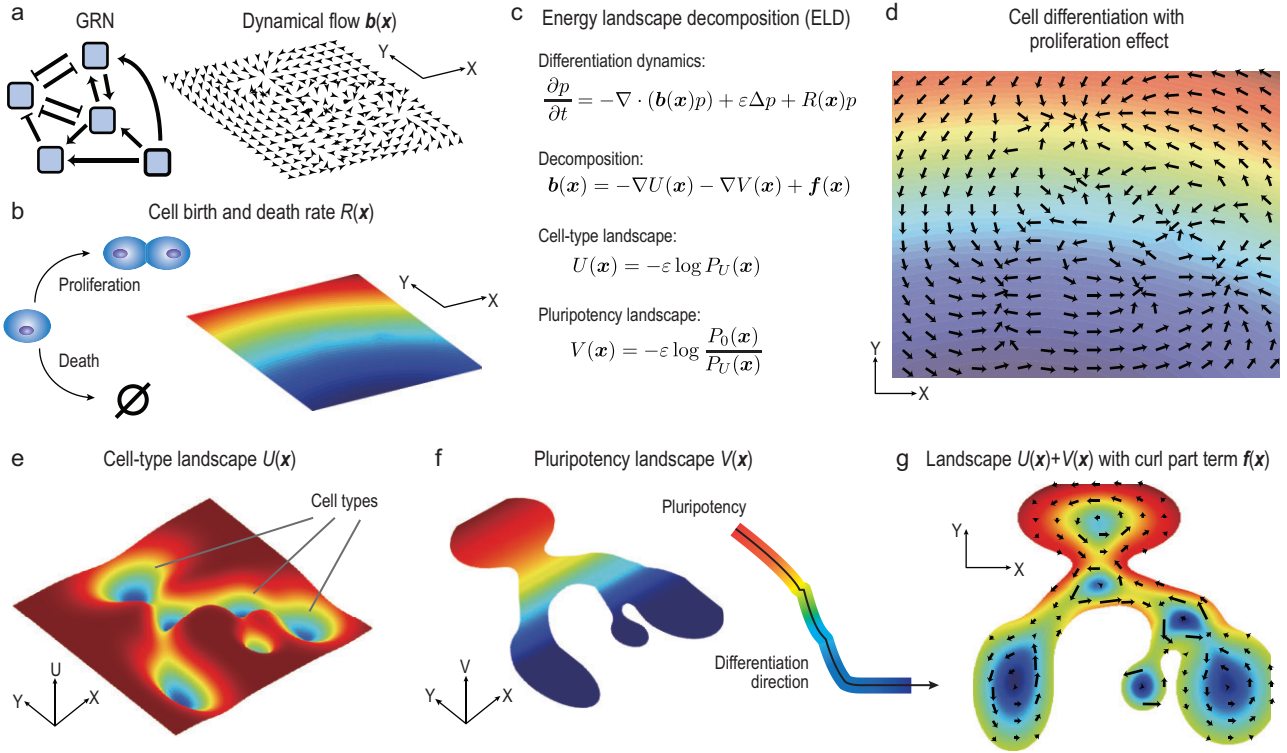**Figure 1.** An illustration of the ELD framework. (a) The GRN, which reins cell differentiation dynamics. Dynamical flow $\boldsymbol{b}(\boldsymbol{x})$ can be modeled using the GRN, and its vector field is drawn in an illustrative space (X, Y). (b) Cells can proliferate or die at a natural birth and death rate $R(\boldsymbol{x})$. Proliferation with $R(\boldsymbol{x}) > 0$ is shown in red, while cell death with $R(\boldsymbol{x}) < 0$ is in blue (jet colormap). (c) The ELD theory in this study. The dynamics $(\boldsymbol{b}(\boldsymbol{x}), R(\boldsymbol{x}))$ can be characterized using two potential terms (energy landscapes) $U(\boldsymbol{x})$ and $V(\boldsymbol{x})$, and a curl part term $\boldsymbol{f}(\boldsymbol{x})$. We define $U(\boldsymbol{x})$ and $V(\boldsymbol{x})$ using two steady PDFs, denoted $P_U(\boldsymbol{x})$ and $P_0(\boldsymbol{x})$. (d) The combination of dynamical flow $\boldsymbol{b}(\boldsymbol{x})$ (arrows) and birth and death rate $R(\boldsymbol{x})$ (jet colored background). (e–g) Illustrations for the cell-type landscape $U(\boldsymbol{x})$, the pluripotency landscape $V(\boldsymbol{x})$ and the curl part term $\boldsymbol{f}(\boldsymbol{x})$, respectively. Metastable states of $U(\boldsymbol{x})$ shown in (e) stand for different cell types. The values of $V(\boldsymbol{x})$ shown in (f) indicate the pluripotency, with its negative gradient depicting the differentiation direction. The color bar in (f) is calculated using the mean value of V along the X axis, which displays an intuitive global differentiation direction. In (g), landscape $U(\boldsymbol{x}) + V(\boldsymbol{x})$ is plotted as the background, while the curl part term $\boldsymbol{f}(\boldsymbol{x})$ is indicated by the vectors.

which drives the system or cells to the steady distribution, and

$$V(\boldsymbol{x}) = -\epsilon \log P_0(\boldsymbol{x}) - (-\epsilon \log P_U(\boldsymbol{x}))$$
$$= -\epsilon \log(P_0(\boldsymbol{x})/P_U(\boldsymbol{x})), \quad (6)$$

which quantifies the change of the potential caused by $R(\boldsymbol{x})$, i.e. the influence induced by cell proliferation and death. The metastable basins in landscape $U(\boldsymbol{x})$ indicate cell types, and the depth of each basin characterizes its stability. The values of V depict the pluripotency and its negative gradient field describes the differentiation direction. From a stem cell state to a differentiated cell state, V decreases gradually. The potential U is the cell-type landscape and V is the pluripotency landscape. Besides the two potential functions, the remaining term $\boldsymbol{f}(\boldsymbol{x})$ is defined as

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{b}(\boldsymbol{x}) + \nabla U(\boldsymbol{x}) + \nabla V(\boldsymbol{x}), \quad (7)$$

which is the curl part that describes the non-gradientness of the considered dynamics. The $\boldsymbol{f}(\boldsymbol{x})$ term satisfies

$$\nabla \cdot (\boldsymbol{f}(\boldsymbol{x}) P_0(\boldsymbol{x})) = 0, \quad (8)$$

which corresponds to the divergence-free condition of the curl flux $\boldsymbol{J}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x}) P_0(\boldsymbol{x})$ defined in [17,18].

In summary, we propose an energy decomposition for the differentiation dynamics characterized by the pair $(\boldsymbol{b}(\boldsymbol{x}), R(\boldsymbol{x}))$ in our ELD framework as

$$\boldsymbol{b}(\boldsymbol{x}) = -\nabla U(\boldsymbol{x}) - \nabla V(\boldsymbol{x}) + \boldsymbol{f}(\boldsymbol{x}). \quad (9)$$

The potentials U and V, together with the curl part $\boldsymbol{f}$, help us have a deep understanding of cell differentiation and cell subtypes along the differentiation pathway. Figure 1 presents an illustration of the whole ELD procedure.

The connection between the ELD framework and the existing landscape theory is as follows. In the case of no birth and death of cells, i.e. $R(\boldsymbol{x}) \equiv 0$, (9) reduces to Wang's landscape decomposition [17–19] with $P_U(\boldsymbol{x}) \equiv P_0(\boldsymbol{x})$ and $V(\boldsymbol{x}) \equiv 0$; if the system is of gradient type with $\boldsymbol{b}(\boldsymbol{x}) = -\nabla F(\boldsymbol{x})$, (9) reduces to the landscape decomposition discussed in PBA and LDD [13,16] with $\boldsymbol{f}(\boldsymbol{x}) \equiv 0$.

In a case that only $p_{ss}(\boldsymbol{x})$ is known, PBA and LDD define the energy landscape V for gradient systems by solving the equation

$$\nabla \log p_{ss}(\boldsymbol{x}) \cdot \nabla V(\boldsymbol{x}) + \Delta V(\boldsymbol{x}) = -R(\boldsymbol{x}). \tag{10}$$

with proper boundary conditions. However, for a non-gradient system, we show that (10) is not always valid, unless the curl part $\boldsymbol{f}$ satisfies $\boldsymbol{f}(\boldsymbol{x}) \perp \nabla V(\boldsymbol{x})$ almost everywhere (see the online supplementary material for details). In this sense, we provide a more general definition of V using (6). More detailed discussion about the ELD can be found in the online supplementary material.

Numerical construction of energy landscapes

According to the definitions of potentials $U(\boldsymbol{x})$ in (5) and $V(\boldsymbol{x})$ in (6), we can compute the energy landscapes once the steady distributions $P_U(\boldsymbol{x})$ and $P_0(\boldsymbol{x})$ are obtained. Next we discuss the numerical methods for the low-dimensional case and the high-dimensional case separately.

Solving FPE for the low-dimensional case

For the low-dimensional case (usually in less than three dimensions), we used the finite difference or finite element method to solve (1) numerically until the steady state. For a small parameter ϵ, the spatial scale h of the grids in traditional methods is needed to be set as $h \ll \epsilon$ to resolve the dynamical behavior of such a convection-dominated equation, which would be impractical. We utilized the streamline diffusion method [28] that can avoid the numerical oscillations even when $h > \epsilon$. The detailed procedure of the streamline diffusion method is described in the online supplementary material in a two-dimensional case. It is also applied in our numerical examples in the Results section.

Mean-field approximation in the high-dimensional case

For the high-dimensional case, it is not feasible to directly solve (1). Thus, we start from the single-cell dynamics (3) and propose an MFA approach to reduce the computational complexity from exponential to polynomial time. Overall, the potentials P_U and P_0 are approximated by the Gaussian mixtures in the MFA, i.e.

$$p(\boldsymbol{x}) = \sum_{k=1}^{K} \rho^{(k)} p_{(k)}(\boldsymbol{x}; \boldsymbol{\mu}_{(k)}, \boldsymbol{\Sigma}_{(k)}), \tag{11}$$

where K is the number of components obtained by the counts of stable states in the deterministic dynamics $\mathrm{d}\boldsymbol{x}_t / \mathrm{d}t = \boldsymbol{b}(\boldsymbol{x}_t)$, $p_{(k)}(\boldsymbol{x}; \boldsymbol{\mu}_{(k)}, \boldsymbol{\Sigma}_{(k)})$ is the kth Gaussian component with mean $\boldsymbol{\mu}_{(k)}$ and covariance $\epsilon \boldsymbol{\Sigma}_{(k)}$, and $\rho^{(k)}$ is the mixture weight. These approximations are obtained through asymptotics of (3) with respect to the small parameter ϵ.

Short time asymptotics. We denote by Ω_k the kth attractive basin for the $\boldsymbol{X}_t$ dynamics in (3). In the $O(1)$ timescale (or the transition timescale in Ω_k), there is no state hopping and we derive the approximation $\boldsymbol{X}_t \approx \boldsymbol{\mu}_t^{(k)} + \sqrt{\epsilon}\boldsymbol{Z}_t^{(k)}$, where $\boldsymbol{Z}_t^{(k)} \sim N(0, \boldsymbol{\Sigma}_t^{(k)})$ when the initial $\boldsymbol{Y}_0 \in \Omega_k$. The mean and covariance satisfy the equations

$$\frac{\mathrm{d}\boldsymbol{\mu}_t^{(k)}}{\mathrm{d}t} = L_1\left(\boldsymbol{\mu}_t^{(k)}, \boldsymbol{\Sigma}_t^{(k)}, \epsilon\right), \tag{12a}$$

$$\frac{\mathrm{d}\boldsymbol{\Sigma}_t^{(k)}}{\mathrm{d}t} = L_2\left(\boldsymbol{\mu}_t^{(k)}, \boldsymbol{\Sigma}_t^{(k)}, \epsilon\right), \tag{12b}$$

where L_1 and L_2 are functions derived from the Taylor expansion of $\boldsymbol{b}(\boldsymbol{x})$ around $\boldsymbol{\mu}_t^{(k)}$ until $O(\epsilon)$ terms. The details are shown in the online supplementary material. The parameters $(\boldsymbol{\mu}_{(k)}, \boldsymbol{\Sigma}_{(k)})$ are steady states of (12). Especially, $L_1(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \epsilon) = \boldsymbol{b}(\boldsymbol{\mu})$ in the $O(1)$ approximation (i.e. $\epsilon = 0$) and we recover the classical MFA [17,29].

Long time asymptotics. The determination of the mixture weights $\{\rho^{(k)}\}$ is designed to take into account the basin transitions in a longer time scale like $\log(t) \gtrsim O(1/\epsilon)$. In this regime, the diffusion dynamics of $\boldsymbol{X}_t$ is upscaled to a continuous-time Markov chain, in which the Arrhenius-type transition rates depend on the energy barriers between the corresponding attraction basins [30,31]. We assume that the upscaled transition rate matrix is $\boldsymbol{Q}$ and define $R_k = \int R(\boldsymbol{x}) p_{(k)}(\boldsymbol{x}, \boldsymbol{\mu}_{(k)}, \boldsymbol{\Sigma}_{(k)}) \mathrm{d}\boldsymbol{x}$ as the average BDR. The evolution of weights $\boldsymbol{\rho}$ with/without a birth-death term $R(\boldsymbol{x})$ shows the asymptotics

$$\frac{\mathrm{d}\boldsymbol{\rho}_U}{\mathrm{d}t} = \boldsymbol{Q}^T\boldsymbol{\rho}_U + \boldsymbol{R}\boldsymbol{\rho}_U, \qquad \frac{\mathrm{d}\boldsymbol{\rho}_0}{\mathrm{d}t} = \boldsymbol{Q}^T\boldsymbol{\rho}_0, \tag{13}$$

where $\boldsymbol{R} = \mathrm{diag}(R_k)$, $\boldsymbol{\rho}_U = (\rho_U^{(k)})$ is the mixture weights for $P_U(\boldsymbol{x})$ and $\boldsymbol{\rho}_0 = (\rho_0^{(k)})$ is the mixture weights for $P_0(\boldsymbol{x})$. As accurate $\boldsymbol{Q}$ is difficult to obtain, we perform a Monte Carlo approximation to steady $\boldsymbol{\rho}_0$ by utilizing the equation

$$\rho_0^{(k)} \approx \lim_{t \to +\infty} \mathbb{E}[\delta(\boldsymbol{X}_t(\omega) \in \Omega_k)], \tag{14}$$

where the trajectories are simulated with a uniform initial distribution on a finite domain at $t = 0$ until a suitable finite ending time $t = T$. Such a choice actually gives the same MFA as proposed in [17,18] without the $R(\boldsymbol{x})$ term. With ergodic assumption,

the steady mixture weights $\rho_U^{(k)}$ for $P_U(\boldsymbol{x})$ are derived as

$$\rho_U^{(k)} \approx \lim_{t \to +\infty} \mathbb{E}[\rho_t(\boldsymbol{X}_t(\omega))\delta(\boldsymbol{X}_t(\omega) \in \Omega_k)]$$

$$= \frac{\rho_0^{(k)} q_k}{q_k - R_k}, \qquad (15)$$

where $q_k = -Q_{kk}$ is the exit rate for state k. Under the constraint equation, (2), and an additional assumption that $1/q_k \propto \rho_0^{(k)}$, an approximation for q_k and thus $\rho_U^{(k)}$ is also obtained (see the online supplementary material for details).

With the mean-field approximations $P_0(\boldsymbol{x})$ and $P_U(\boldsymbol{x})$, we obtain the estimations for the landscapes $U(\boldsymbol{x})$ and $V(\boldsymbol{x})$ using (5) and (6), respectively.

Several detailed remarks need to be made. (i) There is also an MFA to (1), but, for multipotential systems, the MFA to (1) can have different numbers of components with different parameters when approximating $P_0(\boldsymbol{x})$ and $P_U(\boldsymbol{x})$, respectively. That can lead to an odd $V(\boldsymbol{x})$. Thus, one of the advantages of the MFA to (3) is that the numbers of components K and $p_{(k)}(\boldsymbol{x}, \boldsymbol{\mu}_{(k)}, \Sigma_{(k)})$ computed using (12) are independent of the BDR $R(\boldsymbol{x})$, and only mixture weights estimated in (14) and (15) are different. (ii) The MFA to (3) is not suitable for monostable systems, as the mixture weight for trajectories in the only attractive basin is supposed to be the same. For the system with a uni-component, we use the MFA to (3), which is also described in detail in the online supplementary material. For real problems, most systems are multi-stable with several different components. (iii) In (12), the MFA is expanded to $O(\epsilon)$ in functions L_1 and L_2. In Wang's framework [17,18] where $R(\boldsymbol{x}) \equiv 0$, it is only expanded to $O(1)$ to estimate energy landscape $U(\boldsymbol{x})$. In the online supplementary material, we state that it is necessary to have the MFA to $O(\epsilon)$ to obtain a proper approximation of $V(\boldsymbol{x})$. Two analytic examples are also used to validate the necessity.

Numerical examples

In this section, we utilize three examples to show how we applied the ELD framework to describe cell differentiation.

Example 1: two-dimensional drift-diffusion process

For the first example, we use a two-dimensional drift-diffusion process as a basic example for simulating cell differentiation [13,16]. In this example, we take $\boldsymbol{b}(\boldsymbol{x}) = -\nabla F(\boldsymbol{x})$ to be a gradient field with

$$F(\boldsymbol{x}) = \left(\frac{x_1^2}{2} + \frac{x_2}{2}\right)^2 + \frac{(x_2^2 - 1)^2}{2}, \qquad (16)$$
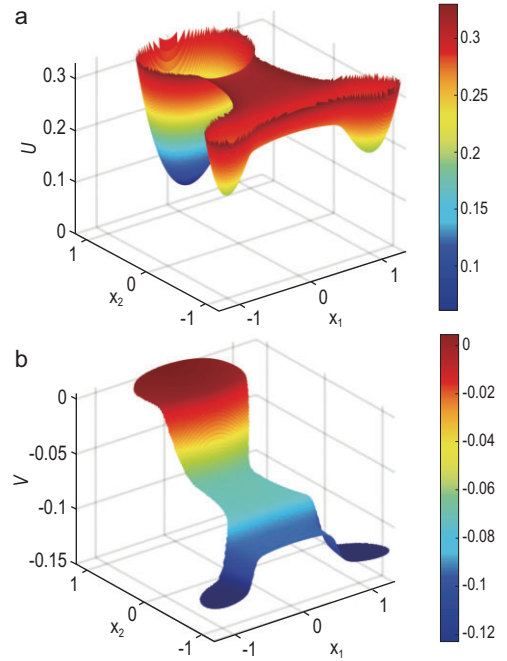

Figure 2. The landscapes for the two-dimensional drift-diffusion process. (a) The landscape *U*. Three metastable states represent three cell types. (b) The landscape *V*. The cells differentiate from the state at a high value of *V* to a lower value.

and set the rate function $R(\boldsymbol{x}) = 0.3[(x_1^2 - 1)^2 + (x_2 + 1)^2 - 1]$, where $\boldsymbol{x} = (x_1, x_2)^T$. The noise parameter is $\epsilon = 0.01$. There exist three stable equilibrium points of the potential $F(\boldsymbol{x})$, i.e. $\boldsymbol{x}_A = (0, \sqrt{3}/2)^T$, $\boldsymbol{x}_B = (-1, -1)^T$ and $\boldsymbol{x}_C = (1, -1)^T$, which correspond to three cell types. According to $R(\boldsymbol{x})$, most cells proliferate around $\boldsymbol{x}_A$, and die around $\boldsymbol{x}_B$ and $\boldsymbol{x}_C$. Thus, cell type A around $\boldsymbol{x}_A$ stands for stem cells with high pluripotency, while cell types B around $\boldsymbol{x}_B$ and C around $\boldsymbol{x}_C$ represent differentiated states. Using the streamline diffusion method to solve (1) with and without $R(\boldsymbol{x})$, we obtain the distributions P_U and P_0, respectively. Furthermore, the energy landscapes U and V are constructed as shown in Fig. 2. Three cell types are easily identified in $U(\boldsymbol{x})$. The effect of the birth and death rate $R(\boldsymbol{x})$ determines the differentiation direction of the process, and the values of $V(\boldsymbol{x})$ characterize the stemness from high to low. Figure S1 within online supplementary material also shows the original potential $F(\boldsymbol{x})$, together with the two-dimensional projections of U and V.

Example 2: the two-gene fate decision system

The second example models a binary cell fate decision controlled by the interaction between two genes [18,32,33]. Figure 3a shows the regulatory relationships between the two genes. It is modeled using a
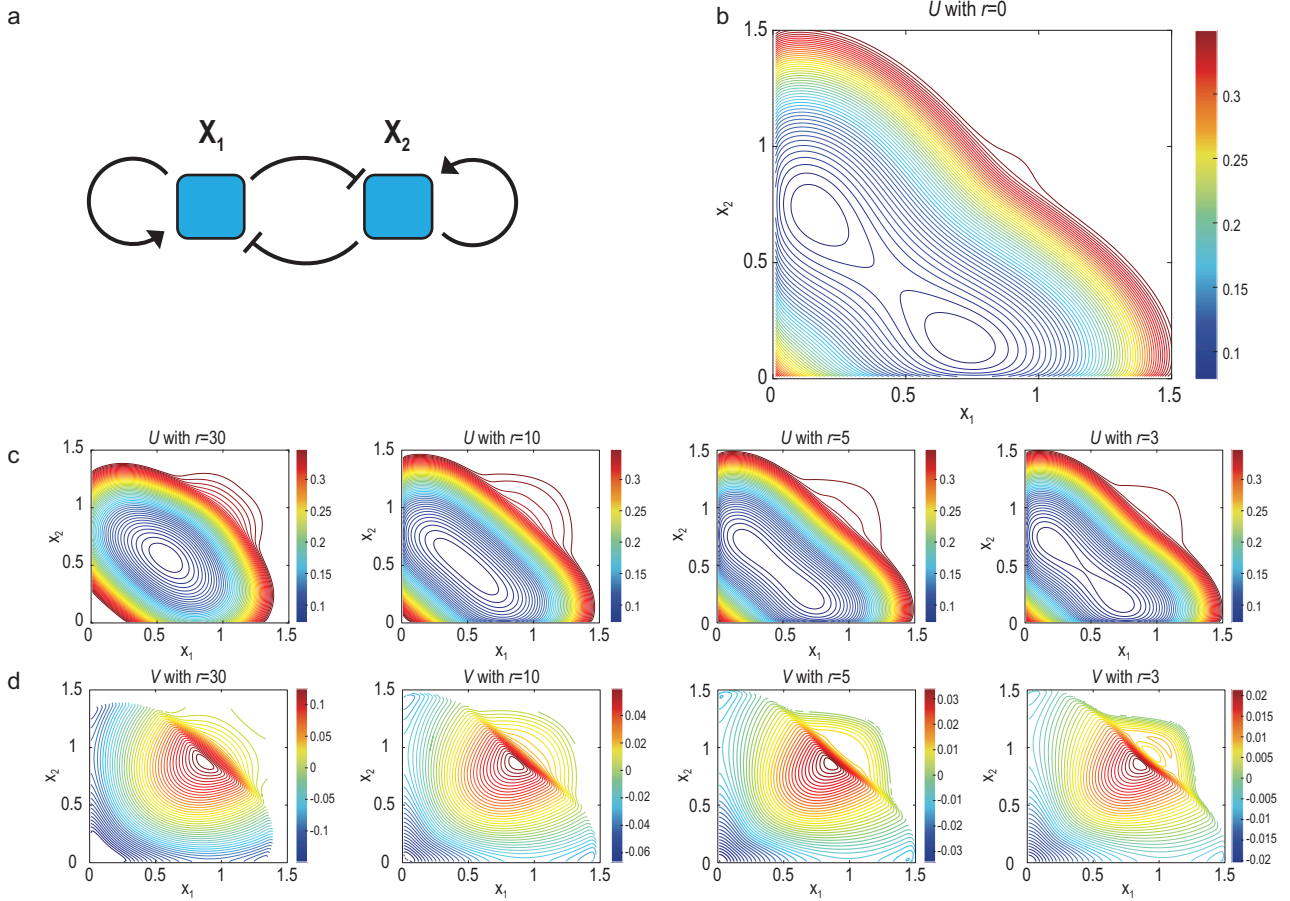
a

b

c

d

Figure 3. (a) The interaction between two genes X_1 and X_2 in the fate decision system. Self-activation and inter-inhibition are observed. (b) The landscape U when the amplitude of the birth and death rate is set as $r = 0$. (c) The landscape U changes when r decreases from 30 to 3. (d) The landscape V changes when r decreases from 30 to 3.

non-gradient Hill dynamics as

$$b(x) = \begin{bmatrix} \frac{\alpha_1 x_1^n}{S^n + x_1^n} + \frac{\beta_1 S^n}{S^n + x_2^n} - k_1 x_1 \\ \frac{\alpha_2 x_2^n}{S^n + x_2^n} + \frac{\beta_2 S^n}{S^n + x_1^n} - k_2 x_2 \end{bmatrix}, \quad (17)$$

where the parameters are set as $\alpha_1 = \alpha_2 = 0.3$, $\beta_1 = \beta_2 = 0.5$, $n = 4$, $k_1 = k_2 = 1$ and $S = 0.5$. Through self-activation and inter-inhibition, the two genes X_1 and X_2 (such as GATA1 and PU.1) are coexpressed in the pluripotent stem cell, and one gene gradually dominates over the other when the system is committed to two different lineages. In Wang's framework, α_1, α_2, β_1 and β_2 vary to control the differentiation [18]. However, in the current viewpoint of cell differentiation, ELD theory claims that, even if these parameters are fixed, the cell fate decision can be adjusted according to the change in BDR $R(x)$. To construct the landscapes U and V, we set the noise amplitude $\epsilon = 0.01$, and the birth and death rate $R(x) = -r[(x_1 - 1)^2 + (x_2 - 1)^2 - 0.5]$. The rate amplitude r is changed from 30 to 0. Figure 3c and d demonstrate how the landscapes change when r decreases, and Fig. 3b

shows the landscape of U when $r = 0$ (the landscape $V \equiv 0$). We obtain the following results through our computations: (i) when the BDR is high ($r = 30$), there is only one cell type characterized by U, where X_1 and X_2 are coexpressed; (ii) this single state splits into two as the amplitude of BDR decreases, i.e. differentiation; (iii) in the two separated cell types when $r = 0$, one gene dominates over the other; (iv) the value of V characterizes the pluripotency of cells, and $-\nabla V$ indicates the differentiation direction. Overall, instead of changing the interaction strength between genes, the BDR term $R(x)$ might also be responsible for explaining cell differentiation, which controls the cell fate decision.

Example 3: T-cell differentiation

The third example is used to describe the T-cell differentiation in high-dimensional space [34]. Four genes x_1 (TCF-1), x_2 (PU.1), x_3 (GATA3) and x_4 (BLC11B) interact with each other through activation or inhibition (see Fig. 4a). The dynamical term $b(x)$ is modeled using Hill functions and the

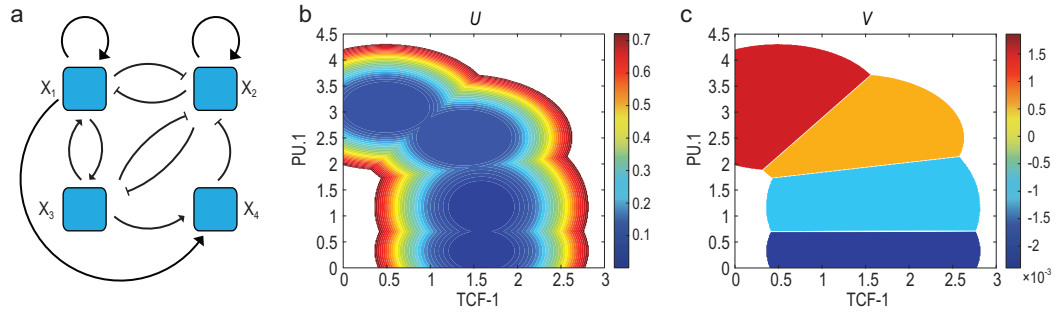**Figure 4.** (a) The gene regulatory network in the T-cell differentiation process. The four genes X_1, X_2, X_3 and X_4 are TCF-1, PU.1, GATA3 and BLC11B, respectively. (b) The landscape U and (c) the landscape V. Metastable states represent the T-cell stages, i.e. ETP/DN1, DN2a, DN2b and DN3. The potential V indicates the pluripotency and differentiation direction.

parameters are listed in the online supplementary material. We set the BDR as $R(x) = 30[4.2 - x_1^2 - (x_2 - 4)^2]$. Using the MFA equation, (11), the landscapes are constructed as shown in Fig. 4b and c. The potential U corresponds to the steady-state landscape with birth-death term R. The four metastable states stand for the four stages of development of T cells (ETP/DN1, DN2a, DN2b and DN3). The potential V in Fig. 4c shows the pluripotency of cells with its value and differentiation direction with its negative gradient field. As applied to MFA, the BDR is also averaged for each Gaussian component, so the variation of $V(x)$ within each metastable state/cell type is much smaller than that shown by U. Accordingly, there exists a sharp variation of V between two adjacent cell types, as shown in Fig. 4c.

Besides the four-dimensional example for T-cell differentiation, we also conducted ELD on a high-dimensional example with 10 variables shown in Sec. IV(D) and Figure S2 of the online supplementary material. These results support the fact that the ELD and MFA are practical in studying the metastable states by energy landscape U and the pluripotency by energy landscape V.

CONCLUSIONS AND DISCUSSIONS

In this paper, we have proposed the ELD framework to describe cell differentiation with proliferation effect. Two energy landscapes, $U(x)$ and $V(x)$, can explain the dynamical behavior of the system during differentiation. The potential U depicts the attractors standing for cell types, while V characterizes the pluripotency and differentiation direction. The consideration of BDR is important to construct V. With an additional V, ELD theory is a generalization of traditional energy landscape theory, and it is a natural realization for Waddington's epigenetic landscape with birth-death terms. The numerical construction of ELD, especially its mean-field approximation in

high-dimensional cases, is introduced. Simulated examples demonstrate the applicability of ELD. However, there are still several issues that need to be further discussed and studied.

First, the ELD framework is different from other models for constructing landscapes for cell differentiation. PBA and LDD are data driven, which are based on the scRNA-seq data and difficult to handle the non-gradient system. Wang's landscape does not consider the birth and death of cells during the differentiation. ELD is model based and analyzes a landscape V caused by cell proliferation and death to display the differentiation direction.

Second, the BDR $R(x)$ in our model is known a priori. For practical cases, R can be estimated from experiments (as in PBA, [13]) or approximated for each cell type (as in LDD, [16]). If only R for cell types matters, which means that R is a constant for each metastable state, the corresponding V will change little within one state while varying sharply between two adjacent states.

Third, the estimated mixture weights $\rho_0^{(k)}$ in (14) are only rough approximations according to the size of attractive basins (percentages of trajectories falling into one meta-stable state). A more rational approach is to get the transition rate matrix $\mathbf{Q}$ between different states [31,35–37]. However, it is difficult to catch rare transitions within the limited time of the simulation. A quick and accurate way to estimate mixture weights is still an open question.

Fourth, constructing landscapes for systems with limit cycles is also possible once a proper BDR is given. According to the approximation in [17], $p_{(k)}(x) = \lim_{s \to +\infty} (1/T) \int_s^{s+T} p(x, \mu_t, \Sigma_t)\, dt$ can be used as the kth component in (11), where T is the period for the limit cycle and $p(x; \mu_t, \Sigma_t)$ is the Gaussian mean-field approximation of the PDF at t by simulating (12). Constructing landscapes for real systems with limit-cycle behavior will be our future work.

Finally, in this paper we used FPE with a BDR term to study the dynamical behavior of the differentiation process in normal cells. For cancer cells that developed from normal cells, the pluripotency and proliferation rates may be restored by external factors. We leave the study of the energy landscape for cancer or tumor cells to future work.

Overall, the energy landscape is a universal concept to characterize the dynamical behavior of a system, and the proposed ELD in this study can help understand systems with proliferation and death, beyond pure reactions. This work can also be applied to model-based and data-based dynamical analyses of various biological systems [38–42].

MATERIALS AND METHODS

Detailed methods are available in the online supplementary material.

SUPPLEMENTARY DATA

Supplementary data are available at *NSR* online.

FUNDING

AUTHOR CONTRIBUTIONS

J.S., T.L., L.C. and K.A. designed the research; J.S. and T.L. performed the research; J.S. analyzed the data; and J.S., T.L., L.C. and K.A. wrote the paper.

Conflict of interest statement. None declared.

REFERENCES

1. Davidson EH. *The Regulatory Genome: Gene Regulatory Networks in Development and Evolution.* New York: Academic Press, 2006.
2. Waddington CH. *Principles of Development and Differentiation.* New York: Macmillan, 1966.
3. Waddington CH. *The Strategy of the Genes.* London: Routledge, 2014.
4. Tang F, Barbacioru C and Wang Y *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* 2009; **6**: 377–82.
5. Saliba AE, Westermann AJ and Gorski SA *et al.* Single-cell RNA-seq: advances and future challenges. *Nucleic Acids Res* 2014; **42**: 8845–60.
6. Shi J, Teschendorff AE and Chen W *et al.* Quantifying Waddington's epigenetic landscape: a comparison of single-cell potency measures. *Brief Bioinformatics* 2020; **21**: 248–61.
7. Cannoodt R, Saelens W and Saeys Y. Computational methods for trajectory inference from single-cell transcriptomics. *Eur J Immunol* 2016; **46**: 2496–506.
8. Saelens W, Cannoodt R and Todorov H *et al.* A comparison of single-cell trajectory inference methods. *Nat Biotechnol* 2019; **37**: 547–54.
9. Trapnell C, Cacchiarelli D and Grimsby J *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* 2014; **32**: 381–6.
10. Haghverdi L, Buettner F and Theis FJ. Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics* 2015; **31**: 2989–98.
11. Teschendorff AE and Enver T. Single-cell entropy for accurate estimation of differentiation potency from a cell's transcriptome. *Nat Commun* 2017; **8**: 15599.
12. Jin S, MacLean AL and Peng T *et al.* scEpath: energy landscape-based inference of transition probabilities and cellular trajectories from single-cell transcriptomic data. *Bioinformatics* 2018; **34**: 2077–86.
13. Weinreb C, Wolock S and Tusi BK *et al.* Fundamental limits on dynamic inference from single-cell snapshots. *Proc Natl Acad Sci USA* 2018; **115**: E2467–76.
14. Tusi BK, Wolock SL and Weinreb C *et al.* Population snapshots predict early haematopoietic and erythroid hierarchies. *Nature* 2018; **555**: 54–60.
15. Briggs JA, Weinreb C and Wagner DE *et al.* The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science* 2018; **360**: eaar5780.
16. Shi J, Li T and Chen L *et al.* Quantifying pluripotency landscape of cell differentiation from scRNA-seq data by continuous birth-death process. *PLoS Comput Biol* 2019; **15**: e1007488.
17. Wang J, Li C and Wang E. Potential and flux landscapes quantify the stability and robustness of budding yeast cell cycle network. *Proc Natl Acad Sci USA* 2010; **107**: 8195–200.
18. Wang J, Zhang K and Xu L *et al.* Quantifying the Waddington landscape and biological paths for development and differentiation. *Proc Natl Acad Sci USA* 2011; **108**: 8257–62.
19. Li C and Wang J. Quantifying cell fate decisions for differentiation and reprogramming of a human stem cell network: landscape and biological paths. *PLoS Comput Biol* 2013; **9**: e1003165.
20. Zhao L and Wang J. Uncovering the mechanisms of Caenorhabditis elegans ageing from global quantification of the underlying landscape. *J R Soc Interface* 2016; **13**: 20160421.
21. Zhou JX, Aliyu M and Aurell E *et al.* Quasi-potential landscape in complex multi-stable systems. *J R Soc Interface* 2012; **9**: 3539–53.

22. Ge H and Qian H. Landscapes of non-gradient dynamics without detailed balance: stable limit cycles and multiple attractors. *Chaos* 2012; **22**: 023140.

23. Huang S. The molecular and mathematical basis of Waddington's epigenetic landscape: a framework for post-Darwinian biology? *BioEssays* 2012; **34**: 149–57.

24. Ao P. Potential in stochastic differential equations: novel construction. *J Phys A-Math Gen* 2004; **37**: L25.

25. Yin L and Ao P. Existence and construction of dynamical potential in nonequilibrium processes without detailed balance. *J Phys A-Math Gen* 2006; **39**: 8593.

26. Zhou P and Li T. Construction of the landscape for multi-stable systems: potential landscape, quasi-potential, A-type integral and beyond. *J Chem Phys* 2016; **144**: 094109.

27. Qian H. Fitness and entropy production in a cell population dynamics with epigenetic phenotype switching. *Quant Biol* 2014; **2**: 47–53.

28. Johnson C. *Numerical Solution of Partial Differential Equations by the Finite Element Method*. New York: Dover Publication Inc., 2012.

29. E W, Li T and Vanden-Eijnden E. *Applied Stochastic Analysis*. Rhode Island: American Mathematical Society, 2019.

30. Gayrard V, Bover A and Eckhoff M *et al.* Metastability in reversible diffusion processes I: sharp asymptotics for capacities and exit times. *J Eur Math Soc* 2004; **6**: 399–424.

31. Hänggi P, Talkner P and Borkovec M. Reaction-rate theory: fifty years after Kramers. *Rev Mod Phys* 1990; **62**: 251.

32. Huang S, Guo YP and May G *et al.* Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Dev Biol* 2007; **305**: 695–713.

33. Zhou JX and Huang S. Understanding gene circuits at cell-fate branch points for rational cell reprogramming. *Trends Genet* 2011; **27**: 55–62.

34. Ye Y, Kang X and Bailey J *et al.* An enriched network motif family regulates multistep cell fate transitions with restricted reversibility. *PLoS Comput Biol* 2019; **15**: e1006855.

35. Pearce P, Woodhouse FG and Forrow A *et al.* Learning dynamical information from static protein and sequencing data. *Nat Commun* 2019; **10**: 5368.

36. Henkelman G and Jónsson H. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J Chem Phys* 2000; **113**: 9978–85.

37. E W, Ren W and Vanden-Eijnden E. String method for the study of rare events. *Phys Rev B* 2002; **66**: 052301.

38. Shi J, Aihara K and Chen L. Dynamics-based data science in biology. *Natl Sci Rev* 2021; **8**: nwab029.

39. Liu X, Chang X and Leng S *et al.* Detection for disease tipping points by landscape dynamic network biomarkers. *Natl Sci Rev* 2019; **6**: 775–85.

40. Chen C, Li R and Shu L *et al.* Predicting future dynamics from short-term time series using an anticipated learning machine. *Natl Sci Rev* 2020; **7**: 1079–91.

41. Yang B, Li M and Tang W *et al.* Dynamic network biomarker indicates pulmonary metastasis at the tipping point of hepatocellular carcinoma. *Nat Commun* 2018; **9**: 678.

42. Chen P, Liu R and Aihara K *et al.* Autoreservoir computing for multistep ahead prediction based on the spatiotemporal information transformation. *Nat Commun* 2020; **11**: 4568.