Cell fate reprogramming using transcription factor feedback overexpression

by

Nithin Senthur Kumar

Submitted to the Department of Mechanical Engineering
in partial fulfillment of the requirements for the degree of

Mechanical Engineer's Degree

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2018

Signature redacted

Author .....
Department of Mechanical Engineering
May 11, 2018

Signature redacted

Certified by .....
Domitilla Del Vecchio
Professor
Thesis Supervisor

Signature redacted

Accepted by .....
Rohan Abeyaratne
Chairman, Department Committee on Graduate Theses

Cell fate reprogramming using transcription factor feedback overexpression

by

Nithin Senthur Kumar

Submitted to the Department of Mechanical Engineering
on May 14, 2018, in partial fulfillment of the
requirements for the degree of
Mechanical Engineer's Degree

Abstract

Recent advances in stem cell research has demonstrated that the fate of a terminally differentiated cell can be reverted back to pluripotency. The ability to reprogram a differentiated cell back to its undifferentiated, pluripotent state would be a significant breakthrough for regenerative medicine. For example, lost or damaged cells could be replaced by patient-specific reprogrammed cells, thus providing on-demand, compatible, high-quality cells of any required type. However, current protocols for reprogramming rely on simplified models that do not wholly capture system dynamics and on inefficient transcription factor overexpression. We study a gene regulatory network that determines the cell fate in the hematopoietic lineage and demonstrate that a deterministic model cannot capture the experimentally observed system dynamics. We also propose the use of feedback control to address inefficient reprogramming and implement two configurations of the controller on both deterministic and stochastic models of the Oct4-Nanog network. We also address practical issues such as placement of the regulator and consider the effect of inducing or constitutively producing microRNA on the protein steady-state distribution.

Thesis Supervisor: Domitilla Del Vecchio
Title: Professor

Contents

List of Figures

List of Tables

Chapter 1

Introduction

There have been several recent advances in the stem cell field, with demonstrations that the fate of a terminally differentiated cell, contrary to what was traditionally believed, could be reverted back to pluripotency or directly converted to other differentiated cell types. The ability to "reprogram" a differentiated cell back to its undifferentiated, pluripotent state would be a significant breakthrough for regenerative medicine. For example, lost or damaged cells could be replaced by patient-specific reprogrammed cells, thus providing on-demand, compatible, high-quality cells of any required type. Current reprogramming protocols are largely based on prefixed overexpression of suitable transcription factors (TFs), with the rationale that this overexpression could trigger transitions among the states of the gene regulatory networks (GRNs) that take part in cell fate determination. Yet, despite a decade of remarkable progress, the efficiency of these protocols remains low and the quality of produced cells is often unsatisfactory. These issues pose a formidable obstacle to the practical use of pluripotent stem cells.

A better understanding of the GRNs involved in this reprogramming process and tighter control of TFs could potentially alleviate the current bottleneck. My thesis work addresses both these factors: firstly, I studied and modeled one branch of the hematopoietic cell lineage and secondly, I modeled and simulated a feedback controller for better control of TF expression.

Hematopoietic stem cells (HSCs) have the ability to differentiate into any blood

cell and hematopoietic cell fate is thought to be determined by a series of discrete, stable progenitor cells [42, 11, 31]. Here we study the GRN of the first branching step of HSC differentiation to platelets, specifically the PU.1 vs GATA-1 network. Current deterministic models of this GRN do not predict experimentally observed multistablity and we demonstrate that stochasticity is necessary to explain this behavior. A good understanding of this GRN is crucial for developing protocols that could provide an unlimited supply of safer and higher quality platelet transfusion product from human-induced pluripotent stem cells (hiPSCs). Platelet transfusions from donors are currently the most popular method of treating thrombocytopenia and other ailments that require an external platelet source. However, donor-derived platelets have side effects (hemolysis, infection) and logistic problems (short shelf-life). Using hiPSCs to directly produce platelets would address these issues, but a better understanding of the GRN is required to guide an effective protocol.

We also seek to improve current reprogramming protocols that rely on TF overexpression by considering closed-loop feedback overexpression. This study is performed on the Oct4-Nanog GRN, which is a subset of the Yamanaka factors (Oct4, Sox2, Klf4, c-Myc) that have been shown to reprogram fibroblasts to pluripotency. We created deterministic and stochastic models that incorporate regulator dynamics and cell-cell variability. Simulation results demonstrate the efficacy of feedback control compared to open-loop overexpression. The stochastic model was also used to optimize the placement of the regulator gene and the type of expression (constitutive vs inducible) of microRNA. This approach is novel and could replace current prefixed overexpression protocols, leading to a fundamental shift in current cell fate reprogramming strategies.

Chapter 2

Modeling hematopoietic cell fate determination

2.1 Background

All mature, specialized blood cells have been shown to be derived from hematopoietic stem cells (HSCs) through a differentiation process called hematopoiesis [5, 40]. Though there is currently no consensus on the mechanism that guides this process, there is evidence supporting that hematopoiesis occurs in discrete steps where each step is determined by the relative levels of certain transcription factors (TFs) [42, 11, 31]. A stochastic increase in the relative expression level of one TF over another at a cell lineage junction leads the differentiating cell down that corresponding lineage [37, 26, 44]. A diagram of the hematopoietic cell lineage with multipotent progenitor and specialized cells is shown in Fig. 2-1. HSCs have also been shown to self-renew, meaning they not only give rise to all blood cell types, but also have the ability to give rise to HSCs themselves without the need for differentiation, in a process called maintenance. HSC transplants from bone marrows of healthy donors are currently used in treatments of immune system disorders and some cancers because of these two properties.

A good understanding of the gene regulatory network (GRN) that governs hematopoiesis is critical to developing a model that can accurately predict the resulting specialized

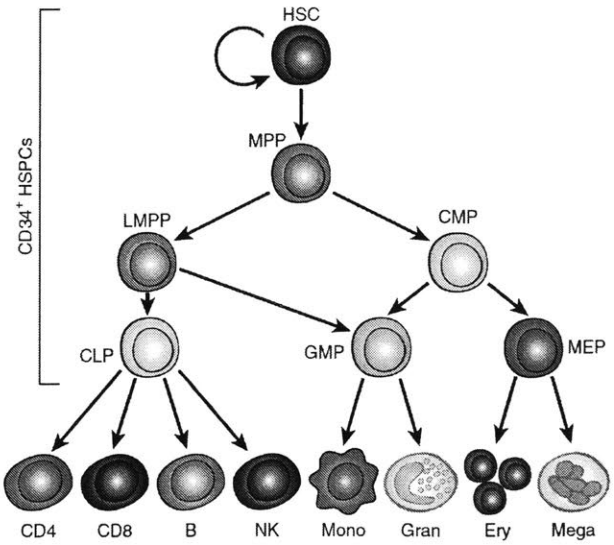

Figure 2-1: HSC lineage tree studied in [9]. In this study, we will discuss the lineage determinant GRN of the Common Myeloid Progenitor (CMP) as it differentiates to either the Megakaryocyte-Erythroid Progenitor (MEP) or the Granulocyte-Macrophage Progenitor (GMP).

cell phenotype from an undifferentiated progenitor cell, subject to overexpression of certain TFs. Here we study the GRN at the Common Myeloid Progenitor (CMP) state, at which point the cell can either differentiate to the Megakaryocyte-Erythroid Progenitor (MEP) or the Granulocyte-Macrophage Progenitor (GMP). Commitment to the MEP or GMP phenotype results in the differentiation to the erythroid or myeloid/lymphoid lineages, respectively [35]. Commitment to the MEP lineage results in the production of megakaryocytes, which shed into platelets.

Platelets are a component of blood that are primarily used to stop bleeding by aggregating and clotting blood vessel injuries and are also responsible for vascular integrity, immunity, and neoangiogenesis (formation of new blood vessels). Platelet transfusions from donors are currently the most popular method of treating thrombocytopenia and other ailments that require an external platelet source. In the US alone, over 2 million platelet doses are transfused every year [1]. However, donor-derived platelets have major potential side effects including hemolysis, infection, sepsis, allergic reactions, and graft-vs-host diseases [43]. Furthermore, platelets must be stored at room temperature; cooling them causes them to lose their clotting ability. How-

ever, room temperature storage increases the risk of bacterial infection leading to a shelf life of just a few days and so donors are continuously being sought to renew supplies that often go to waste. An alternative is to produce platelets directly from human-induced pluripotent stem cells (hiPSCs) derived from universal donor type platelets. The advantage of this method is that hiPSCs, like HSCs, have the ability to self-renew, and so once a colony of hiPSCs is grown, they can, in theory, be allowed to proliferate without a need to replenish their supply. Secondly, since these hiPSCs would be obtained from universal donor type platelets, there would be no issue with infections or graft-vs-host diseases that need to be considered when using current platelet transfusions. Lastly, this could potentially provide an unlimited supply of safer and higher quality platelet transfusion, if there were a protocol to force the hiPSCs to differentiate into platelets (as opposed to other specialized blood cells). Here we study the GRN that governs the differentiation of the HSC or hiPSC at the CMP state into either MEP and GMP.

Literature indicates that the TFs PU.1 and GATA-1 regulate the differentiating HSC into either the myeloid/lymphoid or erythroid lineages, respectively [54, 33]. These 2 TFs have been shown to transcriptionally self-activate and interact antagonistically [54, 33, 53] as pictorially depicted in Fig. 2-2. Arrows and blunt-end arrows represent transcriptional activation and repression, respectively. The expression level of GATA-1 has been shown to increase down the hematopoietic lineage from HSCs to MEPs and the TF is a key regulator of erythroid genes [8, 3]. Similarly, PU.1 expression increases as cells differentiate to the myeloid/lymphoid lineage [35] and has been shown to be critical for myeloid cell regulation. Early progenitor cells around the CMP stage have been shown to have relatively low levels of GATA-1 and PU.1 compared to more differentiated cells [3, 23]. A stochastically-driven increase in the concentration of either PU.1 or GATA-1 may perturb this balance and tip lineage commitment to either myelopoiesis or erythropoiesis [53, 13].

The PU.1-GATA-1 network has been extensively studied since the early 2000s [54, 33, 53]. Previous models of the PU.1-GATA-1 GRN in the literature make assumptions on the biological interaction between the TFs to explain the biologically

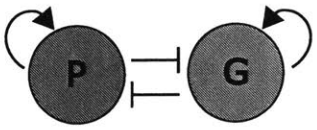

Figure 2-2: PU.1-GATA-1 GRN demonstrating self-activation and mutual repression.

expected bistable property of this GRN, in which one stable state corresponds to high levels of GATA-1 and low levels of PU.1 and vice-versa. However, some of these assumptions have not been experimentally validated. Specifically, the studies in [41] and [22] assume high cooperativity of TFs (n=2 and 4, respectively), though mutual repression and self-activation have been shown to occur primarily in their monomeric form [20, 10, 54]. The model presented in [6] introduces a hypothetical protein X that transcriptionally represses PU.1 and is activated by GATA-1, which has not been experimentally identified. The model studied in [48] assumes both TFs can directly bind to each others' promoter sites and transcriptionally repress each other but this has not been shown. In this paper, we consider a set of biomolecular reactions for the system, in which none of these assumptions are made. Interestingly, we mathematically demonstrate that the corresponding ODE model is monostable, which does not agree with experimental results demonstrating that the network should be capable of exhibiting two phenotypes, high GATA-1, low PU.1 (MEP) and low GATA-1, high PU.1 (GMP).

2.2 Modeling assumptions for the PU.1 vs GATA-1 GRN

The list of modeling assumptions (and evidence) used to derive the deterministic model are:

- PU.1 and GATA-1 transcriptionally self-activate their respective production [36, 34].
- PU.1 represses GATA-1 production by binding to the complex formed by GATA-1 and its promoter. This complex forms repressive chromatin structure effec-

tively silencing transcription (the PU.1-GATA-1 complex has been shown to be present at repressed GATA-1 target genes) [45, 28].

- Similarly, GATA-1 represses production of PU.1 by binding to it on its target genes and prevents the recruitment of co-activators (such as cJun), which are critical for PU.1-mediated transcriptional activation [28, 54].

- PU.1 cannot directly bind to the GATA-1 promoter. [54] reports that PU.1 blocks GATA-1 activation without affecting GATA-1 mRNA, protein expression, or nuclear translocation.

- GATA-1 cannot directly bind to the PU.1 promoter. [7] reports two GATA-1 binding sites on the PU.1 locus, a -18 kb site, which has not been shown to have a functional regulatory role, and a -17 bp site that potentially transcriptionally represses PU.1 production. However, [36] reports that the -14 kb PU.1 URE (to which GATA-1 cannot directly bind) is significantly more critical than the proximal promoter (which includes the -17 bp GATA-1 binding site) in myeloid cell line 416B for PU.1 expression. Therefore, here we only consider the contribution of the PU.1 URE (either activation or repression when bound with GATA-1).

- All TF interaction occurs in their monomeric form. ETS TFs, such as PU.1 typically bind as monomers (both to DNA and other proteins) [20]. Though there is evidence of GATA-1 dimerization [10], it exists primarily in its monomeric form. In particular, GATA-1 self-activation and binding to PU.1 occurs only in its monomeric form [10, 54].

- The promoters of both TFs are leaky [36, 34].

- Protein production occurs in a one-step process (no intermediary mRNA dynamics).

2.3 Modeling primer

Here we demonstrate an example of how to model biochemical reactions using mass action kinetics. Consider the following reaction where two species, A and B interact to form species C at a forward rate a and reverse rate d:

$$\mathrm{A} + \mathrm{B} \underset{\mathrm{d}}{\overset{\mathrm{a}}{\rightleftharpoons}} \mathrm{C} \tag{2.1}$$

The most common model for reaction networks is a deterministic model [17]. It assigns to each species a state variable corresponding to its concentration. The time-evolution of a species' concentration is given by an ordinary differential equation (ODE). Mass action kinetics dictates that the rate of a chemical reaction is proportional to the product of the concentrations of the reactants. Therefore, the dynamics of this system is given by:

$$\dot{A} = dC - aAB, \tag{2.2}$$

$$\dot{B} = dC - aAB, \tag{2.3}$$

$$\dot{C} = aAB - dC. \tag{2.4}$$

Several TFs have been shown to transcriptionally activate or repress the production of other TFs by interacting with the promoter sites of the target TFs. Activation occurs when the complex formed by the TF and the promoter increases the transcription rate of the target TF. Conversely, repression occurs when the previous complex silences transcription. When modeling the concentration of promoters in their on or off states, we typically assume that DNA is conserved (i.e., the total promoter concentration is constant). Consider protein A that transcriptionally activates itself as shown in Fig. 2-3. Let the unbound promoter be represented by P_0 and the protein A and promoter complex given by P_1. By conservation of DNA we find $P_0 + P_1 = P_T$, where P_T is the

total number of promoter sites. The reactions for this system are given below:

$$\mathrm{P_0} + \mathrm{A} \underset{\mathrm{d}}{\overset{\mathrm{a}}{\rightleftharpoons}} \mathrm{P_1}, \tag{2.5}$$

$$\mathrm{P_1} \xrightarrow{\alpha} \mathrm{P_1} + \mathrm{A}, \tag{2.6}$$

$$\mathrm{A} \xrightarrow{\delta} \emptyset. \tag{2.7}$$

Equation 2.5 refers to reversible reaction formed by the unbound promoter P_0 and protein A to form P_1. Equation 2.6 describes the production of protein A. Note that here we model this as one-step protein production, but protein is generally produced from messenger RNA (mRNA) in a process called translation, which is in turn produced from the gene, in a process called transcription. Equation 2.7 refers to dilution of protein A. Since cells grow and their volumes increase, a species with a constant number of molecules decreases in concentration. This is modeled with a dilution term. The ODEs that describe this system are given below:

$$\dot{A} = \alpha P_1 - \delta A + dP_1 + aAP_0, \tag{2.8}$$

$$\dot{P}_0 = dP_1 - aAP_0, \tag{2.9}$$

$$\dot{P}_1 = aAP_0 - dP_1. \tag{2.10}$$

We are often concerned with reducing the dimension of our ODE model. Since protein dynamics are significantly slower than binding reactions [4], we can reduce this system to a single ODE that describes the dynamics of A. Using the conservation of DNA,

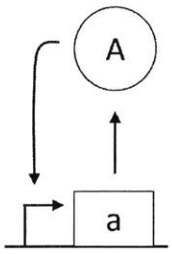

Figure 2-3: Self-activating protein.

we can calculate the steady-state values of P_0 and P_1.

$$P_0 + P_1 = P_T, \tag{2.11}$$

$$P_1 = \frac{dAP_0}{a}, \tag{2.12}$$

$$\frac{dAP_0}{a} + P_0 = P_T, \tag{2.13}$$

$$P_0 = \frac{P_T}{1 + A/K_d}, \tag{2.14}$$

$$\tag{2.15}$$

This leads to the dynamics of A given by:

$$\dot{A} = \alpha P_T \frac{A/K_d}{1 + A/K_d} - \delta A. \tag{2.16}$$

We note the Hill function form of the activation term ($\frac{A/K_d}{1+A/K_d}$). This indicates that for $A/K_d << 1$, we operate in a linear regime where the activation term is roughly $\frac{\alpha P_T A}{K_d}$. If there is a lot of protein A in the system, we find that the Hill function saturates ($\frac{A/K_d}{1+A/K_d} \approx 1$) and the activation term approaches αP_T. In this regime, additional A does not increase its production rate since its promoter is saturated.

2.4 Model reactions

In this section, we use the same modeling principles in Section 2.3 on the biological system described in Section 2.2. The reactions that describe this system are given

below. Here p_0, shown in Fig. 2-4(a), and g_0 are the unbound promoters of TFs PU.1 (P) and GATA-1 (G):

$$p_0 + P \underset{d_0}{\overset{a_0}{\rightleftharpoons}} p_1, \tag{2.17}$$

$$g_0 + G \underset{d_0'}{\overset{a_0'}{\rightleftharpoons}} g_1, \tag{2.18}$$

$$p_1 + G \underset{d_1'}{\overset{a_1'}{\rightleftharpoons}} p_2, \tag{2.19}$$

$$g_1 + P \underset{d_1}{\overset{a_1}{\rightleftharpoons}} g_2, \tag{2.20}$$

$$p_0 \xrightarrow{\alpha_{p0}} p_0 + P, \tag{2.21}$$

$$g_0 \xrightarrow{\alpha_{g0}} g_0 + G, \tag{2.22}$$

$$p_1 \xrightarrow{\alpha_{p1}} p_1 + P, \tag{2.23}$$

$$g_1 \xrightarrow{\alpha_{g1}} g_1 + G, \tag{2.24}$$

$$P \xrightarrow{\delta_P} \emptyset, \tag{2.25}$$

$$G \xrightarrow{\delta_G} \emptyset. \tag{2.26}$$

The reversible binding reactions between the unbound promoters (p_0 and g_0) and their TFs (P and G) to form complexes p_1 (shown in Fig. 2-4(b)) and g_1, respectively are given by (2.17)-(2.18). Reactions (2.19)-(2.20) describe the formation of complexes p_2 and g_2 by reversible binding of G(P) with $p_1(g_1)$, respectively. These complexes represent the "off" state of the promoter wherein the gene is silenced as shown in Fig. 2-4(c) for p_2. Reactions (2.21) and (2.22) describe the leaky promoter one-step production of P and G with rates α_{p0} and α_{g0}, respectively. P(G) is produced at rate $\alpha_{p1}(\alpha_{g1})$ when it is bound to its promoter as shown in reactions (2.23)-(2.24). Lastly, reactions (2.25)-(2.26) describe the decay of both transcription factors. Since the genes are self-activating, we have $\alpha_{p0} < \alpha_{p1}$ and $\alpha_{g0} < \alpha_{g1}$. To simplify the model, we assume that there is no expression from the repressed p_2 and g_2 configurations.

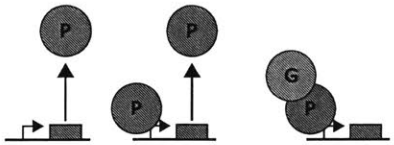

Figure 2-4: Promoter states: (a) p_0 representing leaky production, (b) p_1 with self-activated production, and (c) p_2 when the promoter is fully repressed.

2.5 Deterministic Model

Using mass action kinetics, the ODEs that describe this system are given below:

$$\dot{P} = \alpha_{p1} p_1 - \delta_P P + d_0 p_1 - a_0 p_0 P + d_1 g_2 - a_1 g_1 P, \tag{2.27}$$

$$\dot{G} = \alpha_{g1} g_1 - \delta_G G + d_0^{'} g_1 - a_0^{'} g_0 G + d_1^{'} p_2 - a_1^{'} p_1 G, \tag{2.28}$$

$$\dot{g}_0 = d_0^{'} g_1 - a_0^{'} g_0 G, \tag{2.29}$$

$$\dot{g}_1 = a_0^{'} g_0 G - d_0^{'} g_1 + d_1 g_2 - a_1 g_1 P, \tag{2.30}$$

$$\dot{g}_2 = a_1 g_1 P - d_1 g_2, \tag{2.31}$$

$$\dot{p}_0 = d_0 p_1 - a_0 p_0 P, \tag{2.32}$$

$$\dot{p}_1 = a_0 p_0 P - d_0 p_1 + d_1^{'} p_2 - a_1^{'} p_1 G, \tag{2.33}$$

$$\dot{p}_2 = a_1^{'} p_1 G - d_1^{'} p_2. \tag{2.34}$$

Here we once again use the conservation of DNA given by:

$$p_0 + p_1 + p_2 = p_T, \tag{2.35}$$

$$g_0 + g_1 + g_2 = g_T. \tag{2.36}$$

Once again using the fact that protein dynamics occurs on a slower scale than binding reactions, we obtain the reduced model describing the dynamics of PU.1 and GATA-1.

$$\dot{P} = \frac{\alpha_{p0} p_T + \frac{\alpha_{p1} p_T}{K_d} P}{1 + \frac{P}{K_d} + \frac{PG}{K_d K_{d'}}} - \delta_P P, \tag{2.37}$$

$$\dot{G} = \frac{\alpha_{g0} g_T + \frac{\alpha_{g1} g_T}{K_{d'}} G}{1 + \frac{P}{K_d} + \frac{PG}{K_d K_{d'}} + \frac{G}{K_{d'}}} - \delta_G G, \tag{2.38}$$

The number of steady states of this reduced system is the same as the number of steady states in the original network. Setting the derivatives to zero and performing algebraic manipulations results in a quintic equation. It is not easy to show directly that this equation has a unique solution for all possible choices of kinetic parameters. Thus, in order to determine the number of positive equilibria, we use the advanced deficiency algorithm developed in [14, 16, 24], and implemented in the "Chemical Reaction Network" toolbox [15]. When the algorithm is applied to our network, it shows that it cannot admit multiple positive equilibria for any combination of kinetic parameters. Hence, the deterministic model cannot explain the bistable behaviour observed experimentally [35, 8, 3].

2.6 Conclusions

Deterministic models are usually justified under the assumptions of sufficiently large volume and sufficiently large number of molecules [27], or, under some conditions such as fast promoter kinetics [25]. In such cases, an ODE model captures the system's dynamics, and it produces a similar qualitative behaviour to the one produced by the stochastic model. However, these assumptions are not usually satisfied in practice due to the fact that cell-fate GRNs have usually very low gene copy numbers. Furthermore, in such cells, transcriptional regulation is often mediated by an additional regulation layer dictated by DNA methylation and histone modifications, commonly referred to as chromatin dynamics. For example, the presence of nucleosomes makes binding sites less accessible to TFs and therefore TF-gene binding/unbinding is modu-

lated by the stochastic process of chromatin opening [38, 30, 50]. Several experiments have confirmed the role of the aforementioned complex transcription processes in slow promoter kinetics [39, 49, 29, 52]. Therefore, the qualitative behaviour produced by deterministic models can be erroneous. We are currently preparing a paper that considers the stochastic version of this chemical network and analytically construct the stationary distribution to show that this distribution is indeed capable of admitting a multiplicity of modes.

Chapter 3

Feedback overexpression to reprogram the Oct4-Nanog GRN

3.1 Background

The state of a multistable gene regulatory network (GRN) can be determined by the concentration of the transcription factors (TFs) that make up the network. Waddington's view of cell differentiation [51] consists of cells, represented by balls, that roll down a landscape of bifurcating valleys as shown in Fig. 3-1. Each valley represents a possible cell fate and the ridges between the valleys maintain the cell fate once it has been chosen. A cell in its pluripotent state occupies the top of this landscape. Normal development consists of a pluripotent or multipotent progenitor cell differentiating to a committed cell state naturally over time. Reprogramming consists of forcing a differentiated cell up the landscape to its pluripotent state. Transdifferentiation consists of directly reprogramming a cell to a different state without intermediary reprogramming to the pluripotent state.

According to Waddington, each stable state of a GRN can be represented by a particular cell phenotype and the transitions between the phenotypes are triggered by external stimuli or noise. Our ability to trigger these transitions is dependent upon how well we can artificially stimulate the GRN [21]. The current method of triggering changes in cell phenotypes is by overepxressing certain TFs above their endogenous

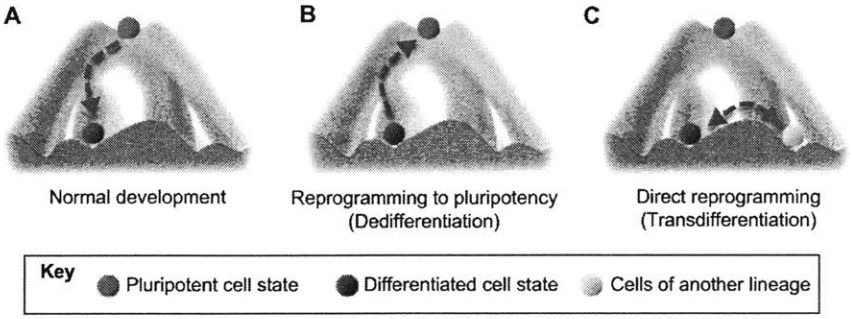

Figure 3-1: Waddington's view of cell differentiation [46]. A: Normal development consists of a pluripotent or multipotent progenitor cell differentiating to a differentiated cell state naturally. B: Reprogramming consists of forcing a differentiated cell up the landscape to its pluripotent state. C: Transdifferentiation consists of directly reprogramming a cell to a different state without intermediary reprogramming to the pluripotent state.

level [18]. Here "endogenous" refers to the cell's natural GRN without any artificial modifications. However, the success rate of protocols that rely on overexpression remains low and once an experiment begins, the TF levels cannot be adjusted [32]. [12] suggests that the low reprogramming efficiency from TF overexpression is because there is no general guarantee that a GRN's dynamics will allow transitions to a desired target state under the imposed stimuli. Here we propose a method for artificially enabling transitions between stable states that does not depend on the natural network's dynamics and would allow for more efficient reprogramming.

The goal of this project is to reprogram fibroblasts to efficiently generate high quality human-induced pluripotent stem cells (hiPSCs) by using feedback overexpression of certain TFs, in particular Oct4, Sox2, Klf4, and c-Myc (OSKM factors). These 4 OSKM TFs have been shown to induce reprogramming to iPSCs from fibroblasts [47]. Here we study a model of the Oct4-Nanog GRN and develop a feedback controller to arbitrarily place the steady-state of Oct4 and Nanog. The purpose of using models is to address questions regarding the optimization of virus construction and to study how various parameters affect the distribution.

3.2 Endogenous system model

We began this analysis using a tristable network incorporating the Nanog-Oct4 GRN, developed in [12] and pictorially represented in Fig. 3-2. Arrows represent transcriptional activation and blunt-end arrows represent transcriptional repression. The exogenous inputs u_1 and u_2 in Fig. 3-2 are the synthetic signals that directly increase the concentration of Nanog and Oct4, respectively. The list of reactions for this GRN is given in [12] and the coupled ODEs that describe the dynamics endogenous network with the exogenous inputs are given below (O=Oct4, N=Nanog):

$$\dot{O} = \frac{\kappa_O}{\delta_O}\left[\frac{\alpha_{O0} + \alpha_{O1}(\frac{O}{k_{O1}})^2 + \alpha_{O2}(\frac{N}{k_{O2}})^2 + \alpha_{O3}(\frac{NO}{k_{O1}k_{O2}})^2}{1 + (\frac{O}{k_{O1}})^2 + (\frac{N}{k_{O2}})^2 + (\frac{NO}{k_{O1}k_{O2}})^2}\right] + u_2 - \gamma_O O \tag{3.1}$$

$$= \frac{\kappa_O}{\delta_O} H_O + u_2 - \gamma_O O, \tag{3.2}$$

$$\dot{N} = \frac{\kappa_N}{\delta_N}\left[\frac{\alpha_{N0} + \alpha_{N1}(\frac{N}{k_{N1}})^2 + \alpha_{N2}(\frac{O}{k_{N2}})^2 + \alpha_{N3}(\frac{NO}{k_{N1}k_{N2}})^2}{1 + (\frac{N}{k_{N1}})^2 + (\frac{O}{k_{N2}})^2 + C_{N1}(\frac{O}{k_{N2}})^4 + (\frac{NO}{k_{N1}k_{N2}})^2}\right] + u_1 - \gamma_N N, \tag{3.3}$$

$$= \frac{\kappa_N}{\delta_N} H_N + u_1 - \gamma_N N, \tag{3.4}$$

where the Hill function terms for Oct4 and Nanog are represented by H_O and H_N, respectively. The above system is tristable and the three stable equilibrium points corresponds to the trophectoderm (TR), the desired pluripotent state (PL), and primitive endoderm (PE). Fig. 3-3 plots the transfer curves for Nanog and Oct4, which show the stable steady states for each level of inducer concentration. Since the addition of inducer to the system changes the GRN's dynamics, the steady states themselves change. Both the Nanog and Oct4 transfer curves demonstrate that the PE state is accessible for all inducer values and that the TR and PL states are only accessible for sufficiently low inducer concentration. Furthermore, note that the PE state represents the high steady state for Oct4 but is intermediate in terms of Nanog expression (there is no requirement for a steady-state to have the same relative expression levels of all TFs in the GRN). Also note the disparity in the change of inducer for the system to have multiple steady states to switch to a single steady state. For u_1 (Nanog overex-

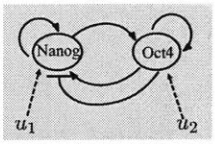

Figure 3-2: Nanog-Oct4 GRN showing transcriptional interactions between the TFs [12]. u_1 and u_2 are the inducers for Nanog and Oct4, respectively.

pression), the TR state is lost for low inducer values and the PL state is attainable for a wide range of u_1. However, the behavior of both TR and PL states are quite similar for u_2.

Fig. 3-4 plots the steady-state that a trajectory initialized at 0 converges to for a given inducer concentration. The single circle around 40 nM for Oct4 represents the PL state demonstrating that it would be extremely difficult to guide the system to pluripotency if one were to use u_2 to control the system.

3.3 Controller dynamics

Here we propose the use of feedback control to arbitrarily place the steady-state of TFs, so that upon removal, the system converges to the desired steady-state. The reason we choose to not use direct overexpression is because we have shown that overexpressing Oct4 or Nanog may lead the system to overshoot and converge to the undesired PE state. Implementation of this feedback control requires the use of lentiviral constructs that are inserted into cells. These constructs are pieces of DNA that use cell machinery to produce proteins, which can act as input signals (activation or repression) for other processes in the host cell. Specifically, feedback overexpression involves using microRNA to degrade the mRNA of the TF we wish to control (high gain negative feedback) and overexpression of the same TF.

For simplicity, we will consider a bistable system of a protein X, that transcriptionally activates itself. We would like to control the concentration of X from its low state by feedback overexpression so that it is within the stability region of its high state and so once the controller is removed the endogenous system converges to the desired high state. Though for this toy system, overexpressing X sufficiently

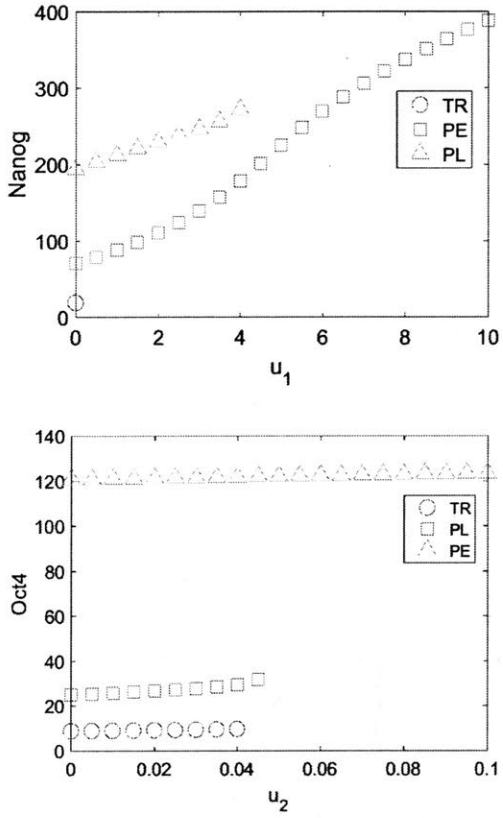

Figure 3-3: Nanog (**top**) and Oct4 (**bottom**) transfer curves, representing the stable steady states accessible for each inducer level (u_1 or u_2).

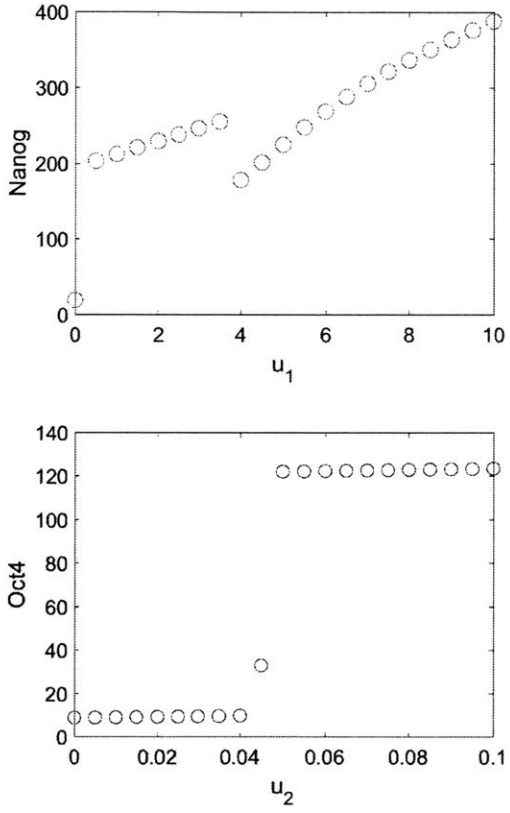

Figure 3-4: Nanog (**top**) and Oct4 (**bottom**) transfer curves, representing the stable steady state that the system converges to when initialized at 0.

will cause it to converge to its high state we use this model as an example on how to implement similar control for the Oc4-Nanog GRN.

The endogenous system is bistable and is initialized with its concentration in the stability region of the low state. The controller, which consists of the microRNA gene and synthetic x_s gene with copy number n_0 is introduced making the system monostable. We aim to arbitrarily place the concentration of the total TF in the stability region of the high state by only controlling the production rate of the synthetic mRNA. Upon removal of the controller (setting the synthetic TF to 0), we hope to show that the steady-state value of the endogenous TF concentration approaches that of the high state.

The reactions describing this system are given below. Here x_e is the endogenous gene that produces mRNA, m_e at a rate α_0 as shown in Equation 3.5. Equations 3.6 and 3.9 represent the production of X from m_e and m_s, respectively at rate κ. The TF dimerizes (Equation 3.7) and binds to x_e in Equation 3.10. Equation 3.8 describes the production of synthetic mRNA, m_s. Here since we have assumed that it is constitutively produced, we represent its production by the copy number of its DNA, which is n_0. Its rate of production is α, which is the input to the system. Here we assume that the promoter is inducible by an external inducer and that α is proportional to the inducer concentration. The microRNA, s is constitutively produced from a different vector with copy number n_1 (Equation 3.14) and enzymatically degrades both synthetic and endogenous mRNA, m_s and m_e respectively, after reversible binding in Equations 3.12 and 3.11. The dilution reactions are given in Equations 3.15-3.19.

$$x_e \xrightarrow{\alpha_0} x_e + m_e, \tag{3.5}$$

$$m_e \xrightarrow{\kappa} m_e + X, \tag{3.6}$$

$$X + X \underset{d}{\overset{a}{\rightleftharpoons}} X_2, \tag{3.7}$$

$$n_0 \xrightarrow{\alpha} n_0 + m_s, \tag{3.8}$$

$$m_s \xrightarrow{\kappa} m_s + X, \tag{3.9}$$

$$X_2 + x_e \underset{d^*}{\overset{a^*}{\rightleftharpoons}} C_1, \tag{3.10}$$

$$m_e + s \underset{d'}{\overset{a'}{\rightleftharpoons}} C_3 \xrightarrow{\delta_C} s, \tag{3.11}$$

$$m_s + s \underset{d'}{\overset{a'}{\rightleftharpoons}} C_4 \xrightarrow{\delta_C} s, \tag{3.12}$$

$$C_1 \xrightarrow{\alpha_1} C_1 + m_e, \tag{3.13}$$

$$n_1 \xrightarrow{\alpha_2} n_1 + s, \tag{3.14}$$

$$m_e \xrightarrow{\delta_m} \emptyset, \tag{3.15}$$

$$m_s \xrightarrow{\delta_m} \emptyset, \tag{3.16}$$

$$s \xrightarrow{\delta_s} \emptyset, \tag{3.17}$$

$$X \xrightarrow{\delta_X} \emptyset, \tag{3.18}$$

$$X_2 \xrightarrow{\delta_X} \emptyset. \tag{3.19}$$

The ODEs that describe the endogenous and controller system are given below:

$$\dot{x}_e = \mathrm{d}^* \, \mathrm{C}_1 - \mathrm{a}^* \, \mathrm{x}_e \, \mathrm{X}_2, \tag{3.20}$$

$$\dot{m}_e = \alpha_0 \, \mathrm{x}_e + \alpha_1 \, \mathrm{C}_1 - \delta_m \, \mathrm{m}_e + \mathrm{d}' \, \mathrm{C}_3 - \mathrm{a}' \, \mathrm{m}_e \, \mathrm{s}, \tag{3.21}$$

$$\dot{\mathrm{X}} = \kappa(\mathrm{m}_e + \mathrm{m}_s) - \delta_{\mathrm{X}} \, \mathrm{X} + 2 \, d \, \mathrm{X}_2 - 2 \, a \, \mathrm{X}^2, \tag{3.22}$$

$$\dot{X}_2 = a \, \mathrm{X}^2 - d \, \mathrm{X}_2 + \mathrm{d}^* \, \mathrm{C}_1 - \mathrm{a}^* \, \mathrm{x}_e \, \mathrm{X}_2 - \delta_{\mathrm{X}} \, \mathrm{X}_2, \tag{3.23}$$

$$\dot{\mathrm{m}}_\mathrm{s} = \alpha \mathrm{n}_0 - \delta_m \, \mathrm{m}_s + \mathrm{d}' \, \mathrm{C}_4 - \mathrm{a}' \, \mathrm{m}_s \, \mathrm{s}, \tag{3.24}$$

$$\dot{\mathrm{C}}_1 = \mathrm{a}^* \, \mathrm{x}_e \, \mathrm{X}_{e2} - \mathrm{d}^* \, \mathrm{C}_1, \tag{3.25}$$

$$\dot{\mathrm{C}}_3 = \mathrm{a}' \, \mathrm{m}_e \, \mathrm{s} - \mathrm{d}' \, \mathrm{C}_3 - \delta_{\mathrm{C}} \, \mathrm{C}_3, \tag{3.26}$$

$$\dot{\mathrm{C}}_4 = \mathrm{a}' \, \mathrm{m}_s \, \mathrm{s} - \mathrm{d}' \, \mathrm{C}_4 - \delta_{\mathrm{C}} \, \mathrm{C}_4, \tag{3.27}$$

$$\dot{\mathrm{s}} = \alpha_2 \, \mathrm{n}_1 - \delta_s \, \mathrm{s} + \mathrm{d}' \, \mathrm{C}_3 - \mathrm{a}' \, \mathrm{m}_e \, \mathrm{s} + \delta_{\mathrm{C}} \, \mathrm{C}_3 + \mathrm{d}' \, \mathrm{C}_4 - \mathrm{a}' \, \mathrm{m}_s \, \mathrm{s} + \delta_{\mathrm{C}} \, \mathrm{C}_4. \tag{3.28}$$

Using conservation of DNA, $x_e + C_1 = x_T$ we find the equilibrium value of the species ($d \gg \delta_X$):

$$s = \frac{\alpha_2 n_1}{\delta_s}, \tag{3.29}$$

$$C_4 = \frac{a' m_s s}{d + \delta_C}, \tag{3.30}$$

$$m_s = \frac{\alpha n_0}{\delta_m \gamma_1}, \tag{3.31}$$

$$x_e = \frac{x_T}{1 + \frac{X^2}{K_{d1} K_d^*}}, \tag{3.32}$$

$$C_1 = \frac{x_e X^2}{K_{d1} K_d^*}, \tag{3.33}$$

$$X_2 = \frac{X^2}{K_{d1}}, \tag{3.34}$$

$$m_e = (\frac{x_T}{\gamma_2}) \frac{\alpha_0 + \frac{\alpha_1}{K_{d1} K_d^*} (X^2)}{1 + \frac{X^2}{K_{d1} K_d^*}}, \tag{3.35}$$

where $\gamma_1 = \delta_m + a' s_{|eq} (1 - \frac{d'}{d + \delta_C})$, $\gamma_2 = \delta_m + a' s_{|eq} (1 - \frac{d'}{d' + \delta_C})$. This leads to the

following reduced system:

$$\dot{X} = \frac{\kappa\alpha n_0}{\gamma_1} + (\frac{x_T\kappa}{\gamma_2})\frac{\alpha_0 + \frac{\alpha_1 X^2}{K_{d1}K_d^*}}{1 + \frac{X^2}{K_{d1}K_d^*}} - \delta_X X \tag{3.36}$$

The first term in $\dot{X}$ is from the production of X by the controller (synthetic X) and the second term is Hill function representing the transcriptional endogenous activation by X. We can rewrite this as:

$$\dot{X} = \frac{K_1\alpha}{\gamma_1} + \frac{K_2}{\gamma_2}H(X) - \delta_X X \tag{3.37}$$

with constants $K_1 \sim n_0, K_2$. Increasing the amount of microRNA and the synthetic promoter strength corresponds to increasing n_1 and α, respectively because $\gamma_1, \gamma_2 = C_1 + C_2 n_1$ for positive constants C_1, C_2. Therefore, for large n_1, α we have:

$$\dot{X} \approx \frac{K_1\alpha}{\gamma_1} - \delta_X X \tag{3.38}$$

and so $X_{|eq} \approx \frac{K_1\alpha}{\delta_X\gamma_1}$. In order to drive X to a desired equilibrium point, we require α and n_1 sufficiently large and that α/γ_1 becomes a non-zero constant. As a sanity check we can see if $n_1 = 0$, $\gamma_1, \gamma_2 = \delta_m$ and we recover the endogenous system with constitutive synthetic production. Now, if we remove the microRNA degradation of m_s (remove Equation 3.12) so that the controller only degrades m_e, the X dynamics are given by:

$$\dot{X} = \frac{\kappa\alpha n_0}{\delta_m} + (\frac{x_T\kappa}{\gamma_2})\frac{\alpha_0 + \frac{\alpha_1 X^2}{K_{d1}K_d^*}}{1 + \frac{X^2}{K_1 K_1^*}} - \delta_X X \tag{3.39}$$

Rewriting this as:

$$\dot{X} = \frac{K_1\alpha}{\delta_m} + \frac{K_2}{\gamma_2}H(X) - \delta_X X \tag{3.40}$$

Here just increasing n_1 is sufficient to ensure that:

$$\dot{X} \approx \frac{K_1\alpha}{\delta_m} - \delta_X X \tag{3.41}$$

and so $X_{|eq} \approx \frac{K_1\alpha}{\delta_X \delta_m}$, which is unique and controllable from α. Note here there is no requirement for increasing α with γ_1, γ_2. Now, if we remove microRNA degradation of m_e (remove equation 3.11), we would expect to not be able to control $X_{|eq}$. The X dynamics are of the form:

$$\dot{X} = \frac{K_1\alpha}{\gamma_1} + \frac{K_2}{\delta_m}H(X) - \delta_X X \tag{3.42}$$

This matches our intuition since here we shut down the synthetic production of X by increasing both α and n_1.

Applying the same controller reactions (for endogenous mRNA degradation only) to the endogenous Nanog-Oct4 network results in the following dynamics :

$$\dot{O} = \frac{\kappa_O}{G_O}H_O + K_O u_2 - \gamma_O O, \tag{3.43}$$

$$\dot{N} = \frac{\kappa_N}{G_N}H_N + K_N u_1 - \gamma_N N. \tag{3.44}$$

The parameters, G_O and G_N are increasing functions of microRNA steady state and so for sufficiently large amount of microRNA, the endogenous network is shut down:

$$\dot{O} \approx K_O u_2 - \gamma_O O, \tag{3.45}$$

$$\dot{N} \approx K_N u_1 - \gamma_N N. \tag{3.46}$$

The steady-state of this system is linear with respect to the inputs u_1, u_2. For the system with microRNA degradation of both endogenous and synthetic mRNA, the protein dynamics are of the form:

$$\dot{O} = \frac{\kappa_O}{G_O}H_O + \frac{K_O u_2}{G_O} - \gamma_O O, \tag{3.47}$$

$$\dot{N} = \frac{\kappa_N}{G_N}H_N + \frac{K_N u_1}{G_N} - \gamma_N N. \tag{3.48}$$

Here we note that for large microRNA degradation, the inputs u_1, u_2 also need to be relatively large to have any appreciable change on the Oct4 or Nanog steady-state.

From this analysis, it appears that the controller with endogenous mRNA degradation has the benefit of not only being able to produce linear input-output system response but also requiring significantly less input than the system with degradation to both synthetic and endogenous mRNA. However, in reality, the vector copy number varies from cell to cell (i.e., not every cell will receive n_0 copies of the synthetic gene) and so we need to address the robustness of this controller to this form of stochasticity.

3.4 Multiplicity of infection

The multiplicity of infection (MOI) is the ratio of the total number of infectious agents (e.g., lentiviruses with synthetic Oct4 and microRNA gene) to infection targets (e.g., cells). The protocol to infect a colony of cells with a desired MOI is given below (adopted from [2]):

- Seed a fixed number of cells into each well (e.g. 75,000 cells in each well of a 6-well dish) by:
    1. Diluting cells in media (e.g. dilute 525,000 cells into 14 mL of DME media)
    2. Mixing well by pipetting/inverting
    3. Adding 2 mL of cell suspension to each well (e.g. (2mL/14mL) $\times$ 525,000 cells = 75,000 cells)
- Incubate the cells overnight.
- If using freshly collected virus, use a filter to remove cells/debris. If using frozen virus, rapidly thaw in warm water bath. At this point the concentration of the virus stock is unknown.
- Prepare dilutions of virus (1:10, 1:100, 1:1000, etc.) in same DME media.
- Add the viral dilution to each well of cells that were grown overnight and incubate for 48-72 hrs.

- Calculate the percent of fluorescent-positive cells in each well (MOI). Only consider wells with <40% fluorescent-positive cells. Here we assume 1 integration event per cell because for >40% fluorescent-positive cells, we risk counting cells with multiple integration events per cell.

- Calculate the titer (transduction units/mL) of the original virus stock using the dilution factors (method 1) or virus volume (method 2).

    1. Method 1: TU/mL = (Number of cells transduced x Percent fluorescent x Dilution Factor)/(Transduction Volume in mL). Method 1 example: If the 1:100 well has 25% fluorescent cells and 150,000 cells were originally transduced, then there are $(150,000 \times 0.25 \times 100)/(1.5 \text{ mL}) = 2.5 \times 10^6$ TU/mL

    2. Method 2: TU/mL = (Number of cells transduced $\times$ Percent fluorescent)/(Virus volume in mL). Method 2 example: If 15 μL of virus is added to 150,000 cells resulted in 25% fluorescent cells, then there are $(150,000 \times 0.25)/(0.015 \text{ mL}) = 2.5 \times 10^6$ TU/mL

    3. For an accurate titer, take the average of multiple dilutions.

- Now that the titer is known, to get a desired MOI for a certain number of cells, pick the appropriate volume of the virus stock (e.g. say titer was calculated to be 10^5 TU/mL and we want MOI=10 for 100 cells, we need 10^5 TU so take 1 mL of virus stock and add it to plate of cells (volume of virus stock = MOI $\times$ cells/titer)).

The variation in vector copy number is modeled as a Poisson distribution with the multiplicity of infection (MOI) as the Poisson parameter. The intuitive reason for this is because if we had 100 balls (infectious particles) that are randomly thrown in 10 boxes (cells), the probability that a given box has n balls is Poisson with parameter 10.

3.5 Modeling cell-cell variability using Poisson parameter

A numerical study was performed with the vector copy number having a Poisson distribution. Initially, two configurations of the controller (for both Oct4 and Nanog) were studied: degradation of endogenous mRNA only and degradation of both endogenous and synthetic mRNA. For simplicity, we assume that the microRNA and TF genes are on the same vector with copy number n and we adjust the promoter strength of the microRNA to change the microRNA steady-state. We found that for sufficiently strong mRNA degradation, there is approximately a linear input-output relation between the inducer concentration and the mean of the protein distribution for both controllers as shown in the deterministic model. However, in the system with just endogenous mRNA degradation, the distribution widens as inducer concentration increases. This occurs because the microRNA shuts down the endogenous network but there is no control of the variation in vector copy number of the synthetic gene as shown by the following dynamics:

$$\dot{O} = \frac{\kappa_O}{G_O} H_O + \frac{K_O u_2}{\delta_m} - \gamma_O O, \tag{3.49}$$

$$\dot{N} = \frac{\kappa_N}{G_N} H_N + \frac{K_N u_1}{\delta_m} - \gamma_N N. \tag{3.50}$$

Once again G_O, G_N are increasing functions of the microRNA steady-state (for Oct4 and Nanog, respectively) and so the Oct4 steady-state is of the form:

$$O = \frac{K_O u_2}{\delta_m \delta_O}, \tag{3.51}$$

$$= C n u_2, \tag{3.52}$$

for some positive constant C. Here we find that the Oct4 steady-state (and Nanog) is proportional to the Poisson random variable n. In particular, there is no way to control for this variation from the input u_2.

In the case with both endogenous and synthetic mRNA degradation, this variation

in copy number is accounted for leading to a tighter distribution. This is reflected in the system dynamics of the form:

$$\dot{O} = \frac{\kappa_O}{G_O} H_O + \frac{K_O u_2}{G_O} - \gamma_O O, \tag{3.53}$$

$$\dot{N} = \frac{\kappa_N}{G_N} H_N + \frac{K_N u_1}{G_N} - \gamma_N N. \tag{3.54}$$

For large G_O, G_N, the Oct4 steady-state is given by:

$$O = \frac{K_O u_2}{G_O \gamma_O}, \tag{3.55}$$

$$\approx \frac{n}{1+n} \frac{u_2}{\alpha_2}, \tag{3.56}$$

where α_2 is the microRNA production rate and is assumed to be large. Therefore, the cost of degrading the synthetic mRNA is that the inducer concentration (u_1, u_2) needs to be significantly larger to produce any appreciable change in the Oct4 steady state value. Here the parameters, G_O and G_N are of the form $C_1 + C_2 n$ ($C_1, C_2 > 0$), and K_O and K_N are of the form $C_3 n$ ($C_3 > 0$).

The Oct4 steady state concentration for a cell population (N=500) infected with a lentivirus (MOI=10) containing both the synthetic Oct4 gene and the microRNA gene as a function of inducer concentration (u_2) is shown in Fig. 3-5. The left plot is the open-loop case (no microRNA), the middle plot is for endogenous mRNA degradation only, and the right plot is for both endogenous and synthetic mRNA degradation. For no inducer, the distribution of all 3 plots indicate the cells are in the initial TR state. For the open-loop case, increasing the inducer slightly results in a small proportion of cells converging to the desired PL state (parameter values were chosen so that this population was roughly 1% of cells). Increasing the inducer concentration resulted in the cell population converging to the undesired PE state. For the system with endogenous mRNA degradation, there is a spreading out effect in cell population. This is because the Oct4 steady-state is a scaled version of a Poisson distribution, which only has probabilities associated for non-negative integers. The distribution of the model with degradation to both endogenous and synthetic mRNA retains the

input-output linear response of the mean but is significantly tighter.

3.6 Regulator dynamics

Until this point of analysis, the inducer was modeled to directly increase the Oct4 steady-state concentration without limit. This assumption was then relaxed to accommodate a more realistic model of gene activation. A regulator gene is also inserted in the lentivirus that produces a protein TetR, which dimerizes in the presence of the inducer (the true input to the system). The complex of 2 TetR molecules and 1 Dox molecule can bind to the promoter of and activate the synthetic Oct4 gene. Fig. 3-6 demonstrates the two experimental choices for where to place the regulator: on the same or different vector as the Oct4 gene. In the different vector case, the copy number of the regulator vector is n_0 and that of the Oct4 gene is n_1. This was done to answer if it was better to place the regulator and gene on the same (1V) or different vectors (2V).

3.6.1 Regulator model

Here we consider a toy model incorporating the regulator and inducer that produces synthetic Oct4 (i.e., no endogenous GRN). TetR (T) can dimerize only when bound to the inducer, Dox (D) and the resulting complex binds to the synthetic Oct4 promoter (P_0). The activated promoter (P_1) produces mRNA (m). For simplicity we also assume no inducer dynamics and so binding to the promoter does not affect its concentration: this is reasonable if D is large and we are operating in a saturating inducer region (explained later). The reactions for the same vector configuration are

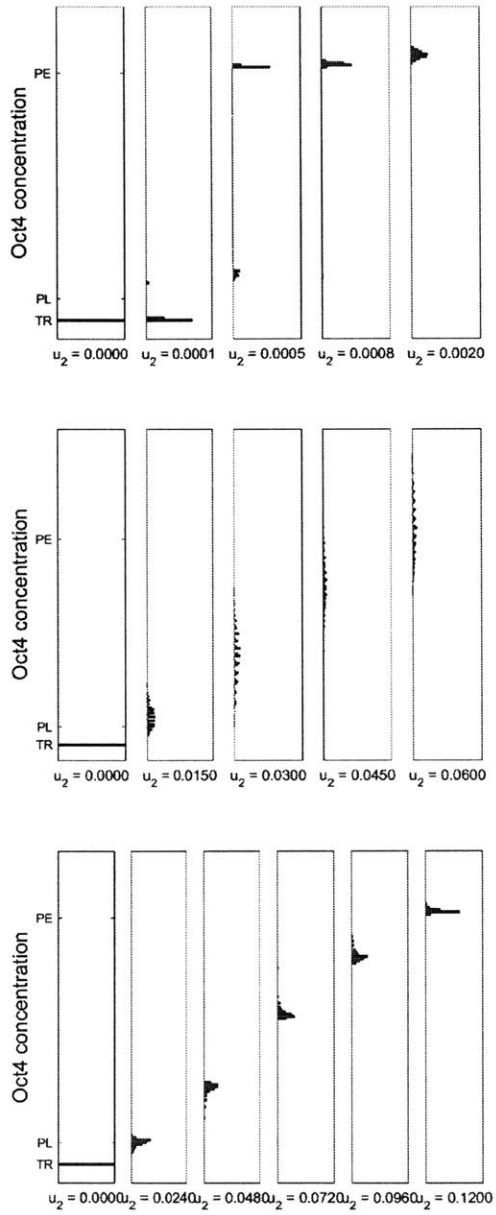

Figure 3-5: Distribution of steady-state Oct4 concentration in cells as a function of Oct4 inducer overexpression for **top:** no microRNA, **middle:** endogenous Nanog and Oct4 mRNA degradation **bottom:** endogenous and synthetic Nanog and Oct4 mRNA degradation. The steady-state vales of the tristable endogenous systems are also shown (TR = trophectoderm, PL = pluripotent, PE = primitive endoderm). N=500, MOI=10.

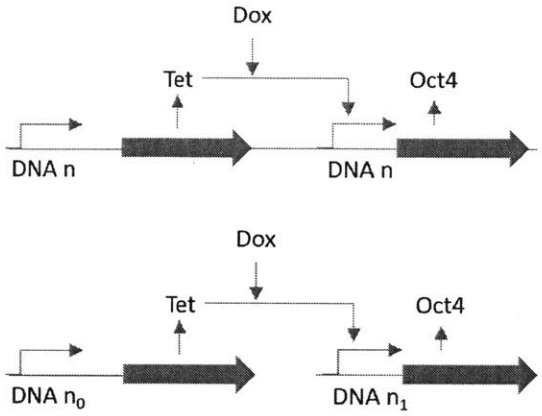

Figure 3-6: Schematic of regulator and inducer used to control expression of synthetic Oct4. Regulator and Oct4 on same (**top**) and different (**bottom**) vector.

given below:

$$\mathrm{n} \xrightarrow{\alpha_T} \mathrm{n} + \mathrm{T}, \tag{3.57}$$

$$\mathrm{T} + \mathrm{D} \underset{\mathrm{d}}{\overset{\mathrm{a}}{\rightleftharpoons}} \mathrm{C}_1, \tag{3.58}$$

$$\mathrm{T} + \mathrm{C}_1 \underset{\mathrm{d}}{\overset{\mathrm{a}}{\rightleftharpoons}} \mathrm{C}_2, \tag{3.59}$$

$$\mathrm{C}_2 + \mathrm{P}_0 \underset{\mathrm{d}}{\overset{\mathrm{a}}{\rightleftharpoons}} \mathrm{P}_1, \tag{3.60}$$

$$\mathrm{P}_1 \xrightarrow{\alpha_m} \mathrm{P}_1 + \mathrm{m}, \tag{3.61}$$

$$\mathrm{m} \xrightarrow{\alpha_O} \mathrm{m} + \mathrm{O}, \tag{3.62}$$

$$\mathrm{O} \xrightarrow{\delta_O} \emptyset, \tag{3.63}$$

$$\mathrm{m} \xrightarrow{\delta_m} \emptyset, \tag{3.64}$$

$$\mathrm{T} \xrightarrow{\delta_T} \emptyset, \tag{3.65}$$

$$\mathrm{C}_1 \xrightarrow{\delta_C} \emptyset, \tag{3.66}$$

$$\mathrm{C}_2 \xrightarrow{\delta_C} \emptyset. \tag{3.67}$$

The ODEs that describe this system include:

$$\dot{\mathrm{O}} = \alpha_{\mathrm{O}}\,\mathrm{m} - \delta_{\mathrm{O}}\,\mathrm{O}, \tag{3.68}$$

$$\dot{\mathrm{m}} = \alpha_m\,\mathrm{P}_1 - \delta_m\,\mathrm{m}, \tag{3.69}$$

$$\dot{P}_1 = a\,\mathrm{C}_2\,\mathrm{P}_0 - d\,\mathrm{P}_1, \tag{3.70}$$

$$\dot{\mathrm{T}} = \alpha_{\mathrm{T}}\,\mathrm{n} - \delta_{\mathrm{T}}\,\mathrm{T} + d\,\mathrm{C}_1 - a\,\mathrm{T}\,\mathrm{D} + d\,\mathrm{C}_2 - a\,\mathrm{T}\,\mathrm{C}_1, \tag{3.71}$$

$$\dot{C}_1 = a\,\mathrm{T}\,\mathrm{D} - d\,\mathrm{C}_1 + d\,\mathrm{C}_2 - a\,\mathrm{T}\,\mathrm{C}_1 - \delta_{\mathrm{C}}\,\mathrm{C}_1, \tag{3.72}$$

$$\dot{C}_2 = a\,\mathrm{T}\,\mathrm{C}_1 - d\,\mathrm{C}_2 + d\,\mathrm{P}_1 - a\,\mathrm{C}_2\,\mathrm{P}_0 - \delta_{\mathrm{C}}\,\mathrm{C}_2, \tag{3.73}$$

$$\dot{\mathrm{D}} = d\,\mathrm{C}_1 - a\,\mathrm{T}\,\mathrm{D} - \delta_{\mathrm{D}}\,\mathrm{D}. \tag{3.74}$$

Using conservation of DNA ($P_0 + P_1 = n$) we can show ($K_d = \frac{d}{a}$):

$$P_0 = \frac{n}{1 + C_2/K_d}, \tag{3.75}$$

$$P_1 = n\frac{C_2/K_d}{1 + C_2/K_d}. \tag{3.76}$$

For binding significantly faster than dilution ($d >> \delta$), we can show the equilibria of C_2 and C_1 as:

$$C_1 = \frac{TD}{K_d}, \tag{3.77}$$

$$C_2 = \frac{TC_1}{K_d} = \frac{T^2D}{K_d^2}. \tag{3.78}$$

From $T_{eq} = \frac{\alpha_T n}{\delta_T}$ we find:

$$O_{eq} = (\frac{\alpha_O \alpha_m}{\delta_O \delta_m})n\frac{\frac{\alpha_T^2 n^2 D}{\delta_T^2 K_d^2}}{1 + \frac{\alpha_T^2 n^2 D}{\delta_T^2 K_d^2}}, \tag{3.79}$$

$$\approx n\frac{n^2 D}{1 + n^2 D} \tag{3.80}$$

For the case of different vectors the form of the equilibrium O is given by:

$$O_{eq} \approx n_1(\frac{n_0^2 D}{1 + n_0^2 D}) \tag{3.81}$$

Here we note a difference from our previously assumed activation term given by:

$$\dot{O} = K_O n D - \gamma_O O. \tag{3.82}$$

In particular, activation is now a Hill function term and so for low inducer levels ($n_0^2 D << 1$) it is linear, but after the promoter is saturated ($n_0^2 D >> 1$), increasing the inducer concentration does not change the Oct4 steady-state. To summarize, the dynamics for the two configurations in consideration are given (for the 1V case):

$$\dot{O} = K_O n \frac{D n^2}{1 + D n^2} - \gamma_O O, \tag{3.83}$$

where n is the copy number of the vector. The 2V dynamics is given by:

$$\dot{O} = K_O n_1 \frac{D n_0^2}{1 + D n_0^2} - \gamma_O O, \tag{3.84}$$

where n_0 and n_1 are the copy numbers of the regulator and Oct4 gene vectors, respectively. A configuration was deemed better if it had a smaller coefficient of variation (CV) for a given mean and is given by:

$$CV(X) = \frac{\sqrt{E[X^2] - E[X]^2}}{E[X]} \tag{3.85}$$

for a random variable X (in our case, it would be the steady-state value of Oct4). We simplified the analysis by assuming that the microRNA had completely shut down the endogenous network so we are only concerned with the steady-state distribution of the synthetically produced Oct4. The CV for the 1V model is calculated by assigning each probability of a Poisson distribution to the corresponding Oct4 steady-state. For example, let the Oct4 steady-state be a function, f, given by (for the 1V and 2V

cases, respectively):

$$f(n, u_2) = n\frac{n^2 D}{1 + n^2 D}, \tag{3.86}$$

$$f(n, u_2) = n_1\frac{n_0^2 D}{1 + n_0^2 D}. \tag{3.87}$$

Let the MOI=λ. The probability mass function (pmf) of the random variable n is Poisson and given by:

$$P(n = n^*) = \frac{\lambda^n e^{-\lambda}}{n^*!} \tag{3.88}$$

The pmf of $f(n)$ is given by $P(n)$ since the transformation from $n = 0, 1, ...$ to $f(n)$ is one-to-one ($f(n)$ is monotone increasing in n).

The CV for the 2V model involves making a grid of $n_0 = 0, ..., N_0$ and $n_1 = 0, ..., N_1$ values that we would expect to contain significant probabilities ($N_0, N_1 >> MOI$). The join distribution is calculated for every combination of n_0, n_1 and multiplying the value at every grid location with the corresponding Oct4 steady-state gives us the mean. Fig. 3-7 plots the ratio of the CV for the 2V to 1V models as a function of MOI and inducer concentration (u).

Here MOI is plotted from 1-10, which are experimentally realizable values over a wide range of inducer concentrations. For large MOI and inducer values, we find that the ratio approaches 1. This agrees with intuition since for large MOI and inducer concentrations, the Hill function term of $f(n)$ given by $\frac{n^2 D}{1+n^2 D}$ saturates and $f(n) \approx n$, $f(n_0, n_1) \approx n_1$. Since both n and n_1 are the same distribution, it would make sense that their CVs match. It is perhaps difficult to explain the trend in other regions of the CV plot until we look at Fig. 3-8. Here we plot the ratio of the means of the 2V and 1V models. To make a valid comparison between both models, the means must be equal. This is only satisfied by the large inducer and high MOI region of the plot where the mean ratio is 1. At every other point, we find that for the same inducer concentration, the mean for the 2V model is lower than its corresponding 1V configuration. A lower mean can skew the CV to be larger for the same standard

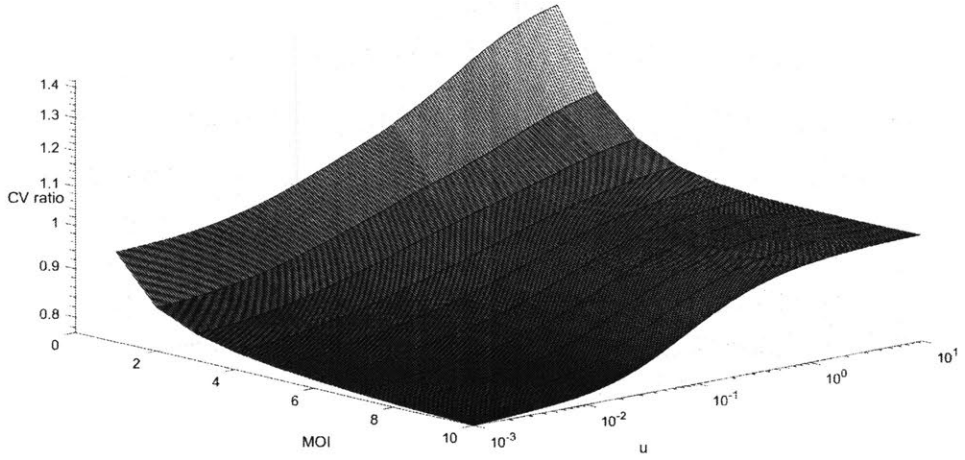

Figure 3-7: CV ratio (2V/1V) for regulator dynamics as a function of MOI and inducer concentration (u).

deviation.

Instead of numerically finding inducer levels that yield the same mean for a given MOI across the 1V and 2V models, we instead plot the analytic pmfs of both models to find a trend. Examples of these plots are shown in Fig. 3-9, where the MOI is changed from MOI=1,5,10. The pmfs of both the 1V and 2V models are shown in red and blue, respectively. The inducer values are adjusted to ensure that the mean for the 1V and 2V models are the same. We find that for MOI=1 and MOI=5, the 2V model has a smaller CV compared to the 1V model for the same mean. We find that if the MOI is not sufficiently large, the 2V case can give a lower CV compared to 1V because there is a higher probability of having values closer to the mean of the distribution. Consider the comparison of the CV of two models $Y_1 = X^2$ and $Y_2 = X_1X_2$, where X, X_1, X_2 are the same distribution (Poisson with same parameter) but independent. Note that the values of Y_1 can only take on squares $0, 1, 4, 9, 16, 25, 36, ...$ while Y_2 can take any non-negative integer value. This means there exist combinations of X_1, X_2 that have relatively high probabilities that are much closer to the mean compared to the Y_1 case and a result we would expect it to have a smaller CV for a given mean.

We now proceed to study the CV of the 2V model. In particular we ask if it is

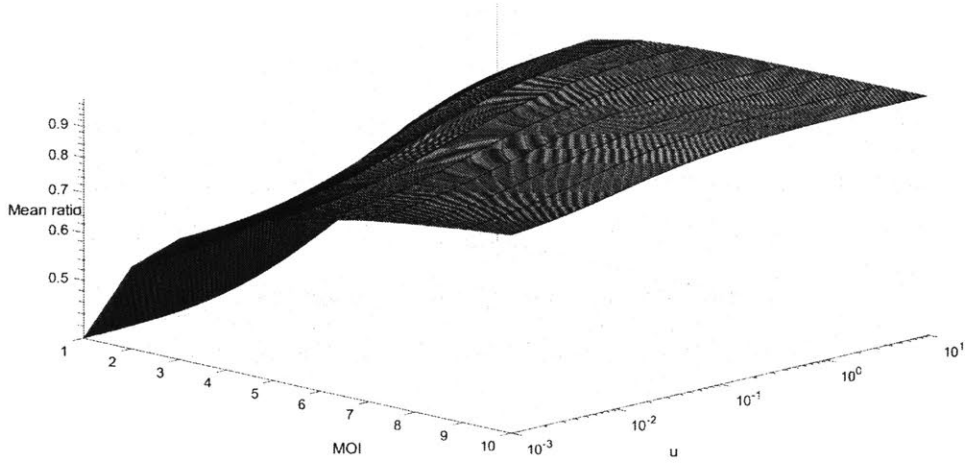

Figure 3-8: Mean ratio (2V/1V) for regulator dynamics as a function of MOI and inducer concentration (u).

favorable to have the regulator on the lower or higher copy number vector compared to the synthetic Oct4 gene vector. To do so, distributions from the 2V model were chosen with roughly the same mean as shown in Fig. 3-10. The 3D plot is of $Y_2 = n_1 \frac{n_0^2 D}{1+n_0^2 D}$ and represents the form for the steady-state of Oct4 with n_0, n_1 being the copy number for the regulator vector and Oct4 vector, respectively. We use this plot to determine combinations of n_0 and n_1 that result in a similar mean.

We find that the CV of the 2V case is more sensitive to the MOI of the regulator and decreases rapidly as this MOI is increased as shown in Fig. 3-11. This is because, for low inducer levels, the Oct4 steady state is roughly proportional to the square of the regulator MOI (D=inducer concentration):

$$O = n_1 \frac{n_0^2 D}{1 + n_0^2 D}, \tag{3.89}$$

$$\approx n_1 n_0^2 D. \tag{3.90}$$

For low MOI (MOI<5) there is a relatively high probability for the Oct4 steady state to be zero even though the mean is relatively large. This Oct4 zero steady-state concentration increases the CV of the distribution. This is captured in the yellow

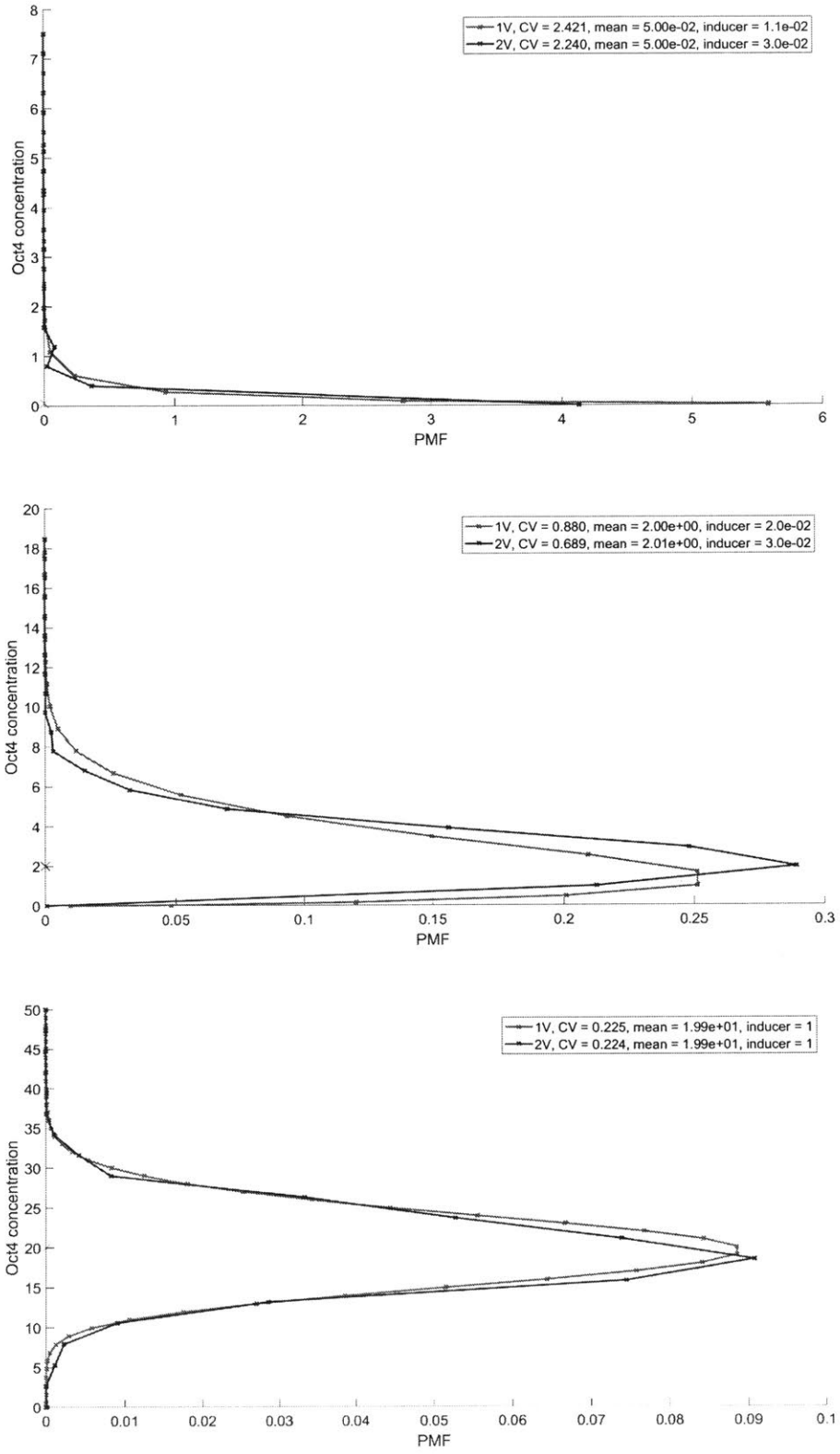

Figure 3-9: Comparison of pmfs for regulator models on 1V vs 2V. We find that the 1V model has a higher CV for the same mean but the difference between the CVs decreases for large MOI.

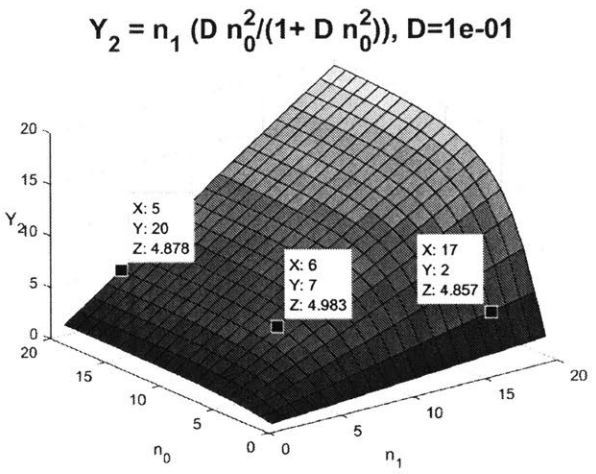

Figure 3-10: Oct4 steady-state as a function of n_0, n_1 used to determine copy numbers of the regulator and synthetic gene vectors yielding the same mean.

curves of Fig. 3-11, where n_0 is relatively small (=2) and so there is a significant probability that the vector copy number of the regulator is 0. When this happens, there can be no Oct4 in the system and so its steady-state is 0. However, since $O \sim n_0^2$, there is still a significant probability for $n_0^2 = 4$ and this "stretches" the distribution, increasing the CV for a given mean. However, we find that the Oct4 steady-state is not significantly affected by n_1. The blue plot represents relatively low copy number for the Oct4 gene and even for $n_1 = 3$, we find that the CV of the distribution is comparable to the distribution with intermediate values of both n_0, n_1 as shown in red. This is once again because of the difference between the linear and squared relationship for n_1 and n_0. Fig. 3-12 plots the CV as a function of MOI, which is calculated as follows (E is the expected value):

$$CV(n) = \frac{\sqrt{E[n^2] - E[n]^2}}{E[n]}, \tag{3.91}$$

$$CV(n^2) = \frac{\sqrt{E[n^4] - E[n^2]^2}}{E[n^2]}. \tag{3.92}$$

We find that for a given MOI, the CV of n is always less than n^2 and that for both cases, the CV decreases monotonically as MOI increases.

In summary, for sufficiently large MOI, the CV is identical for the 1V and 2V

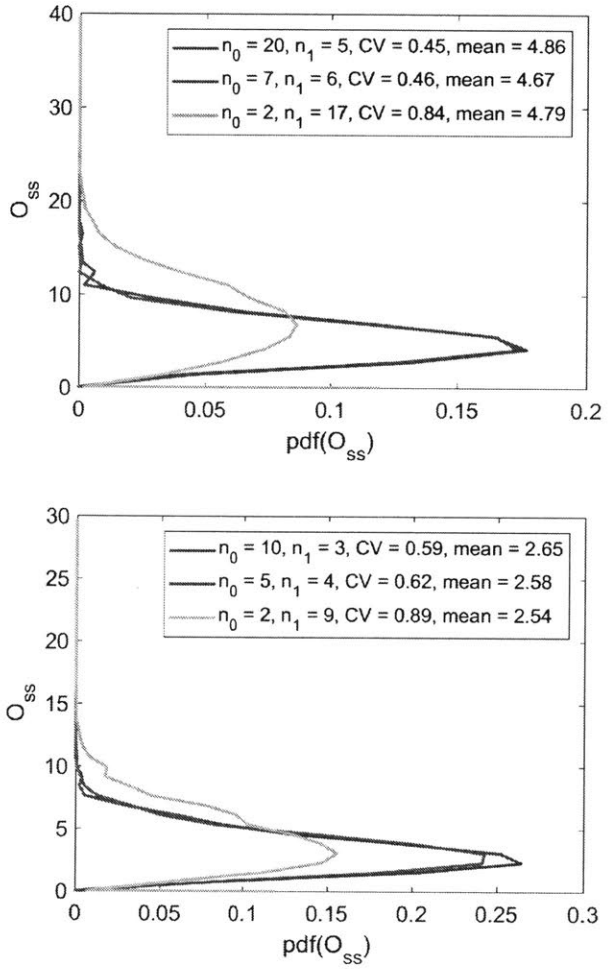

Figure 3-11: Analytic pmf of steady-state Oct4 distribution for 2V for various combinations of regulator (n_0) and Oct4 gene (n_1). **Top:** For roughly the same mean, we find the CV for the yellow curve with $n_0 = 2$ to be roughly twice that of distributions with intermediate MOI. **Bottom:** CV is not as sensitive to low Oct4 vector MOI (blue) compared to regulator vector MOI (yellow).

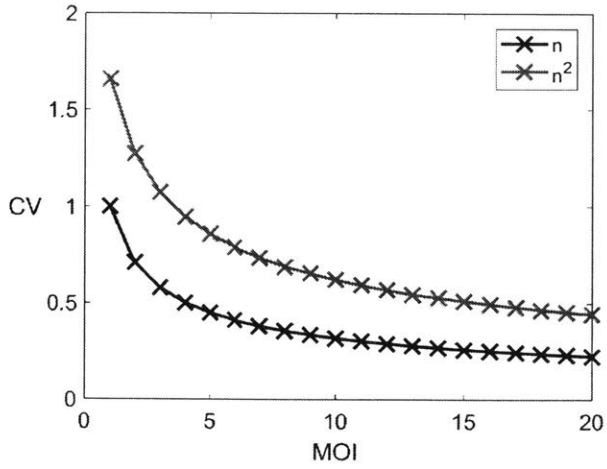

Figure 3-12: CV for Poisson random variables n and n^2 as a function of its parameter (MOI)

models. If the MOI and inducer concentration are sufficiently low, the CV for the 2V model is lower than the 1V case. For the 2V case, the CV is more sensitive to the MOI of the regulator and decreases rapidly as this MOI increases.

3.7 MicroRNA dynamics

The next question to answer is whether to constitutively express microRNA or have it inducible by the same Dox-TetR complex that activates the Oct4 gene. Here we derive a model for constitutively-expressed microRNA (μ) from the biochemical reactions, assuming the microRNA is on a different vector (with copy number n_0) than the Oct4 gene (with copy number n_1):

$$\mathrm{n}_0 \xrightarrow{\alpha_0} \mathrm{n}_0 + \mu, \tag{3.93}$$

$$\mathrm{n}_1 \xrightarrow{\alpha_1} \mathrm{n}_1 + \mathrm{m}, \tag{3.94}$$

$$\mathrm{m} + \mu \underset{\mathrm{d}}{\overset{\mathrm{a}}{\rightleftharpoons}} \mathrm{C} \xrightarrow{\kappa} \mu, \tag{3.95}$$

$$\mathrm{m} \xrightarrow{\alpha_O} \mathrm{m} + \mathrm{O}, \tag{3.96}$$

$$\mathrm{O} \xrightarrow{\delta_O} \emptyset, \tag{3.97}$$

$$\mathrm{C} \xrightarrow{\delta_C} \emptyset, \tag{3.98}$$

$$\mathrm{m} \xrightarrow{\delta_m} \emptyset, \tag{3.99}$$

$$\mu \xrightarrow{\delta_\mu} \emptyset. \tag{3.100}$$

The ODEs that describe this system is given by:

$$\dot{\mu} = \alpha_0 n_0 - \delta_\mu \mu + dC - am\mu + \kappa C, \tag{3.101}$$

$$\dot{m} = \alpha_1 n_1 - \delta_m m + dC - am\mu, \tag{3.102}$$

$$\dot{C} = am\mu - dC - \kappa C - \delta_C C, \tag{3.103}$$

$$\dot{O} = \alpha_O m - \delta_O O. \tag{3.104}$$

Using the model reduction techniques used in previous sections, the dynamics that describe the concentration of the mRNA with constitutively-expressed microRNA is given by:

$$\dot{m} = \alpha_1 n_1 - \delta_m m - \kappa n_0 \frac{m/K_m}{1 + m/K_m}, \tag{3.105}$$

where $K_m = \frac{d+\kappa+\delta_C}{a}$. The mRNA dynamics with the regulator described in the previous section and constitutively-expressed microRNA is given by (assuming the regulator gene is on the same vector as that Oct4 gene):

$$\dot{m} = \alpha_1 n_1 \frac{n_1 D^2}{1 + n_1^2 D} - \delta_m m - \kappa n_0 \frac{m/K_m}{1 + m/K_m}, \tag{3.106}$$

Similarly, for inducible microRNA, the dynamics of mRNA is given by:

$$\dot{m} = \alpha_1 n_1 \frac{n_1 D^2}{1 + n_1^2 D} - \delta_m m - \kappa n_0 \frac{m/K_m}{1 + m/K_m} \frac{n_0 D^2}{1 + n_0^2 D}, \tag{3.107}$$

We study these two models of microRNA (Dox-inducible and constitutive) and compare their CVs with the model without microRNA and the results are summarized below.

3.7.1 Comparison of CV for models with constitutive and no microRNA

We find that the CV of the constitutive microRNA model is lower than that of the model without microRNA for a given mean if $m/K_m << 1$ or if κ is sufficiently large. If $m/K_m << 1$, this reduces Equation 3.106 to:

$$\dot{m} = \alpha_1 n_1 H(D) - \delta_m m - \kappa n_0 \frac{m}{K_m}, \tag{3.108}$$

where $H(D) = \frac{n_1 D^2}{1+n_1^2 D}$. Solving for the mRNA steady-state yields:

$$m = \frac{\alpha_1 n_1 H(D)}{\delta_m + \frac{\kappa n_0}{K_m}}. \tag{3.109}$$

The mRNA steady-state of the system without microRNA is given by:

$$m = \frac{\alpha_1 n_1 H(D)}{\delta_m}. \tag{3.110}$$

We find that the $\frac{n_1}{1+n_0}$ term is effective in decreasing the CV of the distribution when compared to the free n_1 term in the model without microRNA. κ is the degradation catalytic constant and needs to be sufficiently large for the $\kappa n_1 \frac{m/K_m}{1+m/K_m}$ term to be significant. In simulation, this was satisfied as long as κ was larger than dilution, δ_m, which is realistic.

The conditions for the CV of the constitutive microRNA model to be greater than that of the no microRNA model for a given mean include $m/K_m \sim, >> 1$, low

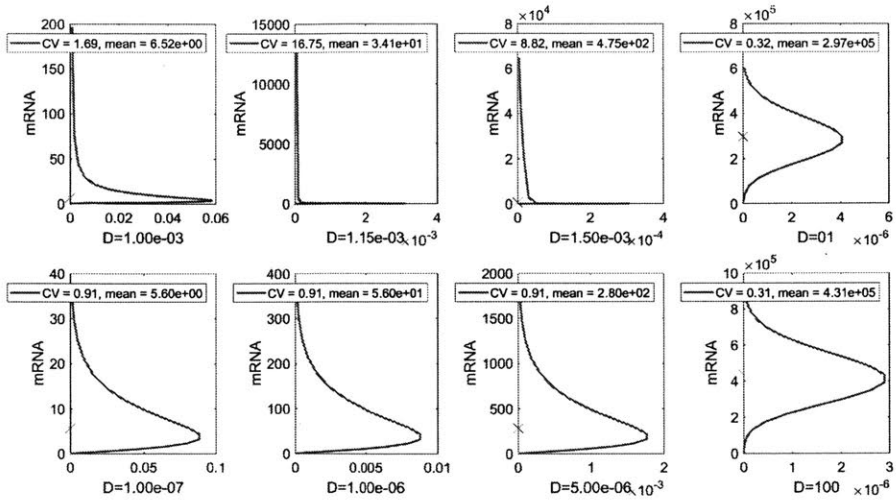

Figure 3-13: Comparison of analytical pmfs of Oct4 mRNA steady-state distribution for models with constitutive microRNA (**top row**) and without microRNA (**bottom row**). The CVs and mean are shown for each distributions and we compare the two models having the same mean (column by column comparison). We find that there are parametric regions where the CV for the constitutive microRNA case is larger than that of the CV for the model without microRNA.

inducer concentration ($H(D)$ should not be saturating) and small κ. When all these conditions are met, the tail of the distribution for the constitutive microRNA case becomes large as shown in Fig. 3-13. Here the 4 plots on the first row are the mRNA pmfs with constitutive microRNA and the 4 plots on the second row are the mRNA pmfs with no microRNA for approximately the same mean.

To see why this happens, we plot the mRNA steady-state as a function of n. The steady-state for Equation 3.106 is given by (assuming microRNA, Oct4, and regulator genes are all on the same vector with copy number n):

$$m = \frac{-\gamma + \sqrt{\gamma^2 + 4\alpha n H(D)\delta_m/K_m}}{2\delta_m/K_m}, \tag{3.111}$$

$$\gamma = \delta_m + \frac{\kappa n - \alpha n H(D)}{K_m}. \tag{3.112}$$

To compute the mean of this distribution, we perform a dot product of the vector $m(n)$ for $n = 0, 1, ..., N$ with the Poisson pmf vector of parameter MOI and sum

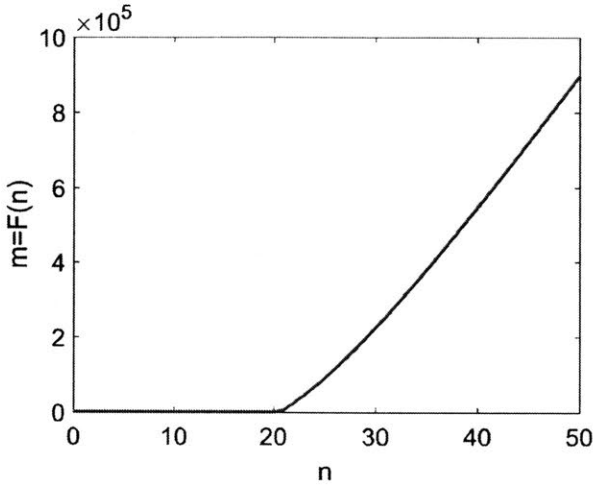

Figure 3-14: mRNA steady-state as a function of n for $K_m = 10$ demonstrating the offset ramp function form.

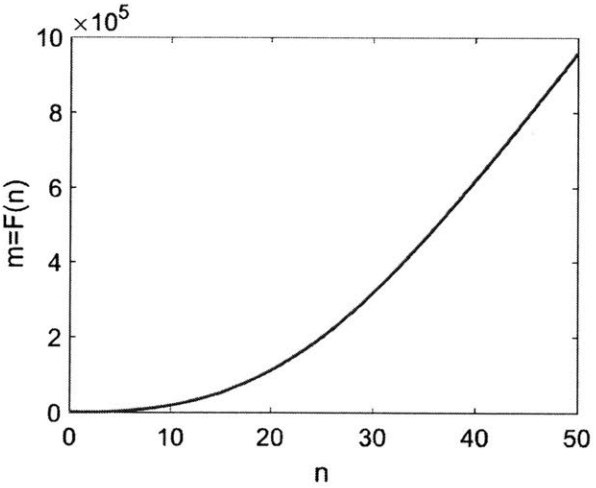

Figure 3-15: mRNA steady-state as a function of n for $K_m = 10^5$ demonstrating a smoother, monotonically increasing function.

the resultant vector ($N >> MOI$ to ensure we capture the whole distribution). Typically, as n increases, $m(n)$ increases, but the probability of these values for large n are so small, they do not affect significantly affect the mean (and CV). However, for the conditions above, we find the mRNA steady-state vs n plot resembles an offset ramp function as shown in Fig. 3-14 for MOI=10. At $n = 10$, $m(n) = 4$ but for $n = 20$, $m(n) = 200$ and the mRNA steady-state values quickly reaches 10^5. For the same parameters (inducer, κ) but large $K_m = 10^5$, we find this plot is significantly smoother as shown in Fig. 3-15. Here the corresponding values for $n = 10, 20$ are $m(n) = 1.9 \times 10^4, 1.1 \times 10^5$, which is less than an order of magnitude closer than the previous case. The CV for this distribution is, as a consequence, significantly lower than the model with intermediate K_m.

3.7.2 Comparison of CV for models with constitutive, inducible, and no microRNA

We find that the CV for the constitutive microRNA case is the lowest, followed by the inducible microRNA and then the no microRNA case if $m/K_m << 1$. In this regime, the microRNA degradation term in $\dot{m}$ is smaller for inducible microRNA than constitutive microRNA since it is scaled by 0<H(n)<1 and not as effective in degrading the mRNA as reflected in the equations:

$$m_{constitutive} = \frac{\alpha_1 n_1 H(D)}{\delta_m + \frac{\kappa n_0}{K_m}}, \tag{3.113}$$

$$m_{inducible} = \frac{\alpha_1 n_1 H(D)}{\delta_m + \frac{H(D)\kappa n_0}{K_m}}. \tag{3.114}$$

Depending on the inducer concentration (i.e., if the promoter is far from saturation) the difference in CV between the constitutive and inducible microRNA case can be significant. Conversely, if the target mRNA level reached is reached for saturating inducer ($H(D) \sim 1$), the CV for both models are similar.

If $m/K_m \sim, >> 1$, we find that the CV for the inducible microRNA model is lower than the CV for the constitutive microRNA for the same mean. Once again,

this is because in this regime the mRNA vs n plot for the constitutive microRNA case resembles the offset ramp function. This behavior is not observed in the system with inducible microRNA because the microRNA level scales with mRNA (unlike constitutive microRNA system) since they are activated by same complex.

3.7.3 Realistic parameter values

We find that typical values of $K_m = 8$ nM [19]. We would like to be operating in the saturating inducer regime when the mRNA levels reach the PL state, which corresponds to $m_{target} = 2$ nM. This corresponds to $m/K_m = 0.25$, which satisfies the $m/K_m < 1$ requirement. Preliminary experiments have shown up to 60% reduction in mRNA level in pluripotent stem cells, and so $\kappa = 5$ nM was used in simulation. For these parametric values, we find that the CV for both inducible and constitutive microRNA cases are similar and are lower than the case without microRNA. Fig. 3-16 shows representative pmfs of mRNA steady state at the target level (2 nM) for the models with no microRNA, constitutive and inducible microRNA, respectively. We find that the CV for both microRNA models is significantly lower (0.3) compared to the model with no microRNA (0.8).

3.8 Conclusions

The goal of this study was to determine the best configuration for a feedback controller to reprogram a differentiated cell back to pluripotency. We considered the Oct4-Nanog GRN, a crucial component of the OSKM transcription factors that have been shown experimentally to trigger a transition from adult cells to induced pluripotent stem cells (iPSCs). iPSCs derivation is typically an inefficient process ($0.01 - 0.1\%$) and this study seeks to improve the current protocol of overexpressing the 4 TFs. We propose the implementation of a feedback controller by lentiviral constructs that shut down the endogenous GRN by microRNA-mediated mRNA degradation and synthetic gene overexpression. We developed a deterministic model and studied two controllers: one with pure endogenous mRNA degradation and the other with degra-

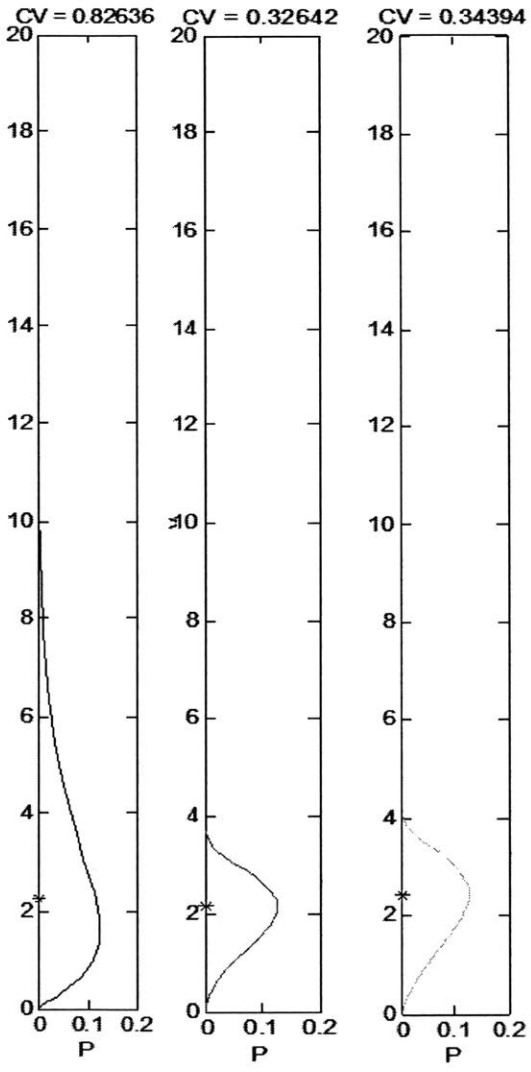

Figure 3-16: Analytical pmfs of models with no microRNA (**left**), constitutive (**middle**), and inducible microRNA (**right**).

dation to endogenous and synthetic mRNA. For this deterministic scenario, we found that both controllers shut down the endogenous network for sufficiently large microRNA copy number and have a linear input-output relationship from inducer concentration to protein steady-state. However, the system with degradation to both mRNA requires significantly higher inducer concentration for any appreciable change in protein steady-state since we are effectively degrading the same protein that we are inducing.

Stochasticity in the form of the vector copy number as a Poisson random variable is then introduced. We find that the coefficient of variation (CV), defined as the ratio of the standard deviation to the mean, is large for the system with just degradation to the endogenous network. This is due to the microRNA not being able to control (via degradation) the synthetically produced protein copy number. Conversely, the model with degradation to both mRNA produces a much tighter distribution of protein steady-state.

We revisit the assumption of being able to increase the production rate of the synthetic protein in a linear manner without bound. Modeling the dynamics of the regulator and inducer and how the complex interacts with the promoter of the Oct4 gene leads to a Hill function activation term, which is approximately linear for low inducer concentration but eventually saturates. We then ask if it is beneficial to place the regulator on the same vector as the synthetic gene or a different one. We find that for sufficiently large MOI, the CV is identical for the 1V and 2V models. If the MOI and inducer concentration are sufficiently low, the CV for the 2V model is lower than the 1V case. For the 2V case, the CV is more sensitive to the MOI of the regulator and decreases rapidly as this MOI increases.

We then study whether it is beneficial to either constitutively express microRNA or place it under the control of the same complex that activates the synthetic gene. We find that under certain parametric regimes, the model for constitutively expressed microRNA demonstrates significantly large CV, which is undesirable. This does not appear in the case for the inducible microRNA. For the parametric region of interest, we find that there is no practical difference between the two configurations of the

microRNA in terms of the CV, which is lower than the no microRNA case.

Bibliography

[1] The 2009 national blood collection and utilization survey report (us department of health and human services, office of the assistant secretary for health). https://www.aabb.org/research/hemovigilance/bloodsurvey/Documents/09-nbcus-report.pdf. Accessed: 2018-05-11.

[2] Fluorescence titering assay for lentivirus. http://www.addgene.org/protocols/fluorescence-titering-assay/. Accessed: 2018-05-05.

[3] Koichi Akashi, David Traver, Toshihiro Miyamoto, and Irving L Weissman. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature*, 404(6774):193, 2000.

[4] Uri Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits.* 2006.

[5] Alexander Birbrair and Paul S Frenette. Niche heterogeneity in the bone marrow. *Annals of the New York Academy of Sciences*, 1370(1):82–96, 2016.

[6] Vijay Chickarmane, Tariq Enver, and Carsten Peterson. Computational modeling of the hematopoietic erythroid-myeloid switch reveals insights into cooperativity, priming, and irreversibility. *PLoS computational biology*, 5(1):e1000268, 2009.

[7] Stella T Chou, Eugene Khandros, L Charles Bailey, Kim E Nichols, Christopher R Vakoc, Yu Yao, Zan Huang, John D Crispino, Ross C Hardison, Gerd A Blobel, et al. Graded repression of pu. 1/sfpi1 gene transcription by gata factors regulates hematopoietic cell fate. *Blood*, 114(5):983–994, 2009.

[8] M Ryan Corces, Jason D Buenrostro, Beijing Wu, Peyton G Greenside, Steven M Chan, Julie L Koenig, Michael P Snyder, Jonathan K Pritchard, Anshul Kundaje, William J Greenleaf, et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nature genetics*, 48(10):1193–1203, 2016.

[9] M. Ryan Corces, Jason D. Buenrostro, Beijing Wu, Peyton G. Greenside, Steven M. Chan, Julie L. Koenig, Michael P. Snyder, Jonathan K. Pritchard, Anshul Kundaje, William J. Greenleaf, Ravindra Majeti, and Howard Y. Chang. Lineage-specific and single-cell chromatin accessibility charts human

hematopoiesis and leukemia evolution. *Nature Genetics*, 48:1193 EP –, Aug 2016. Article.

[10] Merlin Crossley, Menie Merika, and Stuart H Orkin. Self-association of the erythroid transcription factor gata-1 mediated by its zinc finger domains. *Molecular and Cellular Biology*, 15(5):2448–2456, 1995.

[11] Ana Cvejic. Mechanisms of fate decision and lineage commitment during haematopoiesis. *Immunology and cell biology*, 94(3):230, 2016.

[12] Domitilla Del Vecchio, Hussein Abdallah, Yili Qian, and James J. Collins. A blueprint for a synthetic genetic feedback controller to reprogram cell fate. *Cell Systems*, 4(1):109–120.e11, Jan 2017.

[13] Louis C Doré and John D Crispino. Transcription factor networks in erythroid cell and megakaryocyte development. *Blood*, 118(2):231–239, 2011.

[14] Phillipp Ellison and Martin Feinberg. How catalytic mechanisms reveal themselves in multiple steady-state data: I. basic principles. *Journal of Molecular Catalysis A: Chemical*, 154(1):155–167, 2000.

[15] Phillipp Ellison, Martin Feinberg, Haixia Ji, and D Knight. Chemical reaction network toolbox v2.3. *Available online at* `http://www.crnt.osu.edu/CRNTWin`, 2011.

[16] Phillipp Raymond Ellison. *The advanced deficiency algorithm and its applications to mechanism discrimination.* PhD thesis, The University of Rochester, 1998.

[17] Péter Érdi and János Tóth. *Mathematical models of chemical reactions: theory and applications of deterministic and stochastic models.* Manchester University Press, 1989.

[18] Thomas Graf and Tariq Enver. Forcing cells to change lineages. *Nature*, 462:587 EP –, Dec 2009. Review Article.

[19] Benjamin Haley and Phillip D. Zamore. Kinetic analysis of the rnai enzyme complex. *Nature Structural &Amp; Molecular Biology*, 11:599 EP –, May 2004. Article.

[20] Peter C Hollenhorst, Lawrence P McIntosh, and Barbara J Graves. Genomic and biochemical insights into the specificity of ets transcription factors. *Annual review of biochemistry*, 80:437–471, 2011.

[21] Sui Huang. Reprogramming cell fates: reconciling rarity with robustness. *BioEssays*, 31(5):546–560.

[22] Sui Huang, Yan-Ping Guo, Gillian May, and Tariq Enver. Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Developmental biology*, 305(2):695–713, 2007.

[23] Hiromi Iwasaki and Koichi Akashi. Myeloid lineage commitment from the hematopoietic stem cell. *Immunity*, 26(6):726–740, 2007.

[24] Haixia Ji. *Uniqueness of equilibria for complex chemical reaction networks.* PhD thesis, The Ohio State University, 2011.

[25] Mads Kærn, Timothy C Elston, William J Blake, and James J Collins. Stochasticity in gene expression: from theories to phenotypes. *Nature Reviews Genetics*, 6(6):451–464, 2005.

[26] Diane S Krause. Regulation of hematopoietic stem cell fate. *Oncogene*, 21(21):3262, 2002.

[27] Thomas G Kurtz. The relationship between stochastic and deterministic models for chemical reactions. *The Journal of Chemical Physics*, 57(7):2976–2978, 1972.

[28] Chu Wai Liew, Kasper D Rand, Raina JY Simpson, Wendy W Yung, Robyn E Mansfield, Merlin Crossley, Mette Proetorius-Ibba, Claus Nerlov, Flemming M Poulsen, and Joel P Mackay. Molecular analysis of the interaction between the hematopoietic master transcription factors gata-1 and pu. 1. *Journal of Biological Chemistry*, 281(38):28296–28306, 2006.

[29] Luca Mariani, Edda G Schulz, Maria H Lexberg, Caroline Helmstetter, Andreas Radbruch, Max Löhning, and Thomas Höfer. Short-term memory in gene induction reveals the regulatory principle behind stochastic il-4 expression. *Molecular systems biology*, 6(1):359, 2010.

[30] Kathryn Miller-Jensen, Siddharth S Dey, David V Schaffer, and Adam P Arkin. Varying virulence: epigenetic control of expression noise and disease processes. *Trends in biotechnology*, 29(10):517–525, 2011.

[31] Victoria Moignard, Iain C Macaulay, Gemma Swiers, Florian Buettner, Judith Schütte, Fernando J Calero-Nieto, Sarah Kinston, Anagha Joshi, Rebecca Hannah, Fabian J Theis, et al. Characterisation of transcriptional networks in blood stem and progenitor cells using high-throughput single cell gene expression analysis. *Nature cell biology*, 15(4):363, 2013.

[32] Samantha A. Morris and George Q. Daley. A blueprint for engineering cell fate: current technologies to reprogram cell identity. *Cell Research*, 23:33 EP –, Jan 2013. Review.

[33] Claus Nerlov and Thomas Graf. Pu. 1 induces myeloid lineage commitment in multipotent hematopoietic progenitors. *Genes & development*, 12(15):2403–2412, 1998.

[34] Keizo Nishikawa, Makoto Kobayashi, Atsuko Masumi, Susan E Lyons, Brant M Weinstein, P Paul Liu, and Masayuki Yamamoto. Self-association of gata1 enhances transcriptional activity in vivo in zebra fish embryos. *Molecular and cellular biology*, 23(22):8295–8305, 2003.

[35] Stephen L Nutt, Donald Metcalf, Angela D'Amico, Matthew Polli, and Li Wu. Dynamic regulation of pu. 1 expression in multipotent hematopoietic progenitors. *Journal of Experimental Medicine*, 201(2):221–231, 2005.

[36] Yutaka Okuno, Gang Huang, Frank Rosenbauer, Erica K Evans, Hanna S Radomska, Hiromi Iwasaki, Koichi Akashi, Francoise Moreau-Gachelin, Youlin Li, Pu Zhang, et al. Potential autoregulation of transcription factor pu. 1 by an upstream regulatory element. *Molecular and cellular biology*, 25(7):2832–2845, 2005.

[37] Stuart H Orkin. Transcription factors and hematopoietic development. *Journal of Biological Chemistry*, 270(10):4955–4958, 1995.

[38] A Paldi. Stochastic gene expression during cell differentiation: order from disorder? *Cellular and molecular life sciences*, 60(9):1775–1778, 2003.

[39] Arjun Raj, Charles S Peskin, Daniel Tranchina, Diana Y Vargas, and Sanjay Tyagi. Stochastic mrna synthesis in mammalian cells. *PLoS Biol*, 4(10):e309, 2006.

[40] Michael A Rieger and Timm Schroeder. Hematopoiesis. *Cold Spring Harbor perspectives in biology*, 4(12):a008250, 2012.

[41] Ingo Roeder and Ingmar Glauche. Towards an understanding of lineage specification in hematopoietic stem cells: a mathematical model for the interaction of transcription factors gata-1 and pu. 1. *Journal of theoretical biology*, 241(4):852–865, 2006.

[42] Frank Rosenbauer and Daniel G Tenen. Transcription factors in myeloid development: balancing differentiation with transformation. *Nature reviews. Immunology*, 7(2):105, 2007.

[43] Hemlata Sahu and A. Verma. Adverse events related to blood transfusion. *Indian J Anaesth*, 58:543–551, 2014. Article.

[44] Michael H Sieweke and Thomas Graf. A transcription factor party during blood cell differentiation. *Current opinion in genetics & development*, 8(5):545–551, 1998.

[45] Tomas Stopka, Derek F Amanatullah, Michael Papetti, and Arthur I Skoultchi. Pu. 1 inhibits the erythroid program by binding to gata-1 on dna and creating a repressive chromatin structure. *The EMBO journal*, 24(21):3712–3723, 2005.

[46] Kazutoshi Takahashi. Cellular reprogramming – lowering gravity on waddington's epigenetic landscape. *Journal of Cell Science*, 125(11):2553–2560, 2012.

[47] Kazutoshi Takahashi and Shinya Yamanaka. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, 126(4):663 - 676, 2006.

[48] Tianhai Tian and Kate Smith-Miles. Mathematical modeling of gata-switching for regulating the differentiation of hematopoietic stem cell. *BMC systems biology*, 8(1):S8, 2014.

[49] Tsz-Leung To and Narendra Maheshri. Noise can induce bimodality in positive transcriptional feedback loops without bistability. *Science*, 327(5969):1142–1145, 2010.

[50] Ty C Voss and Gordon L Hager. Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nature Reviews Genetics*, 15(2):69–81, 2014.

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

[52] Lei Yuan, Gary C Chan, David Beeler, Lauren Janes, Katherine C Spokes, Harita Dharaneeswaran, Anahita Mojiri, William J Adams, Tracey Sciuto, Guillermo Garcia-Cardeña, et al. A role of stochastic phenotype switching in generating mosaic endothelial cell heterogeneity. *Nature communications*, 7, 2016.

[53] Pu Zhang, Gerhard Behre, Jing Pan, Atsushi Iwama, Nawarat Wara-Aswapati, Hanna S Radomska, Philip E Auron, Daniel G Tenen, and Zijie Sun. Negative cross-talk between hematopoietic regulators: Gata proteins repress pu. 1. *Proceedings of the National Academy of Sciences*, 96(15):8705–8710, 1999.

[54] Pu Zhang, Xiaobo Zhang, Atsushi Iwama, Channing Yu, Kent A Smith, Beatrice U Mueller, Salaija Narravula, Bruce E Torbett, Stuart H Orkin, and Daniel G Tenen. Pu. 1 inhibits gata-1 function and erythroid differentiation by blocking gata-1 dna binding. *Blood*, 96(8):2641–2648, 2000.