A Monte Carlo method for in silico modeling and visualization of Waddington' s epigenetic landscape with intermediate details

Zhang, Xiaomeng; Chong, Ket Hing; Zhu, Lin; Zheng, Jie

2020

Zhang, X., Chong, K. H., Zhu, L., & Zheng, J. (2020). A Monte Carlo method for in silico modeling and visualization of Waddington' s epigenetic landscape with intermediate details. *Biosystems*, 198, 104275-. doi:10.1016/j.biosystems.2020.104275

<https://hdl.handle.net/10356/144897>

<https://doi.org/10.1016/j.biosystems.2020.104275>

© 2020 Elsevier B.V. All rights reserved. This paper was published in *Biosystems* and is made available with permission of Elsevier B.V.

Downloaded on 26 Apr 2023 08:04:18 SGT

A Monte Carlo method for *in silico* modeling and visualization of Waddington’s epigenetic landscape with intermediate details

Xiaomeng Zhang^{a,c}, Ket Hing Chong^{a,c}, Lin Zhu^b, Jie Zheng^{b,*}

^a*Biomedical Informatics Lab, School of Computer Science and Engineering, Nanyang Technological University, 639798, Singapore*

^b*School of Information Science and Technology, ShanghaiTech University, Pudong District, Shanghai 201210, China*

^c*These authors contributed equally to this work.*

Abstract

Waddington’s epigenetic landscape is a classic metaphor for describing the cellular dynamics during the development modulated by gene regulation. Quantifying Waddington’s epigenetic landscape by mathematical modeling would be useful for understanding the mechanisms of cell fate determination. A few computational methods have been proposed for quantitative modeling of landscape; however, to model and visualize the landscape of a high dimensional gene regulatory system with realistic details is still challenging. Here, we propose a Monte Carlo method for modeling the Waddington’s epigenetic landscape of a gene regulatory network (GRN). The method estimates the probability distribution of cellular states by collecting a large number of time-course simulations with random initial conditions. By projecting all the trajectories into a 2-dimensional plane of dimensions i and j , we can approximately calculate the quasi-potential $U(x_i, x_j, *) = -\ln P(x_i, x_j, *)$, where $P(x_i, x_j, *)$ is the estimated probability of an equilibrium steady state or a non-equilibrium state. Compared to the state-of-the-art methods, our Monte Carlo method can quantify the global potential landscape (or emergence behavior) of GRN for a high dimensional system. The potential landscapes show that not only attractors represent stability, but the

*Corresponding author

Email address: zhengjie@shanghaitech.edu.cn (Jie Zheng)

paths between attractors are also part of the stability or robustness of biological systems. We demonstrate the novelty and reliability of our method by plotting the potential landscapes of a few published models of GRN.

Keywords: Waddington’s epigenetic landscape; Monte Carlo; gene regulatory network; dynamical systems

1. Introduction

The Waddington’s epigenetic landscape has been recognized as a powerful metaphor for explaining the phenomena of embryonic development and cellular differentiation in biology (Waddington, 1957; Goldberg et al., 2007; Slack, 2002; Yamanaka, 2009; Yu et al., 2020). The essence of the conceptual model proposed by Waddington is the ability to explain the emergent properties of cell fate decisions (Jamniczky et al., 2010). At least two types of approaches based on dynamical systems theory for quantifying the Waddington’s epigenetic landscape have been used. The first is the discrete formalism of Boolean network modeling (Davila-Velderrain et al., 2015a; Zhou et al., 2014) and the second is continuous modeling in the form of ordinary differential equations (ODEs) (Davila-Velderrain et al., 2015b; Huang, 2009, 2012). This paper is focused on the second approach.

Recently, a few methods for quantifying and plotting Waddington’s epigenetic landscape based on gene regulatory networks (GRNs) have been proposed (Bhattacharya et al., 2011; Ferrell Jr, 2012; Li and Wang, 2013; Wang et al., 2011; Zhou et al., 2012). A key step in these methods is the formulation of a potential (or quasi-potential) value for the dynamical system of GRN. For example, Bhattacharya et al. (2011) proposed a method for mapping aligned trajectories of the dynamical system of GRN in ODEs to a “quasi-potential” surface in the x - y phase space. When investigating mathematical models of two important processes in development, cell-fate induction and lateral inhibition (Ferrell Jr, 2012), Ferrell showed that the unique formulation of the potential surface for the lateral inhibition model can produce a pitchfork bifurcation,

25 which is consistent with Waddington’s epigenetic landscape where a ball repre-
26 senting a cell is moving down the hill and then bifurcates into two valleys (i.e.
27 two stable states).

28 Later, Zhou et al. (2012) proposed a theoretical framework for the decom-
29 position of vector fields which enables the computation of a “quasi-potential
30 function” for multi-attractor systems. Among recent methods for quantifying
31 Waddington’s epigenetic landscape is the one proposed by Li and Wang (2013)
32 who used statistical mechanics to quantify the potential landscape through self-
33 consistent mean field approximation. Their formulation of the potential based
34 on the probability distribution of steady states captures the global potential
35 landscape and the global barrier height measured by the “potential difference
36 between the two attractor minimums and the saddle point on landscape” (Li
37 and Wang, 2013). Li and Wang (2013) also used a path integral method to ob-
38 tain the kinetic paths of transition between attractors. The self-consistent mean
39 field approximation method has been implemented as a part of a software pack-
40 age named “NetLand” by our group to facilitate the drawing of Waddington’s
41 epigenetic landscape (Guo et al., 2017).

42 Although the method proposed by Li and Wang can quantify potential land-
43 scapes for high dimensional GRNs, their method has limitation in the lack of
44 realistic details of the landscape. Moreover, the high dimensionality of the GRN
45 as measured by the number of genes poses challenges for modeling, analysis and
46 visualization; for example, the methods proposed by Bhattacharya et al. (2011)
47 and Ferrell Jr (2012) allow two variables only. In this paper we propose a simple
48 Monte Carlo method for quantifying the Waddington’s epigenetic landscape of
49 GRNs of more than two genes. Our algorithm projects the time-course trajecto-
50 ries into a 2-dimensional plane of dimensions i and j to calculate the probability
51 distribution and potential $U(x_i, x_j, *) = -\ln P(x_i, x_j, *)$, where $P(x_i, x_j, *)$ is
52 the estimated probability of an equilibrium steady state or a non-equilibrium
53 state. We demonstrate unique features of the proposed method by plotting the
54 landscapes of a few case studies of GRN from two-dimensional to higher dimen-
55 sional models. In our Monte Carlo method a large number of random initial

56 conditions drawn from the state space are used to calculate time-series trajec-
57 tories based on ODEs. The time-series trajectories are then projected into a plane
58 that is divided into grid boxes to estimate the probability distribution and the
59 quasi-potentials of cell states. Surprisingly, such a simple method can capture
60 detailed features of the dynamical system, such as basin of attraction and un-
61 stable manifolds connecting two attractors as a kinetic path in the potential
62 landscape.

63 Testing on a few published models of GRN showed that our Monte Carlo
64 method can successfully quantify global potential landscapes consistent with
65 the state-of-the-art methods. The case studies demonstrate the power of our
66 computational method in uncovering the detailed dynamical behaviors of GRNs
67 that other methods fail to capture. In addition, we have also used the stochas-
68 tic approach of Chemical Langevin Equation (CLE) to estimate the probability
69 distribution of states by collecting simulated time series. The potential land-
70 scape constructed by the stochastic approach turned out to be consistent with
71 the landscape by the deterministic approach. Our analysis indicates that the
72 structure of a GRN (or in other words model interactions) can contribute to the
73 robustness of the attractors to noise. Moreover, we argue that in the Wadding-
74 ton’s epigenetic landscape not only the attractors represent stability, but that
75 the paths between attractors characterised by unstable manifolds of saddle point
76 or stable manifolds to attractor also contribute to the stability or robustness of
77 gene regulatory systems for cell fate decision.

78 2. Methods

79 In this paper, we propose a Monte Carlo method to quantify the quasi-
80 potential landscape of equilibrium steady states and non-equilibrium states of a
81 GRN. First, we derive the quasi-potential from chemical master equation (CME)
82 using the Monte Carlo method. Instead of solving the CME using the Gillespie’s
83 algorithm (also known as the stochastic simulation algorithm) which is compu-
84 tationally costly, we use numerical simulations of ODE mainly because there are

many efficient ODE solvers to obtain time-course trajectories. The Monte Carlo method is used to: (1) generate a large number of random initial conditions that are used for time-course simulations, and (2) estimate probability distribution from time-course trajectories projected into a 2-dimensional plane to quantify the quasi-potential U . A summary of the Monte Carlo method for quantifying Waddington’s epigenetic landscape is given in Algorithm 1. Secondly, we discuss how to use stochastic simulations with CLE, which is an improved computation of tau-leaping method for approximate execution of Gillespie’s algorithm (Gillespie, 2007). Using the time-course trajectories generated from CLE we can also quantify the Waddington’s epigenetic landscape for a GRN. Our method is based on several assumptions, some of which were inherited from previous publications and others extracted from our own empirical observations. For readers’ convenience, we list most of the assumptions that we can think of, along with their justifications and references, in Table 1.

2.1. Quantifying Waddington’s epigenetic landscape using deterministic ODE models

2.1.1. Derivation of average state probability used in quasi-potential

To derive the quasi-potential for the dynamics of gene expression driving cell state transition, we start from the definition of the CME (Gillespie, 1977; Toral and Colet, 2014)

$$\frac{dP(\mathbf{x}, t)}{dt} = \sum_{j=1}^m a_j(\mathbf{x} - \boldsymbol{\nu}_j)P(\mathbf{x} - \boldsymbol{\nu}_j, t) - \sum_{j=1}^m a_j(\mathbf{x})P(\mathbf{x}, t), \quad (1)$$

where $\mathbf{x} = (x_1, x_2, \dots, x_n)$ is the state of the system under study (e.g. a GRN), x_i is the copy number of the i th molecular species, n is the number of molecular species, m is the number of reactions, and $P(\mathbf{x}, t)$ is the probability of the system being in state $\mathbf{x}$ at time t . The function $a_j(\cdot)$ defines the propensity function for the j th reaction and $\boldsymbol{\nu}_j$ is the stoichiometric transition vector for the j th reaction. The CME defines the time-evolution of the function $P(\mathbf{x}, t)$ (Gillespie, 1977). The CME can be interpreted as that the flow of the probability

112 of a system being in state $\mathbf{x}$ at time t is given by the probability of arriving at
 113 $\mathbf{x}$ when reaction j occurs, $a_j(\mathbf{x} - \boldsymbol{\nu}_j)P(\mathbf{x} - \boldsymbol{\nu}_j, t)$ subtracted by the probability
 114 of the system leaving $\mathbf{x}$ when reaction j fires, $a_j(\mathbf{x})P(\mathbf{x}, t)$ (Sunkara, 2009).
 115 Summing up all the possible reactions for j from 1 to m gives Eq. (1). For an
 116 introduction of CME readers are refer to Higham (2008). From Eq. (1), we can
 117 calculate the probability of a cell in state $\mathbf{x}$ at time t

$$118 \quad P(\mathbf{x}, t) = \int \sum_{j=1}^m [(a_j(\mathbf{x} - \boldsymbol{\nu}_j)P(\mathbf{x} - \boldsymbol{\nu}_j, t) - a_j(\mathbf{x})P(\mathbf{x}, t))]dt + C \quad (2)$$

120 In theory, when the time t is large, e.g. approaching infinity, the probability
 121 $P(\mathbf{x}, t)$ will approach the steady-state probability P_{ss} (Wang et al., 2008a,
 122 2010). Wang and co-workers (Wang et al., 2008a, 2010) proposed a formulation
 123 for the quasi-potential as $U = -\ln P_{ss}$. However, to study the dynamics of a
 124 biological system we also calculate the average probability of state at $\mathbf{x}$ over the
 125 time from 0 to T , as

$$126 \quad P(\mathbf{x}) = \frac{\int_0^T P(\mathbf{x}, t)dt}{T} \quad (3)$$

127 According to the definition of the quasi-potential U proposed in (Wang et al.,
 128 2008b, 2010), we approximate the quasi-potential U by using $P(\mathbf{x})$ from Eq.(3):

$$U = -\ln P(\mathbf{x}) \quad (4)$$

129 2.1.2. Using Monte Carlo simulation to estimate quasi-potential

130 It is difficult to solve $U = -\ln P(\mathbf{x})$ analytically. Thus, we can use a Monte
 131 Carlo method to get an approximate solution. Monte Carlo methods use com-
 132 puter simulations with random sampling to approximate the exact solutions
 133 (Metropolis and Ulam, 1949). It has been widely used for solving a variety
 134 of problems including landscape modeling. For example, two recent works used
 135 Monte Carlo simulations to model epigenetic landscape and cell type transitions.
 136 Nakagawa and Narikiyo (2010) proposed a minimal modeling of epigenetic land-
 137 scape based on the fitness of interacting cells. Wang et al. (2014) proposed a

138 Monte Carlo method based on an ensemble of parameters to simulate the global
 139 dynamics of the epigenetic state network. Our Monte Carlo method is different
 140 from the above two methods, in that we use a large number of random initial
 141 conditions for simulating trajectories and then estimate the probability $P(\mathbf{x})$ by
 142 aggregating the trajectories.

143 First, from Eq. (3) we discretize the formulation from integral into summa-
 144 tion:

$$P(\mathbf{x}) = \frac{\sum_{0 \leq t_k \leq T} P(\mathbf{x}, t) \Delta t_k}{\sum_{0 \leq t_k \leq T} \Delta t_k} \quad (5)$$

145 where Δt_k is the increment of time for $P(\mathbf{x}, t)$. Suppose the initial condition is
 146 uniformly distributed, and x_i is not larger than X_i , which is a positive integer
 147 fixed by a modeler based on the maximum value in the dynamical system. Then
 148 the probability averaged over the initial conditions is given by:

$$P(\mathbf{x}, 0) = \begin{cases} \frac{1}{X_1 \cdot X_2 \cdot \dots \cdot X_n}, & 1 < x_i \leq X_i, (i = 1, 2, \dots, n) \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

150 Next, we randomly choose N initial states $\{\mathbf{x}_1(0), \mathbf{x}_2(0), \dots, \mathbf{x}_N(0)\}$ which
 151 are uniformly distributed in every dimension. Based on these random initial
 152 conditions we simulate N trajectories using the Gillespie's algorithm (Gillespie,
 153 1977) which is again a Monte Carlo method. Let us denote the i th trajectory
 154 by $S_i = \{\mathbf{x}_i(t_{i,0}), \mathbf{x}_i(t_{i,1}), \mathbf{x}_i(t_{i,2}), \dots, \mathbf{x}_i(t_{i,m_i})\}$, where $\mathbf{x}_i(t_{i,m_i})$ is the state at
 155 time t_{i,m_i} for the i th trajectory. From these trajectories we can estimate the
 156 probability in Eq. (5) as follows:

$$P(\mathbf{x}) = \frac{\sum_{i=1}^N \sum_{j=1}^{m_i} [\mathbf{x}_i(t_{i,j}) = \mathbf{x}] \cdot (t_{i,j} - t_{i,j-1})}{\sum_{i=1}^N \sum_{j=1}^{m_i} (t_{i,j} - t_{i,j-1})}, \quad (7)$$

158 where $[A]$ is the Iverson bracket defined by

$$[A] = \begin{cases} 1, & \text{if } A \text{ is true} \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

160 From this Monte Carlo method we can estimate the probability distribution
 161 of the state at $\mathbf{x}$ given by $P(\mathbf{x})$ and then using Eq. (4) we can obtain the
 162 quasi-potential U for the dynamical system.

163 2.1.3. Improving the speed for quasi-potential estimation by using ODEs

164 Using the Gillespie’s algorithm we can simulate the time-evolution trajecto-
 165 ries for CME. However, it incurs very high computational cost for simulating ev-
 166 ery event of the chemical reactions (Li et al., 2008). According to the Gillespie’s
 167 work (Gillespie, 2001), when the number of molecules present in the biochemical
 168 reactions is large the stochastic and deterministic simulation results are almost
 169 equivalent with negligible random fluctuations. Assuming that the number of
 170 molecules present in the biochemical reactions is large, we use a numerical solu-
 171 tion of ODEs to speed up the computation. We also randomly choose N initial
 172 conditions $\{\mathbf{x}_1(0), \mathbf{x}_2(0), \dots, \mathbf{x}_N(0)\}$ which are uniformly distributed in every di-
 173 mension. However, these initial conditions are measured in concentration levels
 174 of the molecular species. From these random initial conditions we simulate N
 175 trajectories.

176 After numerically solving the ODEs, the output trajectories can be dis-
 177 cretized from the continuous time into specific time steps. As such, we can
 178 use Eq. (7) to obtain $P(\mathbf{x})$ for calculating the quasi-potential of the dynamical
 179 system. The only difference here is that $\mathbf{x}$ is measured in concentration of each
 180 molecular species instead of the number of molecules as in the Gillespie’s algo-
 181 rithm. We choose to use an ODE solver with a fixed time increment, i.e. Δt is
 182 constant. Let us assume the time difference is given by

$$183 \quad \Delta t = t_{i,j} - t_{i,j-1} \quad (9)$$

184 Then, substituting Eq. (9) into Eq. (7) and making simplification, we obtain

$$185 \quad P(\mathbf{x}) = \frac{\sum_{i=1}^N \sum_{j=1}^{m_i} [\mathbf{x}_i(t_{i,j}) = \mathbf{x}]}{N \cdot m_i}. \quad (10)$$

186 With Eq. (10) we have improved the speed for calculating the quasi-potential

187 value. However, an accurate estimation of $P(\mathbf{x})$ still involves high computational
 188 cost due to the large state space of the system. In order to further improve the
 189 speed of computation we apply a coarse graining step defined by

$$\begin{aligned}
 P(\mathbf{x}) &\approx P(\mathbf{x} - \Delta\mathbf{x} \leq \mathbf{x} \leq \mathbf{x} + \Delta\mathbf{x}) \\
 &= \frac{\sum_{i=1}^N \sum_{j=1}^{m_i} [\mathbf{x}_i(t_{i,j}) \in (\mathbf{x} - \Delta\mathbf{x}, \mathbf{x} + \Delta\mathbf{x})]}{N \cdot m_i}.
 \end{aligned} \tag{11}$$

191 The coarse graining above is implemented as the division of a 2-D plane into
 192 grid boxes.

193 2.1.4. Plotting landscape

194 From previous sections we have derived Eq. (7) and Eq. (11) for calcu-
 195 lating $P(\mathbf{x})$. However, these equations are for dynamical systems with high-
 196 dimensional state spaces. In order to plot the potential landscape in 3 dimen-
 197 sions for human viewers to understand, we need to reduce the dimensions of a
 198 system. Many dimensionality reduction methods can be applied. Here, we first
 199 propose a simple method of projecting the trajectories into a 2-dimensional plane
 200 which is later confirmed by principal component analysis (PCA). By marginal-
 201 izing out all the variables except for the i th and j th variables, we can obtain
 202 the probability distribution of the states in the i th and j th dimensions as

$$P(x_i, x_j, *) = \sum_{s^1} \sum_{s^2} \dots \sum_{s^{i-1}} \sum_{s^{i+1}} \dots \sum_{s^{j-1}} \sum_{s^{j+1}} \dots \sum_{s^n} P(x_1=s^1, x_2=s^2, \dots, x_{i-1}=s^{i-1}, x_i, x_{i+1}=s^{i+1}, \dots, x_{j-1}=s^{j-1}, x_j, x_{j+1}=s^{j+1}, \dots, x_n=s^n) \tag{12}$$

204 Then, we can substitute Eq. (12) into Eq. (4) to obtain the quasi-potential
 205 of a state in two dimensions

$$U(x_i, x_j, *) = -\ln P(x_i, x_j, *). \tag{13}$$

207 With x_i , x_j and $U(x_i, x_j, *)$ as the x , y and z coordinates respectively, we can
 208 plot the landscape in 3 dimensions. Since the calculation of $P(x_i, x_j, *)$ includes
 209 the time points from t_0 to t_{m_i} , our method can be used to analyze the properties
 210 of transient states, rather than being limited to steady states, and thereby can
 211 reveal more dynamical details.

Our Monte Carlo method is summarized in Algorithm 1. Essentially, the algorithm collects N simulated time-course trajectories from random initial conditions in the state space. Here, we use a fixed time step of $\Delta t = 0.1$ for numerical simulation, and thereby we can use the number of time points along a trajectory instead of the length of continuous time to estimate the probability of a state. Then from these time-course data, we project the points of the trajectories into a phase plane of two selected variables of interest and estimate the probabilities of states. For example, Figure 1 shows how the probability values can be estimated in a plane that has been divided into grid boxes. A grid box with a locally maximal number of points on the trajectories represents an attractor. In Figure 1 there are two yellow grid boxes with locally maximal numbers of points on the trajectories, and therefore they represent two attractors. These results are then used to estimate the quasi-potential $U(x_i, x_j, *) = -\ln P(x_i, x_j, *)$. Compared with other methods (e.g. Li and Wang (2013)) our method includes both equilibrium state and non-equilibrium states. A non-equilibrium state quantifies the transient behaviors of the system such as the intermediate flows in a vector field, whereas an equilibrium state quantifies a repeller (unstable steady state) or an attractor (stable steady state) (Tyson et al., 2001). This formulation and approximate calculation of the quasi-potential enables us to plot the Waddington’s epigenetic landscape with details.

Algorithm 1. Monte Carlo steps for quantifying quasi-potential landscape

1. generate N random initial conditions
2. select 2 marker genes (proteins) and split the 2-dimensional plane into grid boxes
3. for each initial condition:
 - (i) generate one trajectory
 - (ii) project the trajectory to the 2-dimensional grid boxes
 - (iii) from the trajectory calculate the number of points in each grid box
4. estimate probability of each grid box $P(x_i, x_j, *)$ as given by Eq. (12)

Figure 1: Illustration of the Monte Carlo method for approximating the probability distribution and identifying attractors. The projection of the time-course data into a plane with grid boxes enables the estimation of the probabilities of cellular states. In this example 10 trajectories (blue lines) are shown and the plane is split into 8 x 8 grid boxes. A grid box with a locally maximal number of points of trajectories corresponds to an attractor. In this landscape, there are two attractors as indicated by the two yellow boxes.

5. calculate quasi-potential of each grid box using $U(x_i, x_j, *) = -\ln P(x_i, x_j, *)$
6. plot the landscape in 3 dimensions

2.2. Stochastic modeling of landscape based on Chemical Langevin Equation (CLE)

To test the role of noise in gene expression and the robustness of GRNs we also investigate a second type of potential landscape that is based on stochastic time-course simulations. The algorithm for obtaining the potential landscape is similar to Algorithm 1. The key difference is that the stochastic approach uses Chemical Langevin Equation (CLE) (Gillespie, 2000), an improved version of the tau-leaping method for approximation of the Gillespie's algorithm (Gillespie, 2007), in simulating the time-series trajectories. CLE differs from ODEs in that the biochemical reactions are simulated using the stochastic part in the CLE (Higham, 2008). We adapted the CLE code from Higham (2008) with

the model reactions from the GRNs of Li and Wang (2015) and Gardner et al. (2000). We used a larger noise than in the original code of Higham (2008), by decreasing the volume of the system from $V = 10^{-15}$ to $V = 10^{-20}$. Since CLE is a well-established method for stochastic simulation of biochemical reactions (Gillespie, 2007) we will not discuss it in detail here. The CLE models and their implementation in MATLAB code are given in Supplementary materials. Below we will illustrate how to construct CLE by converting from ODEs. The definitions of the rates of change of molecular species are different between ODEs and CLE. In ODEs the molecular species are measured in concentration, whereas in CLE the molecular species are measured in the number of molecules. To explain how to convert the deterministic rate constant k_i in ODEs to the stochastic reaction rate constant c_i in CLE, let us look at Eq. (16), an example ODE model equation for protein x with self-activation (the first term on the right hand side) and spontaneous degradation (the second term):

$$\frac{dx}{dt} = k_1 \frac{x^n}{k_2^n + x^n} - k_3 x. \quad (14)$$

Let us use X to denote the copy number of protein x . Hereafter we will use the same symbol x to denote the concentration of protein x , in the unit of μM . The relationship between x and X is given by:

$$X = x \cdot N_A \cdot V, \quad (15)$$

where N_A is the Avogadro's number which equals 6.023×10^{23} and V is the volume of the system in liters (Higham, 2008). Next, denote $B = N_A \cdot V$, then Eq. (15) becomes

$$X = x \cdot B.$$

Multiply B to both sides of Eq. (14), and we get

$$\begin{aligned} \frac{dx}{dt} B &= B \cdot k_1 \frac{x^n}{k_2^n + x^n} - k_3 x \cdot B, \\ \frac{dx}{dt} B &= B \cdot k_1 \frac{B^n \cdot x^n}{B^n (k_2^n + x^n)} - k_3 x \cdot B, \end{aligned}$$

$$\frac{d(x \cdot B)}{dt} = B \cdot k_1 \frac{(x \cdot B)^n}{(Bk_2)^n + (x \cdot B)^n} - k_3(x \cdot B).$$

Since $X = x \cdot B$, we obtain the following equation:

$$\frac{dX}{dt} = B \cdot k_1 \frac{(X)^n}{(Bk_2)^n + (X)^n} - k_3 X. \quad (16)$$

In the stochastic approach the rate of change for X is defined by:

$$\frac{\Delta X}{\Delta t} = c_1 \frac{X^n}{c_2^n + X^n} - c_3 \cdot X. \quad (17)$$

Comparing Eq. (18) and Eq. (19), we deduce that,

$$c_1 = B \cdot k_1 = N_A \cdot V \cdot k_1,$$

$$c_2 = B \cdot k_2 = N_A \cdot V \cdot k_2,$$

$$c_3 = k_3$$

From the derivations above, we demonstrate the relationship between k_i and c_i which enables us to obtain the c_i for the stochastic simulation using CLE.

3. Results

3.1. Quantifying the potential landscape of non-equilibrium and equilibrium states

To demonstrate the capability of the proposed method, we selected three GRN models: (1) a bistable synthetic toggle switch (Gardner et al., 2000), (2) a model of cancer attractors (Li and Wang, 2015) and (3) a stem cell differentiation and reprogramming model (Li and Wang, 2013). For the model equations, readers may refer to the original papers or the source code included in our MATLAB package (see the Additional material). These case studies illustrate that our method can capture distinct details of dynamical systems, e.g. attractor, repeller, unstable manifold and saddle point. Furthermore, we compare our method with another constructing landscape method (based on the NetLand (Guo et al., 2017)) from computing costs and results.

316 3.1.1. Case Study 1: Bistable synthetic toggle switch from Gardner et al. (2000)

317 In the first case study, our method is used to analyse the transient properties
 318 of a synthetic genetic toggle switch proposed by Gardner et al. (2000). The
 319 model equations and parameter values we used are the same as given by Segel
 320 and Edelstein-Keshet (2013). By applying the Monte Carlo method which is
 321 presented as Algorithm 1 in this paper (see Methods), we obtained the landscape
 322 as shown in Figure 2A. The potential landscape shows two attractors and there
 323 is a valley (or kinetic path) connecting the two attractors. Between the two
 324 attractors, there is a saddle point (Figure 2B & 2C).

325 In Figure 2D phase plane analysis shows that the system displays two attrac-
 326 tors and one saddle point. A saddle point is formed by both stable and unstable
 327 manifolds (Gardner et al., 2000; Segel and Edelstein-Keshet, 2013; Chong et al.,
 328 2015). By comparing the potential landscape with the phase plane (Figure 2D),
 329 the valley (or kinetic path) in Figure 2A is found to be formed by the unstable
 330 manifolds of saddle point or stable manifolds, which cannot be generated by
 331 analysing steady states only.

332 This simple model of GRN shows that our method of formulating the poten-
 333 tial can capture some transient properties of a dynamical system. For example,
 334 when there is a valley between two attractors in the potential landscape, the
 335 kinetic path is formed by unstable manifolds. The saddle point sets a thresh-
 336 old for the barrier height that can separate the two attractors. The potential
 337 landscape displays a three-dimensional view of the phase plane and shows the
 338 attractors and saddle point more clearly than a two-dimensional plane. In par-
 339 ticular, the saddle point is shown to have one convex (local minimum) and one
 340 concave down (local maximum) in the nearly orthogonal directions. The result
 341 from this example suggests that not only attractors represent stability but the
 342 kinetic paths are also part of the stability of GRNs.

351 3.1.2. Case Study 2: Cancer attractors from Li and Wang (2015)

352 To test if our method can handle a network of more than two genes, we
 353 choose a six-gene network model proposed by Li and Wang (2015). This gene

Figure 2: **Potential landscape for Case Study 1.** (A) 3D view of a genetic toggle switch-based landscape: The landscape displays two attractors that are connected with a kinetic path formed by unstable manifolds. (B) Side view of the genetic toggle switch-based landscape in Case 1: The landscape displays two basins of attraction that are connected by the unstable manifolds. Separating the two attractors, the saddle point is a tipping point (or barrier height) for transition along the kinetic path between the basins of attraction. (C) Top view indicates the kinetic path formed by unstable manifolds. (D) Phase plane plot shows the unstable manifolds (blue arrows) from the saddle point form the kinetic path.

network was reported to produce four attractors representing cancer stem cells, stem cells, cancer cells and normal cells (Li and Wang, 2015). The potential landscape of this model generated by our algorithm, however, shows four pairs of attractors in which the two attractors in each pair are close to each other (Figure 3). Our analysis of the end points of the trajectories confirmed that there are 8 attractors (data not shown). The result implicates that Li and Wang (2015) may have joined the two nearby attractors into one attractor and thus resulted in only four attractors.

362 This result indicates that the potential landscape obtained using our method
363 can capture more details of attractors. In addition, the landscape also contains 6
364 valleys (Figure 3). These valleys are formed by unstable manifolds, as discussed
365 in the first example, and contain the information about saddle points as barrier
366 heights for separating two attractors. These valleys are similar to the kinetic
367 paths in Case Study 1 in illustrating the transition from one attractor to another
368 attractor in the landscape.

369 To visualize the Waddington’s epigenetic landscape of a gene regulatory net-
370 work of more than two genes in three dimensions using dimensionality reduction
371 is a problem that we will explore next. We tested the Monte Carlo method with
372 PCA (principal component analysis) where all the trajectories were arranged
373 vertically as each column represents a gene (or a protein) to form a feature ma-
374 trix and then apply PCA to obtain the two principal components that were later
375 used as the x-axis and y-axis. The Waddington’s epigenetic landscape obtained
376 from the Monte Carlo method with PCA is given in Figure 4. The landscape
377 plotted with the first two principal components has preserved the information
378 from the original data. It displays 8 attractors and the landscape is consistent
379 in the number of attractors with that in Figure 3. This result demonstrates that
380 the Monte Carlo method with PCA enables us to plot Waddington’s epigenetic
381 landscape in three dimensions without having to select two marker genes and
382 marginalize out all other genes.

399 3.1.3. Case Study 3: Stem cell differentiation and reprogramming from Li and 400 Wang (2013)

401 Finally, to test if our method can model high-dimensional gene regulatory
402 networks, we selected a 52-gene network model proposed by Li and Wang (2013)
403 for quantifying the stem cell differentiation and reprogramming. Li and Wang
404 (2013) identified two attractors in the state space of gene expression: stem cell
405 attractor and differentiated cell attractor. To plot the potential landscape we
406 chose two marker genes GATA6 and NANOG as in Li and Wang (2013), which
407 play pivotal roles in regulating stem cell fates, and used their expression levels

Figure 3: **Potential landscape for Case Study 2.** (A) 3D view of the landscape. Validation of our method by reproducing the potential landscape of Li and Wang (2015) in Case 2. By comparing the locations of the attractors labeled by Li and Wang (2015), we identified the attractors and their corresponding cell states. The landscape displays four main basins of attraction for normal cells, cancer cells, stem cells and cancer stem cells. However, our landscape is slightly different from that in the original paper (Li and Wang, 2015): each of the main basins of attraction is divided into two attractors which are connected by kinetic paths. (B) Top view of the landscape. These kinetic paths and their directions are plotted according to Li and Wang (2015). These kinetic paths suggest that cell changes state according to different directions of the kinetic paths (Li and Wang, 2015). The color corresponds to the value of the quasi-potential U .

as the coordinates of the 2D panel. The shape of our potential landscape is consistent with that in Li and Wang (2013), e.g. both showing two attractors (Figure 5). The potential landscape shows that the stem cell attractor has a bigger basin of attraction and lower potential value than the differentiated cell attractor.

The result in Figure 5 suggests that our method can capture the transient non-equilibrium states and the attractors of the equilibrium steady states with more details than the results reported by Li and Wang (2013). The potential landscape displays one dominant attractor shown as the stem cell attractor, which implies that the probability of getting attracted to this stem cell attractor

Figure 4: **Potential landscape for Case Study 2 plotted using the Monte Carlo method with PCA.** (A) 3D view of the Waddington's epigenetic landscape plotted based on the first two principal components. The landscape indicates there are eight attractors which is consistent with the potential landscape obtained using the two selected genes P53 and ZEB as shown in Figure 3. (B) Top view of the landscape.

is higher than the differentiated cell attractor. As with Case Study 2 we also use PCA to plot a potential landscape in three dimensions based on the first two principal components as the coordinates of the 2D panel, as shown in Figure 6. Again, PCA has preserved the presence and locations of two attractors as in Figure 5. Moreover, it suggests that the proposed Monte Carlo method can be used to visualize Waddington's epigenetic landscape for high-dimensional gene regulatory networks.

3.1.4. Comparison of our Monte-Carlo method with NetLand

To assess the reliability and advantages of our Monte Carlo method, we compare it with NetLand, a state-of-the-art software tool for landscape modeling and visualization (Guo et al., 2017). First, we test and compare the computational efficiency between the two methods. They were tested on the ODE models of the three case studies, running on a ThinkPad Laptop, with Intel Core i7-5500U (CPU 2.4GHz) and memory size of 12GB. As shown in Table 2, our method is faster than NetLand, but needs more memory. Note that NetLand

Figure 5: **Potential landscapes for Case Study 3.** (A) 3D view of the Waddington's epigenetic landscape based on a 52-gene network model (Li and Wang, 2013), which shows two basins of attraction. By comparing the locations of the attractors labeled by Li and Wang (2013), we observe that the stem cell attractor located on the left is the dominant attractor, i.e. bigger and deeper (blue color) than the differentiation attractor located on the right. (B) Top view of the landscape.

is implemented in Java, whereas our Monte Carlo method is in MATLAB. Thus, it is likely that our method can be more efficient when it is re-implemented in another computer language in the future. From the constructed landscape results (as shown in Figure 3 and Figure 7), both methods can capture the steady states of the system. Furthermore, our method can spontaneously reveal kinetic paths between attractors. By contrast, in the output of NetLand, the shape of a basin of attraction is round likely due to the Gaussian approximation and the kinetic paths need to be calculated separately, which has not been implemented in NetLand. Figure 7 is a screenshot of the output of NetLand on Case Study 2. Although the numerical results of NetLand show that there are 8 attractors, it is difficult to find more than 4 attractors in the landscape plot. This example demonstrates the advantage of showing more intermediate details offered by our Monte Carlo method.

Figure 6: **Potential landscape for Case Study 3 plotted using PCA.** (A) 3D view of the Waddington's epigenetic landscape plotted based on the first two principal components. The landscape shows there are two attractors which is consistent with the potential landscape obtained using the two selected genes GATA6 and NANOG as shown in Figure 5. This landscape is plotted with 60,000 trajectories with random initial conditions due to memory limitation, whereas other landscapes in this paper were plotted with 100,000 trajectories. (B) Top view of the landscape.

Figure 7: A screenshot of output of NetLand on Cast Study 2 (Li and Wang, 2015).

459 3.2. Stochastic modeling of landscape based on Chemical Langevin Equation 460 (CLE)

461 The models of Waddington’s epigenetic landscape in the previous section
462 were constructed based on deterministic simulations using ODEs. Here, we
463 also investigate the landscape model for Li and Wang (2015) constructed using
464 a stochastic approach based on Chemical Langevin Equation (CLE) (Gillespie,
465 2000; Higham, 2008). We applied the same procedure as described in Algorithm
466 1 in the Methods section with the only difference that the numerical simulation
467 of the time-course data was done with CLE instead of ODEs. For the 6-gene net-
468 work model in Case Study 2 from Li and Wang (2015), we defined 22 reactions
469 (16 network interactions of activation and inhibition plus 6 self-degradations)
470 and converted the rate constants from the deterministic k_i in ODEs to the
471 stochastic rate constants c_i , as explained in Methods. We simulated 100,000
472 trajectories with random initial conditions. Two runs of the time-course simu-
473 lation using CLE with random initial conditions are given in Figure 8A. Each
474 trajectory is projected to a 2D plane of dimensions i and j to estimate the
475 marginalized probability $P(x_i, x_j, *)$ and the quasi-potential $U(x_i, x_j, *) = -\ln$
476 $P(x_i, x_j, *)$ as in the ODE model. A potential landscape for the model of Li and
477 Wang (2015) obtained from CLE is shown in Figure 8B, which is comparable to
478 the one obtained using deterministic ODE-based simulation (Figure 3A). The
479 stochastic noise in the CLE only slightly perturbed the dynamic simulations as
480 the shapes of the attractors in the potential landscape were almost unchanged
481 (even when we used a larger noise than in the original code of Higham (2008) by
482 changing the volume of the system from $V = 10^{-15}$ to $V = 10^{-20}$). However,
483 the attractors in the potential landscape from CLE display slightly larger basins
484 of attraction. As a result, two of the attractors that are located close to each
485 other in the ODE-based landscape have been merged into one in the CLE-based
486 landscape (Figure 8B). We also set the volume of the system $V = 10^{-15}$ to
487 obtain a large number of molecules, and the landscapes plotted based on ODE
488 and CLE are almost identical (data not shown). The computational time for
489 generating the potential landscape using the CLE is 57.9 minutes, which is much

longer than using the ODEs (12 minutes). This is expected as the stochastic simulation normally takes a longer time due to the high computational cost for simulating all the events of biochemical reactions (Li et al., 2008).

We also performed stochastic simulations using the synthetic toggle switch model of Gardner et al. (2000) and we obtained similar results, i.e. the potential landscape is not affected by the noise (Supplementary Figure S9). These results implicate some degree of robustness of GRNs.

Figure 8: **(A)** Two realizations of time-course simulation of the 6-gene network in Case Study 2 generated from Chemical Langevin Equation (CLE) with random initial conditions. These time-course simulations show that, starting from two different initial conditions, two trajectories with some randomness can eventually converge to the same attractor. **(B)** Waddington's epigenetic landscape of the 6-gene network model from Li and Wang (2015), which shows four basins of attraction. The potential landscape was generated by using CLE-based dynamic simulations, and its shape is consistent with that of the landscape obtained using ODEs shown in Figure 3. It implicates that the attractors in the potential landscape are robust to noise.

4. Discussion

In this paper, we present a simple Monte Carlo method for quantitatively modeling and visualization of Waddington's epigenetic landscape based on dynamical simulation of GRNs. This method uses a large number of initial conditions randomly sampled from a uniform distribution in the state space, and

510 the collected time-series data are then used to estimate the probability values
 511 and the quasi-potential values of cell states in the landscape. One key advantage of our method is that it can quantify the potential landscape to display
 512 both global and local dynamics of the cell state transitions for both equilibrium
 513 states (e.g. steady states or attractor states) and non-equilibrium states (e.g.
 514 transient states along a kinetic path). Dimensionality reduction is an important
 515 part of landscape plotting in three dimensions (just as GPLVM used in NetLand
 516 (Guo et al., 2017), but it is not a trivial task. We used PCA to plot Waddington's landscapes by using the first two principal components as the coordinates
 517 of state space, for Case Study 2 in Figure 4 and for Case Study 3 in Figure
 518 6. The potential landscape from the first two principal components shows a
 519 consistent picture of the potential landscape using the two selected genes of
 520 P53 and ZEB as shown in Figure 3, e.g. both landscapes show that there are 8
 521 attractors. However, choosing marker genes according to prior biological knowledge can have more interpretability than using dimensionality reduction. Our
 522 results suggest that not using dimensionality reduction will often not affect the
 523 3D view of Waddington's epigenetic landscape, since choosing the right pair of
 524 genes as the x- and y-axes we could plot a meaningful landscape. Particularly,
 525 how genes for x- and y-axes should be chosen are based on key biomarkers in
 526 the biological system that contribute to cell fate decisions or cell states as commonly practiced by modelers (Bhattacharya et al., 2011; Li and Wang, 2013;
 527 Huang, 2012; Li and Wang, 2015). The landscape provides a 3D view of the
 528 dynamical features (e.g. attractor, saddle point, unstable manifold and stable manifold) of the system, whereas conventional dynamical system analysis uses
 529 2D views such as bifurcation diagram, phase plane and vector field. Thus, our
 530 computational method can be used for detailed modeling and 3D visualization
 531 of dynamical systems of cells.

537 While the method proposed by Li and Wang (2013) formulates the potential as $U = -\ln P_{ss}$ (where P_{ss} is the steady state probability) and uses the
 538 self-consistent mean field approximation, our method uses the estimated probability of any state, either an equilibrium steady state or non-equilibrium tran-
 539

541 sient states). Li and Wang (2013) also used 100,000 time-course simulations
 542 from random initial conditions in the state space and inferred the steady state
 543 probability distribution with multi-variable Gaussian distribution to plot the
 544 potential landscape, which displays a smooth surface with basins of attraction.
 545 Between their method and ours, the locations of attractors are essentially the
 546 same. However, our method can capture more detailed information than their
 547 method (Li and Wang, 2013, 2015), as demonstrated in Case 2 (with four pairs
 548 of attractors as shown in Figure 3) and in Case Study 3 (with one dominant
 549 attractor as shown in Figure 5). In Case Study 2, the four pairs of attractors
 550 (Figure 3) can be explained by the kinetic paths for the transition between at-
 551 tractors. As for Case Study 3, we are not aware of any biological reason for
 552 the landscape to display one dominant stem cell attractor and one minor dif-
 553 ferentiated cell attractor (Figure 5), which should be investigated in the future.
 554 Moreover, our Monte Carlo method is powerful in that it can capture the kinetic
 555 path without using the path integral method as in Li and Wang (2013). The
 556 kinetic paths proposed by Li and Wang (2013) or the epigenetic valleys as in
 557 Ashwin and Sasai (2015) are barrier pathways that describe the intermediate
 558 states and the differentiated states. The kinetic path (or valley) between two
 559 attractors can give biological insight into the transition from one attractor to
 560 another where the intermediate state transitions must follow this path towards
 561 the final stable state. The kinetic path in the Waddington’s epigenetic land-
 562 scape can explain why the cell differentiation in embryonic development follows
 563 a deterministic path (Waddington, 1957; Lanctôt, 2015) which was called by
 564 Waddington himself as “chreod” (Slack, 2002; Jaeger and Monk, 2014).

565 The potential landscape modeling based on stochastic simulations of Langevin
 566 dynamics has also been conducted by Li and Wang (2013). Also, to plot 3D
 567 views of landscape, they used the method of root mean square distance (RMSD),
 568 i.e. using coordinates of two attractors (with locally minimum potentials) as
 569 two reference points to reduce a multi-dimensional space into two dimensions of
 570 RMSD1 and RMSD2. Using the Langevin dynamics and RMSD they obtained
 571 a landscape with the same number of attractors as the landscape based on their

self-consistent mean field approximation method. However, the topographies of the two landscapes (i.e. from self-consistent mean field approximation and Langevin dynamics with RMSD) reported by Li and Wang Li and Wang (2013) are different. Here, we used Chemical Langevin Equation to obtain the time-course trajectories and applied our Monte Carlo method to directly obtain a landscape. Using the stochastic approach of CLE we can obtain a Waddington’s epigenetic landscape consistent with that using the deterministic approach of ODEs. These results of computer simulation highlight the robustness of the gene regulatory network to noise (Wang et al., 2008b; Thattai and Van Oudenaarden, 2001).

5. Conclusion

The Monte Carlo method for plotting potential landscapes for multi-dimensional GRNs allows us to study the links between genotype and phenotype as initially proposed by Conrad Waddington over 60 years ago (Waddington, 1957; Goldberg et al., 2007). Through studies of real biological networks, we demonstrated the simplicity and power of the method for plotting Waddington’s epigenetic landscape. It can facilitate our understanding of cellular dynamics, e.g. the differentiation and reprogramming of stem cells as well as other biological processes. In general, the algorithm proposed here for cellular dynamics can also be applied to studying other types of dynamical systems such as social networks.

Acknowledgments

The authors would like to thank Chunhe Li and Sudin Bhattacharya for answering our questions regarding their published methods. We also thank Bard Ermentrout for the permission to use the XPPAUT software for plotting the phase plane.

597 **Availability of data and materials**

598 All the data and materials are provided in the additional files.

599 **Additional files**

600 available at [https://github.com/MCLand-NTU/Zhang_et_al_Supplementary_](https://github.com/MCLand-NTU/Zhang_et_al_Supplementary_Information/)
601 [Information/](https://github.com/MCLand-NTU/Zhang_et_al_Supplementary_Information/)

602 *Additional file 1: Supplementary Information*

603 This contains supplementary methods and instructions on how to use the
604 MATLAB code provided to simulate and visualize the Waddington’s epigenetic
605 landscapes presented in this paper.

606 *Additional file 2: Supplementary Code.*

607 This file contains the source code in MATLAB format for the implementation
608 of the algorithm used to plot the Waddington’s epigenetic landscapes.

609 **Competing interests**

610 The authors declare that they have no competing interests.

611 **Funding**

612 This work was supported by the MOE AcRF Tier 1 grant (2015-T1-002-094),
613 MOE AcRF Tier 1 Seed Grant on Complexity (RGC 2/13, M4011101.020), and
614 MOE AcRF Tier 2 Grant (ARC39/13, MOE2013-T2-1-079), Ministry of Edu-
615 cation Singapore, and the start-up grant of ShanghaiTech University, Shanghai,
616 China.

617 **Author’s contributions**

618 XZ and KHC designed the method and implemented the algorithm and case
619 studies. LZ contributed in the implementation and test of PCA. JZ supervised
620 the design of the method and the study. XZ, KHC and JZ wrote the manuscript.

621 References

- 622 Ashwin, S., Sasai, M., 2015. Effects of collective histone state dynamics on
623 epigenetic landscape and kinetics of cell reprogramming. *Scientific reports* 5,
624 16746.
- 625 Bhattacharya, S., Zhang, Q., Andersen, M.E., 2011. A deterministic map of
626 waddington’s epigenetic landscape for cell fate specification. *BMC Syst. Biol*
627 5, 85.
- 628 Chong, K.H., Samarasinghe, S., Kulasiri, D., Zheng, J., 2015. Computational
629 techniques in mathematical modelling of biological switches, in: Weber, T.,
630 McPhee, M., Anderssen, R.S. (Eds.), 21st International Congress on Mod-
631 elling and Simulation (MODSIM 2015), p. 578.
- 632 Davila-Velderrain, J., Juarez-Ramiro, L., Martinez-Garcia, J.C., Alvarez-
633 Buylla, E.R., 2015a. Methods for characterizing the epigenetic attractors
634 landscape associated with boolean gene regulatory networks. *arXiv preprint*
635 *arXiv:1510.04230* .
- 636 Davila-Velderrain, J., Martinez-Garcia, J.C., Alvarez-Buylla, E.R., 2015b. Mod-
637 eling the epigenetic attractors landscape: toward a post-genomic mechanistic
638 understanding of development. *Front. Genet* 6, 160.
- 639 Ferrell Jr, J.E., 2012. Bistability, bifurcations, and waddington’s epigenetic
640 landscape. *Curr. Biol.* 22, R458–R466.
- 641 Gardner, T.S., Cantor, C.R., Collins, J.J., 2000. Construction of a genetic toggle
642 switch in *Escherichia coli*. *Nature* 403, 339.
- 643 Gillespie, D.T., 1977. Exact stochastic simulation of coupled chemical reactions.
644 *J. Phys. Chem.* 81, 2340–2361.
- 645 Gillespie, D.T., 2000. The chemical langevin equation. *J. Chem. Phys.* 113,
646 297–306.

647 Gillespie, D.T., 2001. Approximate accelerated stochastic simulation of chemi-
648 cally reacting systems. *The Journal of chemical physics* 115, 1716–1733.

649 Gillespie, D.T., 2007. Stochastic simulation of chemical kinetics. *Annu. Rev.*
650 *Phys. Chem.* 58, 35–55.

651 Goldberg, A.D., Allis, C.D., Bernstein, E., 2007. Epigenetics: a landscape takes
652 shape. *Cell* 128, 635–638.

653 Guo, J., Lin, F., Zhang, X., Tanavde, V., Zheng, J., 2017. Netland: quanti-
654 tative modeling and visualization of waddington’s epigenetic landscape using
655 probabilistic potential. *Bioinformatics* 33, 1583–1585.

656 Higham, D.J., 2008. Modeling and simulating chemical reactions. *SIAM review*
657 50, 347–368.

658 Huang, S., 2009. Reprogramming cell fates: reconciling rarity with robustness.
659 *Bioessays* 31, 546–560.

660 Huang, S., 2012. The molecular and mathematical basis of waddington’s epi-
661 genetic landscape: A framework for post-darwinian biology? *Bioessays* 34,
662 149–157.

663 Jaeger, J., Monk, N., 2014. Bioattractors: dynamical systems theory and the
664 evolution of regulatory processes. *J. Physiol.* 592, 2267–2281.

665 Jamniczky, H.A., Boughner, J.C., Rolian, C., Gonzalez, P.N., Powell, C.D.,
666 Schmidt, E.J., Parsons, T.E., Bookstein, F.L., Hallgrímsson, B., 2010. Redis-
667 covering waddington in the post-genomic age. *Bioessays* 32, 553–558.

668 Lanctôt, C., 2015. Single cell analysis reveals concomitant transcription of
669 pluripotent and lineage markers during the early steps of differentiation of
670 embryonic stem cells. *Stem Cells* 33, 2949–2960.

671 Li, C., Wang, J., 2013. Quantifying cell fate decisions for differentiation and
672 reprogramming of a human stem cell network: landscape and biological paths.
673 *PLOS Comput. Biol* 9, e1003165.

674 Li, C., Wang, J., 2015. Quantifying the landscape for development and cancer
675 from a core cancer stem cell circuit. *Cancer Res.* 75, 2607–2618.

676 Li, H., Cao, Y., Petzold, L.R., Gillespie, D.T., 2008. Algorithms and software
677 for stochastic simulation of biochemical reacting systems. *Biotechnol. Prog.*
678 24, 56–61.

679 Metropolis, N., Ulam, S., 1949. The Monte Carlo method. *J. Am. Stat. Assoc.*
680 44, 335–341.

681 Nakagawa, M., Narikiyo, O., 2010. Epigenetic landscape of interacting cells: A
682 model simulation for developmental process. *Biosystems* 101, 156–161.

683 Segel, L.A., Edelstein-Keshet, L., 2013. *A Primer in Mathematical Models in*
684 *Biology.* Siam.

685 Slack, J.M., 2002. Conrad Hal Waddington: the last renaissance biologist? *Nat.*
686 *Rev. Genet.* 3, 889.

687 Sunkara, V., 2009. The chemical master equation with respect to reaction
688 counts, in: *Proc. 18th World IMACS/MODSIM Congress*, pp. 703–707.

689 Thattai, M., Van Oudenaarden, A., 2001. Intrinsic noise in gene regulatory
690 networks. *Proc. Natl. Acad. Sci. U.S.A.* 98, 8614–8619.

691 Toral, R., Colet, P., 2014. *Stochastic numerical methods: an introduction for*
692 *students and scientists.* John Wiley & Sons.

693 Tyson, J.J., Chen, K., Novak, B., 2001. Network dynamics and cell physiology.
694 *Nat. Rev. Mol. Cell Biol.* 2, 908.

695 Waddington, C., 1957. *The strategy of the genes.*

696 Wang, J., Xu, L., Wang, E., 2008a. Potential landscape and flux framework
697 of nonequilibrium networks: Robustness, dissipation, and coherence of bio-
698 chemical oscillations. *Proc. Natl. Acad. Sci. U.S.A.* 105, 12271–12276.

699 Wang, J., Xu, L., Wang, E., 2008b. Robustness, dissipations and coherence of
700 the oscillation of circadian clock: potential landscape and flux perspectives.
701 PMC Biophys 1, 7.

702 Wang, J., Xu, L., Wang, E., Huang, S., 2010. The potential landscape of genetic
703 circuits imposes the arrow of time in stem cell differentiation. Biophys. J. 99,
704 29–39.

705 Wang, J., Zhang, K., Xu, L., Wang, E., 2011. Quantifying the waddington
706 landscape and biological paths for development and differentiation. Proc.
707 Natl. Acad. Sci. U.S.A. 108, 8257–8262.

708 Wang, P., Song, C., Zhang, H., Wu, Z., Tian, X.J., Xing, J., 2014. Epigenetic
709 state network approach for describing cell phenotypic transitions. Interface
710 focus 4, 20130068.

711 Yamanaka, S., 2009. Elite and stochastic models for induced pluripotent stem
712 cell generation. Nature 460, 49.

713 Yu, C., Liu, Q., Chen, C., Wang, J., 2020. Quantification of the underlying
714 mechanisms and relationship among cancer, metastasis and differentiation/
715 development. Front. Genet. 10.

716 Zhou, J.X., Aliyu, M.D.S., Aurell, E., Huang, S., 2012. Quasi-potential land-
717 scape in complex multi-stable systems. J. Royal Soc. Interface 9, 3539–3553.

718 Zhou, J.X., Qiu, X., Fouquier d’Hérouël, A., Huang, S., 2014. Discrete gene
719 network models for understanding multicellularity and cell reprogramming:
720 from network structure to attractor landscape, in: Kriete, A., Eils, R. (Eds.),
721 Computational Systems Biology. Elsevier.

Table 1: A list of assumptions used in our method.

No.	Assumption	Justification	References
1.	100,000 uniform distribution for initial conditions.	We used the uniform distribution for the initial conditions, because we try to avoid any prior bias about initial states of trajectories in the epigenetic landscape. As such, we sample the initial cellular states from the interval of the minimum to maximum gene expression values uniformly, to cover all possible trajectories as completely as possible. The convergence of trajectories to attractors should be driven by the self-organization of GRN, rather than initial conditions. We believe that setting the number of trajectories to 100,000 would be appropriate. Our experiments showed that, with fewer trajectories, the landscape shape will be almost the same, whereas simulating more trajectories won't change the landscape shape very much, although more costly computationally.	Li and Wang (2013)
2.	When the number of molecules present in the biochemical reactions is large the stochastic and deterministic simulation results are almost equivalent.	According to the works of Daniel T. Gillespie, when the populations of reactant molecules are large, any noticeable change of the propensity function in a stochastic model would involve a large number of reaction events. In other words, if this condition is satisfied, then the effect of stochastic variability involving only a few molecules is negligible. As such, stochastic simulations can be approximated by deterministic simulations with high accuracy.	Gillespie (1977, 2001)
3.	The definition of the quasi-potential landscape U is given by the negative of probability distribution of state, $U = -\ln P(x)$.	This is a fundamental assumption based on analogy between molecular kinetics such as for cell fate regulation and statistical physics. The idea, definition and formulation of the quasi-potential stem from dynamical systems theory and statistical physics.	Wang et al. (2008b, 2010)
4.	The probability distribution of state $P(x)$ can be estimated with a coarse-grained method based on the division of a 2-D plane into grid boxes.	The cellular process of transcription is also discretized (with bursting) rather than continuous. Also, there are intrinsic and extrinsic noises affecting the cell state transitions. Therefore, the coarse-graining based on division of 2-D plane into grid boxes won't bring much bias or error to the probability estimation.	N.A.

Table 2: Comparison of our Monte Carlo method with NetLand. We compare them from three aspects, including run time, peak memory usage, and the number of detected attractors.

Case studies		Monte Carlo method	NetLand
(Gardner et al., 2000)	Run time (s)	34	33
	Memory usage (GB)	4.5	4.0
	No. of attractors	2	2
(Li and Wang, 2015)	Run time (s)	32	58
	Memory usage (GB)	4.6	4.3
	No. of attractors	8	8
(Li and Wang, 2013)	Run time (s)	90	622
	Memory usage (GB)	5.6	4.6
	No. of attractors	2	2