INTERFACE

royalsocietypublishing.org/journal/rsif

Research

Cite this article: Huang B, Lu M, Galbraith M, Levine H, Onuchic JN, Jia D. 2020 Decoding the mechanisms underlying cell-fate decision-making during stem cell differentiation by random circuit perturbation. *J. R. Soc. Interface* **17**: 20200500.
http://dx.doi.org/10.1098/rsif.2020.0500

Received: 24 June 2020
Accepted: 20 July 2020

Subject Category:
Life Sciences–Physics interface

Subject Areas:
biophysics, systems biology, computational biology

Keywords:
stem cell, gene regulatory circuit, random circuit perturbation, hierarchical structure, systems biology

Authors for correspondence:
Herbert Levine
e-mail: h.levine@northeastern.edu
Jose N. Onuchic
e-mail: jonuchic@rice.edu
Dongya Jia
e-mail: dj9@rice.edu

†These authors contributed equally to this work.

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.5082480.

THE ROYAL SOCIETY PUBLISHING

Decoding the mechanisms underlying cell-fate decision-making during stem cell differentiation by random circuit perturbation

Bin Huang[1,†], Mingyang Lu[5,†], Madeline Galbraith[1,2], Herbert Levine[1,6,7], Jose N. Onuchic[1,2,3,4] and Dongya Jia[1]

[1]Center for Theoretical Biological Physics, [2]Department of Physics and Astronomy, [3]Department of Chemistry, and [4]Department of Bioengineering, Rice University, Houston, TX 77005, USA
[5]The Jackson Laboratory, 600 Main St, Bar Harbor, ME 04609, USA
[6]Department of Bioengineering, and [7]Department of Physics, Northeastern University, Boston, MA 02115, USA

ML, 0000-0001-8158-0593; HL, 0000-0002-8819-9055; JNO, 0000-0002-9448-0388; DJ, 0000-0002-6307-8580

Stem cells can precisely and robustly undergo cellular differentiation and lineage commitment, referred to as stemness. However, how the gene network underlying stemness regulation reliably specifies cell fates is not well understood. To address this question, we applied a recently developed computational method, *ra*ndom *ci*rcuit *pe*rturbation (RACIPE), to a nine-component gene regulatory network (GRN) governing stemness, from which we identified robust gene states. Among them, four out of the five most probable gene states exhibit gene expression patterns observed in single mouse embryonic cells at 32-cell and 64-cell stages. These gene states can be robustly predicted by the stemness GRN but not by randomized versions of the stemness GRN. Strikingly, we found a hierarchical structure of the GRN with the Oct4/Cdx2 motif functioning as the first decision-making module followed by Gata6/Nanog. We propose that stem cell populations, instead of being viewed as all having a specific cellular state, can be regarded as a heterogeneous mixture including cells in various states. Upon perturbations by external signals, stem cells lose the capacity to access certain cellular states, thereby becoming differentiated. The new gene states and key parameters regulating transitions among gene states proposed by RACIPE can be used to guide experimental strategies to better understand differentiation and design reprogramming. The findings demonstrate that the functions of the stemness GRN is mainly determined by its well-evolved network topology rather than by detailed kinetic parameters.

1. Introduction

Embryonic stem cells (ESCs) can differentiate into cells of specialized types in a precise and organized manner, and dysregulation in stem cell differentiation results in early fetal death or severe disease [1–3]. Due to its essential role in survival for all multicellular organisms, stem cell differentiation must be highly conserved in order to allow for precise decisions at each step of lineage commitment. However, recent experimental results suggest that some transcription factors (TFs) such as Nanog exhibit heterogeneous expression levels at the single-cell level in mouse ESCs [4–8]. It remains largely unknown how the regulatory machinery of stemness performs its robust function in the presence of significant cell-to-cell heterogeneity. The answer to this question will shed light on the regulatory mechanism of stem cell differentiation, a crucial step toward better cellular reprogramming and stem cell-based therapies.

A substantial amount of research has been conducted to identify key TFs and their roles in directing stem cell differentiation [9–11]. These accumulated data enable us to map the underlying gene regulatory networks (GRNs). To elucidate the operating principles of these GRNs, computational approaches have been applied [12–16]. In particular, some of these computational studies adopted a bottom-up approach to construct GRNs with a small set of master regulators, and assume that the decision-making of stem cell differentiation is driven by the master regulators, such as the TFs Oct4, Sox2 and Cdx2. The dynamics of the GRNs can be simulated by either deterministic [14,15,17] or stochastic approaches [12,13,18–21]. These studies have indeed provided valuable insights into the regulatory mechanism underlying stem cell differentiation. However, these studies typically suffer from three issues. First, the modelling analysis typically focuses on only a standalone gene circuit, and therefore the effects of other genes and heterogeneous microenvironments cannot be included. Second, the exact values of kinetic parameters needed for modelling are largely unavailable, and since the modelling results depend on the estimated parameters, this issue can severely limit the predictive power of the models. Third, most studies do not provide a systematic way to quantify the robustness and plasticity of GRNs.

To address these issues, we here have applied a recently developed mathematical modelling algorithm, *r*andom *ci*rcuit *per*turbation (RACIPE) [22,23], to explore the robust dynamical behaviours of a proposed core GRN governing stemness. RACIPE was developed to elucidate the robust gene expression patterns (also referred to as gene states) and generic features of transcriptional regulatory networks [22–24] (figure 1). Unlike traditional approaches, RACIPE takes the topological information of a network as the only input, and generates an ensemble of mathematical models. Each mathematical model is simulated by the same set of chemical rate equations with different sets of kinetic parameters representing heterogeneous signalling and epigenetic states. The parameters of each model can differ by up to one or two orders of magnitude and are generated randomly under a specially designed sampling scheme (e.g. half-functional rule). Multiple initial conditions are used for each model in order to identify all possible steady-state solutions. The parameters and the corresponding stable steady-state solutions generated by the ensemble models are collected and subject to statistical analysis, by which the robust gene states are elucidated. It has been shown that RACIPE successfully identifies the gene states enabled by circuit motifs (i.e. toggle-switch-like circuit, repressilator, coupled toggle-switches) and GRNs governing epithelial–mesenchymal transition (EMT) and B cell development [22–24].

Here, we use RACIPE to analyse a core stemness GRN which contains eight master regulatory TFs involved in stem cell differentiation (figure 2*a*). We find that applying RACIPE to the stemness GRN can recapitulate the gene expression patterns of mouse ESCs at 32-cell and 64-cell stages. These gene expression patterns are conserved robust features of the stemness GRN but disappear when the network topology is randomized. Furthermore, through performing *in silico* perturbation analysis, we show that (i) the presence of external signals can exclude the accessibility of some of the gene states enabled by the stemness GRN; (ii) the stemness GRN has a hidden hierarchical structure, which enables a two-step decision-making process with the Oct4/Cdx2 motif functioning as the first decision-making module and the Gata6/Nanog motif as the second one. In summary, we

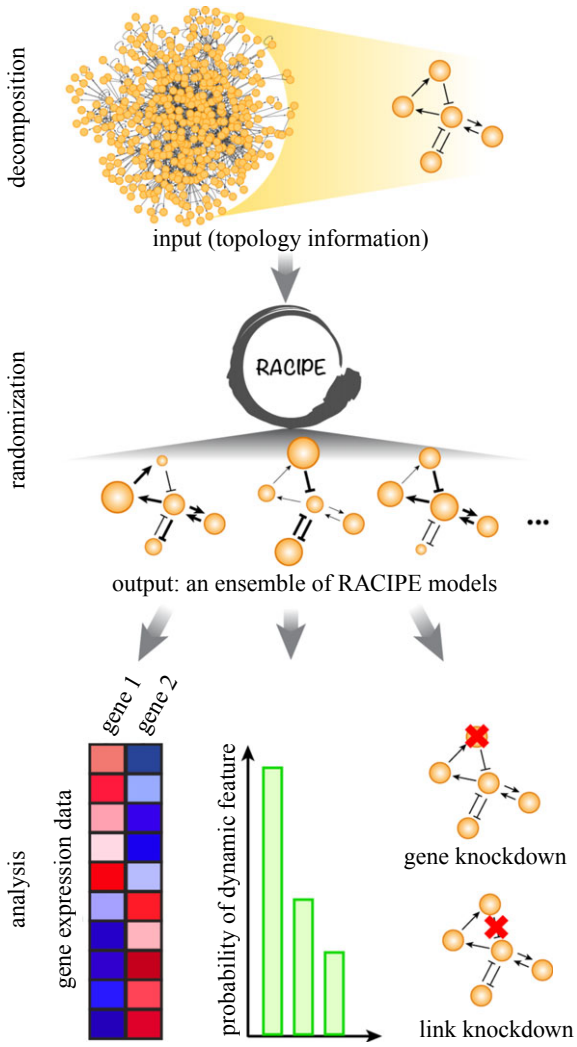

Figure 1. Schematic illustration of random circuit perturbation (RACIPE). The gene regulatory network governing a specific cellular function can be divided into two parts—a core decision-making module and the rest functioning as input signals to the core. Through randomization, RACIPE generates an ensemble of mathematical models, each of which is simulated by the same set of chemical rate equations but with randomly sampled parameters. The simulation results of the model ensemble are subject to statistical analysis, such as hierarchical clustering analysis (HCA), and *in silico* gene/link perturbation analysis.

demonstrate that the robustness of stemness regulation is mainly determined by the topology of the stemness GRN, and RACIPE can be applied straightforwardly to elucidate the hierarchical decision-making of stem cell differentiation.

Before proceeding, it is important to clarify the connection between RACIPE model ensembles and actual cell populations. Ever since the seminal work of Elowitz and coworkers on 'external' noise, it has become clear that cell populations exhibit distributions of phenotypic properties due to differing single-cell values of the kinetic parameters governing the genetic circuitry. RACIPE, with one slight change from its baseline algorithm (see Material and methods), also can be used to create a nominal cell population, making use of the null model choice of uniform uncorrelated parameter distributions over relevant physiological ranges. This will then need to be modified to account for correlations among the parameters. It is nonetheless useful to exhibit results for this uncorrelated ensemble, just to provide a guide as to what experimental data would indicate the need for a more biologically accurate formulation. Thus, we provide below initial results for distributions with the aforementioned goal.

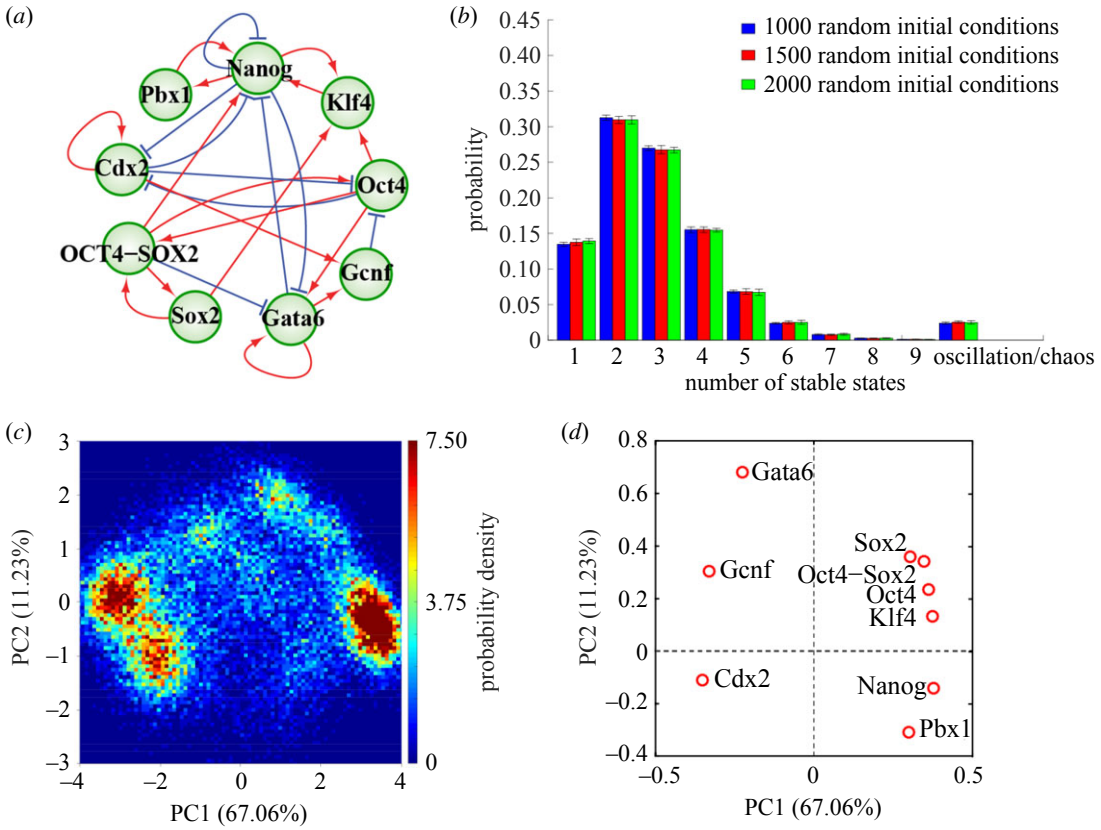

Figure 2. The RACIPE method uncovers robust gene states allowed by the stemness GRN. (*a*) Diagram of the core gene regulatory network governing stem cell differentiation. Red arrows represent excitatory regulation; blue bar-headed arrows represent inhibitory regulation. (*b*) Probability distribution of the number of stable steady states generated by 10 000 RACIPE models. Different colours represent different cases characterized by different numbers of initial conditions (blue: 1000 times, red: 1500 times and green: 2000 times) that are used to simulate each RACIPE model. Each case was repeated 10 times to estimate the mean and the standard deviation of the distribution. (*c*) 2D probability density map of the RACIPE-predicted gene expression profiles projected onto the 1st and 2nd principal component (PC1 and PC2) axes. (*d*) Contribution of each gene to PC1 and PC2. The PCs were obtained by performing the principal component analysis (PCA) using the gene expression profiles from all 10 000 RACIPE models.

2. Material and methods

2.1. Mathematical modelling of the stemness GRN

In this study, the dynamical behaviour of the stemness GRN (network diagram illustrated in figure 2*a*, details illustrated in §3.1) was studied by RACIPE. RACIPE is a free open source software distributed under (Apache 2.0) license and can be downloaded from GitHub (https://github.com/simonhb1990/RACIPE-1.0). Specifically, the RACIPE procedure creates an ensemble of models in each of which the temporal dynamics of the eight TFs (Oct4, Sox2, Cdx2, Gata6, Gcnf, Pbx1, Klf4 and Nanog) and the protein complex (Oct4–Sox2) are simulated by a set of ordinary differential equations (ODEs) accounting for their production, degradation and regulatory interactions. The transcriptional regulation between these TFs is modelled by the shifted Hill function [25]. The full details about the mathematical model and the implementation can be found in electronic supplementary material, §S1 and table S1. To account for the binding/unbinding reactions between the TFs Oct4 and Sox2, which is not captured in the original RACIPE, we generalize the algorithm by modifying the rate equations to capture association and disassociation of the protein complex Oct4–Sox2. Full details regarding generalized RACIPE and its implementation can be found in electronic supplementary material, §S2 and table S2.

2.2. Analysis of the RACIPE-generated gene expression data and experimental data

The details about the normalization of RACIPE-generated gene expression data can be found in electronic supplementary material, §S3. Clustering analysis has been performed on the normalized RACIPE-generated data to identify the gene expression patterns (electronic supplementary material, §S4). The details of the method to compare RACIPE-generated data with experimental data can be found in electronic supplementary material, §S5 and S6.

2.3. The RACIPE population model

To generate a null model of a cell population replete with phenotypic heterogeneity, we modify the RACIPE method to properly count states found by the algorithm. In particular, when a given set of parameters gives rise to multiple steady states, each steady state is weighted by the proportion of the initial conditions leading to that particular state.

3. Results

3.1. RACIPE identifies robust gene states enabled by the stemness GRN

We integrate the master gene regulators of stemness characterized by previous studies [12,13,26], and construct a core stemness GRN. The GRN is composed of eight TFs (Oct4, Sox2, Cdx2, Gata6, Gcnf, Pbx1, Klf4 and Nanog) and one protein complex (Oct4–Sox2) (figure 2*a*). Due to the complexity of the GRN, elucidating its dynamical behaviours can be difficult through traditional modelling approaches. Here, we use RACIPE to identify the robust gene states enabled by the stemness GRN. As all regulatory links in the stemness GRN are transcriptional except for the binding/unbinding process between the TFs Oct4 and Sox2, for the simplicity, we initially

model the temporal dynamics of the Oct4–Sox2 complex in the same manner as the other TFs. Later, we will generalize RACIPE to include binding/unbinding reactions and verify that none of the results change in any meaningful ways.

In our approach, instead of finding a representative set of kinetic parameters, we randomly generate 10 000 sets of parameters (i.e. 10 000 RACIPE models) within their given biologically reasonable ranges. For each model, we numerically solve the governing ODEs with 1000 random initial conditions so as to thoroughly identify all possible stable steady-state solutions. These 1000 initial conditions give rise to 1000 solutions, based on which we identify the number of distinct stable states and their corresponding gene expression profiles. We show that 1000 initial conditions are sufficient, as increasing the number of initial conditions to 1500 or 2000 generates consistent probability distributions of the number of stable states (figure 2*b*) and stable states (electronic supplementary material, figure S1a) of the 10 000 models. Similarly, we show that 10 000 models are sufficient to capture the robust gene states of the stemness GRN, as the consistency of the top 14 clusters of gene expression profiles from three distinct 10 000 sets of parameters is statistically significant (electronic supplementary material, figure S1b,c).

For the majority of RACIPE models (approx. 98%), the stemness GRN allows one to six stable steady states (figure 2*b*). There are rare occasions (less than 1%) where RACIPE models generate more than six stable steady states. Since they are not statistically significant, we excluded these data for further analysis. The circuit also has approximately 2% chance of having oscillatory or chaotic dynamics (e.g. time-dependent dynamics), which are not the focus of this manuscript and therefore are excluded from further analysis. We collected the gene expression profiles from all 10 000 RACIPE models and constructed a data matrix, where each column represents a gene and each row represents a stable steady-state solution. These data resemble experimental gene expression data, thus inspiring us to apply similar statistical methods.

Since the kinetic parameters of the circuit are randomized with large variations (up to one or two orders of magnitude), one might expect that the gene expression profiles from different models would be very different. Strikingly, we found that the gene expression profiles can be segregated into only a few clusters when projected onto two independent components by the commonly used principal component analysis (PCA) (figure 2*c*). Regarding the first principal component (PC1), Oct4, Sox2 and Nanog contribute positively while Cdx2 and Gata6 contribute negatively (figure 2*d*), indicating an anti-correlation of the activity between these two sets of TFs, which is consistent with the experimental observations [27]. We have also shown that different ranges for parameter randomization and different types of the random sampling distribution (Uniform or Gaussian) in RACIPE all generate largely consistent probability density maps (electronic supplementary material, figures S2 and S3), as compared with the one shown in figure 2*c*.

3.2. RACIPE-generated gene expression profiles are consistent with experimental observations and match single-cell gene expression data of mouse ESCs at 32-cell and 64-cell stages

To identify the pattern of RACIPE-generated gene expression profiles, we applied hierarchical clustering analysis (HCA) to the RACIPE-generated gene expression data. We found that the RACIPE-generated data form several major clusters, representing different gene expression patterns (figure 3*a*,*b*). We found that most of the TFs exhibit bi-modal distributions, which indicates the up/downregulation of these TFs is associated with different phenotypes during stem cell differentiation (figure 3*a*, electronic supplementary material, figure S4). To evaluate how well RACIPE can recapitulate the characteristic gene expression of various cellular phenotypes during stem cell differentiation, we compared the RACIPE-generated gene expression profiles with those observed experimentally. We found that the most significant RACIPE-generated gene states, as determined by the reproducibility in different ensembles of RACIPE models, recapitulate the gene expression patterns measured by experiments (figure 3*c*, electronic supplementary material, figures S1 and S5). We compare the RACIPE-generated data with the gene expression data of single mouse embryo cells at various stages during development [8] (figure 3*a*, electronic supplementary material, §S6, figures S6–S8). Interestingly, the RACIPE-generated gene states match those during the late stage of embryo development (greater than or equal to 32 cells), where totipotent cells start to differentiate into trophectoderm (TE) and inner cell mass (ICM) [28], but not those from the early stage (less than or equal to 16 cells). The experimental verified gene states—$Cdx2^{Hi}$, $Gata6^{Hi}/Nanog^{Hi}/Oct4^{Hi}/Sox2^{Hi}$, $Nanog^{Hi}/Oct4^{Hi}/Sox2^{Hi}$ and $Gata6^{Hi}/Oct4^{Hi}/Sox2^{Hi}$—are among the most probable gene states identified by RACIPE (figure 3*b*). Notably, the matching between RACIPE data and experimental data is statistically significant (electronic supplementary material, figures S7 and S8).

Indeed, the RACIPE-generated gene expression patterns can characterize multiple developmental stages. For example, the first cell fate determination during embryonic development happens at the blastocyst stage when the ICM and TE are formed [28]. Oct4 and Sox2 were reported to be expressed throughout ICM (**State 1**, the state numbers are indicated in figure 3*c*) [29,30]. At the early stage of ICM differentiation, Gata6 and Nanog exhibit coexpression (**State 6**) [31,32], but Nanog, Oct4 and Sox2 are required for cells to commit to epiblast and reach the ground state of pluripotency (**States 1** and **12**) [29,30]. Further differentiation of mESCs into mesendoderm requires Nanog and Oct4 but not Sox2 (**State 8**) [33]. Upon Gata6 induction, mESCs exhibit a step-wise pluripotency factor disengagement, starting with downregulation of Nanog and Esrrb, then Sox2, and finally Oct4, along with a step-wise coexpression of extraembryonic endoderm (ExEn) genes (**State 5**) [30]. On the other hand, downregulation of Oct4 induces the differentiation of mESCs into trophoblast characterized with high expression of Cdx2 and Gata6 (**State 2**) [26,34]. Overexpression of Cdx2 is sufficient to generate proper trophoblast stem cells (**State 3**) [8,27]. TE could further differentiate into ExEn, where Cdx2, Gata6 and Sox2 can be all expressed (**States 7, 9, 11** and **15**) [27,35,36]. A summary of the correspondence between RACIPE-generated gene states and experimental observed gene expression profiles during development can be found in electronic supplementary material, figure S5.

These results suggest that RACIPE can identify the gene expression patterns of various cellular phenotypes especially those of the late developmental stages and also characterize potential additional gene states during stem cell differentiation.

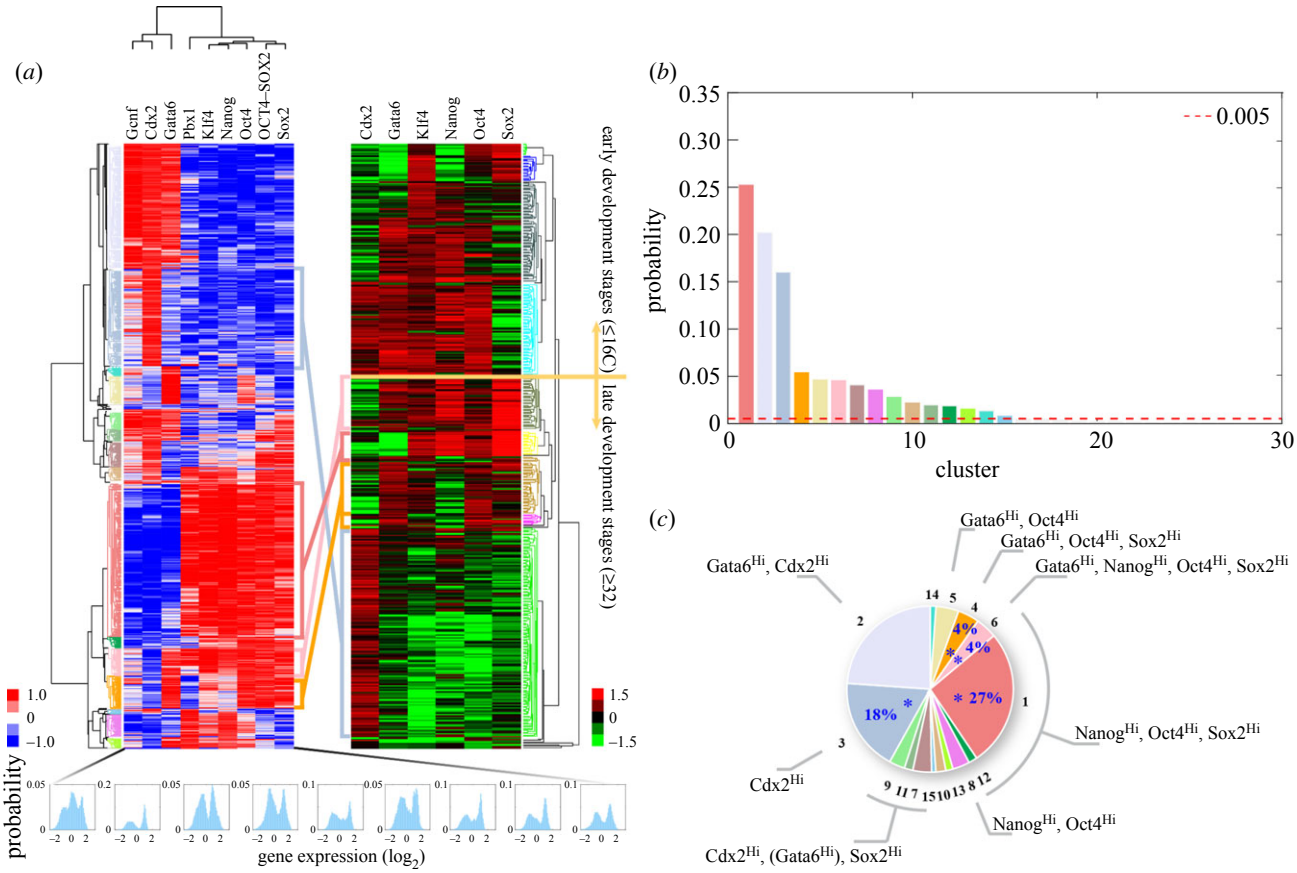

Figure 3. Comparison of the RACIPE-generated gene expression profiles and single-cell gene expression data of mouse embryo. (*a*) Robust clusters (gene states, coloured hierarchical trees) were identified for both datasets by unsupervised HCA. Four RACIPE-predicted gene states match those from the late stage single-cell gene expression data. The histogram of the predicted expression levels for each gene is shown at the bottom (blue, 50 bins in each histogram). In both heat maps, each column represents a gene; each row represents the gene expression profile of a stable steady state of a RACIPE model (left) or that for a single cell (right). (*b*) A total of 54 gene clusters (only show 30 here) were identified by HCA. With a minimum probability cut-off of 0.005, we identified 15 clusters, referred to as major gene states. The colouring scheme for these 15 clusters is consistent with that used in (*a*), and the other clusters are shown in grey. (*c*) The characteristic gene expression of each gene state ranked by the likelihood in the RACIPE models. The four gene states that match the experimental data are highlighted by blue asterisks and are shown with their likelihoods. (The method to classify the gene states in the presence of external signal can be found in electronic supplementary material, §S8.)

3.3. The topology of the stemness GRN determines the robust gene states

To further investigate the role of the topology of the stemness GRN in maintaining the robust gene states, we compare the stemness GRN with two types of its randomized versions (Type I including 10 networks and Type II including 10 networks) with randomly generated connections among genes. Both Type I and Type II randomized networks preserve the total number of inward and outward links for each gene, and the binding of Oct4 and Sox2 (electronic supplementary material, figure S9). For the Type I randomized network, we also keep the same numbers of excitatory and inhibitory inward links for each gene. We then apply RACIPE to these two types of randomized networks and compare their dynamic behaviours with those of the actual stemness GRN.

Neither Type I nor Type II randomized networks can generate the aforementioned robust gene states and recapitulate the experimentally observed gene expression features. Compared to the stemness GRN, both Type I and Type II randomized networks are much more likely to generate oscillatory or chaotic dynamics but not multi-stable states for each RACIPE model. Instead, both randomized networks are more likely to have only one stable state for each RACIPE model (electronic supplementary material, figure S10). When the gene states from all RACIPE models are combined, the histogram of each gene expression generated by the stemness GRN typically exhibits multi-modal distributions but those generated by the randomized networks mainly exhibit mono-modal distribution (electronic supplementary material, figure S11). In addition, we also find that it is difficult to cluster the gene expression data generated by the randomized networks (electronic supplementary material, figure S12), partly because the stemness GRN has much higher local density of the RACIPE-generated gene expression data relative to the randomized networks (electronic supplementary material figure S11). Furthermore, the RACIPE-predicted gene expression profiles of the stemness GRN are significantly better than those of the randomized networks in recapitulating the experimental observation of the gene expression features of various stages during development, and especially the single-cell gene expression data of mouse embryo (figure 4 and electronic supplementary material, figure S13). Notably, even though randomized networks can contain a similar number of self-excitatory and mutually inhibitory feedback loops, they still cannot compete with the stemness GRN in achieving multi-stable behaviours and in recapitulating the experimental observations. This indicates that the

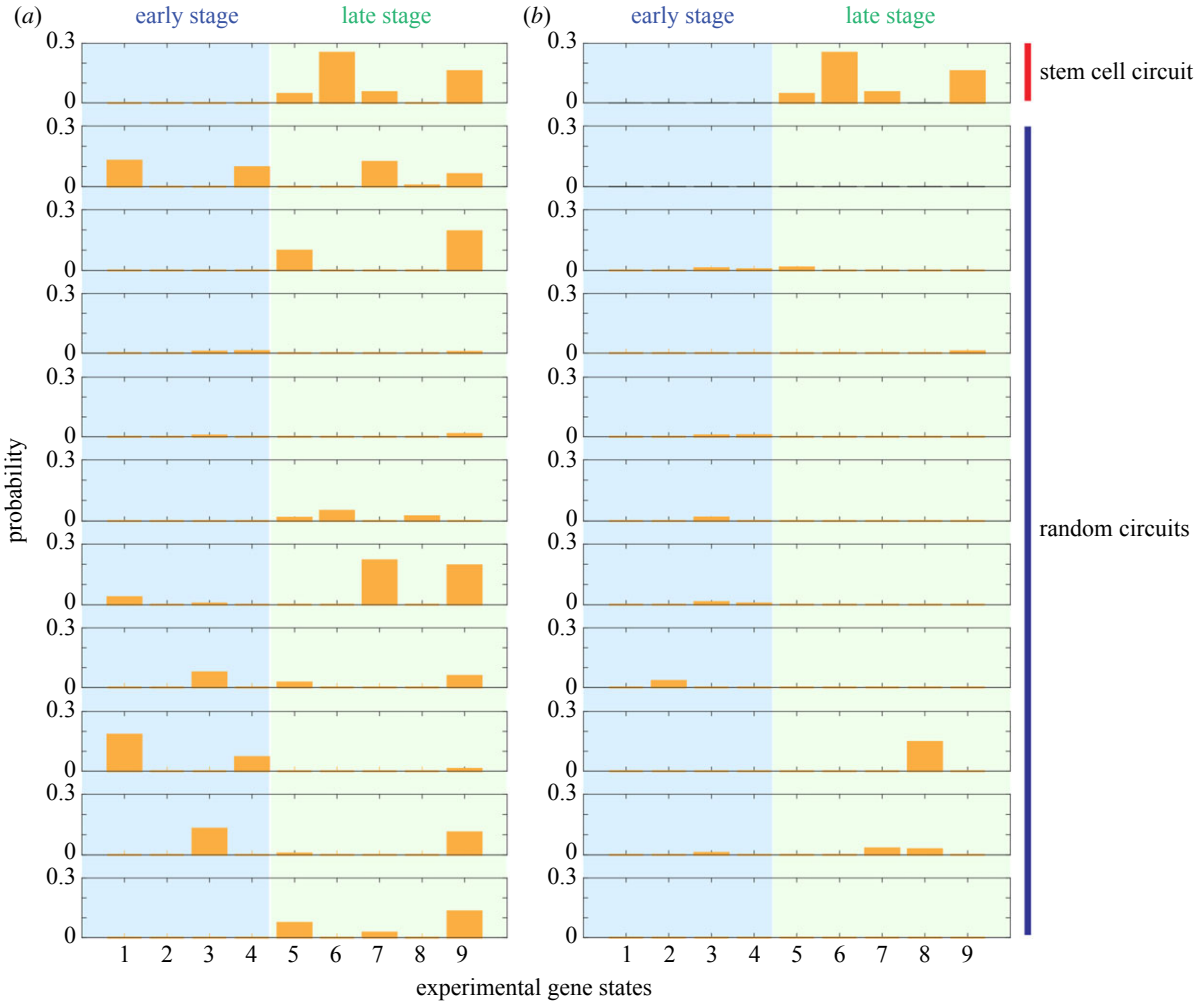

Figure 4. Comparison between the stemness GRN and the randomized networks (10 for Type I (*a*) and 10 for Type II (*b*)). Percentage of the RACIPE-predicted gene expression data matching each experimental gene state shown in figure 3*a* (right) for the stemness GRN and random networks. The details of the 10 Type I randomized networks and the 10 Type II networks can be found in electronic supplementary material, figure S9.

stemness GRN may indeed exhibit special properties beyond what can be expected based on counting of simple motifs [37]. In summary, these results show that the topology of the stemness GRN has been well-evolved to be robust in regulating stem cell differentiation in the presence of perturbations.

3.4. RACIPE identifies the key parameters representing key stimuli that promote transitions between different phenotypes during stem cell differentiation

In addition to characterizing the gene expression profiles of different cell phenotypes, RACIPE enables us to uncover the key parameters representing physiological conditions that mediate the phenotypic transitions during stem cell differentiation. From the output of RACIPE simulations, we can identify the kinetic parameters that are significantly changed between gene states (electronic supplementary material, §S7, figure 5*a*). We will discuss two examples of this feature in the following paragraphs.

To characterize the most differential kinetic processes between gene state 1 (representing the pluripotent epiblast stage) and gene state 2 (representing the trophoblast stage), we quantify the change of the mean values of each parameter in gene state 2 relative to gene state 1. We identify the parameters whose values increase the most (e.g. the degradation rate of Sox2 and the degradation rate of Oct4–Sox2), and the parameters whose values decrease the most (e.g. the production rate of Sox2, the degradation rate of Gata6 and the threshold of Gata6 self-activation) in gene state 2 relative to gene state 1 (figure 5*b*). The results indicate the transition from gene state 1 to gene state 2 is characterized by increased degradation of Sox2, decreased production of Sox2 and accumulation of Gata6. This indication is consistent with the experimental result that loss of Sox2 results in the differentiation of ESCs as characterized by the upregulation of trophoblast markers [38].

To characterize the stimuli that can promote the transition from gene state 1 (representing the pluripotent epiblast stage) to gene state 6 (representing the early stage of ICM differentiation), we quantify the changes of each parameter in gene state 6 relative to gene state 1. We identify the parameters whose values increase most (the maximum production rate of Gata6) and the parameters whose values decrease most (the threshold for Gata6 self-activation and fold-change of the inhibition of Gata6 by Nanog) in gene state 6 relative to gene state 1 (figure 5*a*,*c*). The results indicate that transition from gene state 1 to gene state 6 requires the upregulation of Gata6 that can be accomplished by either increasing Gata6 production or decreasing the inhibition of Gata6 by Nanog. This simulation result is again consistent with the experimental

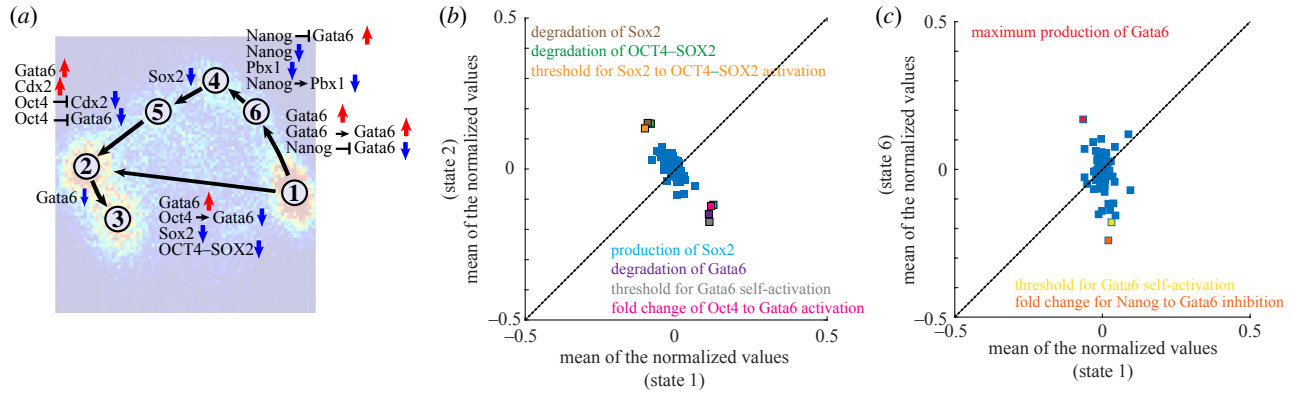

Figure 5. Key parameters that are involved in the transitions among certain gene states. (*a*) A summary of the results depicted on top of the probability density map (figure 2*c*) of the RACIPE-generated gene expression data. In (*a*), along with each transition, the key parameters that have shifted the most have been marked. Red up arrows represent upregulation and blue down arrows represent downregulation. The mean of the normalized values for each parameter for the two corresponding gene states (*x*-axis for the first gene state, and *y*-axis for the second gene state) are shown in (*b*) (states 1 and 2) and (*c*) (states 1 and 6). The change of the parameters values between any of the transitions shown in (*a*) can be found in electronic supplementary material, figure S14.

observation that Gata6 is highly expressed at the early stage of ICM differentiation [31]. The full details regarding the key parameters shifts between any two of the gene states, as shown in figure 5*a*, can be found in electronic supplementary material, figure S14.

In summary, RACIPE can identify the parameters that differ the most between different gene states, such as the production/degradation rates and the threshold/fold-change of a regulatory link. Consequently, the biological process represented by these parameters can be considered as the regulatory target needed to drive cells to undergo certain phenotypic transitions.

3.5. The stemness GRN exhibits a hierarchical decision-making structure

RACIPE also enables *in silico* perturbation analysis of the stemness GRN, including knocking out genes and removing links, by which we can understand the role of the knocked-out gene or the removed link in the dynamical behaviours of the stemness GRN. Here, we perform two types of perturbation analyses, knocking out a gene each time or removing a regulatory link each time (electronic supplementary material, §S8 and S9). In both types of perturbation analyses, we quantify the change of the probability distribution of the multi-stability exhibited by the stemness GRN using the Kullback–Leibler (KL) divergence (electronic supplementary material, §S10).

Through analysing the gene knockout results, we found that the knockout of TFs Cdx2, Oct4, Sox2 and the complex OCT4–SOX2 leads to the most significant changes in the probability distribution of the multi-stable behaviour of the GRN (figure 6*a*). Strikingly, removal of the regulatory links among these specific TFs also leads to the most significant changes in the probability distribution of the number of stable states (figure 6*b*). These TFs and the regulatory links between them indeed form a sub-network, representing the first decision-making module, referred to as the Oct4/Cdx2 module (figures 6*c*,*d*). The rest of the TFs and regulatory links form a sequential second sub-network, referred to as the Gata6/Nanog module (figure 6*d*). The RACIPE simulation results are consistent with experimental observations showing that the TFs Oct4 and Cdx2 govern the commitment of totipotent cells to either the ICM ($Oct4^{high}$) or the TE ($Cdx2^{high}$) lineages and the Gata6/Nanog module governing the commitment of ICM cells to either epiblast ($Nanog^{high}$) or primitive endoderm ($Gata6^{high}$) lineages [39].

To compare the behaviours of these two sub-networks, we apply RACIPE to each of them and then compared the RACIPE-generated gene expression profiles with those generated by applying RACIPE to the full network. In other words, we want to evaluate how the dynamic behaviours of the two sub-networks change upon removal of the regulatory links connecting them. We find that the gene expression profiles determined by the first decision-making module Oct4/Cdx2 is conserved while those determined by the second module Gata6/Nanog are largely disrupted, upon the removal of the regulatory links connecting these two modules (electronic supplementary material, figure S15). The results here support the hierarchical structure of these two decision-making modules.

3.6. Gene states of the stemness GRN can become inaccessible upon external signals

As stem cell differentiation is a cascading event induced by signalling, we would like to analyse how the stemness GRN responds to various external signals (e.g. signals acting on the TFs in the network but getting no feedback from the circuit) by RACIPE. The effects of external signals on a certain gene are simulated by scaling the production rate of that gene to 50-times larger (representing excitatory signal) or smaller (representing inhibitory signals). We then calculated the probability distribution of gene states upon the imposition of external signals (excitatory or inhibitory) on each gene. As we showed before, without any external signals, the stemness GRN allows 15 robust gene states, referred to as the wild-type (WT) (figure 6*e*). Relative to the WT, upregulation of Cdx2 restricts most RACIPE models to acquire the stable states with $Cdx2^{Hi}$ (representing the TE stage) [27], while upregulation of Oct4 restricts most RACIPE models to acquire the states with $Oct4^{hi}$ (representing the ICM stage) [31,39] (figure 6*e*). After cells reach the ICM stage, additional signals acting on Nanog and Gata6 can convert the GRN largely to either the $Nanog^{Hi}$ state (representing the epiblast stage) or the $Gata6^{Hi}$ state (representing the primitive endoderm stage) [39] (figure 6*e*). These simulation results indicate that external signals often do not create new gene states but instead make a subset of gene states more accessible and the rest less

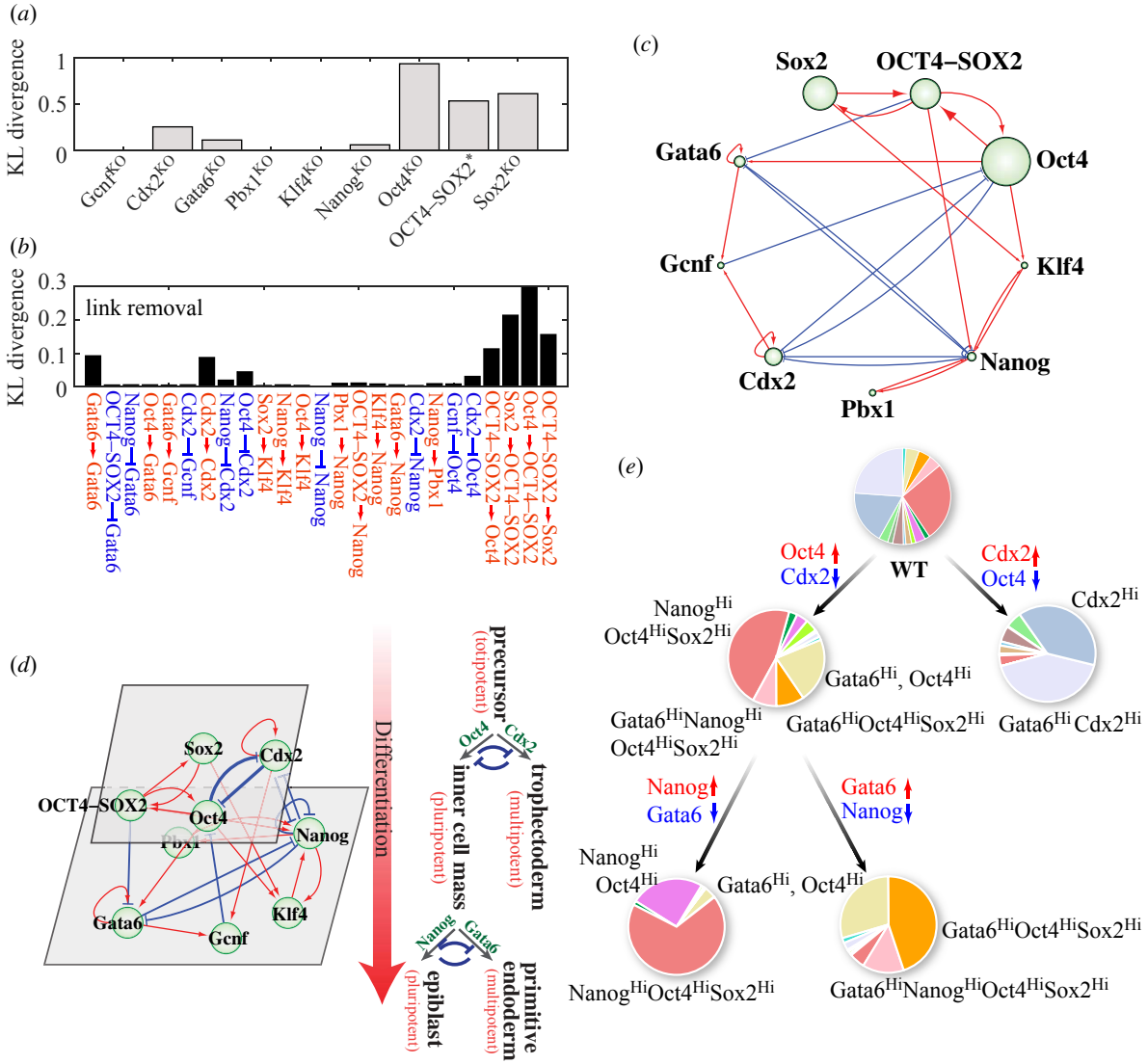

Figure 6. Hierarchical structure of the stemness GRN inferred from the perturbation analysis. (*a*) The Kullback–Leibler (KL) divergence between the probability distributions of the number of stable states for each RACIPE model computed before and after the knockout (KO) of each gene. 'Oct4–Sox2*' represents the removal of the protein complex Oct4–Sox2. (*b*) Similar to (*a*), but the KL divergences are between the distributions before and after removal of each regulatory link. (*c*) Schematic diagram of the stemness GRN highlighting the important genes and regulatory links. The larger the gene element and the thicker the regulatory link, the more important the component to the network behaviour, as inferred from the analyses in (*a*) and (*b*). (*d*) The hierarchical structure of the stemness GRN (left) is consistent with the two-step decision-making of mouse embryonic development (right). (*e*) The roadmap of stem cell differentiation inferred from the RACIPE simulations. All the original RACIPE models (WT) were treated by activating (↑) or inhibiting (↓) the maximum production rate of the corresponding genes by 50-fold. The probability of different gene states is proportional to the area in the pie chart.

accessible as also observed in figure 5 in our previous work [23]. In other words, a step-wise administration of external signals is able to restrict the gene expression of the stemness GRN to specific cellular states.

As already mentioned, the usefulness of RACIPE does not require that the RACIPE ensemble of models corresponds to the actual cell population. Getting the latter feature correct requires more input information specifically regarding parameter correlations, information that goes beyond the topology of the governing GRN. Hence, we have made no predictions regarding the actual population structure, for example, the percentage of cells in each of the allowed states as a function of external conditions. It is nonetheless useful to present some results under the null assumption that no such correlations exist, so as to provide benchmark data to which real populations can be compared and thereby to detect the possible need for this additional input information.

With this in mind, we wish to quantify the effects that various signals would have on the population heterogeneity as generated by the *uncorrelated* RACIPE approach. To accomplish this, we employed information entropy theory (electronic supplementary material, §S11). As one set of parameters can give rise to multiple stable state solutions, a weight factor proportional to the percentage of initial conditions leading to a stable state solution was multiplied to that solution. We systematically simulated the external signals acting on each TF by scaling the maximum production rate of that TF from 1/100 to 100 of the base level, representing inhibitory signals and excitatory signals respectively. For each TF, we apply RACIPE to the stemness GRN considering 20 different scenarios representing 10 inhibitory and 10 excitatory signals with varying strengths on that TF. We then applied the entropy-based index to quantify the heterogeneity of the ensemble stable states generated by RACIPE in each scenario for each TF (electronic supplementary material, figure S16). As we show that the WT stemness GRN exhibits the highest entropy, e.g. highest heterogeneity, and external signals that either upregulate or downregulate the maximum production

rate of a TF usually decreases the entropy, leading to decreased heterogeneity of the nominal cell population. The decrease of heterogeneity partially results from the limited access to only a few gene states instead of all due to external signals.

Altogether, our results suggest that RACIPE can explore the possible roles of external signals in the dynamic behaviours of the stemness GRN. These external signals may indeed restrict the gene states that can be accessible and may thereby lower population entropy. Indeed, we expect that it should be a generic property for cellular networks that the number of the stable states will decrease when strong perturbation signals are imposed.

3.7. Toward generalizing RACIPE by including the binding/unbinding details

RACIPE provides a straightforward way to identify the robust dynamic behaviours of the stemness GRN. As we discussed before, RACIPE was originally developed for transcriptional regulation. In the stemness GRN, in addition to the majority of the links representing transcriptional regulation, there is one binding/unbinding process, between the TFs Oct4 and Sox2. We have, therefore, extended the RACIPE framework to explicitly model this binding/unbinding process to analyse how that may affect the network behaviour. We performed a parallel analysis of the stemness GRN using the updated RACIPE including binding/unbinding details (referred to as RACIPE-wb) (figure 7*a*, electronic supplementary material, figures S17 and S18). We observed consistent gene states by RACIPE-wb (figure 7*b*, electronic supplementary material, figures S19–S21) relative to those acquired by RACIPE (figure 2*c*). We show that the RACIPE-wb generated gene expression profiles are quantitatively consistent with the RACIPE-generated ones (figure 7*c*). By RACIPE-wb, we performed perturbation analysis by knocking out genes and removing regulatory links one by one. Consistent with the result by RACIPE (figure 6*a*,*b*), we found that knocking out the TFs Oct4, Sox2, Cdx2 or Oct4–Sox2 or removing the regulatory links among these genes have the most pronounced effects on the multi-stable behaviours of the stemness GRN (figures 7*d*,*e*). Indeed, RACIPE-wb amplifies the differences observed under link removal relative to RACIPE (figures 6 and 7). Specifically, blockade of the binding between Oct4 and Sox2 (KL divergence = 0.32) in RACIPE-wb is so pronounced that it is equivalent to a full removal of the protein complex Oct4–Sox2 in the stemness GRN (KL divergence = 0.32). Blockade of the unbinding of Oct4–Sox2 (KL divergence = 1.57) in RACIPE-wb is so pronounced that it is approximately equivalent to knocking out both Oct4 (KL divergence = 0.95) and Sox2 (KL divergence = 0.64) (figure 7*d*,*e*). The effects of link removal for the rest of the regulatory interactions are in general smaller in RACIPE-wb than those in the original RACIPE. But the overall trends are similar for both modelling algorithms. Among these interactions, we found that removal of the inhibitory link from Cdx2 to Oct4 exhibits the highest KL divergence (electronic supplementary material, figure S22). The result by RACIPE-wb also indicates a similar hierarchical decision-making structure of the stemness GRN with the Oct4/Cdx2 module forming the first decision-making module and the rest forming the second decision-making module. In summary, when the detailed binding/unbinding processes are considered, the RACIPE-wb characterized dynamic behaviours of the stemness GRN remain consistent with RACIPE characterized ones.

4. Discussion

By applying RACIPE to the stemness GRN, we have identified a significant number of distinct gene states. The four most probable gene states generated by RACIPE match the gene expression patterns of single mouse embryonic cells at 32-cell and 64-cell stages quantitatively (figure 3*a*), and 11 RACIPE-generated gene states are consistent with experimental results qualitatively (electronic supplementary material, figure S5). We also elucidate a hierarchical structure of the stemness GRN by RACIPE, which is consistent with experimental observations. The fact that our results reproduce known findings is a positive indicator of RACIPE's validity in analysing the dynamical behaviours of GRNs in general, including both transcriptional regulation and binding/unbinding processes.

There are two important insights generated by RACIPE. First, multiple RACIPE-generated gene states are associated with one experimentally observed stage. This is because the developmental stages are often characterized by expression of a few genes and these few genes can have similar expression patterns in various gene states. Our findings thus predict that future work using the expression of additional stemness genes to characterize developmental stages may help further delimit possible sub-developmental stages. Second, RACIPE can identify the specific parameters that differ the most between different gene states, such as the production/degradation rates of genes and the threshold/fold-change of regulatory links. Consequently, the biological process represented by these parameters can be considered as the regulatory targets needed to drive cells to undergo certain phenotypic transitions.

We found that the external signals acting on the stemness TFs can restrict the stemness GRN to acquire only certain gene states corresponding to differentiated cell phenotypes. The result may be related to a popular interpretation of Waddington's epigenetic landscape [40,41] for stem cell differentiation. Of course, our RACIPE ensemble is only a lowest-order approximation to an actual cell population. However, taking this approximation seriously, at least in the absence at the moment of a more complete picture, our RACIPE simulation results indicate an interesting perspective. Specifically, a stem cell population, instead of consisting of cells in a highly specific 'stemness' state, might instead be regarded as a heterogeneous population of cells with variable gene expression patterns corresponding to a mixture of differentiated lineage with a distinctive gene expression pattern (figure 8) [42]. Cells with high cell potency are plastic and able to interconvert into the various cell states stochastically by both the intrinsic factors (gene expression noise, a fast process) and the extrinsic factors (transient epigenetic regulation and cell signalling, a slow process). However, when cells are subject to stable perturbations by external signals, they lose the capacity to access certain cellular states, therefore, making the population less heterogeneous, i.e. having smaller information entropy, and differentiated. Our view is consistent with the observation in experiments that the stem cell progenitors of either totipotency or pluripotency have highly heterogeneous gene expression, and several cell sub-populations of differentiated types, called lineage priming, have been identified in cell culture [43–45]. Some of the new gene states proposed by RACIPE share similarities with the gene states reported in [46]. For example, the state 2 characterized by $GATA6^{high}CDX2^{high}NANOG^{low}$ proposed by RACIPE (figure 3) is consistent with the intermediate state 2 during transdifferentiation reported in [46].

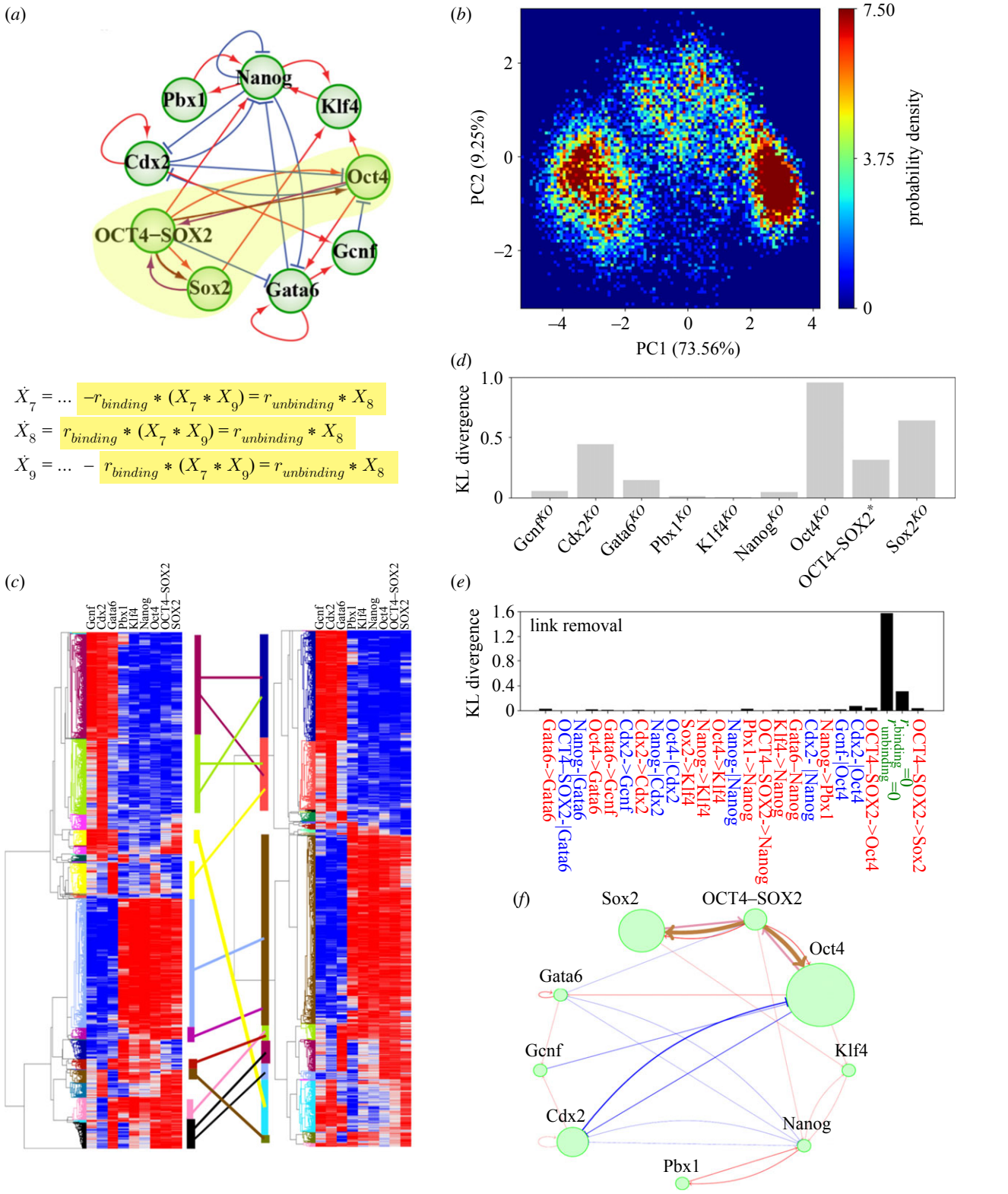

Figure 7. Analysis results elucidated by applying the extended RACIPE (RACIPE-wb) to the stemness GRN reveals consistent characterization. (*a*) Top panel: Diagram of the core stemness GRN highlighting the binding interactions between Oct4, Sox2, and the OCT4–SOX2 complex. Bottom panel: The main changes in the mathematical equations simulating the dynamics of Sox2 (X9), Oct4 (X7), OCT4–SOX2 (X8) to capture their binding/unbinding interactions. The full equations for RACIPE-wb are listed in electronic supplementary material, §S2. (*b*) The 2D probability density map of the results for the RACIPE-wb model projected onto the first two principal components. (*c*) Unsupervised HCA for RACIPE and RACIPE-wb, left and right, respectively. Clusters were identified using a probability cut-off of 0.005. The lines between clusters show the majority of clusters from the original RACIPE are also present in RACIPE-wb. Additionally, some of the clusters obtained by using RACIPE-wb were seen to be over/underrepresented as compared to results using the original RACIPE framework; the number of solutions belonging to that cluster is shown by the coloured vertical bars on the left and right of the middle which correspond the dendrograms of HCA for RACIPE and RACIPE-wb, respectively. (*d*) The KL divergence of RACIPE-wb distributions before and after knocking out a gene. 'Oct4–Sox2$^{*\prime}$ represents the removal of the protein complex Oct4–Sox2. (*e*) The KL divergence of RACIPE-wb distributions before and after the removal of a regulatory link. Also included are the blocking of binding between Oct4 and Sox2 and the blocking of the unbinding of OCT4–SOX2. (*f*) A schematic diagram depicting the relative importance of each gene and link as inferred by the analysis in (*d–e*).

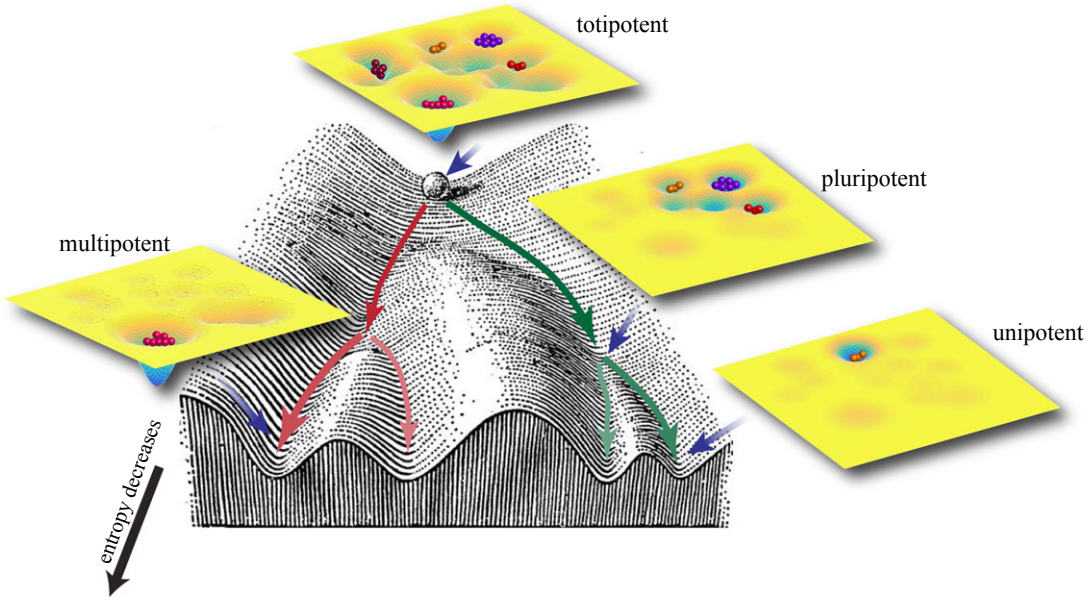

Figure 8. Schematic illustration of the revised Waddington's epigenetic landscape for stem cell differentiation. For each cell potency, the accessible cell types are shown by the attractors. Stem cells are induced by external signals toward differentiation along the valleys (highlighted by arrow lines with different colours) with the decrease of cell potency.

We showed that RACIPE-generated gene expression patterns recapitulate the gene expression profiles of mouse embryos at late developmental stages (greater than or equal to 32 cells) while not those at early stages. This may be due to the incompleteness of the decision-making network of stemness. Although the TFs in the stemness GRN have been shown to be the master regulators governing stem cell differentiation by experimental studies, and the stemness GRN is largely consistent with stemness regulatory networks proposed by other studies [12,13,47], it is possible that there are other important molecular regulators that are not included here. An experimentally validated GRN could be constructed by combining genomics data such as ChIP-Seq with biochemistry experiments [48–50]. However, it still remains a challenge to reliably construct reasonably large GRNs. Notably, RACIPE mainly considers the large variations in parameters without including the effect of internal gene expression noise, which can be critical in cellular dynamics. Although we do not expect the inclusion of stochastic effects to change in any important way the possible states of the system can acquire, including the effect of noise will become important if we want to focus on dynamical processes such as transitions between those gene states. Some early efforts have already been made in this direction [51,52].

To conclude, by applying RACIPE to a core stemness GRN, we showed that the network topology plays an essential role in cell fate decision-making during stem cell differentiation. This result is analogous to the findings from protein structure modelling, where conformational motions have been found to be determined by the overall molecular shape [50] and protein folding process by native residue contacts [50,53]. RACIPE allows the interrogation of the robust dynamical behaviours of GRN by parametric randomization, from which we can identify the operating principles underlying the GRN functions.

Ethics. This article does not present research with ethical considerations

Data accessibility. The gene expression data of single mouse embryo cells as shown in figure 3*a* can be accessed via reference [8].

Authors' contributions. Conceptualization: M.L., H.L., J.N.O., D.J. Data acquisition and analysis: B.H., M.L., M.G., D.J. Data interpretation and manuscript writing: B.H., M.L., H.L., J.N.O., D.J. All authors gave final approval for publication and agree to be held accountable for the work performed therein.

Competing interests. We declare we have no competing interests.

Acknowledgements. This work was supported by the National Science Foundation (NSF) Center for Theoretical Biological Physics (NSF PHY-2019745) and NSF grant nos. PHY-1522550, DMS-1361411 and CHE-1614101. J.N.O. is a CPRIT Scholar in Cancer Research sponsored by the Cancer Prevention and Research Institute of Texas (CPRIT). D.J. is supported by a training fellowship from the Gulf Coast Consortia, on the Computational Cancer Biology Training Program (CPRIT grant no. RP170593). M.L. is supported by a start-up fund from The Jackson Laboratory, by the National Cancer Institute of the National Institutes of Health under Award number P30CA034196, and by the National Institute of General Medical Sciences of the National Institutes of Health under award number R35GM128717. M.G. is supported by the NSF GRFP number 1842494. This work was supported in part by the Big-Data Private-Cloud Research Cyberinfrastructure MRI-award funded by NSF under grant CNS-1338099 and by Rice University.

References

1. Young RA. 2011 Control of the embryonic stem cell state. *Cell* **144**, 940–954. (doi:10.1016/j.cell.2011.01.032)
2. Basilico S, Göttgens B. 2017 Dysregulation of haematopoietic stem cell regulatory programs in acute myeloid leukaemia. *J. Mol. Med.* **95**, 719–727. (doi:10.1007/s00109-017-1535-3)
3. Ilic D, Ogilvie C. 2017 Concise review: human embryonic stem cells—what have we done? What are we doing? Where are we going? *Stem Cells* **35**, 17–25. (doi:10.1002/stem.2450)
4. Chambers I *et al.* 2007 Nanog safeguards pluripotency and mediates germline

development. *Nature* **450**, 1230–1234. (doi:10.1038/nature06403)

5. Hayashi K, de Sousa Lopes SMC, Tang F, Lao K, Surani MA. 2008 Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. *Cell Stem Cell* **3**, 391–401. (doi:10.1016/j.stem.2008.07.027)
6. Kumar RM *et al*. 2014 Deconstructing transcriptional heterogeneity in pluripotent stem cells. *Nature* **516**, 56–61. (doi:10.1038/nature13920)
7. Singer ZS, Yong J, Tischler J, Hackett JA, Altinok A, Azim Surani M, Cai L, Elowitz MB. 2014 Dynamic heterogeneity and DNA methylation in embryonic stem cells. *Mol. Cell* **55**, 319–331. (doi:10.1016/j.molcel.2014.06.029)
8. Guo G, Huss M, Tong GQ, Wang C, Li Sun L, Clarke ND, Robson P. 2010 Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Dev. Cell* **18**, 675–685. (doi:10.1016/j.devcel.2010.02.012)
9. Kashyap V, Rezende NC, Scotland KB, Shaffer SM, Persson JL, Gudas LJ, Mongan NP. 2009 Regulation of stem cell pluripotency and differentiation involves a mutual regulatory circuit of the NANOG, OCT4, and SOX2 pluripotency transcription factors with polycomb repressive complexes and stem cell microRNAs. *Stem Cells Dev.* **18**, 1093–1108. (doi:10.1089/scd.2009.0113)
10. Boyer LA *et al*. 2005 Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–956. (doi:10.1016/j.cell.2005.08.020)
11. Yeo J-C, Ng H-H. 2013 The transcriptional regulation of pluripotency. *Cell Res.* **23**, 20–32. (doi:10.1038/cr.2012.172)
12. 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**, e1003380. (doi:10.1371/journal.pcbi.1003380)
13. Zhang B, Wolynes PG. 2014 Stem cell differentiation as a many-body problem. *Proc. Natl Acad. Sci. USA* **111**, 10 185–10 190. (doi:10.1073/pnas.1408561111)
14. Chickarmane V, Troein C, Nuber UA, Sauro HM, Peterson C. 2006 Transcriptional dynamics of the embryonic stem cell switch. *PLoS Comput. Biol.* **2**, e123. (doi:10.1371/journal.pcbi.0020123)
15. Herberg M, Roeder I. 2015 Computational modelling of embryonic stem-cell fate control. *Development* **142**, 2250–2260. (doi:10.1242/dev.116343)
16. Huang S, Guo Y-P, May G, Enver T. 2007 Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Dev. Biol.* **305**, 695–713. (doi:10.1016/j.ydbio.2007.02.036)
17. Chickarmane V, Peterson C. 2008 A computational model for understanding stem cell, trophectoderm and endoderm lineage determination. *PLoS ONE* **3**, e3478. (doi:10.1371/journal.pone.0003478)
18. Kalmar T, Lim C, Hayward P, Muñoz-Descalzo S, Nichols J, Garcia-Ojalvo J, Martinez Arias A. 2009 Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells. *PLoS Biol.* **7**, e1000149. (doi:10.1371/journal.pbio.1000149)
19. 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. (doi:10.1371/journal.pone.0011238)
20. Herberg M, Kalkan T, Glauche I, Smith A, Roeder I. 2014 A model-based analysis of culture-dependent phenotypes of mESCs. *PLoS ONE* **9**, e92496. (doi:10.1371/journal.pone.0092496)
21. Huang S. 2011 Systems biology of stem cells: three useful perspectives to help overcome the paradigm of linear pathways. *Phil. Trans. R. Soc. B* **366**, 2247–2259. (doi:10.1098/rstb.2011.0008)
22. Huang B, Lu M, Jia D, Ben-Jacob E, Levine H, Onuchic JN. 2017 Interrogating the topological robustness of gene regulatory circuits by randomization. *PLoS Comput. Biol.* **13**, e1005456. (doi:10.1371/journal.pcbi.1005456)
23. Huang B, Jia D, Feng J, Levine H, Onuchic JN, Lu M. 2018 RACIPE: a computational tool for modeling gene regulatory circuits using randomization. *BMC Syst. Biol.* **12**, 74. (doi:10.1186/s12918-018-0594-6)
24. Jia D, Jolly MK, Harrison W, Boareto M, Ben-Jacob E, Levine H. 2017 Operating principles of tristable circuits regulating cellular differentiation. *Phys. Biol.* **14**, 035007. (doi:10.1088/1478-3975/aa6f90)
25. Lu M, Jolly MK, Levine H, Onuchic JN, Ben-Jacob E. 2013 MicroRNA-based regulation of epithelial–hybrid–mesenchymal fate determination. *Proc. Natl Acad. Sci. USA* **110**, 18 144–18 149. (doi:10.1073/pnas.1318192110)
26. Chen L, Yabuuchi A, Eminli S, Takeuchi A, Lu C-W, Hochedlinger K, Daley GQ. 2009 Cross-regulation of the Nanog and Cdx2 promoters. *Cell Res.* **19**, 1052–1061. (doi:10.1038/cr.2009.79)
27. Niwa H, Toyooka Y, Shimosato D, Strumpf D, Takahashi K, Yagi R, Rossant J. 2005 Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell* **123**, 917–929. (doi:10.1016/j.cell.2005.08.040)
28. Marikawa Y, Alarcón VB. 2009 Establishment of trophectoderm and inner cell mass lineages in the mouse embryo. *Mol. Reprod. Dev.* **76**, 1019–1032. (doi:10.1002/mrd.21057)
29. Silva J *et al*. 2009 Nanog is the gateway to the pluripotent ground state. *Cell* **138**, 722–737. (doi:10.1016/j.cell.2009.07.039)
30. Wamaitha SE, del Valle I, Cho LTY, Wei Y, Fogarty NME, Blakeley P, Sherwood RI, Ji H, Niakan KK. 2015 Gata6 potently initiates reprograming of pluripotent and differentiated cells to extraembryonic endoderm stem cells. *Genes Dev.* **29**, 1239–1255. (doi:10.1101/gad.257071.114)
31. Schrode N, Saiz N, Di Talia S, Hadjantonakis A-K. 2014 GATA6 levels modulate primitive endoderm cell fate choice and timing in the mouse blastocyst. *Dev. Cell* **29**, 454–467. (doi:10.1016/j.devcel.2014.04.011)
32. Ohnishi Y *et al*. 2014 Cell-to-cell expression variability followed by signal reinforcement progressively segregates early mouse lineages. *Nat. Cell Biol.* **16**, 27–37. (doi:10.1038/ncb2881)
33. Thomson M, Liu SJ, Zou L-N, Smith Z, Meissner A, Ramanathan S. 2011 Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. *Cell* **145**, 875–889. (doi:10.1016/j.cell.2011.05.017)
34. Hay DC, Sutherland L, Clark J, Burdon T. 2004 Oct-4 knockdown induces similar patterns of endoderm and trophoblast differentiation markers in human and mouse embryonic stem cells. *Stem Cells* **22**, 225–235. (doi:10.1634/stemcells.22-2-225)
35. Avilion AA, Nicolis SK, Pevny LH, Perez L, Vivian N, Lovell-Badge R. 2003 Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev.* **17**, 126–140. (doi:10.1101/gad.224503)
36. Strumpf D, Mao C-A, Yamanaka Y, Ralston A, Chawengsaksophak K, Beck F, Rossant J. 2005 Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. *Development* **132**, 2093–2102. (doi:10.1242/dev.01801)
37. Tripathi S, Kessler DA, Levine H. In press. Biological regulatory networks are minimally frustrated. *Phys. Rev. Lett*.
38. Masui S *et al*. 2007 Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat. Cell Biol.* **9**, 625–635. (doi:10.1038/ncb1589)
39. Wang H, Dey SK. 2006 Roadmap to embryo implantation: clues from mouse models. *Nat. Rev. Genet.* **7**, 185–199. (doi:10.1038/nrg1808)
40. Waddington CH. 1957 The strategy of the genes: a discussion of some aspects of theoretical biology. New York, NY: Macmillan.
41. Wang J, Zhang K, Xu L, Wang E. 2011 Quantifying the Waddington landscape and biological paths for development and differentiation. *Proc. Natl Acad. Sci. USA* **108**, 8257–8262. (doi:10.1073/pnas.1017017108)
42. Loh KM, Lim B. 2011 A precarious balance: pluripotency factors as lineage specifiers. *Cell Stem Cell* **8**, 363–369. (doi:10.1016/j.stem.2011.03.013)
43. Morgani SM, Canham MA, Nichols J, Sharov AA, Migueles RP, Ko MSH, Brickman JM. 2013 Totipotent embryonic stem cells arise in ground-state culture conditions. *Cell Rep.* **3**, 1945–1957. (doi:10.1016/j.celrep.2013.04.034)
44. Piotrowska-Nitsche K, Perea-Gomez A, Haraguchi S, Zernicka-Goetz M. 2005 Four-cell stage mouse blastomeres have different developmental properties. *Development* **132**, 479–490. (doi:10.1242/dev.01602)
45. Chang HH, Hemberg M, Barahona M, Ingber DE, Huang S. 2008 Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* **453**, 544–547. (doi:10.1038/nature06965)
46. Li C, Wang J. 2013 Quantifying Waddington landscapes and paths of non-adiabatic cell fate decisions for differentiation, reprogramming and transdifferentiation. *J. R. Soc. Interface* **10**, 20130787. (doi:10.1098/rsif.2013.0787)
47. Li C, Wang J. 2013 Quantifying cell fate decisions for differentiation and reprogramming of a human stem cell network: landscape and biological paths.

PLoS Comput. Biol. **9**, e1003165. (doi:10.1371/journal.pcbi.1003165)

48. Blais A. 2005 Constructing transcriptional regulatory networks. *Genes Dev.* **19**, 1499–1511. (doi:10.1101/gad.1325605)
49. Wang P, Qin J, Qin Y, Zhu Y, Wang LY, Li MJ, Zhang MQ, Wang J. 2015 ChIP-Array 2: integrating multiple omics data to construct gene regulatory networks. *Nucleic Acids Res.* **43**, W264–W269. (doi:10.1093/nar/gkv398)
50. Lu M, Ma J. 2005 The role of shape in determining molecular motions. *Biophys. J.* **89**, 2395–2401. (doi:10.1529/biophysj.105.065904)
51. Li C. 2018 Landscape of gene networks for random parameter perturbation. *Integr. Biol.* **10**, 92–99. (doi:10.1039/C7IB00198C)
52. Kohar V, Lu M. 2018 Role of noise and parametric variation in the dynamics of gene regulatory circuits. *NPJ Syst. Biol. Appl.* **4**, 40. (doi:10.1038/s41540-018-0076-x)
53. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. 1995 Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **21**, 167–195. (doi:10.1002/prot.340210302)