Quantifying the interplay between genetic and epigenetic regulations in stem cell development

Yunbin Zhang1,2, Ning Liu2, Wei Lin1,2,3,4, and Chunhe Li1,3

1 Shanghai Center for Mathematical Sciences, Fudan University, Shanghai, People’s Republic of China
2 School of Mathematical Sciences, Fudan University, Shanghai, People’s Republic of China
3 Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, People’s Republic of China
4 Shanghai Key Laboratory for Contemporary Applied Mathematics, LMNS (Fudan University) and LCNBI (Fudan University), Shanghai, People’s Republic of China

E-mail: chunheli@fudan.edu.cn

Keywords: gene networks, stem cell development, histone modifications, epigenetic landscape, transdifferentiation, mathematical models

Supplementary material for this article is available online

Abstract
Waddington epigenetic landscape, as a classic metaphor, has been used to explain cellular development and differentiation. However, it remains challenging to quantify the epigenetic landscape. Especially, a key issue arises as what are the underlying mechanisms for the interplay between genetic and epigenetic regulations to govern cell fate decisions in development. Based on a developmental epigenetic model combining histone modifications and gene regulations, we studied state switching mechanisms of histone modifications for stem cell development, and uncovered corresponding epigenetic landscape. The topography of landscape provides a quantitative measure for the relative stability of different attractors or phenotypes. We showed that histone regulations facilitate the occurrence of intermediate states or multistability. From the epigenetic landscape of stem cell differentiation, we identified key cellular states characterized by attractors, including pluripotent stem cell state, differentiated state and intermediate states. We also quantified representative kinetic transition paths for differentiation, reprogramming and transdifferentiation, which agree well with previous experimental observations. Specifically, previous experiments indicate that transdifferentiation can go through a mixed, unspecific intermediate or progenitor-like state. By calculating the kinetic transition paths, our developmental epigenetic models are able to replicate all these three experimental results, and therefore provide theoretical explanations for these experimental observations. We propose that epigenetic regulations play critical roles on the kinetic transitions for differentiation, reprogramming and transdifferentiation, which also provide a source for the heterogeneity of gene expressions observed in developmental process. Our work provides new insights into the roles of epigenetic modifications on controlling gene expression and stem cell differentiation, and facilitates our mechanistic understanding for the cell fate determinations regarding the interplay between genetic and epigenetic regulations.

Introduction

Cells change their phenotypes during development, giving rise to different lineages and differentiated cell types. This cell fate decision dynamics during development has been illustrated by Waddington using a metaphor, known as the famous Waddington epigenetic landscape [1]. Recently, it has been found that differentiated cells can be reprogrammed to induced pluripotent stem cells or other differentiated cell types by manipulating certain key transcription factors [2–5]. However, the mechanisms for cellular development and reprogramming have not been fully elucidated.
The Waddington landscapes for gene regulatory networks have been quantified to investigate the stochastic dynamics of cell fate decisions and phenotypic stability [6–14]. From the landscape perspective, different phenotypes are delineated as the basins of attraction on a potential surface, and the cell fate determination process is viewed as a ball rolling from one basin to the other on the landscape surface by surmounting certain barriers. The potential barrier heights separating the attractors or basins quantify the degrees of feasibility for cells to switch from one cell type to the other. Therefore, the landscape provides a way to quantify the stochastic dynamics of cell fate decisions in development based on underlying gene regulatory networks.

On the other hand, epigenetic regulations (such as histone modifications) are associated with active and repressed chromatin states, and therefore control gene expressions [15, 16]. Histone modifications are proposed to act as heritable marks, i.e. the chromatin state can be maintained through DNA replications and cell divisions [17]. However, the previous landscape models usually focus on the gene regulations, and did not explicitly consider the epigenetics (such as histone modifications) [8, 12].

Therefore, a critical issue arises as to how to combine the gene regulations and epigenetic modifications in developmental model and get the epigenetic developmental landscape. Computational models have been developed to study the epigenetic control mechanisms [18–26]. For example, Berry et al proposed a mathematical model to study the interplay between histone modifications and gene regulations [21]. Polycomb repressive complex 2 (PRC2) is a multiprotein complex that can methylate histone H3 at Lys-27 (H3K27) forming methylation state (H3K27me3). The positive feedback between H3K27me3 and PRC2 is suggested to contribute to the maintenance of H3K27-methylated chromatin states [27, 28]. Also, in mammalian cells PRC2 and H3K27me3 can accumulate in response to transcriptional repression and can be removed by transcriptional activation [29]. Therefore, the transcription and the methylation of H3K27 form a mutual-repression regulatory structure. It was suggested that the transcription regulations can be combined with bistable chromatin state to regulate gene expressions, and provide robust mechanisms for epigenetic memory [21]. Also epigenetic modifications are suggested to play critical roles in lineage commitment [30–32]. Despite these advances, the stochastic dynamics and global stability of the epigenetic models remain to be elucidated. More importantly, the mechanisms for how genetics coordinate epigenetic effects to control stem cell development have yet to be clarified.

Here, we aim at applying the landscape theory to investigate the stochastic dynamics of epigenetic regulations (histone modifications) for stem cell development by integrating genetic and histone regulations, and obtained the Waddington epigenetic landscape. We compare landscapes for different gene network structures with histone modifications. For a self-activation gene network model, we found that epigenetic regulations promote the occurrence of intermediate states. For a simplified stem cell differentiation epigenetic model (with a mutual repressed self-activation structure), we identified different attractors representing different cell states, including pluripotent stem cell state, differentiated state and intermediate states. This provides a possible explanation for the heterogeneous population distribution of the marker gene expressions observed experimentally. We also found that dynamical processes for differentiation, reprogramming and transdifferentiation can take different routes in the space of gene expression levels. For example, the differentiation and reprogramming can be finished directly or go through other intermediate states, while the transdifferentiation can go through a mixed (MX), unspecific intermediate (UI) or progenitor-like (PG) state. These predictions agree well with previous studies [33].

Our work establishes a general approach to investigate the interplay between genetic and epigenetic regulations, and can be applied to other gene regulatory network models. Our studies provide new insights into the roles of epigenetic regulations in stem cell development, and provide a possible explanation for the heterogeneity in cellular populations. This will further our understanding for the mechanism of genetic and epigenetic regulations in cell fate decision processes of development.

Models and methods

Mathematical model of chromatin
To combine genetic and epigenetic regulations, we first construct the Chromatin model based on previous work [21]. We considered the H3K27 methylation, demethylation using epigenetic model [21], and coupled this epigenetic model to different gene network model, including single gene transcription model, self-activation model and mutual repressed self-activation model (simplified developmental model). For the coupling between epigenetic model and simplified developmental model, we aim to obtain an epigenetic model to illuminate the interplay between genetic and epigenetic regulations in development.

First, we consider the single gene transcription model coupled with epigenetic regulations. In our model, the simulated region of the chromatin domain contains N/2 nucleosomes, each of which consists of a pair of histones (H3). Here, N denotes the number of histones (we consider N = 60 histones here). Histones outside
the simulated region are not considered. We chose histone number \( N = 60 \) by following [18, 21]. This is roughly corresponding to an average gene size (about 10 kb). The value of \( N \) can be adjusted for a specific gene.

In the activated state, Pol II can carry H3K27-demethylases (KDM) to promote demethylation and nucleosome exchange. In the repressed state, H3K27me3 (orange hexagons) can recruit PRC2, which can promote H3K27me3. This forms a positive feedback loop. (B) Diagrammatic representation for the epigenetic model of single gene transcription and self-activated gene. The elliptical nodes from left to right (me0 to me3) represent the methylation state of H3K27. Here, four histone methylation states (me0/me1/me2/me3) can switch to each other by methylation or demethylation reactions, histone methylation states affect transcription rate \( f \) (orange repression arrows) and methylation reactions of their neighbors (orange active arrows). Transcription also promotes histone demethylation (green arrows). The module with yellow arrows represent the process of gene expression for a self-activated gene. Black arrows represent state transitions (STs). DNA* represents the DNA with protein bound to the promoter. \( f_\text{on} \) and \( f_\text{off} \) indicate the transcription rate for gene without and with protein bound, respectively. (C) Mathematical description of model. Sum over neighbors in \( E_i \) includes the other histone on the same nucleosome, and four histones on the neighboring nucleosomes. So the total number of the neighbors for each histone \( i \) is 5. \( \beta_{\text{me}_2,\text{me}_3} \) is the fraction of H3 histones carrying K27me2 or K27me3. \( \alpha > 1 \) is the self-activation constant. A detailed description of each parameter can be found in table 3.

Figure 1. Chromatin model of PRC2 Target-Gene. (A) Illustration diagram of two different chromatin states (methylation and demethylation). In the active state, Pol II can carry H3K27-demethylases (KDM) to promote demethylation and nucleosome exchange. In the repressed state, H3K27me3 (orange hexagons) can recruit PRC2, which can promote H3K27me3. This forms a positive feedback loop. (B) Diagrammatic representation for the epigenetic model of single gene transcription and self-activated gene. The elliptical nodes from left to right (me0 to me3) represent the methylation state of H3K27. Here, four histone methylation states (me0/me1/me2/me3) can switch to each other by methylation or demethylation reactions, histone methylation states affect transcription rate \( f \) (orange repression arrows) and methylation reactions of their neighbors (orange active arrows). Transcription also promotes histone demethylation (green arrows). The module with yellow arrows represent the process of gene expression for a self-activated gene. Black arrows represent state transitions (STs). DNA* represents the DNA with protein bound to the promoter. \( f_\text{on} \) and \( f_\text{off} \) indicate the transcription rate for gene without and with protein bound, respectively. (C) Mathematical description of model. Sum over neighbors in \( E_i \) includes the other histone on the same nucleosome, and four histones on the neighboring nucleosomes. So the total number of the neighbors for each histone \( i \) is 5. \( \beta_{\text{me}_2,\text{me}_3} \) is the fraction of H3 histones carrying K27me2 or K27me3. \( \alpha > 1 \) is the self-activation constant. A detailed description of each parameter can be found in table 3.
activating PRC2 [21]. So, our model considers the methylation and demethylation (transition reactions among me0, me1, me2 and me3), the effects of transcription on methylation (green arrows), the effects of methylation states on transcription (orange repressive arrows), and PRC2-mediated positive feedback promoting methylation (orange arrows). To our knowledge, the mechanism for how the transcription influences the histone methylation remains to be clarified. So, for simplification, we modeled demethylation and histone exchange processes only depending on the default transcription rate $f$.

Let $S_i$ denotes the state of each H3 histone ($S_i \in \{\text{me0, me1, me2, me3}\}$, $i \in [1, N]$). The state of $S_i$ will change as each methylation or demethylation reaction happens (figure 1(B)). For $1 \leq i \leq N$, the rates of methylation for the $i$th histone (denoted by $r^{\text{me}}_i$) are calculated as [21]:

\[
\begin{align*}
    r^{\text{me}}_i = & \delta_{S_i,\text{me0}}(\gamma_{\text{me0-1}} + k_{\text{me0-1}}E_i) + \delta_{S_i,\text{me1}}(\gamma_{\text{me1-2}} + k_{\text{me1-2}}E_i) + \delta_{S_i,\text{me2}}(\gamma_{\text{me2-3}} + k_{\text{me2-3}}E_i), \\
    E_i = & \sum_{j \in M_i}(\rho_{\text{me2}}\delta_{S_j,\text{me2}} + \delta_{S_j,\text{me3}}), S_j \in \{\text{me0, me1, me2, me3}\}.
\end{align*}
\]

(1)

Here, $k_{\text{me0-1}}, k_{\text{me1-2}}$ and $k_{\text{me2}}$ are PRC2-mediated methylation rate and $\gamma_{\text{me0-1}}, \gamma_{\text{me1-2}}$ and $\gamma_{\text{me2-3}}$ are noisy methylation rate. $E_i$ incorporates the positive feedback from neighboring me2/me3, and $\rho_{\text{me2}} = 0.1$ accounts for the reduced efficiency of PRC2 activation from H3K27me2. $\delta_{i,j}$ is the Kronecker delta function, equal to 1 if $i = j$ and 0 otherwise, and $E_i$ is summed over ‘neighboring’ histones, where

\[
M_i = \begin{cases} 
    \{i-2, i-1, i+1, i+2, i+3\}, & \text{if $i$ odd,} \\
    \{i-3, i-2, i-1, i+1, i+2\}, & \text{if $i$ even.}
\end{cases}
\]

Here, each histone $i$ ($3 \leq i \leq N - 2$) has 5 neighbors except the 4 histones ($i = 1, 2, 58, 59$) in the boundary nucleosomes. For these 4 histones ($i = 1, 2, 58, 59$), each histone has three neighbors. We consider that this boundary effect will be small since the region of chromatin domain we are simulating is relatively large (60 histones) relative to the boundaries (4 histones).

The rates of transcription-independent demethylation, $r^{\text{dem}}_i$ are calculated as:

\[
r^{\text{dem}}_i = \gamma_{\text{dem}}(\delta_{S_i,\text{me1}} + \delta_{S_i,\text{me2}} + \delta_{S_i,\text{me3}}),
\]

where $\gamma_{\text{dem}}$ is noisy demethylation rate. Transcription-dependent demethylation results in the removal of methylation groups at each histone and also replacement of each nucleosome. The rates of removal of methylation groups and histone exchange rate are represented by $f_{\text{dem}}^{\text{in}}$ and $f_{\text{ex}}^{\text{in}}$, respectively (figure 1(C)).

**ST approach for simulations**

Stochastic reaction kinetics can be simulated by the Stochastic Simulation algorithm (SSA) [35], but the SSA can be low efficient for large chemical reaction systems. Based on the above epigenetic model, we improved the simulation approach by tracking the number of histones in different methylation states within a simulated region of chromatin instead of considering the state of each histone explicitly. Starting from the current state, we only consider which reaction occurs at the next moment without considering the specific histone location for the reaction. In this case, the number of possible reactions is reduced from $2N$ (N methylation reactions and N demethylation reactions) to 2 (methylation and demethylation) for a system including N histones. This significantly reduces the computational complexity of the system.

We will use methylation and histone exchange as examples to to describe the improved approach, which we call the ST approach. In the original model, the propensities of methylation are given by $r^{\text{me}}_i$ ($i \in [1, N]$). If we don’t consider where the reaction occurs, there are three possibilities for methylation reactions, which occur at me0, me1, and me2 state, respectively. Let $Z_i$ denote the number of histones in me state, and $z_i = Z_i/N$ ($l \in \{0, 1, 2, 3\}$) represent the percentage of histones in four different methylation states. The propensities of methylation are recalculated as:

\[
\begin{align*}
    r_{\text{me0}} &= \sum_{i=1}^{N}\delta_{S_i,\text{me0}}(\gamma_{\text{me0-1}} + k_{\text{me0-1}}E_i), \\
    r_{\text{me1}} &= \sum_{i=1}^{N}\delta_{S_i,\text{me1}}(\gamma_{\text{me1-2}} + k_{\text{me1-2}}E_i), \\
    r_{\text{me2}} &= \sum_{i=1}^{N}\delta_{S_i,\text{me2}}(\gamma_{\text{me2-3}} + k_{\text{me2-3}}E_i),
\end{align*}
\]

(2)

where $E_i$ incorporates the positive feedback from neighboring me2/me3. We use expected value to estimate this effect, and we have:
where $r_{m0} = (\gamma_{m0.1} + k_{m0.1}E_{me}) \times Z_0$.

$r_{me} = (\gamma_{me1.2} + k_{me1.2}E_{me}) \times Z_1$.

$r_{me2} = (\gamma_{me2.3} + k_{me2.3}E_{me}) \times Z_2$.

Similarly, for histone exchange reactions, there are totally 16 possible situations (figure 1(C)). We take one of the situations as an example. If the reaction occurs at $m_{0}$ at a certain location, the histone status at the neighbor position is $m_{0}$, the propensity of this situation is calculated as $fp_{e} \times z_{i} \times z_{o}$.

We made simulations by tracking the state of the system which is given by $(Z_0, Z_1, Z_2, Z_3)$. The reaction propensity function and state change of each reaction channel are shown in table 1. With above propensities, we can calculate the transition probability and state change for each reaction channel.

Table 1. Propensity functions and state change for the histone modification.

| Reaction type | Reaction channel | Propensity function | State change |
|---------------|------------------|---------------------|--------------|
| Methylation   | $m_0 \rightarrow m_1$ | $(\gamma_{m0.1} + k_{m0.1}E_{me}) \times Z_0$ | $(-1, 1, 0, 0)$ |
|               | $m_1 \rightarrow m_2$ | $(\gamma_{m1.2} + k_{m1.2}E_{me}) \times Z_1$ | $(0, -1, 1, 0)$ |
|               | $m_2 \rightarrow m_3$ | $(\gamma_{m2.3} + k_{m2.3}E_{me}) \times Z_2$ | $(0, 0, -1, 1)$ |
| Demethylation | $m_1 \rightarrow m_0$ | $(\gamma_{dm} + \beta_{dm}) \times Z_0$ | $(1, -1, 0, 0)$ |
|               | $m_2 \rightarrow m_1$ | $(\gamma_{dm} + \beta_{dm}) \times Z_1$ | $(0, 1, -1, 0)$ |
|               | $m_3 \rightarrow m_2$ | $(\gamma_{dm} + \beta_{dm}) \times Z_2$ | $(0, 0, 1, -1)$ |

Histone exchange

| (me0, me0)   | $fp_{e} \times Z_0 \times Z_0$ | $(0, 0, 0, 0)$ |
| me0, me1    | $fp_{e} \times Z_0 \times Z_1$ | $(1, -1, 0, 0)$ |
| me0, me2    | $fp_{e} \times Z_0 \times Z_2$ | $(0, 1, -1, 0)$ |
| me0, me3    | $fp_{e} \times Z_0 \times Z_3$ | $(0, 0, 1, -1)$ |
| me1, me0    | $fp_{e} \times Z_1 \times Z_0$ | $(1, -1, 0, 0)$ |
| me1, me1    | $fp_{e} \times Z_1 \times Z_1$ | $(2, -1, 0, 0)$ |
| me1, me2    | $fp_{e} \times Z_1 \times Z_2$ | $(2, 0, -1, 0)$ |
| me1, me3    | $fp_{e} \times Z_1 \times Z_3$ | $(2, 0, 0, -1)$ |
| me2, me0    | $fp_{e} \times Z_2 \times Z_0$ | $(1, -1, 0, 0)$ |
| me2, me1    | $fp_{e} \times Z_2 \times Z_1$ | $(2, -1, 0, 0)$ |
| me2, me2    | $fp_{e} \times Z_2 \times Z_2$ | $(2, 0, -1, 0)$ |
| me2, me3    | $fp_{e} \times Z_2 \times Z_3$ | $(2, 0, 0, -1)$ |
| me3, me0    | $fp_{e} \times Z_3 \times Z_0$ | $(0, 1, -1, 0)$ |
| me3, me1    | $fp_{e} \times Z_3 \times Z_1$ | $(1, 0, 0, -1)$ |
| me3, me2    | $fp_{e} \times Z_3 \times Z_2$ | $(1, 0, -1, 0)$ |
| me3, me3    | $fp_{e} \times Z_3 \times Z_3$ | $(1, 0, 0, -2)$ |

$E_{me} = E[E_{i}] = E[\sum_{j \in M_{i}} (\rho_{me2} \delta_{S_{j}, me2} + \delta_{S_{j}, me3})] = \sum_{j \in M_{i}} (\rho_{me2}E[\delta_{S_{j}, me2}] + E[\delta_{S_{j}, me3}]) = \sum_{j \in M_{i}} (\rho_{me2}P(S_{j} = me2) + P(S_{j} = me3)) = 5(\rho_{me2}z_2 + z_3)$. (4)

Through this approach, the calculation speed of the algorithm can be significantly improved compared to using SSA. The improved ST approach can be considered as an approximation from original SSA models. By making comparisons for calculating distributions using the ESG model as an example (figure S1 available online at stacks.iop.org/NJP/21/103042/mmedia), we show that the relative error (1.87%) is very small for using our ST approach compared to using SSA. For estimating average transition time between high protein level state and low protein level state, ST approach is not as accurate as calculating landscape compared with SSA approach (with a relative error of 8.7% and 38.4%, respectively for high to low protein level and low to high protein level transition, see figure S1). However, considering the approximation used in ST approach and complexity that exists in our dynamical models, we believe that ST approach, as an approximation approach, provides a useful way to study landscape and transition time.
increasing the maximal transcription rate, in which the formula changes to:

\[
\text{methylation, as indicated in above formula. In the second way, self-activated transcription is modeled as also}
\]

assumed the transcriptional rate  of gene as

dimer and activate transcriptional regulation of gene expression, we hypothesized that the protein molecules bind to DNA as a

by combining the transcriptional regulation module using direct Gillespie stochastic algorithm and histone

cation module using improved ST approach, we can perform the stochastic simulations for the full epigenetic model of self-activated gene (see parameters values in table 3). We did not recorded the DNA state in our simulations here. We believe that this does not influence our conclusions, but warrants further explorations.

| Reaction channel | Propensity function | State change |
|------------------|---------------------|--------------|
| DNA \(\rightarrow\) mRNA | \(f\) | \((1, 0)\) |
| DNA* \(\rightarrow\) mRNA | \(f_h\) | \((1, 0)\) |
| mRNA \(\rightarrow\) \(\varnothing\) | \(d_m \times m\) | \((-1, 0)\) |
| mRNA \(\rightarrow\) Protein | \(s_p \times m\) | \((0, 1)\) |
| Protein \(\rightarrow\) \(\varnothing\) | \(d_p \times p\) | \((0, -1)\) |
| DNA \(\rightarrow\) DNA* | \(\alpha_0 (p - 1)/2\) | \((0, -2)\) |
| DNA* \(\rightarrow\) DNA | \(\alpha_1\) | \((0, 2)\) |

Table 2. Propensity functions and state change for transcriptional self-activation

Table 3. Parameter values for single gene epigenetic model. For the epigenetic model of self-activated gene, we keep the parameter values same except setting \(p_{dom} = 0.0012\) and \(\alpha = 2\).

| Parameter | Description | Value | References |
|-----------|-------------|-------|------------|
| \(k_{me}\) | PRC2-mediated methylation rate (me2 to me3) (histone\(^{-1}\) s\(^{-1}\)) | \(10^{-5}\) | Estimated |
| \(p_{dom}\) | Demethylation probability (histone\(^{-1}\) transcription\(^{-1}\)) | \(1.4 \times 10^{-3}\) | Estimated |
| \(p_{es}\) | Histone exchange probability (histone\(^{-1}\) transcription\(^{-1}\)) | \(1.4 \times 10^{-2}\) | Estimated |
| \(\rho_{me2}\) | Relative activation of PRC2 by H3K27me2 | 0.1 | |
| \(k_{me1-0}\) | PRC2-mediated methylation rate (me0 to me1) (histone\(^{-1}\) s\(^{-1}\)) | \(9k_{me}\) | [21] |
| \(k_{me1-2}\) | PRC2-mediated methylation rate (me1 to me2) (histone\(^{-1}\) s\(^{-1}\)) | \(6k_{me}\) | [21] |
| \(\gamma_{me0-1}\) | Noisy methylation rate (me0 to me1) (histone\(^{-1}\) s\(^{-1}\)) | \(k_{me0-1}\) | [21] |
| \(\gamma_{me1-2}\) | Noisy methylation rate (me1 to me2) (histone\(^{-1}\) s\(^{-1}\)) | \(k_{me1-2}\) | [21] |
| \(\gamma_{me2-3}\) | Noisy methylation rate (me2 to me3) (histone\(^{-1}\) s\(^{-1}\)) | \(k_{me}\) | [21] |
| \(\gamma_{dem}\) | Noisy demethylation rate (histone\(^{-1}\) s\(^{-1}\)) | \(f_{me0, p_{dom}}\) | [21] |
| \(N\) | Number of histones | 60 | |
| \(f_{max}\) | Maximum transcription initiation rate (mRNA s\(^{-1}\)) | \(4 \times 10^{-3}\) | [21] |
| \(f_{min}\) | Minimum transcription initiation rate (mRNA s\(^{-1}\)) | \(10^{-4}\) | [21] |
| \(d_m\) | mRNA degradation rate (h\(^{-1}\)) | 0.5 | [21] |
| \(s_p\) | Protein production rate (protein h\(^{-1}\)) | 0.5 | [21] |
| \(d_p\) | Protein degradation rate (h\(^{-1}\)) | 0.25 | Estimated |
| \(\alpha\) | Self-activation constant | 1 | Estimated |
| \(\alpha_0\) | protein-DNA binding rate for self-activation (protein\(^{-1}\) s\(^{-1}\)) | \(k_r\) | Estimated |
| \(\alpha_1\) | protein-DNA unbinding rate for self-activation (protein s\(^{-1}\)) | \(10^{-4}\) | Estimated |
| \(k_r\) | Equilibrium constant of self-activation (protein\(^{-2}\)) | 0.01 | Estimated |

Epigenetic model of self-activated gene
The mechanistic basis of transcriptional repression by PRC2 and H3K27me2/me3 is poorly understood. We assume the transcriptional rate \(f\) a simple linear function of the proportion of H3K27me2/me3 marked histones at the gene as \(f = f_{max} - P_{me2, me3}(f_{max} - f_{min})\), where \(P_{me2, me3}\) is the proportion of me2/me3 marks [21]. In the transcriptional regulation of gene expression, we hypothesized that the protein molecules bind to DNA as a dimer and activate (or inhibit) gene expression. In our model, DNA* represents the DNA state with promoter bound to the protein (figure 1(B)). For the self-activation regulation, the transcription rate takes the form as

\[
\bar{f}_h = f_{max} - P_{me2, me3}(f_{max} - f_{min})/\alpha,
\]

where \(\alpha > 1\) is defined as the self-activation constant, and we have \(P_{me2, me3} = z_2 + z_3\). Here, we modeled transcriptional activation as a factor to weaken the role of histone methylation, as indicated in above formula. In the second way, self-activated transcription is modeled as also increasing the maximal transcription rate, in which the formula changes to:

\[
\bar{f}_h = f_{max} f_{\alpha_m} - P_{me2, me3}(f_{max} / \alpha_m - f_{min}) / \alpha.
\]

Here, \(\alpha_m < 1\) represents that self-activation will also increase the maximal transcription rate (figure S2).

For the transcriptional regulation module, we use \(m\) and \(p\) to represent numbers of mRNA and protein respectively. The reaction propensity function and state change of each reaction channel are shown in table 2. By combining the transcriptional regulation module using direct Gillespie stochastic algorithm and histone modification module using improved ST approach, we can perform the stochastic simulations for the full epigenetic model of self-activated gene (see parameters values in table 3). We did not recorded the DNA state in our simulations here. We believe that this does not influence our conclusions, but warrants further explorations.
Results

Epigenetic landscape for single gene and self-activated gene
To study the key properties of epigenetic regulations interacting with gene regulations, we first examine a simple single gene example, in which only the transcription and translation of one gene are considered, and couple it with the epigenetic regulations (figure 1(B), no yellow arrows). We call this model epigenetic model for single gene (ESG model). Figure 2 shows the simulation results for the ESG model. In this model, no gene regulation is considered, so only epigenetic modifications take roles. Here we plot both trajectories (figures 2(A)–(D)) for different components and potential landscape with different pairs of components (figures 2(E)–(I)). From the trajectories, we see that the system switches between two different states (low and high protein level) under fluctuations. However, the global stability is not easy to study only from trajectories. So we quantified the potential energy landscape based on the statistics for the long time trajectories [9]. Here, the potential landscape is defined as: \( U = -\log P_{ss} \), \( P_{ss} \) is steady state probability distribution. The landscape surface is characterized by different colors, in which the blue region represents lower potential or higher probability, and the red region represents higher potential or lower probability. On the landscape the blue region represents lower potential or higher probability, and the red region represents higher potential or lower probability.

![Figure 2. Trajectories and landscape for the epigenetic model of single gene transcription (figure 1(B), without yellow arrows). (A)–(D) Trajectories of mRNA, Protein, me3 and transcription rate f. (E)–(G) Two dimensional landscape for different pairs of components using me3/mRNA, me0/Protein and me0/me3. (H)–(I) Three dimension landscape for me3 versus mRNA and me3 versus Protein. The values of the parameters are shown in table 3. The potential landscape is defined as: \( U = -\log P_{ss} \), \( P_{ss} \) is steady state probability distribution. On the landscape the blue region represents lower potential or higher probability, and the red region represents higher potential or lower probability.](image-url)
exhibits some of the hallmarks found in developmental systems because many self-activation regulations have been suggested from developmental systems [8, 36]. We call this model self-activated epigenetic (SAE) model. Figure 3 shows the trajectories of different components and the potential landscape (using different pairs as the coordinates) for the SAE model. We found that three stable states or attractors emerge on the epigenetic landscape (figures 3(E), (F), (H) and (I)), i.e. the self-activated gene can have three different expression levels (also see figure S2 for a different modeling strategy, where self-activated transcription is modeled as also increasing the maximal transcription rate). Specifically, in the methylation state (me3 high), the mRNA and protein both have two different levels (figures 3(E) and (F)), i.e. the gene or protein can have a middle level state. It can be imagined that the middle expression state will occur when the methylation and self-activation are both on because the methylation inhibits gene expression while the self-activation promotes gene expression, which results in a middle expression level state. Previous work has suggested that a self-activated gene can give rise to bistability (low and high expression) [37], where the epigenetic modifications (such as histone modifications) were not been explicitly explored. Here, our landscape results suggest that the epigenetic regulations can promote the multistability or the intermediate state. In the single gene case, gene can have low, high and middle expression level state. Of note, the intermediate state we identified here is related with our regulatory network coupling genetic and histone regulations (figure 1(B)). Our histone modification network is based on the work of Berry et al [21], and a major difference compared with previous histone modification models [18] is that we explicitly considered the interplay between transcription and histone modifications.

**Mutual repressed self-activation model (simplified developmental epigenetic model)**

Epigenetic factors have been long related to the cell fate decisions in development and differentiation [30–32]. But the mechanisms for the roles of epigenetic regulations in this process remain elusive. To address this problem, we developed an epigenetic model for development based on a gene circuit consisting of a pair of self-
activated and mutually repressed genes. This self-activated mutual repression gene circuit has been commonly studied as an example for developmental system [6, 38]. In fact, this gene regulatory circuit has been found in many tissues where a pluripotent stem cell needs to go through a binary cell fate determination for the lineage commitment [38]. For example, in the multipotent common myeloic progenitor cell (CMP), a binary cell fate decision has to be made between the myeloid and the erythroid fate. The transcription factors PU.1, and GATA1, promote the myeloid and the erythroid fates, respectively, which forms such a circuit [5]. We need to stress that for the model of self-activation and mutual repression with self-activation, only a few of the vast literature are cited.

Here, we developed a simplified developmental epigenetic model by combining the histone modification model with the mutual repressed self-activation gene network structure (gene A and B mutually repress each other). In this model, two genes mutually repress each other and self-activate themselves, and each gene (A and B) is combined with the epigenetic modification model as illustrated in figure 4. The simplified developmental epigenetic model can be expressed by the following chemical reactions:

\[
D_{A}^{00} + 2p_{A} \xrightarrow{\alpha_{A}} D_{A}^{01}, D_{A}^{10} + 2p_{A} \xrightarrow{\beta_{A}} D_{A}^{11} \\
D_{A}^{00} + 2p_{B} \xrightarrow{\alpha_{B}} D_{B}^{01}, D_{B}^{10} + 2p_{B} \xrightarrow{\beta_{B}} D_{B}^{11} \\
D_{B}^{00} + 2p_{A} \xrightarrow{\alpha_{A}} D_{A}^{01}, D_{A}^{10} + 2p_{A} \xrightarrow{\beta_{A}} D_{A}^{11} \\
D_{B}^{00} + 2p_{B} \xrightarrow{\alpha_{B}} D_{B}^{01}, D_{B}^{10} + 2p_{B} \xrightarrow{\beta_{B}} D_{B}^{11} \\
D_{A}^{00} \xrightarrow{f} m_{A}, D_{A}^{01} \xrightarrow{f} m_{A}, D_{A}^{10} \xrightarrow{f} m_{A}, D_{A}^{11} \xrightarrow{f} m_{A} \\
D_{B}^{00} \xrightarrow{f} m_{B}, D_{B}^{01} \xrightarrow{f} m_{B}, D_{B}^{10} \xrightarrow{f} m_{B}, D_{B}^{11} \xrightarrow{f} m_{B} \\
m_{A} \xrightarrow{\gamma_{A}} p_{A}, m_{B} \xrightarrow{\gamma_{B}} p_{B}.
\]

Here, \(D_{A}^{ij}\) and \(D_{B}^{ij}\) represent gene A and gene B, respectively. \(m_{A}\) and \(m_{B}\) represent the mRNA of gene A and gene B, respectively. \(p_{A}\) and \(p_{B}\) represent the protein of gene A and gene B, respectively. Due to the mutual repression and self-activation structure, each gene can have four possible states (fig. 4). Taking gene A as an example, for the gene state index \(i\), the first index \(i = 0(1)\) stands for gene A unbound (bound) to protein B (repression); the second index \(j = 0(1)\) stands for gene A unbound (bound) to protein A (self-activation). Here, \(\alpha_{A}(\alpha_{B})\) represent the binding (unbinding) rate constant for self-activation, and \(\beta_{A}(\beta_{B})\) represent the binding (unbinding) rate constant for repression. For these four states, the rates of transcription become:

\[
\begin{align*}
\tilde{f} &= f_{\max} - \frac{p_{\text{me2,me3}}(f_{\max} - f_{\text{min}})}{\alpha_{A}}, \\
\tilde{f}_{s} &= f_{\max} - \frac{p_{\text{me2,me3}}(f_{\max} - f_{\text{min}})}{\alpha_{B}}, \\
\tilde{f}_{i} &= f_{\max} / \beta_{A} - \frac{p_{\text{me2,me3}}(f_{\max} / \beta_{A} - f_{\text{min}})}{\alpha_{B}},
\end{align*}
\]

where \(\alpha, \beta > 1\) are defined as the self-activation and repression constant, respectively, and \(p_{\text{me2,me3}} = \delta_{Z} + \delta_{Z_{b}}\).

Here, \(f\) represents the transcription rate for gene without bound to any protein, whose range is \([f_{\text{min}}, f_{\text{max}}]\); \(f_{b}\) represents transcription rate for the gene only bound to the activator protein without bound to the repressor protein, whose range is \([f_{\max} - (f_{\max} - f_{\min})/\alpha_{B}, f_{\text{max}}]\), which has the highest transcription rate, and \(f_{i}\) represents transcription rate for the gene bound to the repressor protein, whose range is \([f_{\min}, f_{\text{max}} / \beta_{A}]\), which has the lowest transcription rate.

In our algorithm, we divide the reactions of the system into three groups. Where the first group is for the reactions of gene A, the second group is for the reactions of gene B, and the third group is for the reactions related to the interplay between gene A and gene B. In the process of simulation, we first determine in which group the next reaction will occur, and then determine which reaction will occur in corresponding group. Let \(p_{A}, p_{B}\) denote the level of protein A and protein B, respectively. The system state is given by \((p_{A}, p_{B})\). With above reactions, we combined SSA (for transcription and translation) and our ST approach (for histone methylation) to perform the simulations for the developmental epigenetic model (see parameter values in table 4).

**Epigenetic landscape for development**

We further quantified the landscape for the developmental epigenetic model (fig. 4). Figure 5 shows the trajectories and landscape for gene A and B, respectively. In figure 5, the trajectories and landscape for gene A (figures 5(A)–(F)) or gene B (figures 5(G)–(L)) are similar, since this is a simplified symmetric model. Take gene A as an example, we found that gene A can have four different states in the space of protein level and me3 level (figure 5(E)), which also suggests that protein A can have three different expression levels (high, low and middle). We can further analyze the reasons for the appearance of attractors. Take gene A as an example, we see four
attractors on the landscape (figures 5(D) and (E)). For the attractor in the up-left corner, the transcriptional repression (from gene B) and the histone methylation are both off, so that gene A can have highest expression level. For the attractor in the bottom-left corner, the methylation is off but the transcriptional repression is on, so the combination effects lead to a middle expression level for gene A. For the attractor in the bottom-right corner,
Table 4. Parameter values for mutual repressed self-activation model (simplified developmental epigenetic model). For figures 7(M–P), the parameters are changed as $k_n = 20, k_p = 0.001$ while keeping other parameters the same.

| Parameter       | Description                                                                 | Value          | References |
|-----------------|------------------------------------------------------------------------------|----------------|------------|
| $k_{me}$        | PRC2-mediated methylation rate (me2 to me3) (histone$^{-1}$ s$^{-1}$)         | $10^{-5}$      | Estimated  |
| $p_{sem}$       | Demethylation probability (histone$^{-1}$ transcription$^{-1}$)               | $1.8 \times 10^{-3}$ | Estimated  |
| $p_{ex}$        | Histone exchange probability (histone$^{-1}$ transcription$^{-1}$)            | $1.4 \times 10^{-2}$ | Estimated  |
| $\rho_{me2}$   | Relative activation of PRC2 by H3K27me2                                        | 0.1            | [21]       |
| $k_{me0\rightarrow1}$ | PRC2-mediated methylation rate (me0 to me1) (histone$^{-1}$ s$^{-1}$)     | $9k_{me}$      | [21]       |
| $k_{me1\rightarrow2}$ | PRC2-mediated methylation rate (me1 to me2) (histone$^{-1}$ s$^{-1}$)     | $6k_{me}$      | [21]       |
| $\gamma_{me0\rightarrow1}$ | Noisy methylation rate (me0 to me1) (histone$^{-1}$ s$^{-1}$)   | $k_{me0\rightarrow1}$ | [21]       |
| $\gamma_{me1\rightarrow2}$ | Noisy methylation rate (me1 to me2) (histone$^{-1}$ s$^{-1}$)   | $k_{me1\rightarrow2}$ | [21]       |
| $\gamma_{me2\rightarrow3}$ | Noisy methylation rate (me0 to me1) (histone$^{-1}$ s$^{-1}$)   | $k_{me2\rightarrow3}$ | [21]       |
| $\gamma_{dem}$ | Noisy demethylation rate (histone$^{-1}$ s$^{-1}$)                           | $f_{max}P_{smax}$ | [21]       |
| $f_{max}$       | Maximum transcription initiation rate (mRNA s$^{-1}$)                        | $4 \times 10^{-5}$ | [21]       |
| $f_{min}$       | Minimum transcription initiation rate (mRNA s$^{-1}$)                        | $10^{-4}$      | [21]       |
| $d_m$           | mRNA degradation rate (h$^{-1}$)                                              | 0.3            | Estimated  |
| $p_M$           | Protein production rate (h$^{-1}$)                                            | 0.3            | Estimated  |
| $d_P$           | Protein degradation rate (h$^{-1}$)                                            | 1/6            | Estimated  |
| $\alpha$        | Self-activation constant                                                     | 10             | Estimated  |
| $\beta$         | Mutual repression constant                                                    | 2              | Estimated  |
| $c_0$           | protein-DNA binding rate for self-activation (protein$^{-1}$ s$^{-1}$)       | $k_n\alpha$    | Estimated  |
| $c_1$           | protein-DNA unbinding rate for self-activation (protein$^{-1}$ s$^{-1}$)     | $10^{-5}$      | Estimated  |
| $\beta_0$       | protein-DNA binding rate for mutual repression (protein$^{-1}$ s$^{-1}$)     | $k_m\beta_1$   | Estimated  |
| $\beta_1$       | protein-DNA unbinding rate for mutual repression (protein$^{-1}$ s$^{-1}$)   | $10^{-6}$      | Estimated  |
| $k_2$           | Equilibrium constant of self-activation (protein$^{-2}$)                     | 0.2            | Estimated  |
| $k_3$           | Equilibrium constant of mutual repression (protein$^{-2}$)                   | 0.2            | Estimated  |

Figure 5. Trajectories and landscape for the developmental model of stem cell differentiation. (A)–(C) Trajectories of mRNA, Protein, me3 for gene A. (D)–(F) Two dimensional landscape for different pairs of components for gene A, me3/mRNA, me0/Protein and me0/me3. (G)–(I) Trajectories of mRNA, Protein, me3 for gene B. (J)–(L) Two dimensional landscape for different pairs of components for gene B, me3/mRNA, me0/Protein and me0/me3. The values of the parameters are shown in table 4. The potential landscape is defined as: $U = -\log P_s$, and $P_s$ is steady state probability distribution. On the landscape the blue region represents lower potential or higher probability, and the red region represents higher potential or lower probability.
the methylation is on and there is no self-activation regulations, which leads to the lowest expression level for gene A. On the contrary, for the attractor in the up-right corner, the self-activation regulation is on and repression is off, which results in a high level expression for gene A although the methylation is also on.

From the color on the landscape, which denotes the potential U, we also see that the middle expression level state (for both A and B) is very stable as it occupies a very deep landscape basin (figures 5(D) and (E)). When we further plot the landscape (figures 6(A) and (B)) and the steady state distribution (figures 6(C) and (D)) of stem cell differentiation in terms of gene A and B, we identify the middle expression state as the pluripotent stem cell state (SC), which indicates that stem cell state is a stable cellular state. In fact, we discovered six attractors emerging on the landscape, including pluripotent stem cell state (SC), differentiated states (DA and DB) and some intermediate states (IA, IB and IO). This again suggests that the epigenetic regulations promote multistability or intermediate states. So, epigenetic regulation provides a source for the heterogeneity observed in the populations of development. Here, the intermediate expression level state is defined as stem cell state because this state is in a symmetrical position, which can go left (express A) or right (express B), and therefore corresponds to the pluripotent properties for stem cell. When the symmetry is broken, the cell has to choose to go to a state where gene A is expressed (DA) or a state where gene B is expressed (DB), which corresponds to a cell fate decision process in lineage commitment. Besides these stable pluripotent and differentiated cell states (SC, DA and DB), we found some other attractor states, which we define as the intermediate states (IA, IB and IO). The intermediate states have been suggested to play critical roles during cell lineage transitions [9, 39–41] and cancer progression [11].

**Figure 6.** Epigenetic landscape for stem cell developments. Landscape and probability distribution are separately shown in 3-dimensional (A for landscape and C for probability distribution) and 2-dimensional (B for landscape and D for probability distribution) figures. SC represents pluripotent stem cell, DA and DB represent two differentiated cell types, IA, IB and IO represent three intermediate states. The potential landscape is defined as: \( U = -\log P_{ss} \), and \( P_{ss} \) is steady state probability distribution. On the landscape the blue region represents lower potential or higher probability, and the red region represents higher potential or lower probability.

**Kinetic transition paths for differentiation, reprogramming and transdifferentiation**

From the simulation trajectories of the developmental epigenetic model, we can obtain some representative transition paths between different cell types (basins). These paths will provide us the information on how the
transition happens in terms of the changes in protein (gene) levels for different transition process. Specifically, we acquired some typical transition paths for differentiation (figures 7(A)–(D)), reprogramming (figures 7(E)–(H)), and transdifferentiation (I)–(P), individually. Each column represents a typical transition path for differentiation (A)–(D), reprogramming (E)–(H), and transdifferentiation (I)–(P), individually. Magenta, white, and red lines correspond to three stages (early, middle and late stages) for the transitions in temporal order. These kinetic paths indicate that transdifferentiation can go through a mixed (MX), unspecific intermediate (UI) or progenitor-like (PG) state. The potential landscape is defined as $U = -\log P_x$, and $P_x$ is steady state probability distribution. On the landscape the blue region represents lower potential or higher probability, and the red region represents higher potential or lower probability. For (M)–(P), the parameters are changed as $k_x = 20$, $k_y = 0.001$ while keeping other parameters the same.

In the same way, we can discover the transition paths for transdifferentiation. Here, to see more possible transdifferentiation path, we use two different sets of parameter values. Specifically, in figures 7(I)–(L) we use default parameter values (table 4), and in figures 7(M)–(P) the parameters are changed as $k_y = 20$, $k_z = 0.001$ while keeping other parameters unchanged. We found that the transdifferentiation process can go through stem cell state (figure 7(I)), going through stem cell state and intermediate state (figure 7(J)), or going through both}

---

**Figure 7.** Typical kinetic transition paths are projected on the landscape background for differentiation (A)–(D), reprogramming (E)–(H), and transdifferentiation (I)–(P), individually. Each column represents a typical transition path for differentiation (A)–(D), reprogramming (E)–(H), and transdifferentiation (I)–(P), individually. Magenta, white, and red lines correspond to three stages (early, middle and late stages) for the transitions in temporal order. These kinetic paths indicate that transdifferentiation can go through a mixed (MX), unspecific intermediate (UI) or progenitor-like (PG) state. The potential landscape is defined as $U = -\log P_x$, and $P_x$ is steady state probability distribution. On the landscape the blue region represents lower potential or higher probability, and the red region represents higher potential or lower probability. For (M)–(P), the parameters are changed as $k_x = 20$, $k_y = 0.001$ while keeping other parameters the same.
intermediate states (figure 7(K)). Interestingly, we found that the transdifferentiation can also go through a mixed intermediate state (figures 7(N)–(P)), where both lineage marker genes are expressed (high A and high B). These results agree well with recent studies showing that cell could pass a mixed (MX), unspecific intermediate (UI) or progenitor-like (PG) transition states. For example, single cell RNA-seq analysis showed that fibroblast undergoes transdifferentiation to neurons, which passes a progenitor-like state [44]. Additionally, experiments also suggest that transdifferentiation can also go through a ‘mixed’ state where both differentiation state marker gene are expressed. For example, single-cell transcriptome analysis has suggested that the Transdifferentiation of mouse cardiac fibroblasts to induced cardiomyocytes (iCMs) can go through several ‘mixed’ states where both cardiomyocyte and fibroblast-specific markers are expressed [45]. We also calculated the possibility (percentage) for the occurrence of each of these transition path and the distribution of average transition time of each transition path for differentiation, reprogramming and transdifferentiation, respectively (figure S3, see text S1 for details).

To summarize our transition path results, we provide an illustration figure for the comparisons between experimental observation and our simulations results (figure 8). Figure 8(A) gives an illustration for the experimental observations based on classic Waddington landscape [33]. Specifically, experiments suggest that transdifferentiation can go through a mixed (MX), unspecific intermediate (UI) or progenitor-like (PG) state. Figure 8(B) gives a summary for our major results of the transition paths for differentiation and transdifferentiation based on the epigenetic model of development (figure 7). Our simulation results agree well with the experimental observations in that three possible routes for transdifferentiation suggested by experiments all can be identified from our landscape model (see the red paths in figures 8(A) and (B)).

The adiabatic approximation (much faster protein-DNA binding/unbinding rate compared with epigenetic regulations) has been used in modeling developmental process [23]. However, the roles of non-adiabaticity in developmental process have also been emphasized [7, 9]. To see how the adiabaticity affects our results, we calculated landscape for different $\alpha_1$ (unbinding rate for self-activation) and $\beta_1$ (unbinding rate for mutual repression) while keeping the equilibrium constant ($k_\alpha$ and $k_\beta$) for the binding/unbinding reactions the same (figure 9). Figure 9 shows that the unbinding rate $\beta_1$ does not affect the landscape results significantly. For example, along the vertical direction from down to top, $\beta_1$ increases from $10^{-3}$ to $10^{-5}$, but the landscape does not change a lot. When we look at the horizontal direction from left to right ($\alpha_1$ increases from $10^{-5}$ to $10^{-3}$), the landscape does change, and the number of basins (stable states) decreases and intermediate states are inclined to disappear. So, the time scale for the binding/unbinding rate of self-activation, i.e. the degree of adiabaticity, does influence the landscape or the number of cell states. However, as shown in figure 8, we believe that the slower binding/unbinding rate for self-activation (e.g. $\alpha_1 = 10^{-5}$, corresponding to figures 9(A), (D) and (G)) provides a better fit to experimental data because it can generate the intermediate states as suggested in experiments (figure 8).

Therefore, although the adiabatic approximation is important in modeling development [23], we believe that it is possible that cell can use different mechanisms for multiple fate transitions, including using the non-adiabatic regime to generate intermediate states and increase plasticity. It warrants further explorations about
whether models using adiabatic approximation can also generate multiple intermediate states, consistent with experimental observations (figure 8(A)).

**Interplay between genetic and epigenetic regulations governs cell fate decisions in development**

To study the influence of gene regulations on the stem cell differentiation, we quantified the landscape at different regulatory strengths. Specifically, we changed two critical parameters related to regulatory strengths to see how the landscape is affected (figure 10). Here, $k_{\alpha}$ represents the equilibrium constant for self-activation and $k_{\beta}$ represents the equilibrium constant for mutual repression. So, $k_{\alpha}$ and $k_{\beta}$ can quantify the activation and repression strength, respectively. We found that when the repression strength decreases, the stem cell state becomes less stable (figure 10, vertical direction from up to down), indicated by the landscape change where stem cell attractor becomes more and more shallow and finally disappears. Also, as the repression strength declines, more attractor states occur on the landscape. When the activation strength increases, more intermediate states emerge on the landscape (figure 10, horizontal direction from left to right). This suggests that the decrease of repression strength or increase of activation strength might promote the occurrence of heterogeneity in the populations of differentiated cells. Therefore, our results indicate that the change in gene regulations can lead to the differentiation or reprogramming. In current case, the decrease of repression strength provides a possible way to induce the cellular differentiation. On the contrary, the increase of repression strength might provide a way to induce reprogramming.

Here, we can use two different ways to explain the landscape results. In the first way, we have a fixed landscape (with fixed parameters), and different cell states (e.g. pluripotent stem cell state and differentiated state) can make transitions among each other (figure 6). In the second way, the landscape changes as parameter values change (figure 10). Here the point is that cell fate transitions in biological systems are driven by signals
corresponding to parameters in our model), and different parameter values will generate different landscape shapes, which determine the stability of cell state attractors. Take differentiation as an example, if we look at the vertical direction from top to down for the first column in figure 10, we will see that as repression strength decreases, the landscape shape changes and differentiated states become more stable. That is, the decrease of repression strength might promote differentiation. As a result, when the repression strength is very small, differentiated states are very stable over time because the two differentiated states are much more stable than other states (figure 10(D)). In summary, it is the landscape changing with signals, rather than fixed landscape, that can explain the distinct stability of cell types driven by signals.

We also calculated the mean first passage time (MFPT) for differentiation, reprogramming, and transdifferentiation (figure 11). MFPT reflects the average transition time of the system switching from one basin to another basin on the landscape, and therefore can quantify the feasibility of the cells switching from one cell type to another one (for example, from stem cell to differentiated cell in development and from differentiated cell to stem cell in reprogramming). We estimated the MFPT based on the simulation trajectories (see text S1 for details) [9]. As \( k_β \) increases, the MFPT for differentiation becomes longer while the MFPT for reprogramming becomes shorter (figure 11(B), corresponding to third column in figure 10, \( k_α \) is fixed as 0.2). This is consistent with the landscape results showing that larger \( k_β \) leads to a more stable stem cell state (figure 10). As \( k_α \) increases, the MFPT for differentiation becomes a little bit shorter but the MFPT for reprogramming does not change too much (figure 11(A), corresponding to second row in figure 10, \( k_β \) is fixed as 0.2). So, the change of \( k_α \) does not change the landscape significantly. Similarly, the MFPT for transdifferentiation becomes shorter as \( k_α \) increases but longer as \( k_β \) increases (figures 11(C) and (D)). This shows that larger activation strength \( k_α \) and smaller repression strength \( k_β \) (right bottom figure in figure 10)
promotes transdifferentiation, where the two differentiation states are very stable compared to other states. Of note, the relative time scales for differentiation, reprogramming, and transdifferentiation depend on the landscape shape (here depending on the parameter \( k_\alpha \) and \( k_\beta \)). So, we also calculated the MFPT results for different parameter regions of \( k_\alpha \) and \( k_\beta \) (see figure S4).

To further study the effects of epigenetic regulation on the cell fate decision process in stem cell development, we changed the epigenetic strength (represented by the parameter \( pdem \)), to see the landscape change. We found that in the case without epigenetic regulations (figure 12, increasing \( pdem \)), the number of attractors on the landscape becomes less (compare figures 10 and 12). This agrees well with our conclusion that epigenetic regulations promote the occurrence of intermediate states. Additionally, in the case without epigenetic regulations, we also obtain the consistent results showing that the decrease of repression strength might induce the cellular differentiation.

**Discussion**

Waddington epigenetic landscape, as a famous metaphor, has been used to explain cellular differentiation. However, it remains challenging to quantify the epigenetic landscape. For example, a critical issue is how to integrate genetic and epigenetic factors to study the mechanisms of development in a holistic perspective. Previous studies have been devoted to quantifying the landscape for development. However, the epigenetic modifications (such as histone modifications) have not been investigated explicitly in previous landscape models [6, 8, 12]. In this work, by combining genetic and epigenetic regulations, we constructed a developmental epigenetic model and quantified the epigenetic landscape for cellular development. We identified some attractors on the landscape characterizing different cell states, including pluripotent stem cell state, differentiated states, and some intermediate states. Our results show that epigenetic regulations can promote
multistability or intermediate states. This suggests that epigenetic regulations might play critical roles in stem cell development, which can be used to intervene the reprogramming.

We also quantified the kinetic transition paths for differentiation, reprogramming and transdifferentiation. Intriguingly, the kinetic transition paths for transdifferentiation from our model predictions agree well with experimental observations, showing that transdifferentiation can go through a mixed ($MX$), unspecific intermediate ($UI$) or progenitor-like ($PG$) state. Therefore, our modeling results based on a simplified developmental epigenetic model recapitulate experimental observations and provide insights into the underlying mechanisms of cell fate decisions in development.

To study the roles of gene regulations on the differentiation or reprogramming, we changed the gene regulatory strengths to see how the landscape is influenced. We found that the decrease of repression strength makes stem cell state less stable. Therefore, decreasing repression strengths provides a way to promote differentiation, and increasing repression strengths provides a way to induce reprogramming. We also calculated the MFPT, which provides another measure for the feasibility of cell STs. The results of MFPT give the consistent results with the landscape results. From calculating MFPT, we found that the increase of activation strength or decrease of repression strength promotes transdifferentiation. Additionally, by changing the epigenetic strength, we found that the number of attractors decreases as the epigenetic strength declines. This is consistent with our findings that epigenetic modifications promote the occurrence of intermediate states. In our models, the rates of histone modifications are much lower than the degradation or synthesis rate of protein (tables 3 and 4). This corresponds to the non-adiabatic regulations, where the effective binding/unbinding
process can be comparable or even slower than the synthesis and degradation of proteins. Therefore, our results in this work are reminiscent of previous studies showing that non-adiabatic effects can facilitate the occurrence of more cell types [9]. Here, the non-adiabaticity comes from the slow epigenetic dynamics (histone modifications).

In this work, we considered the H3K27 (repressive) as an example for histone modifications in gene regulations. It would be straightforward to extend our models to integrate other forms of histone modification, e.g., the lysine 4 (active) on histone H3 (H3K4). Future work can incorporate other forms of histone modifications and/or DNA methylation to study more intricate mechanisms for epigenetic regulations on gene expressions in development. In our simplified developmental model, we use a two-gene network to simulate stem cell differentiation and the SC state and PG state are treated as one single state. However, in a more realistic developmental model, which usually involves more genes, the SC and PG state could be different, i.e. there could be many different SC/PG like states appearing in development and differentiation. It would be interesting to integrate the epigenetic control to more realistic developmental gene network models [8], and examine these epigenetic cell states. That will reveal more intricate epigenetic control mechanisms in development. Furthermore, current model is in the level of single cell, and it is desirable to develop multiscale approaches formulating different models at different levels (genetic, epigenetic, cellular etc). Also, it is of great importance to integrate the genetic and epigenetic data [46] into the stem cell development model. In our opinion this is a very challenging but crucial task in the fields of systems biology.

The epigenetic mechanisms for cell fate decisions also have indications for cancer. Although cancer is traditionally considered as a genetic disease, epigenetic plasticity and alterations are suggested to contribute to hallmarks of cancer [47]. Our modeling framework can be applied to construct the epigenetic model for cancer and study the epigenetic control mechanisms in tumorigenesis. In summary, our developmental epigenetic model provides new insights into the interplay between genetic and epigenetic factors on controlling lineage commitment in development. Our approach provides a general way to study the epigenetic cell fate decisions in biological systems, by integrating transcriptional, translational and epigenetic players.

Acknowledgments

CL is supported by the National Key R&D Program of China (2018YFC0910500), National Natural Science Foundation of China (11771098), Natural Science Foundation of Shanghai, China (17ZR1444500), and partially supported by the key project of Shanghai Science & Technology (No. 16JC1420402). WL is supported partly by the National Science Foundation of China (NSFC) (Grant Nos. 11322111 and 61773125), by the National Key R&D Program of China (Grant No. 2018YFC0116600), by the Science and Technology Commission of Shanghai Municipality Grant (Grant No. 18DZ1201000), and by the State Key Laboratory of Medical Neurobiology and MOE Frontiers Center for Brain Science of Fudan University.

Author contributions

CL conceived and designed the research. YZ and CL performed the research. YZ, NL, WL and CL analyzed the results, and wrote the manuscript.

ORCID iDs

Wei Lin https://orcid.org/0000-0002-1863-4306
Chunhe Li https://orcid.org/0000-0002-9127-3930

References

[1] Waddington C H 1957 The Strategy of the Genes: A Discussion of some Aspects of Theoretical Biology (London: Allen and Unwin) p 290
[2] Takahashi K and Yamanaka S 2006 Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors Cell 126 663–76
[3] Zhou Q and Melton D 2008 Extreme makeover: converting one cell into another Cell Stem Cell 3 382–8
[4] Saha K and Jaenisch R 2009 Technical challenges in using human induced pluripotent stem cells to model disease Cell Stem Cell 5 384–95
[5] Graf T and Enver T 2009 Forcing cells to change lineages Nature 462 587–94
[6] Wang J, Zhang K, Xu L and Wang E K 2011 Quantifying the waddington landscape and biological paths for development and differentiation Proc. Natl Acad. Sci. USA 108 8257–62
[7] Zhang B and Wolynes P G 2014 Stem cell differentiation as a many-body problem Proc. Natl Acad. Sci. USA 111 10185–90
[8] Li C and Wang J 2013 Quantifying cell fate decisions for differentiation and reprogramming of a human stem cell network: landscape and biological paths PLoS Comput. Biol. 9 e1003165
[9] Li C and Wang J 2013 Quantifying waddington landscapes and paths of non-adiabatic cell fate decisions for differentiation, reprogramming and transdifferentiation J. R. Soc. Interface 10 20130787

[10] Margaret J T, Chu B K, Gallivan C P and Read E L 2018 Rare-event sampling of epigenetic landscapes and phenotype transitions PLoS Comput. Biol. 14 e1006336

[11] Li C and Balazsi G 2018 A landscape view on the interplay between emt and cancer metastasis NPJ Syst. Biol. Appl. 4 34

[12] Brackston R, Lakatos E and Stumpf M P 2018 Transition state characteristics during cell differentiation PLoS Comput. Biol. 14 e1006405

[13] Zhou P and Li T 2016 Construction of the landscape for multi-stable systems: potential landscape, quasi-potential, a-type integral and beyond J. Chem. Phys. 144 094109

[14] Sasi M, Kawabata Y, Makishi K, Itoh K and Terada T P 2013 Time scales in epigenetic dynamics and phenotypic heterogeneity of embryonic stem cells PLoS Comput. Biol. 9 e1003380

[15] Moazed D 2011 Mechanisms for the inheritance of chromatin states Cell 146 510–8

[16] Furey T S and Sethupathy P 2013 Genetics driving epigenetics Science 342 705–6

[17] Angel A, Song J, Dean C and Howard M 2011 A polycistron-based switch underlying quantitative epigenetic memory Nature 476 105

[18] Dodd I B, Michaelisen M A, Steppen K and Thon G 2007 Theoretical analysis of epigenetic cell memory by nucleosome modification Cell 129 813–22

[19] Zhang H, Tian X J, Mukhopadhyay A, Kim K and Xing J 2014 Statistical mechanics model for the dynamics of collective epigenetic histone modification Phys. Rev. Lett. 112 068101

[20] Ashwin S and Sasi M 2015 Effects of collective histone state dynamics on epigenetic landscape and kinetics of cell reprogramming Sci. Rep. 5 16746

[21] Berry S, Dean C and Howard M 2017 Slow chromatin dynamics allow polycistron target genes to filter fluctuations in transcription factor activity Cell Syst. 4 444–57

[22] Cortini R et al 2016 The physics of epigenetics Rev. Mod. Phys. 88 025002

[23] Folguera-Blasco N, Pérez-Carrasco R, Cuyàs E, Menéndez J A and Alarcón T 2018 Epigenetic regulation of cell fate reprogramming in aging and disease: a predictive computational model PLoS Comput. Biol. 14 e1006052

[24] Bintu L et al 2016 Dynamics of epigenetic regulation at the single-cell level Science 351 720–4

[25] Huang R and Lei J 2019 Cell-type switches induced by stochastic histone modification inheritance Discrete Continuous Dyn. Syst. B 24 5601–19

[26] Hansen K H et al 2008 A model for transmission of the H3K27me3 epigenetic mark Nat. Cell Biol. 10 1291

[27] Kuzmichev A, Nishioka K, Erdjument-Bromage H, Tempst P and Reinberg D 2002 Histone methyltransferase activity associated with a human multiprotein complex containing the enhancer of zeste protein Genes Dev. 16 2893–905

[28] Riising E M et al 2014 Gene silencing triggers polycistron repressive complex 2 recruitment to cpg islands genome wide Mol. Cell 55 347–60

[29] Hemberger M, Dean W and Reik W 2009 Epigenetic dynamics of stem cells and cell lineage commitment: digging waddington’s canal Nat. Rev. Mol. Cell Biol. 10 526

[30] Rugg-Gunn P J, Cox B J, Ralston A and Rossant J 2010 Distinct histone modifications in stem cell lines and tissue lineages from the early mouse embryo Proc. Natl Acad. Sci. 107 10783–90

[31] Yang X et al 2018 Silencing of developmental genes by H3K27me3 and dna methylation reflects the discordant plasticity of embryonic and extraembryonic lineages Cell Res. 28 593

[32] Reid A and Tursun B 2018 Transdifferentiation: do transition states lie on the path of development? Curr. Opin. Syst. Biol. 11 18–23

[33] Margueron R et al 2009 Role of the polycistron protein eed in the propagation of repressive histone marks Nature 461 762

[34] Gillespie D T 1977 Exact stochastic simulation of coupled chemical reactions J. Phys. Chem. 81 2340–61

[35] Besciari A, Séraphin B and Serrano L 2001 Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion EMBO J. 20 5256–35

[36] Feng H, Han B and Wang J 2011 Adiabatic and non-adiabatic non-equilibrium stochastic dynamics of single regulating genes J. Phys. Chem. B 115 1254–61

[37] Huang S, Guo Y, May G and Enver T 2007 Exploring intermediate cell states through the lens of single cells Curr. Opin. Syst. Biol. 9 32–41

[38] ikawa T et al 2010 An essential developmental checkpoint for production of the T cell lineage Science 329 93–6

[39] Ye Y, Kang X, Bailey J, Li C and Hong T 2019 An enriched network motif family regulates multistep cell fate transitions with restricted reversibility PLoS Comput. Biol. 15 e1006855

[40] Plath K and Lowry E W 2011 Progress in understanding reprogramming to the induced pluripotent state Nat. Rev. Genet. 12 253

[41] Wang P et al 2014 Epigenetic state network approach for describing cell phenotypic transitions Interface Focus 4 20130068

[42] Treutlein B et al 2016 Dissecting direct reprogramming from fibroblast to neuron using single-cell rna-seq Nature 534 391

[43] Liu Z et al 2017 Single-cell transcriptomics reconstructs fate conversion from fibroblast to cardiomyocyte Nature 551 100

[44] Dureen Z, Chen X, Jiang R, Wang Y and Wong W H 2017 Modeling gene regulation from paired expression and chromatin accessibility data Proc. Natl Acad. Sci. 114 E491–23

[45] Flavahan W A, Gaskell E and Bernstein B E 2017 Epigenetic plasticity and the hallmarks of cancer Science 357 eaal2380