

# Time Scales in Epigenetic Dynamics and Phenotypic Heterogeneity of Embryonic Stem Cells

Masaki Sasai<sup>1,2,3,4\*</sup>, Yudai Kawabata<sup>2</sup>, Koh Makishi<sup>1</sup>, Kazuhito Itoh<sup>1</sup>, Tomoki P. Terada<sup>1,2</sup>

**1** Department of Computational Science and Engineering, Nagoya University, Nagoya, Japan, **2** Department of Applied Physics, Nagoya University, Nagoya, Japan, **3** School of Computational Sciences, Korea Institute for Advanced Study, Seoul, Korea, **4** Okazaki Institute for Integrative Bioscience, Okazaki, Japan

## Abstract

A remarkable feature of the self-renewing population of embryonic stem cells (ESCs) is their phenotypic heterogeneity: Nanog and other marker proteins of ESCs show large cell-to-cell variation in their expression level, which should significantly influence the differentiation process of individual cells. The molecular mechanism and biological implication of this heterogeneity, however, still remain elusive. We address this problem by constructing a model of the core gene-network of mouse ESCs. The model takes account of processes of binding/unbinding of transcription factors, formation/dissolution of transcription apparatus, and modification of histone code at each locus of genes in the network. These processes are hierarchically interrelated to each other forming the dynamical feedback loops. By simulating stochastic dynamics of this model, we show that the phenotypic heterogeneity of ESCs can be explained when the chromatin at the *Nanog* locus undergoes the large scale reorganization in formation/dissolution of transcription apparatus, which should have the timescale similar to the cell cycle period. With this slow transcriptional switching of *Nanog*, the simulated ESCs fluctuate among multiple transient states, which can trigger the differentiation into the lineage-specific cell states. From the simulated transitions among cell states, the epigenetic landscape underlying transitions is calculated. The slow *Nanog* switching gives rise to the wide basin of ESC states in the landscape. The bimodal Nanog distribution arising from the kinetic flow running through this ESC basin prevents transdifferentiation and promotes the definite decision of the cell fate. These results show that the distribution of timescales of the regulatory processes is decisively important to characterize the fluctuation of cells and their differentiation process. The analyses through the epigenetic landscape and the kinetic flow on the landscape should provide a guideline to engineer cell differentiation.

**Citation:** Sasai M, Kawabata Y, Makishi K, Itoh K, Terada TP (2013) Time Scales in Epigenetic Dynamics and Phenotypic Heterogeneity of Embryonic Stem Cells. PLoS Comput Biol 9(12): e1003380. doi:10.1371/journal.pcbi.1003380

**Editor:** Marc A. Marti-Renom, CNAG - Centre Nacional d'Anàlisi Genòmica and CRG - Centre de Regulació Genòmica, Spain

**Received:** February 6, 2013; **Accepted:** October 11, 2013; **Published:** December 12, 2013

**Copyright:** © 2013 Sasai et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by JSPS Grants-in-Aid for Scientific Research, 24244068 and 23654147. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: sasai@cse.nagoya-u.ac.jp

## Introduction

Embryonic stem cells (ESCs) are pluripotent having the ability to differentiate into a variety of lineages, while in suitable culture conditions they proliferate indefinitely by maintaining pluripotency. These self-renewing ESCs are distinguished by the marker proteins including Sox2, Oct4 and Nanog (SON) [1–4]. SON are transcription factors (TFs) which directly or indirectly promote the expression of themselves by constituting an overall positive feedback network [5–10], among which Nanog is an essential factor working as a gatekeeper for pluripotency [11,12]. Here, a remarkable feature is the large cell-to-cell variation of the level of Nanog in the self-renewing isogenic population of ESCs [13–15]. Since a distinct downregulation of Nanog is associated with the differentiation of ESCs into mesendoderm or neural ectoderm lineages [16], the heterogeneous Nanog expression can be intimately related to the process of fate decision of individual cells [14,17]. The molecular mechanism and biological implication of this phenotypic fluctuation of ESCs, however, have not yet been clarified. In this paper we address this problem by constructing a model of the regulatory network of core genes in mouse ESCs.

One can figure out, at a glance, several scenarios which may explain the phenotypic heterogeneity. A simple scenario relies on

the possible enhancement of fluctuation of the signal received by a cell: Since the reception of factors such as leukemia inhibitory factor (Lif) by a cell is stochastic, it necessarily bears fluctuation, which might be enhanced through the signal cascade to stochastically activate *Nanog* [18]. The other possible mechanism is based on the presumed self-activation of *Nanog* [6,19], which may lead to the fluctuating pulsative expression of Nanog [14,20]. With these mechanisms, however, another key factor, Oct4, should also exhibit the large fluctuation since *Oct4* is activated by the reception of Lif and the Oct4 expression is maintained through mutually activating interactions among SON. Contrary to this expectation, the observed expression of Oct4 is rather homogeneous [14,17]. A possible resolution of this inconsistency is to assume that some unknown factors which can bind to the *Oct4* locus suppress fluctuation of the Oct4 expression [20]. There has been, however, no direct experimental observation yet for the existence of such regulatory factors, and therefore, in this paper we look for the other mechanism without relying on this assumption.

For modeling the gene regulatory dynamics, not only the topological wiring diagram among genes but also the rates of reactions in the regulatory network should be quantified. These estimated rates, however, have very different values depending on the type of reactions, and hence it is strongly desired to develop the

## Author Summary

Embryonic stem cells (ESCs) can proliferate indefinitely by keeping pluripotency, i.e., the ability to differentiate into any cell-lineage. ESCs, therefore, have been the focus of intense biological and medical interests. A remarkable feature of ESCs is their phenotypic heterogeneity: ESCs show large cell-to-cell fluctuation in the expression level of *Nanog*, which is a key factor to maintain pluripotency. Since *Nanog* regulates many genes in ESCs, this fluctuation should seriously affect individual cells when they start differentiation. In this paper we analyze this phenotypic fluctuation by simulating the stochastic dynamics of gene network in ESCs. The model takes account of the mutually interrelated processes of gene regulation such as binding/unbinding of transcription factors, formation/dissolution of transcription apparatus, and histone-code modification. We show the distribution of timescales of these processes is decisively important to characterize the dynamical behavior of the gene network, and that the slow formation/dissolution of transcription apparatus at the *Nanog* locus explains the observed large fluctuation of ESCs. The epigenetic landscapes are calculated based on the stochastic simulation, and the role of the phenotypic fluctuation in the differentiation process is analyzed through the landscape picture.

theoretical framework to treat effects of coexistence of the distributed timescales [21,22]. In simple bacterial cells, for example, the DNA-protein binding/unbinding is often much faster than the protein-copy number change, so that the fast DNA-state change can be regarded as equilibrated and the dynamical interference between the fluctuation of gene switching and the fluctuation of protein-copy number can be neglected. By borrowing the wording from condensed-matter physics, this separation of fast and slow processes should be referred to as the “adiabatic” separation. Theoretical studies have shown that when the adiabatic limit is not the case, the kinetic flow of the coupled stochastic dynamics of gene switching and protein-copy number change is described as “eddy current” [23], which gives rise to a variety of unexpected dynamical effects in gene regulation [23–28]. Indeed, it has been suggested that the transition of *Bacillus subtilis* into the competence period should be due to the non-adiabatic gene switching in the excitatory self-activating gene network [29]. In eukaryotic cells, processes of gene switching are much more complex including the assembly of transcriptional apparatus (TA) [30–33], the transition from the poised state of TA to the elongation state [34], chemical modifications of nucleosomes [35–39], and the structural reorganization of chromosomes [40–42]. Such epigenetic change of the gene state can be much slower than the bacterial DNA state change, and their timescales are often comparable with or longer than the timescale of the protein-copy number change, so that the non-adiabatic effects should play significant roles in eukaryotic cells.

Many marker genes of ESCs have been identified [43], among which SON regulate many other genes, and hence, the SON network has been regarded as the central network to maintain pluripotency [8,9,43]. Models of the core SON network of ESCs have been developed [14,16,20,43,44], but all of these models have been based on the assumption that the gene state is determined by the fast equilibrated binding/unbinding of TF to/from the gene locus: The assumption of the adiabatic limit has been adopted in all the previous models and the slow non-adiabatic switching dynamics has not been explicitly taken into account. In this paper, we discuss ESCs by focusing on the non-

adiabatic effects, the effects of slow epigenetic processes, and we propose a hypothesis that the non-adiabatic switching in the core gene-network explains the large fluctuation of *Nanog* expression. By using the landscape picture, we discuss the roles of this non-adiabatic switching in the cell-fate decision of ESCs.

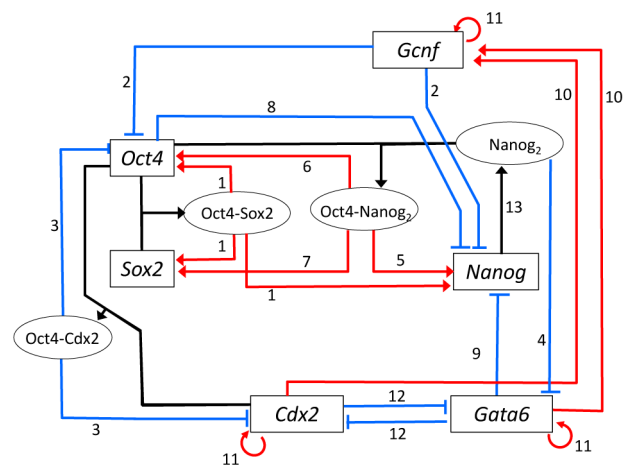
## Results

Before starting the explanation of the simulated results, we briefly explain the interaction network among genes considered in the present model and discuss the dynamics of each gene in the network in subsection *Gene network and epigenetic dynamics*. Coexistence of multiple timescales in the eukaryotic gene dynamics is the focus of the present study.

### Gene network and epigenetic dynamics

**The interaction network.** Interactions among genes have been inferred from observations on how the expressions of genes are correlated with each other and how one factor binds to the genetic locus of the other factor. It is not straightforward, however, to identify each elementary interaction embedded in the complex web of regulation. Indeed, consensus has not been obtained on the role of *Nanog*: Though both *Sox2* and *Oct4* loci have the *Nanog* binding sites [6] and the assumption of positive regulation of *Sox2* and *Oct4* by *Nanog* is reasonable [5,8], the weak correlation of levels of *Sox2* and *Oct4* to the heterogeneous *Nanog* expression has cast doubts on the direct positive regulation by *Nanog* [11,45]. The network model of Fig. 1 is based on the assumptions that *Nanog* directly activates *Sox2* and *Oct4*. We show that even with such assumptions, the weak correlation of *Sox2* and *Oct4* to the heterogeneous *Nanog* expression can be explained when epigenetic dynamics is considered explicitly.

In the model of Fig. 1, the representative lineage-trigger genes are also taken into account. Mouse ESCs can directly differentiate into either of trophoblast, primitive endoderm, or primitive ectoderm [10]. Though majority of ESCs differentiate into primitive ectoderm under the condition of the minimal external



**Figure 1. The network model.** Interactions among marker genes of ESCs, *Sox2*, *Oct4*, and *Nanog*, and lineage-specific genes, *Gata6*, *Cdx2*, and *Gcnf* are shown with black lines for formation of complexes of factors synthesized from the corresponding genes, red lines for activating the target genes, and blue lines for repressing the target genes. Experimental data used to infer the interactions are designated by numbers on lines, which are explained with the corresponding numbering in subsection *Inference of the network from experimental observations* in section *Methods*.

doi:10.1371/journal.pcbi.1003380.g001

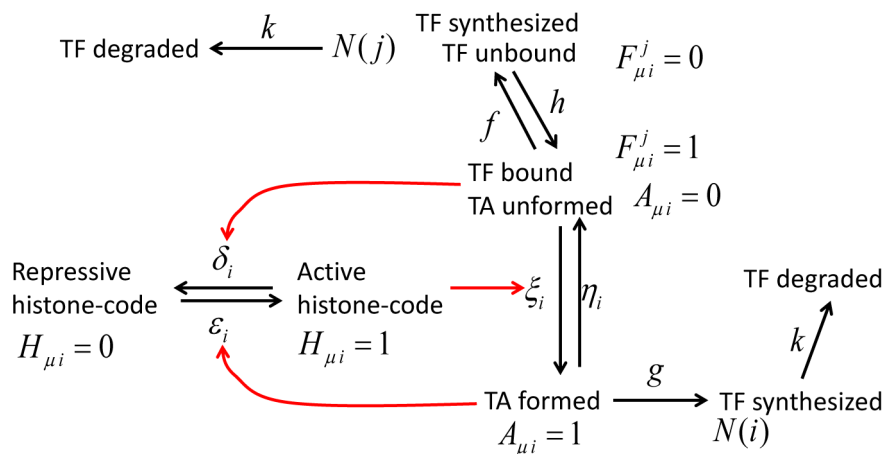
stimuli [46], the core genes to guide ESCs to primitive ectoderm are still elusive, and hence in this paper, in order to analyze the effects of non-adiabaticity on differentiation, we focus on the rest two routes to trophoctoderm and primitive endoderm. Lineage specific factors, *Cdx2* for trophoctoderm and *Gata6* for primitive endoderm, repress expression of SON [1,2,47] by their direct binding to the enhancers [10,47] or the indirect action through *Gcnf* [48], while SON repress *Cdx2* and *Gata6* [13,47,49]. The network of Fig. 1 resembles those proposed in [10] and analyzed in [50]. See subsection *Inference of the network from experimental observations* in section *Methods* for the explanation of the experimental data used to infer the network.

**Epigenetic dynamics.** Shown in Fig. 2 is a sketch of how each gene is regulated through epigenetic dynamics in the model: Changes in the gene activity are triggered by binding/unbinding of TF to/from DNA. We write  $F_{\mu i}^j = 1$  ( $F_{\mu i}^j = 0$ ) when the  $j$  th TF is bound on (unbound from) the regulatory region of DNA at the locus of allele  $\mu$  of gene  $i$ , where each gene consists of two alleles  $\mu = \alpha$  and  $\beta$ . When the activator TF binds to the locus, the subsequent assembly of other molecules should take place to form a TA. The TA becomes ready for transcription when the chromatin structure is altered to accommodate the assembled factors and those factors are chemically suitably modified. For simplicity of description, we refer to the combined multiple processes to make TA active as “TA formation” and the processes to make TA inactive as “TA dissolution”. We write  $A_{\mu i} = 1$  when the TA is formed and becomes ready for transcription, and  $A_{\mu i} = 0$  when the TA is dissolved and not ready for transcription. Changes in  $A_{\mu i}$  are regulated not only by the direct interactions among TFs and other factors on the gene locus but also by methylation or acetylation of nucleosomes at around  $\mu i$ . Since the ensemble of nucleosomes are methylated or acetylated cooperatively to form a bistable switch [35,39], we write the state of the ensemble of nucleosomes, i.e., the histone code, as  $H_{\mu i} = 1$  when the histone code promotes formation of the active TA, and  $H_{\mu i} = 0$  when it is inhibitory from transcription. It has been recently observed that only a single allele of *Nanog* is active in ESCs cultivated with Lif though two alleles can be active for other genes [51], which

indicates that the *Nanog* expression is regulated through the large scale organization of chromosome architecture. This allelic regulation is represented in the model by fixing the values as  $F_{\beta i}^j = A_{\beta i} = H_{\beta i} = 0$  and only the dynamical change of the allele  $\alpha$  is considered for  $i = \textit{Nanog}$ .

The rate of binding of the  $j$  th TF (the rate to turn  $F_{\mu i}^j$  from 0 to 1),  $h$ , depends on the copy number of the  $j$  th TF,  $N(j)$ . The rate to change  $N(j)$  depends on the rate of synthesis of the TF,  $g$ , the average burst size,  $X$  (see Eq.1 for the definition of burst synthesis of proteins), and the degradation rate of the TF,  $kN(j)$ , where the synthesis rate  $g$  depends on whether the TA is ready for transcription ( $A_{\mu i} = 1$ ) or not ( $A_{\mu i} = 0$ ). Rates of the TA formation/dissolution (rates to change  $A_{\mu i}$ ) depend on the TF binding status  $F_{\mu i}^j$  and the histone code  $H_{\mu i}$ , and rates to change  $H_{\mu i}$  should depend on  $F_{\mu i}^j$  and  $A_{\mu i}$  because the enzymes to rewrite the histone code are recruited by the TF bound on the locus. In this way, the model has the hierarchically interrelated processes to change variables,  $F_{\mu i}^j$ ,  $A_{\mu i}$ ,  $H_{\mu i}$ , and  $N(i)$ .

These different processes should have different timescales;  $\tau_N$  for the protein-copy-number change,  $\tau_F$  for binding/unbinding of TF,  $\tau_A$  for formation/dissolution of a TA on chromatin, and  $\tau_H$  for modification of the histone code. Their precise values are not known, but the histone code is often heritable across generations of cell cycle [35] and modification of the histone code is most frequent at the late phase of mitosis [52,53]. Therefore, when the DNA methylation does not inhibit the histone-code modification,  $\tau_H$  should be larger than or near to the period of cell cycle,  $\tau_{cc}$ , as  $\tau_H > \tau_{cc} \approx 4 \times 10^4$  sec. When DNA is methylated, on the other hand, the timescale to make the histone code active (the timescale to turn  $H_{\mu i}$  from 0 to 1) could be much longer as  $\tau_H \gg \tau_{cc}$ , so that in general,  $\tau_H$  depends on the chemical state of DNA of each gene as  $\tau_H(i)$ . Since methylation of DNA is erased after fertilization, we can expect that DNA is not yet methylated at most of the gene loci in the early phase of differentiation we treat in the model. After the *de novo* methylation in the developmental process, the distribution pattern of methylated regions of DNA often lasts through the lifetime of the organism, showing its timescale is very long [54].



**Figure 2. A sketch of the epigenetic switching in the model.** The state of the allele  $\mu$  of gene  $i$  is described by  $F_{\mu i}^j$ ,  $A_{\mu i}$ , and  $H_{\mu i}$ . Binding of TF ( $F_{\mu i}^j = 1$ ) triggers the formation of TA ( $A_{\mu i} = 1$ ), which enables synthesis of the  $i$  th TF to increase its copy number  $N(i)$ . Rates of reactions,  $f$ ,  $h$ ,  $\xi_i$ ,  $\eta_i$ ,  $\epsilon_i$ ,  $\delta_i$ ,  $g$ , and  $k$ , are written on the corresponding arrows.  $\xi_i$  is large when the histone code is active for transcription ( $H_{\mu i} = 1$ ),  $\delta_i$  is large when the bound repressor at  $\mu i$  recruits the enzymes to rewrite the histone code, and  $\epsilon_i$  is large when the bound activator TF recruits the necessary enzymes (red arrows). The adiabaticity parameters are defined by  $\omega_F = f/k$ ,  $\omega_A(i) = \eta_i/k$ , and  $\omega_H(i) = \delta_i/k$ . See subsection *Reactions in the model* in section *Methods* for the precise definition of reactions and rates and *Parameters in Methods* for the values of rate constants.  
doi:10.1371/journal.pcbi.1003380.g002

Therefore, we here do not consider the dynamics of DNA methylation explicitly, but treat it as a given static condition in our model.

From the observed rate of decrease of the amount of Oct4 [4] or Nanog [16],  $\tau_N$  can be estimated as  $\tau_N \approx 1 \sim 2 \times 10^4$  sec, which indicates  $\tau_N < \tau_{cc}$ . In many eukaryotic cells, the fast binding/unbinding of TF is necessary for cells to respond to fast environmental fluctuations, and hence we assume that also in ESCs its timescale  $\tau_F$  is much shorter than the timescale of protein-copy number change  $\tau_N$  as  $\tau_F \ll \tau_N < \tau_{cc}$ . The timescale of TA formation/dissolution ( $\tau_A$ ) is not known but we assume that the timescale is correlated with complexity of the process. Building a TA with DNA looping and recruitment of molecules is more complex than the TF binding, so that we assume  $\tau_F \ll \tau_A \leq \tau_{cc}$ . For the simple cases, we expect  $\tau_A < \tau_N$ , but when the large scale chromatin modification or the chromosome reorganization is necessary for building the TA,  $\tau_A$  may become as large as  $\tau_A \approx \tau_{cc}$ . We write  $\tau_A$  as  $\tau_A(i)$  to emphasize its dependence on the type of gene.

The turnover time of TF binding/unbinding is  $\tau_F \sim 1/h + 1/f$ , where  $h$  is the rate of binding and  $f$  is the rate of unbinding. The turnover time of TA formation/dissolution is  $\tau_A(i) \sim 1/\xi_i + 1/\eta_i$ , where  $\xi_i$  is the rate of formation and  $\eta_i$  is the rate of dissolution. The turnover time of the histone-code modification is  $\tau_H(i) \sim 1/\varepsilon_i + 1/\delta_i$ , where  $\varepsilon_i$  is the rate to turn the histone code positive for transcription and  $\delta_i$  is the rate to turn the histone code repressive for transcription. We can expect in many cases that the balance is achieved between forward and backward processes for the dynamically fluctuating phase of the early differentiation, or in other words,  $h \sim f$ ,  $\xi_i \sim \eta_i$ , and  $\varepsilon_i \sim \delta_i$ . We therefore approximate times for turnover as  $\tau_F \sim 1/f$ ,  $\tau_A(i) \sim 1/\eta_i$ , and  $\tau_H(i) \sim 1/\delta_i$ . For  $i = \text{Nanog}$ , as will be discussed later in *Results* section, we have  $\xi_i > \eta_i$  for the parameterization quantitatively consistent with the experimentally observed data, and hence we use the relation  $\tau_A(i) \sim 1/\xi_i + 1/\eta_i \sim 1/\eta_i$  also for the *Nanog* locus.

Here, we define ‘‘adiabaticity’’ as the ratio of the rate of the process to the rate of protein copy-number change; adiabaticity is larger (smaller) than 1 when the process is adiabatic (non-adiabatic). We use three sets of adiabaticity parameters,  $\omega_F$ ,  $\omega_A(i)$ , and  $\omega_H(i)$ . Using the rate constant of the protein degradation  $k \sim \tau_N^{-1}$  as a measure, the rate of unbinding of TF from DNA,  $f \sim \tau_F^{-1}$ , should be larger than  $k$ , so that the adiabaticity of TF binding/unbinding is  $\omega_F = f/k \sim \tau_N/\tau_F \gg 1$ . The rate to change the histone code from active to repressive state,  $\delta_i \sim \tau_H(i)^{-1}$ , leads to the smaller adiabaticity as  $\omega_H(i) = \delta_i/k < 1$ . The rate of dissolution of TA should depend on the gene as  $\eta_i \sim \tau_A(i)^{-1}$ . We expect that the corresponding adiabaticity parameter is  $\omega_A(i) = \eta_i/k > 1$  when the large scale reorganization of chromosome is not necessary, but  $\omega_A(i) < 1$  otherwise. Thus,  $\omega_A(i)$  is a key parameter to determine the dynamical features of the whole switching process of the gene.

As in the previous works [29,30], we do not explicitly consider mRNA and treat transcription and translation as one combined process. In the same spirit, we do not explicitly consider processes of the post-translational modification of TF, the transport of TF, or the actions of micro RNA. Though regulations through these processes can indeed affect the noise level [55–57], we here focus on the transcriptional regulations to clarify the effects of non-adiabatic gene switching. With this simplification, the states of six genes in Fig. 1 are described by  $\{F_{\mu i}^j\}$ ,  $\{A_{\mu i}\}$ , and  $\{H_{\mu i}\}$ , and the coupled stochastic dynamics of genes and  $\{N(i)\}$  is simulated by using the Gillespie algorithm [58]. In this way, the present model has the resolution intermediate between the simplistic Boolean

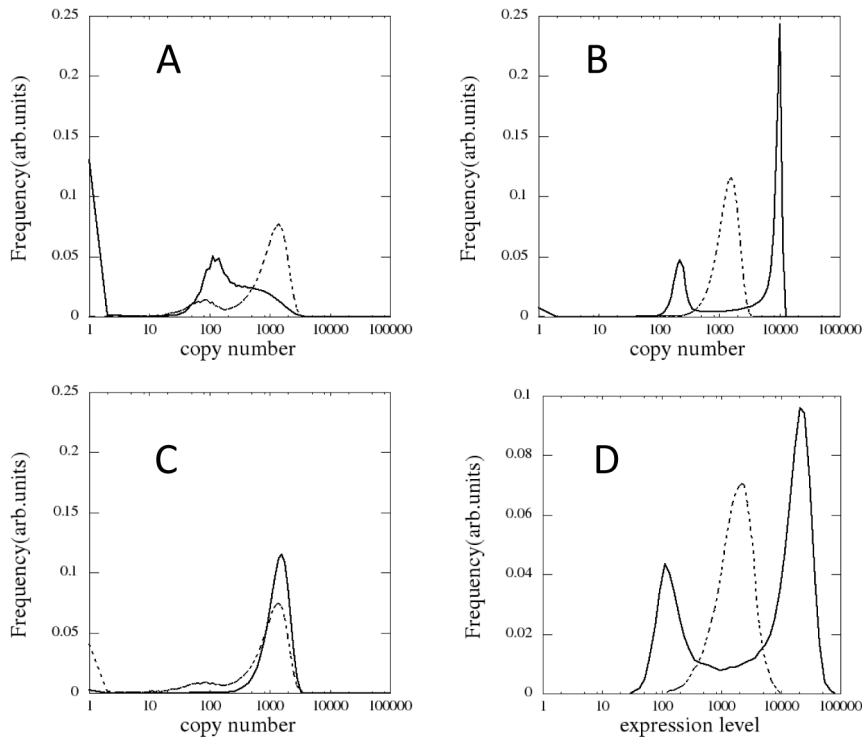
models [59] and the models which integrate the further detailed molecular processes. Through this simplified modeling and simulations, we propose a hypothesis that the hierarchically designed adiabaticity,  $\omega_F \gg \omega_A(i) > 1 > \omega_H(i)$  or  $\omega_F \gg 1 > \omega_A(i) > \omega_H(i)$ , decisively affects the self-renewal of ESCs and differentiation. In *Parameters* subsection of *Methods* section of this paper, values of  $\omega_F$  and  $\omega_H(i)$  are estimated by referring the experimental data, but the value of  $\omega_A(i)$  is largely undetermined. In the following part of *Results* section, we focus on the effects of varying  $\omega_A(i)$  on cell dynamics.

## Phenotypic heterogeneity

We first discuss ESCs in media containing Lif and other agents. Lif activates *c-Myc* [60], which activates SON by keeping the histone code of lineage-specific genes repressive [34]. We simulate this culture by adopting the null rate for turning the histone-code active as  $\varepsilon_i = 0$  for  $i = \text{Gata6}$ , *Cdx2* and *Genf* (See Fig. 2 and *Methods* for the definition of parameters).

First simulated is the case that the formation/dissolution of TA is adiabatic with  $\omega_F \gg \omega_A(i) > 1$  with  $\omega_F = 10^3$ . As a typical value to satisfy this inequality, we use  $\omega_A(i) = 10$  for all  $i$ . Distributions of the expression level of SON simulated with this parameterization are shown in Figs. 3A and 3C. We can see that the simulated distribution of the expression level of Nanog shows a single peak and the simulated distributions of Sox2 and Oct4 are double peaked at their finite values of expression level with some additional populations at the zero expression. These features are different from those observed in experiments: Compared are the distributions of cell population in a culture plotted as functions of the expression level of SON. The observed distribution of Oct4 is single peaked (Fig. 3D) [14], the distribution of Sox2 is similar to that of Oct4 [14], and the observed distribution of Nanog shows two peaks (Fig. 3D) [11,14]. The observed two-peak distribution of Nanog indicates that the fluctuation of Nanog is dominated by transitions between two states; the high-Nanog (HN) and low-Nanog (LN) states [11,33]. The simulated Nanog distribution with the adiabatic TA formation/dissolution apparently disagrees with this observed two-peaked Nanog distribution.

The assumption of the adiabatic TA formation/dissolution with  $\omega_A(\text{Nanog}) > 1$  used in the above simulation is questionable when we consider the following features of Nanog expression: First, the TA of Nanog consists of the fairly large ( $\approx 160$  kb) region of DNA [61], which should make the formation/dissolution of TA a rather complex process. Second, the allelic regulation of Nanog [51] indicates that the chromosome organization on the nuclear scale regulates the Nanog expression. This observation is also consistent with the recent finding that the loci of genes of the pluripotent factors are spatially in proximity to the *Nanog* locus in an ESC-specific manner [62], indicating that the nuclear scale organization of chromosomes is involved in the activation of *Nanog* in ESCs. For such complex and spatially extended processes for TA at the *Nanog* locus, it should be reasonable to assume that the timescale of TA formation/dissolution is as long as the cell cycle period. To find the plausible values for the rate of TA formation ( $\xi_i$ ) and the rate of TA dissolution ( $\eta_i$ ) at the  $i = \text{Nanog}$  locus, we performed a massive parameter search by generating more than 1,000 scattered points on the two-dimensional plane of  $\log_{10}(\xi_i/k)$  and  $\log_{10}(\eta_i/k)$  with  $i = \text{Nanog}$ . The score for each of generated parameter sets was calculated by averaging 10,000 trajectories simulated with the corresponding parameter set, where the score is the number of the experimentally observed features that the simulated data reproduced. The features used to count the score include (1) bimodality of the distribution of expression level of Nanog, (2) the ratio of the copy-number at the HN state to that at

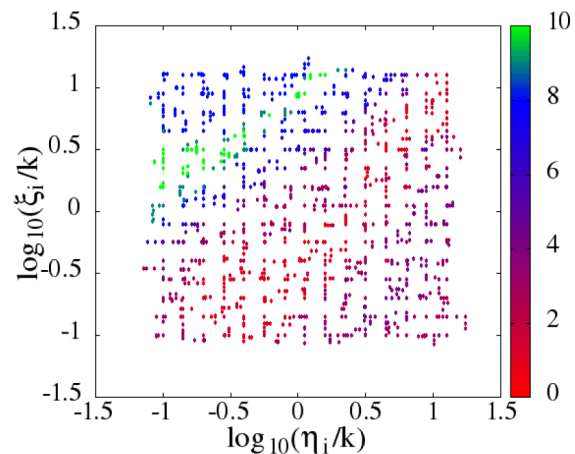


**Figure 3. Distributions of the expression level of SON in the isogenic population of ESCs.** Distributions of the simulated copy number of (A) Nanog (real line) and Oct4 (dashed line) with fast *Nanog* switching, (B) Nanog (real line) and Oct4 (dashed line) with the slow *Nanog* switching, and (C) *Sox2* with the slow (real line) and fast (dashed line) *Nanog* switching. (D) The observed population of cells in a culture as a function of their expression level (in arbitrary units) of Nanog (real line, taken from Fig.S3 of [14]) and that of Oct4 (dashed line, modified from Fig. 6A of [14]). (A–C) Simulated by sampling 10,000 cells during 11.5 days with  $\omega_A(i)=10$  for all  $i$  for the fast *Nanog* switching, and with  $\omega_A(\text{Nanog})=0.5$  and  $\omega_A(\text{Sox2})=\omega_A(\text{Oct4})=10$  for the slow *Nanog* switching. For convenience of the graphical representation, the populations at the zero expression level ( $N(i)=0$ ) are binned with the data of  $N(i)=1$ . doi:10.1371/journal.pcbi.1003380.g003

the LN state, (3) the ratio of the peak height at the HN state to that at the LN state, (4) the single-peak distribution of expression level of Oct4, and so on. The score calculated in this way is plotted in Fig. 4 for 1,125 parameter sets. See *Massive parameter search* subsection in *Methods* section for more details on the definition of the score. Search of the other parameter set is shown in *Fig.S1*.

Results of Fig. 4 indicate that the normalized rate of TA formation  $\xi_i/k$  should be around  $1 < \xi_i/k < 5$  and the normalized rate of TA dissolution  $\eta_i/k$  should be  $0.1 < \eta_i/k < 1$  for  $i = \text{Nanog}$  to reproduce the experimentally observed heterogeneous expression levels in ESCs. Here, the result of  $\xi_i/k > \eta_i/k$  was needed to reproduce the observed feature that the HN peak height is larger than the LN peak height. Since the biologically reasonable lower bound of  $\eta_i$  is the frequency of cell cycle  $1/\tau_{cc}$ , we use the lowest allowed value of  $\omega_A(\text{Nanog}) = \eta_i/k = (1/\tau_{cc})/k = 0.5$  in the subsequent analyses by keeping  $\omega_A(i) = 10$  for  $i = \text{Sox2}$ , and *Oct4*. The precise values of other parameters are explained in *Parameters* subsection in *Methods* section. The simulated distributions of SON with this non-adiabatic switching of *Nanog* are shown in Figs. 3B and 3C. The simulated width of peaks is narrower than the observed one because in simulation the extrinsic noise due to the cell-cycle oscillation and the fluctuating reception of *Lif* are neglected for simplicity. The overall features of the distributions, however, agree well with the experimental data [14]: Nanog shows a clear two-peak distribution and the Oct4 distribution has an asymmetric single major peak.

Shown in Figs. 5A and 5B is the temporal change of distributions of Nanog calculated by starting from the ensemble



**Figure 4. Search for a range of parameter set  $\xi_i$  and  $\eta_i$  for  $i = \text{Nanog}$ .** The ability of the simulated trajectories to reproduce the experimentally observed data of distributions of the expression level of SON is evaluated by the score  $S$  which is defined by Eq.20 in *Methods* section. The score was evaluated for each of 1,125 parameter sets scattered on the two-dimensional plane of  $\log_{10} \eta_i/k$  and  $\log_{10} \xi_{ia}/k$  for  $i = \text{Nanog}$ , where  $\xi_{ia}$  is a conditional value of  $\xi_i$  explained in *Methods* section and  $\xi_i$  is the rate of TA formation.  $\eta_i$  is the rate of TA dissolution, and  $k$  is the rate coefficient of protein degradation. doi:10.1371/journal.pcbi.1003380.g004



of cells either in the HN or LN state at Day 0. Within several days, the single-peaked distribution of cells in either of the HN or LN state recovers the two-peak features, which reproduces the experimentally observed temporal relaxation [11,14]. This relaxation indicates that ESCs show dynamical transitions between HN and LN states with timescale of a few days. The agreement between the observed and simulated timescales of transitions between HN and LN states indicates the validity of the small  $\omega_A(Nanog)$  for the slow switching at the *Nanog* locus, and hence the data in Fig. 5 should rule out the other hypothetical models which can yield a bimodal *Nanog* distribution but with the large  $\omega_A(Nanog)$ .

A possible origin of the slow non-adiabatic switching of *Nanog* is the large scale chromatin reorganization in the formation/dissolution of TA of *Nanog*. This assumption of slow switching explains the observed two-peak distribution and the dynamical transition of the expression level of *Nanog*, and is also consistent with the single-peak distribution of Oct4. Thus, the assumption of the slow non-adiabatic switching of *Nanog* explains the observed phenotypic heterogeneity of ESCs.

### Diagram of transitions among cell states

Given the consistent model for the heterogeneity of ESCs, it is interesting to analyze how cells initiate differentiation. To simulate cells that can differentiate, the rate to turn the histone-code active,  $\varepsilon_i$ , is increased to have a finite value for  $i = Gata6$  and  $Gcnf$ .  $\varepsilon_i$  for  $i = Cdx2$  is also turned finite but kept small because in embryo, the distinct expression of *Cdx2* is the event prior to the formation of inner cell mass from which ESCs are prepared, so that it is plausible to assume that the methylated DNA or the collective action of regulating factors inhibits the histone code of *Cdx2* from being active in ESCs (See subsection *Parameters* in *Methods*).

Examples of trajectory simulated with this parametrization are shown in Figs. 6A and 6B. The trajectory in Fig. 6A wanders around several transient states but neither *Cdx2* nor *Gata6* dominates during this wandering: Cells are jumping among the states by maintaining the features of ESC. In Fig. 6B, on the other hand, the trajectory escapes from the ESC states to reach the *Gata6* dominant state which is a gateway to the primitive endoderm lineage.

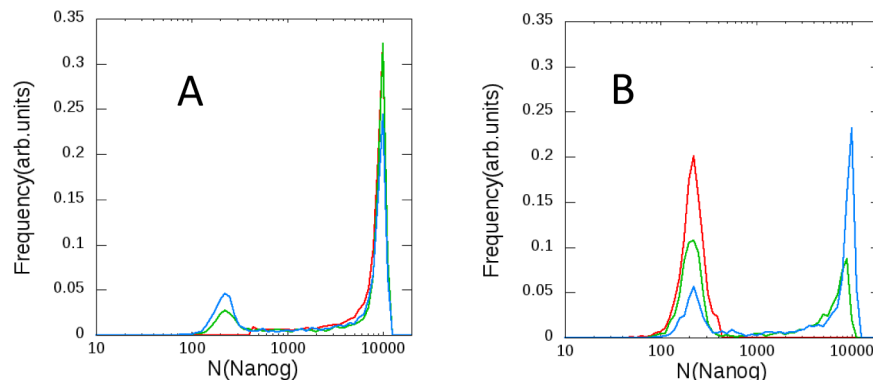
In both Figs. 6A and 6B, the trajectories are not the continuous drifts but consist of sojourns and jumps. This feature allows us to represent each trajectory as a sequence of transitions among “cell states”: Using the feature that the copy number of each factor,

$N(i)$ , shows a multiple-peak distribution (Fig.S2), we divide each distribution into a few parts, each of which is named in an abbreviated way as HN (high-level *Nanog*), MN(middle-level *Nanog*), LN (low-level *Nanog*), S (high-level *Sox2*), LS (low-level *Sox2*), etc. The thresholds used to divide the distributions are summarized in Table 1. Then, cell states are defined by thus discretized distributions and also by a set of the histone states  $\{H_{\mu i}\}$ . The trajectory is regarded as a sequence of transitions among those cell states. With this coarse-grained representation of trajectories, the mean waiting time for transition from the  $i$ th to the  $j$ th cell states can be estimated as  $\bar{\tau}_{i \rightarrow j}$ , and the mean transition rate is defined by  $q_{i \rightarrow j} = 1/\bar{\tau}_{i \rightarrow j}$  (See subsection *Transition diagram* in *Methods* for the detailed explanation on  $q_{i \rightarrow j}$ ).

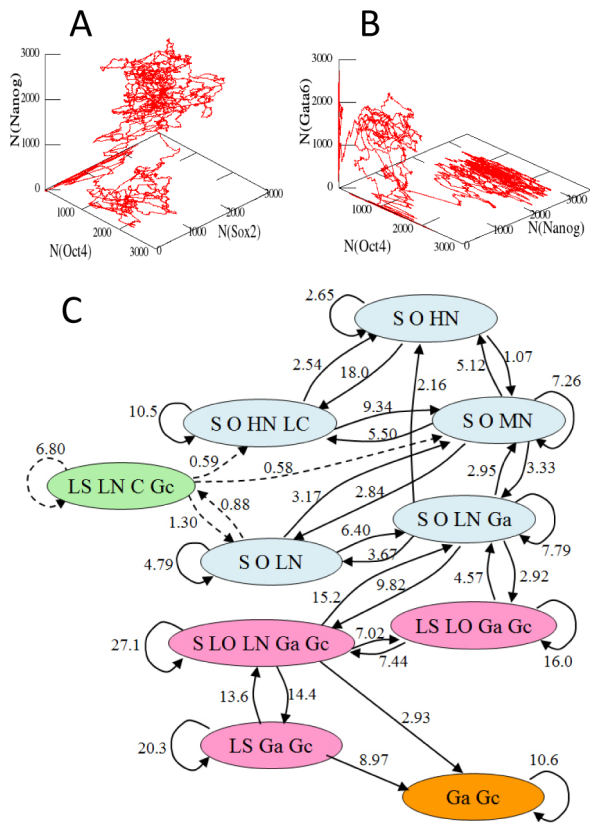
In the case that the trajectory stays for a long duration at each cell state to erase its dynamical memory, this coarse-grained dynamics can be regarded as Markovian, or in other words, the transition probability from the  $i$ th to  $j$ th states is not affected by which state the trajectory visited before reaching the  $i$ th state. It is suggested from Figs. 6A and 6B that the trajectories stay at each cell state long enough to show many oscillations during the stay, but the more quantitative test should be necessary to judge whether the coarse-grained dynamics is indeed Markovian or not. We leave this test as a future problem and proceed further in this paper to show how the transition diagram and the landscape view capture the important features of transitions among cell states.

Drawn in Fig. 6C is the diagram of transitions among thus defined cell states, where the value of  $q_{i \rightarrow j}$  is shown on the link from the  $i$ th to  $j$ th cell states. In Fig. 6C the cell states in which all of *Sox2*, *Oct4*, and *Nanog* (SON) are expressed are regarded as the pluripotent states (or the ESC states) though the level of *Nanog* fluctuates largely among these states and sometimes *Cdx2* or *Gata6* coexists with SON. These ESC states are connected by loops of transitions and hence the cells wander among ESC states to wait for a chance to escape from the ESC states. Trajectories that have escaped from the ESC states go through the network of transitions among the intermediate states in which one or two of SON are lacking. From these intermediate states, cells reach the state in which *Gata6* dominates. In some cell states, *Cdx2* appears as fluctuation but the small value of  $\varepsilon_{cdx2}$  prevents *Cdx2* from dominating the state.

It should be interesting to examine the validity of these predictions with the experimental observations: By quantifying the expression level of important factors, we will be able to define cell states from the experimental data. Then, we can check



**Figure 5. The simulated temporal change of distributions of the expression level of *Nanog*.** Distributions starting from (A) the ensemble of 10,000 cells in the HN state, and (B) from those in the LN state. The distribution at the initial time (red), 1 day later (green), and 5 days later (blue).  $\omega_A(Nanog) = 0.5$  and  $\omega_A(Sox2) = \omega_A(Oct4) = 10$ . doi:10.1371/journal.pcbi.1003380.g005



**Figure 6. Simulated trajectory of differentiation can be regarded as a sequence of transitions.** Example trajectories of 11.5 day length drawn with coordinates of (A)  $N(\text{Nanog})$ ,  $N(\text{Oct4})$ , and  $N(\text{Sox2})$ , and (B)  $N(\text{Gata6})$ ,  $N(\text{Oct4})$ , and  $N(\text{Nanog})$ . (C) Diagram of transitions among cell states. Cells wandering among ESC states (light blue) start differentiation through intermediate states (pink) or the  $\text{Cdx2}$  expressing state (light green) to reach the  $\text{Gata6}$  dominant state (orange), which leads to the primitive endoderm lineage. The transition rate  $q_{i \rightarrow j}$  is written on the link connecting the  $i$ th to  $j$ th cell states in units of 1/day. Arrows returning to the same states represent transitions via the infrequently appearing states which are not shown in the diagram according to the criterion described in the subsection *Transition diagram* in *Method* section. Each state is designated by the expressed factors; S (Sox2), LS (Low-level Sox2), O (Oct4) LO (Low-level Oct4), HN (High-level Nanog), MN (Middle-level Nanog), LN (Low-level Nanog), Ga (Gata6), Gc (Gcnf), and C (Cdx2). The diagram was obtained by sampling 10,000 cells simulated during 11.5 days. Transitions frequent more than the threshold defined in subsection *Transition diagram* in section *Methods* are extracted from trajectories. Transitions to/from the LS-LN-C-Gc state (dashed lines) were less frequent than the threshold but are drawn here to show the topological connectivity of the diagram.  
doi:10.1371/journal.pcbi.1003380.g006

whether the differentiation is the process of jumping among these states. Though there is a global trend of kinetic flows from the ESC states to the differentiated states, the predicted pathways are not single but comprise the network of flows. It should be important to compare the predicted distribution of pathways as in Fig. 6C with the distribution of pathways experimentally observed by following the fate of individual cells in the culture.

### Epigenetic landscapes

To analyze dynamics of differentiation, the epigenetic landscape that underlies transitions among cell states should provide a useful perspective [63,64]. Here, the landscape is derived from the

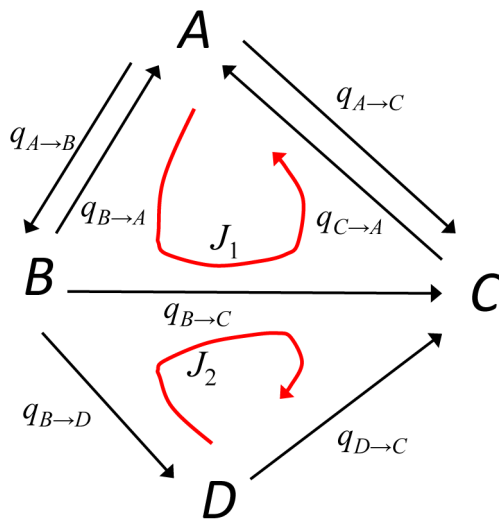
**Table 1. Definition of cell states and their labels.**

$N(\text{Sox2})$	0–10	10–200	200 -
$L(\text{Sox2})^*$	0	2	4
$N(\text{Oct4})$	0–20	20–200	200 -
$L(\text{Oct4})$	0	2	4
$N(\text{Nanog})$	0–40	40–400	400–4000 4000 -
$L(\text{Nanog})$	0	1	2 4
$N(\text{Cdx2})$	0–20	20–200	200 -
$L(\text{Cdx2})$	0	2	4
$N(\text{Gata6})$	0–100	100 -	
$L(\text{Gata6})$	0	4	
$N(\text{Gcnf})$	0–100	100 -	
$L(\text{Gcnf})$	0	4	

\*The values of  $L(i)$  were determined for the visibility of Fig. 8.  
doi:10.1371/journal.pcbi.1003380.t001

transition diagram by using the analogy with the free energy surface in equilibrium dynamics. In equilibrium dynamics, by using the transition-state theory formula, the rate of transition from  $i$  to  $j$ th states should be proportional to  $\exp(-\Delta G_{ij}^\ddagger)$  where  $\Delta G_{ij}^\ddagger = G_{ij}^\ddagger - G_i$ , and  $G_i$  and  $G_{ij}^\ddagger$  are the dimension-less free-energy like quantity at the  $i$ th state and at the transition state between the  $i$  and  $j$ th states, respectively. We use this analogy to equilibrium dynamics by fitting the calculated rates  $q_{i \rightarrow j}$  to this transition-state theory formula to obtain the free-energy like quantity  $G_i$  and  $G_{ij}^\ddagger$ . When the transition diagram has a tree-like structure without a loop, we can determine  $G$  of each state one by one by fitting the simulated rates to this formula. We use this analogy with equilibrium dynamics as far as possible to draw the landscape  $G$  of non-equilibrium transitions. This method of fitting, however, apparently breaks down when the transition network contains one or more loops: When the transition network contains a loop, for example, we may attempt to determine  $G$  of states in the loop one by one by starting from the  $i$ th state in the loop with the landscape value  $G_i$ , but at the end of traverse along the loop, we return to the initial  $i$ th state with a different value of  $G$  from the original  $G_i$ . In this way, the fitting to the transition-state theory formula is inconsistent along loops. This inconsistency can be resolved when we explicitly consider the non-equilibrium feature of dynamics by introducing the curl flux of transition kinetics [65–68]. Thus, the kinetic process along each loop can be expressed by the combination of the landscape and a kinetic flow curling along the loop. Transitions, therefore, are described by the combined representation of landscape and non-equilibrium curl flux. An example of a looped diagram having curl fluxes is shown in Fig. 7. From  $q_{i \rightarrow j}$  of this diagram, the free-energy like quantity  $G_i$  at the  $i$ th cell state and  $G_{ij}^\ddagger$  at the barrier between the  $i$  and  $j$ th cell states are calculated for  $i = A, B, C$ , and  $D$ , and curl fluxes  $J_1$  and  $J_2$  are obtained simultaneously. See subsection *Epigenetic landscape* in *Methods* for the explanation on how to calculate  $G_i$ ,  $G_{ij}^\ddagger$ ,  $J_1$  and  $J_2$  from  $q_{i \rightarrow j}$  of Fig. 7.

In Fig. 8 the landscapes and curl fluxes calculated from the simulated  $q_{i \rightarrow j}$  in the differentiation processes are shown on the two-dimensional plane with the coordinates of  $D(\text{TE}) = L(\text{Cdx2}) + L(\text{Gcnf}) - L(\text{Sox2}) - L(\text{Oct4}) - L(\text{Nanog})$  and  $D(\text{PrE}) = L(\text{Gata6}) + L(\text{Gcnf}) - L(\text{Sox2}) - L(\text{Oct4}) - L(\text{Nanog})$ . Here,  $L(i)$  is the label of the discretized expression level of the  $i$ th factor, which is defined to have the larger value for the higher expression

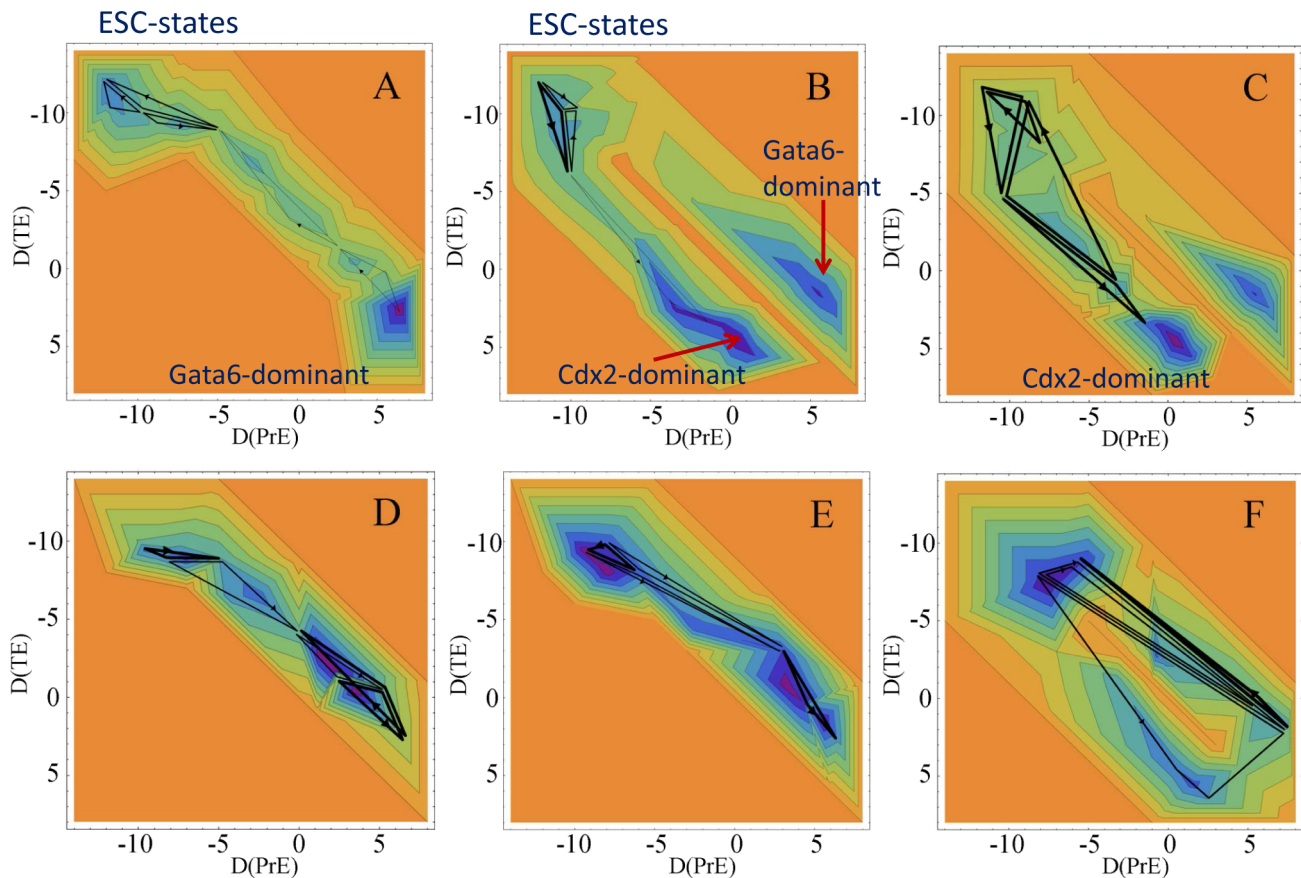


**Figure 7. An example of looped diagram among cell states,  $A$ ,  $B$ ,  $C$ , and  $D$ .** This diagram contains two loops. For two loops, two non-equilibrium curl fluxes,  $J_1$  and  $J_2$ , are introduced. Calculation of the epigenetic landscape and curl fluxes based on this diagram is explained in subsection *Epigenetic landscape* in section *Methods*. doi:10.1371/journal.pcbi.1003380.g007

level.  $D(TE)$  and  $D(PrE)$ , therefore, represent the degree of closeness to the trophoctoderm and primitive endoderm lineages, respectively. The precise values of  $L(i)$  are chosen for obtaining good visibility of Fig. 8, and are explained in Table 1. In Fig. 8, the calculated  $G_i$  and  $G_{ij}^{\ddagger}$  are plotted by assigning  $D(TE)$  and  $D(PrE)$  for  $i$  and  $ij$ , and  $G_i$  and  $G_{ij}^{\ddagger}$  are interpolated by a smooth surface in the two-dimensional space of  $D(TE)$  and  $D(PrE)$ .

The landscape corresponding to the diagram of Fig. 6C is shown in Fig. 8A. We see that the ESC states distribute on a flat basin in the region of small  $D(TE)$  and  $D(PrE)$ : ESCs wander around this basin driven by both the fluctuations satisfying the balance between the forward and reverse transitions and the kinetic flow of curl flux that breaks the balance. ESCs start differentiation as they move along the valley stretching toward the Gata6 dominant state. Transitions among intermediate states along this valley are also accompanied by the weak non-equilibrium curl flux.

In Figs. 8B and 8C, the artificial depletion of Oct4 is simulated with the decreased rate of synthesis of Oct4. Since Oct4 and Cdx2 work in an antagonistic way, the depletion of Oct4 results in the stronger expression of Cdx2, which leads ESCs to the trophoctoderm lineage: With the decrease of the rate of Oct4 synthesis to 25% (Fig. 8B) and 10% (Fig. 8C) of the value in Fig. 8A, the landscape changes its shape by extending the valley toward the



**Figure 8. Epigenetic landscapes and non-equilibrium curl fluxes derived from the transition diagrams in the differentiation process.** Contour maps of landscapes are shown on the two-dimensional plane of  $D(TE)$ , the degree of closeness to the trophoctoderm lineage, and  $D(PrE)$ , the degree of closeness to the primitive endoderm lineage, and curl fluxes are shown with black lines on the landscapes with the width of lines representing the strength of fluxes; thick lines ( $0.1\text{day}^{-1} < J$ ), lines with the middle width ( $0.01\text{day}^{-1} < J < 0.1\text{day}^{-1}$ ), and thin lines ( $J < 0.01\text{day}^{-1}$ ). (A–C) For the slow *Nanog* switching with  $\omega_A(\text{Nanog}) = 0.5$  and  $\omega_A = 10$  for other genes, and (D–F) for the fast *Nanog* switching with  $\omega_A = 10$  for all genes. The rate of Oct4 synthesis is (A and D) same as other factors, and (B and E) 25% and (C and F) 10% of other factors. doi:10.1371/journal.pcbi.1003380.g008



*Cdx2* dominant state. In Fig. 8B two valleys to primitive endoderm and trophoctoderm coexist with the curl flux on the basin of ESC states remaining, and in Fig. 8C the valley to trophoctoderm dominates. These results are consistent with the experimentally observed induction of the trophoctoderm lineage through the reduction of Oct4 [31].

Shown in Figs. 8D–8F are landscapes calculated with the assumption of the fast *Nanog* switching:  $\omega_A(\text{Nanog}) = 10$ . With this fast *Nanog* switching, the flat basin of the ESC states disappears, the curl flux in ESC states becomes localized, and ESCs quickly differentiate toward primitive endoderm (Fig. 8D). The curl flux on the ESC basin, therefore, originates from the slow *Nanog* switching. In other words, the eddy current associated with the non-adiabatic switching [23] manifests itself in the curl flux on the epigenetic landscape.

Difference between the slow and fast *Nanog* switching becomes more evident upon the reduction of Oct4 (Figs. 8E and 8F). With the fast *Nanog* switching, two valleys do not represent the distinct cell fate but they are directly connected to each other by the frequent transdifferentiation (Fig. 8F). This obscured differentiation arises from the averaged intermediate amount of Nanog synthesis under the fast *Nanog* switching. With the intermediate level of Nanog, the alleles of the lineage-specific genes tend to take the intermediate histone code as  $H_{\alpha i} = 1$  and  $H_{\beta i} = 0$  or  $H_{\alpha i} = 0$  and  $H_{\beta i} = 1$ . This intermediate level of activation of both *Cdx2* and *Gata6* increases the frequency of the transdifferentiation. With the slow *Nanog* switching, on the other hand, the histones of *Gata6* and *Cdx2* become either active with  $H_{\alpha i} = H_{\beta i} = 1$  or repressive with  $H_{\alpha i} = H_{\beta i} = 0$ , and such a clear-cut histone switching decreases the probability of the mixed expression of *Cdx2* and *Gata6*. In this way the simulated results suggest that the distinct cell fate decision is based on the slow *Nanog* switching, so that the phenotypic heterogeneity of ESCs is necessary for the stable differentiation.

The present quantification of epigenetic landscapes showed that the model naturally reproduces the observed differentiation to primitive endoderm [10]. The model also reproduces the induced differentiation to trophoctoderm observed when the Oct4 expression is artificially suppressed [3]. It should be interesting to further examine possibility of the predicted transdifferentiation due to the fast *Nanog* switching.

## Discussion

We developed a model of epigenetic dynamics and proposed a hypothesis that the timescale of formation/dissolution of TA decisively affects the self-renewal and differentiation of mouse ESCs. These effects can be checked experimentally by artificially varying the timescale of formation/dissolution of TA. The slower rate of formation/dissolution of TA for Oct4, for example, should give rise to the multi-peak distribution of Oct4, which should also affect the epigenetic landscape and non-equilibrium curl fluxes on the landscape. Further important is the application of the present ideas to engineering differentiation. Overexpression or repression of specific genes should alter the epigenetic landscape and curl fluxes, so that the calculation and observation of landscape and curl fluxes should provide a guideline for designing the process of cell differentiation.

An intriguing question is the effect of variation of the number of working alleles in a cell. In the present simulation, following the report for the single non-silenced *Nanog* allele in each ESC [51], only the single *Nanog* locus was considered in the simulation, which explained the bimodal Nanog distribution when the *Nanog* switching was slow. Assuming that both two alleles are working independently owing to the invalidated allelic regulation, we have

three peaks in the Nanog distribution corresponding to the ‘high-high’, ‘high-low’ and ‘low-low’ levels of expression for two alleles of Nanog with the slow *Nanog* switching as shown in Fig.S3. This predicted three-peak distribution could be experimentally tested in ESCs, though the more careful investigation is needed on the possible correlation between the allelic regulation and the regulation of the timescale of gene switching.

The core part of the network relations among genes in the present paper was built from the experimental observations, but there are experimental suggestions still not taken into account in the present model. For example, a recent report suggested the auto-repression of Nanog [45]. This suggested interaction can affect the transition dynamics between the HN and LN states, which should be examined by simulation. The validity of the assumptions used in the present modeling of epigenetic dynamics should be checked by examining how the results are modified when the model is further extended. In the present model, three processes having the different timescales were considered; TF binding/unbinding, TA formation/dissolution, and the histone code modification. Each of these processes consists of multiple sub-processes, and therefore if the model is extended with the finer resolution, the involved timescales should have more variety [69]. The TA formation/dissolution, for example, may involve assembly of mediators and RNA polymerase, phosphorylation of these factors, chromatin looping, and the large scale change in the chromosome positioning in nucleus. In the present model, we treat them in a coarse-grained manner by representing the TA state with  $A_{\mu i}$  which takes a value between 0 and 1. By treating these multiple processes explicitly, we may be able to construct a more quantitative model that can be compared with experiments in more details, and the validity of the level of coarse-graining in the present model could be checked through such comparison.

We should stress, however, that the main conclusions on the importance of design of timescales of regulations and the usefulness of combined representation of landscape and non-equilibrium curl fluxes do not depend on the molecular details. Indeed, the simplified mathematical or statistical physical models to capture the essential features of landscapes and curl fluxes should be useful. The dynamical systems models, for example, emphasize the importance of oscillations in the gene network [70], which conforms with the view presented here on the importance of rotating curl fluxes.

Another important direction to improve the present model is to take into account the core genes that guide ESCs to primitive ectoderm which further differentiates into the primary germ layers. To develop the reliable models, the effects of cell-cell communication and cell cycles should be also taken into account. Especially, the cell-cell communication should play important roles to stabilize the cell type of colony of interacting cells [70,71]. The model developed in the present paper was based on the assumption that the partial effect of cell-cell communication is implicitly taken into account by the mutual inhibition between *Cdx2* and *Gata6* (See *Methods* section). In order to analyze the differentiation process more quantitatively, the model needs to be extended to explicitly treat the effects of cell-cell communication. Those more elaborate models, together with the simplified statistical mechanical models, should reveal the rich phenomena in ESCs and differentiation processes.

## Methods

### Inference of the network from experimental observations

The model consists of interactions among six genes. Those interactions are inferred from the experimental data, which are

complemented with various levels of assumptions as explained below. In the following, the assumptions used are categorized into Level A, Level B, Level C, and Others. The aim of the present study is not to claim the validity of those assumptions, but to clarify the mechanisms of epigenetic dynamics by using a set of biologically consistent assumptions. The interactions considered in Fig. 1 were inferred from the discussions below, which are numbered in the same way as interactions designated in Fig. 1:

**Level A.** Microarray or other genetic experimental techniques revealed the correlation or anti-correlation between expression levels of two genes, and the chromatin immunoprecipitation or other biochemical data showed the binding of one factor to the locus of the other gene. These data support the assumption that the transcription factor (TF) synthesized from one gene directly regulates the other gene. The Level A assumptions give the backbone of the present model of the regulatory network.

1. Each of *Oct4*, *Sox2* and *Nanog* loci has the Oct/Sox enhancer region [7,72], on which Oct4 and Sox2 bind together to form the Oct4-Sox2 complex to activate *Oct4*, *Sox2*, or *Nanog* [4,7,72]. There are two possible ways of binding though they are not mutually exclusive; The Oct4-Sox2 complex is formed *before* they bind to DNA, or Oct4 and Sox2 bind to the adjacent sites of DNA to form the complex *after* binding. These two ways of binding are different in their cooperativity in the binding process. However, since the cooperativity of binding is masked by the cooperative formation/dissolution of transcription apparatus (TA) in the present model, these two ways of binding do not give significant difference in the switching behavior. We use, for simplicity, the latter assumption of forming complex *after* binding to DNA, but represent the effects of complex formation by assuming that the binding of either one of Oct4 or Sox2 is not enough but the binding of both two factors are needed for forming TA (We assume that the formation of the Oct4-Nanog complex is another route to form TA).

2. *Gcnf* binds to the *Oct4* and *Nanog* loci to repress them [48].

3. The Oct4-Cdx2 complex represses both *Oct4* and *Cdx2* [47,73].

4. Nanog binds directly to the *Gata6* locus to repress it [13].

5. Because the binding of Oct4 to the *Nanog* locus is necessary for forming the higher order structure of chromatin at the *Nanog* locus [61] and the binding site of Oct4 is adjacent to the binding site of Nanog at the *Nanog* locus [6], we expect that the Oct4-Nanog complex formed on the chromatin is necessary for building the TA of *Nanog*.

6. Nanog promotes the expression of Oct4 [5] and both Nanog and Oct4 directly bind to the *Oct4* locus [6]. Because the binding site of Oct4 is in proximity of the binding site of Nanog at the *Oct4* locus [6], we assume the promotion of the formation of TA of Oct4 through the binding of Oct4-Nanog complex on the *Oct4* locus.

7. Nanog is suggested to promote expression of Sox2 [8,74] and both Nanog and Oct4 directly bind to the *Sox2* locus [6]. Because the binding site of Oct4 is in proximity of the binding site of Nanog at the *Sox2* locus [6], we assume that the formation of TA of Sox2 is promoted by the binding of Oct4-Nanog complex on the *Sox2* locus.

**Level B.** Genetic experimental data showed the correlation or anti-correlation between expression levels of two genes, but the direct evidence for the physical interactions between two genes are not yet obtained. In this case, the interactions can be indirect through the other unidentified factors. Even in that case, we may assume the hypothetical direct interaction between two genes in the model. Such assumption is reasonable in the coarse-grained

model, in which the multiple detailed molecular processes are summarized into one process.

8. Excess expression of Oct4 reduces the expression level of Nanog [75]. We assume in the model that the *Nanog* locus has multiple binding sites of Oct4 and the occupation of the part of those sites is necessary for the formation of TA, but the occupation of all sites increases the rate for making the histone-code repressive.

9. The Nanog-null ESCs differentiate into cells similar to those induced by Gata6-positive cells [2,76]. Since Gata6 and Nanog work in an antagonistic way [73,77], we assume that Gata6 and Nanog are mutual repressors. Though it is not clear whether the repression of Nanog by Gata6 is direct or indirect through the other factors, we represent the interaction as a direct one by following the spirit of the coarse-grained approximation.

10. *Gcnf* is positively regulated by Gata6 and Cdx2 [10]. We assume for simplicity that *Gcnf* is activated directly by Gata6 and Cdx2 in the model.

**Level C.** No precise genetic data is available on the correlation or anti-correlation of gene activities, but from functional or biological observations, it is reasonable to assume the relation between two genes. The assumed interaction on such phenomenological basis might be a summary of the action of the larger network, but its representation as a single hypothetical process is useful to make the model behavior biologically reasonable.

11. Upon the removal of Lif or other agents from the culture, ESCs start differentiation. Then, each cluster of differentiated cells do not return to ESCs spontaneously. This stabilization of the differentiated cells may be enhanced by the positive feedback among the lineage-specific genes as was assumed by [63]. This effect of the regulatory network is represented in the model as auto-activation of lineage-specific genes, *Gata6*, *Cdx2* and *Gcnf*. These auto-activating interactions may be phenomenological or hypothetical interactions.

12. The cluster of cells differentiated into one lineage do not spontaneously transdifferentiate into the other lineages. This inhibition of transdifferentiation may arise from the reception of the external factor that is secreted or exhibited by the neighbor cells in the cluster. Such effect of the cell-cell communication is phenomenologically represented in the model by the mutual repression of genes specific to the different lineages.

**Others.** Other biochemical or biophysical data showed the existence of interactions.

13. Nanog dimerization is essential for the self-renewal of ESCs [19]. Nanog dimerization can be the faster process than its binding to the loci, so that we assume in the model that all the interactions between Nanog and chromatin are through the dimerized Nanog.

As assumed in the above argument, each locus of gene has multiple binding sites of TFs. From Fig. 1 we define the binding sites in the model as in Table 2.

## Reactions in the model

The state of the allele  $\mu$  of gene  $i$  in the model is represented by variables  $H_{\mu i}$ ,  $F_{\mu i}^j$ , and  $A_{\mu i}$ , where  $H_{\mu i}$  represents whether the histone code is active ( $H_{\mu i}=1$ ) or repressive ( $H_{\mu i}=0$ ),  $F_{\mu i}^j$  represents whether the  $j$ th TF is bound ( $F_{\mu i}^j=1$ ) or unbound ( $F_{\mu i}^j=0$ ), and  $A_{\mu i}$  represents whether the TA of the locus  $\mu i$  is formed and ready for transcription ( $A_{\mu i}=1$ ) or unformed and not ready ( $A_{\mu i}=0$ ). TA may be partially formed when the incomplete number of TFs are bound on a locus, and hence we write  $A_{\mu i}=0.5$  to represent such a partially formed TA. The copy number of the  $i$ th protein is represented by  $N(i)$ . The temporal changes of  $H_{\mu i}$ ,

**Table 2.** The binding sites of TFs on each locus of gene in the model.

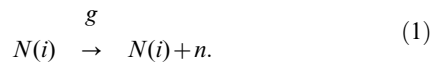
locus	activator binding site			repressor binding site		
	1	2	3	1	2	3
Sox2	Sox2	Oct4	Nanog <sub>2</sub>	-	-	-
Oct4	Sox2	Oct4	Nanog <sub>2</sub>	Cdx2	Gcnf	-
Nanog	Sox2	Oct4	Nanog <sub>2</sub>	Oct4*	Gata6	Gcnf
Cdx2	Cdx2	-	-	Oct4	Gata6	-
Gata6	Gata6	-	-	Nanog <sub>2</sub>	Cdx2	-
Gcnf	Cdx2	Gata6	Gcnf	-	-	-

\*The secondary Oct4 bound on the *Nanog* locus works in repressive ways only in the reactions of Eqs15 and 16.

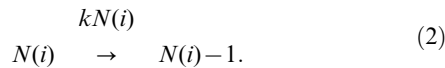
doi:10.1371/journal.pcbi.1003380.t002

$F_{\mu i}^j$ ,  $A_{\mu i}$ , and  $N(i)$  are numerically followed by using the Gillespie algorithm [58], which simulates the reactions explained below.

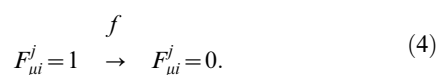
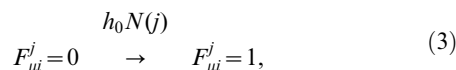
The  $i$ th protein is synthesized from the locus  $\mu i$  in a burst-like fashion with the rate  $g$  as



Here, we assume that the burst size  $n$  stochastically fluctuates in each burst with the probability of the distribution,  $P(n)=(X^n/n!)e^{-X}$ , with  $X$  being the average burst size. Though the distribution of the burst size was reported to obey the geometric distribution in bacteria [78], the precise distribution of the burst size in higher organisms is not known [79]. We here used the Poisson distribution to highlight the effects of the burst size, but as shown in Fig.S1, the model behavior does not sensitively depend on the burst size  $X$ , and hence we expect that the difference in distribution does not much affect the results. The  $i$ th protein is degraded with the rate  $kN(i)$  as



From Eqs.1 and 2, we can see that the representative copy number of each protein is  $2gX/k$ , where the factor 2 comes from two available alleles. The synthesized  $j$ th protein can bind to and unbind from the  $i$ th site if the locus  $\mu i$  has the binding site for the  $j$ th protein (Table 2). The binding rate,  $h$ , depends on the copy number of the protein to be bound. We assume the simplest linear relation by introducing a constant  $h_0$  as  $h=h_0N(j)$ . When the TF cooperatively binds in a form of oligomer, the contribution of higher orders of  $N(j)$  should be taken into account in  $h$ . However, in the present model, unlike the bacterial cases, the modification of  $h$  with the higher order terms of  $N(j)$  does not affect the model behaviors significantly because the cooperativity of switching is dominated by the formation/dissolution of TAs. We therefore assume

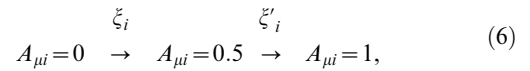


We should note that for the frequent switching between  $F_{\mu i}^j=1$  and 0 to take place, the ratio  $h_0N(j)/f$  should be around 1, or  $h_0N(j)/f \approx 2h_0gX/(kf) \approx 1$ . For  $j=\text{Nanog}$ , dimerization should be much faster than other processes, so that we use the copy number of Nanog dimer,  $N(\text{Nanog}_2)$ , as  $h=h_0N(\text{Nanog}_2)$  in Eq.3 instead of the total copy number of Nanog,  $N(\text{Nanog})$ .  $N(\text{Nanog}_2)$  can be estimated from the equilibrium relation as

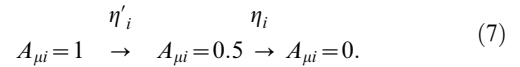
$$N(\text{Nanog}_2) = \frac{N(\text{Nanog})}{2} + \frac{1}{8K} - \sqrt{\frac{1}{8K} \left[ N(\text{Nanog}) + \frac{1}{8K} \right]}, \quad (5)$$

where  $K$  is the equilibrium constant of dimerization. In Eq.2, the copy number of the monomeric Nanog,  $N(\text{Nanog}_1) = N(\text{Nanog}) - 2N(\text{Nanog}_2)$  is used to define the degradation rate as  $kN(\text{Nanog}_1)$ .

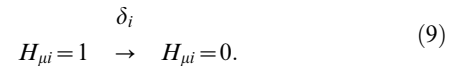
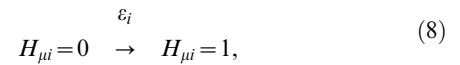
The binding of activator TFs triggers the formation of TA; starting from  $A_{\mu i}=0$  through the intermediate state  $A_{\mu i}=0.5$  to reach  $A_{\mu i}=1$  as



and the TA is resolved stochastically as



These formation/dissolution of TA should be largely affected by the state of histones at that locus. The change of the histone code is simulated by switching between the active and repressive states as



In Eqs.6–9 the rates  $\xi$ ,  $\xi'$ ,  $\eta$ ,  $\eta'$ ,  $\varepsilon$ , and  $\delta$  are represented with a suffix  $i$  to emphasize their dependences on the type of gene.

We next explain how the rates defined in Eqs.1–9 depend on the gene state,  $H_{\mu i}$ ,  $F_{\mu i}^j$ , and  $A_{\mu i}$ .  $g$  in Eq.1 depends on whether the TA is formed or not, which is represented by the variable  $A_{\mu i}$ . By using constants  $g_0 > 0$  and  $a > 1$ , we write  $g$  as

$$g = \begin{cases} 0 & (A_{\mu i}=0) \\ g_0 & (A_{\mu i}=0.5) \\ ag_0 & (A_{\mu i}=1) \end{cases} \quad (10)$$

Notice that with this rule the protein is synthesized only when the TA is formed at least partially as  $A_{\mu i} \geq 0.5$ .

The rates of TA formation and dissolution depend on how TFs are bound on the locus. Since TFs can be either of repressor or activator when they bind on the particular locus, we distinguish the binding sites  $j$  by writing  $j=ja$  when the  $j$ th TF is an activator of  $i$ , and  $j=jr$  when the  $j$ th TF is a repressor of  $i$ . With this representation, the rate for the first step of TA forming,  $\xi_i$  in Eq.6, is represented as follows;

$$\xi_i = \begin{cases} \xi_{ia} & (F_{\mu i}^{ja} = 1 \text{ for } \exists ja, F_{\mu i}^{jr} = 0 \text{ for } \forall jr, \text{ and } H_{\mu i} = 1) \\ \xi_{i0} & (F_{\mu i}^{ja} = 0 \text{ for } \forall ja, F_{\mu i}^{jr} = 0 \text{ for } \forall jr, \text{ and } H_{\mu i} = 1) \\ \xi_{ir} & (F_{\mu i}^{jr} = 1 \text{ for } \exists jr, \text{ and } H_{\mu i} = 1) \\ 0 & (H_{\mu i} = 0). \end{cases} \quad (11)$$

Notice that the TA is formed with  $\xi_i \neq 0$  only when the histone code is active ( $H_{\mu i} = 1$ ). Here,  $\xi_{ia}$  is the rate when some activator TF is bound on the locus but there is no bound repressor,  $\xi_{ir}$  is the rate when some repressor is bound on the locus, and  $\xi_{i0}$  is the rate when no activator or repressor in the model is bound on the locus. Even in the case there is no bound activator in the model, other TFs which are not represented in the model may bind on the locus to promote the TA forming, and hence we assume that the basal background rate for the TA formation,  $\xi_{i0}$ , is finite. Considering these definitions, we expect  $\xi_{ia} > \xi_{i0} > \xi_{ir}$ . For the second step of TA forming, we expect that all possible TFs should bind on the locus to complete the TA formation, so that we have

$$\xi'_i = \begin{cases} \xi'_{ia} & (F_{\mu i}^{ja} = 1 \text{ for } \forall ja, F_{\mu i}^{jr} = 0 \text{ for } \forall jr, \text{ and } H_{\mu i} = 1) \\ 0 & (\text{Otherwise}). \end{cases} \quad (12)$$

For *Gcnf* and *Sox2*, we do not consider the repressor explicitly, so the rule is simplified as

$$\xi'_i = \begin{cases} \xi'_{ia} & (F_{\mu i}^{ja} = 1 \text{ for } \forall ja \text{ and } H_{\mu i} = 1) \\ 0 & (\text{Otherwise}). \end{cases} \quad (13)$$

The rates of the histone-code modification,  $\varepsilon_i$  in Eq.8 and  $\delta_i$  in Eq.9, should also depend on the state of gene. We assume that the histone code can be turned active only when enzymes to modify the histone code are recruited by activators and are not inhibited by repressors. Therefore,  $\varepsilon_i$  becomes  $\varepsilon_i \neq 0$  when the gene state is similar to the situation for  $\xi = \xi_{ia}$ , i.e., when some activator TF binds and no repressor binds on the locus;

$$\varepsilon_i = \begin{cases} \varepsilon_{i0} & (F_{\mu i}^{ja} = 1 \text{ for } \exists ja, \text{ and } F_{\mu i}^{jr} = 0 \text{ for } \forall jr) \\ 0 & (\text{Otherwise}). \end{cases} \quad (14)$$

For  $i = Sox2, Oct4$  and *Nanog*, the histone-code activation is promoted by binding the Oct4-Sox2 or Oct4-Nanog complex. To represent cooperativity due to this complex formation, we modify the rule of Eq.14 as

$$\varepsilon_i = \begin{cases} \varepsilon_{i0} & \left( F_{\mu i}^{ja} = 1 \text{ for } ja = Oct4 \text{ and } Sox2, \right. \\ & \left. \text{or } ja = Oct4 \text{ and } Nanog_2, \text{ and } F_{\mu i}^{jr} = 0 \text{ for } \forall jr \right) \\ 0 & (\text{Otherwise}). \end{cases} \quad (15)$$

We assume that  $\delta_i$  becomes  $\delta_i \neq 0$  in the opposite situation, i.e., when some repressor binds and no activator binds on the locus. For all genes except *Gcnf* and *Sox2*, we have

$$\delta_i = \begin{cases} \delta_{i0} & (F_{\mu i}^{jr} = 1 \text{ for } \exists jr, F_{\mu i}^{ja} = 0 \text{ for } \forall ja, \text{ and } A_{\mu i} = 0) \\ 0 & (\text{Otherwise}). \end{cases} \quad (16)$$

No repressor is assumed to bind on the *Gcnf* or *Sox2* locus in the model, so that for  $i = Gcnf$  or *Sox2* we have the rule,

$$\delta_i = \begin{cases} \delta_{i0} & (F_{\mu i}^{ja} = 0 \text{ for } \forall ja \text{ and } A_{\mu i} = 0) \\ 0 & (\text{Otherwise}). \end{cases} \quad (17)$$

Eqs.1–9 were simulated with the Gillespie algorithm with the rates defined by Eqs.10–17. The simulation was started from the following initial condition which represents the pluripotent state,

$$F_{\mu i}^j = 0 \text{ (for } \forall j), A_{\mu i} = 0, \\ H_{\mu i} = 1, N(i) = 1 \times 10^3, \quad (18)$$

for  $i = Sox2, Oct4$ , and *Nanog* and

$$F_{\mu i}^j = 0 \text{ (for } \forall j), A_{\mu i} = 0, \\ H_{\mu i} = 0, N(i) = 0, \quad (19)$$

for  $i = Cdx2, Gata6$ , and *Gcnf*. Starting from this initial condition, the simulation for  $4 \times 10^5$  sec was performed by keeping  $\varepsilon_{i0} = 0$  for  $i = Cdx2, Gata6$ , and *Gcnf* to reach the steady ESC state. This first  $4 \times 10^5$  sec trajectory was discarded and the data were sampled after that for the statistical analysis by keeping  $\varepsilon_{i0} = 0$  for  $i = Cdx2, Gata6$ , and *Gcnf* to simulate ESCs or by turning  $\varepsilon_{i0}$  to be  $\varepsilon_{i0} \neq 0$  for  $i = Cdx2, Gata6$ , and *Gcnf* to simulate the differentiation process.

## Parameters

The model has parameters,  $g_0$ ,  $a$ , and  $X$  for protein synthesis,  $k$  for protein degradation,  $h_0$  and  $f$  for binding and unbinding of TF,  $K$  for Nanog dimerization,  $\xi_{ia}$ ,  $\xi_{ir}$ ,  $\xi_{i0}$ ,  $\xi'_{ia}$ ,  $\eta_i$  and  $\eta'_i$  for formation and dissolution of TA, and  $\varepsilon_{i0}$  and  $\delta_{i0}$  for the histone-code modification. For simplicity of description, the suffix  $a$  of  $\xi_{ia}$  and  $\xi'_{ia}$ , and the suffix 0 of  $\varepsilon_{i0}$  and  $\delta_{i0}$  are suppressed in the previous sections of *Model* and *Results*, in Fig. 2, and in Table 3. Parameters were determined according to the following guideline: (1) Parameter values were not tuned in a precise way but only their orders of magnitude were taken care of. (2) The same parameter was set to have the same value for different genes as far as possible. In the following, in order to determine the ranges that parameters can take with this guideline, we first discuss two basic quantities, the period of cell cycle  $\tau_{cc}$  in the following items 1 and 2, and the typical copy number of each protein in a cell in the item 3.

1. A basic timescale for the description of cellular processes is the period of cell cycle. The period of cell cycle of ESCs is  $\tau_{cc} \approx 4 \times 10^4$  sec  $\approx 0.5$  day, which is shorter than that of somatic cells. Though the periodic modulation of cells along the cell cycle is not explicitly treated in the present paper,  $\tau_{cc}$  can be used as a measure of timescales of other processes. The timescale of binding/unbinding of TF should be much shorter than  $\tau_{cc}$ , or in other words, the rates of binding/unbinding of TF should be much larger than  $1/\tau_{cc}$ . The timescale of TA formation/dissolution should not exceed the cell cycle period, so that the rates of TA formation/dissolution should be similar to or larger than  $1/\tau_{cc}$ . Since the histone code is often inherited across the cell cycle [35], the timescale of the histone-code modification should be larger than  $\tau_{cc}$ , or in other words, the rates of histone-code modification should be smaller than  $1/\tau_{cc}$ .



**Table 3.** Order of magnitude of parameters in the model.

$2ag_0X/k^*$	$10^3 - 10^4$	$\eta_i/k$	$5 \times 10^{-1} - 10$
$K$	$10^{-4} - 10^{-3}$	$\eta'_i/k$	$\sim \eta_i/k$
$h_0/k$	$\sim 1$	$\xi_i/k$	$5 \times 10^{-1} - 10$
$f/k$	$\sim 10^3$	$\zeta'_i/k$	$\sim \xi_i/k$
$\delta_i/k^\dagger$	$\sim 5 \times 10^{-2}$		
$\varepsilon_i/k$	$\sim 5 \times 10^{-2}$		

\* $k = 1 \times 10^{-4} \text{sec}^{-1}$ .

†Subscript 0 or  $a$  in  $\delta_{i0}$ ,  $\varepsilon_{i0}$ ,  $\xi_{ia}$ , and  $\zeta'_{ia}$  is omitted for simplicity.

doi:10.1371/journal.pcbi.1003380.t003

- Though the cell cycle period is a convenient measure of biological processes, it is not explicitly used in the model. Alternatively, we use the lifetime of proteins, or the rate constant of protein degradation  $k$ , as an explicit measure of processes in the model. From the observed variation of the copy number of Oct4 [4] and Nanog [16], we have  $k \approx 1 - 2 \times 10^{-4} \text{sec}^{-1}$ , which is larger than the frequency  $1/\tau_{cc}$  of cell division of ESCs;  $k \approx 2/\tau_{cc} - 4/\tau_{cc}$ . In the following, instead of  $1/\tau_{cc}$ , we use  $k$  as a measure to quantify the other parameters and to define adiabaticity.
- Another important quantity is the typical copy number of proteins in a cell. It has been observed that this copy number is  $\sim 10^4 - 10^5$  [80], which is much larger than that in bacteria ( $10 - 10^3$ ) [81] due to the much larger cell volume and the larger number of working ribosomes in eukaryotic cells. The copy number of TFs which concentrate in nucleus, however, should be smaller than that of proteins working in cytosol because the volume of nucleus is about 10% of the whole cell volume. Therefore, by assuming that the similar concentration to the one of cytosolic proteins is transported and accumulated in the nucleus, the typical value of the copy number of TFs should have the range of  $10^3 - 10^4$ . Therefore, we should have  $2ag_0X/k \approx 10^3 - 10^4$ .
- The rate of TF binding/unbinding may not be much different from that in bacteria, so that  $f/k \sim 10^3$ . For the sensitive gene switching, the probability that a TF binds to the locus should not be so close to 1 or 0, so that  $h_0N(i) \sim f$ . Given  $N(i) \approx 10^3$ , we have  $h_0/k \approx 1$ .
- The Nanog dimerization constant is  $K = N(\text{Nanog}_2)/N(\text{Nanog}_1)^2$ . By assuming  $N(\text{Nanog}_2) < N(\text{Nanog}_1) \approx 10^3 - 10^4$ , we have  $K \approx 10^{-4} - 10^{-3}$ .
- The rate to change the chemical state of a nucleosome may be as large as  $h$  or  $f$  because each reaction to add (remove) the methyl or acetyl group to (from) a nucleosome should be catalyzed by enzymes recruited by TFs [9,10]. The chemical state of an array of nucleosomes along each of gene loci, however, is changed in a cooperative way and should show the much longer lifetime. Indeed, the chemical state of nucleosomes, i.e., the histone code is often inherited across generations of cell cycle, and it has been observed that the timescale to change the histone code of *Oct4* in ESCs is several days, which is approximately  $10\tau_{cc}$  [35]. We assume, therefore, the rate to change the histone code to be  $\varepsilon_{i0}/k \approx \delta_{i0}/k \approx (10\tau_{cc})^{-1}/(2/\tau_{cc}) \approx 5 \times 10^{-2}$ .
- The timescale of TA formation/dissolution is not known, but we assume it to be slower than TF binding/unbinding and faster than the cell cycle. The latter assumption of the larger

rate of TA formation/dissolution than the cell-cycle frequency should be reasonable because the TA structure should be reset and resolved in the large scale reorganisation of chromatin structure at the time of cell division. The former assumption of the slower rate than the TF binding/unbinding is also reasonable when we consider that TA formation is the much more complex process including assembly of many factors, looping of chromatin, and chemical modification of RNA polymerase. Therefore, we examined the wide range of  $10 > \xi_{ia}/k \approx \eta_{ia}/k > \tau_{cc}^{-1}/(2/\tau_{cc}) \approx 0.5$ . The completion of TA may be the similar process to the initiation of the formation of TA, so that we assume  $\zeta'_{ia} \approx \xi_{ia}$  and  $\eta'_i \approx \eta_i$ . Other conditional values of  $\xi_i$  are defined in Eq.11 as  $\xi_{ir} < \xi_{i0} < \xi_{ia}$ .

From the above consideration, we can estimate the orders of magnitude of parameters, which are summarized in Table 3. In order to further analyze the meaning of difference in the order of magnitude of these parameters, we define the dimension-less parameters, adiabaticities, as measures of relative rates of individual processes to the rate of the protein copy-number change: We have three adiabaticities in our model of epigenetic dynamics,  $\omega_F = f/k$  which measures the relative frequency of the TF binding/unbinding,  $\omega_A(i) = \eta_i/k$  which measures the relative frequency of the TA formation/dissolution, and  $\omega_H(i) = \delta_i/k$  which measures the relative frequency of the histone-code modification. From the above estimation, the orders of magnitude of adiabaticity are  $\omega_F \gg \omega_A(i) > 1 > \omega_H(i)$  or  $\omega_F \gg 1 > \omega_A(i) > \omega_H(i)$ . Therefore, the TF binding/unbinding is strongly adiabatic, and the histone-code modification is non-adiabatic. The TA formation/dissolution is adiabatic or non-adiabatic depending on the type of gene, which characterises the dynamic behavior of the present model.

The parameters used in Figs. 3, 5, 6, and 8 in *Results* section are shown in Table 4. We can see that the values of Table 4 are within the range shown in Table 3. In Table 4, the dependence of parameters on the type of genes is minimized: The specific values deviating from the most common values are  $\varepsilon_{i0}$  for  $i = Cdx2$ , and  $\eta_i$ ,  $\eta'_i$ ,  $\xi_i$ , and  $\zeta'_i$  for  $i = Nanog$ . Since the differentiation to trophoctoderm takes place prior to the formation of inner cell mass (ICM) in the early embryo and ESCs are prepared from ICM, it is reasonable to assume that the differentiation to trophoctoderm is somehow suppressed in ESCs. We express this tendency by using the smaller value of  $\varepsilon_{i0}$  for  $i = Cdx2$ , the marker protein for trophoctoderm. The small value of  $\varepsilon_{i0}$  for  $i = Cdx2$  represents the possible inherent effects of silencing *Cdx2* in ESCs.

The smaller values of  $\eta_i$ ,  $\eta'_i$ ,  $\xi_i$ , and  $\zeta'_i$  for  $i = Nanog$  are the manifestation of the slow switching dynamics of the *Nanog* locus, which is a main feature of the present model of epigenetic dynamics as explained in Figs. 3B, 3C, 5, 6, and 8A–8C. To clarify the effects of this slow switching, we also calculate for comparison by using the same values of  $\eta_i$ ,  $\eta'_i$ ,  $\xi_i$ , and  $\zeta'_i$  for  $i = Nanog$  as those for *Sox2* or *Oct4* as shown in Figs. 3A, 3C, and 8D–8F.

### Massive parameter search

From Table 3, the order of magnitude of most of parameters are determined and their typical example values can be adopted as in Table 4. In Table 3, however, values of some parameters are undetermined yet. We perform a massive parameter search for values of these parameters to find the consistent values with the experimentally observed results.

An important undetermined set of parameters are  $\eta_i$  and  $\xi_i$ . We adopt the values in Table 4 for other parameters, and as discussed in *Parameters* subsection, we impose the constraints  $\eta'_i = \eta_i$  and  $\xi'_{ia} = \xi_{ia}$ . Considering the constraint of  $\xi_{ia} > \xi_{i0} > \xi_{ir}$ , we assume a

**Table 4.** The parameter set used in Figs.3, 5, 6, and 8 in the section *Results*.

	<i>Sox2</i>	<i>Oct4</i>	<i>Nanog</i>	<i>Cdx2</i>	<i>Gata6</i>	<i>Gcnf</i>
$g_0/k$ *	$2.0 \times 10$	$2.0 \times 10$	$2.0 \times 10$	$2.0 \times 10$	$2.0 \times 10$	$2.0 \times 10$
$a$	$2.0 \times 10$	$2.0 \times 10$	$2.0 \times 10$	$2.0 \times 10$	$2.0 \times 10$	$2.0 \times 10$
$X$	10	10	10	10	10	10
$K$	—	—	$2.0 \times 10^{-4}$	—	—	—
$h_0/k$	5.0	5.0	5.0	5.0	5.0	5.0
$f/k$	$1.0 \times 10^3$	$1.0 \times 10^3$	$1.0 \times 10^3$	$1.0 \times 10^3$	$1.0 \times 10^3$	$1.0 \times 10^3$
$\eta_i/k$	$1.0 \times 10$	$1.0 \times 10$	$5.0 \times 10^{-1} \ddagger$	$1.0 \times 10$	$1.0 \times 10$	$1.0 \times 10$
$\eta'_i/k$	$1.0 \times 10$	$1.0 \times 10$	$5.0 \times 10^{-1} \ddagger$	$1.0 \times 10$	$1.0 \times 10$	$1.0 \times 10$
$\xi_{ia}/k$	$1.0 \times 10$	$1.0 \times 10$	$5.0 \ddagger$	$1.0 \times 10$	$1.0 \times 10$	$1.0 \times 10$
$\xi_{i0}/k$	5.0	5.0	$3.0 \ddagger$	5.0	5.0	5.0
$\xi_{ir}/k$	$1.0 \times 10^{-1}$	$1.0 \times 10^{-1}$	$1.0 \times 10^{-1}$	$1.0 \times 10^{-1}$	$1.0 \times 10^{-1}$	$1.0 \times 10^{-1}$
$\xi'_{ia}/k$	$1.0 \times 10$	$1.0 \times 10$	$5.0 \ddagger$	$1.0 \times 10$	$1.0 \times 10$	$1.0 \times 10$
$\delta_{i0}/k$	$5.0 \times 10^{-2}$	$5.0 \times 10^{-2}$	$5.0 \times 10^{-2}$	$5.0 \times 10^{-2}$	$5.0 \times 10^{-2}$	$5.0 \times 10^{-2}$
$\varepsilon_{i0}/k$	$5.0 \times 10^{-2}$	$5.0 \times 10^{-2}$	$5.0 \times 10^{-2}$	$2.0 \times 10^{-2} \ddagger$	$7.0 \times 10^{-2} \ddagger$	$7.0 \times 10^{-2} \ddagger$

\* $k = 1 \times 10^{-4} \text{sec}^{-1}$ .

†These values are turned to 0 to simulate ESCs cultivated in media that can keep cells in the self-renewing pluripotent state.

‡Simulations with the fast Nanog switching were performed by making these values identical to the values for *Sox2* or *Oct4*.

doi:10.1371/journal.pcbi.1003380.t003

small value for  $\xi_{ir}$  as in Table 4 and also assume  $\xi_{ia} \approx 2\xi_{i0}$ . Then, the parameters  $\eta_i$  and  $\xi_{ia}$  are left undetermined. In Figs.3A, 3C, and 8D–8F, we used the values  $\eta_i/k = \xi_{ia}/k = 10$  to represent the situation that the TA formation/dissolution is much slower than the TF binding/unbinding but faster than the protein copy-number change, i.e.,  $f/k = 10^3 \gg \eta_i/k \approx \xi_{ia}/k > 1$ . This choice of values for  $\eta_i$  and  $\xi_{ia}$ , however, is not consistent with the experimentally observed distributions of expression level of SON as is shown in Fig. 3, and hence we examined the other values of  $\eta_i$  and  $\xi_{ia}$  for  $i = \text{Nanog}$ .

We generated about 1,000 parameter sets scattered on the two-dimensional plane of  $\log \eta_i$  and  $\log \xi_{ia}$ . For each of these generated parameter sets, we calculated 10,000 trajectories and obtain the distributions of expression level of SON by averaging over the trajectories. We then evaluated the score  $S$  as

$$\begin{aligned}
 S = & T(\text{number of Nanog peaks}) + \\
 & T(\text{positions of Nanog peaks}) \\
 & + T(\text{height of Nanog peaks}) + T(\text{zero level of Nanog}) \\
 & + T(\text{number of Sox2 peak}) + T(\text{zero level of Sox2}) \\
 & + T(\text{number of Oct4 peak}) + T(\text{position of Oct4 peak}) \\
 & + T(\text{zero level of Oct4}), \quad (20)
 \end{aligned}$$

where each term of Eq.20 is  $T \neq 0$  when the simulated data agrees with the corresponding feature of the observed data as shown in Fig. 3D, and  $T = 0$  otherwise:

- $T(\text{number of Nanog peaks}) = 1$  when the simulated distribution of expression level of Nanog has two peaks (HN and LN peaks).

- $T(\text{positions of Nanog peaks}) = 1$  when the ratio of the expression level at the HN peak to that at the LN peak,  $R_{\text{position}}$ , is  $5 < R_{\text{position}} < 50$ .
- $T(\text{height of Nanog peaks}) = 1$  when the ratio of the height of the HN peak to that of the LN peak,  $R_{\text{height}}$ , is  $3 < R_{\text{height}} < 4$  or  $6 < R_{\text{height}} < 7$ , and  $T(\text{height of Nanog peaks}) = 2$  when  $4 \leq R_{\text{height}} \leq 6$ .
- $T(\text{zero level of Nanog}) = 1$  when the population at the zero expression of Nanog is less than 2%.
- $T(\text{number of Sox2(Oct4) peak}) = 1$  when the distribution of the expression level of Sox2 (Oct4) is single peaked.
- $T(\text{zero level of Sox2(Oct4)}) = 1$  when the population of zero expression of Sox2 (Oct4) is less than 2%.
- $T(\text{position of Oct4 peak}) = 1$  when the expression level at the peak of Oct4 distribution is in between expression levels at the HN and LN peaks.

In this way,  $S = 10$  when all the observed features of the distributions of expression level of SON in ESCs are reproduced by the simulated data. In Fig. 4,  $S$  for 1,125 parameter sets is plotted on the plane of  $\log \eta_i$  and  $\log \xi_{ia}$  with  $i = \text{Nanog}$ .

The other undetermined set of parameters in Table 3 are the bare rate of protein synthesis,  $g_0$ , the ratio of the rate of synthesis at the completed TA to that of the partially formed TA,  $a$ , and the average burst size,  $X$ . Since we have a constraint  $2ag_0X/k = 10^3 - 10^4$  as in Table 3, we fixed  $2ag_0X/k$  to the value  $2ag_0X/k = 8 \times 10^3$  as in Table 4 and searched the values of  $a$  and  $X$  extensively by modifying  $g_0$  according to the constraint  $2ag_0X/k = 8 \times 10^3$ . 1,125 parameter sets were generated as scattered points on the two-dimensional plane of  $\log a$  and  $\log X$  and  $S$  were calculated by averaging over 10,000 trajectories for each of the parameter sets. The calculated  $S$  is plotted in Fig.S1, which shows that  $a$  should be within the range of  $5 < a < 50$

especially to satisfy  $T(\text{positions of Nanog peaks})=1$  and the results are not sensitively dependent on the burst size  $X > 1$ .

### Transition diagram

The total time length,  $t_i(k)$ , during which the  $k$ th trajectory stayed at a certain cell state  $i$  is calculated. By sampling  $N = 10,000$  trajectories of  $T = 11.5$  days, the averaged frequency of the appearance of the state  $i$ ,  $P_i$ , is obtained as  $P_i = \frac{1}{N} \sum_{k=1}^N t_i(k)/T$ . We can see in Fig. S4 that the small number of states appear much more frequently than the other many states. We disregard the rarely appearing states and draw the transition diagrams among the states whose  $P_i$  are larger than a threshold value  $P_i^{\text{thr}}$ . Here, we choose different thresholds for different transition diagrams because the difficulty to solve simultaneous equations for landscape and fluxes depends on the topology of the diagram. The larger threshold makes the diagram simpler to increase the solvability of equations, but we use the smallest possible threshold;  $P_i^{\text{thr}} = 0.014$  for Fig. 8A, 0.016 for Figs. 8B–8E, and 0.029 for Fig. 8F.

Then, the time of the trajectory needed for the transition from the  $i$ th to  $j$ th states is monitored and recorded as  $\tau_{i \rightarrow j}$ .  $\tau_{i \rightarrow j}$  is averaged along the trajectory and over the ensemble of trajectories to obtain  $\bar{\tau}_{i \rightarrow j}$ . The transition rate is defined by  $q_{i \rightarrow j} = 1/\bar{\tau}_{i \rightarrow j}$ . The link  $i \rightarrow j$  between two cell states  $i$  and  $j$  is drawn in the transition diagram when the transition  $i \rightarrow j$  is observed more frequently than the threshold times, which are 200 (Fig. 8F), 600 (Figs. 6C, 8A, 8B, and 8E), 700 (Fig. 8C), or 900 (Fig. 8D) times in the sampled ensemble of trajectories. The transition diagram of Fig. 6C is drawn by connecting the cell states with thus defined links of transitions.

### Epigenetic landscape

The epigenetic landscape,  $\{G_i\}$  with  $\{G_{ij}^\ddagger\}$ , is calculated with the following rules:

- Given  $G_i$  at the  $i$ th state,  $G_{ij}^\ddagger$  at the saddle between the  $i$ th and  $j$ th states is calculated by fitting the expression  $q_{i \rightarrow j} = q_0 \exp[-(G_{ij}^\ddagger - G_i)]$  with  $q_0 = 2/\text{day}$  to the simulated rate by regarding  $G$  as a quantity analogous to free energy.
- After obtaining  $G_{ij}^\ddagger$ ,  $G_j$  is calculated by fitting the expression  $q_{j \rightarrow i} = q_0 \exp[-(G_{ij}^\ddagger - G_j)]$  to the simulated rate of the reverse transition.
- When the reverse transition does not appear in the simulated trajectories, a threshold value  $G_{th} = 4.6$  ( $\exp(-G_{th}) = 0.01$ ) is used to represent the high enough barrier to inhibit the reverse transition within the simulated timescale. We then have  $G_j = G_{ij}^\ddagger - G_{th}$ .
- The above three rules can not be applied for the looped part in the transition diagram. This problem is solved by combining landscape and curl flux to represent the kinetic flow on the landscape.

The last rule can be explained by using an example of Fig. 7, in which four cell states  $A$ ,  $B$ ,  $C$ , and  $D$  are connected by directed links that represent transitions among cell states. For the diagram of Fig. 7, we should solve the following equations:

$$q_{A \rightarrow B} = q_0 \exp[-(G_{AB}^\ddagger - G_A)] + J_1,$$

$$q_{B \rightarrow A} = q_0 \exp[-(G_{AB}^\ddagger - G_B)] - J_1,$$

$$q_{B \rightarrow C} = q_0 \exp[-(G_{BC}^\ddagger - G_B)] + J_1 + J_2,$$

$$q_0 \exp[-G_{th}] = q_0 \exp[-(G_{BC}^\ddagger - G_C)] - J_1 - J_2,$$

$$q_{C \rightarrow A} = q_0 \exp[-(G_{CA}^\ddagger - G_C)] + J_1,$$

$$q_{A \rightarrow C} = q_0 \exp[-(G_{CA}^\ddagger - G_A)] - J_1,$$

$$q_{B \rightarrow D} = q_0 \exp[-(G_{BD}^\ddagger - G_B)] - J_2,$$

$$q_0 \exp[-G_{th}] = q_0 \exp[-(G_{BD}^\ddagger - G_D)] + J_2,$$

$$q_{D \rightarrow C} = q_0 \exp[-(G_{DC}^\ddagger - G_D)] - J_2,$$

$$q_0 \exp[-G_{th}] = q_0 \exp[-(G_{DC}^\ddagger - G_C)] + J_2. \quad (21)$$

These equations have 11 variables. When we fix  $G_n$  at a certain state  $n$ , the relative height  $G_i - G_n$  at other three states,  $G_{ij}^\ddagger - G_n$  at five saddles, and two currents  $J_1$  and  $J_2$  are obtained by solving the above 10 equations simultaneously.

### Supporting Information

#### Figure S1 Search for a range of parameter set $a$ and $X$ .

The ability of the simulated trajectories to reproduce the experimentally observed data of distributions of the expression level of SON is evaluated by the score  $S$  which is defined in Eq. 20 in *Methods* section. The score was evaluated for each of 1,125 parameter sets scattered on the two-dimensional plane of  $\log_{10} a$  and  $\log_{10} X$ , where  $a$  is the ratio of the rate of protein synthesis from the fully formed TA and that from the partially formed TA, and  $X$  is the average burst size.

(TIF)

#### Figure S2 Simulated distributions of the copy number of protein factors which appeared in trajectories of the differentiation process from the ESC states to the Gata6-dominant state.

Distributions are divided to define the cell states by introducing thresholds designated by arrows. The abbreviations used to refer the cell states in Fig. 6C of the main text are written on each panel. 10,000 trajectories for 11.5 days were used for sampling the data.

(TIF)

#### Figure S3 Distribution of expression level of Nanog in the case that either of two alleles of Nanog is not silenced through the allelic regulation.

(A)  $\omega_A(\text{Nanog}) = 10$  and (B)  $\omega_A(\text{Nanog}) = 0.5$ . Other parameters are the same as those used in Fig. 3 of the main text.

(TIF)

#### Figure S4 Probability $P_i$ of the appearance of the cell state $i$ in the 10,000 simulated trajectories are shown in the rank order of $P_i$ .

(TIF)

## Author Contributions

Conceived and designed the experiments: MS KI TPT. Performed the experiments: MS YK KM. Analyzed the data: MS YK KM KI TPT.

Contributed reagents/materials/analysis tools: MS YK KM TPT. Wrote the paper: MS KM TPT.

## References

- Chambers I, Colby D, Robertson M, Nichols J, Lee S, et al. (2003) Functional expression cloning of nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* 113: 643–655.
- Mitsui K, Tokuzawa Y, Itoh H, Segawa K, Murakami M, et al. (2003) The homeoprotein nanog is required for maintenance of pluripotency in mouse epiblast and es cells. *Cell* 113: 631–642.
- Niwa H, Miyazaki J, Smith AG (2000) Quantitative expression of oct-3/4 defines differentiation, dedifferentiation or self-renewal of es cells. *Nat Genet* 24: 372–376.
- Chew JL, Loh YH, Zhang W, Chen X, Tam WL, et al. (2005) Reciprocal transcriptional regulation of *pou5f1* and *sox2* via the oct4/*sox2* complex in embryonic stem cells. *Mol Cell Biol* 25: 6031–6046.
- Wang J, Rao S, Chu J, Shen X, Levasseur DN, et al. (2006) A protein interaction network for pluripotency of embryonic stem cells. *Nature* 444: 364–368.
- Loh YH, Wu Q, Chew JL, Vega VB, Zhang W, et al. (2006) The oct4 and nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* 38: 431–440.
- Masui S, Nakatake Y, Toyooka Y, Shimosato D, Yagi R, et al. (2007) Pluripotency governed by *sox2* via regulation of *oct3/4* expression in mouse embryonic stem cells. *Nat Cell Biol* 9: 625–635.
- Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, et al. (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 122: 947–956.
- Boyer LA, Mathur D, Jaenisch R (2006) Molecular control of pluripotency. *Curr Opin Genet Dev* 16: 455–462.
- Niwa H (2007) How is pluripotency determined and maintained. *Development* 134: 635–646.
- Chambers I, Silva J, Colby D, Nichols J, Nijmeijer B, et al. (2007) Nanog safeguards pluripotency and mediates germline development. *Nature* 450: 1230–1235.
- Silva J, Nichols J, Theunissen TW, Guo G, van Oosten AL, et al. (2009) Nanog is the gateway to the pluripotent ground state. *Cell* 138: 722–737.
- Singh AM, Hamazaki T, Katherine E, Hankowski NT (2007) A heterogeneous expression pattern for nanog in embryonic stem cells. *Stem Cells* 25: 2534–2542.
- Kalmar T, Lim C, Hayward P, Muñoz-Descalzo S, Nichols J, et al. (2009) Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells. *PLoS Biol* 7: e1000149.
- Eldar A, Elowitz MB (2010) Functional roles for noise in genetic circuits. *Nature* 467: 167–173.
- Thomson M, Liu SJ, Zou LN, Smith Z, Meissner A, et al. (2011) Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. *Cell* 145: 875–889.
- Canham MA, Sharov AA, Ko MSH, Brickman JM (2010) Functional heterogeneity of embryonic stem cells revealed through translational amplification of an early endodermal transcript. *PLoS Biol* 8: e1000379.
- Niwa H, Ogawa K, Shimosato D, Adachi K (2009) A parallel circuit of *lif* signalling pathways maintains pluripotency of mouse es cells. *Nature* 460: 118–122.
- Wang J, Levasseur DN, Orkin SH (2008) Requirement of nanog dimerization for stem cell self-renewal and pluripotency. *Proc Natl Acad Sci USA* 105: 6326–6331.
- Glauche I, Herberg M, Roeder I (2010) Nanog variability and pluripotency regulation of embryonic stem cells - insights from a mathematical model analysis. *PLoS One* 5: e11238.
- Sasai M, Wolynes PG (2003) Stochastic gene expression as a many body problem. *Proc Natl Acad Sci USA* 100: 2374–2379.
- Walczak AM, Sasai M, Wolynes PG (2005) Self-consistent proteomic field theory of stochastic gene switches. *Biophys J* 88: 828–850.
- Walczak AM, Onuchic JN, Wolynes PG (2005) Absolute rate theories of epigenetic stability. *Proc Natl Acad Sci USA* 102: 18926–18931.
- Yoda M, Ushikubo T, W WI, Sasai M (2007) Roles of noise in single and coupled multiple genetic oscillators. *J Chem Phys* 126: 115101.
- Shi PZ, Qian H (2007) A perturbation analysis of rate theory of self-regulating genes and signaling networks. *J Chem Phys* 134: 065104.
- Feng H, Han B, Wang J (2011) Adiabatic and non-adiabatic non-equilibrium stochastic dynamics of single regulating genes. *J Phys Chem B* 115: 1254–1261.
- Feng H, Han B, Wang J (2012) Landscape and global stability of nonadiabatic and adiabatic oscillations in a gene network. *Biophys J* 102: 1001–1010.
- Zhang K, Sasai M, Wang J (2013) Eddy current and coupled landscapes for nonadiabatic and nonequilibrium dynamics of complex systems. *Proc Natl Acad Sci USA* 110: 14930–14935.
- Schultz D, Jacob EB, Onuchic JN, Wolynes PG (2007) Molecular level stochastic model for competence cycles in *bacillus subtilis*. *Proc Natl Acad Sci USA* 104: 17582–17587.
- Blake WJ, Kærn M, Cantor CR, Collins JJ (2003) Noise in eukaryotic gene expression. *Nature* 422: 633–637.
- Raser JM, O'Shea EK (2004) Control of stochasticity in eukaryotic gene expression. *Science* 304: 1811–1814.
- Zhang J, Chen L, Zhou T (2012) Analytical distribution and tunability of noise in a model of promoter progress. *Biophys J* 102: 1247–1257.
- Tokuda N, Sasai M, Chikenji G (2011) Roles of dna looping in enhancer-blocking activity. *Biophys J* 100: 126–134.
- Rahl PB, Lin CY, Seila AC, Flynn RA, McQuine S, et al. (2010) c-myc regulates transcriptional pause release. *Cell* 141: 432–445.
- Hathaway NA, Bell O, Hodges C, Miller EL, Neel DS, et al. (2012) Dynamics and memory of heterochromatin in living cells. *Cell* 149: 1447–1460.
- Dodd IB, Micheelsen MA, Sneppen K, Thon G (2007) Theoretical analysis of epigenetic cell memory by nucleosome modification. *Cell* 129: 813–822.
- Sneppen K, Micheelsen MA, Dodd IB (2008) Ultrasensitive gene regulation by positive feedback loops in nucleosome modification. *Mol Syst Biol* 4: 182.
- Artyomov MN, Meissner A, Chakraborty AK (2010) A model for genetic and epigenetic regulatory networks identifies rare pathways for transcription factor induced pluripotency. *PLoS Comput Biol* 6: e1000785.
- Sneppen K, Dodd IB (2012) A simple histone code opens many paths to epigenetics. *PLoS Comput Biol* 8: e1002643.
- Becskei A, Kaufmann BB, van Oudenaarden A (2005) Contributions of low molecular number and chromosome positioning to stochastic gene expression. *Nat Genet* 37: 937–944.
- Tokuda N, Terada TP, Sasai M (2012) Dynamical modeling of three-dimensional genome organization in interphase budding yeast. *Biophys J* 102: 296–304.
- Tanizawa H, Iwasaki O, Tanaka A, Capizzi JR, Wickramasinghe P, et al. (2010) Mapping of long-range associations throughout the fission yeast genome reveals global genome organization linked to transcriptional regulation. *Nucleic Acids Res* 38: 8164–8177.
- MacArthur BD, Maayan A, Lemischka IR (2009) Systems biology of stem cell fate and cellular reprogramming. *Nat Rev Mol Cell Biol* 10: 672–681.
- Chickarmane V, Trocin C, Nuber UA, Sauro HM, Peterson C (2006) Transcriptional dynamics of the embryonic stem cell switch. *PLoS Comput Biol* 2: e123.
- Navarro P, Festuccia N, Colby D, Gagliardi A, Mullin NP, et al. (2012) Oct4/*sox2*-independent nanog autorepression modulates heterogeneous nanog gene expression in mouse es cells. *EMBO J* 31: 4547–4562.
- Kamiya D, Banno S, Sasai N, Ohgushi M, Inomata H, et al. (2011) Intrinsic transition of embryonic stem-cell differentiation into neural progenitors. *Nature* 470: 503–510.
- Niwa H, Toyooka Y, Shimosato D, Strumpf D, Takahashi K, et al. (2005) Interaction between oct3/4 and *cdx2* determines trophoblast differentiation. *Cell* 123: 917–929.
- Gu P, LeMenuet D, Chung ACK, Mancini M, Wheeler DA, et al. (2005) Orphan nuclear receptor *gcnf* is required for the repression of pluripotency genes during retinoic acid-induced embryonic stem cell differentiation. *Mol Cell Biol* 25: 8507–8519.
- Yasuda S, Tsuneyoshi N, Sumi T, Hasegawa K, Tada T, et al. (2006) Nanog maintains self-renewal of primate es cells in the absence of a feeder layer. *Genes to Cells* 11: 1115–1123.
- Chickarmane V, Peterson C (2008) A computational model for understanding stem cell, trophoblast and endoderm lineage determination. *PLoS One* 3: e3478.
- Miyazawa Y, Torres-Padilla ME (2012) Control of ground-state pluripotency by allelic regulation of *nanog*. *Nature* 483: 470–475.
- Egli D, Birkhoff G, Eggan K (2008) Mediators of reprogramming: transcription factors and transitions through mitosis. *Nat Rev Mol Cell Biol* 9: 505–516.
- Orford KW, Scadden DT (2008) Deconstructing stem cell self-renewal: genetic insights into cell cycle regulation. *Nat Rev Genet* 9: 115–128.
- Bergman Y, Cedar H (2013) Dna methylation dynamics in health and disease. *Nat Str Mol Biol* 20: 274–281.
- Singh A, Bokes P (2012) Consequences of mrna transport on stochastic variability in protein levels. *Biophys J* 103: 1087–1096.
- Sun J, Gong X, Puro B, Zhao Z (2012) Uncovering microRNA and transcription factor mediated regulatory networks in glioblastoma. *PLoS Comput Biol* 8: e1002488.
- Mitarai N, Benjamin JAM, Krishnab S, Semsey S, Csiszovszkie Z, et al. (2009) Dynamic features of gene expression control by small regulatory RNAs. *Proc Natl Acad Sci USA* 106: 10655–10659.
- Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* 81: 2340–2361.
- Peter IS, Faure E, Davidson EH (2012) Predictive computation of genomic logic processing functions in embryonic development. *Proc Natl Acad Sci USA* 109: 16434–16442.



60. Cartwright P, McLean C, Sheppard A, Rivett D, Jones K, et al. (2005) *Lif/stat3* controls es cell self-renewal and pluripotency by a myc-dependent mechanism. *Development* 132: 885–896.
61. Levasseur DN, Wang J, Dorschner MO, Stamatoyannopoulos JA, Orkin SH (2008) Oct4 dependence of chromatin structure within the extended nanog locus in es cells. *Genes Dev* 22: 575–580.
62. de Wit E, Bouwman BAM, Zhu Y, Klous P, Splinter E, et al. (2013) The pluripotent genome in three dimensions is shaped around pluripotency factors. *Nature* 501: 227–231.
63. Huang S (2009) Reprogramming cell fates: reconciling rarity with robustness. *BioEssays* 31: 546–560.
64. Huang S (2010) Cell lineage determination in state space: A systems view brings flexibility to dogmatic canonical rules. *PLoS Biol* 8: e1000380.
65. Ao P (2004) Potential in stochastic differential equations: novel construction. *J Phys A* 37: L25–L30.
66. Qian H (2006) Open-system nonequilibrium steady state: Statistical thermodynamics, fluctuations, and chemical oscillations. *J Phys Chem B* 110: 15063–15074.
67. Wang J, Xu L, Wang E (2008) Potential landscape and flux framework of nonequilibrium networks: Robustness, dissipation, and coherence of biochemical oscillations. *Proc Natl Acad Sci USA* 105: 12271–12276.
68. Xing J (2010) Mapping between dissipative and hamiltonian systems. *J Phys A* 43: 375003.
69. Coulon A, Chow CC, Singer RH, Larson DR (2013) Eukaryotic transcriptional dynamics: from single molecules to cell populations. *Nat Rev Genet* 14: 572–584.
70. Furusawa C, Kaneko K (2012) A dynamical-systems view of stem cell biology. *Science* 338: 215–217.
71. White DE, Kinney MA, McDevitt TC, Kemp ML (2013) Spatial pattern dynamics of 3d stem cell loss of pluripotency via rules-based computational modeling. *PLoS Comput Biol* 9: e1002952.
72. Okumura-Nakanishi S, Saito M, Niwa H, et al (2005) Oct-3/4 and sox2 regulate oct-3/4 gene in embryonic stem cells. *J Biol Chem* 18: 5307–5317.
73. Hough SR, Clements I, Welch PJ, et al (2006) Differentiation of mouse embryonic stem cells after rna interference-mediated silencing of oct4 and nanog. *Stem Cells* 24: 1467–1475.
74. Jaenisch R, Young R (2008) Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* 132: 567–582.
75. Pan G, Li J, Zhou Y, et al (2006) A negative feedback loop of transcription factors that controls stem cell pluripotency and self-renewal. *FASEB J* 20: 1730–1732.
76. Fujikura J, Yamato E, Yonemura S, Hosoda K, Masui S, et al. (2002) Differentiation of embryonic stem cells is induced by gata factors. *Genes Dev* 16: 784–789.
77. Ralston A, Rossant J (2002) Genetic regulation of stem cell origins in the mouse embryo. *Clin Genet* 68: 106–112.
78. Ingram PJ, Stumpf MPH, Stark J (2008) Nonidentifiability of the source of intrinsic noise in gene expression from single-burst data. *PLoS Comput Biol* 4: e1000192.
79. Dar RD, Razoooky BS, Singh A, Trimeloni TV, McCollum JM, et al. (2012) Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proc Natl Acad Sci USA* 109: 17454–17459.
80. Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, et al. (2011) Global quantification of mammalian gene expression control. *Nature* 473: 337–342.
81. Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, et al. (2010) Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* 329: 533–538.