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

February 3, 2020

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, 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 methylation plays in enhancing the stability of the silenced gene state.

1 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* [51, 54]) provides a way to visualize this process. Well-known examples of CFNs are the PU.1/GATA.1 gene regulatory network in hematopoietic progenitors [10, 19], the pluripotent stem cell network [19, 11], and a circuit including SNAIL, miR-34, ZEB1, and miR-200 that regulates epithelial-mesenchymal transitions in tumor metastasis [37].

Phenotypes associated to CFNs had classically been seen as irreversible. However, in their pioneering work [47], 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 [40]. 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 [34, 15]. Such a quantitative understanding could have a huge impact in the field of regenerative medicine and stem cell therapy [35, 53].

Traditionally, CFNs have been modeled within the wider theoretical framework of Gene Regulatory Networks (GRNs) [25, 28]. 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 [3, 14]. When used to model a CFN, a GRN must be able to display *multistability*. This means that the asymptotic expression levels of genes can settle on multiple distinct steady states, each corresponding to a distinct cell lineage. A well-known multistable GRN is the *toggle switch* [22] whose architecture consists of two mutually inhibiting genes. Similar architectures can also give rise to multistability and have been subject of great interest [54].

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 has been both theoretical [30, 20, 21] and experimental [8] 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 [48]. DNA methylation, in particular, has been well-studied in the context of developmental CFNs [32]. Methylation is a highly heritable, hard to reverse, and robust silencer of genes [6].

In this work, we develop mathematical models of GRNs that incorporate DNA methylation, and thus can quantitatively explain its effect on gene silencing. As a metric to quantify the gene silencing abilities of DNA methylation, 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 methylation rates or the time scale of methylation.

We carry out this program computationally, illustrating it with a three-gene model of the pluripotent stem cell network, and quantifying how methylation 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.

Basic biological concepts and definitions

DNA methylation as epigenetic regulation

DNA methylation is one of the main epigenetic regulation mechanisms studied in modern molecular cell biology. Methylation plays a crucial role [13] 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 [39]. Methylation of promoter regions directly affects TF-promoter binding kinetics, and is often associated with transcriptional repression [42].

The current understanding of the DNA methylation 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

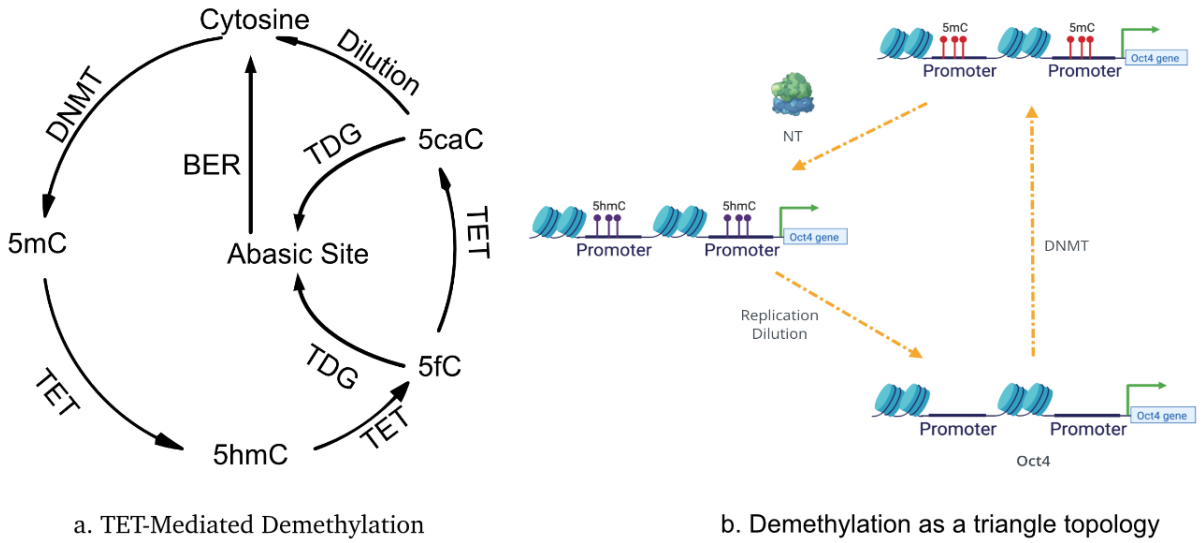


Figure 1: The Methylation Cycle. (a) The TET-mediated Cytosine Demethylation Cycle. (b) A simplified de-methylation cycle for Oct4 in the PSCC.

5caC forms. The cycle closes by a TET/TDG (thymineDNAglycosylase)/BER (base excision repair)-dependent pathway that restores the unmodified cytosine state [31].

Experiments have shown that hydroxymethylation and formylation are relatively stable cytosine modifications in genomic DNA of both dividing and nondividing cells [5, 4]. 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 DNA methylation dynamics is crucial for understanding how epigenetics regulates CFNs. DNA methylation 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 DNA methylation 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 [47, 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 [47], 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 [9], and much research has been devoted to elucidating the interplay between the Nanog, Oct4 and TeT1 proteins [50, 16]. Recent experiments regarding the induction of the pluripotent state in PSCCs have highly heterogeneous results [49, 40]. There is evidence [50] that Oct4 over-expression is sufficient

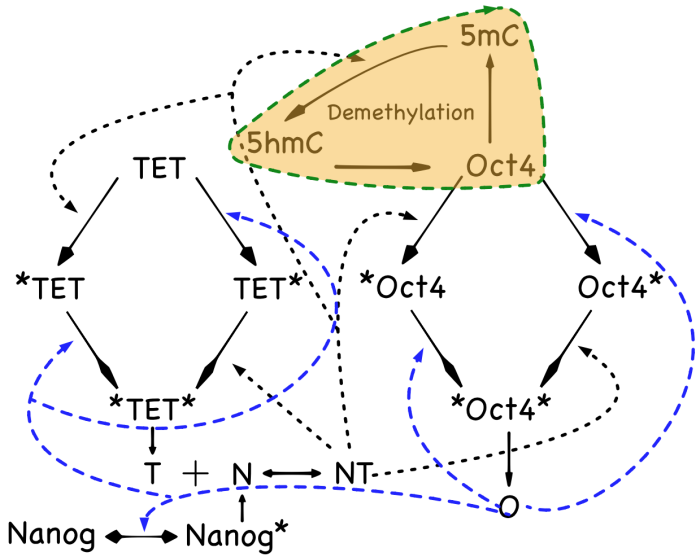


Figure 2: A detailed diagram of the PSCC network showing N, T, O interactions. TET, Oct4, Nanog denote the unbound genes. *TET and *Oct4 denote NT binding to the respective promoter sites, TET* and Oct4* denote Oct4 bound to the respective promoter sites, and *TET*, *Oct4* denote both NT and Oct4 binding 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 colored shaded area emphasizes the DNA demethylation cycle

to induce the pluripotent state from somatic cells. More recent research [12, 43] 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, DNA methylation of the Oct4 gene plays a key role in the regulation of the PSCC. In this work, we model the interplay between epigenetic regulation via the Nanong-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 underling genotypic level interaction. Although the full modeling of epigenetic regulation at the molecular level would involve many additional epigenetic regulation mechanisms beyond methylation, such as histone modifications [31], the timescale of DNA methylation is relatively slow [44], and methylation 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 [28, 33]. To further quantify the stability boundary of such attractors, researchers have been using the concept of basin of attraction which is standard in dynamical systems when analyzing multi-stable systems, and has recently been systematically used by the physics community in that context [36]. 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

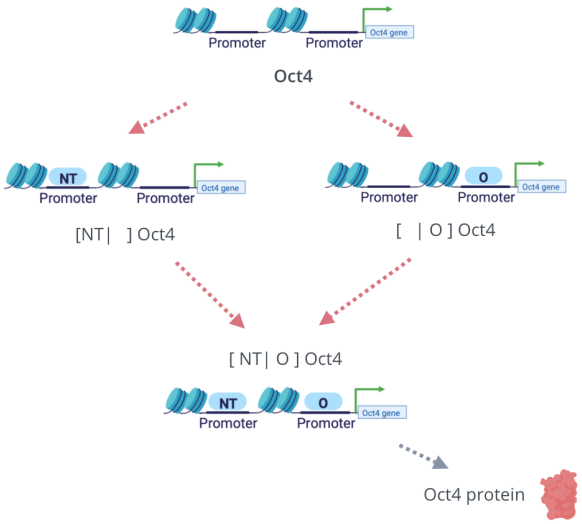


Figure 3: The **Diamond Topology** for double site TF-gene binding. In our model, it describes both Oct4 and TET promoters binding mechanism. Considering Oct4 as an example, there are two pathways for the activation of its promoter. Each of NT and Oct4 can bind to the corresponding Oct4 binding site. When both NT and Oct4 bind to the promoter region of the Oct4 gene, the Oct4 promoter will be activated.

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:

$$\text{BoAp} := \frac{\text{Volume of the BoA of } s_0}{\text{Volume of the region of interest}} . \tag{1}$$

Equipped with this definition, we can study the effect of methylation and demethylation constants on BoAp. 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.

2 Results

2.1 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 [7] 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 Nanong genes has the diamond topology as depicted in Figure 3. Both the Oct4 and the TET gene promoter binding mechanisms will follow the diamond topology, 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 4. The parameter

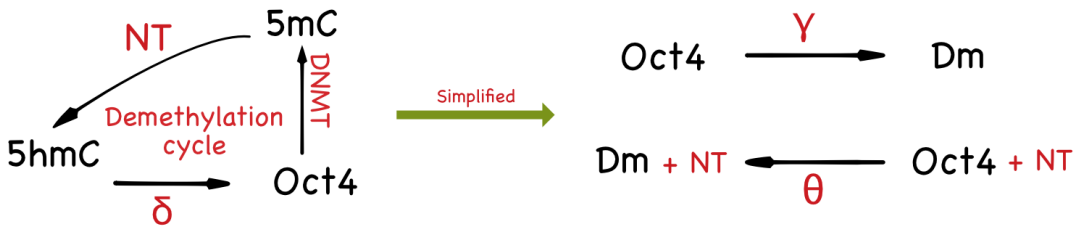


Figure 4: The diagram shows how the triangle topology can be further simplified into a two-species CRN. Since the 5hmC methylation state has a slower dynamics compared to the 5mC methylation state, one may employ a quasi-steady states approximation to reduce the triangle topology to the two-state model depicted on the right.

Oct4 promoter (R1-R4)	$D^O_{00} + [NT] \underset{K_{nt}*a_{nt}}{\overset{a_{nt}}{\rightleftharpoons}} D^O_{10}$ $D^O_{01} + [NT] \underset{K_{nt}*a_{nt}}{\overset{a_{nt}}{\rightleftharpoons}} D^O_{11}$ $D^O_{00} + O \underset{K_O*a_O}{\overset{a_O}{\rightleftharpoons}} D^O_{01}$ $D^O_{10} + O \underset{K_O*a_O}{\overset{a_O}{\rightleftharpoons}} D^O_{11}$	TET promoter (R5-R8)	$D^T_{00} + [NT] \underset{K_{nt}*a_{nt}}{\overset{a_{nt}}{\rightleftharpoons}} D^T_{10}$ $D^T_{01} + [NT] \underset{K_{nt}*a_{nt}}{\overset{a_{nt}}{\rightleftharpoons}} D^T_{11}$ $D^T_{00} + O \underset{K_O*a_O}{\overset{a_O}{\rightleftharpoons}} D^T_{01}$ $D^T_{10} + O \underset{K_O*a_O}{\overset{a_O}{\rightleftharpoons}} D^T_{11}$
Nanog promoter (R9)	$D^N_0 + O \underset{K_O*a_O}{\overset{a_O}{\rightleftharpoons}} D^N_1$	Nanog-TET Dimer (R10)	$N + T \underset{K_d*a}{\overset{a}{\rightleftharpoons}} [NT]$
Protein decay and production (R11-R13)	$N \xrightarrow{\delta} \emptyset; D^N_1 \xrightarrow{\alpha_N} D^N_1 + N$ $T \xrightarrow{\delta} \emptyset; D^T_{11} \xrightarrow{\alpha_T} D^T_{11} + T$ $O \xrightarrow{\delta} \emptyset; D^O_{11} \xrightarrow{\alpha_O} D^O_{11} + O$	Basal Rate (R14-R16)	$\emptyset \xrightarrow{m_1} N$ $\emptyset \xrightarrow{m_2} T$ $\emptyset \xrightarrow{m_3} O$
Methylation (R17)	$D^O_{00} \xrightarrow{\gamma} D_m$	Demethylation (R18)	$D_m + NT \xrightarrow{\theta} D^O_{00} + NT$

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

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 various species in a population.

2.1.1 Full Model Description

Description of each module

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 4 reactions. In our modeling context, we assume that the Oct4 gene has two binding sites. The first binding site is for Oct4 protein itself [52, 24], and the second binding site is for Nanog-TET heterodimer complex [43] as shown in the diamond topology in Figure 3. 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 [50, 43, 31]. 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 [24].

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 [43]. 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, 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 [50].

7. Oct4 promoter Methylation module (R17): This module describes the de novo methylation of the Oct4 gene promoter region with a methylation rate γ.

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 4)

Model reduction via quasi-steady steady approximation

The full model CRN is computationally intensive to simulate and conceptually difficult to interpret its dynamical behavior. Biologically speaking, it is often the case that the DNA methylation dynamics operates on a slower time scale than the protein level's dynamics [44]. Therefore, we make the following reasonable assumptions.

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. Note that, unlike TF-promoter binding/unbinding, DNA methylation kinetics are as slow as protein kinetics or even slower [44].

The full model ODEs are given in the Methods section. The reduced dynamical system's ODEs can be written as below:

$$
\begin{aligned}
\frac{dN}{dt} =& m_1 - \delta N + \frac{\alpha_N O}{K_O + O}, \\
\frac{dT}{dt} =& m_2 - \delta T + \frac{\alpha_T NTO}{(NT + K_{nt}K_d)(K_O + O)}, \\
\frac{dO}{dt} =& \frac{-\left(K_{nt}K_d\delta O^2 + \delta NTO^2 - K_O K_{nt} K_d m_3 - \alpha_O NTO - K_{nt}K_d m_3 O - K_O m_3 NT - m_3 NTO + K_O \delta NTO + K_O K_{nt} K_d \delta O + \alpha_O D_m NTO\right)}{(NT + K_{nt}K_d)(K_O + O)}, \\
\frac{dD_m}{dt} =& \frac{-\left(K_O K_{nt} K_d^2 \gamma D_m - K_O K_{nt} K_d^2 \gamma + \theta K_O N^2 T^2 D_m + \theta N^2 T^2 O D_m + \theta K_O K_{nt} K_d NTD_m + \theta K_{nt} K_d NTOD_m\right)}{K_d (NT + K_{nt}K_d)(K_O + O)},
\end{aligned}
$$

2.1.2 Effect of methylation rate

For the full model, Figure 5 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, α_T, α_O, α_N, δ, γ, θ, 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 higher the effective methylation rate, the more likely it is for the underlying dynamical system to stay at the somatic cell state, which is represented by a larger BoAp. The phenotypic somatic state is characterized [50] by low expression of Oct4, Nanog, TET, and methylated promoters. With simulations performed over 6200 parameter sets as shown in Figure 5, the BoAp vs. γ plot shows that the BoAp is monotonically increasing with respect to γ. Note the different behavior 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 $10^{-1.5}$.

On the other hand, changing the reaction rate θ is equivalent to changing the effective 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 by the NT complex and the rate of oxidized 5hmC reverting to the unbound promoter state via either TDG-BER pathway or dilution via cell replication. In Figure 6, 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. 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 plays in enhancing the stability of the somatic state.

2.1.3 Effect of slow/fast methylation kinetics

To understand how the relative methylation rate affects the BoAp, we again define the methylation association rate as $R = \frac{\gamma}{\theta}$. We simulate the full model across different fixed ratios of R, letting γ range over 4 orders of magnitude.

The simulations in Figure 7 show that BoAp vs. γ has a biphasic trend across 5 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.

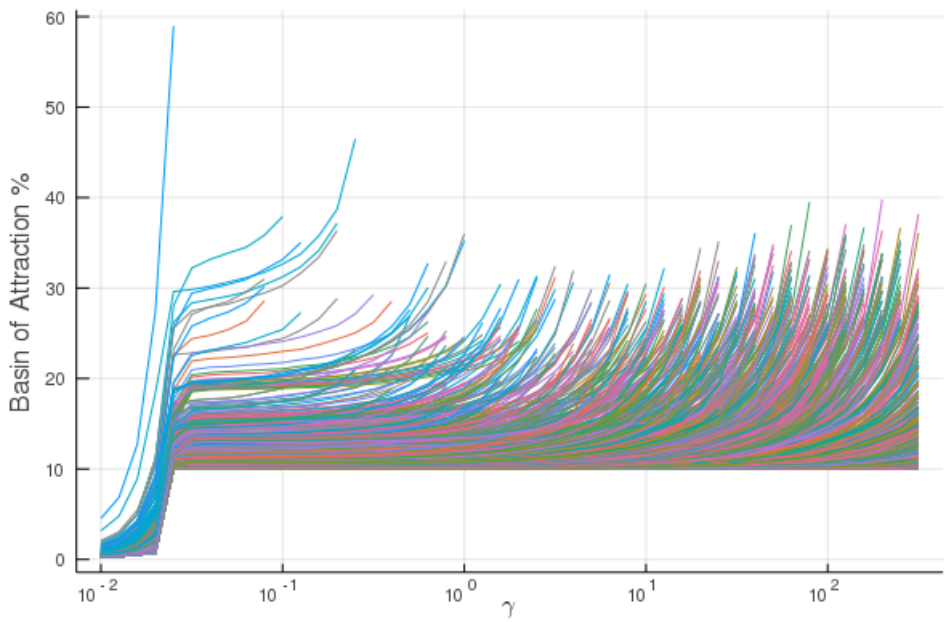


Figure 5: BoAp vs γ simulated with a large sets of different rates for the full model. The BoAp increasing rate exhibits a sudden slowing down at around $10^{-1.5}$

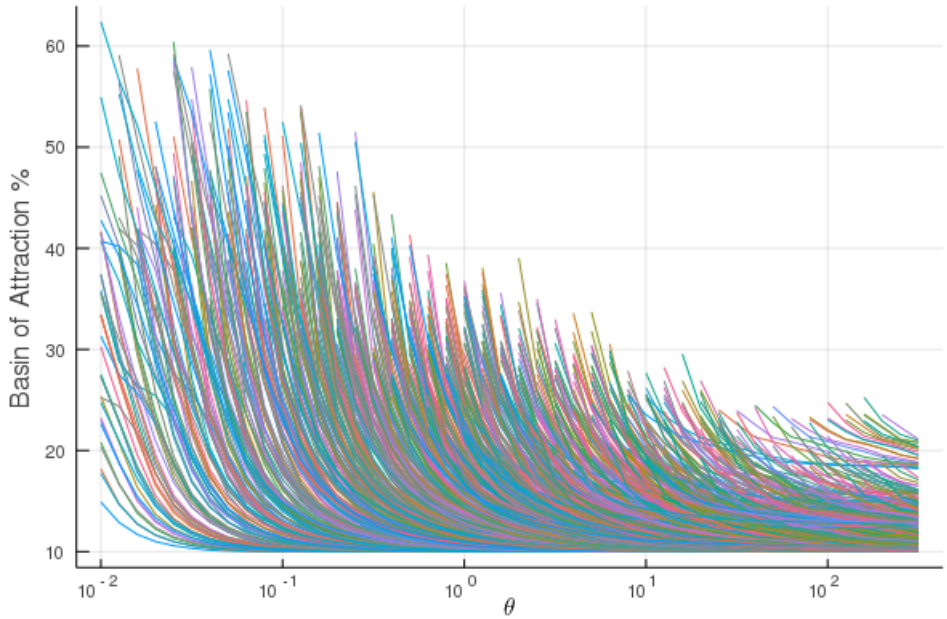


Figure 6: BoAp vs θ for a large set of different rates for the full model. This plot shows BoAp is monotonically decreasing with respect to the effective demethylation rate θ.

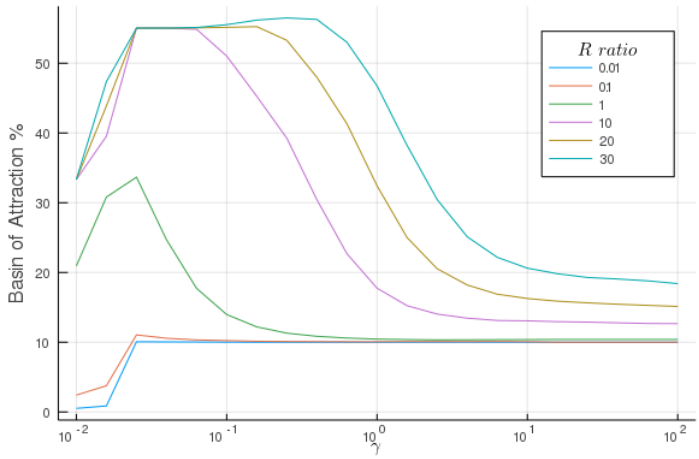


Figure 7: BoAp vs γ while θ is increased so as to keep the ratio R constant. The kinetic rate parameters are: K_O = 0.3, K_{nt} = 0.2, K_d = 0.1, a= 1, a_{nt} = 100, a_O = 100, α_T = 1, α_O = 1, α_N = 1, δ = 1, $R \in \{0.01, 0.1, 1, 10, 20, 30\}$, $\gamma \in [10^{-2}, 10^{2}]$.

The biphasic nature of the trends is not a result of high-dimensional dynamics. Rather, we will re-observe the same behavior in our subsequent study of the single gene model. Furthermore, the biphasic trend for the fixed ratio R can roughly be inferred from the PSCC simulations with varying each γ and θ independently . As per the trend we shown in Figure 5, the BoAp increasing rate exhibits a sudden slowing down at around $10^{-1.5}$, while there is no corresponding change for the demethylation rate θ shown in Figure 6. Therefore, it is not surprising that the methylation association ratio R has a biphasic behaviour.

2.2 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 modeled 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:

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

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

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

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

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.

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:

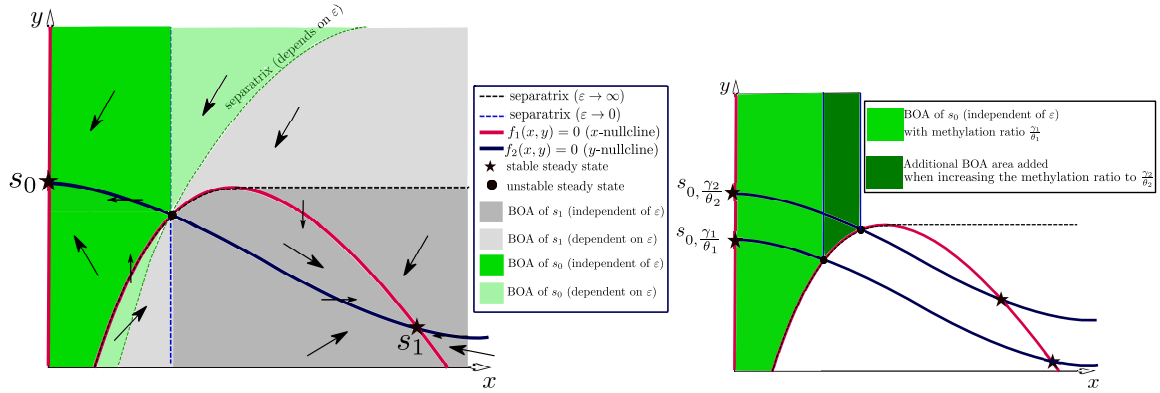
$$\begin{aligned} \dot{x} &= k(1-y)\frac{x^2}{B + x^2} - \delta x \\ \varepsilon\dot{y} &= \gamma(1-y)\frac{B}{B + x^2} - \theta y\,, \end{aligned} \tag{11}$$

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 analyze this reduced system.

2.2.1 Global behavior

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



(a) (b)

Figure 8: 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.

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 analyze the global behavior 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 [45] (see also [46] for an exposition and biological applications). In [26], 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 [45]. 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 analyzed via nullclines. A summary of the analysis is depicted in Figure 8 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 BoA's, is ε-dependent. It is shown that each BoA can be partitioned into ε-dependent and ε-independent regions. Figure 8-b) shows that the area of the ε-independent BoA of s_0 increases as the methylation ratio is increased.

2.2.2 Slow methylation kinetics

Since DNA methylation is a slow process, we focus on case $\varepsilon \to \infty$ in this subsection. Figure 8 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:* The arc is $x = 0$, which gives the slow dynamics of

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

   It is a linear system with asymptotically stable steady state at $y = \frac{B\gamma/\theta}{B\gamma/\theta+1}$. Notice that it is increasing with respect to γ/θ.

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

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

   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 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 is that the quantity under the square root in (12) is real. Hence, we have:

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

Hence, $y(0) > 1 - \frac{2\delta\sqrt{B}}{k}$ automatically implies that the silenced arc represents the dynamics regardless of the initial $x(0)$.

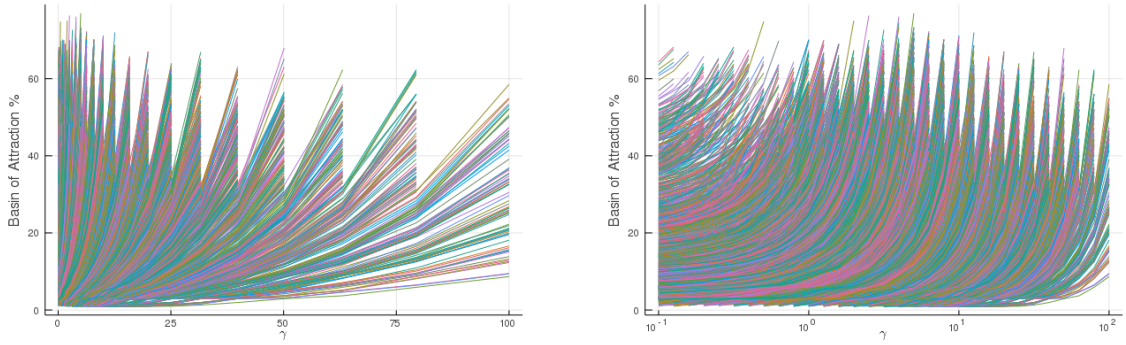
Assuming the necessary condition is satisfied, we need also that

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

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 gray region in Figure 8-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. 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



(a) Linear Scale

(b) Log Scale

Figure 9: Plots of BoAp vs γ, for a large set of different parameters for the single gene model. It can be seen that the BoAp is monotonically increasing with respect to the methylation rate.

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

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

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

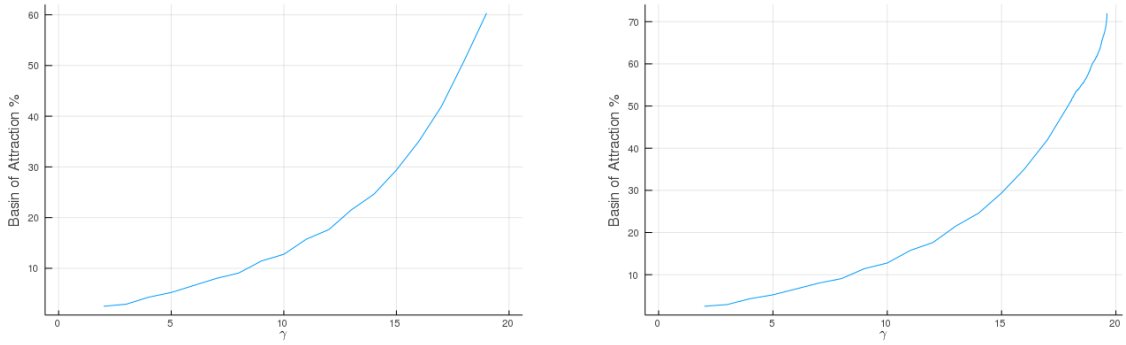
If both of these conditions are satisfied, then overexpression will steer the trajectory to the active arc if the following holds:

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

In a nutshell, eqns. (15), (16) *imply that we need a sufficiently high expression ratio, a sufficiently high overexpression value, and sufficiently low methylation ratio.*

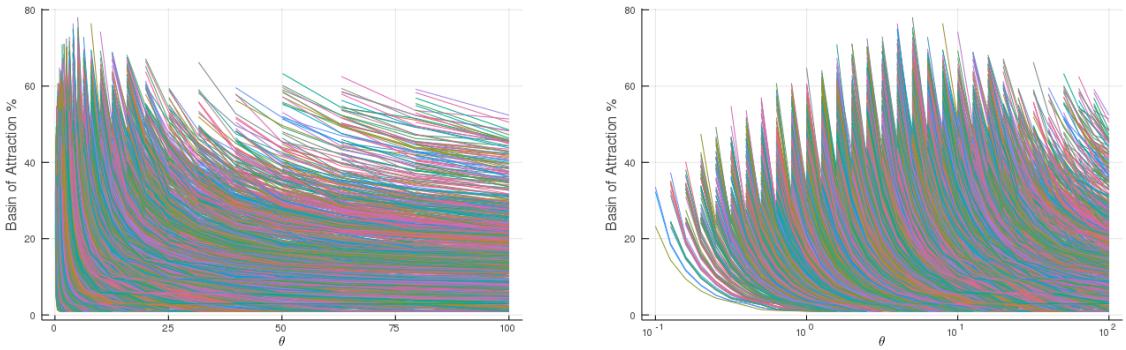
2.2.3 Numerical simulation results for the single-gene model

As a complement to the theoretical discussion, we carried out numerical experimentation to confirm the preceeding nullcline analysis. Figure 9 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 9, 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



(a) Grid 1 (b) Grid 2

Figure 10: 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: β = 5, β_- = 5, k =5, δ = 4, θ = 5. We vary γ from 0 to 20.



(a) Linear Scale (b) Log Scale

Figure 11: BoAp vs θ for a large set of different rates for the single gene model. It can be seen that the BoAp is monotonically decreasing with respect to the demethylation rate.

effective demethylation on the promoter dominates, and therefore a cell would be more likely to stay in a pluripotent state.

We display the simulations both in linear scale, shown in Figure 9(a), and in log scale, shown in Figure 9(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 10, 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 11 over different parameter sets that for : β, β_-, k, γ, θ, δ. We describe the parameter sampling method in the Methods section.

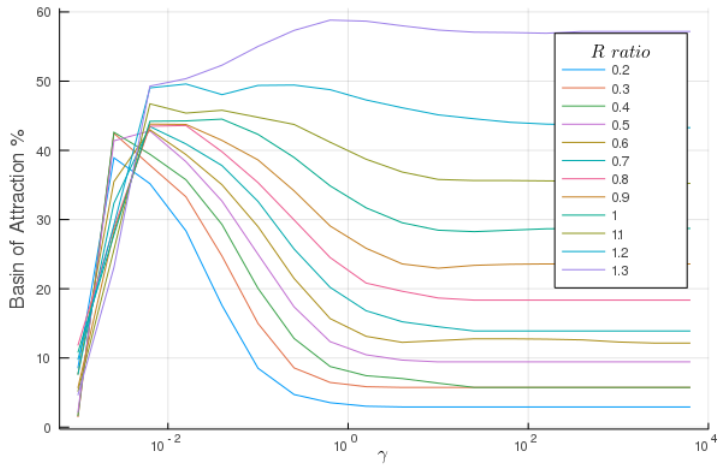


Figure 12: BoAp vs. γ with a range of fixed R values for the single gene model. For each R ratio that allows the bistability of the single gene model, the BOAp vs. γ plot has a biphasic nature based on the time scale of the (de)methylation dynamics

2.2.4 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 12, 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 considered to be large. The simulation in Figure 12 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 8. Indeed, as (de)methylation dynamics becomes very fast, the ϵ in Figure 8, 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 12.

Comparison of the two models. The results for the single gene model 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 Nanong 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 behavior 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 epignetic process –in our case, a meythlation cycle– shifts the stability boundary of the underlying system consistently in the same direction.

3 Discussion

We have proposed a GRN model that combines TF-promoter binding/unbinding and gene silencing via DNA methylation. We are currently working on extending the methodology to other types of epigenetic modifications, including histone modifications and chromatin remodeling.

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 DNA methylation 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 DNA methylation has been proposed in [20]. In addition to lacking mechanistic interpretations, Boolean models do not allow one to quantify the effect of DNA methylation on the stability boundary between the multiple phenotypes. As the role of the TET protein family in gene regulation became clearer [29, 31, 41, 43], various CRN models and ODE models that have partially considered TET-mediated demethylation cycle [50, 1] and core gene-TF regulatory mechanisms have been proposed. Notably in [1], 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 modeled the DNA methylation 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 DNA methylation model in [1] did not include a diamond topology-like mechanism for gene activation, it represented an early attempt to mechanistically link gene regulation and chromatin modifications. In [50], 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 [50] was not spelled-out clearly. In addition to this, without the NT complex involved in the DNA demethylation process, the model in [50] 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 [17], 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 [17] 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 [38], 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 [30]. 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 modeling approach using CRN agrees with the intuition on how DNA methylation affects the cell fate network. The faster the methylation rate the more stable the somatic cell 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.

As for future directions, even though the diamond topology 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 [27]. 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.

4 Methods

4.1 Basic mathematical concepts and definitions

Mathematically, our model is initially described as a Chemical Reaction Network (CRN). CRN's 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 since DNA methylation is commonly understood as a slow process at the level of promoters. Following [2], 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. [18]. A CRN is specified by a set of species $\mathscr{S} = \{Z_1, .., Z_n\}$ and a set of reactions $\mathscr{R} = \{\mathrm{R}_1, ..., \mathrm{R}_\nu\}$. A reaction $\mathrm{R_j}$ 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 $\Gamma \in \mathbb{R}^{n\times\nu}$ is defined elementwise as $[\Gamma]_{ij} = \beta_{ij} - \alpha_{ij}$. Each reaction $\mathrm{R_j}$ can occur with a rate function $R_j : \mathbb{R}^n_{\geq 0} \rightarrow \mathbb{R}^\nu_{\geq 0}$. 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 $z(t) \in \mathbb{R}^n_{\geq 0}$ be the vector of species concentrations at time t, the associated ODE can be written as: $\dot{z} = \Gamma R(z)$, where $R := [R_1, .., R_\nu]^T$.

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.

4.2 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 [23].

4.3 ODE system for the full CRN model

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

[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.

$$
\begin{aligned}
\frac{dN(t)}{dt} &= -aNT + ak_d[NT] + \alpha_N D_1^N - \delta N + m_1 \\
\frac{dT(t)}{dt} &= -aNT + ak_d[NT] + \alpha_T D_{11}^T - \delta T + m_2 \\
\frac{d[NT](t)}{dt} &= aNT - ak_d[NT] - a_{nt}[NT]D_{00}^O + K_{nt}a_{nt}D_{10}^O - a_{nt}[NT]D_{01}^O + K_{nt}a_{nt}D_{11}^O \\
&\quad - a_{nt}[NT]D_{00}^T + K_{nt}a_{nt}D_{10}^T - a_{nt}[NT]D_{01}^T + K_{nt}a_{nt}D_{11}^T \\
\frac{dD_{00}^O(t)}{dt} &= -a_{nt}[NT]D_{00}^O + K_{nt}a_{nt}D_{10}^O - a_O O D_{00}^O + K_O a_O D_{01}^O - \gamma D_{00}^O + \theta D_m[NT] \\
\frac{dD_{10}^O(t)}{dt} &= a_{nt}[NT]D_{00}^O - K_{nt}a_{nt}D_{10}^O - a_O O D_{10}^O + K_O a_O D_{11}^O \\
\frac{dD_{01}^O(t)}{dt} &= -a_{nt}[NT]D_{01}^O + K_{nt}a_{nt}D_{11}^O + a_O O D_{00}^O - K_O a_O D_{01}^O \\
\frac{dD_{11}^O(t)}{dt} &= a_{nt}[NT]D_{01}^O - K_{nt}a_{nt}D_{11}^O + a_O O D_{10}^O - K_O a_O D_{11}^O \\
\frac{dO(t)}{dt} &= -a_O O D_{00}^O + K_O a_O D_{01}^O - a_O O D_{10}^O + K_O a_O D_{11}^O - a_O O D_{00}^T + K_O a_O D_{01}^T \\
&\quad - a_O O D_{10}^T + K_O a_O D_{11}^T - a_O O D_0^N + K_O a_O D_1^N + \alpha_O D_{11}^O - \delta O + m_3 \\
\frac{dD_{00}^T(t)}{dt} &= -a_{nt}[NT]D_{00}^T + K_{nt}a_{nt}D_{10}^T - a_O O D_{00}^T + K_O a_O D_{01}^T \\
\frac{dD_{10}^T(t)}{dt} &= a_{nt}[NT]D_{00}^T - K_{nt}a_{nt}D_{10}^T - a_O O D_{10}^T + K_O a_O D_{11}^T \\
\frac{dD_{01}^T(t)}{dt} &= -a_{nt}[NT]D_{01}^T + K_{nt}a_{nt}D_{11}^T + a_O O D_{00}^T - K_O a_O D_{01}^T \\
\frac{dD_{11}^T(t)}{dt} &= a_{nt}[NT]D_{01}^T - K_{nt}a_{nt}D_{11}^T + a_O O D_{10}^T - K_O a_O D_{11}^T \\
\frac{dD_0^N(t)}{dt} &= -a_O O D_0^N + K_O a_O D_1^N \\
\frac{dD_1^N(t)}{dt} &= a_O O D_0^N - K_O a_O D_1^N \\
\frac{dD_m(t)}{dt} &= \gamma D_{00}^O - \theta D_m[NT]
\end{aligned}
$$

4.4 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.

4.5 Parameters Range

The parameters are constrained to the set which can give rise to bistability.

4.5.1 Kinetic Rates

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 the Table 1. Although the values for the three dissociation parameters K_O, K_{nt}, K_d were given in [50], 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.

4.5.2 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 are depend on the parameter set we choose, but the qualitative trend for all the parameter sets we tested are the same and show a biphasic trend. In the case of plot in Figure 12, we have chosen the following parameter set: $\beta = 4.4$, $\beta_- = 1.8$, $k = 2.5$, $\delta = 4.6$, and we have varied R from 0.2 to 1.3 with γ spanning the range from 10^{-3} to 10^4. For the full model in Figure 7 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.

5 Data availability

N/A

6 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.

References

[1] H. Abdallah. 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.

[2] M. Ali Al-Radhawi, D. Del Vecchio, and E. D. Sontag. Multi-modality in gene regulatory networks with slow promoter kinetics. *PLoS Computational Biology*, 15(2):e1006784, 2019.

[3] U. Alon. *An Introduction to Systems Biology*. CRC press, 2006.

[4] M. Bachman, S. Uribe-Lewis, X. Yang, M. Williams, A. Murrell, and S. Balasubramanian. 5-hydroxymethylcytosine is a predominantly stable DNA modification. *Nature chemistry*, 6(12): 1049, 2014.

[5] M. Bachman, S. Uribe-Lewis, X. Yang, H. E. Burgess, M. Iurlaro, W. Reik, A. Murrell, and S. Balasubramanian. 5-formylcytosine can be a stable DNA modification in mammals. *Nature chemical biology*, 11(8):555, 2015.

[6] S. B. Baylin. DNA methylation and gene silencing in cancer. *Nature Clinical Practice Oncology*, 2(1):S4–S11, 2005. doi: 10.1038/ncponc0354.

[7] S. Ben-Tabou de Leon and E. H. Davidson. Modeling the dynamics of transcriptional gene regulatory networks for animal development. *Developmental biology*, 325(2):317–328, 01 2009. doi: 10.1016/j.ydbio.2008.10.043.

[8] L. Bintu, J. Yong, Y. E. Antebi, K. McCue, Y. Kazuki, N. Uno, M. Oshimura, and M. B. Elowitz. Dynamics of epigenetic regulation at the single-cell level. *Science*, 351(6274):720–724, 2016.

[9] A. Blais and B. D. Dynlacht. Constructing transcriptional regulatory networks. *Genes & Development*, 19(13):1499–1511, 2005. doi: 10.1101/gad.1325605.

[10] P. Burda, P. Laslo, and T. Stopka. The role of PU.1 and GATA-1 transcription factors during normal and leukemogenic hematopoiesis. *Leukemia*, 24:1249–1257, 2010. ISSN 0887-6924. doi: 10.1038/leu.2010.104.

[11] S. T. Chou, E. Khandros, L. C. Bailey, K. E. Nichols, C. R. Vakoc, Y. Yao, Z. Huang, J. D. Crispino, R. C. Hardison, G. A. Blobel, and M. J. Weiss. Graded repression of PU.1/Sfpi1 gene transcription by gata factors regulates hematopoietic cell fate. *Blood*, 114(5):983–994, 07 2009. doi: 10.1182/blood-2009-03-207944.

[12] Y. Costa, J. Ding, T. W. Theunissen, F. Faiola, T. A. Hore, P. V. Shliaha, M. Fidalgo, A. Saunders, M. Lawrence, S. Dietmann, S. Das, D. N. Levasseur, Z. Li, M. Xu, W. Reik, J. C. R. Silva, and J. Wang. NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature*, 495:370, 2013. ISSN 1476-4687. doi: 10.1038/nature11925.

[13] M. Curradi, A. Izzo, G. Badaracco, and N. Landsberger. Molecular mechanisms of gene silencing mediated by DNA methylation. *Molecular and cellular biology*, 22(9):3157–3173, 05 2002. doi: 10.1128/mcb.22.9.3157-3173.2002.

[14] D. Del Vecchio and R. M. Murray. *Biomolecular Feedback Systems*. Princeton Univ Press, 2015.

[15] D. Del Vecchio, H. Abdallah, Y. Qian, and J. J. Collins. A blueprint for a synthetic genetic feedback controller to reprogram cell fate. *Cell systems*, 4(1):109–120, 2017.

[16] D. P. Dongwei Li, Jiekai Chen. The Battle between TET Proteins and DNA Methylation for the Right Cell. *Trends in Cell Biology*, 28:973–975, 2018. ISSN 0962-8924. doi: 10.1016/j.tcb.2018.10.002.

[17] M. B. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335, 2000.

[18] P. Érdi and J. Tóth. *Mathematical models of chemical reactions: theory and applications of deterministic and stochastic models*. Manchester University Press, 1989.

[19] R. C. Fisher and E. W. Scott. Role of PU.1 in Hematopoiesis. *STEM CELLS*, 16(1):25–37, 1998. ISSN 1549-4918. doi: 10.1002/stem.160025.

[20] M. Flöttmann, T. Scharp, and E. Klipp. A stochastic model of epigenetic dynamics in somatic cell reprogramming. *Frontiers in Physiology*, 3:216, 2012.

[21] N. Folguera-Blasco, R. Pérez-Carrasco, E. Cuyàs, J. A. Menendez, and T. Alarcón. A multiscale model of epigenetic heterogeneity-driven cell fate decision-making. *PLoS computational biology*, 15(4):e1006592, 2019.

[22] T. S. Gardner, C. R. Cantor, and J. J. Collins. Construction of a genetic toggle switch in Escherichia coli. *Nature*, 403(6767):339–342, 2000.

[23] K. S. Gritton, J. Seader, and W.-J. Lin. Global homotopy continuation procedures for seeking all roots of a nonlinear equation. *Computers & Chemical Engineering*, 25(7):1003 – 1019, 2001. ISSN 0098-1354. doi: https://doi.org/10.1016/S0098-1354(01)00675-5.

[24] F. Hammachi, G. M. Morrison, A. A. Sharov, A. Livigni, S. Narayan, E. P. Papapetrou, J. O'Malley, K. Kaji, M. S. Ko, M. Ptashne, et al. Transcriptional activation by Oct4 is sufficient for the maintenance and induction of pluripotency. *Cell reports*, 1(2):99–109, 2012.

[25] M. Hecker, S. Lambeck, S. Toepfer, E. van Someren, and R. Guthke. Gene regulatory network inference: Data integration in dynamic models—a review. *Biosystems*, 96(1):86–103, 2009. doi: https://doi.org/10.1016/j.biosystems.2008.12.004.

[26] M. Hirsch. Systems of differential equations which are competitive or cooperative I: limit sets. *SIAM J. Appl. Math*, 13:167–179, 1982.

[27] D. Hnisz, K. Shrinivas, R. A. Young, A. K. Chakraborty, and P. A. Sharp. A phase separation model for transcriptional control. *Cell*, 169(1):13–23, 2017.

[28] S. Huang, G. Eichler, Y. Bar-Yam, and D. E. Ingber. Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Physical Review Letters*, 94:128701, Apr 2005. doi: 10.1103/PhysRevLett.94.128701.

[29] S. Ito, A. C. D'Alessio, O. V. Taranova, K. Hong, L. C. Sowers, and Y. Zhang. Role of TET proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*, 466 (7310):1129, 2010.

[30] W. Jia, A. Deshmukh, S. A. Mani, M. K. Jolly, and H. Levine. A possible role for epigenetic feedback regulation in the dynamics of the epithelial-mesenchymal transition (EMT). *bioRxiv*, page 651620, 2019.

[31] K. H. Kasper Dindler Rasmussen. Role of TET enzymes in DNA methylation, development, and cancer. *Genes & Development*, 30:733–750, 2016. ISSN 0890-9369. doi: 10.1101/gad.276568.115.

[32] K. P. Koh and A. Rao. DNA methylation and methylcytosine oxidation in cell fate decisions. *Current Opinion in Cell Biology*, 25(2):152 – 161, 2013. ISSN 0955-0674. doi: https://doi.org/10.1016/j.ceb.2013.02.014.

[33] A. H. Lang, H. Li, J. J. Collins, and P. Mehta. Epigenetic landscapes explain partially reprogrammed cells and identify key reprogramming genes. *PLoS computational biology*, 10(8): e1003734, 2014.

[34] C. Li and J. Wang. Quantifying cell fate decisions for differentiation and reprogramming of a human stem cell network: Landscape and biological paths. *PLoS Computational Biology*, 9(8): e1003165, 2013. doi: 10.1371/journal.pcbi.1003165PMID-23935477.

[35] R. S. Mahla. Stem cells applications in regenerative medicine and disease therapeutics. *International journal of cell biology*, 2016:6940283–6940283, 2016. doi: 10.1155/2016/6940283.

[36] P. J. Menck, J. Heitzig, N. Marwan, and J. Kurths. How basin stability complements the linear-stability paradigm. *Nature Physics*, 9:89–92, 2013. ISSN 1745-2473. doi: 10.1038/nphys2516.

[37] V. Mittal. Epithelial Mesenchymal Transition in Tumor Metastasis. *Annual Review of Pathology: Mechanisms of Disease*, 13(1):395–412, 2018. doi: 10.1146/annurev-pathol-020117-043854. PMID: 29414248.

[38] T. Miyamoto, C. Furusawa, and K. Kaneko. Pluripotency, differentiation, and reprogramming: a gene expression dynamics model with epigenetic feedback regulation. *PLoS computational biology*, 11(8):e1004476, 2015.

[39] L. D. Moore, T. Le, and G. Fan. DNA methylation and its basic function. *Neuropsychopharmacology*, 38(1):23–38, 01 2013. doi: 10.1038/npp.2012.112.

[40] A. E. Omole and A. O. J. Fakoya. Ten years of progress and promise of induced pluripotent stem cells: historical origins, characteristics, mechanisms, limitations, and potential applications. *PeerJ*, 6:e4370, 2018. doi: 10.7717/peerj.4370PMID-29770269.

[41] K. D. Rasmussen and K. Helin. Role of TET enzymes in DNA methylation, development, and cancer. *Genes & development*, 30(7):733–750, 2016.

[42] A. Razin and H. Cedar. DNA methylation and gene expression. *Microbiological reviews*, 55(3): 451–458, 09 1991.

[43] J. L. Sardina, S. Collombet, T. V. Tian, A. Gómez, B. D. Stefano, C. Berenguer, J. Brumbaugh, R. Stadhouders, C. Segura-Morales, M. Gut, I. G. Gut, S. Heath, S. Aranda, L. D. Croce, K. Hochedlinger, D. Thieffry, and T. Graf. Transcription factors drive TET2-mediated enhancer demethylation to reprogram cell fate. *Cell Stem Cell*, 23(5):727 – 741.e9, 2018. ISSN 1934-5909. doi: https://doi.org/10.1016/j.stem.2018.08.016.

[44] M. Sasai, Y. Kawabata, K. Makishi, K. Itoh, and T. P. Terada. Time scales in epigenetic dynamics and phenotypic heterogeneity of embryonic stem cells. *PLoS computational biology*, 9(12): e1003380, 2013.

[45] H. Smith. *Monotone Dynamical Systems*. AMS, Providence, RI, 1995.

[46] E. D. Sontag. Monotone and near-monotone biochemical networks. *Systems and synthetic biology*, 1(2):59–87, 2007.

[47] K. Takahashi and S. Yamanaka. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*, 126:663–676, 2006. ISSN 0092-8674. doi: 10.1016/j.cell.2006.07.024.

[48] K. Takahashi and S. Yamanaka. A developmental framework for induced pluripotency. *Development*, 142(19):3274–3285, 2015. ISSN 0950-1991. doi: 10.1242/dev.114249.

[49] W. J. Ting Geng, Donghui Zhang. Epigenetic Regulation of Transition Among Different Pluripotent States. *STEM CELLS*, 8 2019. ISSN 1066-5099. doi: 10.1002/stem.3064.

[50] K. S. Victor Olariu, Cecilia Lvkvist. Nanog, Oct4 and TET1 interplay in establishing pluripotency. *Scientific Reports*, 6:25438, 2016. ISSN 2045-2322. doi: 10.1038/srep25438.

[51] C. H. Waddington. *The strategy of the genes*. George Allen and Unwin, 1957.

[52] J. Wang, S. Rao, J. Chu, X. Shen, D. N. Levasseur, T. W. Theunissen, and S. H. Orkin. A protein interaction network for pluripotency of embryonic stem cells. *Nature*, 444(7117):364, 2006.

[53] W. Zakrzewski, M. Dobrzyński, M. Szymonowicz, and Z. Rybak. Stem cells: past, present, and future. *Stem Cell Research & Therapy*, 10(1):68, 2019. doi: 10.1186/s13287-019-1165-5.

[54] J. X. Zhou and S. Huang. Understanding gene circuits at cell-fate branch points for rational cell reprogramming. *Trends in Genetics*, 27(2):55–62, 2011.