Landscape inferred from gene expression data governs pluripotency in embryonic stem cells

Xin Kang [a,b], Chunhe Li [b,c,*]

[a] School of Mathematical Sciences, Fudan University, Shanghai, China
[b] Shanghai Center for Mathematical Sciences, Fudan University, Shanghai, China
[c] Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China

A B S T R A C T

Embryonic stem cells (ESCs) can differentiate into diverse cell types and have the ability of self-renewal. Therefore, the study of cell fate decisions on embryonic stem cells has far-reaching significance for regenerative medicine and other biomedical fields. Mathematical models have been used to study emryonic stem cell differentiation. However, the underlying mechanisms of cell differentiation and lineage reprogramming remain to be elucidated. Especially, how to integrate the computational models with quantitative experimental data is still challenging. In this work, we developed a data-constrained modelling approach, and established a model of mouse embryonic stem cells. We used the truncated moment equations (TME) method to quantify the potential landscape of the ESC network. We identified two attractors on the landscape, which represent the embryonic stem cell (ESC) state and differentiated cell (DC) state, respectively, and quantified high dimensional biological paths for differentiation and reprogramming process. Through identifying the optimal combinations of gene targets based on a landscape control strategy, we offered some predictions about the key regulatory factors that govern the differentiation and reprogramming in ESCs.

© 2020 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Embryonic stem cells have the capacity of self-renewal, high proliferation and multilineage differentiation. Therefore, they have an important and far-reaching influence on new drug discovery and development, disease models, organ transplantation and tissue repair [1,2]. Studies have shown that the reprogramming of human somatic cells and lineage reprogramming can be achieved by regulating several key genes [3–8]. Some key genes that determine whether embryonic stem cells exit or remain in a pluripotent state have been extensively studied [9–14]. However, the underlying mechanisms of cell differentiation and lineage reprogramming have not been fully clarified. It remains challenging to control whether cells exit or remain in the pluripotent state, which needs further explorations.

The pluripotency of mouse embryonic stem cells has been considered to be controlled by underlying gene regulatory networks [15–19]. Dunn et al. developed a data-constrained, computational approach to construct a simplified gene regulation network of embryonic stem cells, which comprises 12 gene components and three input signals. This model can explain the known phenotypes, predict some results of combined genetic perturbations and test the relative stability of embryonic stem cells under different culture conditions [20]. Recently, a simulation approach has been proposed to investigate the stochastic dynamics of pluripotency in mouse embryonic stem cells based on the network of Dunn et al. [21]. However, the global stability of pluripotency network of mouse stem cells has yet to be explored. More importantly, how to identify the optimal combinations of interventions from gene networks to induce reprogramming remains challenging.

Here, we aim to apply the landscape theory to study the dynamical mechanisms of the underlying gene regulatory network of embryonic stem cells. The classic Waddington landscape has been proposed as a metaphor for the development and differentiation of cells [22]. Recently, the epigenetic landscapes for biological networks have been quantified from various approaches [23–29], and employed to study the stochastic dynamics of embryonic development [23,30–34]. From the perspective of landscape, different cell types are depicted as attractors on a potential surface. The cell differentiation process is viewed as a ball rolling from one

* Corresponding author at: Shanghai Center for Mathematical Sciences, Fudan University, Shanghai, China.
  E-mail address: chunheli@fudan.edu.cn (C. Li).

basin to another on the landscape surface by crossing certain barriers. The barrier heights among the attractors quantify the feasibility of the cell transformation from one cell type to the other.

In this work, we try to identify the important factors affecting the differentiation and reprogramming of mouse embryonic stem cells from the perspective of dynamics. Based on an embryonic stem cell network [20], we built models of ordinary differential equations (ODEs) to describe the time evolution of expression levels of different components. One critical issue in constructing a gene regulation network model is how to determine the parameters of the system. Here, we adopted a data constrained modelling strategy, and estimated the parameters in the models to fit the binary experimental data from Dunn et al. [20]. Then we obtained the potential landscape of the network by truncated moment equations (TME) method [35,36]. We identified two attractors on the landscape, which represent the embryonic stem cell (ESC) state and the differentiated cell (DC) state, respectively. We further calculated the minimum action paths (MAPs) to quantify the transition processes of the differentiation and reprogramming. By single factor sensitivity analysis of parameters, we predicted some critical regulations in cell differentiation and reprogramming. From the perspective of gene networks, an effective intervention strategy should be targeting multiple factors in the network, rather than one. By optimizing the transition action from DC state to ESC state, we identified some optimal combinations of key regulations among genes, which can promote reprogramming. These analyses on the landscape and the transition actions will promote our understanding on the mechanisms of differentiation and reprogramming, and help to identify key genes and interactions for inducing reprogramming.

2. Results

2.1. Mathematical model of embryonic stem cell network

Our model is based on the embryonic stem cell network inferred by Dunn et al. [20]. This network includes 26 interactions (regulations) and 15 components (12 genes and 3 signals). The three extra-cellular signals are cytokine leukemia inhibitory factor (LIF) and two selective inhibitors, including glycogen synthase kinase 3 (CH) and mitogen-activated protein kinase (PD) [37].

Some direct transcriptional targets of LIF, CH, and PD have been identified, including Stat3, Gbx2, Klf4, Tcf3, Tfcp2l1, Esrrb, MEKERK and Nanog [38–43] (see Supplemental information Fig. S1 for direct transcriptional targets). The other regulations among genes are predicted from experimental data of mouse embryonic stem cells [20]. Based on this network, we constructed models of ODEs to describe the time evolution of expression levels of each component, which can be written as the following form:

$$\frac{dX_i}{dt} = g_0 + \sum_{j=1}^{N} \frac{A_{ji} \times X_j^n}{S^n + X_j^n} + \sum_{j=1}^{N} \frac{B_{ji} \times S^n}{S^n + X_j^n} - k \times X_i, \qquad (1)$$

where, $X_i, i = 1, \cdots, N$ represents the expression level of gene i. g_0 and k represent the basal production and degradation rates of x_i, respectively.

We used Hill functions to describe the regulations (activations and inhibitions) in the network (links in Fig. 1). The Hill function is a sigmoidal function. It has been commonly used to describe gene regulations in gene network models [44–46]. S in Eq. (1) represents the threshold of a sigmoidal function. n is the Hill coefficient, which determines the steepness of the sigmoidal function. A_{ji} and B_{ji} are the scale factors for the activation and inhibition, respectively. When the scale factor A_{ji} is equal to zero, it means that there is no activation from gene j to gene i. When the scale factor B_{ji} is equal to zero, it means that there is no inhibition from gene j to gene i.

2.2. Estimate parameters of the model from experimental data

We built the models of ODEs based on the embryonic stem cell network. Then, we used experimental data of Dunn et al. to estimate the parameters in our model [20]. The experimental data are binary gene expression data of 12 genes from PCR under five various combinations of external signals (LIF, CH, PD) (Fig. 2C). In the work of Dunn et al., there are totally five different combinations of signals, including 2i + LIF, 2i, LIF + CH, LIF + PD, and no signals. Therefore, we used gene expression data from these five various combinations of external signals to estimate the parameters in our model. In order to reduce the dimension of parameters space, we assumed that all the nonzero A_{ji} are equal to undetermined constant a and nonzero B_{ji} are equal to undetermined
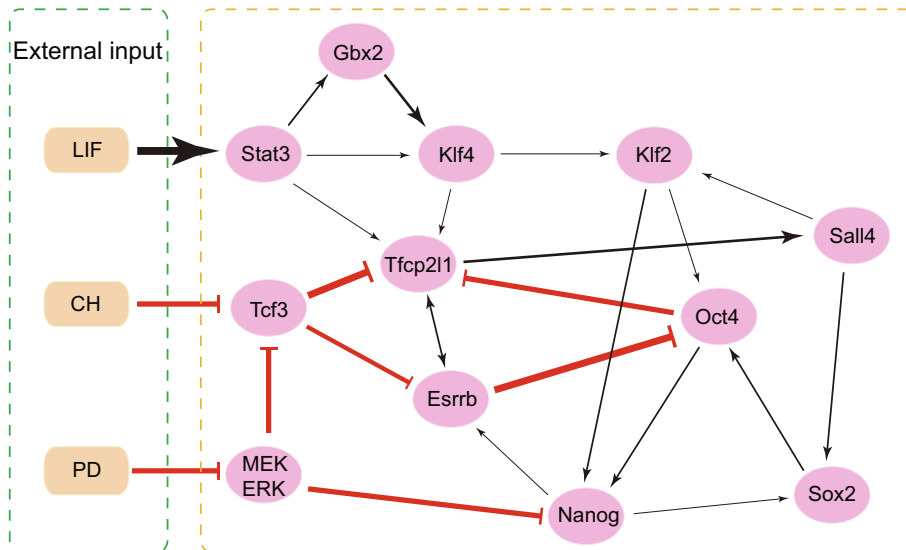
Fig. 1. The network of the core circuit of an embryonic stem cell system including 15 nodes and 26 interaction links (18 activations and 8 inhibitions). Black arrows represent activation and red bars represent inhibition. Orange nodes represent input signals, and pink nodes represent genes. The width of the links represents the sensitivity of the links calculated from the optimization of transition actions (Table 1). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
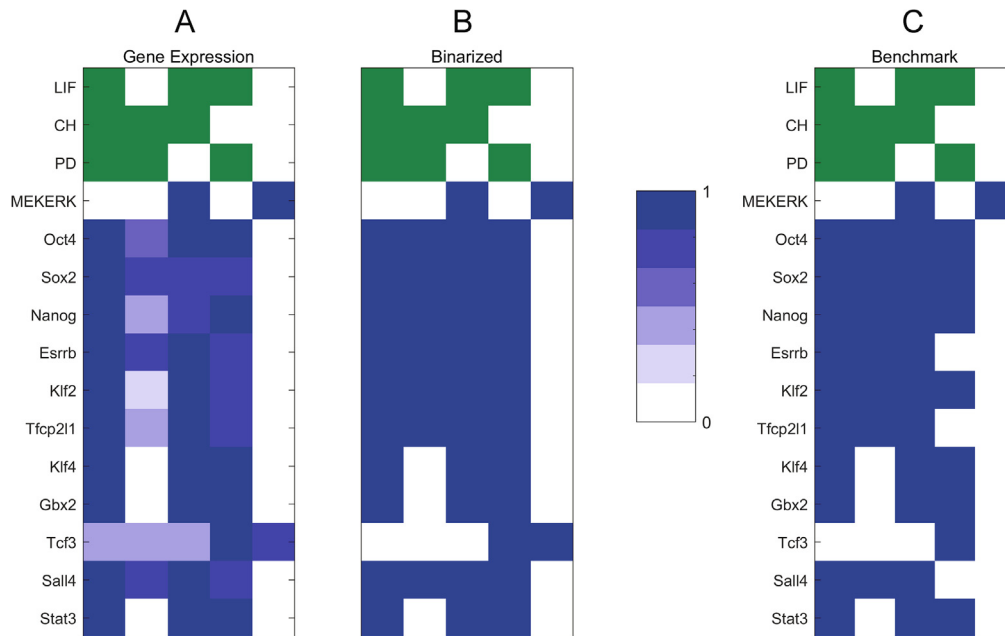
Fig. 2. Comparison of gene expression profiles between modeling and experiments. (A) The gene expression profiles from the model using the picked parameter set. (B) The binary expression profiles corresponding to (A). (C) The gene expression profiles from experiments [20]. Gene expression is discretized to high (blue) or low (white). The input signal is discretized to high (green) or low (white). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

constant b. Moreover, LIF, CH and PD are regarded as signals of the network. Therefore, the unknown parameters in our model include a, b, S, k, n and g_0. Next, we aimed to estimate these parameters from the binary experimental results of Dunn et al. [20]. Since the experimental data are dimensionless, we specified the parameter range based on previous works [23,45]. Here, the Hill coefficient n determines the steepness of the sigmoidal function, i.e. the degree of nonlinearity (the cooperativity for gene regulations). Following previous works [30,46], we assumed $n = 3$ or 4 to represent certain degree of cooperativity for gene regulations. We set the basal production rate $g_0 = 0.01$ and degradation rate $k = 1$, and the ranges of other parameters are $a \in (0, 4), b \in (0, 4), S \in (0, 4)$ and $n \in \{3, 4\}$.

Then, we followed three steps to estimate the parameters of the model:

1. First, we reduced the searching range of parameters based on the biological significance of stable states. A major biological constraint is multistability, because in an embryonic developmental system there should be at least two cell states, i.e. ESC state and DC state. In our model, it means that there will be multiple stable states when $t \rightarrow \infty$ for different initial conditions. Therefore, we performed a search in the parameter space: $a \in (0, 4), b \in (0, 4), S \in (0, 4), n \in \{3, 4\}$ and LIF, CH, PD $\in \{3, 4, 5, 6, 7, 8\}$. For each parameter set, we numerically solved the ODEs under different initial values to acquire stable fixed points of the ODEs, i.e. stable steady states. We recorded the parameter sets that generate bistable states and corresponding stable states. Then, we binarized the stable states by the mean value of stable states, i.e., if the expression level of a gene is larger than the mean value, it is denoted as 1 and if the expression level of a gene is less than the mean value, it is denoted as 0. Further, we picked the parameter sets that can generate biologically significant bistable states (Fig. 3), i.e., one stable state should represent pluripotent ESC state, which has higher expression level (with the expression level of 1 after binarization) of OCT4, Nanog, Sox2, Esrrb and Tfcp2l1 (ESC markers), and the other stable state should represent DC

state, which has lower expression level (with the expression level of 0 after binarization) of OCT4, Nanog, Sox2, Esrrb and Tfcp2l1.

2. For each parameter set including eight parameters: $(a, b, S, n, k, \mathrm{LIF}, \mathrm{CH}, \mathrm{PD})$ obtained in the previous step, we performed simulations under five experimental conditions to find parameter sets that mostly fit the five groups of experimental data. Since embryonic stem cells can be stably cultured as a substantially homogeneous population in different environments [20], if the model has multiple stable states under a certain experimental condition, we choose the steady state with high probability as the expression level under this signal. Here our fitting procedure is based on deterministic ODEs. To consider the heterogeneity of populations, future work can develop methods to fit models based on the distribution of expression data among populations. In this way, we can obtain a 12×5 (expression levels of 12 different gene variables in five different conditions) real-valued matrix. Then we took the average of gene expression levels under different signals as a threshold to acquire a binary matrix.

3. To find parameters that mostly fit the experimental gene expression profile (benchmark data) (Fig. 2C), we calculated the Hamming distance (the discrepancy between two binary matrices) between the experimental gene expression profile and discretized gene expression profile from the model (step 2), and sorted them by hamming distance to find the most suitable parameter set. Fig. 2A shows the gene expression profiles generated from the selected parameter set, and Fig. 2B shows the binary expression profile corresponding to Fig. 2A. Fig. 3 shows the points in parameter space which can generate ESC state and DC state.

2.3. Potential landscape and the minimum action path for the embryonic stem cell model

Lately, several studies have employed the landscape approach to investigate the stochastic dynamics of biological systems [23,25,44]. Here we used the TME method (see Methods) to
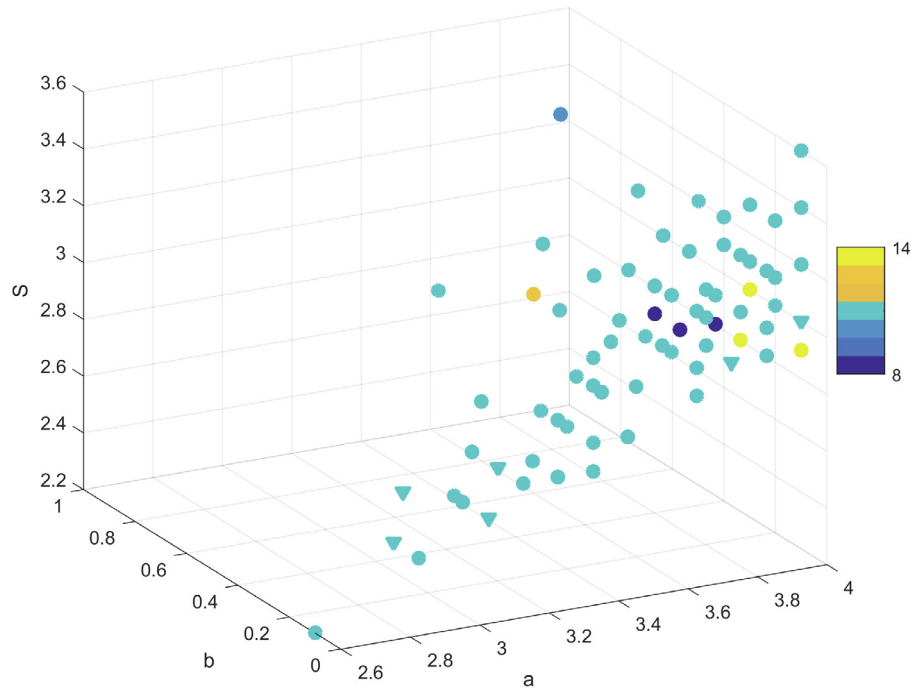
Fig. 3. The points in parameter space which can generate two stable states (ESC state and DC state), from parameter searching. The color of the points denotes the Hamming distance between the benchmark data and 12 × 5 real matrix generated from the model using corresponding parameter set. Blue points represent smaller Hamming distance and yellow points represent larger Hamming distance. The triangular points represent the parameters used for optimization. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

calculate the potential landscape of the embryonic stem cell model. Since it is hard to visualize the landscape in 12-dimensional space, we selected two gene variables as coordinates for landscape and projected the 12-dimensional landscape into two dimensions. Here, we chose two marker genes of embryonic stem cells, OCT4 and Nanog, as the coordinates of landscape. Oct4 is a mammalian POU transcription factor expressed by early embryo cells and germ cells [47]. Nanog's overexpression enables self-renewal of embryonic stem cells, and Nanog is accordingly considered as a core element of the pluripotent transcriptional network [48]. Of note, our main conclusions do not depend on the specific choice of the coordinates because the transition actions among different attractors are based on the 12-dimensional space (see Supplemental information Fig. S2 for landscapes using other pairs of variables as coordinates). In this way, two stable states emerge on the landscape for the embryonic stem cell system (Fig. 4). Here, the ESC state has a high OCT4, Nanog, Sox2, Esrrb and Tfcp2l1 expression level, while the DC state has a low OCT4, Nanog, Sox2, Esrrb and Tfcp2l1 expression level.

To study the transition processes between ESC state and DC state, we calculated the kinetic transition paths between two cell states by minimizing the transition actions, which are called MAPs (see Methods for how to acquire the MAPs). Fig. 4 shows the landscape and MAPs shown in Oct4 and Nanog coordinates. The white line represents the transition path from ESC state to DC state (differentiation), while the magenta line represents the transition path from DC state to ESC state (reprogramming). The transition paths for cell differentiation process and cell reprogramming process are not identical, which is reflected by the disparity between the forward and backward kinetic transition paths. The kinetic paths of the system deviating from the conventionally expected potential gradient paths is caused by the non-gradient force, i.e. curl flux [26].

### 2.4. Single factor sensitivity analysis on model parameters

To further study the effect of interactions among the genes on the dynamics of the embryonic stem cell model, we performed single factor sensitivity analysis on the 26 parameters (18 activation constants and 8 inhibition constants) of embryonic stem cell model. Specifically, we increased or decreased each parameter by 10% to modulate the regulation strengths among genes, and calculated how the transition actions between ESC state and DC state change after these perturbations. In this way, we can identify some critical regulations that govern the transition between DC state and ESC state. From the results of sensitivity analysis (Fig. 5), we conclude that the increase of all activation constants and inhibition constants (the increase of activation strength or the decrease of inhibition strength) will increase the transition actions from ESC state to DC state, and make ESC state more stable. These key regulations are sorted according to their sensitivities (defined as the difference between $\triangle S_{ESC \to DC}$ and $\triangle S_{DC \to ESC}$). The top nine sensitive regulations are shown in Table S1. The sensitivity analysis predicts some key links including LIF → Stat3, Tfcp2l1 → Sall4 and Esrrb → Tfcp2l1. Here → represents activation regulation. The increase of these link strengths can critically enhance the stability of ESC state.

2.5. Cellular reprogramming-promoting interventions

To predict the combination of target regulations among genes which can maintain the totipotency of embryonic stem cells, we used an approach of optimization to predict the critical combinations of parameters [44,49]. The optimization goal is to make ESC state more stable and DC state less stable by tuning the 26 parameters (18 activation constants and 8 inhibition constants) while keeping the number of steady states unchanged (see Methods for details).
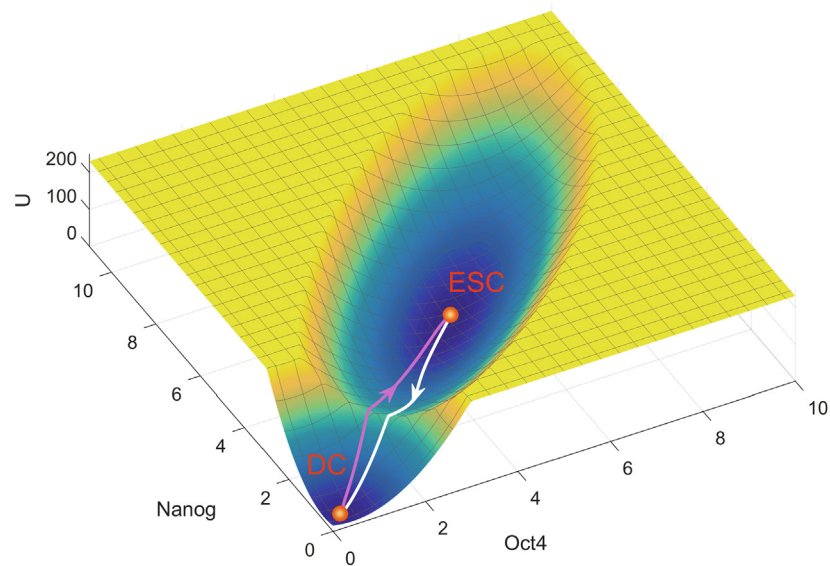
Fig. 4. The landscape and paths for the ESC model shown in Oct4 and Nanog coordinates. The blue region represents higher probability or lower potential and the yellow region indicates lower probability or higher potential. The white line represents the transition path from ESC state to DC state, and the magenta line represents the transition path from DC state to ESC state. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
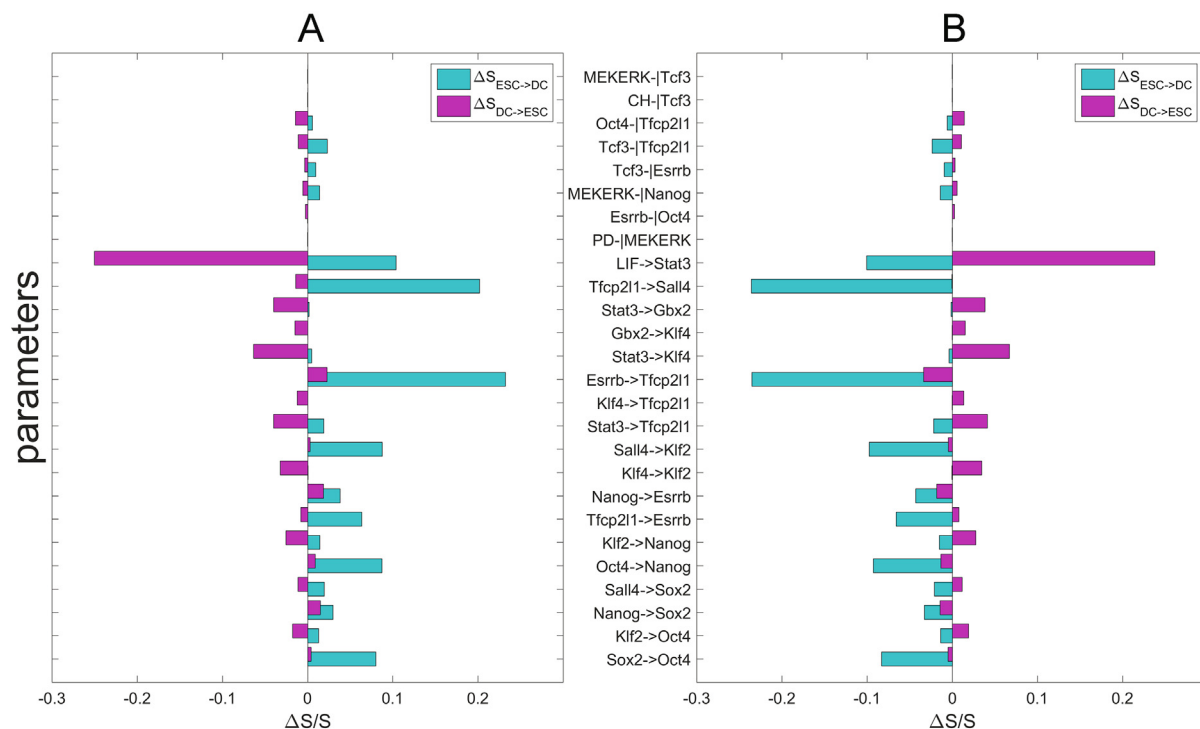

Fig. 5. Sensitivity analysis for the 26 parameters (regulatory strengths, including 18 activation constants and 8 inhibition constants) on the transition actions. Y axis represents the 26 parameters. X axis represents the percentage of the change of the transition action (S) relative to S with default parameters. Here, $S_{ESC->DC}$ represents the transition action from attractor ESC to attractor DC (cyan bars), and $S_{DC->ESC}$ represents the transition action from attractor DC to attractor ESC (magenta bars). (A) Each parameter is increased by 10%, individually. (B) Each parameter is decreased by 10%, individually. ESC: Embryonic stem cell; DC: Differentiated cell.

In this work, we selected six different initial parameter sets to perform optimization (the triangular points in Fig. 3). These six parameter sets can generate two stable states (ESC state and DC state), where the DC state is more stable. Due to the heterogeneity of biological individuals, these six initial parameter sets can be regarded as the corresponding cells of six different individuals. For each initial parameter set, we found an optimal intervention

to induce the transition from DC state to ESC state. The results of optimizing transition actions are shown in Fig. 6 (also marked in Fig. 1).

Specifically, we identified nine critical regulations from 26 regulatory parameters. Table 1 shows the top nine sensitive regulations and the corresponding sensitivity (sensitivity is defined as the average of the changes in the parameters (strength of
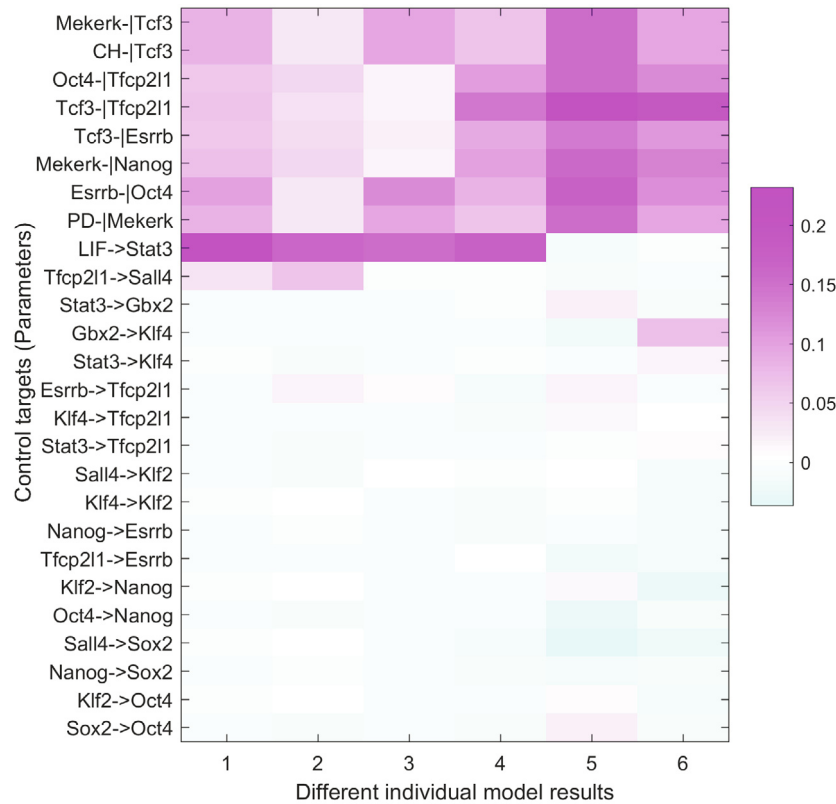
Fig. 6. Predictions for reprogramming-promoting interventions based on the optimization of transition actions. X axis shows 6 different individual model results (different initial parameter sets), reflecting the heterogeneity in stem cell populations, and Y axis shows 26 parameters (link strengths, 18 activation links and 8 inhibition links). Different colors indicate the changes in the parameters (strength of links) caused by each intervention, with purple grids representing the increase of targets (link strength), and cyan grids representing the decrease of targets.

Table 1
Cellular reprogramming-promoting interventions identified from optimization of transition actions for the model. Direction represents the direction of parameter changes for the corresponding interventions. Here, the inhibition constant B_{ji} represents the maximal synthesis rate when the inhibition regulation is completely not working. So, the increase of B_{ji} represents the decrease of the corresponding inhibition strength. Sensitivity is defined as the average of the changes in the parameters (strength of regulations) caused by each intervention across the six parameter sets.

Direction	Regulation	Sensitivity	References
↑	LIF →Stat3	0.123	[60]
↑	Tcf3 ⊣ Tfcp2l1	0.114	Prediction
↑	Esrrb ⊣ Oct4	0.104	[61]
↑	PD ⊣ MEKERK	0.089	[62]
↑	CH ⊣ Tcf3	0.089	Prediction
↑	MEKERK ⊣ Tcf3	0.089	Prediction
↑	MEKERK ⊣ Nanog	0.086	[63]
↑	Oct4 ⊣ Tfcp2l1	0.083	[11]
↑	Tcf3 ⊣ Esrrb	0.077	Prediction

regulations) caused by each intervention across the six parameter sets). From Table 1 and Fig. 6, we found that the important intervention for the transition action $S_{DC \to ESC}$ includes LIF → Stat3, Tcf3 ⊣ Tfcp2l1 and Esrrb ⊣ Oct4, which correspond to the first three rows of Table 1. Therefore, our results suggest a combinatory strategy for reprogramming by increasing the activation of LIF on Stat3, decreasing the inhibition of Tcf3 to Tcfp2l1 and the inhibition of Esrrb to Oct4, simultaneously.

To see the difference between single factor sensitivity analysis and the optimization approach, we compare the results of optimization and single factor sensitivity analysis. From the sensitivity analysis, the top three targets are LIF → Stat3, Tfcp2l1 → Sall4 and

Esrrb → Tfcp2l1. The increase of the strengths of these links can critically enhance the stability of the ESC state. It can be seen from the network that the increase of these link strengths will eventually increase the expression of Sall4. This indicates that these regulations have the similar functions. By contrast, the top three target regulations from optimization are not identical to those from single factor sensitivity analysis. The optimization result shows that we can simultaneously increase the strengths of LIF → Stat3, Tcf3 ⊣ Tfcp2l1 and Esrrb ⊣ Oct4 to inhibit differentiation. Both single factor sensitivity analysis and optimization indicate that an efficient reprogramming-promoting intervention should include LIF → Stat3, while the optimization result includes two other target links, and the function of these two other links are different from LIF → Stat3. Therefore, the optimization result seems to suggests that a better strategy should target links with different functions instead of links with the same functions.

3. Discussion

Embryonic stem cells have the potential to differentiate to any tissue in the body, which provides the motivation for the application of stem cell in regenerative medicine and other fields. However, the mechanisms of cellular differentiation and reprogramming remain to be elucidated. In this study, we established a data-constrained theoretical approach to model the gene regulatory network of mouse embryonic stem cells described by Dunn et al. [20]. Here, we built a model of ODEs for ESC, and proposed a data constrained approach to estimate the parameters of the models from the binary experimental data from Dunn et al. [20]. We identified a parameter set which can fit the experimental

data well, and the models using this parameter set can generate biologically significant bistable states. We calculated the potential landscape of the ESC network and identified two attractors, which represent the pluripotent ESC state and DC state, respectively. We further quantified the MAPs between ESC state and DC state. The results show that the differentiation path and reprogramming path are irreversible. This irreversibility of MAPs is a consequence of non-gradient force, i.e. curl flux [26].

By single factor sensitivity analysis of parameters, we provided some predictions about the key links affecting differentiation and reprogramming. The results from sensitivity analysis indicate that LIF/Stat3 signalling and gene Tfcp2l1 play important roles in maintaining the stability of the ESC state. These predictions agree well with previous experimental studies. For example, LIF is suggested to inhibit differentiation for mouse embryonic stem cells that maintains embryonic stem cells in a totipotent state and stimulates self-renewal [10,50]. Stat3 is seen to be the most important signal transducer following stimulation by LIF and the one which mediates most of the cellular effects. Overexpression of a dominant-negative Stat3 construct in ES cells also leads to both differentiation and a loss of self-renewal [9]. Additionally, the CP2 family transcription factor Tfcp2l1 is a common target in LIF/Stat3- and 2i-mediated self-renewal, and forced expression of Tfcp2l1 can recapitulate the self-renewal-promoting effect of LIF or either of the 2i components. This means that Tfcp2l1 is at the intersection of LIF- and 2i-mediated self-renewal pathways and plays a critical role in maintaining ESC identity [11].

Although single factor sensitivity analysis can uncover some single key parameters in the network, the cell fate decision in biological system is often governed by multiple factors rather than one. In the mean time, the multi-factor sensitivity analysis is not easy to implement due to the large computational cost. Therefore, it is necessary to develop an approach to predict the optimal combinations of multiple regulations that are critical to the dynamics of the embryonic stem cell model. Here, we used an approach of optimization to predict the critical combinations of parameters [44,49]. We identified an optimal combinatory strategy by increasing the activation of LIF on Stat3, decreasing the inhibition of Tcf3 to Tfcp2l1 and the inhibition of Esrrb to Oct4, to induce reprogramming. It is now well established that LIF/Stat3 signaling plays an indispensable role in the self-renewal of mouse embryonic stem cells [9,10,12,50]. Recently, activation of LIF-JAK1-STAT3 signaling to delay contact inhibition has been put forward as a strategy to facilitate engineering of HCEC (human corneal endothelial cells) grafts to solve the unmet global shortage of corneal grafts [12]. In addition, Esrrb and Tcf3 have been demonstrated to be required for efficient self-renewal of embryonic stem cells in vitro [13,14]. According to our model predictions, we propose that relevant experiments can be designed, e.g., to test the reprogramming effects by targeting multiple gene regulations simultaneously (including LIF → Stat3, Tcf3 ⊣ Tfcp2l1 and Esrrb ⊣ Oct4). We expect that these predictions can be tested with the further development of experimental techniques and will provide the guidance for designing the strategies of inducing reprogramming. These analyses on the landscape and transition actions will promote our mechanistic understanding on differentiation and reprogramming in embryonic stem cells.

In the field of systems biology, it remains challenging to combine "model driven" approach (such as gene network models) and "data driven" approach (such as single cell data analysis) in systems biology [51]. In this work, we used binary gene expression data to estimate the parameters in the model. We performed a search in the parameter space based on specific biological constraints. In principle, this approach can be applied to other data sets, such as single cell data. Recently, single cell data of gene expression have been extensively explored from different

approaches [52,53]. One limitation of current methods for single cell analysis is that the molecular mechanisms and kinetic transitions are hard to be studied. Our approach is based on dynamical models, and we use experimental data to infer the parameter regime for the models. This provides a potential way to combine "model-driven" approach and "data-driven" approach. It will be of great interest to develop a single cell data-constrained dynamical modelling approach, and construct more accurate energy landscapes for ESC and other biological networks.

4. Methods

4.1. Truncated Moment Equations (TME)

The time evolution of the gene expression level can be studied as a complex dynamical system. The gene expression level $\mathbf{x}(t) = (x_1(t), x_2(t), \cdots, x_n(t))^T$ can be regarded as a stochastic process, and $\mathbf{f}(\mathbf{x}(t))$ represents the driving force of the system. Then the Langevin equations describing the dynamics of the gene expression level take the form:

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t)) + \mathbf{g}(\mathbf{x}(t))\mathbf{\Gamma}(t), \qquad (2)$$

Here, $\mathbf{g}(\mathbf{x}(t))\mathbf{\Gamma}(t)$ is the noise term of the system, and coefficient matrix $\mathbf{g}(\mathbf{x}(t))$ is a matrix-valued function of $\mathbf{x}(t)$, while $\mathbf{\Gamma}(t) = (\Gamma_1(t), \Gamma_2(t), \cdots, \Gamma_n(t))^T$ is n-dimensional independent Gaussian white noise, which means:

$$\mathbb{E}[\Gamma_i(t)] = 0, \qquad (3)$$

$$\mathbb{E}[\Gamma_i(t)\Gamma_j(t')] = 2D\delta_{ij}\delta_0(t - t'), \qquad (4)$$

where D is constant diffusion coefficient and

$$\delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases} \qquad (5)$$

The time evolution of a dynamical system is determined by probabilistic diffusion equations (Fokker–Planck equation). However, when the drift part of Fokker–Planck equation is nonlinear, it is difficult to find an analytic solution. Nonetheless, the solutions of nonlinear system and linear system have similar form when the diffusion coefficient $D \ll 1$. Therefore, when $D \ll 1$, the solution could be approximated by Gaussian distribution and the mean and covariance satisfy the following equations [54,55]:

$$\dot{\bar{\mathbf{x}}}(t) = \mathbf{f}(\bar{\mathbf{x}}(t)), \qquad (6)$$

$$\dot{\boldsymbol{\sigma}}(t) = \boldsymbol{\sigma}(t)\mathbf{A}^T(t) + \mathbf{A}(t)\boldsymbol{\sigma}(t) + 2\mathbf{D}(\bar{\mathbf{x}}(t)). \qquad (7)$$

$\boldsymbol{\sigma}(t)$ denotes the covariance matrix in time t, and $\mathbf{A}(t)$ is the jacobian matrix of $\mathbf{f}(\mathbf{x})$ when $\mathbf{x}$ is equal to $\bar{\mathbf{x}}(t)$, which means $A_{ij}(t) = \frac{\partial f_i(\mathbf{x})}{\partial x_j}|_{\mathbf{x} = \bar{\mathbf{x}}(t)}$. Furthermore, $\mathbf{D}(\bar{\mathbf{x}}(t))$ is $n \times n$ matrix which is equal to $D \cdot \mathbf{g}(\bar{\mathbf{x}}(t))\mathbf{g}(\bar{\mathbf{x}}(t))^T$. Then the time evolution of density function $p(\mathbf{x}, t)$ for this system can be expressed by Eq. (8).

$$p(\mathbf{x}, t) = \frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{\sigma}(t)|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \bar{\mathbf{x}}(t))^T\boldsymbol{\sigma}^{-1}(t)(\mathbf{x} - \bar{\mathbf{x}}(t))\right\}, \quad (8)$$

In this work, we assume that the noise is homogeneous and only consider the external noise. Therefore, $\mathbf{g}(\mathbf{x})$ is equal to $\mathbf{I}$, which is an identity matrix in n-dimensional space. Then Eq. (7) is modified to:

$$\dot{\boldsymbol{\sigma}}(t) = \boldsymbol{\sigma}(t)\mathbf{A}^T(t) + \mathbf{A}(t)\boldsymbol{\sigma}(t) + 2D \cdot \mathbf{I}. \qquad (9)$$

When $t \to +\infty$, we can get the density function $p(\mathbf{x})$ of the system at the steady state by solving Eqs. (6) and (9). The probability distribution acquired above corresponds to one steady state. If the system has multiple steady states, there should be several proba-

bility distributions localized at each basin with different covariance. Therefore, the total probability is the weighted sum of all these probability distributions.

If we only focus on the distribution of two variables such as x_1 and x_2, we can get the marginal distribution by integrating the other variables in $p(\mathbf{x})$. Finally, we can construct the potential landscape U by $U = -\ln P(\mathbf{x})$ [26,30,44]. Here $P(\mathbf{x})$ is steady state probability distribution, and $U(\mathbf{x})$ is dimensionless potential.

4.2. Minimum action paths

The differential equation $\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t))$ is an autonomous system, which does not explicitly depend on the independent variable t. Assuming that the initial time is 0 and the terminal time is T, we define the path between ith attractor $\mathbf{x}^i$ and jth attractor $\mathbf{x}^j$ as $\mathbf{x}^{ij}(t) = \left(x_1^{ij}(t), x_2^{ij}(t), \cdots, x_n^{ij}(t)\right)^T$ for $t \in [0, T]$, where the path $\mathbf{x}^{ij}(t)$ should satisfy following terminal conditions:

$$\begin{cases} \left(x_1^{ij}(0), x_2^{ij}(0), \cdots, x_n^{ij}(0)\right)^T = \mathbf{x}^i, \\ \left(x_1^{ij}(T), x_2^{ij}(T), \cdots, x_n^{ij}(T)\right)^T = \mathbf{x}^j. \end{cases} \quad (10)$$

The transition action $\mathcal{S}^{ij}$ between $\mathbf{x}^i$ and $\mathbf{x}^j$ is defined as:

$$\mathcal{S}^{ij}\left(\mathbf{x}^{ij}(\cdot)\right) = \int_0^T L^{ij}\left(t, \mathbf{x}^{ij}(t), \frac{d\mathbf{x}^{ij}(t)}{dt}\right) dt. \quad (11)$$

Here, L^{ij} is the Lagrangian function that denotes the distance between the driving force $\mathbf{f}$ and the velocity along the path:

$$L^{ij}\left(t, \mathbf{x}^{ij}(t), \frac{d\mathbf{x}^{ij}(t)}{dt}\right) = \left\| \left(\frac{dx_1^{ij}(t)}{dt}, \frac{dx_2^{ij}(t)}{dt}, \cdots, \frac{dx_n^{ij}(t)}{dt}\right)^T - \mathbf{f}(\mathbf{x}^{ij}(t)) \right\|_2. \quad (12)$$

Following the approaches [56,57] based on Wentzell-Freidlin theory [58], the most probable transition path from ith attractor to jth attractor at time T can be acquired through minimizing the transition action functional over all possible paths:

$$\min_{\mathbf{x}^{ij}(\cdot)} \mathcal{S}^{ij}\left(\mathbf{x}^{ij}(\cdot)\right). \quad (13)$$

We calculated MAPs numerically by applying minimum action methods used in [44,57], and treated the MAPs as the biological paths in our models. In this work, T is set to be 5 and we verified that larger values of T would not improve accuracy of simulations significantly.

4.3. Transition action optimization

For a stochastic differential equation:

$$d\mathbf{x}(t) = \mathbf{f}(\mathbf{x})dt + \sqrt{\varepsilon}d\mathbf{W}(t), \quad (14)$$

the corresponding deterministic components are described by $\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x})$, and the time-independent stable states $\bar{\mathbf{x}}$ can be defined by $\mathbf{f}(\bar{\mathbf{x}}) = 0$. When there are more than one stable states in this system, the transitions between two stable states i and j occur as a Poisson process with a certain transition rate R_{ij}. The transition rate represents the transition probability per unit of time along the most likely transition path(s) between the stable states i and j. They can be obtained by an asymptotic formula [59]:

$$R_{ij} \propto exp\left(-\frac{\mathcal{S}_{ij}}{\varepsilon}\right), \quad (15)$$

where, $\mathcal{S}_{ij}$ is minimum action computed from MAPs. In order to achieve the goal of making ESC state more stable and DC state less stable, we minimized the ratio of $R_{DC \to ESC}$ to $R_{ESC \to DC}$ by tuning the 26 parameters (18 activation constants and 8 inhibition constants). We used the fmincon function of the optimization toolbox of MATLAB to implement the optimization.

CRediT authorship contribution statement

Xin Kang: Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing - original draft, Writing - review & editing. **Chunhe Li:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Writing - original draft, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.csbj.2020.02.004.

References

[1] Davila J, Cezar GG, Thiede M, Strom SC, Miki T, et al. Use and application of stem cells in toxicology. Toxicol Sci 2004;79:214–23.

[2] Perrier AL, Tabar V, Barberi T, Rubio ME, Bruses JL, et al. Derivation of midbrain dopamine neurons from human embryonic stem cells. Proc Natl Acad Sci USA 2004;101:12543–8.

[3] Graf T, Enver T. Forcing cells to change lineages. Nature 2009;462:587.

[4] Saha K, Jaenisch R. Technical challenges in using human induced pluripotent stem cells to model disease. Cell Stem Cell 2009;5:584–95.

[5] Zhou Q, Melton DA. Extreme makeover: converting one cell into another. Cell Stem Cell 2008;3:382–8.

[6] Yamanaka S. Elite and stochastic models for induced pluripotent stem cell generation. Nature 2009;460:49.

[7] Yamanaka S, Blau HM. Nuclear reprogramming to a pluripotent state by three approaches. Nature 2010;465:704.

[8] Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 2006;126:663–76.

[9] Niwa H, Burdon T, Chambers I, Smith A. Self-renewal of pluripotent embryonic stem cells is mediated via activation of stat3. Genes Develop 1998;12:2048–60.

[10] Williams RL, Hilton DJ, Pease S, Willson TA, Stewart CL, et al. Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. Nature 1988;336:684–7.

[11] Ye S, Li P, Tong C, Ying Q. Embryonic stem cell self-renewal pathways converge on the transcription factor tfcp2l1. EMBO J 2013;32:2548–60.

[12] Liu X, Tseng SCG, Zhang M, Chen S, Tighe S, et al. Lif-jak1-stat3 signaling delays contact inhibition of human corneal endothelial cells. Cell Cycle 2015;14:1197–206.

[13] Cole MF, Johnstone SE, Newman JJ, Kagey MH, Young RA. Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. Genes Develop 2008;22:746–55.

[14] Ivanova NB, Dobrin R, Lu R, Kotenko I, Levorse JM, et al. Dissecting self-renewal in stem cells with rna interference. Nature 2006;442:533–8.

[15] Chen X, Xu H, Yuan P, Fang F, Huss M, et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell 2008;133:1106–17.

[16] Lu R, Markowetz F, Unwin RD, Leek JT, Airoldi EM, et al. Systems-level dynamic analyses of fate change in murine embryonic stem cells. Nature 2009;462:358–62.

[17] Som A, Harder C, Greber B, Siatkowski M, Paudel Y, et al. The plurinetwork: an electronic representation of the network underlying pluripotency in mouse, and its applications. PLoS ONE 2010:5.

[18] Den Berg DLCV, Snoek T, Mullin NP, Yates A, Bezstarosti K, et al. An oct4-centered protein interaction network in embryonic stem cells. Cell Stem Cell 2010;6:369–81.

[19] Yeo J, Ng H. The transcriptional regulation of pluripotency. Cell Res 2013;23:20.

[20] Dunn S, Martello G, Yordanov B, Emmott S, Smith A. Defining an essential transcription factor program for naive pluripotency. Science 2014;344:1156–60.

[21] Lin YT, Hufton PG, Lee EJ, Potoyan DA. A stochastic and dynamical view of pluripotency in mouse embryonic stem cells. Plos Comput Biol 2018;14:e1006000.

[22] Waddington CH. The strategy of the genes: a discussion of some aspects of theoretical biology. London: Allen and Unwin; 1957. p. 290.

[23] Wang J, Zhang K, Xu L, Wang E. Quantifying the waddington landscape and biological paths for development and differentiation. Proc Natl Acad Sci USA 2011;108:8257–62.

[24] Liao C, Lu T. A minimal transcriptional controlling network of regulatory t cell development. J Phys Chem B 2013;117:12995–3004.

[25] Lv C, Li X, Li F, Li T. Energy landscape reveals that the budding yeast cell cycle is a robust and adaptive multi-stage process. PLOS Comput Biol 2015;11.

[26] Li C, Wang J. Landscape and flux reveal a new global view and physical quantification of mammalian cell cycle. Proc Natl Acad Sci USA 2014;111:14130–5.

[27] Lu M, Onuchic JN, Benjacob E. Construction of an effective landscape for multistate genetic switches. Phys Rev Lett 2014;113:078102.

[28] Ge H, Qian H. Landscapes of non-gradient dynamics without detailed balance: Stable limit cycles and multiple attractors. Chaos: Interdiscip J Nonlinear Sci 2012;22:023140.

[29] Ge H, Qian H. Mesoscopic kinetic basis of macroscopic chemical thermodynamics: a mathematical theory. Phys Rev E 2016;94:052150.

[30] Li C, Wang J. Quantifying cell fate decisions for differentiation and reprogramming of a human stem cell network: landscape and biological paths. PLoS Comput Biol 2013;9:e1003165.

[31] Li C, Wang J. Quantifying waddington landscapes and paths of non-adiabatic cell fate decisions for differentiation, reprogramming and transdifferentiation. J R Soc Interface 2013;10:20130787.

[32] Tse MJ, Chu BK, Gallivan CP, Read EL. Rare-event sampling of epigenetic landscapes and phenotype transitions. PLoS Comput Biol 2018;14:e1006336.

[33] Brackston R, Lakatos E, Stumpf MP. Transition state characteristics during cell differentiation. PLoS Comput Biol 2018;14:e1006405.

[34] Zhang B, Wolynes PG. Stem cell differentiation as a many-body problem. Proc Natl Acad Sci USA 2014;111:10185–90.

[35] Kang X, Wang J, Li C. Exposing the underlying relationship of cancer metastasis to metabolism and epithelial-mesenchymal transitions. iScience 2019;21:754–72.

[36] Li C, Ye L. Landscape and flux govern cellular mode-hopping between oscillations. J Chem Phys 2019;151:175101.

[37] Ying Q, Wray J, Nichols J, Batllemorera L, Doble BW, et al. The ground state of embryonic stem cell self-renewal. Nature 2008;453:519–23.

[38] Hall J, Guo G, Wray J, Eyres I, Nichols J, et al. Oct4 and lif/stat3 additively induce krüppel factors to sustain embryonic stem cell self-renewal. Cell Stem Cell 2009;5:597–609.

[39] Martello G, Sugimoto T, Diamanti E, Joshi A, Hannah R, et al. Esrrb is a pivotal target of the gsk3/tcf3 axis regulating embryonic stem cell self-renewal. Cell Stem Cell 2012;11:491–504.

[40] Martello G, Bertone P, Smith A. Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor. EMBO J 2013;32:2561–74.

[41] Silva J, Nichols J, Theunissen TW, Guo G, van Oosten AL, et al. Nanog is the gateway to the pluripotent ground state. Cell 2009;138:722–37.

[42] Tai C, Ying Q. Gbx2, a lif/stat3 target, promotes reprogramming to and retention of the pluripotent ground state. J Cell Sci 2013;126:1093–8.

[43] Lanner F, Lee KL, Sohl M, Holmborn K, Yang H, et al. Heparan sulfation-dependent fibroblast growth factor signaling maintains embryonic stem cells primed for differentiation in a heterogeneous state. Stem Cells 2010;28:191–200.

[44] Li C. Identifying the optimal anticancer targets from the landscape of a cancer-immunity interaction network. Phys Chem Chem Phys 2017;19:7642–51.

[45] Huang S, Guo Y, May G, Enver T. Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. Develop Biol 2007;305:695–713.

[46] Lu M, Jolly MK, Levine H, Onuchic JN, Benjacob E. Microrna-based regulation of epithelialchybridcmesenchymal fate determination. Proc Natl Acad Sci USA 2013;110:18144–9.

[47] Nichols J, Zevnik B, Anastassiadis K, Niwa H, Klewenebenius D, et al. Formation of pluripotent stem cells in the mammalian embryo depends on the pou transcription factor oct4. Cell 1998;95:379–91.

[48] Chambers I, Silva JR, Colby D, Nichols J, Nijmeijer B, et al. Nanog safeguards pluripotency and mediates germline development. Nature 2007;450:1230–4.

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

[50] Nicola NA, Babon JJ. Leukemia inhibitory factor (lif). Cytokine Growth Factor Rev 2015;26:533–44.

[51] Cho KH, Lee S, Kim D, Shin D, Joo JI, et al. Cancer reversion, a renewed challenge in systems biology. Curr Opin Syst Biol 2017;2:49–58.

[52] Petropoulos S, Edsgärd D, Reinius B, Deng Q, Panula SP, et al. Single-cell rna-seq reveals lineage and x chromosome dynamics in human preimplantation embryos. Cell 2016;165:1012–26.

[53] Jang S, Choubey S, Furchtgott L, Zou LN, Doyle A, et al. Dynamics of embryonic stem cell differentiation inferred from single-cell transcriptomics show a series of transitions through discrete cell states. Elife 2017;6:e20487.

[54] Hu G. Stochastic forces and nonlinear systems; 1994. pp. 68–74..

[55] Van Kampen N. Stochastic processes in chemistry and physics..

[56] E W, Ren W, Vanden-Eijnden E. Minimum action method for the study of rare events. Commun Pure Appl Math 57; 2004: 637–656..

[57] Zhou X, Ren W, Weinan E. Adaptive minimum action method for the study of rare events. J Chem Phys 2008;128:104111.

[58] Freidlin M, Weber M. Random perturbations of dynamical systems and diffusion processes with conservation laws. Prob Theory Rel Fields 2004;128:441–66.

[59] Wells DK, Kath WL, Motter AE. Control of stochastic and induced switching in biophysical networks. Phys Rev X 2015;5:031036.

[60] Atwood C. Embryonic stem cells: the hormonal regulation of pluripotency and embryogenesis. BoD–Books on Demand; 2011..

[61] Zhang X, Zhang J, Wang T, Esteban MA, Pei D. Esrrb activates oct4 transcription and sustains self-renewal and pluripotency in embryonic stem cells. J Biol Chem 2008;283:35825–33.

[62] Li J, Wang G, Wang C, Zhao Y, Zhang H, et al. Mek/erk signaling contributes to the maintenance of human embryonic stem cell self-renewal. Differentiation 2007;75:299–307.

[63] Hyslop L, Stojkovic M, Armstrong L, Walter T, Stojkovic P, et al. Downregulation of nanog induces differentiation of human embryonic stem cells to extraembryonic lineages. Stem Cells 2005;23:1035–43.