Taylor & Francis
Taylor & Francis Group

RESEARCH PAPER

Check for updates

A mathematical model exhibiting the effect of DNA methylation on the stability boundary in cell-fate networks

Tianchi Chen [a,*], M. Ali Al-Radhawi [b,*], and Eduardo D. Sontag [a,b,c]

aDepartment of Bioengineering, Northeastern University, Boston, MA, USA; bDepartment of Electrical and Computer Engineering, Northeastern University, Boston, MA, USA; cLaboratory of Systems Pharmacology, Program in Therapeutic Science, Harvard Medical School, Boston, MA, USA

ABSTRACT

Cell-fate networks are traditionally studied within the framework of gene regulatory networks. This paradigm considers only interactions of genes through expressed transcription factors and does not incorporate chromatin modification processes. This paper introduces a mathematical model that seamlessly combines gene regulatory networks and DNA methylation (DNAm), with the goal of quantitatively characterizing the contribution of epigenetic regulation to gene silencing. The 'Basin of Attraction percentage' is introduced as a metric to quantify gene silencing abilities. As a case study, a computational and theoretical analysis is carried out for a model of the pluripotent stem cell circuit as well as a simplified self-activating gene model. The results confirm that the methodology quantitatively captures the key role that DNAm plays in enhancing the stability of the silenced gene state.

Introduction

Cell-fate determination in developmental biology is a multi-step biological process in which cellular functions are modified or specialized, ultimately resulting in a differentiated or proliferative state. A central role in this process is played by the dynamics of a core network of genes, often referred to as a *cell-fate network* (CFN). The expression levels of these genes can trigger a cascade of events that determine, in principle irreversibly, the fate of a given cell along a specific lineage. The metaphor of marbles rolling down a hill whose shape is shaped by a CFN (the *Waddington landscape* [1,2]) provides a way to visualize this process. Well-known examples of CFNs are the PU.1/GATA.1 gene regulatory network in haematopoietic progenitors [3,4], the pluripotent stem cell network [4,5], and a circuit including SNAIL, miR-34, ZEB1, and miR-200 that regulates epithelial-mesenchymal transitions in tumour metastasis [6].

Phenotypes associated to CFNs had classically been seen as irreversible. However, in their pioneering work [7], Takahashi and Yamanaka artificially induced a pluripotent state in mouse somatic cells through a process of overexpression of Oct3/4, Sox2, c-Myc, and Klf4. This success notwithstanding, simply overexpressing genes has been experimentally found to be grossly inefficient [8]. Hence, there has been a great interest in obtaining a quantitative theoretical understanding of CFNs in order to guide the process of reprogramming and increase its efficiency [9,10]. Such a quantitative understanding could have a huge impact in the field of regenerative medicine and stem cell therapy [11,12].

Traditionally, CFNs have been modelled within the wider theoretical framework of Gene Regulatory Networks (GRNs) [13,14]. A GRN is defined as a set of genes, each expressing a protein. The expressed proteins can act as Transcription Factors (TFs) by binding to the various promoters in the network to inhibit or enhance expression of the corresponding genes [15,16]. When used to model a CFN, a GRN must be able to display *multistability*. This

---

means that the long-term expression levels of genes can settle on multiple distinct steady state attractors, each corresponding to a distinct cell lineage. A well-known multistable GRN is the *toggle switch* [17] whose architecture consists of two mutually inhibiting genes. Similar architectures can also give rise to multistability and have been subject of great interest [2].

Despite their versatility and variety of applications, GRN models of CFNs do not typically account for epigenetic effects, such as DNA methylation, histone modifications, or chromatin remodelling. There have been both theoretical [18–20] and experimental [21] works aiming at the understanding of how epigenetic regulation and gene regulation interact. It is well known that CFNs employ epigenetic regulation as a mechanism to ensure the irreversibility of the cell lineage [22]. DNA methylation (DNAm), in particular, has been well studied in the context of developmental CFNs [23]. DNAm at the transcriptional start sites (TSSs) is a highly heritable, hard to reverse, and robust silencer of genes [24–26]. In comparison to its well-known silencing function at the regulatory regions, less is known about DNAm's role at intragenic regions [27].

In this work, we develop mathematical models of GRNs that incorporate DNAm at TSSs, and thus can quantitatively explain its effect on gene silencing. As a metric to quantify the gene silencing abilities of DNAm, we consider the shift in the stability boundary of the basin of attraction (BoA) of the silenced steady state. More precisely, we define the BoA percentage (BoAp) of a steady state as the volume fraction of a predetermined region of the state space that is occupied by the BoA. Equipped with this metric, one can then computationally and theoretically study the sensitivity of the BoAp to parameters such as DNAm rates or the time scale of DNAm.

We carry out this program computationally, illustrating it with a three-gene model of the pluripotent stem cell network, and quantifying how DNAm effectively increases the BoAp compared to a standard GRN. In addition, and in order to gain understanding, we also consider an ideal single self-activating gene, for which

a more theoretical study is possible. We view our work as a first step towards integrating epigenetic mechanisms into the standard GRN paradigm studied in systems biology. Furthermore, we envision our model (and its derivatives) as computational guides uncovering the possible behaviours of a network and evaluating the feasibility of the interventions that can be used for regulation. For instance, as we will discuss later, the DNAm rate in our model is proportional to the abundance of DNMT, and hence our mathematical framework can quantitatively predict the effect of DNMT on the relative stability of the silenced (methylated) and active steady states. This would allow our model to function as a computational aid to design the reprogramming protocol in the context of the PSCC. We will elaborate further on the biological implications of our model in the Discussion section.

Basic biological concepts and definitions

DNA methylation as epigenetic regulation

DNAm is one of the main epigenetic regulation mechanisms studied in modern molecular cell biology. Methylation plays a crucial role [28] in understanding the dynamics of gene silencing. It is associated with methylation of Cytosine-Phosphate-Guanine (CpG) islands, which are regions of DNA with a high G +C content and a high frequency of CpG dinucleotides relative to the bulk genome [29]. Methylation of promoter regions directly affects TF-promoter binding kinetics, and is often associated with transcriptional repression [24–26,30].

The current understanding of the DNAm cycle is shown in Figure 1. As depicted, cytosines at DNA promoter regions are initially methylated by DNA methyltransferase (DNMT) to become 5-methylcytosine (5mC). The methylated gene promoter is further oxidized to 5-Hydroxy Methylcytosine (5hmC). The oxidized form of the cytosine can further be oxidized by the TET protein into the 5fC and 5caC forms. The cycle closes by a TET/TDG (thymine DNA glycosylase)/ BER (base excision repair)-dependent pathway that restores the unmodified cytosine state [31,32].
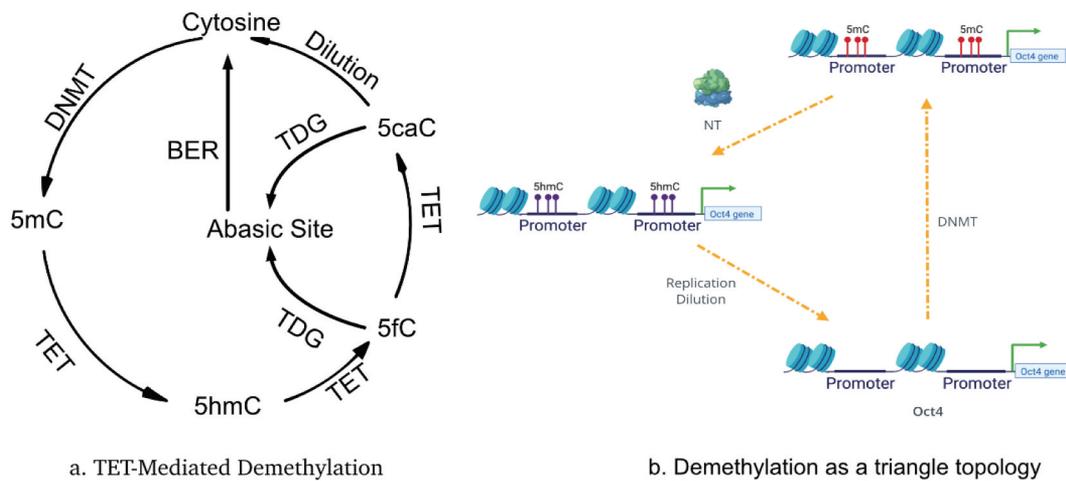
Figure 1. The DNA methylation cycle as currently understood [31,32]. (a) The TET-mediated cytosine DNA demethylation cycle starts with transferring a methyl group onto the C5 position of the cytosine to form 5-methylcytosine (5mC) by the DNA (cytosine-5)-methyltransferase (mainly DNMT3A/B for de novo methylation). The 5mC is then further iteratively oxidized by TET enzymes to form 5hmC, 5fC, and 5caC. The 5fC and 5caC are then recognized by TDG that yields an abasic site that is repaired by BER and returns to the unmodified cytosine state. (b) A simplified de-methylation cycle for Oct4 in the PSCC. Due to the fact that 5mC and 5hmC are two leading abundant oxidized forms of cytosine, we can effectively reduce the full cycle to a three state model.

Experiments have shown that hydroxymethylation and formylation are relatively stable cytosine modifications in genomic DNA of both dividing and nondividing cells [33,34]. The 5hmC and 5fC forms in this enzymatic oxidation process are observed to be comparatively longer lived, though still transient, states. Moreover, it also has been shown [31] that the pool size of the transient oxidized cytosine 5hmC is significantly larger than the other forms of oxidized cytosines (5fC, 5caC). Therefore, in this work, we further simplify the DNA demethylation cycle to the cycle in Figure 1 (b) which consists of just three major steps, which we will refer to as the '*triangle topology.*' The cycle starts with an initial de novo methylation on the unbounded gene promoter, which is followed by a TET oxidization enzymatic process oxidizing 5mC to 5hmC. The last stage corresponds to a replication-dependent dilution bringing 5hmC back to the unbound promoter state.

The underlying DNAm dynamics is crucial for understanding how epigenetics regulates CFNs. DNAm plays a central role not only in early embryogenesis, but also in maintaining the correct pattern of methylation along DNA in somatic cells. As a model system of great interest, we review next how the pluripotent stem cell circuit (PSCC) is affected by the DNAm cycle.

The pluripotent stem cell circuit (PSCC)

Some of the key Transcription Factors (TFs) that are identified as being crucial to the PSCC include Nanog, Oct4, Sox2, Klf4, and the TET protein family [7,31]. Many of such TFs are 'pioneering' transcription factors that are able to bind directly to the condensed chromatin. According to the induction experiments in [7], the overexpression of the aforementioned TFs can lead to the induction of the pluripotent state. For the PSCC, the role of each TF in the GRN has been extensively studied [35]. Much research has been devoted to elucidating the interplay between the Nanog, Oct4 and TeT1 proteins [36,37]. Oct4 is central to the pluripotency machinery, has been very well studied [38], and it can *by itself* reprogram a differentiated cell into pluripotency [39]. Other experiments regarding the induction of the pluripotent state in PSCCs have highly heterogeneous results [8,40]. More recent research [41,42] has shown that, at a mechanistic level, Nanog is able to guide TET protein to bind to a particular region of the methylated region on the chromatin via the formation of a Nang-TET compound that is actively involved in the DNA demethylation process.

Notably, DNAm of the Oct4 gene plays a key role in the regulation of the PSCC. In Yamanaka's lab early works [7,43], it was observed that the
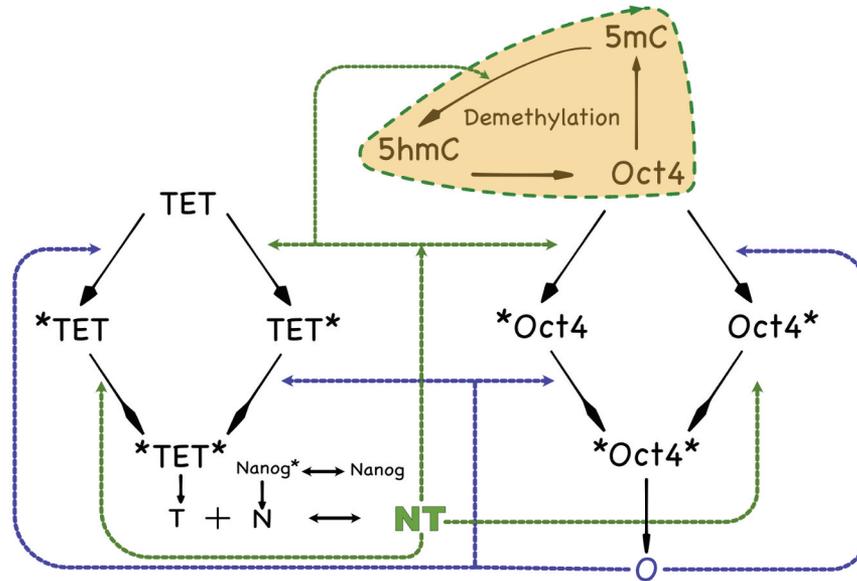
The Pluripotent Stem Cell Circuit (PSCC)

Figure 2. A detailed diagram of the PSCC network showing N, T, O interactions. TET, Oct4, Nanog denote the unbound genes. *TET and *Oct4 denote NT binding to the respective promoter sites, TET* and Oct4* denote Oct4 bound to the respective promoter sites, and *TET*, *Oct4* denote both NT and Oct4 binding to the respective promoter sites. The black dashed line indicates binding locations of the Nanog-guided TET protein complex NT, while the blue dashed line indicates binding locations of Oct4. The coloured shaded area emphasizes the DNA demethylation cycle.

promoters of Oct4 (and other pluripotency-related genes) are unmethylated in embryonic stems (ES) cells, while they are methylated in mouse embryonic fibroblasts. Furthermore, it was observed that Oct4 becomes demethylated just prior to reprogramming [44], and that the differentiation of ES cells is associated with the silencing of Oct4 and de novo methylation of its regulatory region [45,46]. Hence, activation of the PSCC requires demethylation of the corresponding promoters [47], which means that understanding DNAm is a prerequisite of understanding the PSCC. However, very few mathematical models have incorporated DNAm in their GRNs. In this work, we model the interplay between epigenetic regulation via the Nanog-guided TET-mediated active demethylation cycle and the role of pioneering transcription factors such as Oct4 and Nanog. A mechanistic view of this interplay helps one understand how emergent cell states can arise phenotypically from the underlying genotypic level interaction. Although the full modelling of epigenetic regulation at the molecular level would involve many additional epigenetic

regulation mechanisms beyond methylation such as histone modifications [31], DNAm plays a major role in the coarse grained picture of the process of the induction of the pluripotent state which we are mostly interested in.

Basin of attraction percentage (BoAp)

Using high-dimensional attractors, and specifically stable steady states, to represent cell fates in the epigenetic landscape is a widely used paradigm for studying cell fate [14,48]. In a biological context, a steady state attractor corresponds to an observable phenotype, e.g., a somatic state with a silenced Oct4 promoter. In the dynamical cellular environment, perturbations can temporarily change the expression levels away from steady-state. Intuitively, the GRN should be able to return to its attractor (phenotype) for a sufficiently small perturbation, while there is a risk of jumping to another attractor (phenotype) for a large perturbation.

By 'perturbation' we mean an abrupt change in the levels of the dynamical variables of interest such as TF abundances and DNAm levels. An attractor that can withstand a large perturbation is more 'stable'. Hence, to characterize the stability boundary of such attractors we use the concept of Basin of Attraction (BoA). The BoA of a given attractor (phenotype) is the set of all perturbed levels that will return back eventually to the steady state attractor. Figure 3 illustrates these ideas.

The concept of BoA has been frequently used by researchers and it is standard in dynamical systems when analysing multi-stable systems; moreover, it has recently been systematically used by the physics community in that context [49]. In order to quantify the BoA of a given attractor, we propose to use the concept of BoA Percentage (BoAp), which is defined as the ratio of the 'size' of the BoA of interest to the total size of the space of interest. A larger BoAp for a given attractor means that its BoA is larger, and hence it is relatively more stable, while a smaller BoAp means that it is relatively less stable.

In our simulation analysis, we adopt this concept to quantify the stability boundary of the somatic cell state. In a bistable one-dimensional model, the BoAp of the smaller steady state s_0 can be defined as the distance between s_0 and the unstable state relative to the length of a segment of interest. However, this is not an accurate measure for high-dimensional systems. Therefore, we adopt a *volumetric* definition for the BoA. We start by fixing a region of interest in the state space. With respect to such a region, BoAp can be defined as follows:

$$\text{BoAp} : = \frac{\text{Volume of the BoA of } s_0}{\text{Volume of the region of interest}}.$$

(1)

Equipped with this definition, we can study the effect of DNAm and demethylation constants on BoAp. The silenced steady state corresponds to a highly methylated promoter. Hence, as the methylation rate γ increases, the BoA of the silenced steady state expands. On the other hand, as the demethylation rate θ increases, the BoA of the silenced steady state shrinks. Our subsequent results will show that the BoAp of the silenced steady state increases as the methylation rate increases, and decreases as demethylation rate increases. Note that we will use the term 'steady state' to mean a 'steady state attractor' from now on.
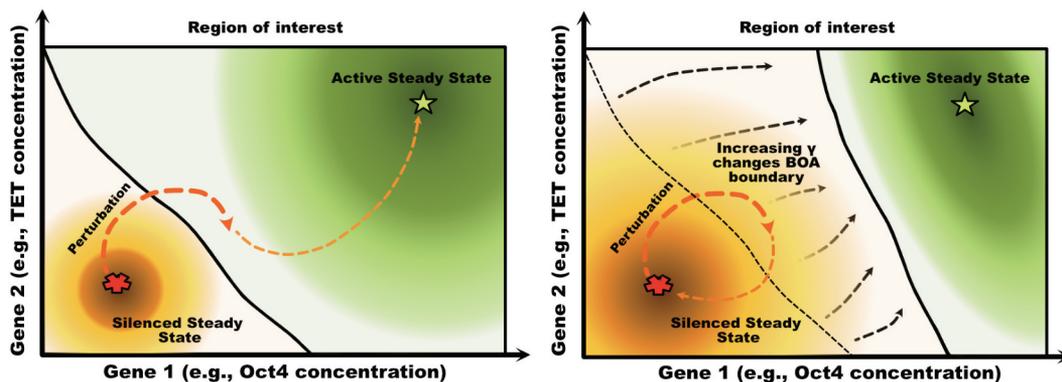

Figure 3. Graphical illustration of how a shift in the boundary separating two regions of attraction can result in phenotype switches under small perturbations. In the left panel, a perturbation of the silenced state leads to a transient configuration that belongs to the basin of attraction of the active steady state, resulting in a phenotype switch to the active steady state (i.e, the somatic state). In the right panel, the boundary has shifted, reflecting that fact that the relative sizes of the two basins of attraction have exchanged dominance. Now the same perturbation keeps the state in the domain of attraction of the silenced state, and the perturbed state eventually returns to that attractor. The shift in boundary could be due, for instance, to an increase in the methylation rate parameter γ.

Results

The PSCC with epigenetic regulation

In our full model description of the PSCC, the network can be viewed as made up of two parts. The first part is the gene-TF regulatory network [50] which is a core underlying network defining how TFs bind and unbind to DNA as shown in Figure 2 without the shaded area. The main regulatory sub-network for the activation of the Oct4 and Nanog genes is depicted within the two-pathway activation scheme in Figure 4. Both the Oct4 and the TET gene promoter binding mechanisms will follow such scheme, which has two pathways that lead to the activation of the gene. Oct4 protein self-binds to the first site while the Nanog-guided TET protein complex binds to the second site. The promoter is active only when both sites are occupied. In this GRN, the Nanog gene is activated by the Oct4 gene. Furthermore, Nanog and TET form a heterodimer which models TET guiding Nanog to the target promoters.

The GRN with DNA demethylation is summarized in Figure 2. The DNA demethylation cycle with the triangle topology can be further simplified into two states, which are shown in Figure 5. The parameter γ is thought of as the effective methylation rate and the parameter θ as the effective demethylation rate. The CRN for the PSCC with a two-state methylation cycle can be written as in Table 1. (For simplicity, we omit explicit variables for mRNA and other intermediate steps such as protein maturation and post-translation modifications.) We model this network deterministically through a system of ODEs. The variables in this system can be thought of as mean numbers for the
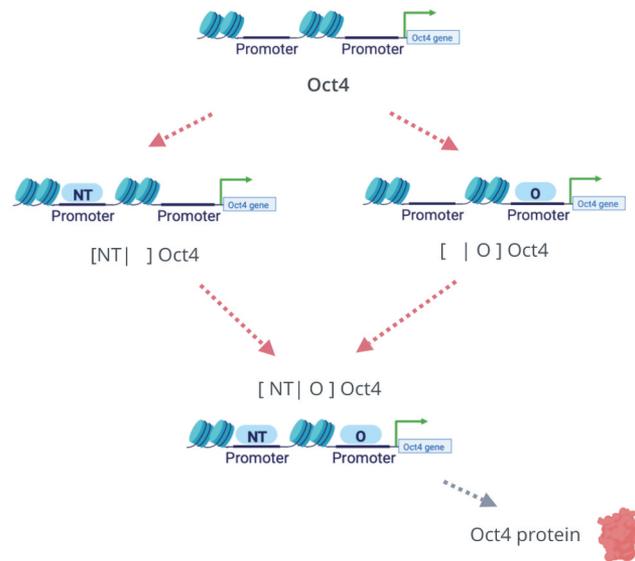

Figure 4. The two-pathway activation scheme for TF-gene binding. In the full model, such an activation scheme describes the promoter binding activation for both Oct4 and TET genes. Considering Oct4 as an example, there are two pathways for the activation of its promoter. Each of NT and Oct4 can independently bind to the promoter of Oct4. When both NT and Oct4 bind to the promoter region of the Oct4 gene, it will be activated.

various species in a population. The full model is described in the Methods section.

Model reduction via quasi-steady state approximation

The full model CRN is computationally intensive to simulate, and it is conceptually difficult to interpret its dynamical behaviour. Hence, we perform a quasi-steady state approximation by distinguishing between two time-scales: the 'fast' time-scale of TF-promoter kinetics (binding/unbinding), and the 'slow' time-scale of protein expression/decay and DNA methylation kinetics. Note that the
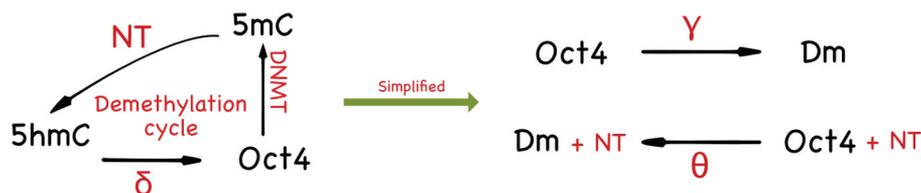

Figure 5. The diagram on the left shows how the triangle topology can be further simplified into a two-species CRN. Since the 5hmC methylation state has a slower dynamics compared to the 5mC methylation state, one may employ a quasi-steady states approximation to reduce the triangle topology to the two-state model depicted on the right.

methylation cycle in our model is a phenomenological reduction of the cycle depicted in Figure 1, which includes a dilution step as well as the coordinated action of many TFs and enzymes. Hence, the methylation cycle is not faster than protein kinetics, which are known to be much slower than TF-promoter kinetics [15]. This is also supported by experimental evidence, as the re-establishment of methylation patterns can take up to 20 minutes [51], and *de novo* methylation kinetics take days which are even slower than protein kinetics as has been observed in various *de novo* methylation studies [52,53]. It can be noted also that inherited methylation patterns have very slow kinetics in general [54,55].

Therefore, we assume that the reaction rates of the promoter-TF binding and unbinding dynamics are faster than the other reaction rates in the CRN Table 1. By applying time scale separation or quasi-steady state approximation, we reduce the full dynamical system model from a 17-dimensional state space to a 4-dimensional state space. The full model ODEs are given in the Methods section. The reduced dynamical system's ODEs can be written as below:

$$\frac{dN}{dt} = m_1 - \delta N + \frac{\alpha_N O}{K_O + O},$$

$$\frac{dT}{dt} = m_2 - \delta T + \frac{\alpha_T NTO}{(NT + K_{nt}K_d)(K_O + O)},$$

$$\frac{dO}{dt} = \frac{-K_{nt}K_d\delta O^2 - \delta NTO^2 + K_O K_{nt}K_d m_3 + \alpha_O NTO + K_{nt}K_d m_3 O}{(NT + K_{nt}K_d)(K_O + O)}$$

$$+ \frac{K_O m_3 NT + m_3 NTO - K_O \delta NTO - K_O K_{nt}K_d \delta O - \alpha_O D_m NTO}{(NT + K_{nt}K_d)(K_O + O)}$$

$$\frac{dD_m}{dt} = \frac{-K_O K_{nt}K_d^2 \gamma D_m + K_O K_{nt}K_d^2 \gamma - \theta K_O N^2 T^2 D_m - \theta N^2 T^2 O D_m}{K_d(NT + K_{nt}K_d)(K_O + O)}$$

$$- \frac{\theta K_O K_{nt}K_d NTD_m + \theta K_{nt}K_d NTOD_m}{K_d(NT + K_{nt}K_d)(K_O + O)}$$

Effect of methylation rate

For the full model, Figure 6 displays how the BoAp changes as one changes the epigenetic reaction rate γ. Each plot in the figure corresponds to one set of parameters, sampled from a wide range. The parameter set for the full model contains the following rates: $K_O, K_{nt}, K_d, a, a_{nt}, a_O, \alpha_T, \alpha_O, \alpha_N, \delta, \gamma, \theta, m_1, m_2$ and m_3. Biologically speaking, changing the reaction rate γ is equivalent to changing the effective methylation rate. Our subsequent study of the single gene model will display the same trend of BoAp's dependence on γ.

The phenotypic somatic state is characterized [36] by low expression of Oct4, Nanog, TET, and methylated promoters. With simulations performed over 6200 parameter sets as shown in Supplementary Figure 1(a), the BoAp vs. γ plot shows that the BoAp is monotonically increasing with respect to γ. In Figure 6, we show a representative sample of these curves. If γ is increased beyond a certain threshold, then bistability is lost and the BoAp of the silenced steady state becomes 100%. Note the different behaviour for intermediate values of γ compared to low or larger values of γ. Another interesting observation is that the BoAp is bounded below by 10% for all parameter sets when γ is greater than $10^{-1.5}$.

Table 1. The CRN model of the PSCC. Reactions have been grouped into eight modules and labelled from R1 to R18.

| Oct4 promoter(R1-R4)		TET promoter (R5-R8)	
	$D_{00}^O + [NT] \overset{K_{nt}*a_{nt}}{\underset{a_{nt}}{\rightleftharpoons}} D_{10}^O$		$D_{00}^T + [NT] \overset{K_{nt}*a_{nt}}{\underset{a_{nt}}{\rightleftharpoons}} D_{10}^T$
	$D_{01}^O + [NT] \overset{K_{nt}*a_{nt}}{\underset{a_{nt}}{\rightleftharpoons}} D_{11}^O$		$D_{01}^T + [NT] \overset{K_{nt}*a_{nt}}{\underset{a_{nt}}{\rightleftharpoons}} D_{11}^T$
	$D_{00}^O + O \overset{K_O*a_O]}{\underset{a_O}{\rightleftharpoons}} D_{01}^O$		$D_{00}^T + O \overset{K_O*a_O}{\underset{a_O}{\rightleftharpoons}} D_{01}^T$
	$D_{10}^O + O \overset{K_O*a_O}{\underset{a_O}{\rightleftharpoons}} D_{11}^O$		$D_{10}^T + O \overset{K_O*a_O}{\underset{a_O}{\rightleftharpoons}} D_{11}^T$
Nanog promoter (R9)	$D_0^N + O \overset{K_O*a_O}{\underset{a_O}{\rightleftharpoons}} D_1^N$	Nanog-TET Dimer (R10)	$N + T \overset{K_d*a}{\underset{a}{\rightleftharpoons}} [NT]$
Protein decay and production (R11-R13)	$N \overset{\delta}{\rightarrow} \emptyset; D_1^N \overset{a_N}{\rightarrow} D_1^N + N$	Basal Rate (R14-R16)	$\emptyset \overset{m_1}{\rightarrow} N$
	$T \overset{\delta}{\rightarrow} \emptyset; D_{11}^T \overset{a_T}{\rightarrow} D_{11}^T + T$		$\emptyset \overset{m_2}{\rightarrow} T$
	$O \overset{\delta}{\rightarrow} \emptyset; D_{11}^O \overset{a_O}{\rightarrow} D_{11}^O + O$		$\emptyset \overset{m_3}{\rightarrow} O$
Methylation (R17)	$D_{00}^O \overset{\gamma}{\rightarrow} D_m$	Demethylation (R18)	$D_m + NT \overset{\theta}{\rightarrow} D_{00}^O + NT$
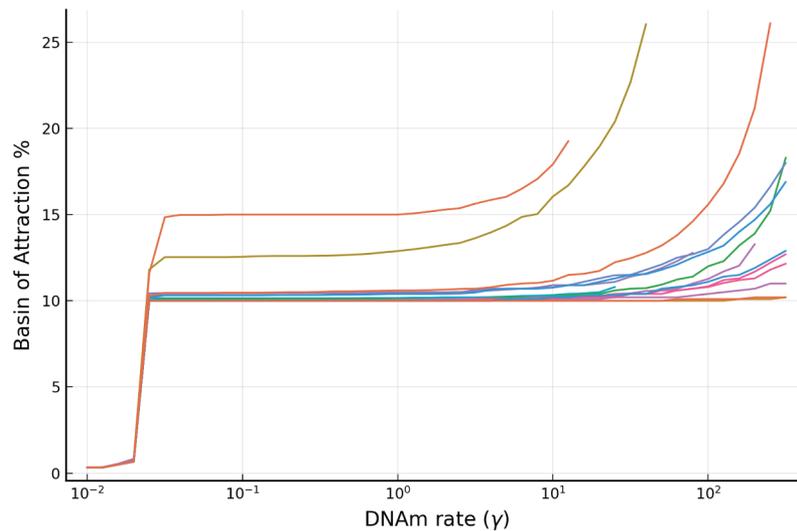
Figure 6. The relative stability (measured by BoAp) of the silenced steady state increases with respect to the DNAm rate (γ). We show simulations for a representative set of different rates for the full model, where extensive simulations are shown in SI-§1.1. Each curve in the figure correspond to a particular set of parameters. Each curve ends when bistability is lost and BoAp jumps to 100%. This behaviour is illustrated in SI-§ 2.1. The simulations also show that, at small γ region, the BoAp exhibits a sudden slowing down at around $\gamma = 10^{-1.5}$.

On the other hand, changing the reaction rate θ means changing the demethylation rate. As we reduced the triangle topology for the DNA demethylation cycle to a two-state cycle in the full model, the effective demethylation rate can be thought of as representing the combination effect of 5mC to 5hmC oxidation rate affected and the rate of oxidized 5hmC reverting to the unbound promoter state via either TDG-BER pathway or dilution via cell replication. In Supplementary Figure 1(b), the simulation shows that the BoAp is monotonically decreasing with respect to θ. Over 6200 parameter sets have been simulated and we plot BoAp vs. θ over 4 orders of magnitude to show the trend. In Figure 7, we show a representative sample of these curves. We choose the value of the range for θ to be the same as the range of γ in the BoAp vs. γ plot, which both are from 10^{-2} to 10^{2}.

These results confirm that our model captures the key role that methylation rate plays in enhancing the stability of the somatic steady state. The higher the effective methylation rate, the larger the BoAp. This means that it is more likely for the underlying dynamical system to stay at the somatic (silenced) steady state, which implies that it is

relatively more stable. Furthermore, the methylation rate can be increased to the point of eliminating the existence of the active steady state (the pluripotent steady state) and making the silenced steady state immune to any abrupt perturbations. Note that the methylation rate is proportional to the abundance of DNMT, and hence all the conclusions above can be interpreted as applying to the DNMT abundances. We will elaborate on the practical implications of our model further in the Discussion section.

Effect of slow/fast methylation kinetics

To understand how the relative methylation rate affects the BoAp, we define the methylation association rate as R = $\frac{\gamma}{\theta}$.

The simulations in Figure 8 show that BoAp vs. γ across 4 orders of magnitude in γ while θ is increased so as to keep the ratio R constant with different ratios of R that span 3 orders of magnitude. Assuming a normalized protein decay rate ($\delta = 1$), Figure 8 allows us to disentangle two effects: the methylation ratio (R) and the time-scale of the methylation cycle. The plots display a biphasic trend, i.e., the BoAp monotonically increases and then monotonically decreases. Therefore, there is a maximum value of the
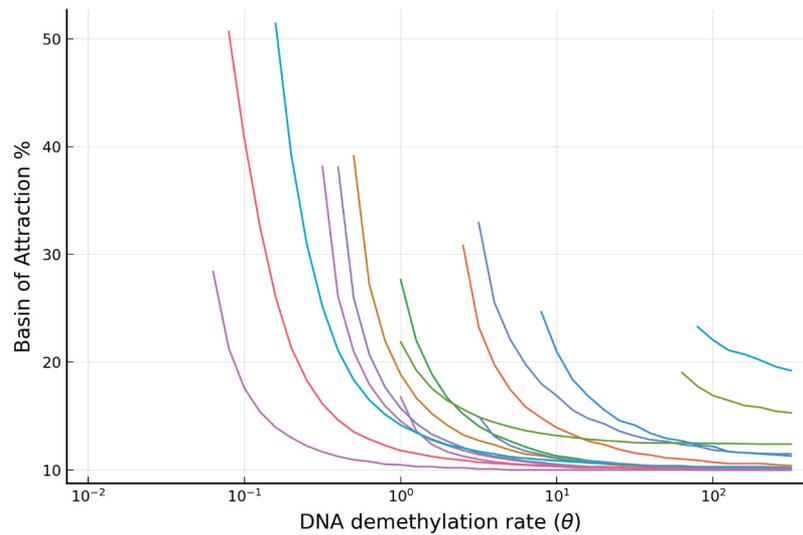
Figure 7. The relative stability (measured by BoAp) of the silenced state decreases with respect to the DNA demethylation rate (θ). Shown are a representative set of simulations for the full model, where extensive simulations are shown in SI-§1.1. Each curve in the figure corresponds to a particular set of parameters.

BoAp which occurs in the slow range $[10^{-1.5}, 10^{-0.5}]$ for various methylation ratios R, while the BoAp decreases as the methylation cycle becomes faster (which is consistent with the analytical picture to be shown in Figure 9). These results mean that the stability of the silenced state (or the efficiency of methylation) is maximized with a higher methylation ratio and a relatively

slower methylation kinetics (roughly an order of magnitude slower than the protein decay rate).

It should be noted that the biphasic nature of the trends is not a result of high-dimensional dynamics. Rather, the same behaviour can be observed in the single gene model (see SI-§ 1.2). Furthermore, the biphasic trend for the fixed ratio R can roughly be inferred from the PSCC simulations by varying each γ and θ independently. As
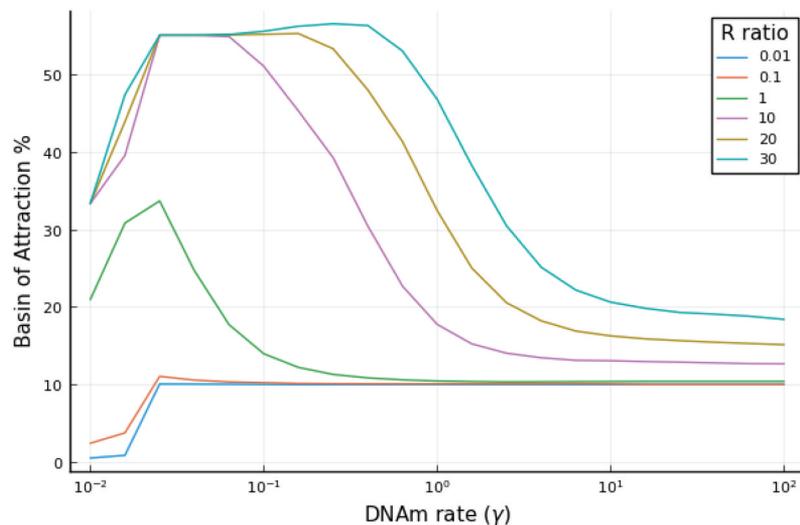

Figure 8. The BoAp of the silenced state shows a biphasic behaviour when the methylation cycle transitions from slow to fast, where R defined as the methylation association rate $R = \frac{\gamma}{\theta}$. The plot shows the BoAp while γ is increased with a fixed methylation ratio, i.e., θ is increased so as to keep the ratio R constant. The system loses bistability when R exceeds 30. The kinetic rate parameters used are: $K_O = 0.3$, $K_{nt} = 0.2$, $K_d = 0.1$, $a = 1$, $a_{nt} = 100$, $a_O = 100$, $a_T = 1$, $a_O = 1$, $a_N = 1$, $\delta = 1$, $R \in \{0.01, 0.1, 1, 10, 20, 30\}$, $\gamma \in [10^{-2}, 10^2]$.
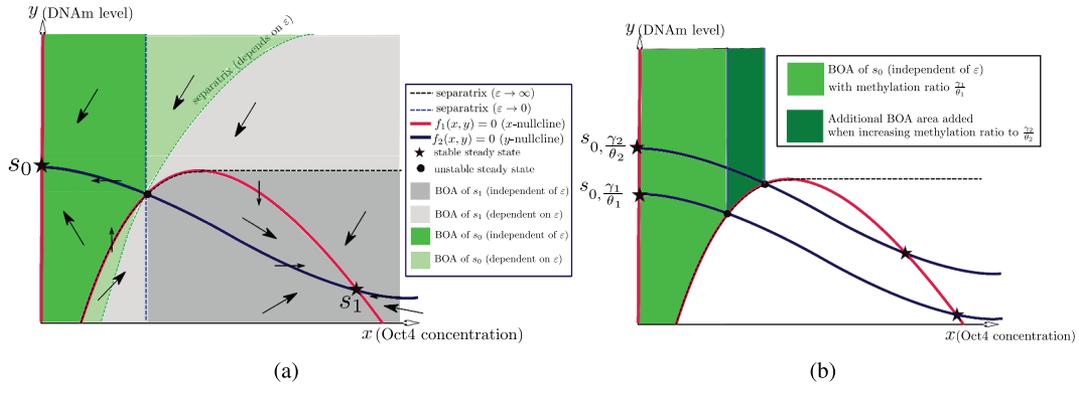
Figure 9. Phase plane analysis of the two-dimensional model (11). (a) The directions of the vector field and the BoAs in the case of bistability with stable steady states s_0, s_1, where s_0 denotes the silenced steady state and s_1 denotes the active steady state. The arrows in each region denote the direction of the vector field. (b) The area of the ε-independent BoA of the silenced steady state s_0 increases when the methylation ratio increases. Recall that $\varepsilon \to \infty$ denotes the case of a slow methylation kinetics (relative to protein kinetics), while $\varepsilon \to 0$ denotes the case of fast methylation kinetics.

per the trend we have shown in Figure 6, the BoAp increasing rate exhibits a sudden slowing down at γ around $10^{-1.5}$, while there is no corresponding change for the demethylation rate θ shown in Figure 7. Therefore, it is not surprising that the methylation association ratio R has a biphasic behaviour.

Model of a single self-activating gene

In order to complement the numerical investigations in the previous sections, we consider next a simplified model which includes Oct4 only. Its interaction with the rest of the network is modelled as a self-activating loop. This model is more amenable to an analytical study as will be discussed below.

Our simplified model consists of a single self-activating gene. A CRN model for it can be written as follows:

$$2X + D_i \underset{\beta_-}{\overset{\beta/2}{\rightleftharpoons}} D_a \tag{2}$$

$$D_a \xrightarrow{k} D_a + X \tag{3}$$

$$D_i \underset{\gamma}{\overset{\theta}{\rightleftharpoons}} D_m \tag{4}$$

$$X \xrightarrow{\delta} \emptyset, \tag{5}$$

where X is the protein expressed from the active promoter state D_a and D_i is the inactive promoter state. (As earlier, we omit explicit variables for mRNA and other intermediate steps such as protein maturation and post-translation modifications.) To model self-activation, we assume that an X homodimer can bind to the inactive promoter site to activate the gene. The inactive promoter region can be methylated to a state that we denote by D_m. Reversely, D_m can be epigenetically modified back to the D_i state. We assume that protein X has degradation rate δ. Our aim is to study the effect of the methylation ratio γ/θ on the dynamics.

Since the total concentration of promoters is conserved, the ODE model can be reduced by writing $D_i = 1 - D_a - D_m$, where we scale the total number of promoters to 1. The system of ODEs that results is given as follows:

$$\frac{dx}{dt} = kD_a - \delta x - \beta x^2 (1 - D_a - D_m) + \beta_- D_a, \tag{6}$$

$$\frac{dD_a}{dt} = \beta x^2 (1 - D_a - D_m) - \beta_- D_a, \tag{7}$$

$$\frac{dD_m}{dt} = \gamma(1 - D_a - D_m) - \theta D_m. \tag{8}$$

The steady states can be evaluated analytically to yield a cubic equation. There always exists a steady state in which $x = 0$ (and also $D_a = 0$). We call such a state the *silenced steady state*. It is easy to verify by linearization that this state is locally stable, for any parameter values. A bifurcation occurs and three steady states exist if and only if the following inequality holds:

$$\frac{k^2}{\delta^2} > \left(\frac{\gamma}{\theta} + 1\right)\frac{4}{B}, \tag{9}$$

where $B := \beta_-/\beta$ is (double) the dissociation ratio, and k/δ and γ/θ are the production and methylation ratios, respectively. One of two additional steady states is always locally asymptotically stable and the other one is unstable (a saddle). We call the second stable steady state *the active steady state*. It follows from condition (9) that a sufficiently high methylation ratio renders the active steady state non-existent. Hence, a point attractor will coincide with the silenced steady state. This recapitulates our observation in Figures 6, 7 that increasing the methylation ratio beyond a certain threshold will eliminate bistability, and hence the active (pluripotent) steady state disappears.

The model can be reduced further by utilizing the fact that transcription factor *binding/unbinding is fast* relative to protein expression and epigenetic modifications. We thus carry out a quasi-steady state (QSS) approximation, in which we set the state D_a to:

$$D_{a,qss} = \frac{(1 - D_m)x^2}{B + x^2}. \tag{10}$$

Substituting this expression into the ODE system, we obtain a two-dimensional system whose two variables are the promoter methylated state $y(t) := D_m(t)$ and the protein x, as follows:

$$\dot{x} = k(1 - y)\frac{x^2}{B + x^2} - \delta x \tag{11}$$

$$\varepsilon\dot{y} = \gamma(1 - y)\frac{B}{B + x^2} - \theta y,$$

where ε is a parameter that represents the relative time-scale separation between protein and methylation kinetics. This system must be studied for $(x, y) \in [0, \infty) \times [0, 1]$. We next analyse this reduced system.

Global behaviour

Observe that all solutions of the reduced system are bounded. Indeed, y is bounded by definition, and $\dot{x} < 0$ whenever $x > k/\delta$, so the rectangle $[0, k/\delta] \times [0, 1]$ is an attractor.

The steady states of the reduced system are in a one-to-one correspondence with the states of the full system, in the sense that for each steady state (x, y) there is a unique steady state (x, D_a, D_m) of the full system, where $D_m = y$ and D_a is given by the expression in Eq. (10). Thus, there are one or more steady states depending on condition (9). We may think of the system (11) as a 'toggle switch' in which Y (the methylated promoter) represses X (protein), and conversely X represses Y. If silencing is stronger, we expect the protein state to remain at zero, while if activation is stronger, we expect protein values to converge to a higher steady state.

In order to analyse the global behaviour of system (11), we first evaluate its Jacobian matrix:

$$\begin{bmatrix} \frac{2kBx(1-y)}{(B+x^2)^2} - \delta & -k\frac{x^2}{B+x^2} \\ \frac{-2\gamma Bx(1-y)}{\varepsilon(B+x^2)^2} & -\frac{1}{\varepsilon}\left(\frac{\gamma B}{B+x^2} + \theta\right) \end{bmatrix}.$$

Both off-diagonal entries are non-positive, which makes this system a monotone system, and specifically, a system that is cooperative with respect to the cone defined by the $(-, -)$ orthant, as discussed in [56] (see also [57] for exposition and biological applications). In [58], M.W. Hirsch proved (Theorem 2.2) that every bounded solution of a planar cooperative system converges to an equilibrium; see also Theorem 2.2 in [56]. Coupled from boundedness of solutions for our system, we conclude that every trajectory converges to an equilibrium (which depends on the initial conditions of the trajectory).

Phase plane analysis

Our two-dimensional reduced model can be analysed via nullclines. A summary of the analysis is depicted in Figure 9 for a bistable model.

The nullclines and the steady states are independent of ε. On the other hand, the separatrix, which is the boundary between the BoAs, is ε-dependent. It is shown that each BoA can be partitioned into ε-dependent and ε-independent regions (Figure 9(b)) shows that the area of the ε-independent BoA of s_0 increases as the methylation ratio is increased. Hence, it can be argued that the BoAp of the silenced steady state increases as the methylation ratio increases, which recapitulates our observations for the full model. Furthermore, we have performed numerical simulations for the single-gene model and they show the same trends. Refer to SI-§ 1.2.

Slow methylation kinetics

We have argued previously that DNAm can be even slower than protein kinetics; hence, we focus on case $\varepsilon \to \infty$ in this subsection. Figure 9 shows that, in this limit, the BoA of s_0 consists only of the ε-independent region. Next, we analytically investigate the aforementioned limit.

Computing the asymptotic separatrix

Setting $\dot{x} = 0$, we see that the x-nullcline consists of two components, which we call the silenced and active arcs, because the dynamics along each arc converge to the silenced and active steady states, respectively. The arcs can be written as follows:

1. *Silenced/Somatic*: The arc is $x = 0$, which gives the slow dynamics of

$$\dot{y} = \gamma B (1 - y) - \theta y.$$

This is a linear system, and it has an asymptotically stable steady state at $y = \frac{B\gamma/\theta}{B\gamma/\theta + 1}$. Notice that the location of this steady state is an increasing function of the methylation ratio γ/θ.

2. *Active/Pluripotent*: The arc is given by the implicit equation: $\frac{k}{\delta}(1 - y)x = B + x^2$. Solving for x in terms of y:

$$x = \frac{k}{\delta}(1 - y) \pm \sqrt{\frac{k^2}{4\delta^2}(1 - y)^2 - B}. \qquad (12)$$

The slow dynamics are given as (note $\limsup_{t \to \infty} y(t) \neq 1$ for all t and any initial condition with $y(0) \neq 1$):

$$\dot{y} = \frac{\gamma \delta B}{k} - \theta y.$$

Hence, $y = \frac{\gamma \delta B}{\theta k}$ is the asymptotic steady state, which is proportional to the methylation ratio γ/θ.

Given an initial condition $(x(0), y(0))$, we want to determine which arc will be representative of the dynamics. A necessary condition for the existence of the second arc (i.e., the chance to converge to the active steady state) is that the quantity under the square root in (12) is real. Hence, we have:

$$y(0) \leq 1 - \frac{2\delta\sqrt{B}}{k}, \quad \text{and} \quad 1 - \frac{2\delta\sqrt{B}}{k} > 0 \qquad (13)$$

Hence, $y(0) > 1 - \frac{2\delta\sqrt{B}}{k}$ automatically implies that the silenced arc represents the dynamics regardless of the initial $x(0)$. Furthermore, if the dissociation ratio is high compared to the protein production ratio (i.e., $k/\delta < \frac{1}{2}\sqrt{B}$) then the dynamics are also represented solely by the silenced arc. Assuming the necessary condition is satisfied, we need also that

$$x(0) > \frac{k}{2\delta}(1 - y) - \sqrt{\frac{k^2}{4\delta^2}(1 - y)^2 - B}. \qquad (14)$$

To summarize, if both (13), (14) are satisfied, then the second arc is representative of the dynamics; otherwise, the first arc represents the dynamics. This conclusion is summarized in the dark grey region in Figure 9(a).

A remark concerning reprogramming by over-expression of X. Our global analysis allows us to intuitively understand the role of X in over-expression experiments (such as the ones carried out in Yamanaka's lab [7]). In this scenario, we assume that $(x(0), y(0)) = (0, \frac{B\gamma/\theta}{B\gamma/\theta + 1})$, i.e., the initial condition is a silenced steady state. Over-expression means that a sudden dose of x will be injected to the system such that $x(0_+) = u$, where u is a constant positive number. Our aim is to drive the system to the BoA of the active steady state.

It follows from the analysis above that

$$\frac{B\gamma/\theta}{B\gamma/\theta + 1} \leq 1 - \frac{2\delta\sqrt{B}}{k}$$

is necessary for reprogrammability. Hence, the expression ratio $\frac{k}{\delta}$ and the methylation ratio $\frac{\gamma}{\alpha}$ must satisfy:

$$\frac{k}{\delta} > 2\sqrt{B}, \quad \frac{\gamma}{\theta} \leq \frac{1}{B}\left(\frac{k}{2\delta\sqrt{B}} - 1\right). \quad (15)$$

If both of these conditions are satisfied, then overexpression will steer the trajectory to the active arc if the following holds:

$$u > \frac{k}{2\delta}\left(\frac{1}{B\gamma/\theta + 1}\right) \\ - \sqrt{\frac{k^2}{4\delta^2}\left(\frac{1}{B\gamma/\theta + 1}\right)^2 - B}. \quad (16)$$

In a nutshell, eqns. (15), (16) imply that we need a sufficiently high Oct4 expression ratio, a sufficiently high overexpression value (u), and a sufficiently low methylation ratio. This again underscores the importance of the methylation ratio.

Comparison of the two models. The results for the single gene model (presented analytically above, and numerically in SI-§ 1.2 parallel those for the four-dimensional model (which was, in turn, obtained through QSSA from the full 17-dimensional CRN)). For example, we see a qualitatively similar behaviour regarding the BoAp's dependence on the (de)methylation rate in both models. In addition, the results in the reduced model with 4D state space again agree with those for the single gene model with 2D state space.

The full model has the advantage of incorporating all the regulation loops and displaying the effect of the methylation cycle on the steady-state values of the Nanog and TET proteins. However, the single gene model can be reduced to a 2D model which enables the analytical computation of the steady states and the application of phase plane methods. Basic trends in the BoAp can be intuitively read from the shift in the separatrix within the predefined cube. Unfortunately, performing similar analysis for the 4D model is intractable, but the numerical simulations have shown the same basic behaviour predicted by the single gene model.

The single gene model is not equipped with the additional regulation loops mediated by the NT heterodimer. Nevertheless, this model demonstrates that the trends displayed by the full model are intrinsic effects of the proposed methylation/demethylation cycle. In both cases, a slow epigenetic process – in our case, a methylation cycle – shifts the stability boundary of the underlying system consistently in the same direction.

Discussion

We have proposed a mathematical modelling paradigm to elucidate the role of DNAm in the activation and silencing of genes. This has been achieved by integrating DNAm into the standard paradigm of GRNs, which rely on TF-promoter binding/unbinding. Our model can be considered as an effective model which considers essential chemical reactions that are relevant to the core cell Oct4 reprogramming network. We have provided a quantitative model of the effect of methylation rates on gene silencing and activation, particularly in the context of ES cell reprogramming. We have shown that increasing the methylation rate increases the BoAp of the silenced state until the active state disappears and the silenced state becomes the only stable steady state. Since the methylation rate is directly proportional to the abundance of DNMT, our results predict that reducing DNMT abundances will lead to gene activation via the emergence of the new active steady state and the loss of stability of the silenced state by BoA shrinkage. Indeed, experimental evidence shows that the application of DNMT inhibitors leads to the reactivation of the associated genes in various contexts [59–62]. In the context of the PSCC, repressing DNMT3B helps in generating pluripotent cells and underscores its role as a regulator of the circuit [63]. Furthermore, downregulation of DNMT3a/b via miRNAs has been suggested as a mechanism for regulating pluripotency [64,65]. On the other hand, DNMTs 1/3a/3b genes can be eliminated without any negative impact on the pluripotent state [66]. Similar results were reproduced in human ES cells with DNMT 3a/b knockout [67]. These results are consistent with our prediction that a highly reduced methylation rate will consolidate the stability of the active steady state. As discussed in the introduction, a smaller basin of

attraction of the silenced state means that natural perturbations in the cell environment will have a greater chance of leading to a state transition to the active state.

Since DNAm and demethylation rates can be changed dynamically during the developmental process, the model can be applied at different stages. Also, it can be used to predict the intervention needed (for instance in terms of DNMT up- or down-regulation) to activate or silence a gene of interest. The relative stability of a certain phenotype (measured by the BoAp) can be also estimated which correlates with the average time (e.g., number of days) needed to transition between different phenotypes. Hence, our model can be used as a quantitative guide to design the reprogramming protocol in the context of the PSCC. Furthermore, the DNAm rate can be measured directly. Experimental tools have been developed to trace the dynamic changes of DNAm at single base resolution [68]. In principle, a time series of the DNAm density at the TSS over the course of the developmental process can be constructed from such data, and the methylation rate can be determined accordingly. Such time-series data have been used in a recent experiment [53], which studied the reprogramming of B cells to induced pluripotent cells. That paper showed that the last stage of the successful pluripotency induction displays successive waves of hydroxymethylation associated with a decrease in DNA methylation and active demethylation. Although such techniques are in their infancy, we predict that our model can be used, in conjunction with more extensive data collection methods, to predict the likelihood of the activation of a gene based on the measurable methylation and demethylation rates.

Our model can also be used in conjunction with more direct experimental manipulation methods. In addition to DNMT inhibitors mentioned above, a recent intriguing technique in synthetic epigenetics is the newly developed synthetic read-write modules [69] implemented in human cells via m6A DNA modification. Such synthetic epigenetic regulatory system can allow us, in principle, to regulate the DNAm patterns and influence the resulting phenotypes by controlling the (de)methylation rate. Using our model, the relative stability of a phenotype is reflected in the duration of the time that a cell takes to successfully reprogram to the pluripotent state due to the stochastic nature of the underlying processes. From an engineering perspective, with the help of synthetic epigenetic read-write modules, one can aim at re-engineering the DNAm patterns to explore how changes of (de)methylation rates affect the reprogramming time during the developmental process. In addition to directly manipulating the DNAm or demethylation rates, one can also think of indirect ways for interacting with DNAm. Evidence suggest that histone modifications like H3K4me1/3, H3K27me3, H3K27ac and the polycomb family of proteins are intricately interacting with DNAm and are playing key roles in recruiting some cofactors and interacting with enhancer regions [70]. Therefore, one can experimentally engineer specific pathways for perturbing the DNAm rate through histone level interventions.

Along with the discovery of the role of TET protein in regulation of the DNA (de)methylation cycle, several theoretical models have been proposed to understand the underlying mechanistic picture of how TFs and DNA interact with each other, and additionally how epigenetic factors such as DNAm and histone modifications affect gene regulatory networks in various biological contexts. In the following, we discuss the relation between these models and ours. Many related core gene regulatory networks have been proposed at various levels of complexity. An early attempt using a probabilistic Boolean network that describes the interplay between gene expression, chromatin modifications, and DNAm has been proposed in [19]. In addition to lacking mechanistic interpretations, Boolean models do not allow one to quantify the effect of DNAm on the stability boundary between the multiple phenotypes. As the role of the TET protein family in gene regulation became clearer [31,42,71,72], various CRN models and ODE models that have partially considered TET-mediated demethylation cycle [36,73] and core gene-TF regulatory mechanisms have been proposed. Notably in [73], an ODE model for a core gene-TF regulatory network of the PSCC has been studied. The system involves Oct4, Sox2 and Nanog proteins and exhibits multistability under certain

conditions. Regarding epigenetic regulation, that work modelled the DNAm cycle on top of Oct4 auto-activation. The model did not consider the interaction between Nanog and TET, nor how NT heterodimer regulates the DNA demethylation cycle. Despite the fact that the DNAm model in [73] did not include a two-pathway activation-like mechanism for gene activation, it represented an early attempt to mechanistically link gene regulation and chromatin modifications. In [36], a mechanism was proposed for the core gene-TF network without epigenetics, and then a two-state model without NT regulating the DNA demethylation process was investigated on top of this core network. However, the underlying chemical reaction model in [36] was not spelled-out clearly. In addition to this, without the NT complex involved in the DNA demethylation process, the model in [36] couples the core gene-TF network and DNA demethylation cycle, which does not reflect the experimental evidence of the role of NT in the regulation of DNA demethylation. Specifically, the model lacks the reaction in which NT complex binds to the 5hmC. In the recent experimental work [74], four chromatin regulators, including DNMT3B which causes de novo methylation of cytosine-guanine dinucleotides (CpGs), have been studied. Though a descriptive three-state model was proposed in [74] for explaining how each chromatin regulator affects the silencing and reactivation of gene expression, a detailed mechanism at a molecular level was not provided. In comparison with all these previous attempts, we model the underlying CFN with a detailed CRN model at a molecular level, and we clearly show how epigenetic regulation, specifically the DNA demethylation cycle, regulates the CFN by the Nanog-guided TET protein complex.

In [75], a purely phenomenological model of how epigenetic feedback affects gene expression dynamics is proposed. The main idea is to represent the kinetic rate parameters by epigenetic variables that have their own dynamic equations. More recently, a similar idea has been employed in the context of epithelial-mesenchymal transition (EMT) and cancer metastasis [18]. Such a view of epigenetic feedback regulation of kinetic rates is a valuable complementary perspective.

So far, we have explored a cell fate network combining gene-TF interactions and a TF-guided DNA demethylation process. For the CRN we proposed, we have considered how the TET-mediated demethylation cycle potentially affects the gene regulatory network, and subsequently affects the cell fate decision associated with the transcription factors. We have simplified the demethylation cycle into a two-state cycle. Our mechanistic modelling approach using CRN agrees with the intuition on how DNAm affects the cell fate network. The faster the methylation rate the more stable the silenced (somatic) state will be. However, in the detailed molecular picture of the cell fate network that involves Oct4, Nanog, TET and potentially many other transcription factors, there might exist all sorts of self or non-self interactions in the network. For example, if one adds Nanog protein self activation, one would expect such a model to potentially exhibit tristability. Although our model explored the simplest case in which the system gives rise to bistability, the basin of attraction boundary of other models with tristability can still be studied by the relative BoAp within a predefined hypercube.

DNAm's role in gene silencing at transcriptional start sites has been well established where genomic maps have shown that regulatory genes are indeed active when they are unmethylated [76]. Furthermore, its repressing roles have been documented in many specific cases such as that of the PSCC [46], X-chromosome inactivation [77], and in ageing and cancer [78,79]. However, it is worth mentioning that a full and universal understanding of its role across different genes and its effect on phenotypic variability is not completely achieved yet. This is since different epigenetic mechanisms at different scales are intricately related to DNAm pattern and dynamics. Additionally, its role can vary in different localized biological contexts [76]. For example, although DNAm generally represses TF binding at transcriptional start sites, methylation in the gene body can have a different role [25]. For instance, it has been shown recently that DNAm can indirectly facilitate TF binding in some special cases [80]. Another issue for debate is whether DNAm is a master regulator or a by-product of the underlying GRN [80,81].

Therefore, the analysis provided in this paper is more predictive for the majority of cases in which DNAm plays a repressing role in a GRN. A future direction will be to develop a more encompassing framework for the various roles of DNAm.

As for future directions, even though the two-pathway activation is a mechanism-based model, it does not capture the full complexity of the whole picture of gene activation. The implicit assumption for this activation process is that genes are only activated by TFs binding to promoters. Hence, we are studying the extension of our methods to the analysis of more sophisticated transcriptional control mechanisms, such as the formation of super-enhancers [82]. We are also interested in combining machine learning approaches with our CRN model so as to discover new mechanisms when the relevant data is available. Finally, when generalizing the BoAp to a stochastic setting, the effective landscape of the dynamical system and the mean first passage time between equilibria are the concepts which are most relevant to the analysis of the stochastic stability boundary. Furthermore, stochastic analysis via effective potentials can be used for defining the BoAp and exploring gene-TF interactions with epigenetic regulation.

Methods

Basic mathematical concepts and definitions

Mathematically, our model is initially described as a Chemical Reaction Network (CRN). CRNs provide a natural formalism for representing biological interactions, and in particular biochemical processes, and CRN descriptions map into systems of Ordinary Differential Equations (ODEs) using standard procedures.

Chemical reaction network (CRN) framework

It is essential for our study to keep track of promoter occupancy, since DNAm at the level of promoters is commonly understood as a slow process compared to TF binding/unbinding, and it can be even slower than protein kinetics. However, our main model assumes that DNAm and protein kinetics are in the same time-scale.

Following [83], each gene is associated with promoter states and one or more a protein states.[1] After the model is set up, we later reduce it by the substitution of quasi steady-state approximations of fast variables, based on appropriate relative time-scales.

We briefly review the general formalism, see, e.g., [84,85]. A CRN is specified by a set of species $S = \{Z_1, .., Z_n\}$ and a set of reactions $R = \{R_1, \ldots, R_v\}$. A reaction R_j can be written as: $\sum_{i=1}^{n} \alpha_{ij} Z_i \rightarrow \sum_{i=1}^{n} \beta_{ij} Z_j$. The associated stoichiometry matrix $\Gamma \in \mathbb{R}^{n \times v}$ is defined element-wise as $[\Gamma]_{ij} = \beta_{ij} - \alpha_{ij}$. Each reaction R_j can occur with a rate function $R_j : \mathbb{R}_{\geq 0}^n \rightarrow \mathbb{R}_{\geq 0}^v$. We assume that R_j takes the form of Mass-Action kinetics: $R_j(z) = \prod_{i=1}^{n} k_j z_j^{\alpha_{ij}}$, where k_j is a kinetic constant. Letting $z(t) \in \mathbb{R}_{\geq 0}^n$ be the vector of species concentrations at time t, the associated ODE can be written as: $\dot{z} = \Gamma R(z)$, where $R := [R_1, .., R_v]^T$.

The goal of our models is to grasp the essence and the most relevant dynamics that give rise to the CFN with bistability. Unfortunately, it is difficult to decide on appropriate ranges of kinetic parameters for such coarse grained level models. One restriction that we have imposed on parameters is to require bistability, meaning (for both the single gene and the full model) that there should be two stable attractors (and a saddle node) for the underlying dynamics. One stable attractor corresponds to the somatic (silenced) state, and the other one corresponds to the active (pluripotent) state.

Full model description

We grouped our CRN for the full model into eight modules with the reactions numbered from R1 to R18 as shown in Table 1.

1. Oct4 promoter module (R1-R4): This module contains four reactions. In our modelling context, we assume that the Oct4 gene has two binding sites. The first binding site is for Oct4 protein itself [86,87], and the second binding site is for Nanog-TET heterodimer complex [42] as shown in the two-pathway activation scheme in Figure 4. Although we have shown NT binding to the first

site and Oct4 binding to the second site in Figure 4, the cases of Oct4 binding to the first site and NT binding to the second site are the same for our modelling. Both cases are considered as a single-site occupancy and are denoted by D_{10}^O in Table 1. As long as two sites are occupied, the corresponding gene will get activated. The parameter α_{NT} is the forward rate for all reactions that involve NT binding to a promoter, and K_{NT} is the dissociation rate for all reversible reactions associated with NT binding to a promoter. The parameter α_O is the forward rate for all reactions that involve Oct4 binding to a promoter, and K_O is the dissociation rate for all reversible reactions associated with Oct4 binding to a promoter.

2. TET promoter module (R5-R8): This module contains 4 reactions. Similar to the Oct4 promoter, the TET promoter region is also able to bind with the NT dimer and the Oct4 protein [31,36,42]. For the reaction rate, we adopt the same rate parameter for TET as for Oct4.

3. Nanog promoter module (R9): This module describes the reversible reaction of Oct4 protein binding to the Nanog promoter [87].

4. Nanog-TET module (R10): This module describes the fact that the TET-mediated DNA demethylation cycle is guided by the Nanog protein in the form of an NT complex [53]. The parameters a_1, K_d are the association and dissociation rates of the heterodimerization of Nanong and TET.

5. Protein decay and production module (R11-R13): In this module, we assume Nanog, TET, and Oct4 all have the same degradation rate δ. When each gene promoter has been fully occupied by TFs, the corresponding gene will be activated. The protein production rates for each activated promoter D_1^N, D_{11}^T, D_{11}^O are given by the reaction rates α_N, α_T and α_O.

6. Protein basal production module (R14-R16): In this module, we assume that TET protein has a basal production rate m_2. The rates m_1 (basal rate for Nanog) and m_3 (basal rate for Oct4) are 0. We used the values of such basal protein production rates from [36].

7. Oct4 promoter Methylation module (R17): This module describes the de novo methylation of

the Oct4 gene promoter region with a methylation rate γ. We assume that the Oct4 promoter can be methylated only if its promoter is unbound. This is supported by experimental observations that show that Oct4 protects its promoter from methylation when it is bound to it [88].

8. Nanog-guided TET demethylation module (R18): In this module, we assume that the triangle topology of DNA demethylation cycle has a slow dynamics on the 5hmC oxidized promoter state [31]. Using a quasi-steady state approximation, we reduced the demethylation cycle from the triangle topology to a two-state CRN with the effective demethylation rate given by θ. (See Figure 5)

In our model, the DNA methylation rate refers to the effective rate combining *de novo* methylation rate, mainly contributed by DNMT3a/b, and maintenance methylation (upon replication) contributed by both DNMT1 and DNMT3a/b [89,90]. DNMT3 is a family of DNA methyltransferases that can methylate hemimethylated and unmethylated CpGs at the same rate, while DNMT1 has the tendency to methylate hemimethylated DNA after replication [90]. In our model of the DNA demethylation cycle, the 5mC state is considered as a fully methylated state, and our model is not affected by the gene methylation state being hemimethylated state for a transient short period.

In order to understand the basic effects of DNAm on silencing, and to maintain the intelligibility and tractability of the model, we assume that the methylation status of a promoter is binary: it is either methylated or not. As we have illustrated in Figure 1, this is a simplified representation of the real picture where methylation is a complex process. Recent results suggest that the methylation of individual cytosines can be important in certain cases [91]. Furthermore, the methylation of Oct4 starts initially from two sites: the proximal enhancer and the distal promoter, spreading later to neighbouring regions [45]. In addition, the average promoter methylation does not usually reach to either 0% or 100% [92]. However, these intricate details are not relevant to our reduced model. For instance, if 80% methylation of

Oct4 promoter is sufficient for silencing, and the cell is somatic at phenotypic level, we assume that an average DNAm level of 80% to be functionally equivalent to a fully methylated promoter.

ODE system for the full CRN model

The corresponding ODE model for the PSCC (Table 1) is given below:

$$\frac{dN(t)}{dt} = -aNT + ak_d[NT] + \alpha_N D_1^N - \delta N + m_1$$

$$\frac{dT(t)}{dt} = -aNT + ak_d[NT] + \alpha_T D_{11}^T - \delta T + m_2$$

$$\begin{aligned}\frac{d[NT](t)}{dt} &= aNT - ak_d[NT] - a_{nt}[NT]D_{00}^O \\ &+ K_{nt}a_{nt}D_{10}^O - a_{nt}[NT]D_{01}^O \\ &+ K_{nt}a_{nt}D_{11}^O \\ &- a_{nt}[NT]D_{00}^T + K_{nt}a_{nt}D_{10}^T - a_{nt}[NT]D_{01}^T \\ &+ K_{nt}a_{nt}D_{11}^T\end{aligned}$$

$$\begin{aligned}\frac{dD_{00}^O(t)}{dt} &= -a_{nt}[NT]D_{00}^O + K_{nt}a_{nt}D_{10}^O - a_OOD_{00}^O \\ &+ K_Oa_OD_{01}^O - \gamma D_{00}^O + \theta D_m[NT]\end{aligned}$$

$$\begin{aligned}\frac{dD_{10}^O(t)}{dt} &= a_{nt}[NT]D_{00}^O - K_{nt}a_{nt}D_{10}^O - a_OOD_{10}^O \\ &+ K_Oa_OD_{11}^O\end{aligned}$$

$$\begin{aligned}\frac{dD_{01}^O(t)}{dt} &= -a_{nt}[NT]D_{01}^O + K_{nt}a_{nt}D_{11}^O + a_OOD_{00}^O \\ &- K_Oa_OD_{01}^O\end{aligned}$$

$$\begin{aligned}\frac{dD_{11}^O(t)}{dt} &= a_{nt}[NT]D_{01}^O - K_{nt}a_{nt}D_{11}^O + a_OOD_{10}^O \\ &- K_Oa_OD_{11}^O\end{aligned}$$

$$\begin{aligned}\frac{dO(t)}{dt} &= -a_OOD_{00}^O + K_Oa_OD_{01}^O - a_OOD_{10}^O \\ &+ K_Oa_OD_{11}^O - a_OOD_{00}^T + K_Oa_OD_{01}^T \\ &- a_OOD_{10}^T + K_Oa_OD_{11}^T - a_OOD_0^N + K_Oa_OD_1^N \\ &+ \alpha_OD_{11}^O - \delta O + m_3\end{aligned}$$

$$\begin{aligned}\frac{dD_{00}^T(t)}{dt} &= -a_{nt}[NT]D_{00}^T + K_{nt}a_{nt}D_{10}^T - a_OOD_{00}^T \\ &+ K_Oa_OD_{01}^T\end{aligned}$$

$$\begin{aligned}\frac{dD_{10}^T(t)}{dt} &= a_{nt}[NT]D_{00}^T - K_{nt}a_{nt}D_{10}^T - a_OOD_{10}^T \\ &+ K_Oa_OD_{11}^T\end{aligned}$$

$$\begin{aligned}\frac{dD_{01}^T(t)}{dt} &= -a_{nt}[NT]D_{01}^T + K_{nt}a_{nt}D_{11}^T + a_OOD_{00}^T \\ &- K_Oa_OD_{01}^T\end{aligned}$$

$$\begin{aligned}\frac{dD_{11}^T(t)}{dt} &= a_{nt}[NT]D_{01}^T - K_{nt}a_{nt}D_{11}^T + a_OOD_{10}^T \\ &- K_Oa_OD_{11}^T\end{aligned}$$

$$\frac{dD_0^N(t)}{dt} = -a_OOD_0^N + K_Oa_OD_1^N$$

$$\frac{dD_1^N(t)}{dt} = a_OOD_0^N - K_Oa_OD_1^N$$

$$\frac{dD_m(t)}{dt} = \gamma D_{00}^O - \theta D_m[NT]$$

Steady state calculation

For the models defined above (the single gene model is (2)-(5), and the full model given by the CRN in Table 1), we first calculate the number of steady states that the model admits given a certain set of parameters. We use a Homotopy Continuation (HC) method to find all possible steady state solutions for our dynamical system models. The Global HC method is currently the best candidate to find all steady states solution for the system [93].

Sampling method

For the single gene model

We calculate the BoAp with various parameter sets that give rise to three steady states. Then, we calculate the BoAp against γ and θ individually. Though the model contains four state variables O, D_i, D_a, D_m, it is essentially a three-dimensional model due to the conservation of the total concentration of promoters. For simplicity, we normalize

total concentration of promoters to an arbitrary constant, say 5.

In order to calculate the BoAp, we need to specify a predefined volume of interest. We have defined the hyper-cube $[0, o] \times [0, d_1] \times [0, d_2]$ where d_1, d_2, o are 1.5 times the corresponding high steady state. Because the conservation law puts a constraint on the total concentration of promoters, for each D_i, D_a and D_m, their concentration should be all less than the total (in the simulation we choose the concentration total = 5). We sample the state space (with a given grid size) and count the number of points in the defined grided space that converges to the low steady state. We then plot BoAp with respect to γ and θ for various numbers of parameter sets and show that they display a consistent trend.

For the full model

Given the CRN from Table 1, we first calculate steady state solutions that show bistability. Then, we use quasi-steady steady approximation to reduce the full model from 17 dimensions to 4 dimensions and simulate the BoAp in 4D space with a predefined hypercube similar to the previous paragraph.

Parameters

The 'free' parameters below are constrained to the set which can give rise to bistability. The following 'free' parameters can be thought of as normalized constants that can be re-scaled to their experimental values when applied to a biological context.

In the single gene model, the kinetic rate parameters are: β, β_-, k, δ, γ, θ. In the simulation of BoAp vs. γ plot, we choose the parameter (β, β_-, k, δ, θ) values ranging from 0 to 5 with a step size of 0.5. For the simulation of BoAp vs θ plot, the same range is used for β, β_-, k, δ, γ.

The range can be chosen to be $[0, a]$, for any $a > 0$. This is since the ODE (2)-(5) follows Mass-action kinetics and all the rates can be scaled to have values from 0 to a for the purpose of BOAp investigation.

In the full model, the kinetic rate parameters that define the system are: K_O, K_{nt}, K_d, a, a_{nt}, a_O, α_T, α_O, α_N, δ, γ, θ, m_1, m_2 and m_3 as shown in

Table 1. Although the values for the three dissociation parameters K_O, K_{nt}, K_d were given in [36], in our full model simulation results, we are varying K_O, K_{nt}, K_d from 0 to 0.5 with 0.1 step size, and δ from 1 to 5 with 0.5 step size, and the parameter θ from 1 to 20 with a step size of 2. We set protein production rates α_T, α_O and α_N for Nanog, TET and Oct4 to be 1, the parameter $a = 1$, and finally the parameters a_{nt} and a_O are set to be 100. For those parameters that are set to be constants, we did extensive simulations that have shown that the BoAp is insensitive to large changes in their values. Hence, our presented simulations only contain parameter sets that have noticeable influence on the BoAp.

Methylation association rate (R)

In the single gene model, we choose the methylation association rate (R) to be from 0.2 to 1.3 so that it covers both the "$R < 1$" regime and the "$R > 1$" regime of the dynamics with a particular parameter set. In the full model, we choose the methylation association rate (R) to span across several orders of magnitude, as possible, ranging from 0.01 to 30. The values of R for each model depend on the parameter set we choose, but the qualitative trend for all the parameter sets we tested is the same and shows a biphasic trend. For the full model in Figure 8 we have chosen the following parameters: $K_O = 0.3$, $K_{nt} = 0.2$, $K_d = 0.1$, $a = 1$, $a_{nt} = 100$, $a_O = 100$, $\alpha_T = 1$, $\alpha_O = 1$, $\alpha_N = 1$, $\delta = 1$, $R \in \{0.01, 0.1, 1, 10, 20, 30\}$, and with γ spanning the range $[10^{-2}, 10^2]$. For those R values outside the range, the system does not exhibit bistability for this particular parameter set.

Note

1. We assume that transcription is fast enough that no explicit consideration of mRNA abundances is required, but it would be straightforward to add mRNA intermediates to the model.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Code availability

The programming code that was used to analyze the raw data that supports the findings of this study is available from the corresponding author upon request.

ORCID

Tianchi Chen http://orcid.org/0000-0001-6040-6537
M. Ali Al-Radhawi http://orcid.org/0000-0002-6761-4520
Eduardo D. Sontag http://orcid.org/0000-0001-8020-5783

References

[1] Waddington CH. The strategy of the genes. George Allen and Unwin; 1957. Oxon, UK.

[2] Zhou JX, Huang S. Understanding gene circuits at cell-fate branch points for rational cell reprogramming. Trends Genet. 2011;27(2):55–62.

[3] Burda P, Laslo P, Stopka T. The role of PU.1 and GATA-1 transcription factors during normal and leukemogenic hematopoiesis. Leukemia. 2010;24 (7):1249–1257.

[4] Fisher RC, Scott EW. Role of PU.1 in hematopoiesis. Stem Cells. 1998;16(1):25–37.

[5] Chou ST, Khandros E, Bailey LC, et al. Graded repression of PU.1/Sfpi1 gene transcription by GATA factors regulates hematopoietic cell fate. Blood. 2009;114 (5):983–994, 07.

[6] Mittal V. Epithelial mesenchymal transition in tumor metastasis. Annu Rev Pathol. 2018;13(1):395–412. PMID: 29414248.

[7] Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 2006;126 (4):663–676.

[8] Omole AE, Fakoya AOJ. Ten years of progress and promise of induced pluripotent stem cells: historical origins, characteristics, mechanisms, limitations, and potential applications. PeerJ. 2018;6:e4370.

[9] Li C, Wang J. Quantifying cell fate decisions for differentiation and reprogramming of a human stem cell network: landscape and biological paths. PLoS Comput Biol. 2013;9(8):e1003165.

[10] Del Vecchio D, Abdallah H, Qian Y, et al. A blueprint for a synthetic genetic feedback controller to reprogram cell fate. Cell Syst. 2017;4(1):109–120.

[11] Mahla RS. Stem cells applications in regenerative medicine and disease therapeutics. Int J Cell Biol. 2016;2016:6940283.

[12] Zakrzewski W, Dobrzyński M, Szymonowicz M, et al. Stem cells: past, present, and future. Stem Cell Res Ther. 2019;10(1):68.

[13] Hecker M, Lambeck S, Toepfer S, et al. Gene regulatory network inference: data integration in dynamic models – a review. Biosystems. 2009;96(1):86–103.

[14] Huang S, Eichler G, Bar-Yam Y, et al. Cell fates as high-dimensional attractor states of a complex gene regulatory network. Phys Rev Lett. 2005 Apr;94 (12):128701.

[15] Alon U. An introduction to systems biology. Boca raton (FL): CRC press; 2006.

[16] Del Vecchio D, Murray RM. Biomolecular feedback systems. Princeton (NJ): Princeton Univ Press; 2015.

[17] Gardner TS, Cantor CR, Collins JJ. Construction of a genetic toggle switch in Escherichia coli. Nature. 2000;403(6767):339–342.

[18] Jia W, Deshmukh A, Mani SA, et al. A possible role for epigenetic feedback regulation in the dynamics of the epithelialmesenchymal transition (EMT). Phys Biol. 2019;16(6):66004.

[19] Flöttmann M, Scharp T, Klipp E. A stochastic model of epigenetic dynamics in somatic cell reprogramming. Front Physiol. 2012;3:216.

[20] Folguera-Blasco N, Pérez-Carrasco R, Cuyàs E, et al. A multiscale model of epigenetic heterogeneity-driven cell fate decision-making. PLoS Comput Biol. 2019;15 (4):e1006592.

[21] Bintu L, Yong J, Antebi YE, et al. Dynamics of epigenetic regulation at the single-cell level. Science. 2016;351(6274):720–724.

[22] Takahashi K, Yamanaka S. A developmental framework for induced pluripotency. Development. 2015;142(19):3274–3285.

[23] Koh KP, Rao A. DNA methylation and methylcytosine oxidation in cell fate decisions. Curr Opin Cell Biol. 2013;25(2):152–161.

[24] Baylin SB. DNA methylation and gene silencing in cancer. Nature Clin Pract Oncol. 2005;2(S1):S4–S11.

[25] Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. 2012;13 (7):484–492.

[26] Dor Y, Cedar H. Principles of DNA methylation and their implications for biology and medicine. Lancet. 2018;392(10149):777–786.

[27] Luo C, Hajkova P, Ecker JR. Dynamic DNA methylation: in the right place at the right time. Science. 2018;361(6409):1336–1340.

[28] Curradi M, Izzo A, Badaracco G, et al. Molecular mechanisms of gene silencing mediated by DNA methylation. Mol Cell Biol. 2002;22(9):3157–3173, 05.

[29] Moore LD, Le T, Fan G. DNA methylation and its basic function. Neuropsychopharmacology. 2013;38:23–38, 01.

[30] Razin A, Cedar H. DNA methylation and gene expression. Microbiol Rev. 1991;55(3):451–458, 09.

[31] Kasper Dindler KH. Rasmussen, "Role of TET enzymes in DNA methylation, development, and cancer,". Genes Dev. 2016;30(7):733–750.

[32] Wu X, Zhang Y. TET-mediated active DNA demethylation: mechanism, function and beyond. Nat Rev Genet. 2017;18(9):517.

[33] Bachman M, Uribe-Lewis S, Yang X, et al. 5-Formylcytosine can be a stable DNA modification in mammals. Nat Chem Biol. 2015;11(8):555.

[34] Bachman M, Uribe-Lewis S, Yang X, et al. 5-Hydroxymethylcytosine is a predominantly stable DNA modification. Nat Chem. 2014;6(12):1049.

[35] Li M, Belmonte JCI. Ground rules of the pluripotency gene regulatory network. Nat Rev Genet. 2017;18(3):180.

[36] Victor Olariu KS, Lvkvist C. Nanog, Oct4 and TET1 interplay in establishing pluripotency. Sci Rep. 2016;6(1):25438.

[37] Li D, Chen J, Pei D. The battle between TET proteins and DNA methylation for the right cell. Trends Cell Biol. 2018;28(12):973–975.

[38] Pesce M, Schöler HR. Oct-4: gatekeeper in the beginnings of mammalian development. Stem Cells. 2001;19(4):271–278.

[39] Kim JB, Sebastiano V, Wu G, et al. Oct4-induced pluripotency in adult neural stem cells. Cell. 2009;136(3):411–419. .

[40] Geng T, Zhang D, Jiang W. Epigenetic regulation of transition among different pluripotent States. Stem Cells. 2019;11:1372–1380.

[41] Costa Y, Ding J, Theunissen TW, et al. NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. Nature. 2013;495(7441):370.

[42] Sardina JL, Collombet S, Tian TV, et al. Transcription factors drive TET2-mediated enhancer demethylation to reprogram cell fate. Cell Stem Cell. 2018;23(5):727–741.e9.

[43] Okita K, Ichisaka T, Yamanaka S. Generation of germline-competent induced pluripotent stem cells. Nature. 2007;448(7151):313–317.

[44] Polo JM, Anderssen E, Walsh RM, et al. A molecular roadmap of reprogramming somatic cells into iPS cells. Cell. 2012;151(7):1617–1632. .

[45] Athanasiadou R, de Sousa D, Myant K, et al. Targeting of de novo DNA methylation throughout the Oct-4gene regulatory region in differentiating embryonic stem cells. PloS One. 2010;5(4):e9937.

[46] Smith ZD, Meer A. DNA methylation: roles in mammalian development. Nat Rev Genet. 2013;14(3):204–220.

[47] Lee D-S, Shin J-Y, Tonge PD, et al. An epigenomic roadmap to induced pluripotency reveals DNA methylation as a reprogramming modulator. Nat Commun. 2014;5(1):5619. .

[48] Lang AH, Li H, Collins JJ, et al. Epigenetic landscapes explain partially reprogrammed cells and identify key reprogramming genes. PLoS Comput Biol. 2014;10(8):e1003734.

[49] Menck PJ, Heitzig J, Marwan N, et al. How basin stability complements the linear-stability paradigm. Nat Phys. 2013;9(2):89–92.

[50] Ben-tabou de Leon S, Davidson EH. Modeling the dynamics of transcriptional gene regulatory networks for animal development. Dev Biol. 2009;325(2):317–328, 01.

[51] Xu C, Corces VG. Nascent DNA methylome mapping reveals inheritance of hemimethylation at CTCF/cohesin sites. Science. 2018;359(6380):1166–1170.

[52] Velicescu M, Weisenberger DJ, Gonzales FA, et al. Cell division is required for de novo methylation of CpG islands in bladder cancer cells. Cancer Res. 2002;62(8):2378–2384.

[53] Sardina JL, Collombet S, Tian TV, et al. Transcription factors drive Tet2-mediated enhancer demethylation to reprogram cell fate. Cell Stem Cell. 2018;23(5):727–741.

[54] Sasai M, Kawabata Y, Makishi K, et al. Time scales in epigenetic dynamics and phenotypic heterogeneity of embryonic stem cells. PLoS Comput Biol. 2013;9(12):e1003380.

[55] Zhou W, Dinh HQ, Ramjan Z, et al. DNA methylation loss in late-replicating domains is linked to mitotic cell division. Nat Genet. 2018;50(4):591–602.

[56] Smith H. Monotone Dynamical Systems. Providence, RI: AMS; 1995.

[57] Sontag ED. Monotone and near-monotone biochemical networks. Syst Synth Biol. 2007;1(2):59–87.

[58] Hirsch M. Systems of differential equations which are competitive or cooperative I: limit sets. SIAM J Appl Math. 1982;13(2):167–179.

[59] Foltz G, Yoon J-G, Lee H, et al. DNA methyltransferase-mediated transcriptional silencing in malignant glioma: a combined whole-genome microarray and promoter array analysis. Oncogene. 2009;28(29):2667–2677.

[60] Lund K, Cole JJ, VanderKraats ND, et al. DNMT inhibitors reverse a specific signature of aberrant promoter DNA methylation and associated gene silencing in AML. Genome Biol. 2014;15(7):406.

[61] Sin-Chan P, Huang A. DNMTs as potential therapeutic targets in high-risk pediatric embryonal brain tumors. Expert Opin Ther Targets. 2014;18(10):1103.

[62] Ziller MJ, Ortega JA, Quinlan KA, et al. Dissecting the functional consequences of de novo DNA methylation dynamics in human motor neuron differentiation and physiology. Cell Stem Cell. 2018;22(4):559–574. .

[63] Wongtrakoongate P, Li J, Andrews PW. DNMT3B inhibits the re-expression of genes associated with induced pluripotency. Exp Cell Res. 2014;321(2):231–239.

[64] Marson A, Levine SS, Cole MF, et al. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell. 2008;134(3):521–533. .

[65] Horii T, Hatada I. Regulation of CpG methylation by DNMT and TET in pluripotent stem cells. J Reprod Dev. 2016;62(4):331–335.

[66] Tsumura A, Hayakawa T, Kumaki Y, *et al.* Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases DNMT1, DNMT3A and DNMT3B. Genes Cells. 2006;11(7):805–814. .

[67] Liao J, Karnik R, Gu H, *et al.* Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. Nat Genet. 2015;47(5):469. .

[68] Stelzer Y, Shivalila CS, Soldner F, et al. Tracing dynamic changes of DNA methylation at single-cell resolution. Cell. 2015;163(1):218–229.

[69] Park M, Patel N, Keung AJ, et al. Engineering epigenetic regulation using synthetic read-write modules. Cell. 2019;176(1–2):227–238.

[70] Epsztejn-Litman S, Feldman N, Abu-Remaileh M, *et al.* De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. Nat Struct Mol Biol. 2008;15(11):1176. .

[71] Ito S, D'Alessio AC, Taranova OV, et al. Role of TET proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature. 2010;466(7310):1129.

[72] Rasmussen KD, Helin K. Role of TET enzymes in DNA methylation, development, and cancer. Genes Dev. 2016;30(7):733–750.

[73] Abdallah H, "The core mammalian pluripotency network in induced pluripotent stem cell (iPSC) formation: models for genetic and epigenetic reprogramming," Master's thesis, Massachusetts Institute of Technology, 2018.

[74] Elowitz MB, Leibler S. A synthetic oscillatory network of transcriptional regulators. Nature. 2000;403(6767):335.

[75] Miyamoto T, Furusawa C, Kaneko K. Pluripotency, differentiation, and reprogramming: a gene expression dynamics model with epigenetic feedback regulation. PLoS Comput Biol. 2015;11(8):e1004476.

[76] Schübeler D. Function and information content of DNA methylation. Nature. 2015;517(7534):321–326.

[77] Augui S, Nora EP, Heard E. Regulation of X-chromosome inactivation by the X-inactivation centre. Nat Rev Genet. 2011;12(6):429–442.

[78] Esteller M. CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future. Oncogene. 2002;21(35):5427–5440.

[79] Michalak EM, Burr ML, Bannister AJ, et al. The roles of DNA, RNA and histone methylation in ageing and cancer. Nat Rev Mol Cell Biol. 2019;20:573–589.

[80] Héberlé É, Bardet AF. Sensitivity of transcription factors to DNA methylation. Essays Biochem. 2019;63(6):727–741.

[81] Zhu H, Wang G, Qian J. Transcription factors as readers and effectors of DNA methylation. Nat Rev Genet. 2016;17(9):551–565.

[82] Hnisz D, Shrinivas K, Young RA, et al. A phase separation model for transcriptional control. Cell. 2017;169(1):13–23.

[83] Ali Al-Radhawi M, Del Vecchio D, Sontag ED. Multimodality in gene regulatory networks with slow promoter kinetics. PLoS Comput Biol. 2019;15(2):e1006784.

[84] Érdi P, Tóth J. Mathematical models of chemical reactions: theory and applications of deterministic and stochastic models. Manchester (UK): Manchester University Press; 1989.

[85] Ali Al-Radhawi M, Angeli D, Sontag ED. A computational framework for a Lyapunov-enabled analysis of biochemical reaction networks. PLoS Comput Biol. 2020;16(2):e1007681.

[86] Wang J, Rao S, Chu J, et al. A protein interaction network for pluripotency of embryonic stem cells. Nature. 2006;444(7117):364.

[87] Hammachi F, Morrison GM, Sharov AA, *et al.* Transcriptional activation by Oct4 is sufficient for the maintenance and induction of pluripotency. Cell Rep. 2012;1(2):99–109.

[88] Shakya A, Callister C, Goren A, et al. Pluripotency transcription factor Oct4 mediates stepwise nucleosome demethylation and depletion. Mol Cell Biol. 2015;35(6):1014–1025.

[89] Chen T, Ueda Y, Dodge JE, et al. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by DNMT3A and DNMT3B. Mol Cell Biol. 2003;23(16):5594–5605.

[90] Hsieh C-L. The de novo methylation activity of Dnmt3a is distinctly different than that of Dnmt1. BMC Biochem. 2005;6(1):6.

[91] Medvedeva YA, Khamis AM, Kulakovskiy IV, *et al.* Effects of cytosine methylation on transcription factor binding sites. BMC Genomics. 2014;15(1):119.

[92] Hattori N, Nishino K, Ko Y-G, et al. Epigenetic control of mouse Oct-4gene expression in embryonic stem cells and trophoblast stem cells. J Biol Chem. 2004;279(17):17063–17069.

[93] Gritton KS, Seader J, Lin W-J. Global homotopy continuation procedures for seeking all roots of a nonlinear equation. Comput Chem Eng. 2001;25(7):1003–1019.

Supplementary Information for
A mathematical model exhibiting the effect of DNA methylation on the stability boundary in cell-fate networks

Tianchi Chen[1,*], M. Ali Al-Radhawi[2,*], and Eduardo D. Sontag[1,2,3,†]

[1]Department of Bioengineering, Northeastern University, Boston, MA, 02115
[2]Department of Electrical and Computer Engineering, Northeastern University, Boston, MA, 02115.
[3]Laboratory of Systems Pharmacology, Program in Therapeutic Science, Harvard Medical School, Boston, MA 02115
[*]These authors contributed equally.
[†]Corresponding author: e.sontag@northeastern.edu

September 23, 2020

1 Additional BoAp simulations

### 1.1 Extensive numerical simulation results for the full model

In the main paper, we show few simulations for the behavior of the BoAp when γ or θ are varied (see Figures 6 and 7). Supplementary Figure 1 shows more extensive simulations corresponding to all the samples generated as described in the Methods section in the main paper. We observe the same behavior, namely the BoAp increases with respect to the methylation rate γ, and decreases with respect to the demethylation rate θ.

### 1.2 Numerical simulation results for the single-gene model

As a complement to the theoretical discussion, we carried out numerical experimentation to confirm the preceeding nullcline analysis. Figure 2 shows, for the three dimensional system defined by (6) - (8), the result of simulations over 4,000 different parameter sets. The parameter set for the single gene model includes β, β_-, k, γ, θ and δ. In Figure 2, each single curve corresponds to a fixed parameter set. We observe that, as expected, the BoAp is a monotonically increasing function of γ. Intuitively, γ is considered to be the effective methylation rate of the Oct4 gene promoter. The methylation rate of the promoter is proportional to the loss of activity of the gene, and hence is proportional to the likelihood of cells staying at their silenced state (defined as the Oct4 protein staying at low concentration level) We view the variable X as representing Oct4 in the CRN of the single gene model. We interpret a low concentration of the Oct4 protein as meaning that the system stays at silenced steady state (somatic cell state). On the other hand, if Oct4 protein achieves high concentration, the system is considered as being in the active steady state (the pluripotent state). Conversely, if we increase the reaction rate θ, the effective demethylation on the promoter dominates, and therefore a cell would be more likely to stay in a pluripotent state.

(a)
(b)
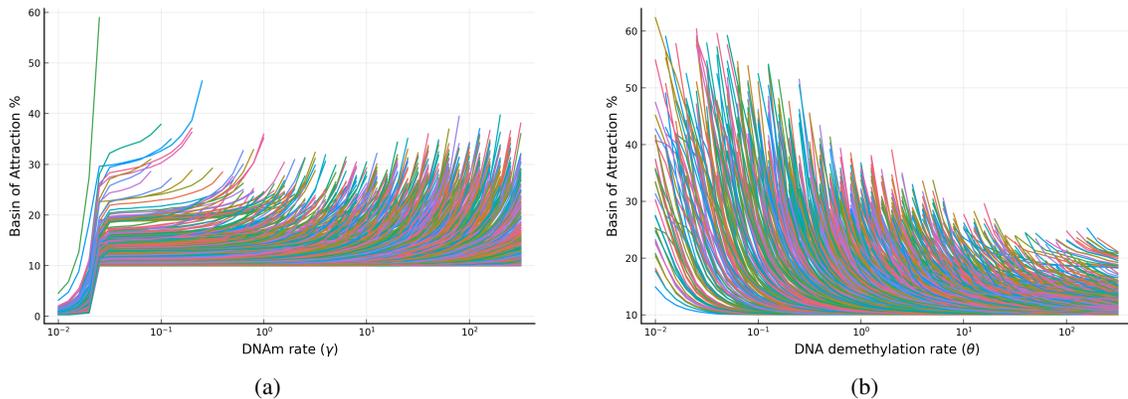
Supplementary Figure 1: The relative stability (measured by BoAp) of the silenced steady state versus the methylation and demethylation rates. We show simulations for a large set of different rates for the full model as described in the Methods section in the main text. Each curve in the figure correspond to a particular set of parameters. (a) The BoAp increases with respect to the DNAm rate (γ). Each curve ends when bistability is lost and BoAp jumps to 100%. This behavior is illustrated in SI-§2.1. The simulations also show that, at small γ region, the BoAp exhibits a sudden slowing down at around $\gamma = 10^{-1.5}$. (b) The BoAp decreases with respect to the demythelation rate θ.

(a) Linear Scale
(b) Log Scale
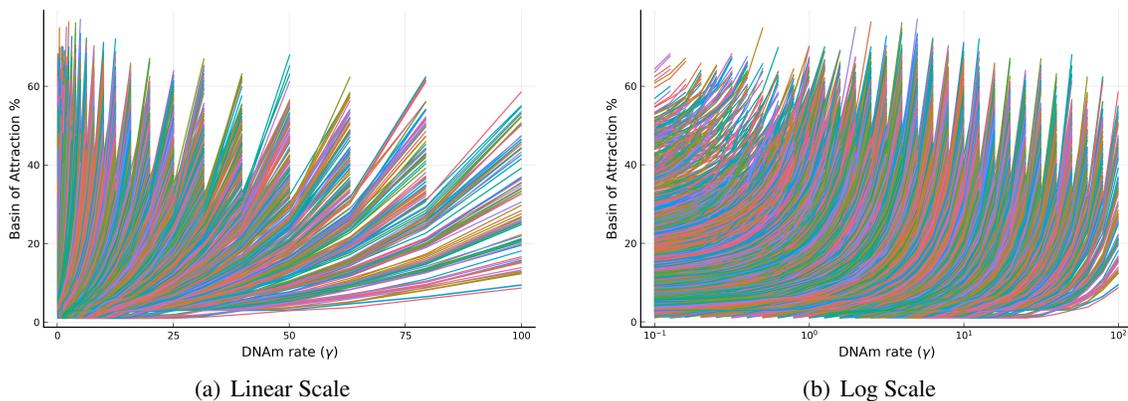
Supplementary Figure 2: The basin of attraction percentage (with respect to silence state) vs DNA methylation rate plot. In this figure we show plots of BoAp vs γ for a large set of different parameters for the single gene model. Each curve in the figure represents a particular set of parameters in the simulation. The color associated with each curve is automatically assigned by the plotting package. We present the BoAp vs γ in both (a) Linear Scale, and (b) Log Scale. It can be seen that the BoAp is monotonically increasing with respect to the methylation rate. For a particular set of parameter, as one increases methylation rate γ, the model will loose bistability after a certain threshold value of γ, which is why both sub-figures show vertical "spike"-like structure.

2

(a) Grid 1   (b) Grid 2
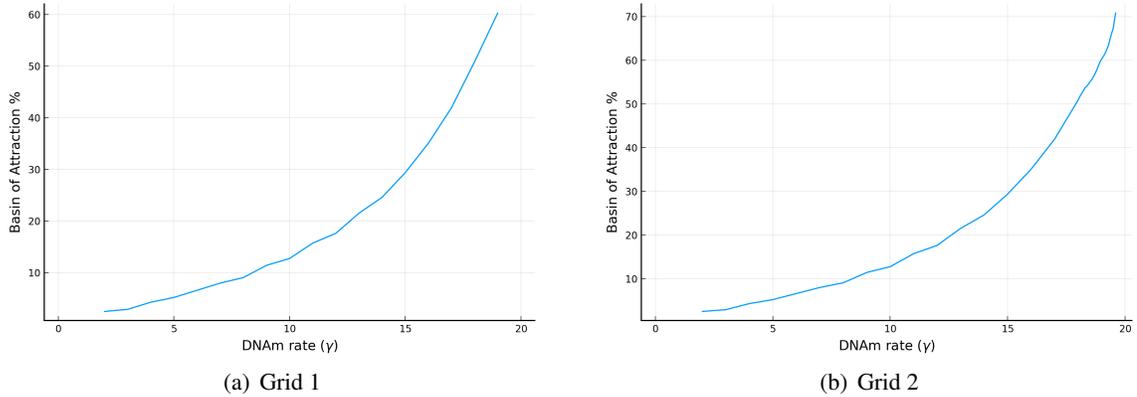
Supplementary Figure 3: BoAp comparison between different grid sizes. The left plot is simulated with grid size = 1. The right plot has a finer grid size = 0.01. Both plots are simulated with the same parameter set: $\beta = 5$, $\beta_- = 5$, $k = 5$, $\delta = 4$, $\theta = 5$. We vary γ from 0 to 20.

We display the simulations both in linear scale, shown in Figure 2(a), and in log scale, shown in Figure 2(b), in order to cover different ranges. The apparent clustering of curves in these plots is due to sampling, as closely related parameters give rise to a similar landscape when changing γ. Observe that subsets of curves in the same group all seem to end at roughly the same value of γ. The reason for this sudden cut-off is that the the system is undergoing a bifurcation and losing bistability. Furthermore, for each curve in the plot, with γ sampled according to a given finite grid, hence the BoAp might achieve a higher value with a finer grid. For example, in Figure 3, the left plot with a coarser grid would indicate that the BoAp reaches only 60%. However, on the right plot, done with a finer grid, the BoAp reaches 72%. We picked a grid size that balances the computational cost involved in the simulation of a large ensemble of parameter sets versus the accuracy of estimation of the BoAp.

In terms of how the DNA demethylation rate θ affects the BoAp, we show a similar simulation in Figure 4 over different parameter sets that for : β, β_-, k, γ, θ, δ. We describe the parameter sampling method in the Methods section.

1.3   Effect of slow/fast methylation kinetics

We have discussed in the previous section how the BoAp changes as we independently change either the methylation rate γ or the demethylation rate θ. Here, we would like understand how the BoAp is affected by the relative changes between γ and θ. In Figure 5, we explore how the BoAp changes with methylation rate γ as one fixes the methylation association ratio defined as

$$R = \frac{\gamma}{\theta}.$$

The methylation association ratio (R) is the ratio of the methylation to the demethylation rates. The larger the value R, the faster the effective methylation process is. Therefore, a value R < 1, indicates that the effective DNA demethylation reaction is faster than the methylation reaction. If instead R > 1, the effective DNA methylation is a faster process. We explore how BoAp changes as one changes γ at different ratios R. We study R across 12 values and simulate as γ and θ span across seven orders of magnitudes.

The value of γ sets the time scale of the dynamics. When γ is small, both methylation and demethylation rate are small. When the value of γ is greater than 10^{-2}, both methylation and demethylation rate are

3

(a) Linear Scale

(b) Log Scale
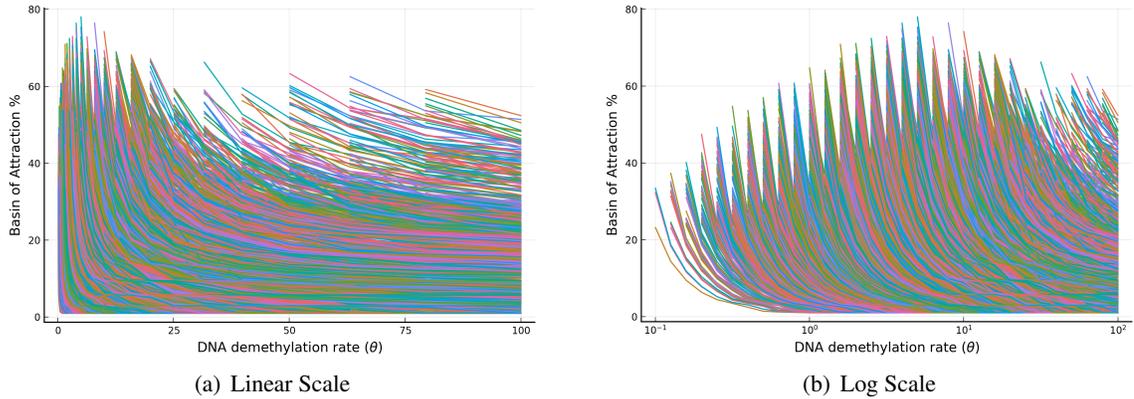
Supplementary Figure 4: The basin of attraction percentage (with respect to silence state) vs DNA demethylation rate plot. In this figure we show plots of BoAp vs θ for a large set of different rates for the single gene model. Each curve in the figure represents a particular set of parameters in the simulation. The color associated with each curve is automatically assigned by the plotting package. We present the BoAp vs θ in both (a) Linear Scale, and (b) Log Scale. It can be seen that the BoAp is monotonically decreasing with respect to the demethylation rate.The vertical "spike"-like structure comes from loosing bistability for the model.
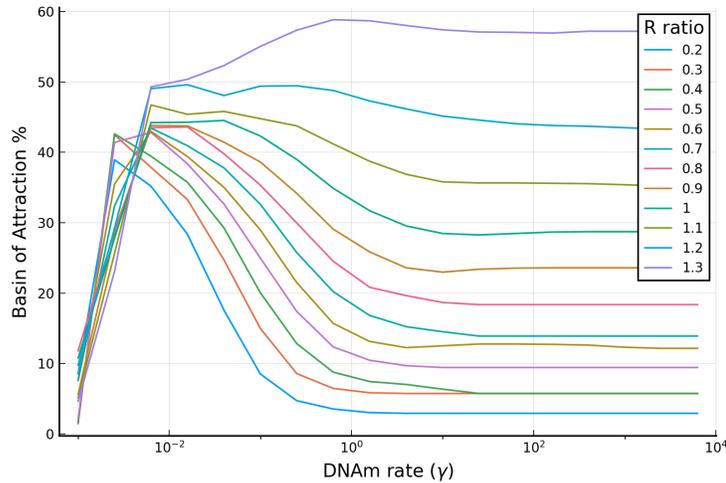
Supplementary Figure 5: The effect of slow/fast methylation kinetics plot. In this plot we show the effect of the timescale of dynamics on the BoAp. We show BoAp vs. γ with a range of fixed R values for the single gene model.The methylation association ratio R set the scale of the dynamics timescale. When R value is large, both γ and θ values are large, therefore, methylation and demethylation dynamics operates at faster timescale. Similar to the full model case, the R we set for the following kinetic rate parameter set was to make sure we explore as wider range as possible as long as they underlying system dynamics still exhibit bistability. For each R ratio (range from 0.2 to 1.3) that allows the bistability of the single gene model, the BOAp vs. γ plot has a biphasic nature (BoAp increases first and then decreases) based on the time scale of the (de)methylation dynamics

4

considered to be large. The simulation in Figure 5 shows a biphasic trend for the underlying dynamics, where $\gamma = 10^{-2}$ set the rough transition barrier. If the (de)methylation rate is sufficiently small, the BoAp is an increasing function with respect to γ, and θ as well since R is fixed. Across different values of R, the BoAp shows saturation for large values of γ. This saturation trend observed in simulations confirms the theoretical analysis shown in Figure 9 in the main text. Indeed, as (de)methylation dynamics becomes very fast, the ϵ in Figure 9 in the main text, which controls the time scale of the dynamics, approaches ∞, and the separatrix becomes a vertical line. Therefore, within the predefined cube, the fraction of the volume that corresponds to the somatic state volume will saturate, hence giving rise to the saturation trend in Figure 5.

2 Full model visualization and additional analysis

2.1 Effect of increasing the methylation rate γ

In the main paper, we have presented how the BoAp changes as a function of the DNAm rate γ. We provide here an alternative visualization. Figure 6 shows the fate of a specific initial state depending on the methylation rate γ. The plots start from a low methylation rate $\gamma = 1$, where the system settles on the active steady state marked by high concentrations of Oct4, TET, and Nanong. As γ increases, the trajectories of the Nanog, TET and Oct4 transition from the active steady state to the silenced steady marked by very low concentrations of the TFs. The lower panel in Figure 6 shows that the methylated Oct4 promoter transitions from a low-methylation to a high-methylation state.

As mentioned in the main text, when the DNA methylation rate increases and passes a threshold value, $\gamma = 50$ in the example above, the system will make a transition from bistability to monostability. Figure 7 shows what happens to the number of steady states during the transition. With $\gamma = 49$, right before the transition, the system retains its bistable character, and when the DNA methylation rate γ increases to 50, the system suddenly loses its bistability and we are left with a single steady state which corresponds to the silenced state as shown in the top right panel of in Figure 7. The lower panel in Figure 7 shows how the BoAp changes as a function of γ with a discontinuity at $\gamma = 50$. When the bifurcation from bi- to mono-stability happens, the BoAp jumps to 100%.

2.2 3D trajectories visualization

We have shown the BoAp vs γ and θ plots in the main paper. We show here 3D trajectories in the state space in the (Nanog,TET,Oct4)-coordinates that illustrates how different initial conditions approach one of the attractors in the 3D state space. Figure 8 shows a system exhibiting bistability. Different trajectories in the state space will converge to either of these two stable attractor based on the initial values.

In addition to the 3D trajectory plots as shown in Figure 8, we show the norm of the velocity vector for the underlying dynamical system in Figure 9. The color represents the norm of the velocity. The velocity plot also shows two stable attractors colored with green. It can be seen also that the norm of the velocity approaches zero close to the attractors.
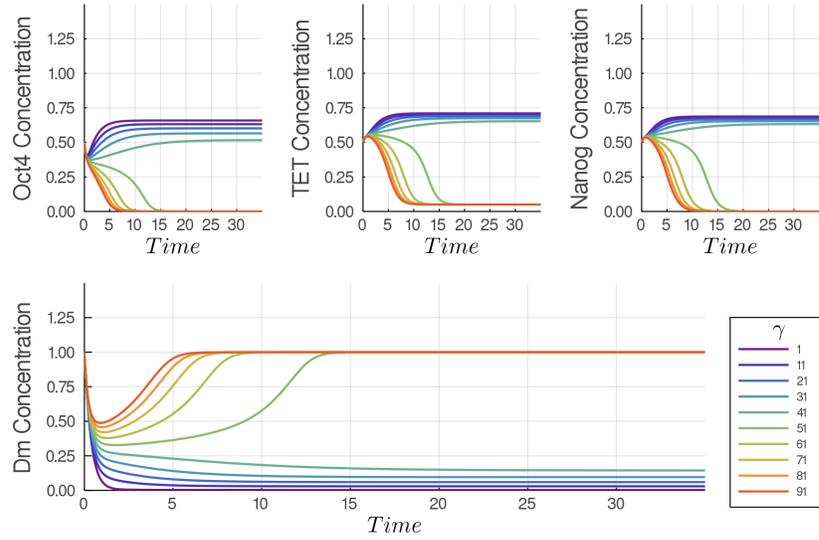
Supplementary Figure 6: Trajectories of Nanog, TET, Oct4 and Dm for different methylation rates γ. All the trajectories have an initial value 0.5 for Nanong, TET and Oct4, while D_m is initialized at 1. The strength of γ is color-coded from blue to red. The parameter set for this plot is given by: $K_O = 0.3$, $K_{nt} = 0.2$, $K_d = 0.1$, $a = 1$, $a_{nt} = 500$, $a_O = 500$, $\alpha_T = 1$, $\alpha_O = 1$, $\alpha_N = 1$, $\delta = 1$, $\theta = 1$, $m1 = 0$, $m2 = 0.05$, and $m3 = 0$.
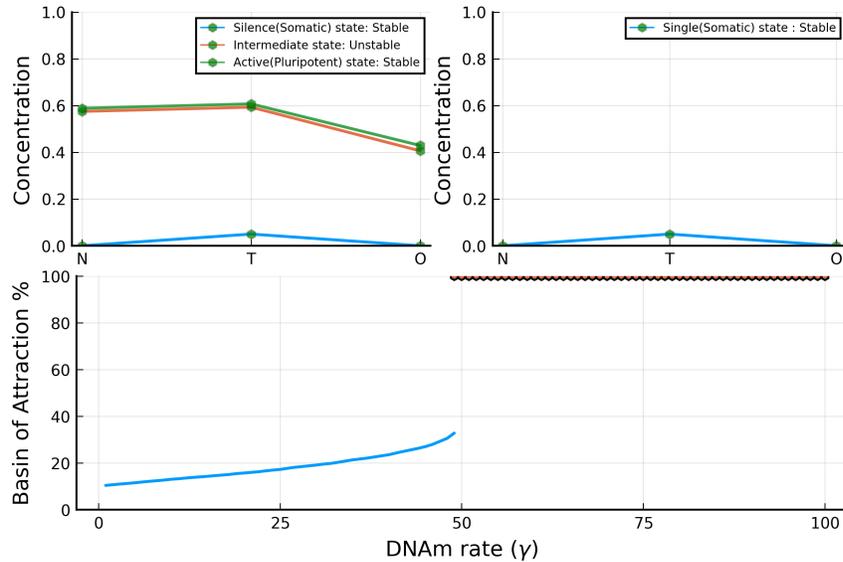
Supplementary Figure 7: Transition from bistability to monostability. The upper left plot shows the steady state values of Nanog, TET, Oct4 displaying bistability at $\gamma = 49$. The upper right plot shows the monostability when γ was increased to 50. The lower panel shows how the BoAp changes as function of γ. Note that when $\gamma > 50$, the BoAp becomes 100% which means that there is only one steady state. The parameter set for this plot is given by: $K_O = 0.3$, $K_{nt} = 0.2$, $K_d = 0.1$, $a = 1$, $a_{nt} = 500$, $a_O = 500$, $\alpha_T = 1$, $\alpha_O = 1$, $\alpha_N = 1$, $\delta = 1$, $\theta = 1$, $m1 = 0$, $m2 = 0.05$, and $m3 = 0$.

6
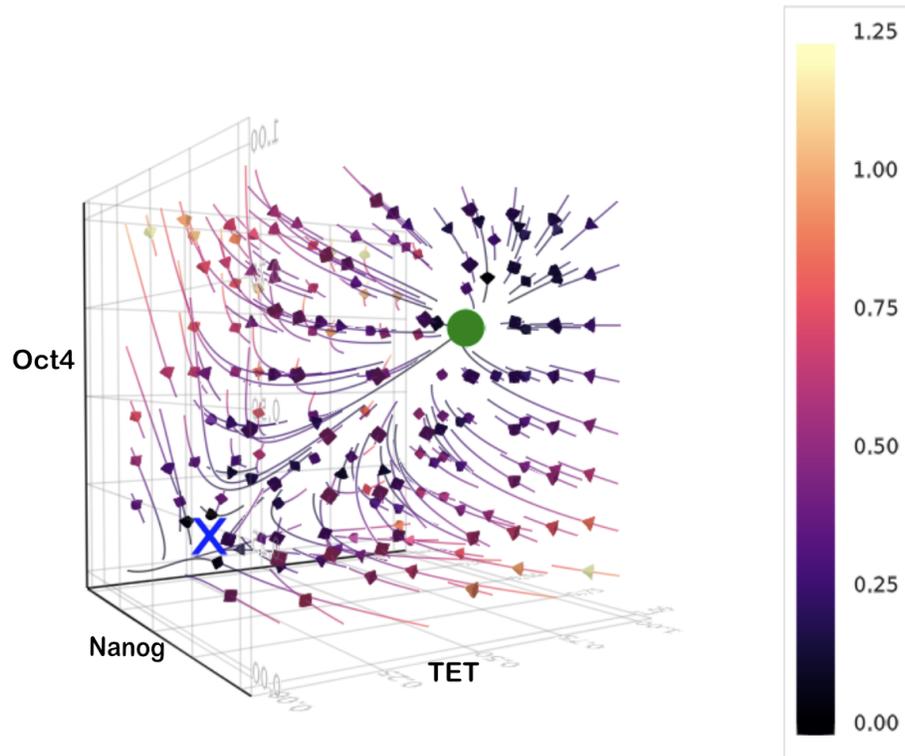
Supplementary Figure 8: 3D trajectories of Nanog, TET and Oct4. There are two steady states in the state space. The blue cross represents the silenced steady state (somatic state) in which all the three genes are lowly expressed. The green dot represents the active state (pluripotent state) where all the three genes are highly expressed. The color range specifies the norm of the 3D velocity vector. A brighter color denotes a higher velocity. The parameters are given in the caption of Figure 7.
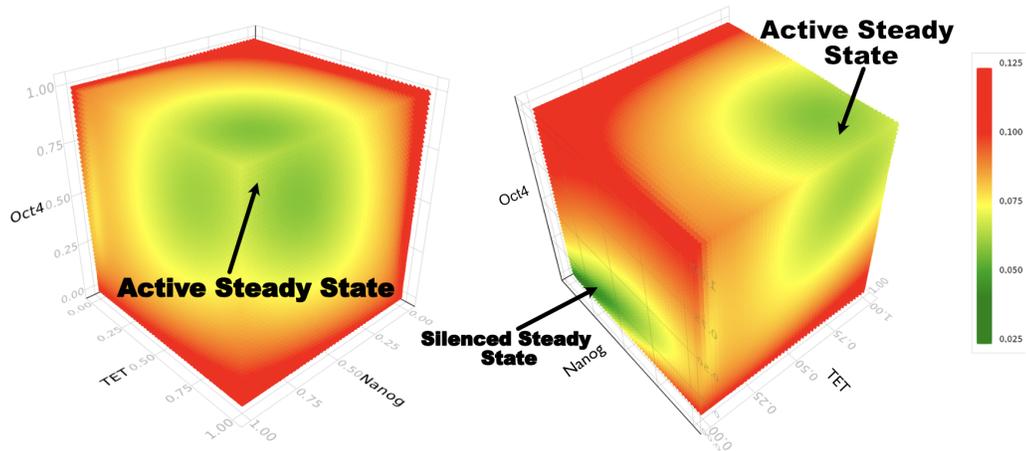
Supplementary Figure 9: The 3D velocity plot. The plot shows the norm of the velocity of the dynamical system. The left panel and right panel are the same visualization from two different angles showing the two stable steady states for the underlying dynamical system. The two green regions correspond to the two steady states. The parameters are given in the caption of Figure 7.