Research article

Single-cell data-driven mathematical model reveals possible molecular mechanisms of embryonic stem-cell differentiation

Xiao Tu[1]**, Qinran Zhang**[1]**, Wei Zhang**[2] **and Xiufen Zou**[1,*]

[1] School of Mathematics and Statistics, Wuhan University, Wuhan, P. R. China

[2] School of Science, East China Jiaotong University, Nanchang, P. R. China

* **Correspondence:** Email: xfzou@whu.edu.cn.

Abstract: Embryonic development is widely studied due to its application in disease treatment. The published literature demonstrated that _Krüppel_-like factor 8(KLF8) plays an important role in modulating mesendoderm to definitive endoderm (DE) differentiation. However, it is not clear how KLF8 interacts with other key genes and affects the differentiation process. To qualitatively and quantitatively explore the molecular mechanisms of KLF8 during the differentiation of human embryonic stem cells (hESCs) in detail, we developed a mathematical model to describe the dynamics between KLF8 and two other significant genes, E-cadherin(CDH1) and Zinc-finger E-box-binding homeobox1(ZEB1). Based on the single-cell RNA-seq data, the model structure and parameters were obtained using particle swarm optimization (PSO). The bifurcation analysis and simulation results reveal that the system can exhibit a complex tristable transition, which corresponds to the three states of embryonic development at the single-cell level. We further predict that the novel important gene KLF8 promotes the formation of DE cells by reciprocal inhibition between CDH1 and KLF8 and promotion of the expression of ZEB1. These results may help to shed light on the biological mechanism in the differentiation process of hESCs.

Keywords: embryonic stem cells(ESCs); single-cell data; bifurcation; molecular mechanisms; tristability

1. Introduction

Embryonic stem cells (ESCs) are pluripotent stem cells that are derived from the inner cell mass of a blastocyst which is an early-stage preimplantation embryo that can be propagated by culturing in an undifferentiated state while maintaining the capacity to generate any cell type in the body [1]. Due to their plasticity and potentially unlimited capacity for self-renewal, human embryonic stem cells (hESCs) play an important role in animal cloning, organ transplants, etc. For instance, embryonic

stem cell therapies have been proposed for regenerative medicine and tissue replacement after injury or disease [2, 3].

The transition from embryonic stem cells to three different types of cells in the ectoderm, mesoderm and endoderm has been the first step in studying how pluripotent cells exit the pluripotent state and give rise to lineage-specific progenitors. From the perspective of molecular biology, the transition is determined by different levels of gene expression. Thus, understanding the regulatory mechanism of the genes underlying differentiation and the cell fate decision of hESCs has become necessary.

In the last few years, genome-wide profiling approaches have started to uncover the molecular programs that drive the developmental processes of hESCs. A number of research works have revealed the key genes or mechanisms underlying the differentiation of hESCs [4–10]. For example, Adamo *et al.* revealed that LSD1 could regulate the balance between self-renewal and differentiation in hESCs [4]. Cheryle *et al.* demonstrated that stable endoderm progenitors can be established from hESCs by the constitutive expression of SOX7 or SOX17, producing extraembryonic endoderm and definitive endoderm progenitors, respectively [9]. Ivanova *et al.* found that OCT4 regulates and interacts with the BNP4 pathway to specify four developmental fates and identified general and cell-line-specific requirements for NANOG, OCT4, and SOX2 in the hESC [10].

Recently, due to the development of single-cell sequencing technology, a new type of data named "single-cell RNA-seq data (scRNA-seq data)" is available and we can now quantify the gene expression of individual cells and analyze the heterogeneity among cells [11, 12]. Through analyzing the scRNA-seq data of hESCs, Chu *et al.* revealed that the *Krüppel*-like factor 8 (KLF8) plays a key role in modulating the mesendoderm, which is an intermediate state before the definitive endoderm and mesoderm are formed, to DE state differentiation [5].

However, the dynamic molecular mechanisms underlying the differentiation of individual human embryonic stem cells are still poorly understood. Moreover, how KLF8 dynamically interacts with other important genes in the core regulatory network and governs the transition from ESCs to the three different cell types in the ectoderm, mesoderm and endoderm remains unclear [13–15].

Mathematical modeling has been demonstrated as a powerful tool for investigating the dynamic mechanisms underlying signal regulatory network. To systematically analyze the dynamical regulatory mechanism underlying hESC differentiation, in this paper, we developed a mathematical model that is based on a possible core regulatory network underlying the differentiation process of hESCs. scRNA-seq data and the particle swarm optimization (PSO) algorithm are used to identify the parameters of the mathematical model. Then, we performed the simulation and dynamic bifurcation analysis that is based on the proposed model and investigated the molecular mechanisms underlying the core regulatory module. More importantly, we proved the reasonability of the proposed regulatory mechanisms through the dynamic analysis results and introduced several strategies that may be feasible for controlling the process of hESC differentiation.

2. Mathematical models and methods

2.1. The core transcriptional regulation underlying human embryonic stem-cell differentiation

It has been documented that epithelial-mesenchymal transition (EMT) underlies cell fate conversions in both reprogramming and differentiation along an endoderm cell fate [16]. HESCs begin differentiation with a near-synchronous EMT, and the differentiation of hESCs is considered an

EMT process [17]. In fact, hESCs are epithelial cells with a high expression of E-cadherin (CDH1), which plays a central role in maintaining epithelial cell-cell adhesion and polarity. Previous work [18, 19] has shown that down regulation of CDH1's transcription is thought to be a primary mechanism that contributes to the onset of EMT. The CDH1 has also been demonstrated to be the the paradigm of epithelial genes. Loss of CDH1 is normally determined to be the hallmark of EMT [20].

In the differentiation of hESCs, KLF8 is recognized to be an essential regulator in modulating mesendoderm to endoderm differentiation. It was observed that overexpression of KLF8 increases the mobility, which was evident by the up regulation of TWIST1, an EMT marker, which indicated that KLF8 plays a specific role in promoting the transition from mesendoderm to endoderm state [5]. Indeed, KLF8 induces EMT by directly binding to the CDH1 promoter through GT boxes and repressing the expression of CDH1 [21].

Zinc finger E-box-binding homeobox1 (ZEB1) has been well studied in cancerous tissues, such as cervical cancer, gastric cancer, lung cancer [22–24]. In addition, it was demonstrated that ZEB1 is an important transcription factor of EMT by inhibiting an E-box gene, including CDH1 [23, 25, 26].

To obtain the appropriate model, we first used single-cell RNA-seq data to compute Spearman's correlation coefficients among three genes. The calculated Spearman's correlation coefficients between ZEB1 and KLF8, ZEB1 and CDH1, and CDH1 and KLF8 are 0.4879, -0.4664 and -0.4851, respectively.

According to the above descriptions, it is obvious that ZEB1 inhibits CDH1 and KLF8 inhibits CDH1. No evidence shows how KLF8 influences ZEB1 and CDH1. It has also been suggested that KLF8 functions as a promoter of DE cells by upregulating many other DE cell marker genes, such as CXCR4 [5]. We assumed several models to describe their different relations and determined that good hypotheses are that ZEB1 is promoted by KLF8 and CDH1 inhibits the expression of KLF8. The connections among the three genes are contained in the following Figure 1.
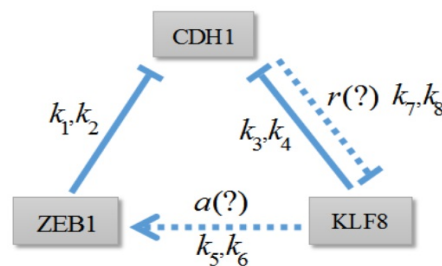
Figure 1. The core regulatory networks we build. The dashed line with (?) means that it is unknown whether the regulation relationship exists. Edges with an arrow stand for activation and edges with a blunt side stand for inhibition.

2.2. The preprocess of single-cell RNA-seq data

The single-cell RNA-seq data we used in this research is obtained from NCBIs Gene Expression Omnibus and are accessible through GEO series accession number GSE75748, which characterizes differentiation from the mesendoderm(a state before the formation of DE cells and mesoderm cells) to definitive endoderm cells [5]. This data set contains a large number of genes expression data collected at 6 time points with 758 cells.

Based on the single cell RNA-seq data, we first reconstructed a single-cell order using the Wave-Crest toolbox under R software. The cell order assigns a specific time point to each cell within the time interval [0 h, 96 h]. There are 758 time points evenly distributed from 0 h to 96 h (each pair of adjacent time points is 96 h/757 apart according to the Wave-Crest algorithm). These 758 time points correspond to the 758 reordered cells. The cell order produces a time-series expression of genes, which was used to calculate the Spearman's correlation coefficients (Table 1). Then, we obtained a pseudo trajectory of the gene-specific expression through polynomial fitting. The normalized scRNA-seq data is obtained by evenly selecting 758 points of the polynomial curve with the same time points. It has been demonstrated that the time series expression of the three considered genes reflect the dynamical differentiation well [5]. After further investigation, we discovered that the parameters estimated by the data not only capture the differentiation from the mesendoderm state to the DE state, but, in fact, changing some of the parameters can simulate the differentiation from ESCs to all three germ layers. We speculate that the function of genes and the interactions between genes at different stages of differentiation remain unchanged. The differentiation trend is mainly determined by specific biological processes.

Table 1. The Spearman's correlation coefficients between gene expressions.

Genes	CDH1&ZEB1	CDH1&KLF8	ZEB1&KLF8
Spearman's correlations	–0.4664	–0.4851	0.4879

2.3. Mathematical modeling of the core regulatory module

We built a mathematical model based on mass action law and Michaelis-Menten equation [27, 28], which is described in the following ordinary differential equations (ODEs). The relationships among the three selected genes has been thoroughly discussed in the network construction, and each term of the equations has been marked accordingly.

$$\frac{d[CDH1]}{d\tau} = \underbrace{\frac{k_1}{k_2 + [ZEB1]^{n_1}}}_{\text{ZEB1 inhibits the transcription of CDH1}} + \underbrace{\frac{k_3}{k_4 + [KLF8]^{n_2}}}_{\text{KLF8 inhibits the transcription of CDH1}} - \underbrace{d_1[CDH1]}_{\text{degradation of CDH1}} \quad (1)$$

$$\frac{d[ZEB1]}{d\tau} = \underbrace{\frac{a \times k_5[KLF8]^{n_3}}{k_6 + [KLF8]^{n_3}}}_{\text{assume KLF8 promotes transcription of ZEB1}} - \underbrace{d_2[ZEB1]}_{\text{degradation of ZEB1}} \quad (2)$$

$$\frac{d[KLF8]}{d\tau} = \underbrace{\frac{r \times k_7}{k_8 + [CDH1]^{n_4}}}_{\text{assume CDH1 inhibits transcription of KLF8}} - \underbrace{d_3[KLF8]}_{\text{degradation of KLF8}} \quad (3)$$

Here k_1, k_3, k_5 and k_7 represent the maximum production rate of CDH1 regulated by ZEB1, the maximum production rate of CDH1 regulated by KLF8, the maximum production rate of ZEB1 regulated by KLF8 and the maximum production rate of KLF8 regulated by CDH1, respectively. k_2, k_4, k_6 and k_8 represent the Michaelis constants that correspond to the production rates k_1, k_3, k_5 and k_7, respectively. d_1, d_2 and d_3 represent the degradation rates of CDH1, ZEB1 and KLF8, respectively. a and r represent weight coefficients to reflect the degree of activation and inhibition, respectively.

To reduce the parameters and make the following analysis more convenient, we nondimensionalized the model by making the following substitutions, whereby we assume that the Hill coefficients are equal to 2.

$$y_1 = \frac{k_2 d_1}{k_1}[CDH1], y_2 = \frac{k_6 d_1}{k_4 k_5}[ZEB1], y_3 = \frac{[KLF8]}{\sqrt{k_4}} \tag{4}$$

Therefore, the simplified ODE model is as following, and all the analysis is based on the nondimensionalized ODEs.

$$\frac{dy_1}{dt} = \frac{1}{1 + p_1 y_2^2} + \frac{p_2}{1 + y_3^2} - y_1 \tag{5}$$

$$\frac{dy_2}{dt} = \frac{a \times y_3^2}{1 + p_3 y_3^2} - p_4 y_2 \tag{6}$$

$$\frac{dy_3}{dt} = \frac{r \times p_5}{1 + p_6 y_1^2} - p_7 y_3 \tag{7}$$

where

$$p_1 = \frac{k_4^2 k_5^2}{k_6^2 d_1^2}, p_2 = \frac{k_2 k_3}{k_1 k_4}, p_3 = \frac{k_4}{k_6}, p_4 = \frac{d_2}{d_1}, p_5 = \frac{k_2^2 k_7 d_1}{k_1^2 \sqrt{k_4}}, p_6 = \frac{k_2^2 k_8 d_1^2}{k_1^2}, p_7 = \frac{d_3}{d_1} \tag{8}$$

The model includes three variables and night parameters. To identify the parameters in the model, we converted the parameter identification into the problem of optimization by minimizing the following objective function. Mathematically, the objective function is defined as the error between the simulation results and the time series experimental data. The formulation can be expressed as

$$\min_P J(P) = \alpha \sum_{k=1}^{K} \sum_{j=1}^{J} (\frac{y_k^D(t_j) - y_k(t_j, P)}{\max_j(y_k^D(t_j))})^2 \tag{9}$$

$y_k^D(t_j)$ represents the measured data of component k at time-point t_j, which, in our case, is the normalized data obtained from the polynomial curves. $y_k(t_j, P)$ represents the kth component of the solutions of ODE at time t_j with parameter set P. Here $J = 758$, since there are 758 pseudo time points. The numerical solution of the ODEs is solved by MATLAB, with the initial value set to be the approximation of the first data point of the normalized data and a, r assumed to be 1. The particle swarm optimization(PSO) [29] algorithm is used to optimize this object function within the optimization intervals that are carefully chosen. The obtained optimization parameters are listed in Table S1.

3. Results and discussion

3.1. The consistency between the simulation results and measured data

We used the proposed model to simulate the dynamic process with time. Figure 2a,c,e demonstrates that the simulation results are consistent with the obtained trajectory of single-cell RNA-seq data, which shows that our model could reflect the main biological process underlying the differentiation of ESC. The fitting of 758 scRNA-seq data is shown in Appendix D.
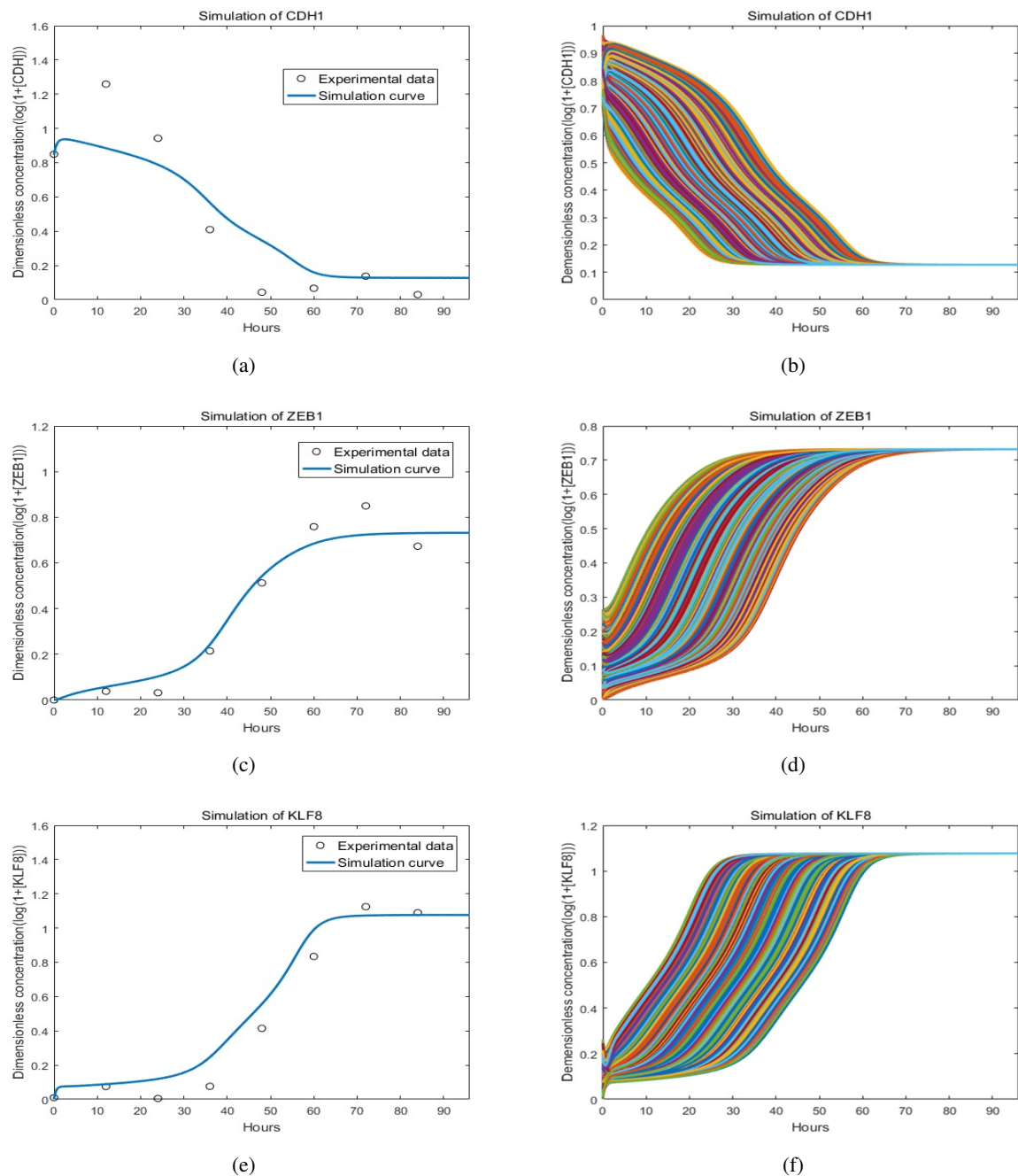
Figure 2. Comparisons between the numerical simulation and the normalized data. We used 8 data points to represent the 758 scRNA-seq data. (a)(c)(e) are the solutions of ODEs and their uniformity with the normalized gene expression date of the $1^{st}, 95^{th}, 189^{th}, 283^{rd}, 377^{th}, 471^{st}, 565^{th}$ and 659th cell, respectively. The corresponding 8 time points are 0 h, 11.9 h, 23.8 h, 35.8 h, 47.7 h, 59.6 h, 71.5 h and 83.4 h. (b)(d)(f) are the solutions of the model with uniformly distributed random perturbations of the initial values.

Since the RNA-seq data we obtained was collected from large amount of cells which may lead to the errors between the real initial gene expression level and the initial expression level we estimated, we gave the initial value random perturbations within 0.3 (Figure 2b,d,f). The results show that the model well reflects the fluctuation of the expression data with different initial values. The simulation is independent of the selection of initial values to a certain extent.

3.2. Bifurcation analysis reveals the threshold dynamics of the relative degradation rate corresponding to different stages of differentiation

We presented dynamic simulations of the system with different values of p_4. The curves in Figure 3a indicate that the dimensionless concentration of CDH1 gradually decreases after a short increase until a steady state is attained during the time evolution. The steady state remains the same with an increase of p_4. Intriguingly, the steady state significantly increases into an extremely high state when p_4 is larger than 0.133. Conversely, a relatively lower state is maintained when p_4 is less than 0.133. It is obvious that the system will sustain more steady states with the change of p_4.

(a)   (b)
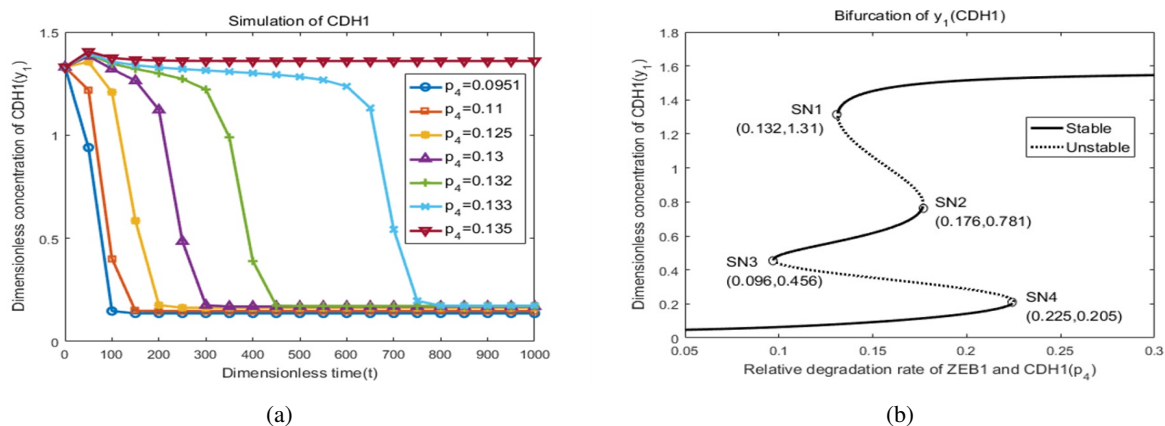
Figure 3. Dynamics of CDH1 with the change of relative degradation p_4. (a) Simulation of the dimensionless concentration of CDH1 with different values of p_4. (b) The one-parameter bifurcation graph for the dimensionless model (Eqs 5–7) with respect to p_4. The solid lines describe the stable steady states of CDH1 versus p_4. The dashed line between the two circles corresponds to the unstable steady states. SN represents the saddle node.

Through the nondimensionalization process, we know that p_4 is the ratio of the degradation rate of ZEB1 to the degradation rate of CDH1, which we named the "relative degradation rate". We hypothesize the existence of thresholds for the relative degradation rate, which eventually distinguishes different levels of CDH1 expression.

The threshold of the relative degradation rate is clearly indicated by Figure 3b. Through qualitative and quantitative analyses, we suggest that a high expression of CDH1, in which relative degradation rate is greater than 0.225, corresponds to the ectoderm state of ESCs. With the relative degradation rate decreasing, the inhibition of ZEB1 targets on CDH1 is enhanced, at least partly, because of the relatively lower degradation rate of ZEB1, which results in weaker adhesion among cells. Thus, EMT is activated to a certain extent, and the system undergoes SN bifurcation and another stable steady

state occurs, which we believe corresponds to the mesoderm state. As the relative degradation rate decreases further (< 0.096), EMT is further activated and the system undergoes another SN bifurcation. Then, the lower expression state of CDH1 occurs, which corresponds to the formation of DE cells. In short, the threshold we discovered is the necessary condition for cell formation in different germ layers. The results also indicate that mesendoderm cells can repossess pluripotency and differentiate into ectodermal cells, and the differentiation process is reversible under certain conditions. There are reasons to believe that the proposed model is able to characterize the differentiation from ESCs to the three embryonic layers.

3.3. The relative dissociation constant distinguishes DE cells from cells in other germ layers

We investigated the effect of inhibition intensity on the dynamic behavior of the system. Figure 4a shows that a bistable phenomenon exists when p_3 varies in the region [9.65, 23]. According to the nondimensionalization process (Eq 8), p_3 is the ratio of k_4 and k_6, which are the dissociation constants that KLF8 targets on CDH1 and ZEB1, respectively. Since p_3 is equal to the ratio of the dissociation rate of the ligand-receptor complex, we named it as the relative dissociation constant, and we found that there is a threshold of the relative dissociation constant. When the dissociation constant of the ligand-receptor complex of CDH1 and KLF8 is approximately 23 times greater than that of ZEB1 and KLF8, the system undergoes a bistable switch from a low-expression state to a high-expression state of CDH1. The bifurcation diagram indicates that the change of p_3 might not be a good way to obtain all three different types of cells during ESC differentiation.

However, we suggest that influencing the relative dissociation constant might be a feasible way to obtain DE cells from cells in other germ layers.

3.4. Bifurcation analysis reveals the connections of feedback coefficients

To further determine how p_3 affects the whole system, we analyzed the two-parameter bifurcation. According to Figure 4b, when the relative dissociation constant p_3 decreases to a small enough amount (for example, $p_3 = 0.3$), the tristable phenomenon disappears, and the system undergoes an irreversible bistable switch, which means that it is impossible to obtain mesoblastema and hard to change the fate of cells from differentiated cell into DE cells. In this case, CDH1 maintains a low level of expression, and KLF8 maintains a high level of expression. Interestingly, when the relative dissociation constant becomes greater, the system exhibits only one stable steady state of high concentration of CDH1.

Qualitatively we find that a large dissociation constant of the ligand-receptor complex of CDH1 and KLF8 and a relatively smaller dissociation constant of the ligand-receptor complex of ZEB1 and KLF8 resulted in the activation of CDH1 (the case where $p_3 = 35$), which is partly due to the situation that the inhibitory effect of CDH1 by KLF8 is weakened by a large dissociation constant. Instead, with a smaller dissociation constant of ligand-receptor complex of CDH1 and KLF8 compared to that of ZEB1 and KLF8, the direct inhibition of CDH1 by KLF8 is enhanced (the case where $p_3 = 0.2$), which leads to a state of low expression of CDH1, although ZEB1 is promoted by KLF8.

Furthermore we have the following equation according to Eq 8.

$$p_1 = p_3^2 \frac{k_5^2}{d_1^2}$$

Under the circumstance that p_3 and d_1 are constants, the decrease of p_1 results from the the decrease of k_5, which characterizes the promotion of ZEB1 by KLF8. It reveals that the system may exhibit only one steady state if the relative dissociation constant p_3 is too large or too small, no matter how the production rate of ZEB1 changes.


(a)

(b)

(c)

(d)
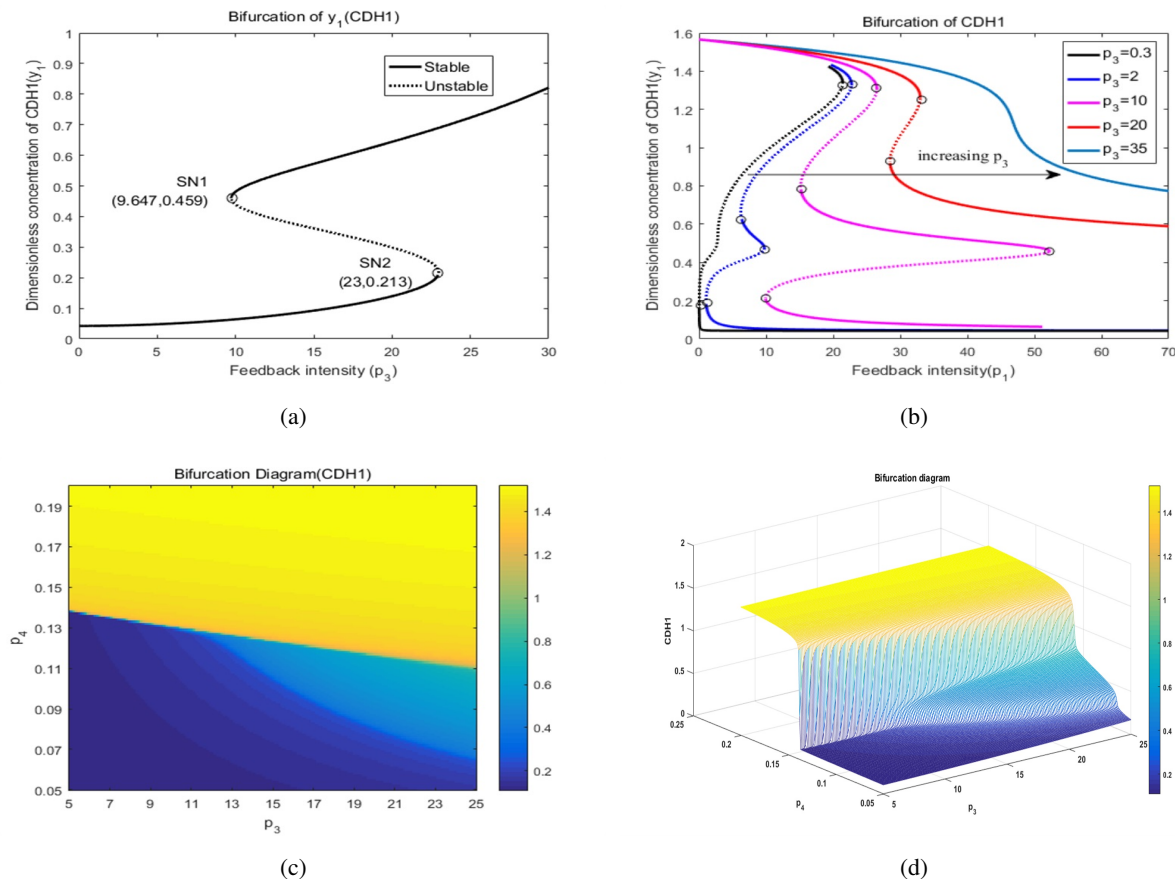
Figure 4. Bifurcation analysis of parameters. (a) One parameter bifurcation graph of p_3. The solid lines describe the steady states of CDH1 versus feedback intensity p_3. The dashed line between two circles corresponds to an unstable state. SN represents the saddle node bifurcation point. (b) The bifurcation diagram of coefficient p_1 when p_3 varies. Solid lines denote stable equilibrium states and dashed lines denote unstable equilibrium states. (c) The bifurcation diagram with respect to p_3, p_4. The vertical coordinates for each pair of (p_3, p_4) shows the dimensionless concentration of the CDH1 expression. (d) The bifurcation diagram with respect toin the parametric plane. The color shows the dimensionless concentration of the CDH1 expression.

Because p_3 and p_4 play important roles in the ESC fate decision, we start to wonder how they synergistically affect the system. Figure 4c,d shows the varying steady states of the model with the two constants in different domains. From the parameter values of p_3, p_4 that correspond to different steady states, we suggest that it is difficult to obtain mesodermal cells with the relatively weakened inhibition effect that KLF8 targets on CDH1 or the weakened promotion that KLF8 targets on ZEB1.

Meanwhile, a high relative degradation rate of ZEB1 to CDH1 determines the irreversible steady state of epithelial cell formation, which corresponds to the differentiation of the ectoderm according to the yellow district of Figure 4d.

In conclusion, the relative dissociation constant p_3 should be neither too large nor too small in case of an irreversible situation of the steady states dominated by KLF8 if three types of embryonic cells are needed.
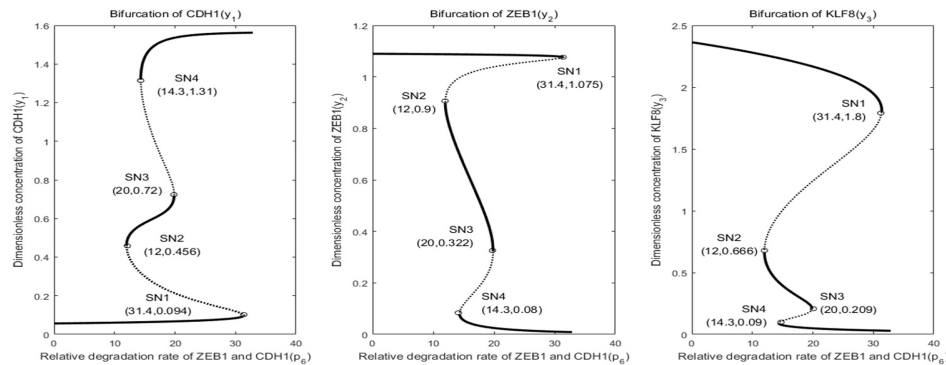
Figure 5. The bifurcation diagrams of three components as a function of the inhibition coefficient that CDH1 targets on the expression of KLF8 (p_6). The solid lines describe the steady states of CDH1, ZEB1 and KLF8 versus p_4, respectively. The dashed line between the two circles corresponds to an unstable state. SN represents the saddle node.

3.5. KLF8 promotes the formation of DE cells possibly by the activation of ZEB1 and the mutual inhibition of KLF8 and CDH1

Previous experimental and dynamical analyses show that KLF8 plays an important role in modulating cell differentiation from mesendoderm to DE. To systematically analyze the role of KLF8 in the differentiation, we plotted the bifurcation diagram (Figure 5) with respect to the inhibition coefficient that CDH1 targets on KLF8 (p_6). With an increase of the inhibition coefficient p_6, CDH1 switches from the monostability of low level to tristable states in a narrow region, then to the high level of bistable states. It reveals that if we want to switch the DE state to other states of hESC, we may enhance the inhibition strength of CDH1 on KLF8, or artificially, we inhibit the expression of KLF8 through medication. Combined with the fact that KLF8 inhibit the expression of CDH1, we assume that KLF8 promotes the formation of DE cells possibly by mutual inhibition of KLF8 and CDH1.

The above analysis is based on the assumption that $a = 1, r = 1$. To further investigate the function of KLF8 in the system, we decreased parameter a to characterize different intensities of CDH1 on KLF8 and plotted Figure 6a to analyze the behavior of KLF8.

With the decrease of a, the tristability phenomenon gradually disappears, and the system undergoes an irreversible trend towards a high expression level of CDH1 (when a reaches the threshold 0.15), which marks the formation of epithelial cells. Similarly, the decrease of r leads to the disappearance of the tristability phenomenon of the system (Figure 6b), which finally resulted in the formation of the ectoderm. The simulation results show similar conclusions (see Figure S2). We reasonably suggest that the promotion of ZEB1 by KLF8 and the inhibition of KLF8 by CDH1 is indispensable for DE cell

formation which is consistent with the complex biological process. More specifically, we assume that the promotion of ZEB1 by KLF8 does not exist, which results in changes of the promotion term from a nonlinear function of KLF8 to a linear constant c. The simulation results (Figure 6c) indicate that when the KLF8-related production rate of ZEB1 degenerates into constants c, the expression of CDH either maintains a high level of expression or reduces it directly, which is inconsistent with experimental observations. Figure 6d shows the same conclusion that the production rate of KLF8 is a nonlinear function of CDH1, as was indicated by our model.
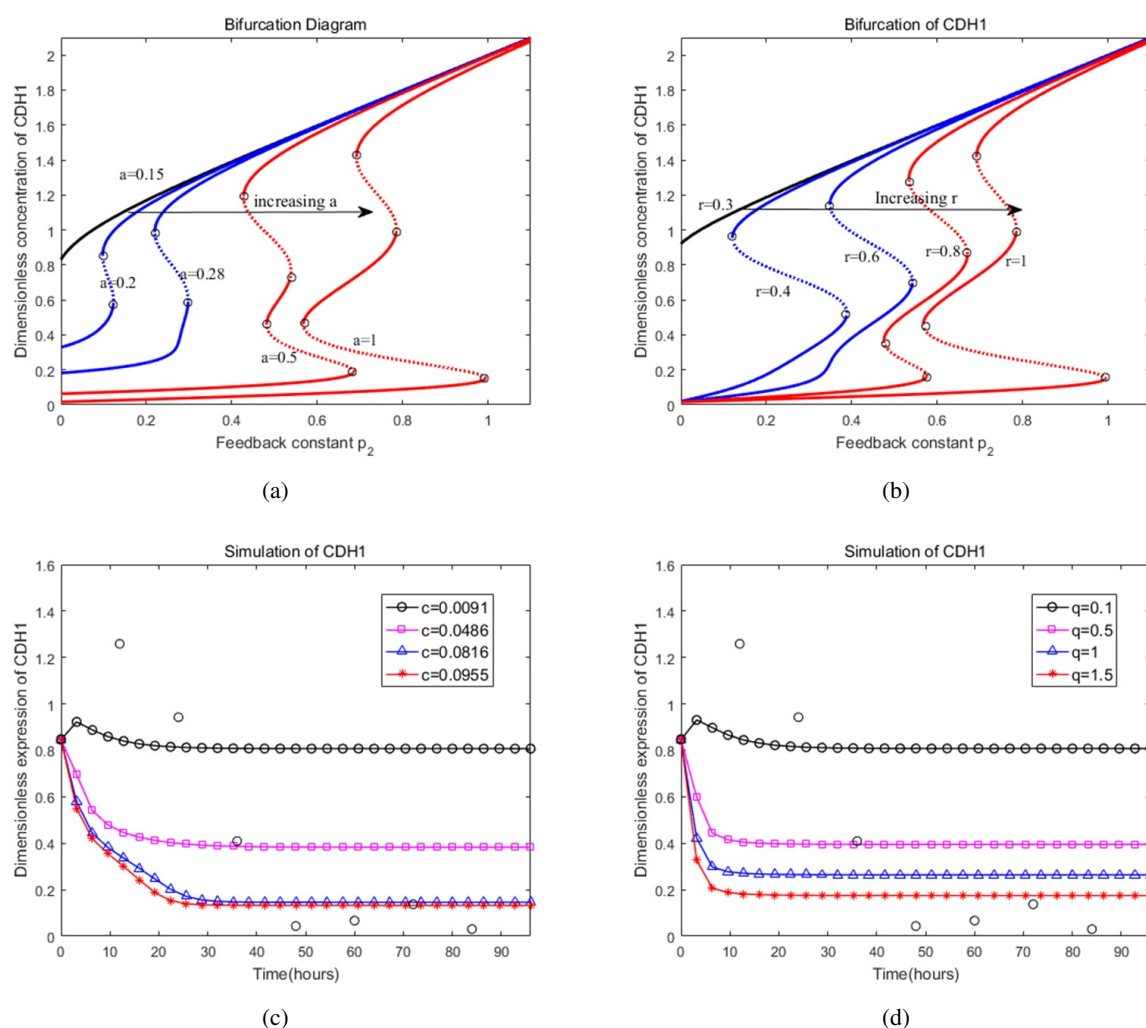
Figure 6. (a)The bifurcation diagram of CDH1 with respect to the feedback constant p_2 and the impact intensity a. (b)The bifurcation diagram of CDH1 with respect to the feedback constant p_2 and the impact intensity r. The solid lines describe the steady states of each gene. The dashed line between two circles corresponds to an unstable state. SN represents the saddle node. (c)The simulation when the KLF8-related production rate of ZEB1 degenerates into the constant c. The production rate c varies from 0 to 1. (d)The simulation where the CDH1-related production rate of KLF8 degenerates into the constant q. The production rate q varies from 0 to 1.

In fact, the intrinsically complex processes require the coordinated dynamic expression of hundreds of genes and proteins in precise response to external signaling cues. Our work only focuses on a fraction of the complex regulatory network, which raises another question about whether the model really reflects the biological differentiation process. To verify the generality and reliability of the proposed model, we analyzed it from three aspects. First, we tried numbers of model settings, e.g., we set the Hill coefficients to 1 and found that the numerical simulations of the model did not fit the data well. We have also tried the linear functions in the model as a regulatory method between two genes. The fitting results remain poor. Secondly, to testify whether the parameters are sensitive to the single-cell data we used in the parameter estimation, we performed 10-fold cross-validation (Appendix E). The parameters that were estimated by 10 data subsets vary slightly and are similar to the parameters listed in Table S1. Third, since cell differentiation is typically heterogeneous and is spatially disordered, the intrinsic fluctuations and extrinsic signal fluctuations may play important roles in modulating the state switch in the differentiation of hESCs. To explore the influence of intrinsic and extrinsic noise on the transition between multiple steady states of the proposed model, we developed corresponding Master equation and Langevin equation models (Appendix F) and performed stochastic simulations using the Gillespie algorithm and Euler-Maruyama algorithm [30], respectively. According to the simulation results (Figures S4 and S5), the Master equation and Langevin equation models displayed similar dynamic behavior as that by the deterministic model.

The above evidences proved that the proposed model can characterize the differentiation from hESCs to three germ layers and we suggest that KLF8 promotes the formation of DE cells possibly with the promotion of ZEB1 and the mutual inhibition of KLF8 and CDH1, which partially answers the unsolved question [5].

4. Conclusions

In this study, based on the single-cell RNA-seq data, we reconstructed a core regulatory network underlying the stem-cell differentiation process, and demonstrated that the core regulatory module shows various behaviors, including the numerical fitting and the three states switch, which is in correspondence with the three types of cell in three germ layers. Thus, we proved that the novel important gene KLF8 [5] affects the differentiation process by up-regulating ZEB1 and down-regulating the expression of CDH1, which forms a coupled feedback loop. In addition, we defined two indexes including the relative degradation rate and the relative disassociation rate which are tightly associated with the complex dynamic behaviors.

We also proposed some possible methods to realize the switch of the cell fate.

- The small relative dissociation constant ensures the formation of DE cells, and the large relative dissociation constant ensures the formation of epithelial cells;
- The inhibition of KLF8 contributes to the formation of epithelial cells by blocking EMT;
- The high relative degradation rate of ZEB1 to CDH1 determines the steady state of epithelial cell formation.

Altogether, we believe that the combination of scRNA-seq analysis and mathematical modeling can well reveal the molecular mechanism of cell fate decisions.

Acknowledgments

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. J. A. Thomson, J. Itskovitz-Eldor, S. S. Shapiro, et al., Blastocysts Embryonic Stem Cell Lines Derived from Human, *Science*, **282** (1998), 1145–1147.

2. C. E. Murry, M. A. Laflamme, X. Yang, et al., Human embryonic-stem-cell-derived cardiomyocytes regenerate non-human primate heart, *Nature*, **510** (2014), 273–277.

3. P. P. Tam and D. A. Loebel, Gene function in mouse embryogenesis: get set for gastrulation, *Nat. Rev. Genet.*, **8** (2007), 368–381.

4. A. Adamo, I. Paramonov, M. J. Barrero, et al., LSD1 regulates the balance between self-renewal and differentiation in human embryonic stem cells, *Nat. Cell Biol.*, **13** (2011), 652–659.

5. L. Chu, J. Zhang, J. A. Thomson, et al., Single-cell RNA-seq reveals novel regulators of human embryonic stem cell differentiation to definitive endoderm, *Genome Biol.*, **17** (2016), 173.

6. S. Larabee, H. Coia, G. Gallicano, et al., miRNA-17 Members that Target Bmpr2 Influence Signaling Mechanisms Important for Embryonic Stem Cell Differentiation In Vitro and Gastrulation in Embryos, *Stem Cell. Dev.*, **24** (2015), 354–371.

7. R. A. Young, L. A. Boyer, T. I. Lee, et al., Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells, *Cell*, **122** (2005), 947–956.

8. L. W. Jeffrey, T. A. Beyer, J. L. Wrana, et al., Switch enhancers interpret TGF- and Hippo signaling to control cell fate in human embryonic stem cells, *Cell Rep.*, **5** (2013), 1611–1624.

9. J. Rossant, J. S. Draper, A. Nagy, et al., Establishment of endoderm progenitors by SOX transcription factor expression in human embryonic stem cells, *Cell Stem Cell*, **3** (2008), 182–195.

10. N. Ivanova, Z. Wang, S. Razis, et al., Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells, *Cell Stem Cell*, **10** (2012), 440–454.

11. A. F. Schier, A. Regev, D. Gennert, et al., Spatial reconstruction of single-cell gene expression data, *Nat. Biotechnol.*, **33** (2015), 495–502.

12. J. L. Rinn, C. Trapnell, T. S. Mikkelsen, et al., The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells, *Nat. Biotechnol.*, **32** (2014), 381–386.

13. J. Tan and X. Zou, Complex dynamical analysis of a coupled network from innate immune responses, *Int. J. Bifurcat. Chaos*, **23** (2013), 1350180.

14. S. Jin, D. Wang and X. Zou, Trajectory control in nonlinear networked systems and its applications to complex biological systems, *SIAM J. Appl. Math.*, **78** (2018), 629–649.

15. S. Jin, F. Wu and X. Zou, Domain control of nonlinear networked systems and applications to complex disease networks, *Discrete Cont. Dyn. B*, **22** (2017), 2169–2206.

16. X. Shu, D. Pei, S. Wei, et al., A sequential EMT-MET mechanism drives the differentiation of human embryonic stem cells towards hepatocytes, *Nat. Commun.*, **8** (2017), 15166.

17. M. A. Nieto, J. P. Thiery, R. Y. Huang, et al., Epithelial-mesenchymal transitions in development and disease, *Cell*, **139** (2009), 871–890.

18. J. P. Thiery and J. P. Sleeman, Complex networks orchestrate epithelial-mesenchymal transitions, *Nat. Rev. Mol. Cell. Biol.*, **7** (2006), 131–142.

19. H. Peinado, F. Portillo and A. Cano, Transcriptional regulation of cadherins during development and carcinogenesis, *Int. J. Dev. Biol.*, **48** (2004), 365–375.

20. A. Voulgari and A. Pintzas, Epithelial mesenchymal transition in cancer metastasis: mechanisms, markers and strategies to overcome drug resistance in the clinic, *Biochim. Biophys. Acta*, **1796** (2009), 75–90.

21. J. Zhao, X. Wang, M. Hung, et al., Krüppel-Like Factor 8 Induces Epithelial to esenchymal Transition and Epithelial Cell Invasion, *Cancer Res.*, **67** (2007), 7184–7193.

22. Y. Ma, X. Zheng, K. Chen, et al., ZEB1 promotes the progression and metastasis of cervical squamous cell carcinoma via the promotion of epithelial-mesenchymal transition, *Int. J. Clin. Exp. Pathol.*, **8** (2015), 11258–11267.

23. D. Chen, Y. Chu, S. Li, et al., Knock-down of ZEB1 inhibits the proliferation, invasion and migration of gastric cancer cells, *Chin. J. Cell. Mol. Immunol.*, **33** (2017), 1073–1078.

24. B. L. Li, J. M. Cai, F. Gao, et al., Inhibition of TBK1 attenuates radiation-induced epithelial-mesenchymal transition of A549 human lung cancer cells via activation of GSK-3β and repression of ZEB1, *Lab. Invest.*, **94** (2014), 362–370.

25. J. Comijn, G. Berx, Roy F. van, et al., The two-handed E box binding zinc finger protein SIP1 down-regulates E-cadherin and induces invasion, *Mol. Cell*, **7** (2001), 1267–1278.

26. M. Moes, E. Friederich, A. Sol, et al., A novel network integrating a mi-RNA 203/SNAI1 feedback loop which regulates epithelial to mesenchymal transition, *PloS One*, **7** (2012), e35440.

27. U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits*, 1st edition, Taylor & Francis Inc, Boca Raton, FL, 2006.

28. D. Chu, N. R. Zabet and B. Mitavskiy, Models of transcription factor binding: Sensitivity of activation functions to model assumptions, *J. Theor. Biol.*, **257** (2009), 419–429.

29. J. Kennedy and R. Eberhart, Particle swarm optimization, *Proceedings of ICNN'95*, International Conference on Neural Networks, Perth, WA, Australia, **4** (1995), 1942–1948.

30. D. Gillespie, Exact Stochastic Simulation of Coupled Chemical Reactions, *J. Phys. Chem.*, **81** (1977), 2340–2361.

31. X. Xiang, Y. Chen, X. Zou, et al., Understanding inhibition of viral proteins on type I IFN signaling pathways with modeling and optimization, *J. Theor. Biol.*, **265** (2010), 691–703.

32. S. Shin, O. Rath, K. Cho, et al., Positive- and negative-feedback regulations coordinate the dynamic behavior of the Ras-Raf-MEK-ERK signal transduction pathway, *J. Cell Sci.*, **122** (2009), 425–435.

33. T. Tian and K. Burrage, Stochastic models for regulatory networks of the genetic toggle switch, *Proc. Natl. Acad. Sci.*, **103** (2006), 8372–8377.

34. M. R. Birtwistle, J. Rauch, B. N. Kholodenko, et al., Emergence of bimodal cell population responses from the interplay between analog single-cell signaling and protein expression noise, *BMC Syst. Biol.*, **16** (2012), 109.

35. L. K. Nguyen, M. R. Birtwistle, B. N. Kholodenko, et al., Stochastic models for regulatory networks of the genetic toggle switch, *Proc. Natl. Acad. Sci.*, **103** (2006), 8372–8377.

Appendix A: Parameter sensitivity analysis

The global sensitivity of parameters reflects how the system responds to the perturbation of parameters in the model. To obtain the sensitivity of the input parameters to all variables in the model, the sensitivity function $s_j(t)$ of the parameter p_j at time t was defined as follows [31, 32]:

$$s_j = \frac{\partial O(t)}{O(t)} \bigg/ \frac{\partial P_j(t)}{P_j(t)} \approx \frac{\|O(P_j + \Delta P_j, t) - O(P_j - \Delta P_j, t)\|}{O(P_j, t)} \bigg/ \frac{2\Delta P_j}{P_j}$$

where $O(t)$ is the model output at time t (T is the total time length), ΔP_j is a small perturbation which is 10% in our situation. $S_j = \int_0^T s_j(t)\, dt$ is the sensitivity value of parameter P_j. From Figure S1, we can see that the perturbation of parameter p_2, p_5 and p_7 have relative large effects on the expression of CDH1 and KLF8.
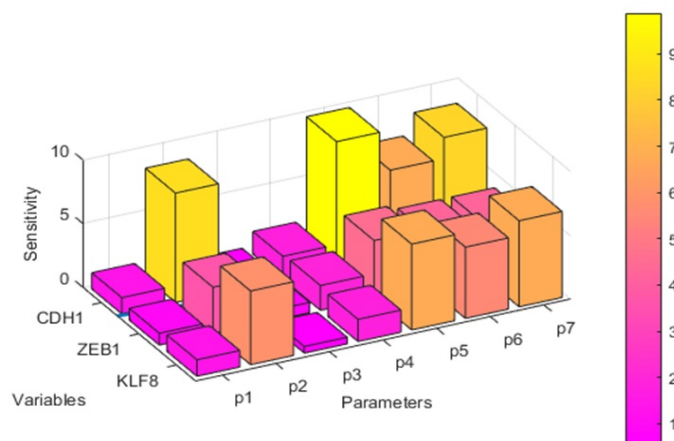
Figure S1. The sensitivity analysis to the perturbation of parameters in the model.

Appendix B: Simulation with different intensity

We simulated the dynamics of CDH1 with different weights that characterize different interactive intensity of CDH1, ZEB1 and KLF8. The variation tendency of the curve indicates that the promotion of ZEB1 by KLF8 and the inhibition of KLF8 by CDH1 is necessary for the formation of DE cells in the differentiation process of ESCs, which is marked by a low expression of CDH1.

(a)   (b)
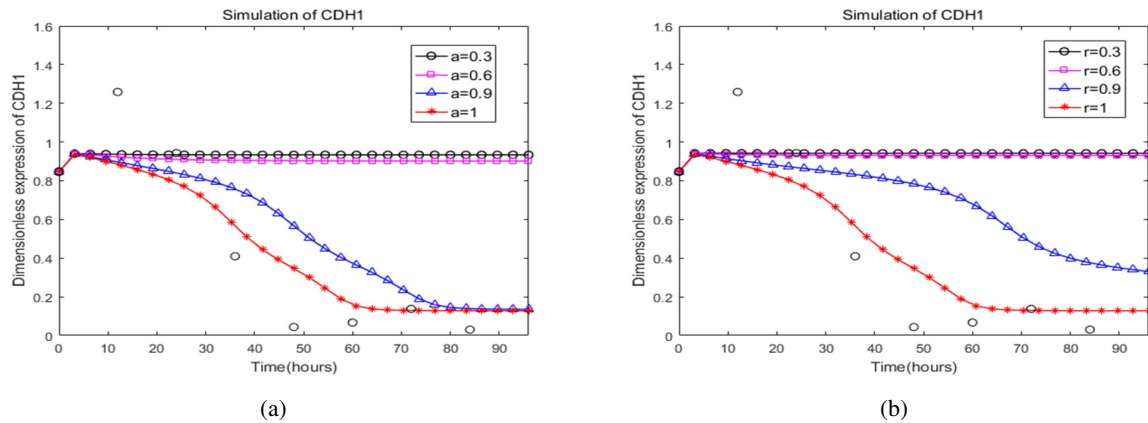
Figure S2. (a)The simulation of CDH1 under different intensity that ZEB1 is promoted by KLF8. (b)The simulation of CDH1 under different intensity that KLF8 is inhibited by CDH1.

Appendix C: Parameters

Table S1. The Optimal Parameters of the Model

Processes	Parameters	Values	Remarks
Associate constant	p_1	49.9611	Fitted
Associate constant	p_2	0.5661	Fitted
Relative dissociation constant	p_3	9.4701	Fitted
Relative degradation constant	p_4	0.0951	Fitted
Associate constant	p_5	2.3591	Fitted
Associate constant	p_6	11.9229	Fitted
Relative degradation constant	p_7	0.9982	Fitted
Hill coefficient	n_1	2	assumed
Hill coefficient	n_2	2	assumed
Hill coefficient	n_3	2	assumed
Hill coefficient	n_4	2	assumed
Intensity of impact	a	1	assumed
Intensity of impact	r	1	assumed

Appendix D: Fitting of the 758 scRNA-seq data

To clearly observe the original 758 single cell gene expression data and to thoroughly investigate the fitting of the model to the data, we plotted the following three graphs. The results indicate that the model reflects the main variation tendency of the gene expression levels. However, the simulation results do not fit the expression levels of ZEB1 and KLF8 well at approximately 90h. We assumed that the cells completed the differentiation process at that time, and the expression of ZEB1 and KLF8 were down-regulated for some reason that are not captured by the model. New regulatory genes need to be considered to study the regulatory mechanisms at the late stage of differentiation. Nevertheless, the present model can serve as a significant foundation for further investigation.

(c)

(d)

(e)
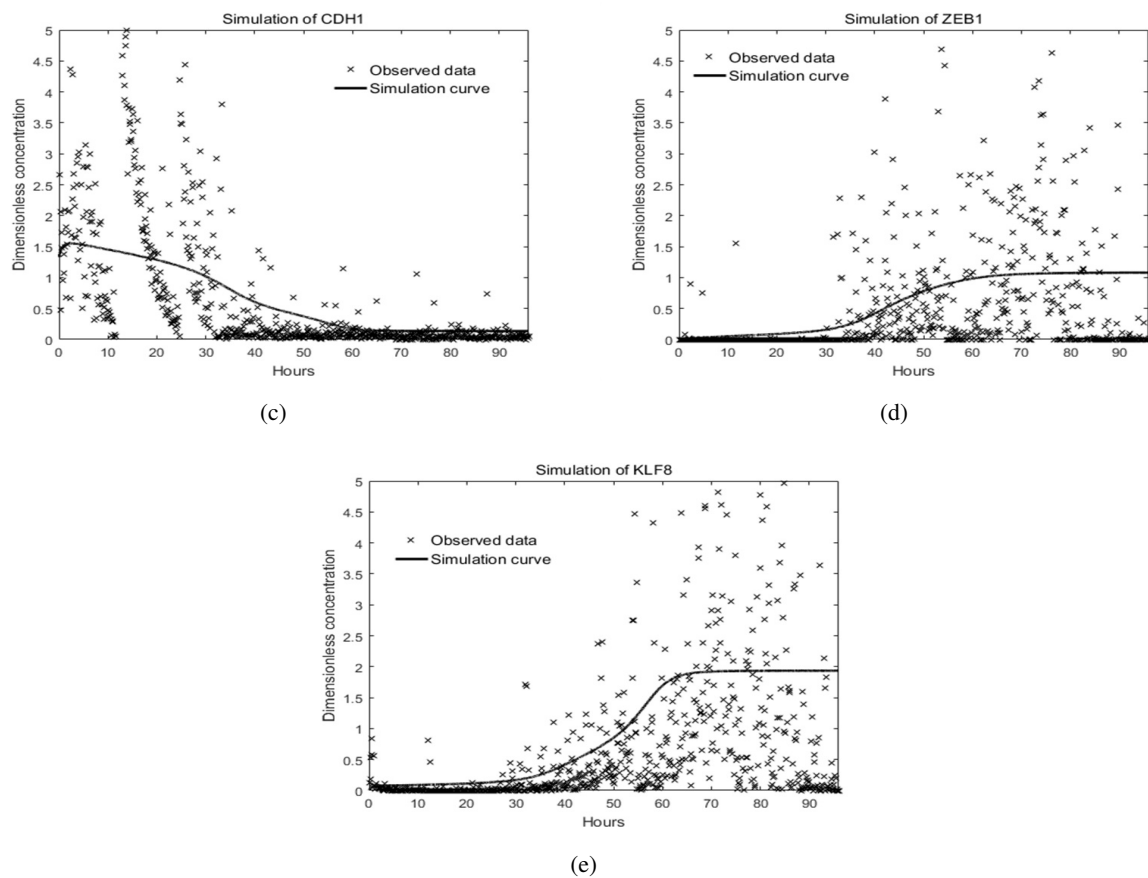
Figure S3. (c)(d)(e)The simulation of CDH1, ZEB1 and KLF8 and their uniformity with the original single cell gene expression levels.

Appendix E: 10-fold cross-validation

Through 10-fold validation, the scRNA-seq data were randomly and evenly divided into 10 parts. 9 of the data subsets were selected to estimate the parameters each time. After 10 optimizations, 10 sets of estimated parameters were listed in Table S2. The parameters estimated by 10 data subsets vary slightly and are similar to the parameters listed in Table S1.

Table S2. The Optimal Parameters of the Model.

Validation	p_1	p_2	p_3	p_4	p_5	p_6	p_7	Error
1	49.99	0.44	17.24	0.06	48.88	15.96	22.47	12.76
2	29.98	0.44	14.06	0.06	5.08	12.51	2.56	12.63
3	50.00	0.46	16.98	0.06	38.76	15.35	17.48	12.63
4	49.98	0.48	15.06	0.07	44.83	14.48	20.13	12.51
5	50.00	0.47	11.68	0.11	31.98	12.80	14.41	12.88
6	49.95	0.44	17.82	0.06	49.99	16.21	22.83	12.71
7	50.00	0.45	18.01	0.06	47.61	15.95	21.04	12.63
8	50.00	0.44	17.00	0.06	23.69	16.05	10.81	12.75
9	50.00	0.44	18.94	0.05	47.20	16.23	21.81	12.74
10	49.99	0.46	16.57	0.06	2.27	14.49	1.04	12.65
Average	47.99	0.45	16.34	0.07	34.03	15.00	15.46	12.69

Appendix F: Stochastic model and simulation

We supposed that the state of the system is $X(t_0) = x_0$ and defined the conditional probability density function $P(x, t|x_0, t_0)$, which is the probability of the system state satisfying $X(t) = x$ at time t. The corresponding Master equation is developed according to the following probability equation:

$$\frac{\partial}{\partial t} P(x, t|x_0, t_0) = \sum_{j=1}^{6} (a_j(x - v_j)P(x - v_j, t|x_0, t_0) - a_j(x)P(x, t|x_0, t_0))$$

where a_j is the propensity function corresponding to the chemical reaction channel j, and v_j is the state-change vector when reaction j occurred. For simplicity, we used the Hill function that was defined in the deterministic model to depict the propensities instead of decomposing this into a set of elementary reaction steps. Following the modeling approaches for stochastic models [33], the propensities were defined in Table S3, and the parameters were set to be the same as those that are defined in Table S1.

Table S3. The propensity function of master equation.

NO.	Reaction	Propensity function
1	$\rightarrow$ CDH1	$\Omega^3/(\Omega^2 + p_1 \times [ZEB1]^2) + p_2 \times \Omega^3/(\Omega^2 + [KLF8]^2)$
2	CDH1 $\rightarrow$	$[CDH1]$
3	$\rightarrow$ ZEB1	$\Omega \times [KLF8]^2/(\Omega^2 + p_3 \times [KLF8]^2)$
4	ZEB1 $\rightarrow$	$p_4 \times [ZEB1]$
5	$\rightarrow$ KLF8	$p_5 \times \Omega^3/(\Omega^2 + p_6 \times [CDH1]^2)$
6	KLF8 $\rightarrow$	$p_7 \times [KLF8]$

The stochastic simulation was performed using the Gillespie algorithm [30], and a steady state was obtained by dividing the system size by the molecular numbers. According to the simulation results (Figure S4), the system displayed similar dynamic behavior as that of the deterministic model.

The Langevin equations were built as follows, where ζ_i is the Gaussian white noises with $\zeta_i(t) > 0, (\zeta_i(t), \zeta_j(t')) \geq \delta_{ij}\delta(t - t')$, and similar results were obtained under a simulation of the chemical Langevin equations (Figure S5).
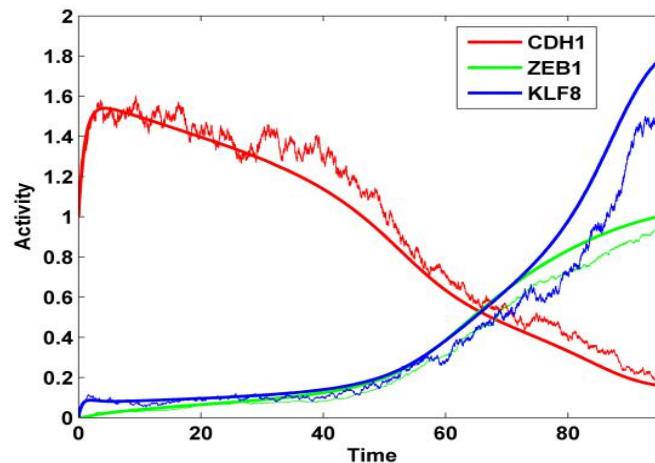
Figure S4. A comparison of the deterministic dynamic behavior with the stochastic simulation results of the Master equation. The bold line corresponds to the deterministic simulation; the lines with fluctuations correspond to the stochastic simulation results of the Master equation with the Gillespie algorithm. All of the other parameters are set to be the same as those in Table S1.

$$[CDH1](t + dt) = [CDH1](t) + \frac{1}{1 + p_1[ZEB1]^2} + \frac{p_2}{1 + [KLF8]^2} - [CDH1] +$$

$$\sqrt{\frac{1}{1 + p_1[ZEB1]^2} + \frac{p_2}{1 + [KLF8]^2}} \zeta_1(t)(dt)^{\frac{1}{2}} - [CDH1]\zeta_2(t)(dt)^{\frac{1}{2}}$$

$$[ZEB1](t + dt) = [ZEB1](t) + \frac{a[KLF8]^2}{1 + p_3[KLF8]^2} - p_4[ZEB1] + \sqrt{\frac{a[KLF8]^2}{1 + p_3[KLF8]^2}} \zeta_3(t)(dt)^{\frac{1}{2}}$$

$$- \sqrt{p_4[ZEB1]}\zeta_4(t)(dt)^{\frac{1}{2}}$$

$$[KLF8](t + dt) = [KLF8](t) + \frac{rp_5}{1 + p_6[CDH1]^2} - p_7[KLF8] + \sqrt{\frac{rp_5}{1 + p_6[CDH1]^2}} \zeta_5(t)(dt)^{\frac{1}{2}}$$

$$- \sqrt{p_7[KLF8]}\zeta_6(dt)^{\frac{1}{2}}$$

Previous studies have demonstrated that the bistable systems can generate bimodal expression patterns of corresponding genes [34, 35]. To explore how the negative feedback strength shapes the KLF8 expression distribution, we simulated the distribution of KLF8 in a collection of 10,000 cells under different values of p_6 for the Master equation. Figure S6 gives the distribution diagram of KLF8 switching from a unimodal distribution of the high state ($p_6 = 13$) via the bimodal distribution in the tristable region ($p_6 = 15$), to the bimodal distribution in the bistable region ($p_6 = 17$) and, then, to the low state with unimodal distribution ($p_6 = 22$). Since the overexpression of KLF8 could accelerate
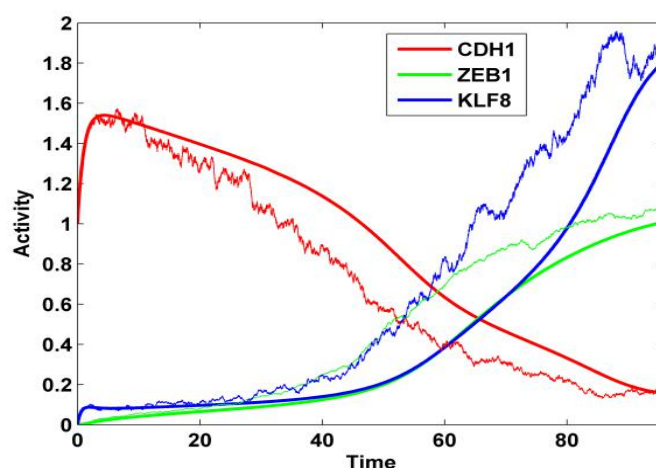
Figure S5. A comparison of the deterministic dynamic behavior with stochastic simulation results of the chemical Langevin equation. The bold line corresponds to the deterministic simulation; the lines with fluctuations correspond to the stochastic simulation results of the chemical Langevin equation with the Gillespie algorithm. All of the other parameters are set to be the same as those in Table S1.

the transition from the mesendoderm cell to DE cell under differentiation [5], the distribution of KLF8 become biomodal ($p_6 = 15$) distribution, which may correspond to the undifferentiated state or differentiate from mesendoderm to DE state. Or the cell may directional differentiate from mesendoderm state to endoderm state ($p_6 = 13$).
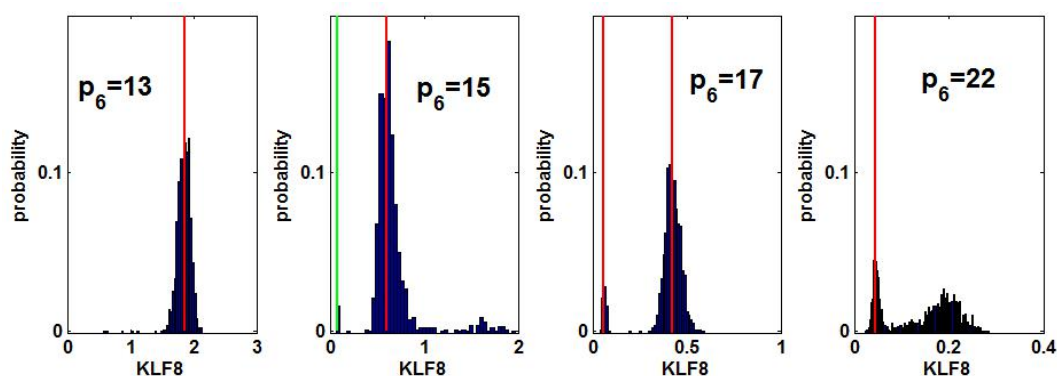
Figure S6. The distribution diagrams of KLF8 under different values of p_6 for the Master Equation. The green and red lines correspond to the low and high steady states of the deterministic model, respectively. All the other parameters are set to be the same as those in Table S1.