1-1-2013

Forward and Reverse Engineering of Cellular Decision-Making

Najaf Ali Shah
University of Pennsylvania, najaf@mail.med.upenn.edu

Forward and Reverse Engineering of Cellular Decision-Making

Abstract

Cells reside in highly dynamic environments to which they must adapt. Throughout its lifetime, an individual cell receives numerous chemical and mechanical signals, communicated through dense molecular networks, and eliciting a diverse array of responses. A large number of these signals necessitate discrete, all-or-none responses. For instance, a cell receiving proliferation signals must respond by committing to the cell-cycle and dividing, or by not initiating the process at all; that is, the cell must not adopt an intermediate route. Analogously, a stem-cell receiving signals for different lineages must commit exclusively to one of these lineages. How individual cells integrate multiple, possibly conflicting, noisy inputs, and make discrete decisions is poorly understood. Detailed insight into cellular decision-making can enable cell-based therapies, shed light on diseases arising out of dysregulation of control, and suggest practical design strategies for implementing this behavior in synthetic systems for research and industrial use.

In this thesis, we have employed both mathematical modeling and experiments to further elucidate the mechanistic underpinnings of decision-making in cells. First, we describe a computational study that assesses the entire space of minimal networks to identify topologies that can not only make decisions but can do so robustly in the dynamic and noisy cellular environment.

Second, via model-driven, quantitative experiments in a megakaryocyte erythroid progenitor line, we demonstrate that a simple network with mutual antagonism and autoregulation captures the dynamics of the master transcription factors at the level of individual cells. Expansion of this model to account for extrinsic cues reconciles the competing stochastic and instructive theories of hematopoietic lineage commitment, and implicates cytokine receptors in broader regulatory roles.

Third, to assess the impact of specific genetic perturbations on the distribution of the population, and on commitment trajectories of individual cells, we implemented the core mutual antagonism and autoregulation topology synthetically in yeast cells. Our approach of using orthogonal variants of a single core protein represents a general, modular design strategy for building synthetic circuits, and model-driven experiments elucidate how gene dosage, repression strength, and promoter architecture can modulate decision-making behavior.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Genomics & Computational Biology

First Advisor
Casim A. Sarkar

FORWARD AND REVERSE ENGINEERING OF CELLULAR

DECISION-MAKING

Najaf A. Shah

A DISSERTATION

in

Genomics and Computational Biology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2013

Supervisor of Dissertation

_____

Casim A. Sarkar

Assistant Professor of Bioengineering

Graduate Group Chairperson

_____

Maja Bucan

Professor of Genetics

Dissertation Committee

Arjun Raj, Assistant Professor of Bioengineering (Committee chair)

Brian Y. Chow, Assistant Professor of Bioengineering

Scott L. Diamond, Professor of Chemical & Biomolecular Engineering

Ravi Radhakrishnan, Associate Professor of Bioengineering

FORWARD AND REVERSE ENGINEERING OF
CELLULAR DECISION-MAKING

This thesis is dedicated to my parents, Kausar and Sharaf.

ACKNOWLEDGEMENTS

As my Penn journey draws to a close, I look back with fondness and gratitude for the six years I have spent here. The opportunity to study challenging, fundamental problems alongside accomplished mentors and colleagues has been nothing less than an extraordinary privilege, one that I am cognizant is afforded to few of the many deserving aspirants.

Foremost amongst the cast of characters that has shaped my experience at Penn is, of course, my advisor, Casim Sarkar, to whom I owe an enormous debt of gratitude. Casim welcomed me into his group despite the fact that I had essentially zero wet-lab experience, gave me a great deal of latitude in choosing projects, entertained many of my naive ideas about experiments, allowed me to fail and learn from my mistakes, and very generously provided scientific guidance and mentorship throughout. In another lab, the path to the thesis might perhaps have been shorter and less challenging, but the breadth and depth of the learning experience, I am convinced, would not have been anywhere near what I have benefited from in Casim's lab.

Life in the Sarkar lab would have been much less colorful without the camaraderie of my uniquely brilliant fellow lab-mates. Santhosh Palani oriented me to the lab, gave useful advice on the topics worth pursuing (Systems Biology) and the ones best avoided (Ribosome Display), provided interesting analyses on cricket and finance, and introduced me to the joys and frustrations of experiments; the fact that this thesis contains any experimental work is due in large part to his willingness and patience in training me. Santhosh's work on erythrocyte commitment also laid the foundation for the erythrocyte/megakaryocyte conflict study in this thesis.

Pamela Barendt generously shared her encyclopedic knowledge of molecular cloning, helped debug many of my failed attempts at PCR, and, overall, challenged me to adopt a more disciplined and rational approach to designing and performing experiments. Apart from her friendship, I will always remember her for her inspiring and intimidating work ethic, pipetting late into the evening, with her daughter playing

happily in a box nearby. Pam also devised the cloning strategies for several constructs for the synthetic circuits project.

Ellen O'Shaughnessy shared her expertise on yeast protocols, introduced me to the versatile TetR system, and provided many of the base plasmids from which the synthetic circuits discussed in this thesis were built. Daphne Ng provided sound advice on a number of topics, especially during the last two years, when she also assumed most of the responsibilities of keeping the lab well-stocked and in code. I have also had the opportunity to work with three very talented undergraduates: Michael Magaraci, Christine Li, and Dalton Banks.

Our work on hematopoiesis has benefited greatly from our collaboration with Arjun Raj's lab. Arjun designed the FISH probes and very generously allowed us to regularly use the instruments in his lab. In Arjun's lab, Marshall Levesque introduced us to the probe-coupling procedure and helped troubleshoot image-analysis issues, and Gautham Nair, Hedia Maamar, Patrick McClanahan, and Yaanik Desai provided useful tips on microscopy.

My committee members, Brian Chow, Scott Diamond, Ravi Radhakrishnan, and Arjun Raj, provided useful feedback and have improved the work presented in this thesis. Many others in the Penn community have also been of assistance. Our neighbors in Matthew Lazzara's lab, Calixte Monast, Alice Macdonald, Janine Buonato, and Chris Furcht shared reagents and advice on cell culture, qPCR, and numerous other topics. JV Vandergrift of Life Technologies suggested using digital PCR for analyzing plasmid copy-numbers and introduced us to the technique along with help from Celia Chang and Sandy Widura at the Wistar Institute. Andrea Stout helped us with live-cell microscopy of yeast.

During my first two years at Penn, I spent most of my time with the GCB graduate group. I am grateful to GCB students, in particular Brett Hannigan, Sameer Soi, Nick Lahens, Nick Stong, Josh Burdick, Jun Chen, and Paul Ryvkin for their friendship. I also thank Maja Bucan, our graduate group's chair, and Hannah Chervitz, our coordinator, for their patience with various requirements and administrative tasks.

Thanks also to Carlo Maley, in whose lab I spent a highly informative and enjoyable rotation.

Before starting graduate school, I spent a couple of years in John Hartman's lab at the University of Alabama at Birmingham (UAB), which in retrospect, served as excellent preparation for graduate school. It was during this time that I realized the importance of developing a clear understanding of how experiments are conducted, and I am grateful to John for this experience. During my time as an undergraduate at UAB, my mentor Anthony Skjellum first introduced me to scientific research and provided invaluable support and advice during my search for a graduate program.

Most of all, I want to take this opportunity to thank my oldest and most loyal support group: my parents, Sharaf and Kausar, and my siblings, Musawir and Huma. My brother has shared with me his passion for computing and has always been on standby to help extricate me from the various health and other situations I have managed to put myself into. My father's strong sense of integrity and service has long been a source of inspiration. My mother has called often, feigned interest in my work, and has harbored an irrational faith in my abilities for as long as I can remember.

My parents have sacrificed much, both in the way of personal comfort and in enduring lengthy separations so that my siblings and I could enjoy the privilege of a good education. Attempting to repay such a profound debt would be futile; I shall instead strive to do justice to the opportunities their sacrifices have enabled me to have.

ABSTRACT

FORWARD AND REVERSE ENGINEERING OF CELLULAR

DECISION-MAKING

Najaf A. Shah

Casim A. Sarkar

Cells reside in highly dynamic environments to which they must adapt. Throughout its lifetime, an individual cell receives numerous chemical and mechanical signals, communicated through dense molecular networks, and eliciting a diverse array of responses. A large number of these signals necessitate discrete, all-or-none responses. For instance, a cell receiving proliferation signals must respond by committing to the cell-cycle and dividing, or by not initiating the process at all; that is, the cell must not adopt an intermediate route. Analogously, a stem-cell receiving signals for different lineages must commit exclusively to one of these lineages. How individual cells integrate multiple, possibly conflicting, noisy inputs, and make discrete decisions is poorly understood. Detailed insight into cellular decision-making can enable cell-based therapies, shed light on diseases arising out of dysregulation of control, and suggest practical design strategies for implementing this behavior in synthetic systems for research and industrial use.

In this thesis, we have employed both mathematical modeling and experiments to further elucidate the mechanistic underpinnings of decision-making in cells. First, we describe a computational study that assesses the entire space of minimal networks to identify topologies that can not only make decisions but can do so robustly in the dynamic and noisy cellular environment.

Second, via model-driven, quantitative experiments in a megakaryocyte erythroid progenitor line, we demonstrate that a simple network with mutual antagonism and autoregulation captures the dynamics of the master transcription factors at the level of individual cells. Expansion of this model to account for extrinsic cues reconciles the competing stochastic and instructive theories of hematopoietic lineage commitment,

and implicates cytokine receptors in broader regulatory roles.

Third, to assess the impact of specific genetic perturbations on the distribution of the population, and on commitment trajectories of individual cells, we implemented the core mutual antagonism and autoregulation topology synthetically in yeast cells. Our approach of using orthogonal variants of a single core protein represents a general, modular design strategy for building synthetic circuits, and model-driven experiments elucidate how gene dosage, repression strength, and promoter architecture can modulate decision-making behavior.

Table of Contents

List of Tables

List of Figures

Chapter 1

Background and Motivation

1.1 Cellular decision-making

Cells from across the three domains of life inhabit a dynamic environment to which they must adapt, at the population and individual-cell levels, and on timescales ranging from that of evolution to the lifetime of an individual organism. Throughout its existence, a unicellular organism may experience numerous environmental changes, such as a shift in pH or temperature, the exertion of a force, or a switch in nutrient source. These chemical, mechanical, and environmental perturbations are sensed by specialized proteins and communicated as molecular messages across networks of signaling components. In multicellular organisms, signaling networks play an expanded role by facilitating communication among cellular components and between different cells, and therefore enable organisms to coordinate spatially and temporally distant processes.

Research aimed at elucidating how cells respond to extrinsic and intrinsic cues has lead to the identification and characterization of a diverse array of messenger molecules and receptor proteins, and has enabled the delineation of signaling pathways that connect individual stimuli to specific cellular responses. Progress in understanding signal propagation has lead to the realization that in many cellular con-

texts, signaling networks perform signal processing; one such situation that arises ubiquitously in biology is the induction of a discrete, all-or-none response [Ferrell Jr., 1998, Kholodenko et al., 2010, Novick and Weiner, 1957]. Despite the diversity in signaling components and the downstream processes they modulate, biological cues eliciting discrete responses share characteristics which place important requirements on the signaling networks that communicate them.

First, most biological cues are continuous, or graded. Hence, if the downstream effect is a discrete, all-or-none event such as apoptosis [Eissing et al., 2004, Bagci et al., 2006, Legewie et al., 2006], the signaling network must establish a threshold such that every possible level of the signal results in one of the two discrete possibilities, with no intermediates.

Second, cues can be short-lived, or fluctuating; a cell may receive a signal only transiently, or the signal strength may change dramatically. In this situation, the network must establish a time threshold and *remember* the signal even after it declines or ceases completely. For instance, if the signal induces entry into the cell cycle, the cell must complete all steps even when the signal drops below the threshold during anaphase [Lisman, 1985, Burrill and Silver, 2010].

Third, cues often conflict with other cues, and hence networks must resolve conflicts between incompatible and mutually exclusive events. For instance, a multipotent progenitor cell might receive high levels of multiple signals, promoting differentiation into opposing lineages [Park et al., 1999].

Networks that fulfill these requirements of transforming multiple, possibly conflicting, graded, and noisy cues into discrete, all-or-none responses in effect perform a type of analog-to-digital conversion that can be broadly referred to as decision-making. Cellular decision-making is ubiquitously present across all organisms, and plays crucial roles in myriad processes including cell-cycle control, differentiation of stem cells, and viral infection [Balázsi et al., 2011]. Since dysregulation of cellular decision-making can have profound consequences on human health, elucidation of the mechanisms that enable and confer robustness to this process can aid in the de-

velopment of therapeutic strategies to address important diseases. In this thesis, we employ both theory and experiments to infer molecular network architectures enabling decision-making, and study decision-making in detail in red-blood cell differentiation, and in synthetic circuits.

1.2   Decision-making in living systems

To illustrate its pervasiveness across diverse cellular contexts, here we describe a few examples of decision-making.

1.2.1   Decision-making in bacteria

One of the earliest examples of decision-making behavior in cells was described in 1957 by Novick and Weiner, who demonstrated that treatment of a genotypically clonal population of *Escherichia coli* cells with the inducer TMG (thiomethyl-β-D-galactoside) results in a dose-dependent increase in the expression of β-galactosidase overall not because each cell produces more product, but because higher concentrations of stimulus increase an individual cell's probability of producing β-galactosidase at a particular rate. Hence, depending on the concentration of the inducer, an individual cell decides to either express or not express β-galactosidase. Furthermore, cells that are treated with high levels of inducer continue to express β-galactosidase at the same rate even as the inducer concentration is subsequently reduced to a sub-threshold level, indicating that the cells retain memory of the induction [Novick and Weiner, 1957].

The example above also illustrates why it can be difficult to identify even relatively simple instances of decision-making behavior: most assays in biological experiments are performed on whole populations and yield bulk measurements that are explained equally well by continuous mechanisms than by all-or-none phenomena, which tend to be less intuitive. However, advances in genetic manipulation techniques, and the in-

troduction of a variety of single-cell assays over the past decade have made monitoring of individual cells more accessible.

1.2.2 Decision-making in mammalian cells

In mammals, cell proliferation is a tightly controlled process, and is regulated at each step by multiple mechanisms; evasion of these mechanisms by individual cells can lead to uncontrolled growth and cancer in the host [Murray, 1992]. Work by Pardee and others demonstrates that the mechanisms regulating cell proliferation effectively impose a 'restriction point,' such that an individual cell is in either a quiescent or a proliferative state. Specifically, recently-divided cells can be shifted to the quiescent state by removing mitogenic cues from the culture medium; however, if these cues are removed after a threshold length of time, the cells stay in the proliferative state and complete all steps of the cell cycle [Pardee, 1974, Malumbres and Barbacid, 2001]. Using flow cytometry and microscopy, Yao and colleagues demonstrated that this decision-making process is mediated by the Rb-E2F pathway which yields discrete responses to continuous serum inputs [Yao et al., 2008].

1.2.3 Decision-making in viral infection

The bacteriophage λ lysis/lysogeny switch is perhaps the best-studied decision-making system. Upon entering a host cell, the virus either enters the lysis state, in which it induces expression of genes which enable it to replicate and lyse the host, or the lysogeny state, in which expression of an alternate set of genes enable it to integrate into the host genome, where it persists until induced by other cues to enter the lysis state. This decision-making behavior is primarily regulated by the repressors, *CI* and *cro* [Golding, 2011].

The HIV virus faces essentially the same two options when it infects a CD4+ lymphocyte: proviral latency and replication. A latent virus switching to the replication state can have profound consequences in a patient who is not under continuous anti-

retroviral therapy. In HIV, the fate decision is regulated by the *Tat* transcriptional activator [Razooky and Weinberger, 2011].

1.3   Mathematical modeling of decision-making

Part of the difficulty in studying decision-making in mechanistic detail stems from the fact that it is an emergent behavior that manifests itself at the cellular or organismal levels, and arises due to relatively simple interactions at the molecular level. Hence, decision-making behavior in a particular system usually cannot be understood through a cursory analysis of network diagrams depicting interactions between the system's components, even though construction of such diagrams represents an essential first step [Kitano, 2002].

Dynamical systems modeling extends the network approach by incorporating kinetic information and representing interactions between components as high-level chemical reactions. For a particular biological process, reactions such as the binding of a ligand to a receptor, or the enzyme-mediated activation of a transcription factor, are modeled by a system of rate equations, which can be simulated for a given set of initial conditions; hence these models can recapitulate observed behaviors and, more importantly, can predict system response for a different configuration of inputs. Systems modeling has been extensively applied to elucidate the mechanisms driving decision-making behavior in a number of systems, perhaps most notably in the phage lysis-lysogeny switch [Golding, 2011, Arkin et al., 1998] and the *lac* operon mentioned previously [Ozbudak et al., 2004].

1.3.1   Decision-making via ultrasensitivity and bistability

Decision-making encompasses the formal mathematical concepts of ultrasensitivity and bistability. Ultrasensitivity is a useful systems-level property in cellular contexts in which a threshold concentration of stimulus triggers entry into a different cellu-

lar state, but intermediate states must be avoided [Goldbeter and Koshland, 1981]. Notable examples of systems exhibiting ultrasensitivity include the MAPK cascade in *Xenopus* oocytes [Huang and Ferrell, 1996], the pathway regulating the mating decision in yeast [Malleshaiah et al., 2010], and the circuit controlling differentiation in the *Drosophila* embryo [Melen et al., 2005]. Ultrasensitivity can arise from several mechanisms: positive feedback [Ferrell, 2002]; cascading [O'Shaughnessy et al., 2011], cooperativity [Koshland et al., 1982], which can result from multimerization [Gardner et al., 2000]; distributive multi-site activation, in which a substrate is released from an enzyme after each activation and must re-bind before the next activation can take place [Markevich et al., 2004]; and zero-order effects, which occurs, for example, when a kinase and phosphatase pair act on a substrate under saturating conditions [Goldbeter and Koshland, 1981, Goldbeter, 2005].

Although ultrasensitive systems can filter the effects of stimulus variation at concentrations far from the switching threshold, minor fluctuations in stimulus concentration near the threshold can cause the system to switch back and forth between the two states. Hence, mechanisms such as cross-antagonism and positive feedback are often employed by a cell to achieve bistability. The hysteresis, or memory effect, that arises as a consequence of bistability enables the system to tolerate stochastic fluctuations in the stimulus and the network species, and in some cases confers irreversibility, allowing the system to lose its dependence on stimulus [Ferrell, 2008, Xiong and Ferrell, 2003, Ferrell, 2002, Becskei et al., 2001, Pomerening, 2008, Ninfa and Mayo, 2004]. Bistability has been observed in numerous biological systems, including the *lac* operon in bacteria [Santillán et al., 2007, Ozbudak et al., 2004, Novick and Weiner, 1957], the network regulating differentiation of erythroid and myelomonocytic lineages [Huang et al., 2007], and the circuit governing exit from quiescence in mammalian cells [Yao et al., 2008]. Bistability has also been engineered in synthetic systems using mechanisms such as cross-antagonism [Gardner et al., 2000] and positive feedback[Palani and Sarkar, 2011], as well as more non-intuitive mechanisms such as negative growth modulation of the host cell [Tan et al., 2009].

Ultrasensitivity and bistability are usually studied by building ordinary differential equation (ODE) models which are solved analytically, or simulated by numerical integration methods. In such studies, the goal is typically to identify the explicit or emergent source of non-linearity that renders the decision-making behavior possible. These models are termed deterministic; that is, for a given set of parameters and initial conditions, every simulation run produces the same output.

1.3.2 Decision-making via noise

Deterministic modeling has been successfully applied to the study of many systems and has yielded important insights; however, a significant drawback of this approach is that it assumes that proteins, nucleic acids, and other species are present in large numbers of molecules. This assumption is violated in a number of important situations which involve interactions between species that are present in small numbers, perhaps most notably in gene regulation, since most cells in diploid organisms contain only two copies (or molecular instances) of each gene.

Furthermore, recent studies have shown that transcription of eukaryotic genes can be highly stochastic; promoters in eukaryotic cells initiate production of mRNA molecules not at a steady rate, but in infrequent bursts [Raj et al., 2006, Raj and van Oudenaarden, 2008, Elowitz et al., 2002]. The increased appreciation of the role of noise has led to the use of the Gillespie algorithm [Gillespie, 1977] in simulating biological systems; this alternative approach simulates each specified interaction explicitly, and converges to the deterministic solution under large numbers of molecules. In the context of decision-making, a number of recent studies have employed stochastic modeling and single-cell experiments to demonstrate that noise alone can effectively render decisions in networks, even in the absence of non-linearity. Notable examples include the circuit deciding between proviral latency and replication in HIV [Razooky and Weinberger, 2011], and a simple autoregulatory loop implemented synthetically in yeast [Lee and Maheshri, 2012].

1.4 Focus of thesis

Cells inhabit a highly dynamic physiological environment, receiving myriad molecular, environmental, and mechanical cues, to which they must mediate appropriate responses. In many important processes, including those governing the cell-cycle or stem cell fate, cells must respond to noisy, possibly conflicting signals with all-or-none decisions.

Despite computational and experimental studies, however, the molecular networks that enable cells to convert graded signals into discrete decisions and to retain memory of these decisions remain poorly understood, due in part to assumptions about specific phenomena that are not universally present, and to the difficulty in decoupling precise contributions of the various components. Detailed insight into cellular decision-making can enable cell-based therapies, shed light on diseases arising out of dysregulation of control, and suggest practical design strategies for implementing this behavior in synthetic systems for research and industrial use.

In this thesis, we have employed three complementary approaches to further elucidate cellular decision-making. First, we performed a computational study that assessed the entire space of minimal networks to identify topologies that can not only make decisions but can do so robustly in the dynamic and noisy cellular environment. Second, via analysis of single-cell data, stochastic simulations, and quantitative dose-response experiments in a megakaryocyte erythroid progenitor (MEP) line, we developed a systems model that reconciles the competing stochastic and instructive theories of hematopoietic lineage commitment. Third, using a synthetic decision-making circuit, we assessed the impact of specific perturbations on the commitment trajectories of individual cells.

Chapter 2

Robust Network Topologies for Generating Switch-Like Cellular Responses

Adapted from Shah N.A., Sarkar C.A. (2011) Robust Network Topologies for Generating Switch-Like Cellular Responses. *PLoS Comput Biol* 7(6): e1002085.

> *Before turning to those moral and mental aspects of the matter which present the greatest difficulties, let the enquirer begin by mastering more elementary problems [...]. Puerile as such an exercise may seem, it sharpens the faculties of observation, and teaches one where to look and what to look for.*
>
> – Sherlock Holmes, *A Study in Scarlet*

2.1 Introduction

Signaling networks enable cells to process information from their surroundings by eliciting temporally and spatially precise responses to environmental cues. The complex and highly interconnected biomolecular interaction networks regulating signal

transmission establish connections between specific molecular effectors and hence delineate pathways through which extrinsic and intrinsic cues integrate to elicit cellular responses [Shen-Orr et al., 2002, Breitkreutz et al., 2010]. However, it is not always apparent what minimal signaling motif is both necessary and sufficient for robustly achieving a specific behavior.

A signaling network that converts a graded input cue into an all-or-none response is said to exhibit switch-like behavior; switching enables the establishment of discrete states which is vital in processes such as cell proliferation and differentiation [Ferrell, 2008, Malleshaiah et al., 2010, Xiong and Ferrell, 2003, Huang et al., 2007]. The term switching encompasses the more formal concepts of ultrasensitivity and bistability (Fig. 2-1), which are described in the previous chapter.

Previous studies have employed a combination of experiments and dynamical systems modeling to demonstrate the existence of ultrasensitivity and bistability in various signaling systems and have contributed to our knowledge of the types of network architectures that can give rise to switch-like behavior [Huang et al., 2007, Huang and Ferrell, 1996, Melen et al., 2005, Ramakrishnan and Bhalla, 2008, Bagowski et al., 2003, Voigt et al., 2005]. However, most studies have been restricted to a few, selected network topologies and have hence explored only a small fraction of the overall space of topologies that can exhibit switch-like behavior. More importantly, the proposed topologies are not necessarily parametrically robust in exhibiting switch-like behavior, since most studies do not account for the uncertain environmental context in which networks must function. Networks that exhibit switch-like behavior only in narrow regimes of the overall biologically relevant parameter space are of diminished utility in understanding natural systems due to intrinsic and extrinsic perturbations that result in changes in species concentrations and interactions with other effectors, which are constrained at both short and evolutionary timescales by the cost-benefit tradeoff for the cell.

An unbiased, comprehensive analysis of networks that robustly generate switch-like responses in living systems would expand our understanding of the types of

circuitry that enable cells to make binary decisions and assume discrete states, and hence may afford a mechanistic understanding of diseases arising out of a loss of control, such as cancer. Furthermore, such an analysis can be useful to synthetic biologists who seek to implement these behaviors as building blocks for engineering robust, complex biological programs.

Here, we simulated all possible two- and three-component networks on random parameter sets, and assessed the resulting response profiles for degree of ultrasensitivity and bistability. Our strategy is partly inspired by a recent analysis of enzymatic networks that enable adaptation in bacteria [Ma et al., 2009]; however, in addition to studying networks with only enzyme components, we expanded our focus to include purely transcriptional networks and hybrid enzymatic/transcriptional networks which enabled us to quantify robustness with respect to both the function of each protein component in the network as well as the interactions among the components.

Our results reveal that network architecture and composition can have a dramatic impact on robustness in generating switch-like behavior. Specifically, compared to other compositional classes studied, hybrid networks are more robust in yielding ultrasensitive and bistable responses. Detailed analysis of network topologies suggests that the zero-order effect arising out of a simple enzymatic activation/inactivation system is a prevalent mechanism for generating robust ultrasensitivity, and hence can act as a building block for switch-like behavior. A global view of network topologies suggests strong clustering into a small number of recurring motifs. Finally, comparison with data from previous studies of natural and synthetic systems demonstrates concordance between these computational results and experimental observations, and highlights the utility of our analysis both as a discovery tool for studying how switching can arise in natural systems and as a design tool for engineering switch-like behavior in synthetic circuits.

2.2 Results and Discussion

2.2.1 Topology search scheme

To enumerate the network architectures that can give rise to switch-like behavior, we considered all possible topologies of two or three components, and assessed them for robustness in generating ultrasensitive and bistable responses. Although switch-like behavior can arise in networks having more than three components, restricting our scope to minimal networks makes the analysis more tractable and the results simpler to interpret. Moreover, many large networks can be reduced to minimal models without significant loss in the spectrum of behaviors observed [Alon, 2007, Shen-Orr et al., 2002, Wolf, 2003].

An overview of the search scheme is illustrated in Fig. 2-2. Each network topology considered consists of an input component, A, an output component, C, and if present, an additional component, B. The input component A is modeled as a receptor that is activated upon binding of the stimulus, S. The output component C is modeled as a downstream effector, and the level of active C is considered the response of the system. Allowing each component to activate, inhibit, or have no impact on the other two components and itself yields 3^9 (19,683) distinct topologies. Within this set, approximately 3,700 topologies lack connections linking the input and output components, and are hence discarded.

Since activation and inhibition in biological systems can occur at both enzymatic and transcriptional levels, an important focus of this study is to compare the robustness in generating switch-like behavior arising out of enzyme and transcription components. Towards this goal, we studied four different categories of networks (Fig. 2-2D): enzyme-only, in which each component is modeled as an activating or inactivating enzyme (EEE); transcription-only, in which each component is modeled as a transcriptional activator or repressor (TTT); and two categories of hybrid networks, one with only C modeled as a transcriptional component (EET), and one with both

12

B and C modeled as transcription components (ETT). While switch-like behavior can arise in networks belonging to other compositional classes, our study is focused on networks that can be directionally described as outside-in signaling (i.e. networks that allow switch-like modulation of a downstream species, such as a master regulatory transcription factor that ushers in a phenotypic change, via an external stimulus).

In our analysis scheme, each component exists in an active form, which carries out the reactions specified by the network, and an inactive form, which only serves as substrate. Enzyme components act by catalyzing the inter-conversion of their targets. For instance, in the EET category, B is an enzyme, and an activation interaction from B to C denotes that B catalyzes the conversion of inactive C into active C; an inhibitory interaction would catalyze the opposite inter-conversion. Similarly, a positive interaction from B to C in the ETT category, in which B is a transcriptional component, denotes that B up-regulates the production of inactive C; an inhibitory interaction denotes B-mediated repression of the synthesis of C. Additionally, since enzymatic auto-regulation in signaling is not a common cellular behavior (e.g., there is a plethora of examples in which a kinase or phosphatase activates or inactivates another type of protein but not many instances in which an enzyme modifies its own species), only transcriptional components are allowed auto-regulatory loops, which reduces the number of network topologies considered for the EEE, EET, and ETT compositional classes. Irrespective of the topology, each component is modeled as being subject to basal synthesis and degradation and basal activation and inactivation by background components assumed to be constant.

A single network topology translates into a system of rate equations in which interactions among the three components are modeled using mass-action kinetics. Assignment of 10^3 random parameter sets to the kinetic constants of this model yields 10^3 different circuits having the same network architecture. Each circuit is simulated on a range of stimulus concentrations, and the resulting steady-state response information is assessed for switch-like behavior by two metrics: the Hill coefficient (n_H), representing the degree of ultrasensitivity [Goldbeter and Koshland, 1981], and the

relative drop in stimulus, or window (W) over which the system remains in the on state (Fig. 2-1). Hence, each network topology yields 10^3 steady-state response plots. Parametric robustness in generating switch-like behavior is quantified by robustness scores representing the percent of plots exhibiting strong ultrasensitivity ($n_H > 2$), and bistability ($W > 5$); for instance, a network that yields more ultrasensitive response profiles on random parameter sets than another is considered to be more robust in generating bistability. In addition to estimating n_H, response steepness was also analyzed by computing the maximum local response coefficient (see Methods) [Kholodenko et al., 1997]. Although both measures show good agreement (Fig. 2-8), since n_H establishes a lower-bound on the steepness, it was used as the primary metric in assessing ultrasensitivity robustness. Our results also demonstrate that simulating 10^3 random parameter sets for each network is sufficient for reliably estimating robustness scores (Fig. 2-9).

2.2.2 Network composition influences robustness in generating switch-like behavior

In outside-in signaling systems, binding of a ligand to a receptor initiates a signaling cascade typically resulting in the activation of downstream transcription factors which can in turn alter the expression program of the cell, thereby ushering in phenotypic change [Kisseleva et al., 2002, Ingham and McMahon, 2001, Hazzalin and Mahadevan, 2002]. Hence, in ligand-activated systems, the switch-like nature of a response is most prominent at the transcriptional level, as is the case for instance in cell differentiation during development [Melen et al., 2005]. However, the actual circuitry enabling switch-like behavior may itself lie further upstream, and may be composed of transcription as well as enzyme components, which have fundamentally different properties and hence generate switch-like behavior via distinct mechanisms.

To assess the extent to which network composition influences robustness in generating switch-like behavior, we performed a global analysis of all network topologies

across four compositional classes. Specifically, each network was simulated under the all-enzyme (EEE) compositional regime, and the resulting response profiles were used to compute a score quantifying the network's robustness in generating ultrasensitivity and bistability (as described above). The network was then re-simulated to obtain robustness scores under all-transcription (TTT) and hybrid (EET, ETT) regimes.

First, across all compositional classes, a significantly larger number of networks demonstrated ultrasensitive behavior than bistable behavior (Fig. 2-3A), in line with the observation in biological systems that bistability is typically accompanied by ultrasensitivity [Xiong and Ferrell, 2003], but ultrasensitivity can also arise in the absence of bistability [Goldbeter and Koshland, 1981, Melen et al., 2005]. Second, within a compositional class, a small proportion of networks exhibit switch-like behavior on a large percentage of random parameter sets. The highly skewed nature of robustness score distributions demonstrates that network architecture alone can impact robustness, and that a particular network's probability of generating switch-like behavior can be dramatically improved with rewiring, and without fine-tuning of kinetic constants such as those associated with binding or catalysis. Third, and most importantly, network composition strongly influences robustness in generating switch-like behavior. Compared to EEE and TTT classes, networks in the hybrid EET and ETT compositional classes yield ultrasensitive responses on a significantly larger proportion of parameter sets, with the most robust networks achieving ultrasensitivity robustness scores as high as 28% in contrast, maximum ultrasensitivity robustness scores in the EEE and TTT classes are 6% and 3%, respectively. For bistability, maximum robustness scores for the EET and ETT compositional classes are approximately 16% and 18%, respectively, while scores for EEE and TTT classes are significantly lower at 3% and 1%, respectively (Fig. 2-3A). Our findings demonstrate that a particular network topology can yield markedly different robustness scores under different compositional regimes, and suggest that minimal networks composed of an enzyme input component, a transcription output component, and an additional enzyme or transcription regulatory node may be optimal for generating switch-like

behavior.

2.2.3 Transcription-only networks are suboptimal in generating switching, even with transcriptional cooperativity

Comparison of network topologies across different compositional classes reveals the unexpected result that purely transcriptional networks are markedly less robust in generating switch-like behavior. Despite the considerably enlarged set of networks analyzed (only transcription components were allowed self-regulatory links, yielding more possible topologies) the most robust TTT networks achieved dramatically lower robustness scores than those achieved by the most robust networks in the optimal EET and ETT categories.

In our analysis scheme, a transcriptional activation interaction represents the binding of a single transcription factor to a regulatory site, and is hence modeled as a linear reaction. However, a large number of transcription factors bind to DNA as dimers, and transcription initiation can itself be inherently cooperative [Sneppen et al., 2008]; both characteristics can directly introduce nonlinearity into a system, and therefore boost the probability of generating switch-like behavior [Gardner et al., 2000, Angeli et al., 2004, Cherry and Adler, 2000]. To further investigate the impact of cooperativity arising out of multimerization and transcription initiation, we re-analyzed the entire set of networks in the TTT compositional class with all transcriptional interactions modeled as cooperative processes ($n_H = 2$). As expected, robustness scores for both ultrasensitivity and bistability were enhanced, with the most robust networks generating ultrasensitive responses on 4%, and bistable responses on 2%, of parameter sets (Fig. 2-3A, slashed bars). However, despite including transcriptional cooperativity only in the TTT class (and not EET or ETT), the best networks in all other classes are still more robust than any network in the $n_H = 2$ TTT class.

Our results suggest that, in terms of generating switch-like behavior, networks composed only of transcription components are inherently suboptimal relative to

16

hybrid or all-enzyme compositional classes.

2.2.4 Transcriptional feedback enhances switch-like behavior in hybrid networks

We now highlight some of the prevalent mechanisms contributing to the robustness differences between circuits in different compositional classes. In particular, we compare two network topologies in which a change in the identity of the output component C (i.e., either an enzyme or transcription component) leads to markedly different robustness scores for ultrasensitivity and bistability.

The network topology depicted in the left-hand column of Fig. 2-3B exhibits an ultrasensitive response on 2% of parameter sets in the EEE compositional context; however, when C is modeled as a transcription component, the robustness score for ultrasensitivity is dramatically higher, at 17%. Since A and B are modeled as enzymes under both EEE and EET regimes, the difference in robustness scores is entirely attributable to the feedback interaction from C to A, suggesting that transcriptional feedback enhances the probability of ultrasensitivity considerably more than activation feedback. To unravel the mechanisms contributing to the difference in robustness scores, we compared modules within this network to known models of ultrasensitivity.

We first examine the network that results when the feedback interaction from C to A is removed from the topology depicted in the left-hand column of Fig. 2-3B. Under both EEE and EET compositional classes, A acts as an enzyme activator for C, and B is effectively a background inactivator for both A and C (since there are no incoming links for B). When the total concentration (inactive and active) of C is much greater than those of active A and B, and the effective Michaelis constant ($K = \frac{k_1 + k_2}{k_0}$, see Methods) values for activation and inactivation interactions are sufficiently small, enzymes A and B operate in a zero-order regime, which in turn causes the system to exhibit ultrasensitive activation of C [Goldbeter and Koshland, 1981]. Furthermore,

transcriptional feedback from C to A can enhance existing ultrasensitivity or confer ultrasensitivity via an independent mechanism described in the next section.

Zero-order ultrasensitivity can also be generated or enhanced by transcriptional feedback merely via a concentration effect: feedback can significantly increase the amount of substrate, which may in turn enable the system to satisfy the conditions for zero-order ultrasensitivity. Hence, the presence of transcriptional feedback broadens the parameter sub-space in which the system yields an ultrasensitive response and boosts the overall probability of generating this behavior. Importantly, although the transcriptional feedback interaction does require minimal tuning to contribute to the overall robustness in generating ultrasensitivity, it does not hinder other mechanisms conferring this behavior.

Enzymatic activation feedback under the EEE compositional regime can give rise to strong ultrasensitivity [Ferrell, 2008]; however, in contrast to transcriptional feedback, activation feedback can also disrupt other interactions and thus narrow the parameter sub-space yielding ultrasensitive behavior. For instance, activation feedback can saturate active A (such that there are no more A molecules that can be converted into active A), thereby diminishing zero-order effects on 4C. Therefore, the network depicted in the left-hand column of Fig. 2-3B achieves a low robustness score, which changes marginally even when the feedback interaction is removed.

To understand mechanisms underlying differing robustness scores for bistability, we examined the network depicted in the right-hand column of Fig. 2-3B. This network generates a bistable response on 3% of parameter sets under the EEE compositional regime, and 8% when C is modeled as a transcription component (EET). This network contains two positive feedback interactions: between B and A, which is enzymatic under both EEE and EET regimes, and between C and A, which is transcriptional under EET and enzymatic under EEE. Removal of the feedback from C to A yields the same circuit under both EEE and EET, which achieves a robustness score of approximately 2%. In contrast, removal of the B to A feedback yields different circuits under EEE and EET, with robustness scores of 3% and 4%, respectively.

Hence, while either feedback is sufficient for conferring bistability to the overall system, their combination leads to a significant increase in robustness under EET, but not under EEE.

A simple two-enzyme dual-activation system can exhibit bistability under certain parameter regimes [Ferrell, 2008]. In the EEE class, the network depicted in the right-hand column of Fig. 2-3B can achieve bistability via two separate enzymatic feedbacks. However, each feedback produces more active A, and can saturate it such that the addition of the second feedback (onto the same target A) has a diminished effect since there is a limited quantity of inactive A that can be activated and hence does not significantly broaden the parameter space for bistable behavior. In contrast, under EET, transcriptional feedback to A produces more inactive A, and hence does not hinder the enzymatic feedback from B to A. Although linear transcriptional feedback alone cannot generate bistability [Ferrell and Xiong, 2001, Angeli et al., 2004], it can help confer this behavior in a network in which the activation interaction is independently ultrasensitive. Hence, under EET, the two feedbacks in the present network confer bistability via distinct mechanisms.

2.2.5 Ultrasensitivity via linear transcriptional feedback and degradation

Transcriptional feedback alone can give rise to modest ultrasensitivity via a mechanism distinct from zero-order ultrasensitivity. To investigate this phenomenon further, we separately modeled a simple system in which a transcription factor C, is activated by an enzyme A, and active C synthesizes more inactive C (Fig. 2-4A). C is synthesized and degraded via background processes, but unlike in our main topology search simulations, C is not subject to any inactivation process, which precludes the possibility of zero-order ultrasensitivity in any parameter regime. Parameter values for binding, dissociation, synthesis and degradation were varied and the resulting systems of ordinary differential equations were numerically integrated on a range of

stimulus concentrations (see Methods for full model details). The resulting curves
were then assessed for ultrasensitivity, and the results are summarized in Fig. 2-4B.

The results show that a simple transcriptional feedback system can generate re-
sponses with characteristic n_H as high as 2, under certain parameter regimes. Interest-
ingly, the extent of ultrasensitivity is independent of the explicit enzymatic binding,
dissociation, and catalysis parameters, and instead is dependent on two dimensionless
quantities. If the maximal feedback synthesis rate, v, is sufficiently greater than the
basal synthesis rate, b (i.e., when $\frac{v}{b} \gg 1$), then n_H reaches a maximum when the
effective feedback synthesis rate constant (where K_F is the concentration of active
C driving additional synthesis of inactive C at rate $\frac{v}{2}$) is approximately equal to
the degradation rate constant k_{deg} (i.e., when $\frac{v/K_F}{k_{deg}} \approx 1$). Hence, when feedback is
strong, proper balance of feedback synthesis and degradation is sufficient to generate
ultrasensitivity.

2.2.6   Minimal architectures for generating ultrasensitivity

Having used our unbiased approach to discover pervasive, yet simple, interactions that
augment the robustness of switch-like responses, we then took a design-centric view of
our results to understand how these interactions could be combined to yield topologies
exhibiting robust ultrasensitivity and bistability. Specifically, we focused on minimal
networks (i.e., networks generating robust switch-like behavior with fewer interactions
and components) for two main reasons. First, networks in biological systems arise via
an evolutionary process, and since there is a cost associated with maintaining each in-
teraction, natural selection is unlikely to maintain those interactions and components
that do not contribute significantly towards enabling a necessary behavior (i.e., do
not affect fitness). Second, minimal networks may suggest practical design strategies
for engineering switch-like behavior in synthetic systems.

To identify minimal networks generating robust switch-like behavior, networks
within each compositional class were ranked by the ultrasensitivity and bistability

robustness scores, and only the top 100 networks in each category were retained. Next, a pruning step was performed. Briefly, within a particular category, each network was compared to every other network to determine if a proper subnetwork of this network having a higher robustness existed, or if this network's robustness score was within 15% of the maximum robustness score. If either was true, the network with more connections was removed from the list. This procedure filtered networks with excessive interactions, and made it easier to identify families of networks. The most robust networks after the filtering step are presented in rank order in Fig. 2-10.

A global view of the resulting topologies (Fig. 2-10) reveals strong consensus patterns and suggests that the set of robust, minimal networks readily clusters into a small number of families. Comparison of ultrasensitive and bistable networks within and across compositional classes reveals that networks with more interactions do not consistently rank higher than sparser networks, indicating that specific mechanisms conferring switch-like behavior cannot necessarily be combined to yield more robust networks, due to the possibility of interference. Despite this, a few simple motifs are particularly prevalent within a given compositional class (e.g., A activating B, which in turn activates C under EEE) and even across compositional classes (e.g., A activating C, which upregulates A under EET and ETT), indicating that such robust motifs can act as modular building blocks for conferring switch-like behavior to a system. In addition, the pruning procedure strikingly reduces each set of the 100 most robust networks to less than 20 networks in all but one compositional class, indicating that the set of networks generating robust switch-like behavior constitutes a very small fraction of the overall network space; below we discuss how this subspace reduces even further to a few distinct mechanisms.

The simplest network considered in our analysis, a two component topology with a positive interaction from A to C, yields an ultrasensitivity robustness score of approximately 5% under the EET compositional regime (Fig. 2-5). The ultrasensitivity exhibited by this circuit is entirely attributable to zero-order effects arising from the enzymatic cycle of induced activation of A and background inactivation. The addi-

tion of a transcriptional interaction from C to A yields a robustness score of 17%; strikingly, the A-to-C-to-A motif is present in all of the 100 most robust circuits in the EET class. An additional auto-regulatory transcriptional interaction onto C instead yields a robustness score of 15%. The combination of both C-to-A and C-to-C feedbacks yields a particularly high robustness score of 26%, making the dual-feedback circuit the most robust in the EET class after filtering. Together, the two feedbacks introduce independent non-interfering mechanisms for generating ultrasensitivity and enhance the probability of zero-order effects in the activation of C via a concentration effect. Thus, our analysis suggests that a simple network with two transcriptional feedbacks is among the most optimal configurations for generating ultrasensitivity.

Although networks in the all-enzyme EEE class yield significantly lower robustness scores, it is worth noting that the pruning procedure drastically trims the list of the 100 most robust networks in the EEE category to three very simple networks (Fig. 2-10). The most robust network, A activating B, which in turn activates C, represents a basic enzyme activation cascade. In the A-to-B-to-C network, ultrasensitivity can arise via two distinct mechanisms. First, the activation of B by A can be ultrasensitive if both A and the background inactivator for B behave in a zero-order manner. The ultrasensitivity can be further enhanced if the activation of C by B is similarly configured. Second, even in the absence of inactivating enzymes (and hence without zero-order effects), this cascade architecture itself can generate ultrasensitivity *de novo* [O'Shaughnessy et al., 2011].

2.2.7 Minimal architectures for generating bistability

Examination of the most robust bistable networks in the ETT category (Fig. 2-6) reveals that although there is no obvious minimal motif conferring bistability, there is a clear bias towards multiple positive transcriptional feedback interactions. However, positive transcriptional feedback alone cannot confer bistability to a system, a point that is affirmed by the observation that the most robust networks in the

22

transcriptional-only TTT category yield drastically lower scores. Closer inspection of the most robust networks reveals that in all of the top 100 networks, A activates C, which upregulates A. This simple hybrid motif of enzymatic activation and transcriptional feedback can yield bistability only if the activation step is independently ultrasensitive. In the space of networks considered in our analysis, bistability can arise via enzymatic activation and transcriptional feedback if the activation of C by A is ultrasensitive due to either zero-order effects or transcriptional autoregulation of C. Under ETT, bistability can also arise due to analogous interactions between A and B.

Importantly, our results also suggest that adding multiple instances of the enzymatic activation and transcriptional feedback motif to a single system does not hinder existing interactions, and can hence boost the probability of exhibiting a bistable response. In contrast, mechanisms such as cross-antagonism do appear in our analysis but are not highly ranked because of their stringent balancing requirements and fragility to interference by other interactions. For instance, in the two-component ETT network in which A activates C, and C upregulates A and itself, around 15% of the parameter sets yield ultrasensitivity but not bistability. To further explore the impact of combining motifs, we duplicated the dual transcriptional feedback motif in the same network by adding analogous interactions between A and B, and simulated the expanded network on the parameter sets that yielded ultrasensitivity but not bistability for the single motif network (parameter values for the added A-B, B-B, and B-A interactions were set to be the same as those for the A-C, C-C, and C-A interactions, respectively). We found that the expanded network with the duplicated motif converted more than 80% of previously ultrasensitive-only responses into strongly bistable responses. Since B and C are not directly connected in the expanded topology, the enhanced robustness can be attributed to increased nonlinearity in the activation response of A. Introduction of additional upregulation interactions from B to C, and C to B, further boosts the overall robustness score from 13% to 18%; this dual upregulation motif can confer bistability to circuits that exhibit only ultra-

23

sensitivity. While it is difficult to ascertain the exact contribution of each interaction in generating bistability as the network connectivity increases, our results point to the overarching principle that layering transcriptional feedback on an independently ultrasensitive activation interaction can act as a reusable building block for conferring bistability.

A noteworthy point about our results is that the robustness scores are bounded due in part to circuits which are otherwise bistable, but yield responses in which the ratio of maximum response to baseline response is low; this can arise in circuits with multiple positive feedbacks, for which basal activation alone is sufficient to switch the system into the on state. However, since our study is primarily focused on networks that can be modulated via an external stimulus, only responses that exhibit $\geq$ 10-fold increase in active C were considered.

2.2.8   Comparison with networks in biological systems

Network families suggested by our analysis exhibit strong resemblance to circuits that have been previously shown to exhibit switch-like behavior in natural systems, and here we discuss a few striking examples of simple, elegant circuits that robustly regulate critical cellular decision-making.

The Drosophila protein Yan is a transcriptional repressor that inhibits differentiation; specifically, in the embryo, ultrasensitivity in Yan phosphorylation enforces a sharp boundary separating developmental domains [Lai, 1992]. Binding of the ligand Spitz to the epidermal growth factor receptor (EGFR) leads to the graded activation of the mitogen-activated protein kinase (MAPK) pathway, and eventually results in the phosphorylation of Yan; Yan dephosphorylation can occur via a separate phosphatase (Fig. 2-7A) [Melen et al., 2005, O'Neill et al., 1994]. Phosphorylation of Yan makes it a target for degradation and thus promotes differentiation. Systematic perturbation of the network demonstrated that its robust ultrasensitivity is attributable to zero-order effects arising from the high levels of Yan relative to the concentrations

of the kinase and phosphatase acting on this substrate [Melen et al., 2005].

MAPK pathways include a core, three-step cascade, and comprise an evolutionarily conserved family that enables eukaryotic cells to respond to a diverse array of signals [Markevich et al., 2004, Chang and Karin, 2001]. Ultrasensitivity has been observed in MAPK cascades in several organisms, most notably in *Xenopus* (Fig. 2-6). Immature *Xenopus* oocytes can be induced into maturation by treatment with the hormone progesterone, which acts via the MAPK signaling cascade: binding of progesterone to its receptor leads to the accumulation of active Mos, which activates MEK, which in turn activates ERK2 (also known as p42 MAPK). Active ERK2 can then activate cyclin B-CDK1 complexes which bring about entry into M-phase, leading to maturation. The three-tier cascade of Mos, MEK, and ERK2 has been demonstrated to exhibit ultrasensitive activation of ERK2 [Ferrell Jr., 1998, Pomerening, 2008]. The architecture of this cascade is essentially the same as the topology in the EEE class that ranks first in terms of robustness in generating ultrasensitivity in our analysis. Although ultrasensitivity in MAPK network can arise via several mechanisms, including zero-order effects and multi-site activation, the cascading architecture itself can amplify existing ultrasensitivity [Brown et al., 1997] and even generate ultrasensitivity where none exists [O'Shaughnessy et al., 2011].

The ERK2 response to progesterone treatment is also bistable. Immature oocytes treated with progesterone proceed to maturation even after progesterone is subsequently removed from the environment. The bistability observed in this system is attributed to a positive feedback from ERK2 that leads to increased synthesis of Mos [Xiong and Ferrell, 2003]. Cdc2, another major driver of oocyte maturation, is involved in a positive feedback loop with Cdc25, and is also connected to the ERK2 system via mutual positive feedback interactions [Xiong and Ferrell, 2003]. While important differences exist, the oocyte maturation system architecturally resembles the family of most robustly bistable topologies in the ETT class, which can yield ultrasensitive activation of B and C via zero-order effects or transcriptional feedback. Robust bistability can be generated by layering positive feedback onto ultrasensitive activa-

tion motifs, with additional minor gains in robustness achieved with positive crosstalk between ultrasensitive nodes (i.e., B and C). Similarly, the oocyte maturation system can generate ultrasensitive activation via cascading and other mechanisms, with robust bistability being achieved by multiple positive feedback interactions.

Another example is the network linking the Erythropoietin receptor (EpoR) to the transcription factor GATA1 (Fig. 2-7B); it exhibits strong ultrasensitivity and helps confer bistability to the circuit regulating commitment to the erythrocyte lineage [Palani and Sarkar, 2008]. Briefly, the binding of the cytokine Erythropoietin (Epo) to EpoR triggers the activation of GATA1, which in turn leads to the initiation of a transcriptional program for erythropoiesis. This circuit contains two feedback loops, with GATA1 transcriptionally up-regulating both EpoR and itself; the EpoR-GATA1 architecture is essentially the same as that depicted in Fig. 2-5 and described in the previous section; it ranks first in robustness (26%) in generating ultrasensitivity and also exhibits strong bistability (13% robustness).

2.2.9 Step-wise dissection of a synthetic circuit

Networks achieving high robustness scores for ultrasensitivity and bistability have increased probabilities of exhibiting switch-like behavior in multiple biological systems and contexts. Although properties of components and the encompassing environment can constrain the effective parameter space and hence alter the ranking, a global analysis of topologies that can generate a desired behavior can help eliminate poor design choices and accelerate the implementation of synthetic circuits. We now highlight a few relevant findings from a separate study by our group which focused on the construction of a circuit exhibiting strong switch-like behavior [Palani and Sarkar, 2011], and we discuss how the topology search method can serve as an effective design tool for synthetic biology.

The synthetic *Saccharomyces cerevisiae* circuit depicted in Fig. 2-7C consists of the heterologously expressed *Arabidopsis thaliana* receptor CRE1 (AtCRE1), the

endogenous SKN7 transcription factor, and GFP as a reporter, and is topologically the same as the ones presented in Fig. 2-5 and 2-7B. Binding of the cytokinin isopentenyladenine (IP) to yeast-expressed AtCRE1 has previously been shown to activate endogenous SKN7 [Inoue et al., 2001, Chen and Weiss, 2005]. In our circuit, active SKN7 was synthetically wired to up-regulate the transcription of itself, AtCRE1, and the reporter GFP. To assess the contributions of specific topological connections in generating ultrasensitivity with respect to IP stimulus, the circuit was implemented in yeast with and without the feedback interactions. In the absence of feedback, the underlying circuit exhibits weak ultrasensitivity ($n_H \approx 2$). Addition of receptor feedback does not impact ultrasensitivity regardless of promoter strength; since the total concentration of SKN7 is low, initial activation saturates active SKN7 levels before the feedback interaction can take effect. Autoregulation of SKN7 alone does non-trivially augment the ultrasensitivity ($n_H \approx 4$); this enhancement arising from the increased concentration of SKN7 can be attributed to the non-linearity introduced by autoregulation (Fig. 2-4) and possibly to more pronounced zero-order effects if endogenous enzymes inactivate this transcription factor (Fig. 2-7A). The complete circuit with both feedback interactions exhibits extremely strong ultrasensitivity ($n_H \approx 20$) and reasonable bistability ($W \approx 2$-3) in response to IP, which is in agreement with our predictions.

The primary objective of this study was to obtain a high-level architectural view of the network topologies yielding robust ultrasensitivity and bistability. To keep the simulations and subsequent analyses tractable, we employed simplifying assumptions which may affect interpretation of our results. First, for protein synthesis, transcription and translation processes were lumped into a single expression which may mask additional dynamics in the case of long-lived mRNA. Second, in our analysis scheme, transcriptional components upregulate the inactive form of their target species, and we find that this type of interaction alone in the TTT class is far less robust in yielding switch-like behavior; however, in some biological systems, transcription factors can effectively act as enzymes by interacting with other co-activators and co-repressors,

and this can increase their ability to yield switch-like behavior. Third, we used simple thresholds for identifying responses as ultrasensitive ($n_H > 2$) and bistable ($W > 5$), and did not focus on the extent of ultrasensitivity or bistability, which may be important in certain biological contexts; however, our general conclusions are not dependent on these specific filtering thresholds.

In conclusion, our analysis shows that although a large number of network topologies exhibit switch-like behavior, only a small fraction of the topologies can be expected to yield ultrasensitive and bistable responses in the context of a noisy and evolving environment. Network motifs generating robust ultrasensitive and bistable responses can help identify circuits with such properties in natural systems and can also suggest design strategies for synthetic implementation of switching behavior.

2.3 Methods

2.3.1 Network construction and modeling

The overall topology search scheme is based in part on a previously described method [Ma et al., 2009]. All possible two- and three-component topologies were constructed, with stimulus and active C considered the input and the response, respectively, for steady-state characterization (Fig. 2-1); networks lacking reachability from A to C were discarded. Depending on the compositional class analyzed, network components (A, B, C) were modeled as either enzymes or transcription factors. All components exist in two forms, inactive and active, which can be either free or bound to another species as part of a complex. Only active forms, denoted with an asterisk, carry out reactions. All species are subject to basal synthesis and degradation, as well as activation and inactivation by background components. For instance, accounting for background reactions leads to the following rate equations for C and $C*$:

$$\frac{dC}{dt} = b_{syn} - k_{deg}C - k_P P \frac{C}{C+K_P} + k_Q Q \frac{C*}{C*+K_Q} + \dots$$
$$\frac{dC^*}{dt} = -k_{deg}C* + k_P P \frac{C}{C+K_P} - k_Q Q \frac{C*}{C*+K_Q} + \dots$$

where P and Q are the background activating and inactivating enzymes, respectively. Enzymatic interactions among main species were modeled using mass-action kinetics; for instance, here active enzyme $B*$ binds to inactive C, forming a complex, W, which can either dissociate or catalyze the activation of C into $C*$:

$$B* + C \xrightarrow{k_{0,B*C}} W$$

$$W \xrightarrow{k_{1,B*C}} B* + C$$
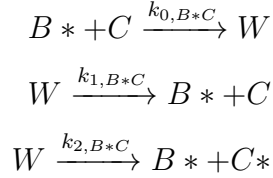
$$W \xrightarrow{k_{2,B*C}} B* + C*$$

This set of interactions, modeled explicitly by law of mass action, yields the following terms in the relevant rate equations:

$$\frac{dC}{dt} = -k_{0,B*C}B*C + k_{1,B*C}W + \dots$$
$$\frac{dW}{dt} = k_{0,B*C}B*C - k_{1,B*C}W - k_{2,B*C}W + \dots$$
$$\frac{dC*}{dt} = k_{2,B*C}W + \dots$$
$$\frac{dB*}{dt} = -k_{0,B*C}B*C + k_{1,B*C}W + k_{2,B*C}W + \dots$$

Inactivation interactions are handled similarly, except that the intermediate complex consists of two active species; for instance, $B*$ can inactivate $C*$ by binding to it and releasing C after catalysis. (For this set of reactions describing the activation of C into $C*$, the effective Michaelis constant is $K = \frac{k_1+k_2}{k_0}$.)

29

The stimulus for the system, S, binds to the receptor, A, in the form of a ligand:

$$\frac{dA}{dt} = -k_{0,SA}SA + k_{1,SA}A * \ldots$$

$$\frac{dA*}{dt} = k_{0,SA}SA - k_{1,SA}A * \ldots$$

The interaction between A and S is in addition to any interactions between A and other components, and background processes that act on all components, modeled by terms analogous to the ones depicted in equations above. Collectively, interactions involving A represent two distinct biological mechanisms. The ligand-mediated activation of A represents a phosphorylation or other modification event immediately downstream; such a modification can also occur without involvement of the ligand, in which case this biological mechanism is modeled using enzymatic reactions.

Transcriptional interactions result in the upregulation of the inactive form of the target component; for instance, here active transcription factor $B*$ upregulates inactive C:

$$\frac{dC}{dt} = v_{BC}\frac{(B*)^{n_H}}{(B*)^{n_H} + (K_{syn,BC})^{n_H}} + \ldots$$

A transcriptional Hill coefficient value of $n_H = 1$ was used for all simulations, except for the re-simulation of circuits in the TTT class where $n_H = 2$ was used, as described. Transcriptional inhibition is modeled as a competitive inhibition interaction; for instance, here $A*$ inhibits the upregulation of C by $B*$:

$$\frac{dC}{dt} = v_{BC} \frac{(B*)^{n_H}}{(B*)^{n_H} + (K_{syn,BC}(1 + A*))^{n_H}} + \dots$$

A scheme similar to Latin hypercube sampling [Iman et al., 1980] was used to generate 10^3 random parameter sets, with non-dimensionalized interaction parameter values (details given in Table 2.1) selected at uniform intervals on a logarithmic scale: k_0 $(10^2, 10^3)$; k_1 $(10^0, 10^4)$; k_2, k_P, k_Q $(10^1, 10^5)$; K_{syn} $(10^1, 10^1)$; K_P, K_Q $(10^3, 10^1)$; v $(10^3, 10^1)$. Application of parameter sets yielded 10^3 circuits for each network. Except where noted, the following parameters were held constant: $b_{syn} = 0.01$, $k_{deg} = 0.01$, $P = 0.01$, $Q = 0.1$.

2.3.2 Simulation and assessment of switch-like behavior

Each nave circuit was simulated to steady-state on a range of stimulus concentrations; levels of A, B, and C at the highest stimulus concentration were recorded and used as initial levels in another round of simulations to assess bistability. For ultrasensitivity, the stimulus levels at which the output reaches 10% and 90% were used to estimate n_H (Fig. 2-1A) [Goldbeter and Koshland, 1981] and the following formula was used to estimate the maximum local response coefficient [Kholodenko et al., 1997]:

$$max(\tfrac{dlnC*}{dlnS})$$

The forward and backward response profiles were used to estimate W (Fig. 2-1B); to be considered part of the bistable window of a response, the ratio of active $C*$ in the forward and backward solves at a particular stimulus concentration had to be at least 5 and the difference had to be greater than 0.1. Activation responses not positively correlated with the stimulus or exhibiting less than a ten-fold increase

from basal levels were not assessed for ultrasensitivity or bistability.

2.3.3   Transcriptional feedback model

The separate transcriptional feedback system described in the text and presented in Fig. 2-4 was modeled as follows. A is an enzyme that catalyzes the conversion of C into $C*$, with the complex Y as an intermediate species. All species, C, Y, and $C*$ are subject to first-order degradation. However, there is no inactivating enzyme, and hence zero-order ultrasensitivity cannot arise.

$$A + C \xrightarrow{k_0} Y$$

$$Y \xrightarrow{k_1} A + C$$

$$Y \xrightarrow{k_2} A + C*$$
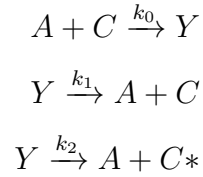
$$\frac{dC}{dt} = b - k_{deg}C - k_0 CA + k_1 Y + \frac{vC*}{K_F + C*}$$

$$\frac{dC*}{dt} = -k_{deg}C* + k_2 Y$$

$$\frac{dY}{dt} = -k_{deg}Y + k_0 CA - k_1 Y - k_2 Y$$
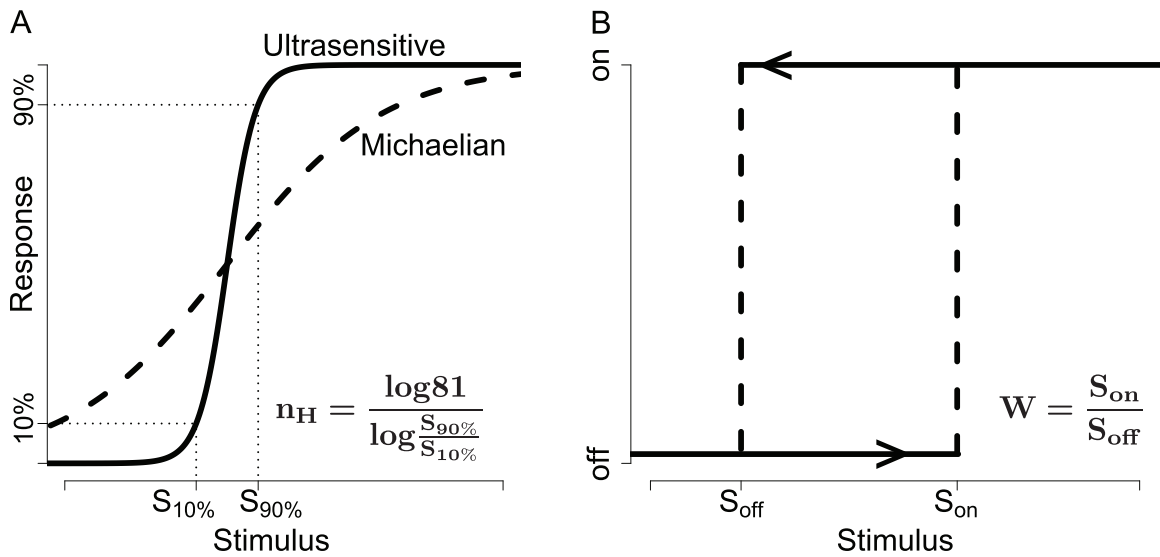
Figure 2-1: **Switch-like behavior.** **A.** A typical Michaelian system $(n_H = 1)$ requires an 81-fold increase in stimulus to increase the response from 10% to 90% of the maximum (i.e., $\frac{S_{90\%}}{S_{10\%}} = 81$) while an ultrasensitive response is more abrupt. **B.** Once triggered into the high, or 'on', state $(S > S_{on})$, a bistable system stays in that state even as the stimulus concentration is decreased, only switching 'off' below a lower threshold stimulus concentration (S_{off}, which is < 0 for irreversible systems).
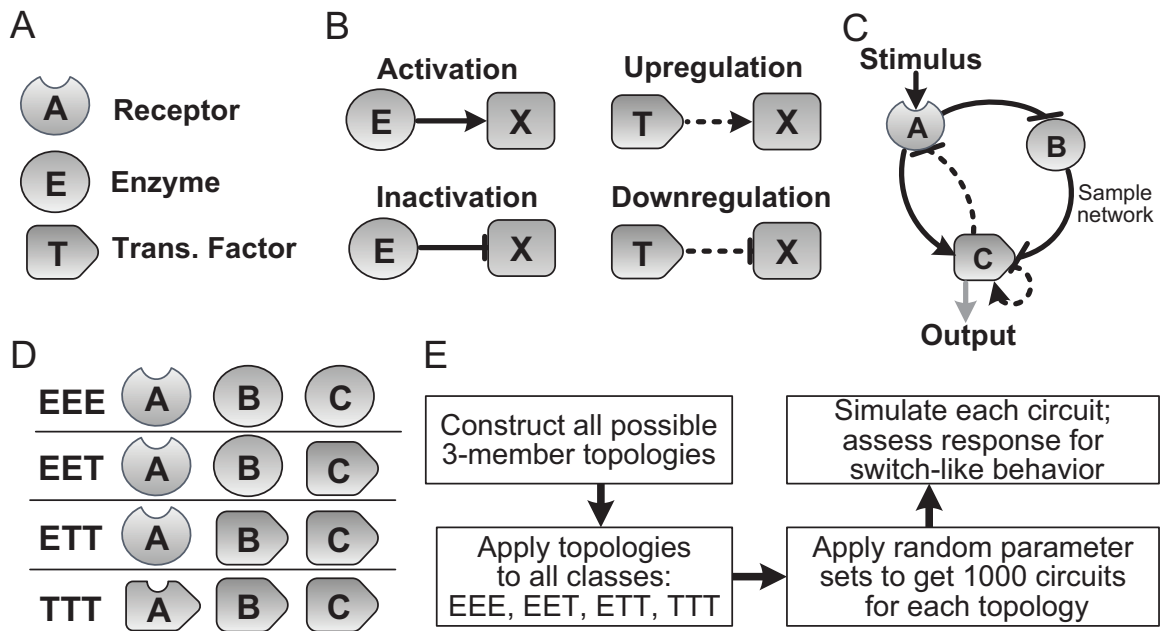
Figure 2-2: **Topology search scheme. A.** Each component is modeled as an enzyme or transcription factor. The input component A is modeled as a receptor to which the stimulus binds. **B.** Enzymatic components can catalyze the activation or inactivation of their targets, denoted as X. Transcriptional components can upregulate or inhibit the synthesis of the inactive forms of their targets. **C.** Sample network illustrating all possible interaction types. **D.** Four compositional classes were studied: EEE, in which A, B, C, are modeled as enzymes; TTT, in which each component is a transcription factor; and hybrid networks, in which only C is a transcription factor (EET) or both B and C are transcription factors (ETT). **E.** Overview of the topology search algorithm.
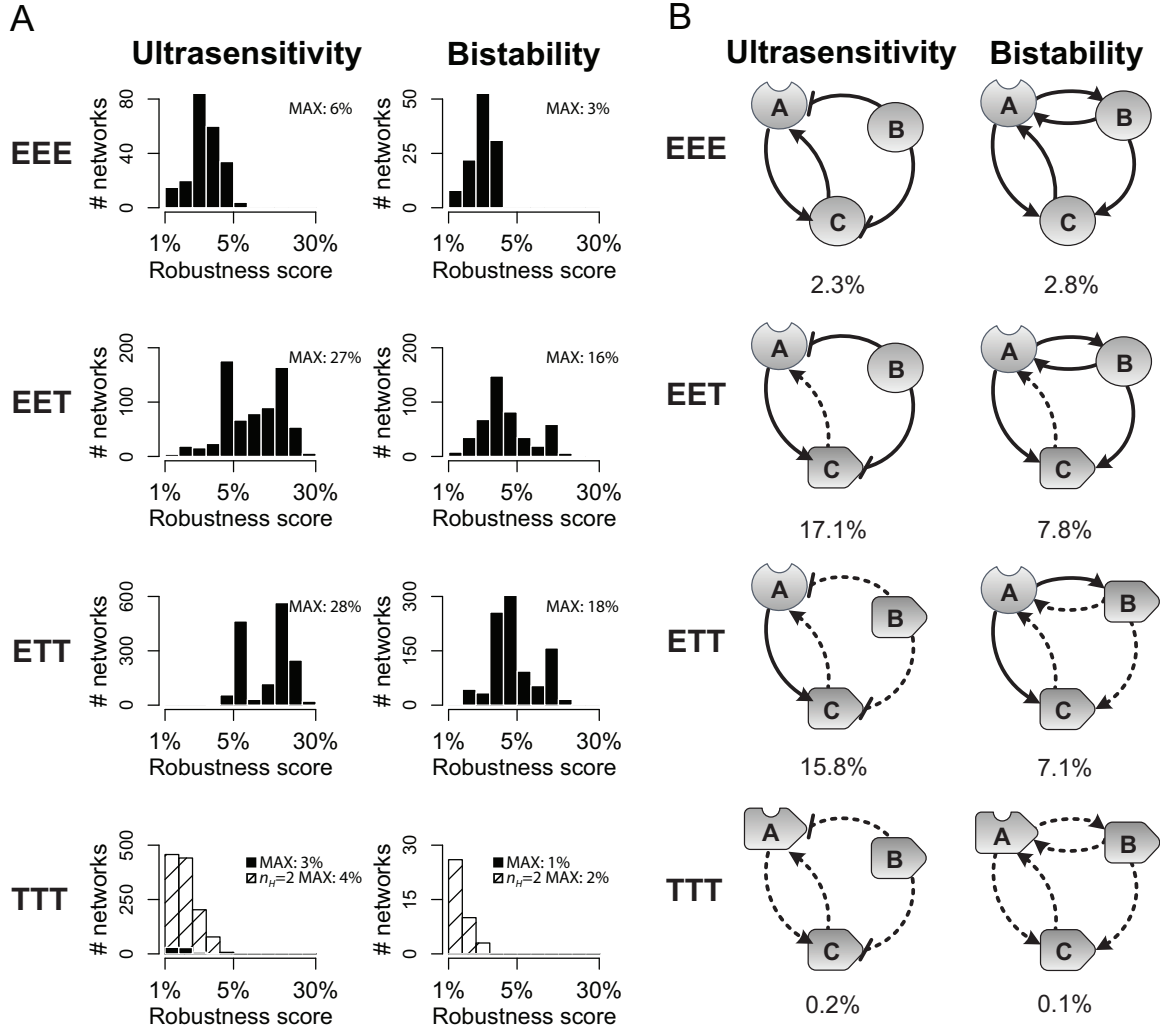
Figure 2-3: **Robustness in switch-like behavior across compositional classes.**
A. All possible network topologies were constructed and simulated; response profiles
were used to compute robustness scores for ultrasensitivity and bistability for each
network topology. This process was repeated for each compositional class. Histograms
depict the distribution of robustness scores for ultrasensitivity and bistability greater
than 1% across all compositional classes; white bars with oblique lines in the TTT
plots depict the distribution of robustness scores when each transcriptional interaction
is modeled as being cooperative ($n_H = 2$). Histograms represent ultrasensitivity
robustness scores for EEE (226 networks), EET (699), ETT (1511), TTT (84), TTT
$n_H = 2$ (1360) and bistability robustness scores for EEE (119 networks), EET (468),
ETT (972), TTT (0), TTT $n_H = 2$ (43). Networks achieving the highest robustness
scores belong to the hybrid classes: the most robust networks in the ETT class
achieve the highest scores for both ultrasensitivity and bistability, and the most robust
networks in EET achieve comparably high scores. **B.** Ultrasensitivity and bistability
robustness scores for two example topologies under different compositional classes;
the same network topology can yield dramatically different robustness scores under
different compositional classes.
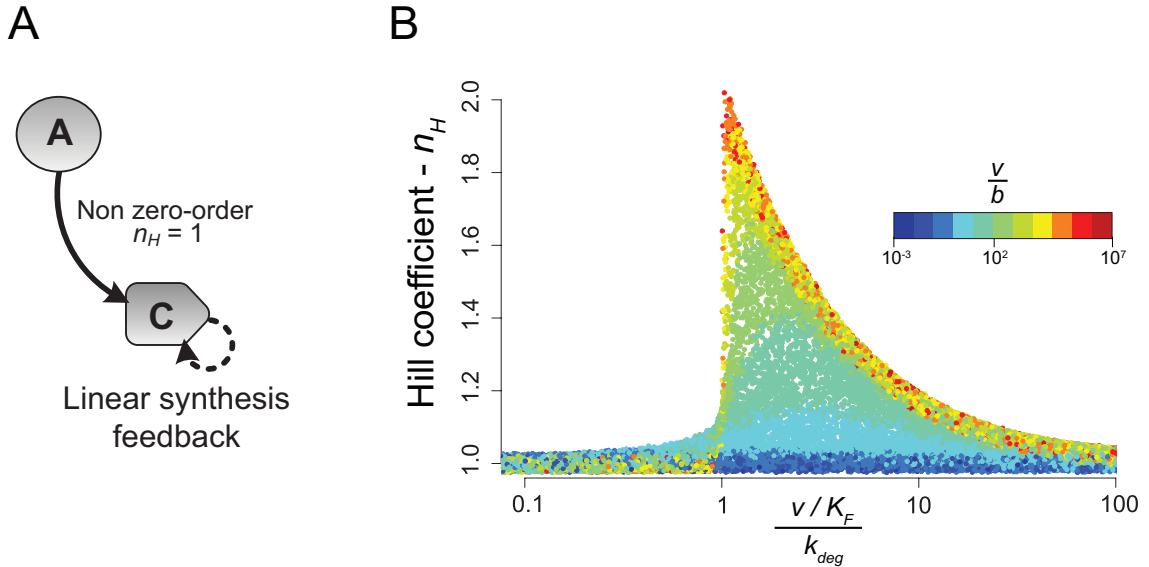
Figure 2-4: **Ultrasensitivity via linear transcriptional feedback and degradation.** A simple linear transcriptional feedback system can give rise to ultrasensitivity even in the absence of an inactivating enzyme. Note that this figure pertains to simulations on a minimal model different from the setup used for the topology search simulations (see Methods). **A.** In this system, the transcription factor C is activated by an enzyme, A. C is subject to basal synthesis and first-order degradation, but not to inactivation. **B.** The model was simulated on 10^6 random parameter sets, and a random subset of the results was plotted. Each dot represents a separate simulation on a random parameter set, and the color of the dot denotes the value of the dimensionless ratio in that parameter set (where b is the basal synthesis rate and v is the maximal feedback synthesis rate). If $\frac{v}{b}$ is sufficiently high, then the Hill coefficient reaches a maximum when the effective feedback synthesis rate constant $\frac{v}{K_F}$ (where K_F is the threshold concentration) is approximately equal to the degradation rate constant k_{deg}.
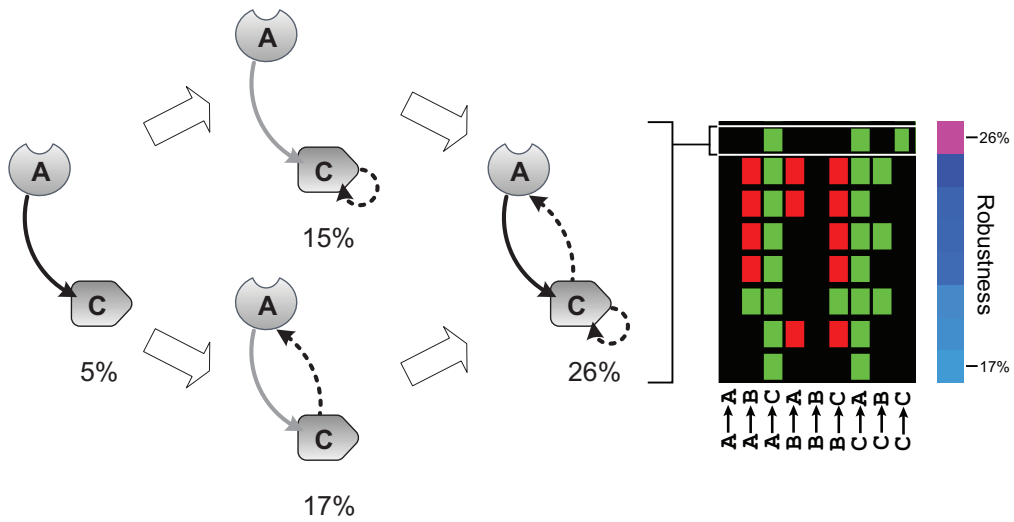
Figure 2-5: **Minimal architecture for generating robust ultrasensitivity** Starting with a simple network, incremental addition of specific interactions significantly improves robustness in generating ultrasensitivity. The map to the right lists the eight most robust network topologies generating ultrasensitivity in the EET class, after pruning; positive, negative, and no interactions are depicted with green, red, and black, respectively.
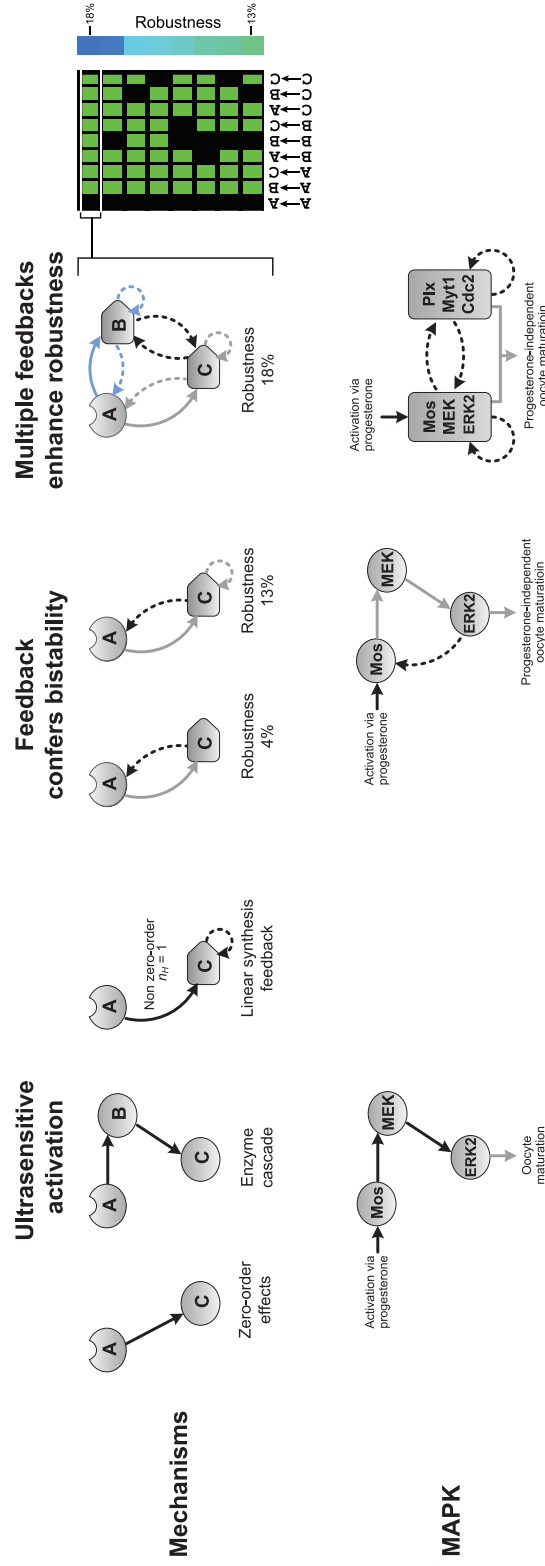
Figure 2-6: **Coupling of ultrasensitive activation and positive synthesis feedback yields robust bistability.** The upper row depicts molecular mechanisms derived from simulation results and the lower row depicts concordant examples in oocyte maturation. In our simulations, ultrasensitivity can arise via zero-order effects, enzyme cascading, and linear synthesis feedback. These motifs can yield bistability when coupled with positive synthesis feedback, and multiple feedbacks contribute to the robustness of this bistability. The map to the right lists the eight most robust network topologies generating bistability in the ETT class.
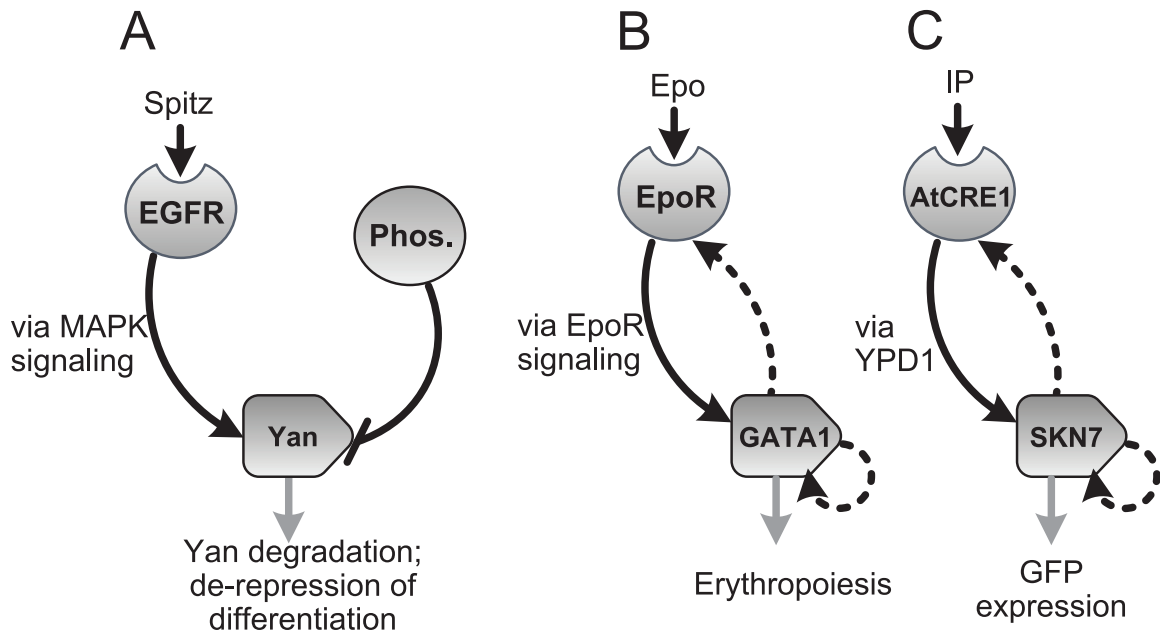
Figure 2-7: **Comparison with natural and synthetic systems. A.** Yan is a critical regulator of differentiation pathways in development, and generates ultrasensitivity via zero-order effects. **B.** The EpoR/GATA1 receptor/transcription factor pair can generate ultrasensity critical to the regulation of commitment to the erythrocyte lineage; this network is architecturally the same as the highest ranking network depicted in (Fig. 2-5). **C.** The synthetic AtCRE1/SKN7 hybrid network depicted exhibits robust switch-like behavior in yeast.
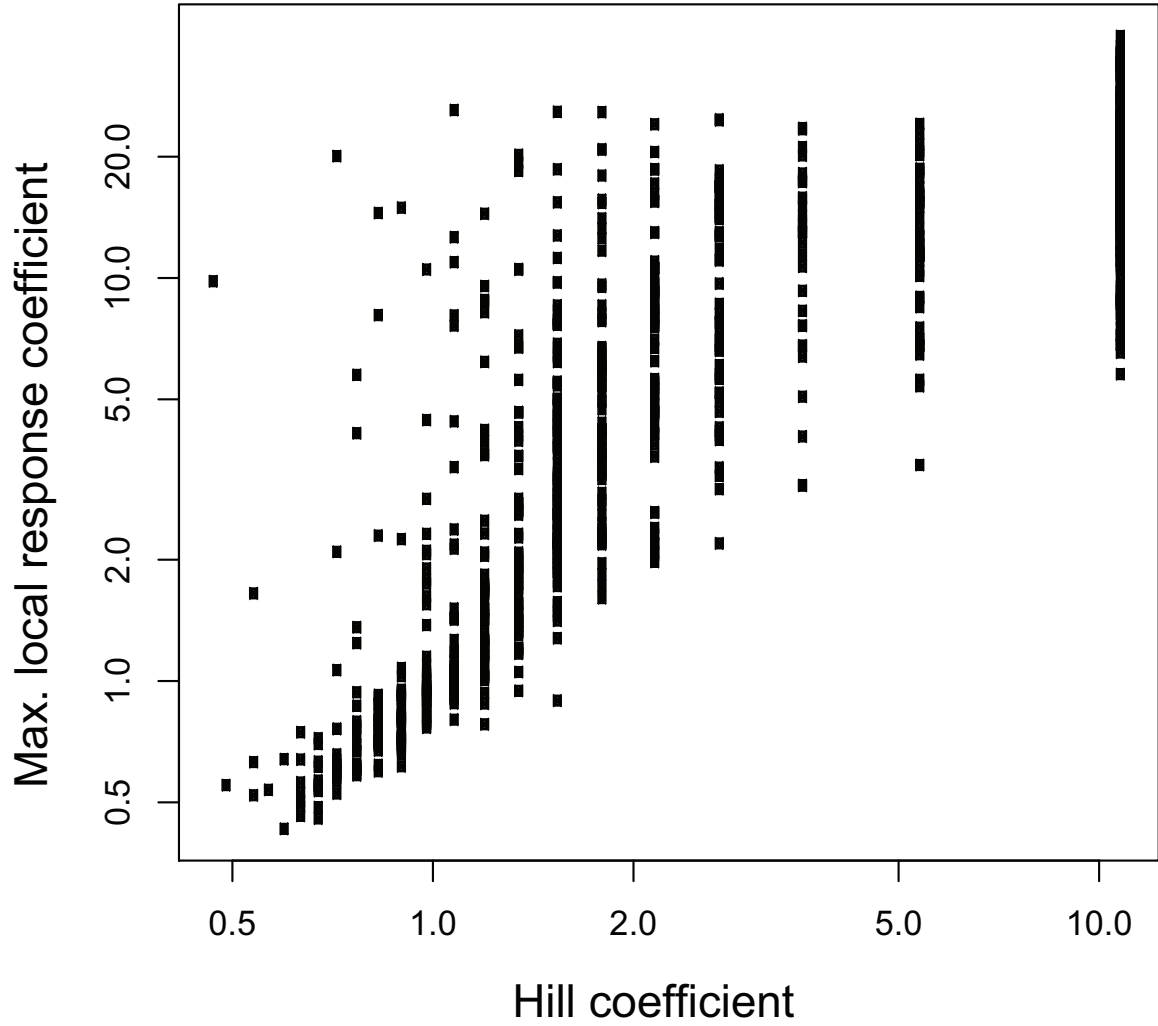
Figure 2-8: **Maximum local response coefficient correlates with estimated Hill coefficient** n_H**.** In addition to estimating n_H, the maximum local response coefficient was also computed for each network. This plot shows how the two metrics compare for all simulations of the double-feedback network topology under EET depicted in Fig. 2-5.
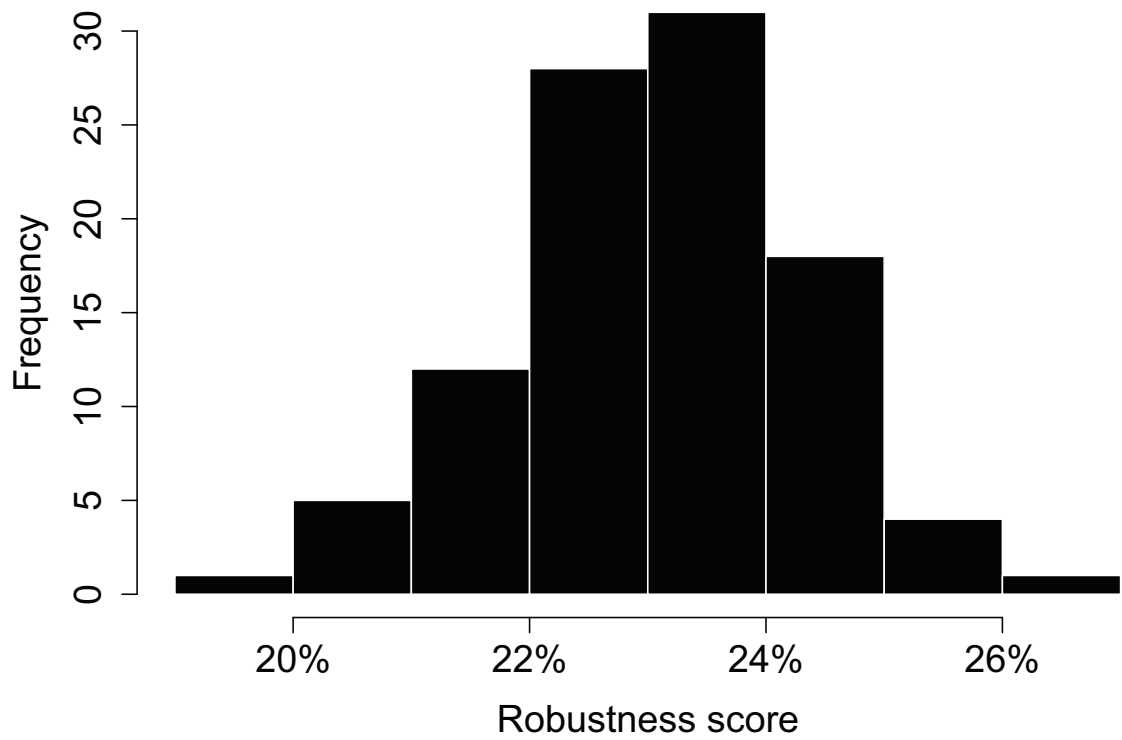
Figure 2-9: Robustness scores converge in 10^3 simulated parameter sets. The double-feedback network depicted in Fig. 2-5 was simulated on 10^3 parameter sets, for 100 runs. The histogram shows the distribution of robustness scores obtained.
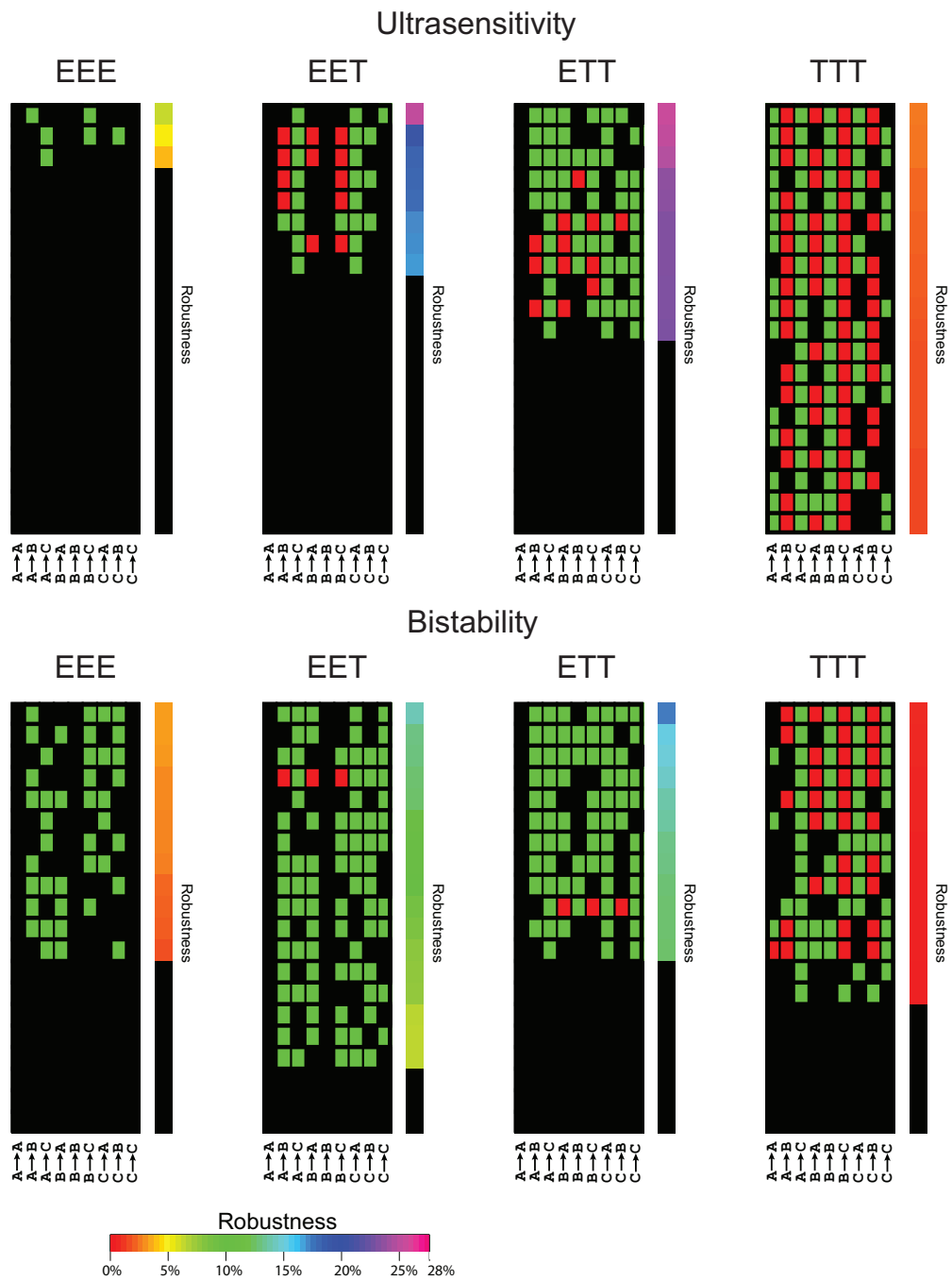
Figure 2-10: Networks within each compositional class were ranked by ultrasensitivity and bistability robustness scores. Only network topologies ranking in the top 100 robust networks in their respective compositional classes were included. Networks with additional, non-contributing interactions were filtered from the list as described in the main text.

Table 2.1: **Parameter ranges and non-dimensionalization.**

Parameter	Description	Biological range		Scaling parameter			Non-dimensional
b_{syn}	basal synthesis rate	10^{-12}	Ms^{-1}	$\alpha\beta$	10^{10}	$M^{-1}s$	0.01
k_{deg}	basal degradation rate constant	10^{-5}	s^{-1}	α	10^{3}	s	0.01
K_P	basal activation Michaelis constant	$10^{-10}\ldots10^{-6}$	M	β	10^{7}	M^{-1}	$10^{-3}\ldots10^{1}$
k_P	basal activation catalytic rate constant	$10^{-2}\ldots10^{2}$	s^{-1}	α	10^{3}	s	$10^{1}\ldots10^{5}$
K_Q	basal inactivation Michaelis constant	$10^{-10}\ldots10^{-6}$	M	β	10^{7}	M^{-1}	$10^{-3}\ldots10^{1}$
k_Q	basal inactivation catalytic rate constant	$10^{-2}\ldots10^{2}$	s^{-1}	α	10^{3}	s	$10^{1}\ldots10^{5}$
k_0	complex association rate constant	$10^{6}\ldots10^{7}$	$M^{-1}s^{-1}$	$\frac{\alpha}{\beta}$	10^{-4}	Ms	$10^{2}\ldots10^{3}$
k_1	complex dissociation rate constant	$10^{-3}\ldots10^{1}$	s^{-1}	α	10^{3}	s	$10^{0}\ldots10^{4}$
k_2	catalytic rate constant	$10^{-2}\ldots10^{2}$	s^{-1}	α	10^{3}	s	$10^{1}\ldots10^{5}$
v	maximal transcription rate	$10^{-13}\ldots10^{-9}$	Ms^{-1}	$\alpha\beta$	10^{10}	$M^{-1}s$	$10^{-3}\ldots10^{1}$
K_{syn}	conc. at which transcription rate is $\frac{v}{2}$	$10^{-8}\ldots10^{-6}$	M	β	10^{7}	M^{-1}	$10^{-1}\ldots10^{1}$
P	basal activator concentration	10^{-9}	M	β	10^{7}	M^{-1}	0.01
Q	basal inactivator concentration	10^{-8}	M	β	10^{7}	M^{-1}	0.1

Chapter 3

Conflict resolution in megakaryocyte-erythroid progenitor commitment

He remarks that, while the individual man is an insoluble puzzle, in the aggregate he becomes a mathematical certainty. You can, for example, never foretell what any one man will do, but you can say with precision what an average number will be up to. Individuals vary, but percentages remain constant. So says the statistician.

– Sherlock Holmes, *The Sign of the Four*

The body of an adult human produces in excess of 10^8 new blood cells each day [Lichtman et al., 2010], with nearly all of the mature cell lineages arising from a single type of multipotent progenitor, the hematopoietic stem cell (HSC), through a complex, multi-step process known as hematopoiesis. Despite the characterization of a large number of molecular components and interactions, a mechanistic understanding of the process by which a progenitor cell becomes increasingly restricted, makes discrete fate decisions, and retains memory of these decisions in the face of noisy, conflicting signals, remains elusive. To investigate these questions, we employed a model-

driven approach to study decision-making in the bipotent Megakaryocyte-Erythroid Progenitor (MEP).

3.1 Lineage commitment in hematopoiesis

Residing primarily in the bone marrow, HSCs can either divide to maintain the population of stem cells, or they can lose their self-renewal capacity and differentiate into either the common myeloid progenitor (CMP) or common lymphoid progenitor (CLP) lineages. Further differentiation of CMP and CLP cells gives rise to bipotent progenitors, which in turn undergo differentiation steps to give rise to mature cells; hence, each differentiation step in effect limits the cell types into which a progenitor can morph [Akashi et al., 2000, Eckfeldt et al., 2005, Kondo et al., 1997]. Apart from this canonical view of lineage commitment, other mechanisms, such as bypassing of certain progenitor states, have also been observed [Adolfsson et al., 2005, Kondo et al., 1997, Kondo et al., 2003].

3.1.1 Intrinsic and extrinsic cues

While precise mechanisms have yet to be elucidated, commitment in hematopoiesis is understood to be controlled by lineage-specific transcription factors [Cantor and Orkin, 2001, Enver et al., 2009, Graf and Enver, 2009, Iwasaki et al., 2003, Orkin, 2000, Rosenbauer and Tenen, 2007] and cytokines [Metcalf, 2008, Murphy, 2005, Rieger et al., 2009]. At each stage, progenitor cells express, at low levels, the master regulatory transcription factors of the lineages to which they can commit [Huang et al., 2007, Laslo et al., 2006]. To commit to a particular lineage, a progenitor cell upregulates the relevant transcription factor, primarily via positive auto-regulation [Chen et al., 1995, Tsai et al., 1991], and down-regulates the transcription factor for the opposing lineage, primarily via cross-antagonism [Cantor and Orkin, 2001, Grass et al., 2003, Liew et al., 2006]. For instance, CMPs 'promiscuously' express both

GATA1 and PU.1, the transcription factors for the erythroid/megakaryocyte and myelomonocytic lineages, respectively, and these transcription factors upregulate themselves and inhibit the expression of each other. In fact, the cross-antagonism and autoregulation motif constitutes a theme in not only hematopoietic lineage commitment, but in diverse biological contexts where a discrete decision between two options is required [Chickarmane and Peterson, 2008, DeGracia et al., 2012, Huang et al., 2007, Liew et al., 2006, Soneji et al., 2007].

Research into the role played by cytokines has stirred debate about whether lineage commitment is a stochastic or instructive process [Murphy, 2005, Enver et al., 1998]. The instructive theory points to strong evidence of dependence on specific cytokines during differentiation, and states that the commitment process is tightly controlled by cytokines [Metcalf, 1998, Metcalf, 2008, Rieger et al., 2009, Robb, 2007]. In contrast, the stochastic theory states that lineage commitment is in essence a function of differences in initial expression levels of lineage-specific transcription factors, and cytokines merely provide survival and proliferation signals after commitment has already occurred; in other words, lineage choice is determined by intrinsic noise in the progenitor cell [Abkowitz et al., 1996, Enver et al., 1998, Losick and Desplan, 2008]. The stochastic theory does not explain the observation that a pool of progenitor cells can be enriched for a particular lineage by treatment with the relevant cytokine [Rieger et al., 2009]; similarly the instructive theory fails to explain the observation that cells can be made to differentiate into particular lineages, even when the relevant receptors are knocked out [Murphy, 2005, Enver et al., 1998]. Recent studies have demonstrated that lineage commitment is in fact influenced by both instructive signals and intrinsic noise [Mossadegh-Keller et al., 2013, Rieger et al., 2009].

3.1.2 Erythropoiesis

The transcription factors EKLF (also known as KLF1) and GATA-1, and the cytokine erythropoietin (Epo) play critical roles in regulating erythroid differentiation. Numer-

ous studies have established the essential and global role of EKLF in erythropoiesis. EKLF-null mice die of anemia by E15, due in part to diminished levels of β-globin. However, this lethal phenotype is not reversed by restoration of β-globin levels via alternative means, suggesting a broader regulatory role for EKLF in erythropoiesis [Nuez et al., 1995, Perkins et al., 1995, Tallack et al., 2012]. In humans, mutations in the EKLF gene are associated with congenital dyserythropoietic anemia, a disease characterized by deficient levels of hemoglobin and red blood cells, as well as non-pathological phenotypes such as the In(Lu) blood group [Helias et al., 2013, Singleton et al., 2008], and elevated levels of fetal hemoglobin [Borg et al., 2010].

The EKLF protein comprises an N-terminal activation domain, and a C-terminal domain conferring binding specificity to the 'CCMCRCCCN' sequence, a motif which occurs in the regulatory regions of a number of red blood cell genes. Interestingly, EKLF can both activate and repress transcription; with the choice of activity thought to be regulated by post-translational modifications. For instance, sumoylated EKLF can act as a transcriptional repressor [Siatecka and Bieker, 2011, Siatecka et al., 2007].

Not expressed at significant levels in the HSC state, EKLF expression continues to increase as cells commit first to the CMP branch, and subsequently to the MEP branch. Movement from the MEP stage toward the erythroid stage coincides with a dramatic upregulation of EKLF; in contrast, commitment to the megakaryocyte lineage brings about a marked decrease in EKLF levels [Frontelo et al., 2007, Siatecka and Bieker, 2011]. In hematopoiesis parlance, the expression of EKLF in multipotent and bipotent progenitors is promiscuous.

The zinc-finger transcription factor GATA-1 also plays a central role in erythropoiesis; its disruption by homologous recombination has been demonstrated to block erythroid differentiation in murine embryonic stem cells [Pevny et al., 1991, Kondo et al., 2010, Martin and Orkin, 1990, Evans and Felsenfeld, 1989]. GATA-1's N-terminal domain confers binding specificity to the '(A/T)GATA(A/G)' sequence, which is is found in the regulatory regions of a large number of red blood cell genes [Cantor and Orkin, 2002, Welch et al., 2004]. Like EKLF, GATA-1 has transcrip-

tional activity that is modulated by post-translational modifications [Constantinescu et al., 1999, Krantz, 1991].

In the absence of functional erythropoietin receptor (EpoR), erythroid progenitors do not mature and also exhibit other phenotypic abnormalities [Wu et al., 1995, Ghaffari et al., 2001]. Epo regulates erythroid differentiation, survival, and proliferation; binding of Epo to EpoR can activate signaling cascades including the Akt, Jak/STAT, and MAPK pathways. The presence of GATA-1 binding sites in the enhancer region of the EpoR gene [Chiba et al., 1991, Kuramochi et al., 1990, Zon et al., 1991] and the regulatory region of GATA-1 itself [Hannon et al., 1991, Iwasaki and Akashi, 2007, Tsai et al., 1991, Zon et al., 1991] leads to the concomitant rise in the expression of both GATA-1 and EpoR during erythroid differentiation [Broudy et al., 1991, Palani and Sarkar, 2012] and establishes a regulatory link between these components.

3.1.3 Megakaryopoiesis

CMP cells can yield colonies of MEP cells, which can in turn commit to the megakaryocyte lineage, and give rise to primitive burst-forming unit-megakaryocyte (BFU-MK) cells, which have high proliferative capacity and can yield colonies of hundreds of cells, and more mature colony-forming unit-megakaryocyte cells (CFU-MK) which give rise to colonies with 3 to 50 mature megakaryocytes. Megakaryocytes undergo endomitosis, increasing in size and reaching ploidy of about 4-64N, and then fragment to give rise to proplatelets, which in turn give rise to platelets. The cytokine thrombopoietin (Tpo) is required from the CMP stage until the megakaryocyte maturation stage [Greer and Wintrobe, 2008, Kaushansky, 2008, Michelson, 2007].

Transcription factors regulating megakaryocyte differentiation and maturation include FLI-1, GATA-1, FOG1, and NF-E2. FLI-1, a member of the Ets transcription factor family, is essential for megakaryocyte generation, and promotes the transcription of megakaryocyte- and platelet-specific genes including PF4, glycoprotein IX,

and glycoprotein IIb. Murine ES cells lacking a functional copy of FLI-1 do not contribute to the megakaryocyte lineage, and mutations in the FLI-1 gene are associated with congenital thrombocytopenia in humans [Bastian et al., 1999, Okada et al., 2011, Shivdasani, 2001]. The Ets domain of FLI-1 facilitates binding to the 'ACCGGAAG/aT/c' sequence [Cui et al., 2009, Mao et al., 1994].

In addition to being essential for erythropoiesis, GATA-1 plays a critical role in megakaryocyte differentiation, and loss or impairment of the GATA-1 gene results in various abnormalities including a dramatic reduction in platelet levels, and significant size and shape changes in platelets. The megakaryocyte-specific actions of GATA-1 are thought to be mediated at least in part through direct interaction with FOG1, and detrimental effects on megakaryocyte proliferation and platelet formation are thought to arise due to the loss of this partnership [Kawada et al., 2001, Shivdasani, 2001].

Produced in the liver, kidneys, and bone marrow, Tpo is the primary regulator of platelet production, and plays a critical role in the survival and expansion of progenitor cells with megakaryocyte potential, and in megakaryocyte maturation; deletion of the Tpo or TpoR genes results in a 90% decrease in platelet levels. Binding of Tpo to TpoR results in the activation of several signaling pathways including the Jak/STAT, Akt, and MAPK cascades, and blocking of these pathways can inhibit the survival and proliferation effect of Tpo. Tpo and its cognate receptor TpoR share homology with the Epo/EpoR ligand-receptor pair, and activate downstream signaling via similar mechanisms [Kaushansky, 2005, Kaushansky, 2008].

A number of megakaryocyte-restricted genes contain tandem GATA-1 and Ets binding sites in their regulatory regions, and these genes are optimally transcribed when both an Ets transcription factor such as FLI-1 and the GATA-1/FOG1 complex are bound to the regulatory region. This mechanism is particularly important for regulation of the TpoR gene, since disruption of the GATA-1/FOG1 interaction results in a substantial decrease in TpoR transcription; additionally, embryonic cells from mice lacking FLI-1 exhibit down-regulation of TpoR. Hence, regulation of TpoR levels by GATA-1, FOG1, and FLI-1 provides a link between the lineage-specific transcription

factors and the receptor [Wang et al., 2002].

3.1.4 Antagonism between erythroid and megakaryoid lineages

In addition to being pro-erythropoiesis, EKLF plays an anti-megakaryopoiesis role. Over-expression of EKLF in murine embryonic stem cell lines results in increased numbers of cells committing toward the erythroid lineage, as well as decreased numbers committing to the alternative megakaryocyte lineage. Additionally, fetal liver cells lacking EKLF show marked increases in megakaryocyte progenitors and products [Siatecka and Bieker, 2011, Bouilloux et al., 2008]. EKLF can inhibit FLI-1-mediated upregulation of glycoprotein IX (a platelet marker), while FLI-1 can inhibit EKLF-mediated upregulation of β-globin. Furthermore, in vitro experiments suggest that EKLF and FLI-1 physically interact, and hence imply a direct mechanism by which the apparent cross-antagonism might be mediated [Siatecka and Bieker, 2011].

3.2 Mathematical modeling of lineage commitment

Despite decades of experimental study at multiple scales, there is no widely-accepted model of the governing mechanism that enables a progenitor cell to integrate extrinsic cytokine cues and intrinsic cues from transcription factors and other cellular components, decide on a particular lineage, and retain commitment to that lineage. Systems modeling offers a powerful framework for understanding lineage commitment in hematopoiesis, and cell differentiation in general. In this view, progenitor and differentiated cells are interpreted as steady attractor states in a multistable system; hence, commitment to a particular lineage constitutes transition from one steady state to another. Models formulated via this approach can recapitulate decision-making behaviors that are hallmarks of the lineage commitment process [Huang et al., 2005, Huang et al., 2007, Huang, 2009]. A recent paper [Foster et al., 2009a] identified

these hallmarks as:

1. Stability: progenitor and mature cells should be considered stable attractors or states of the process dynamics.

2. Branching: movement down the decision tree eliminates differentiation pathways at each step.

3. Directionality: commitment gives rise to specialized cells, and reverts to the progenitor in very rare cases.

4. Exclusivity: mature cells do not express genes specific to other lineages.

5. Promiscuous expression: progenitors express, at low levels, the lineage-specific transcription factors of the subset of lineages to which they can commit.

3.2.1 Cross-antagonism autoregulation model

The cross-antagonism autoregulation (CAA) network topology (Fig. 3-1) has been proposed as a general model for bipotent progenitor commitment. This topology consists of two mutually-repressive autoregulating transcription factors, each driving the expression program for a mature cell lineage. Antagonism between the two lineage-specific transcription factors can confer bistability, meaning that at steady-state, only one of them can be expressed at high levels. Autoregulatory interactions on both transcription factors give rise to a bipotent state, in which both transcription factors can be expressed at medium levels, or promiscuously. Layering of multiple instances of this topology in a hierarchy leads to a systems model of lineage commitment that satisfies the hallmarks described above [Huang, 2009, Foster et al., 2009a].

A frequently observed, yet often overlooked finding during hematopoietic lineage commitment is the upregulation of receptors of the commitment lineage and downregulation of receptors of opposing lineages, due to positive and negative feedback, respectively, from the master transcription factor [Chiba et al., 1991, Hohaus et al.,

1995, Smith et al., 1996, Wang et al., 2002, Zhang et al., 1994]. Results from systems modeling efforts suggest that positive receptor feedback does not merely amplify the cytokine signal, but can act as a core component of the commitment machinery, helping confer bistability [Palani and Sarkar, 2008, Palani and Sarkar, 2009, Palani and Sarkar, 2012, Shah and Sarkar, 2011].

3.2.2 Model validation

The CAA model proposes an elegant explanation of the process by which a bipotent progenitor cell decides on, and stays committed to, a particular mature cell lineage. However, important predictions of this model have been difficult to validate either because most experimental studies aimed at measuring master regulatory transcription factor levels report ensemble averages of entire populations, or employ single-cell assays that are not sufficiently quantitative to allow inference of individual transcription factor distributions and correlations between transcription factors.

Toward the goal of understanding the decision-making behavior of the erythrocyte-megakaryocyte fate switch in mechanistic detail, we performed experiments in the bipotent UT-7/GM cell line, and used fluorescence *in situ* hybridization (FISH) to track the mRNA levels of the EKLF, FLI-1, and GATA-1 transcription factors simultaneously in individual cells. Furthermore, to investigate the role of extrinsic cues in the commitment decision, we tracked a second panel of genes including EpoR, TpoR, and EKLF.

3.3 Results

We used a combination of systems modeling and experiments to elucidate the mechanisms by which bipotent MEP cells process intrinsic and extrinsic cues and make commitment decisions. We performed experiments in UT-7/GM , a bi-potential human cell line, which gives rise to erythrocytes, expressing high levels of hemoglobin,

or megakaryocytes, expressing high levels of platelet factor 4 and platelet glycoprotein GPIIb, when treated with Epo or Tpo, respectively [Komatsu et al., 1997]. Use of a cell line offers an important advantage over primary cells in this study: purification of progenitors from primary cells is not a perfect process, and can erroneously result in the attribution of heterogeneity in receptor and transcription factor expression to noise intrinsic to the commitment system; in contrast, the reduced variability in synchronized populations from a model cell line enhances experimental reproducibility and facilitates quantitative analysis.

3.3.1 Conflicting cue can bias progenitor commitment

To understand how the presence of a conflicting extrinsic cue can impact commitment to the erythrocyte lineage, we first sought to establish the concentrations at which UT-7/GM cells respond to Tpo treatment; a dose-response study to establish analogous concentrations of Epo treatment was performed earlier by our group [Palani and Sarkar, 2012]. UT-7/GM cells maintained in GMCSF (granulocyte macrophage colony-stimulating factor) medium were growth-factor starved for 18 hours, and passaged into media containing different concentrations of Tpo (0.05, 0.5, 5, and 20 ng/mL) on Day 0. Cultures were passaged into fresh medium and tracked for cell viability at regular intervals. Results show that at high concentrations of Tpo (≥ 5 ng/mL) , cultures maintain high viability levels; in contrast, at low concentrations of Tpo, viability levels decrease with time, before recovering to approximately those at high Tpo concentrations (Fig. 3-3). To assess megakaryocyte differentiation, qRT-PCR for the megakaryocyte-specific marker PF4 was performed on RNA extracted from Day 28 cultures; all Tpo-treated cultures show strong PF4 expression, compared to a culture maintained in GMCSF. Together, the viability and PF4 expression data indicate that at decreasing concentrations of Tpo, smaller proportions of UT-7/GM cells commit to the megakaryocyte lineage and survive, resulting in longer recovery times for viability (Fig. 3-3).

Having established the Epo and Tpo concentrations at which UT-7/GM cultures yield erythrocyte- and megakaryocyte-committed cells, we performed a cytokine competition experiment to understand how the presence of Tpo impacts erythrocyte lineage commitment. UT-7/GM cells maintained in GMCSF were growth-factor starved for 18 hours, and passaged into medium containing different cocktails of the two cytokines; three different concentrations each were used for Epo (0.001, 0.01, 0.1 U/mL) and Tpo (0.05, 0.5, 5 ng/mL), yielding a total of nine (Epo and Tpo) combinations. Cultures were passaged into fresh medium on alternate days, and were tracked for viability at regular intervals. To assess erythrocyte differentiation, cells were stained with o-dianisidine and analyzed by microscopy on Day 12. Results show that at least at the population level, the introduction of increasing concentrations of the conflicting cue, Tpo, leads to marked decreases in the proportion of erythrocytes (Fig. 3-4).

3.3.2 Commitment to the erythrocyte lineage is irreversible

Memory is understood to be an essential property of lineage commitment. Cells deciding to commit to a particular branch must remain committed even after the commitment-inducing cue is reduced, or a conflicting cue is introduced; the former has been demonstrated in a previous study by our group [Palani and Sarkar, 2012]. To assess how the introduction of a conflicting cue impacts memory in erythrocyte lineage commitment, UT-7/GM cells maintained in GMCSF were growth-factor starved for 18 hours, and passaged on Day 0 into three different media: medium containing high levels of Epo (Epo-only, 1 U/mL), high levels of Tpo (Tpo-only, 5 ng/mL), or high levels of both Epo and Tpo (Epo+Tpo; Epo: 1 U/mL, Tpo: 5 ng/mL). On Days 3 and 6, part of the Epo-only culture was passaged separately into Epo+Tpo medium; hence, for these two cultures, a competing cytokine was added at different time points, while the original inducing cue was kept constant. All cultures were tracked for viability at regular intervals, and assessed for erythrocyte differentiation via o-dianisidine staining on Day 14.

Results show that 74% of cells in the Epo-only culture, virtually no cells (1%) in the Tpo-only culture, and an intermediate proportion (23%) of cells in the Epo+Tpo culture stain positive. Cells passaged from Epo-only medium into Epo+Tpo on Day 3 yield (37%) of positively-stained cells, a significantly higher proportion than that of cells maintained in Epo+Tpo medium from Day 0. Cells passaged from Epo-only medium into Epo+Tpo medium on Day 6 yield virtually the same proportion of positively-stained cells (71%) as those maintained in Epo-only medium throughout the course of the experiment (Fig. 3-5).

The results show that the later Tpo is introduced, the smaller its impact on erythrocyte differentiation; this indicates that either cells in the starting population committing to the megakaryocyte lineage die because of the absence of a growth signal (consistent with the stochastic theory), or virtually all cells in the starting population commit to the erythrocyte lineage and hence become un-responsive to Tpo (consistent with the instructive theory). Regardless, commitment to the erythrocyte lineage is not reversible by the introduction of a conflicting Tpo cue.

3.3.3   Systems model for bipotent progenitor commitment

Having established the concentrations of cytokines and the time-dynamics of commitment in UT-7/GM cells, we switched our focus from ensemble averages of end-stage differentiation markers to studying the key transcriptional regulators of erythrocyte lineage commitment in individual cells.

To guide our experimental strategy, we first developed a systems model and performed simulations to identify concrete predictions that could be tested. Similar to the CAA model (Fig. 3-1), our model comprises two lineage-specific transcription factors, T_A, T_B, driving commitment to their respective lineages, A, B; each transcription factor upregulates its own synthesis via positive feedback, and inhibits the opposing lineage's transcription factor. However, since the CAA topology does not account for extrinsic cues, we developed an augmented ECAA (extrinsic cross an-

tagonism autoregulation) model (Fig. 3-6) that integrates cytokine cues into the commitment decision, in a manner consistent with published observations. In this model, extrinsic stimuli L_A, L_B promote the up-regulation of T_A, T_B, and the transcription factor for each lineage suppresses the signaling-mediated up-regulation of the opposing lineage's transcription factor.

Analysis of this model demonstrates the existence of three steady-states: two exclusive states where one transcription factor is present at high levels, and the other is expressed only minimally, and a third intermediate state where both transcription factors are expressed at medium levels (Fig. 3-2). In the context of commitment in hematopoiesis, the intermediate state can be considered to represent a bipotent progenitor cell, and the exclusive states can be considered to represent committed, mature cells.

Stochastic simulations of this model via the Gillespie algorithm [Gillespie, 1977] demonstrate that cells starting in the bipotent state eventually settle in one of the exclusive states (Fig. 3-7). Simulations performed with varying combinations of L_A, L_B show that the proportions of cells committing to either lineage can be significantly skewed by extrinsic cues (Figs. 3-8, 3-9, 3-10, 3-11). However, inhibition between the transcription factors establishes an effective separatrix, which when crossed, renders a cell's reversion to the opposing lineage highly improbable, regardless of changes in L_A, L_B levels; this property of the model captures the general behavior of cells in the memory experiment described above.

Analysis of simulated trajectories reveals that as an individual cell commits to a particular lineage (e.g. lineage A), the expression of the relevant master transcription factor, T_A, is up-regulated, while that of the opposing lineage, T_B, is down-regulated; in the context of a population, an anti-correlation develops between T_A, T_B, representing the cell's movement toward the point representing the A lineage, and away from the point representing the B lineage, in gene-expression space. Given that different lineages activate vastly different gene-expression programs, this is an unsurprising consequence; however, model simulations suggest that the anti-correlation between

master transcription factors develops at the outset of the commitment process, and hence significantly before the appearance of clear differentiation markers, such as hemoglobin in erythrocytes. Taken further, if strong bias toward particular lineages exists in individual cells in a bipotent progenitor population (in line with the stochastic theory), then it should be manifest in the levels of master transcription factors; i.e. individual cells would be expected to have low T_A, high T_B, and vice versa, leading to a negative correlation between T_A and T_B,

Taken together, our results demonstrate that the ECAA model can account for extrinsic and stochastic cues, confer discrete commitment decisions, and retain memory, and hence offers a simple framework in which to study decision-making in bipotent progenitors.

3.3.4 Dynamics of EKLF and FLI-1 in MEP commitment

To study model predictions in the context of erythrocyte lineage commitment, we treated UT-7/GM cells with Epo and Tpo and tracked transcript levels of EKLF, FLI-1, and GATA-1 at regular intervals by mRNA FISH, a powerful technique that allows detection of individual mRNA molecules via microscopy, and has been used to gain important insights into gene regulation in a variety of organisms [Raj et al., 2006, Maamar et al., 2007, Raj and van Oudenaarden, 2008, Raj and Tyagi, 2010].

Results show that at the population level, both Epo-only and Tpo-only cultures exhibit congruent up-regulation in the median number of EKLF transcripts until Day 3. Subsequently, EKLF levels decrease significantly in the Tpo-only culture, and increase dramatically in the Epo-only culture. Initial up-regulation of EKLF, a pro-erythropoiesis transcription factor, in Tpo-only medium suggests the movement of a small proportion of cells in the population towards the erythrocyte lineage; the inability of this subset of cells to survive in the absence of Epo can explain the subsequent down-regulation of EKLF (Fig. 3-13).

FLI-1 transcript counts exhibit the opposite pattern. After treatment, FLI-1 levels

in the Epo-only culture continue to decline precipitously, reaching a median value of 8, compared to a median value of 79 in the pre-treatment population. In contrast, FLI-1 levels exhibit a modest increase under Tpo-treatment, reaching a median value of 95 transcripts on Day 12. In the Epo+Tpo culture, median transcription factor levels are generally consistent with the population splitting into the opposing erythrocyte and megakaryocyte lineages. Interestingly, however, the median FLI-1 levels for this culture are significantly higher than those for both Epo-only and Tpo-only cultures at early time-points (Fig. 3-13).

Analysis of transcript counts in individual cells reveals that under most conditions, transcription factor distributions exhibit high variance; for instance, on Day 12 in the Epo-only culture, median 80% transcript counts vary from approximately 47 to 282 (Fig. 3-14). This variability could arise from cells adopting opposing commitment paths, or from stochastic mRNA synthesis [Raj et al., 2006].

3.3.5 Mutual antagonism develops between EKLF, FLI-1 during MEP commitment

Simulations of the CAA model predict that cells moving toward the bipotent state in gene-expression space would exhibit a positive correlation in master transcription factor levels, while cells moving toward a committed state would exhibit an anti-correlation, due to mutual antagonism. UT-7/GM cells can differentiate into erythrocytes and megakaryocytes, and in the context of the CAA model, cells not treated with Epo or Tpo represent the bipotent state.

Examination of EKLF, FLI-1 levels in UT-7/GM cells before cytokine treatment reveals that although both are expressed, no significant correlation exists between the two transcription factors (GF-starved culture: P=0.4, ρ 95% CI [-0.28,+0.16], Fig. 3-15). After treatment with Epo (1 U/mL), individual cells exhibit an increase in EKLF, and a decrease in FLI-1, leading to a significant anti-correlation between the two transcripts on Day 6 (P<0.001, Spearman's rank correlation coefficient ρ 95%

CI [-0.65,-0.33]). On Day 12, cells exhibit a further up-regulation of EKLF, and a dramatic reduction in FLI-1 levels. In contrast, cells treated with high Tpo (5 ng/mL) exhibit down-regulation of EKLF, and modest up-regulation of FLI-1 by Day 12 (Fig. 3-17).

Interestingly, FISH data also show that in Epo-only and Tpo-only cultures, a small proportion of the population appears to traverse the trajectory opposite to the one being induced by the cytokine; given that Epo and Tpo are present at saturating levels, this behavior suggests that commitment is not entirely instructive. In our data, this effect is likely further diminished by the inability of UT-7/GM cells committing to the opposite lineage to continue to survive and proliferate in the absence of the cognate cytokine.

EKLF, FLI-1 data from the Epo+Tpo culture show the population in effect bifurcating into the two commitment trajectories, leading to strong anti-correlation between EKLF and FLI-1 visible on Day 12 (P<0.001, ρ 95% CI [-0.84,-0.65], Fig. 3-17). Consistent with model predictions, the results demonstrate that an anti-correlation between EKLF, FLI-1 levels develops only and quickly after treatment with cytokines, and hence do not suggest the presence of commitment bias mediated by EKLF and FLI-1, as predicted by the stochastic theory.

3.3.6 GATA-1 as a marker for pro-differentiation bias

At the population level, median GATA-1 transcript levels increase significantly immediately after treatment in both Epo-only and Tpo-only cultures, subsequently leveling off in the Tpo-only culture, and continuing to increase in the Epo-only culture (Fig. 3-12). Cells in the Epo-only culture exhibit a significant, and strong positive correlation between EKLF and GATA-1 (Day 12: P<0.001, ρ 95% CI [+0.47,+0.72], Fig. 3-18). Cells in the Tpo-only culture also exhibit a positive correlation between EKLF and GATA-1 that is significant, but weaker in magnitude (Day 12: P<0.001, ρ 95% CI [+0.17,+0.54]).

Interestingly, even before treatment with cytokines, EKLF and FLI-1 are both significantly positively correlated to GATA-1 (EKLF: P<0.001, ρ 95% CI [+0.07,+0.47]; FLI-1: P<0.001, ρ 95% CI [+0.19,+0.56], Fig. 3-18), suggesting that in a pool of un-induced bipotent progenitors, high GATA-1 levels indicate the presence of a pro-differentiation bias. These findings affirm GATA-1's supportive role in erythrocyte and megakaryocyte lineage commitment.

3.3.7 EKLF establishes an irreversibility threshold

As described above, UT-7/GM cells exhibit memory of Epo-treatment, in that the introduction of the competing cytokine Tpo at later timepoints has diminished effects on the proportion of cells committing to the erythrocyte lineage. To understand the role of master transcription factors in establishing this memory, we passaged growth-factor starved UT-7/GM cells into medium containing only Epo (1 U/mL) for 3 days, after which the culture was maintained in medium containing both Epo and the competing cytokine Tpo (5 ng/mL).

Comparison of transcript counts in this culture with one that was maintained in Epo+Tpo throughout reveals that cells passaged from Epo-only medium into Epo+Tpo medium exhibit higher levels of EKLF on Day 6 compared to cells maintained in Epo+Tpo throughout, as expected, since Epo supports the erythrocyte lineage. In contrast, the distributions of FLI-1 levels in the two cultures on Day 6 are very similar, and exhibit significantly higher transcript counts than the Epo-only culture, indicating that the introduction of Tpo allows cells to maintain higher FLI-1 levels (Fig. 3-20).

Given that Epo-only pre-treatment has a definite phenotype effect, and that this bias is not reflected in FLI-1 levels, our results suggest that EKLF effectively establishes a threshold above which the cell irreversibly commits to the erythrocyte lineage, regardless of the presence of Tpo. As explained by our model, an increase in EKLF levels beyond the threshold could activate an autoregulatory feedback, resulting in

further up-regulation of EKLF, and more effective repression of FLI-1. Consistent with model predictions, FLI-1 levels in erythrocyte-committing (EKLF-high) cells are dramatically lowered on Day 12 (Fig. 3-20).

Overall, this analysis demonstrates the utility of multi-stability in progenitor lineage commitment: although intrinsic and extrinsic perturbations can push cells into various parts of gene-expression space, the existence of stable steady-states representing mature cell-types virtually guarantees that each cell eventually settles into a state with a well-defined phenotype.

3.3.8 Correlations between EKLF, EpoR, and TpoR suggest broader regulatory role for cytokine receptors

In the context of hematopoiesis, cytokine receptors allow extrinsic cues to influence the lineage commitment process. Experimental studies suggest that during commitment in some systems, the receptor for the cognate lineage is up-regulated, while that of the opposing lineage is down-regulated [Chiba et al., 1991, Hohaus et al., 1995, Smith et al., 1996, Wang et al., 2002, Zhang et al., 1994].

Systems modeling suggests that more than simply amplifying the signal, receptor up-regulation can help confer bistability, and hence can act as a core component of the decision-making machinery. In this view, master transcription factors up-regulate cognate receptors, leading to a positive correlation between these two species, and a negative correlation between the master transcription factor of one lineage and the receptor for another lineage [Palani and Sarkar, 2008, Palani and Sarkar, 2009, Palani and Sarkar, 2012].

To test these predictions, UT-7/GM cells maintained in GMCSF medium were growth-factor starved for 18 hours, and passaged into Epo-only (1 U/mL) and Tpo-only (5 ng/mL) media. Transcript counting via mRNA FISH reveals that at the population level, median EpoR in Epo-treated cells rises dramatically from 37 immediately before cytokine treatment, to 90 on Day 10; in contrast, median EpoR in

Tpo-treated cells falls to 16 on Day 10. Hence, EpoR levels rise during erythrocyte lineage commitment, and fall during megakaryocyte lineage commitment (Fig. 3-12).

Consistent with the receptor up-regulation model [Palani and Sarkar, 2009], median TpoR levels decrease dramatically in Epo-treated culture, going from 41 immediately before cytokine treatment to 2 on Day 10; however, median TpoR in Tpo-treated culture levels also decrease significantly, reaching 21 on Day 10 (Fig. 3-12).

Analysis of EKLF, EpoR, and TpoR transcript counts in individual cells reveals a significant and strong positive correlation between EKLF and EpoR in both Epo-only (Day 10: $P<0.001$, ρ 95% CI [+0.75,+0.88]) and Tpo-only (Day 10: $P<0.001$, ρ 95% CI [+0.52,+0.76]) cultures (Fig. 3-21). Furthermore, the positive correlation between EKLF and EpoR transcript levels exists even before the addition of cytokine in growth factor-starved cells (Day 0: $P<0.001$, ρ 95% CI [+0.28,+0.62], (Fig. 3-16)). Our results suggest that for an individual progenitor cell, the EpoR level is, in effect, a proxy for the master transcription factor, EKLF; hence EpoR can perhaps be used as a convenient cell-surface marker to identify progenitor cells that are biased towards the erythrocyte lineage.

Taken together, the up-regulation of EpoR and down-regulation of TpoR during erythrocyte lineage commitment, down-regulation of EpoR during megakaryocyte lineage commitment, and the strong positive correlation between EKLF and EpoR support the receptor-feedback model [Palani and Sarkar, 2008, Palani and Sarkar, 2009] for two main reasons. First, the results are consistent with model predictions for erythrocyte lineage commitment, and offer a simple mechanism for generating correlations between EKLF and EpoR. Second, synthesis of a cytokine receptor by the master transcription factor can yield ultrasensitivity to the cytokine, and this interaction, when combined with positive auto-regulation on the master transcription factor, can yield bistability in the absence of explicit non-linearity. Although cooperativity is often explicitly included in models of systems exhibiting bistability (as it is in our model), it may not always exist in biological systems [Palani and Sarkar, 2008].

Overall, our analysis implicates the cytokine receptors in important regulatory roles beyond the mere forwarding of extrinsic cues.

3.4 Methods

3.4.1 Cell culture

All experiments were performed in UT-7/GM cells [Komatsu et al., 1997], which were obtained from Dr. Kenneth Kaushansky (University of California, San Diego) and Dr. Norio Komatsu (University of Yamanashi). Before beginning differentiation experiments, cells were cultured in medium composed of IMDM (Iscove's modified Dulbecco's medium), 10% FBS (fetal bovine serum, Hyclone), and 1 ng/mL GMCSF, after which they were washed with PBS (phosphate buffered saline), and cultured in medium lacking GMCSF for 18 hours. Cells were then washed again with PBS, and passaged into medium containing different concentrations of Epo and Tpo, as appropriate for the experiment. Epo was purchased from Applichem, and Tpo, GMCSF were purchased from Peprotech. The first day of treatment with either Epo or Tpo is referred to as Day 0.

3.4.2 Viability

Cell viability measurements were made either with Trypan blue (Mediatech) in conjunction with microscopy, or with the ViaCount (Millipore) reagent on a Guava Easy-Cyte flow cytometer.

3.4.3 Hemoglobin staining

Erythrocyte differentiation was assessed by staining with o-dianisidine; briefly, cells were washed and re-suspended in IMDM containing 0.3% acetic acid, 0.3% hydrogen peroxide, and 3,3'-dimethoxybenzidine, Fast Blue B (Sigma). After a 20-minute

incubation at room-temperature, cells were analyzed by microscopy.

3.4.4 qRT-PCR for PF4

Total RNA was extracted from fresh cultures with the RNeasy Mini Kit (Qiagen) according to the manufacturer's standard protocol, including the additional QIAshredder and DNase (RNase-Free DNase Set) treatment steps. Purified RNA was stored at -80°C. Reverse transcription was performed using the High Capacity cDNA Reverse Transcription Kit (Life Technologies). Finally, qPCR was performed on the cDNA using the SYBR Green PCR Master Mix reagent (Life Technologies) on an Applied Biosystems 7500 Fast Real-Time PCR System.

Primers used in qPCR include 5'-GCG CTG AAG CTG AAG AAG AT-3', 5'-AGC AAA TGC ACA CAC GTA GG-3' for PF4, and 5'-GCA CCA CGT CCA ATG ACA T-3', 5'-GTG CGG CTG CTT CCA TAA-3' for the endogenous control, POLR2A [Radonić et al., 2004].

3.4.5 Transcript counting via mRNA FISH

Fresh cultures were washed twice with PBS, fixed in 3.7% formaldehyde in PBS for 10 minutes, washed twice again with PBS, resuspended in 70% ethanol, and stored at 4°C.

FISH reactions were performed in solution overnight at 37°C, according to established protocols [Raj et al., 2008, Batish et al., 2011] using two panels of oligonucleotide probes. Panel 1 includes probes targeting EKLF (coupled to Alexa 647N), FLI-1 (coupled to Alexa 594), and GATA-1 (coupled to Cy3); Panel 2 includes EKLF from Panel 1, and probes targeting EpoR (coupled to Cy3), and TpoR (coupled Alexa 594). Coupling of probes with fluorescent dyes was performed as described previously [Batish et al., 2011]. Probes were designed using a custom method developed by Arjun Raj's laboratory (University of Pennsylvania), and were ordered from Biosearch Technologies.

Imaging was performed on a Nikon Ti-E microscope equipped with a 100 x objective, a cooled CCD camera, and the appropriate filter sets. Images for each probe set were acquired separately in stacks spaced $0.3\mu m$ apart. Image segmentation and analysis was performed using a custom Matlab pipeline developed by Arjun Raj's laboratory.

3.4.6 Statistical methods

For FISH data, the significance of correlation between the counts of two transcripts was assessed by comparing Spearman's rank correlation coefficient, ρ, for the dataset to coefficients obtained from 1000 random permutations, from which a P-value was computed. Bootstrapping was used to estimate confidence intervals for ρ. Briefly, transcript counts for a pair of genes at a given timepoint were sampled with replacement to yield a dataset the size of the original dataset, and ρ was computed for this bootstrapped dataset. This process was repeated 1000 times to infer a distribution for ρ.

3.4.7 Systems model for bipotent lineage commitment

The canonical cross-antagonism auto-regulation topology used to model bipotent progenitor commitment [Graf and Enver, 2009, Huang, 2009] was augmented into the ECAA model to include mechanisms by which extrinsic cytokine cues can modulate decision-making. The model consists of two transcription factors, T_A and T_B (represented by A and B in the model equations), driving the expression programs of separate lineages. Each transcription factor up-regulates its own synthesis via positive feedback, and inhibits the synthesis of the other transcription factor. Under certain parameter regimes, this system can give rise to bistability.

b_0, b_1 represent basal synthesis rates, d_0, d_1 represent degradation rates, v_0, v_1 represent maximal feedback-mediated synthesis rates, K_0, K_1 represent transcription factor values at which feedback-mediated synthesis is half-maximal (in the absence of

inhibition), and I_0, I_1 represent repression constants for feedback-mediated synthesis. S_0, S_1 terms represent synthesis of A, B induced by extrinsic cues L_A, L_B; signaling between the cognate receptor and transcription factor is inhibited by the opposing transcription factor. The complete model is stated below.

$$\frac{dA}{dt} = b_0 - d_0 A + \frac{v_0}{\left(\frac{K_0}{A+S_0}\right)^2 \left(1 + \frac{B}{I_0}\right) + 1} \tag{3.1}$$

$$\frac{dB}{dt} = b_1 - d_1 B + \frac{v_1}{\left(\frac{K_1}{B+S_1}\right)^2 \left(1 + \frac{A}{I_1}\right) + 1} \tag{3.2}$$

$$S_0 = \left(\frac{v_{S_0}}{\frac{1}{L_A} + 1}\right) \left(\frac{1}{1 + \frac{B}{\gamma_0}}\right) \tag{3.3}$$

$$S_1 = \left(\frac{v_{S_1}}{\frac{1}{L_B} + 1}\right) \left(\frac{1}{1 + \frac{A}{\gamma_1}}\right) \tag{3.4}$$

Model simulations were performed using a custom implementation of the Gillespie algorithm [Gillespie, 1977].

3.4.8 Estimation of potential energy landscape

In the context of the ECAA systems model, a simulated cell can be interpreted as being subject to different forces, depending on its location in T_A-T_B space. For instance, at (high T_A, low T_B), a cell would be 'drawn' to the A-lineage attractor much more strongly than to the B-lineage attractor. Based on this idea, we developed an algorithm to infer the potential-energy landscape for the system under different conditions. The algorithm works as follows. Given a set of parameters and L_A, L_B for the systems model, the phase-space for T_A-T_B is binned into a 100 x 100 grid; this process yields 10,000 unique (T_A-T_B) uniformly-distributed points.

Using each of the 10,000 points as initial levels for T_A-T_B, 100 stochastic simu-

lations are performed. Each simulation proceeds, until the value of either T_A or T_B changes by more than a specified input parameter, λ, at which point it is halted, and the simulation time is recorded. If neither T_A or T_B change by more than λ after a specified threshold length of time, the simulation is halted, and the simulation time is recorded. This process yields 100 simulation times for each of the points in the $(T_A$-$T_B)$-space; computing the median yields one value, U, for the simulation time for each of the points in the phase-space. Finally, the set of (T_A, T_B, U) points are log-transformed and forwarded to a plotting program.
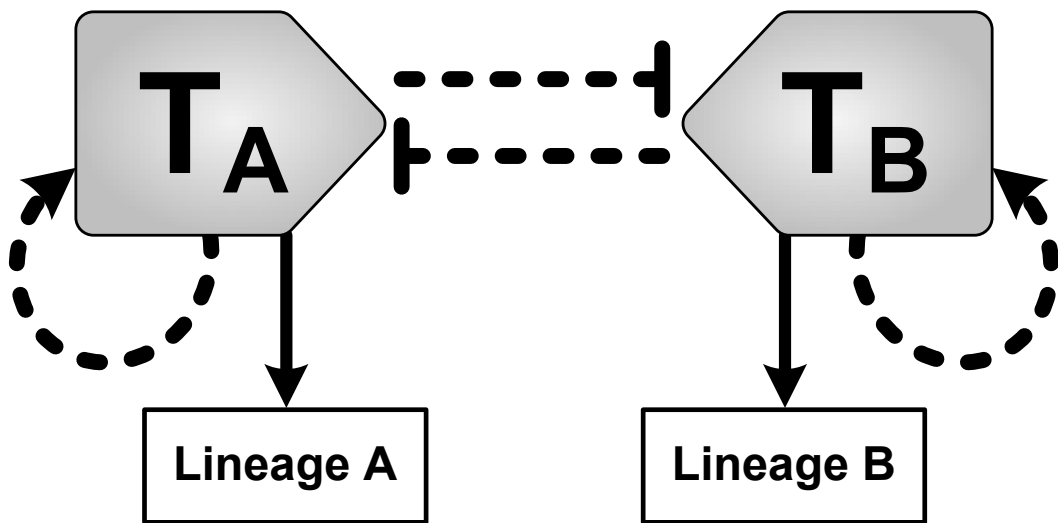
Figure 3-1: **Cross-antagonism auto-regulation motif.** The CAA motif has been proposed as a simple systems model for bipotent progenitor commitment. Transcription factors T_A, T_B repress each other, and up-regulate their own synthesis.
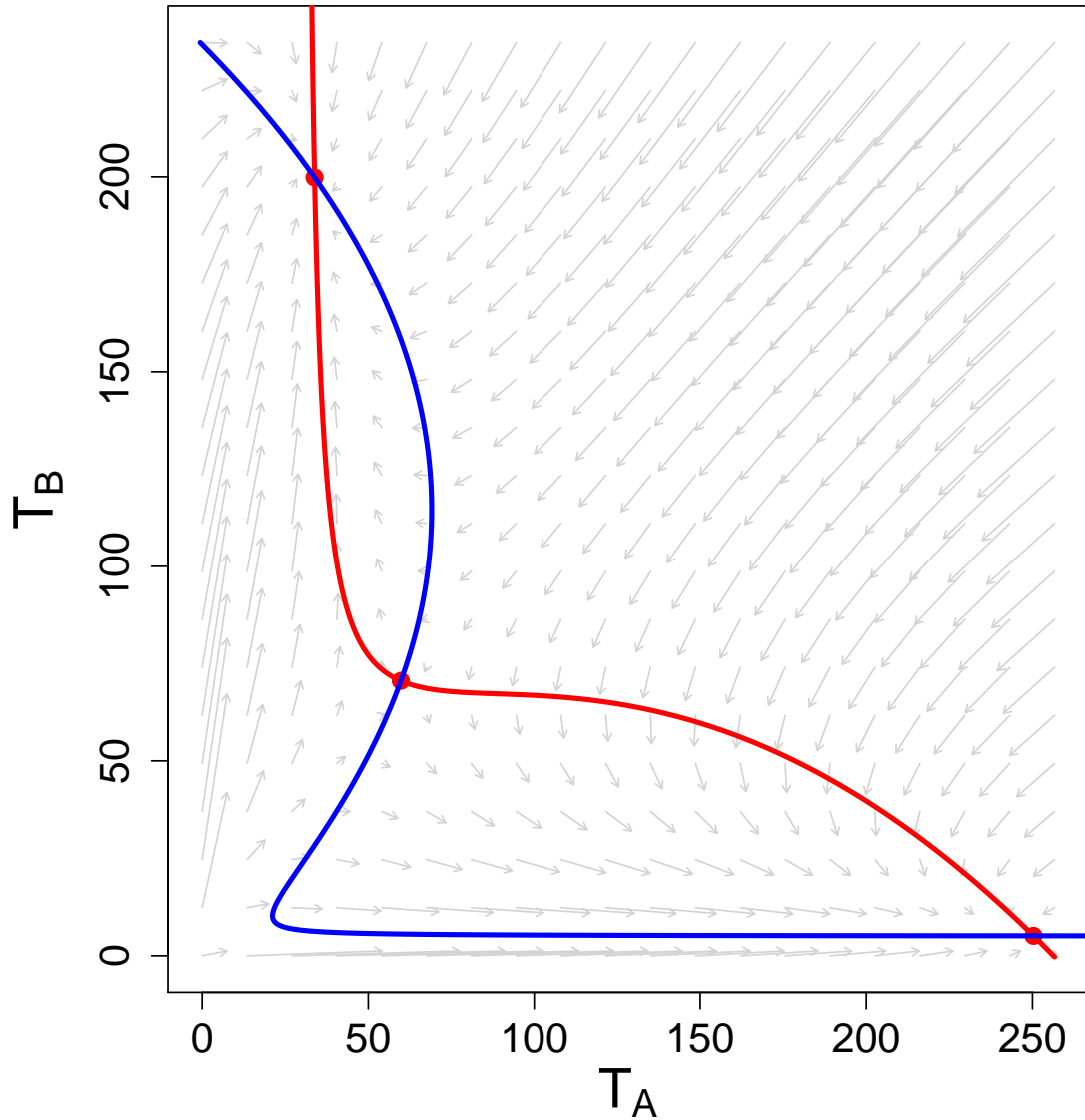
Figure 3-2: **Phase plot for CAA model.** The vector-field and nullclines for the CAA model demonstrate the existence of three steady-states for transcription factors T_A, T_B: two stable states representing committed cells, with one transcription factor maintained at high expression and the other only basally, and one unstable steady state representing progenitor cells, with both transcription factors expressed at intermediate levels. Nullclines for T_A and T_B are plotted in red and blue, respectively.
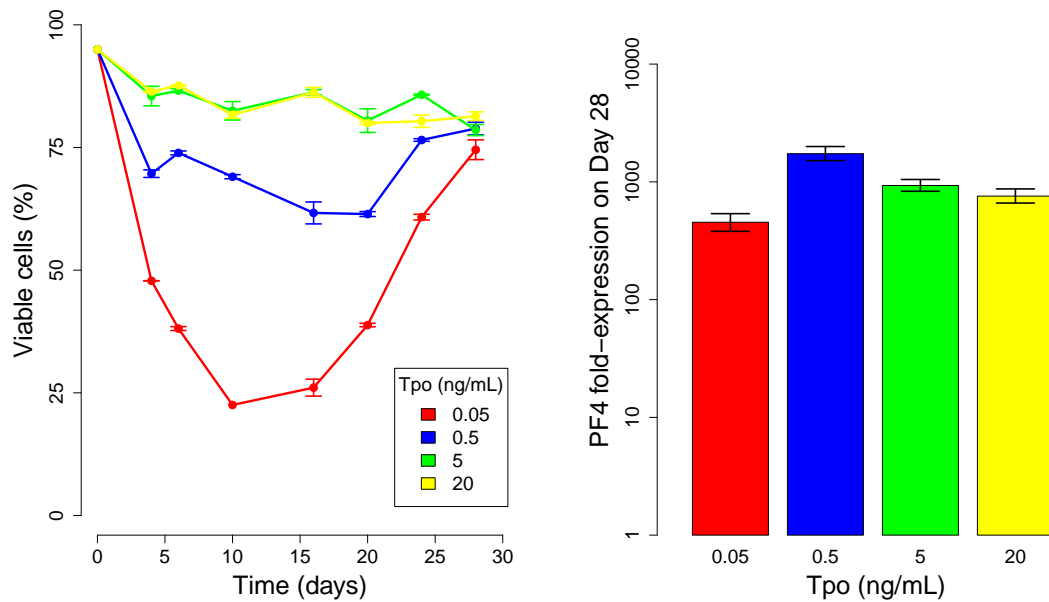
Figure 3-3: **Tpo-induced megakaryocyte differentiation.** UT-7/GM cells were cultured in different concentrations of Tpo. Cultures were tracked for viability by flow cytometry. Day-28 cultures were assessed for expression of PF4, a megakaryocyte marker, via qRT-PCR; bars represent fold-expression of PF4 over cells maintained in GMCSF medium. All dose experiments were performed in duplicate cultures, and qRT-PCR reactions were run in triplicate for each culture.
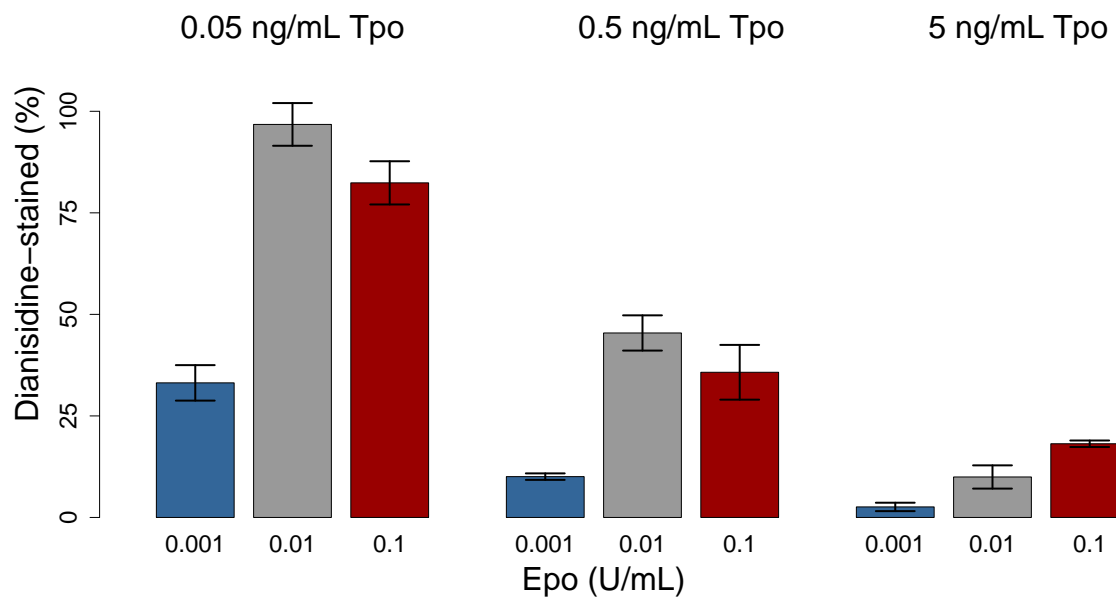
Figure 3-4: **Impact of Tpo on Epo-induced erythrocyte differentiation.** The presence of a conflicting cue, Tpo, can bias progenitor commitment. UT-7/GM cells were cultured in nine different combinations of Epo and Tpo. Erythrocyte differentiation was analyzed on Day 14 via o-dianisidine staining and microscopy. Bars represent proportions of erythrocytes in different cultures. Error bars represent the standard error from duplicate cultures.
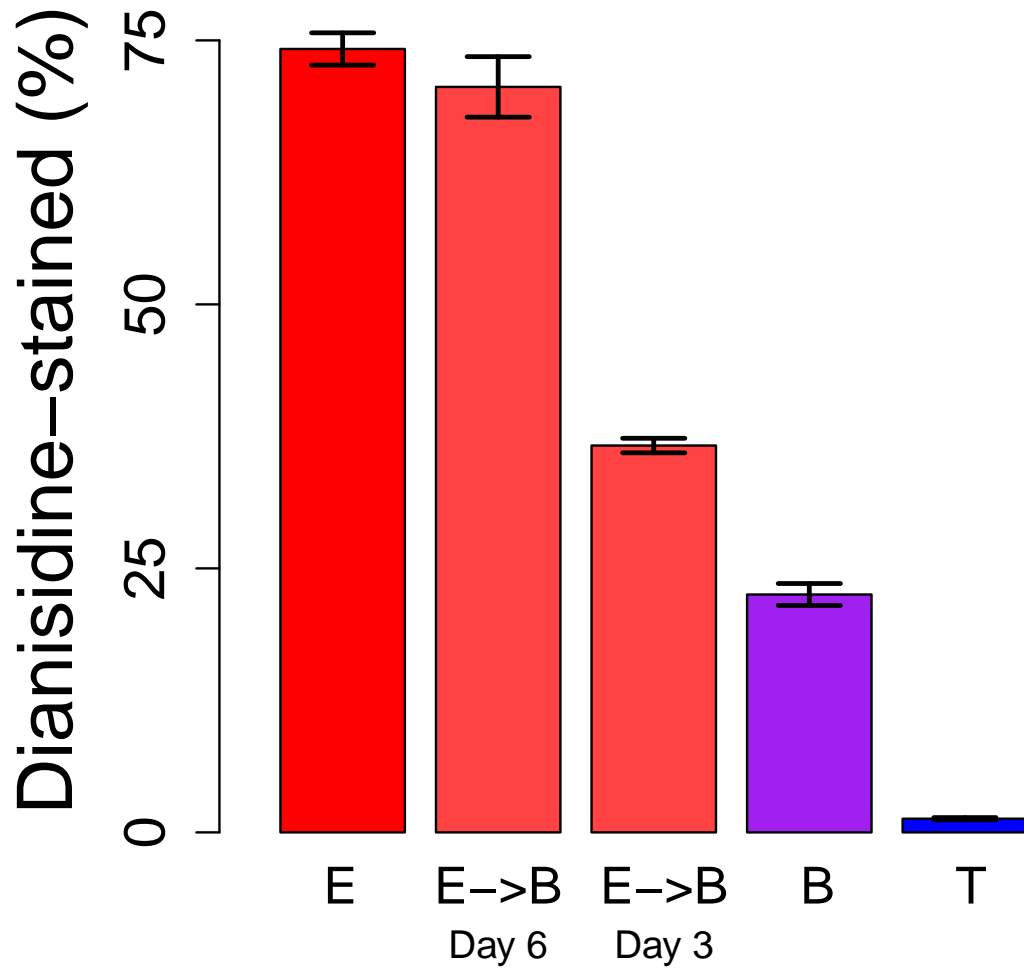
Figure 3-5: **Irreversibility of erythrocyte lineage commitment.** UT-7/GM cells were passaged into medium containing only Epo (1 U/mL, labeled E), only Tpo (5 ng/mL, labeled T), or both Epo and Tpo (1 U/mL Epo and 5 ng/mL Tpo, labeled B). On Days 3 and 6, part of the Epo culture was passaged separately into medium containing Epo and Tpo (1 U/mL Epo and 5 ng/mL, labeled E->B). All cultures were assessed for erythrocyte differentiation by o-dianisidine staining on Day 14. Results demonstrate that the conflicting cytokine Tpo has diminished impact on differentiation when introduced at later times.
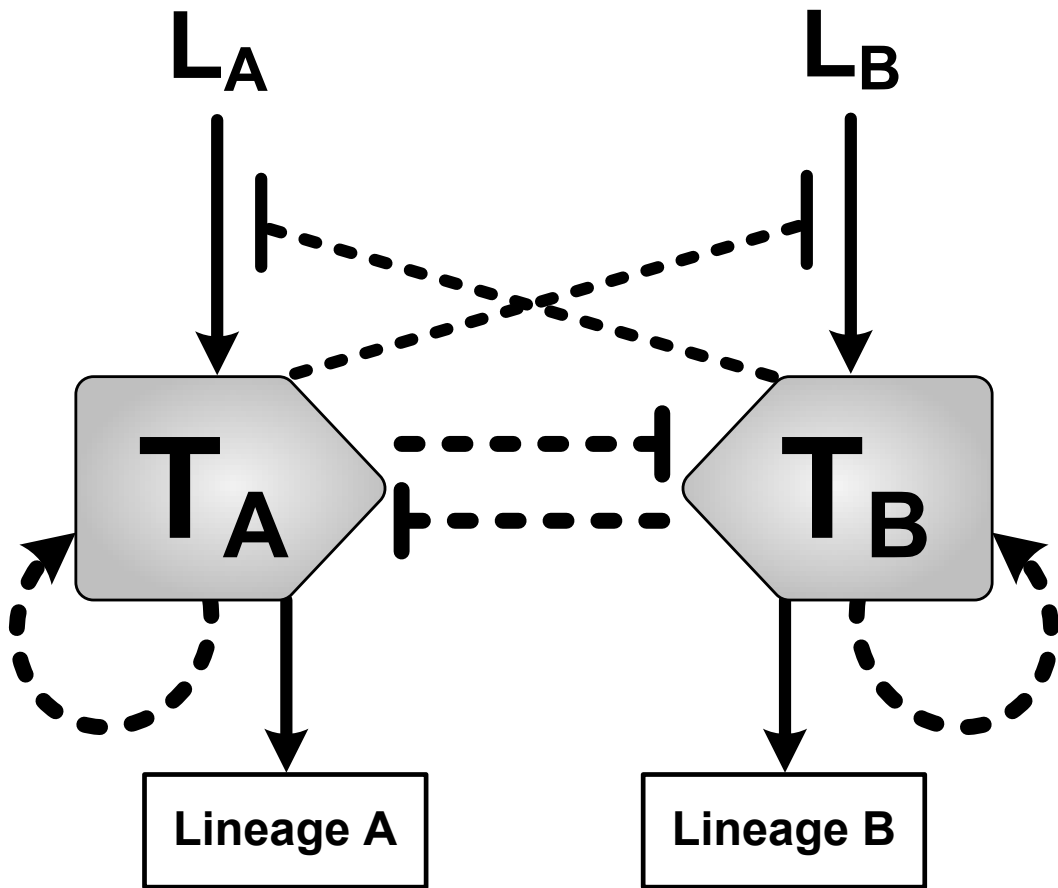
Figure 3-6: **ECAA systems model augmented to account for extrinsic cues.** The CAA systems model was expanded into ECAA to allow extrinsic cues, L_A, L_B to promote the synthesis of their cognate transcription factors. Extrinsic cue-mediated synthesis is inhibited by the opposing transcription factor.
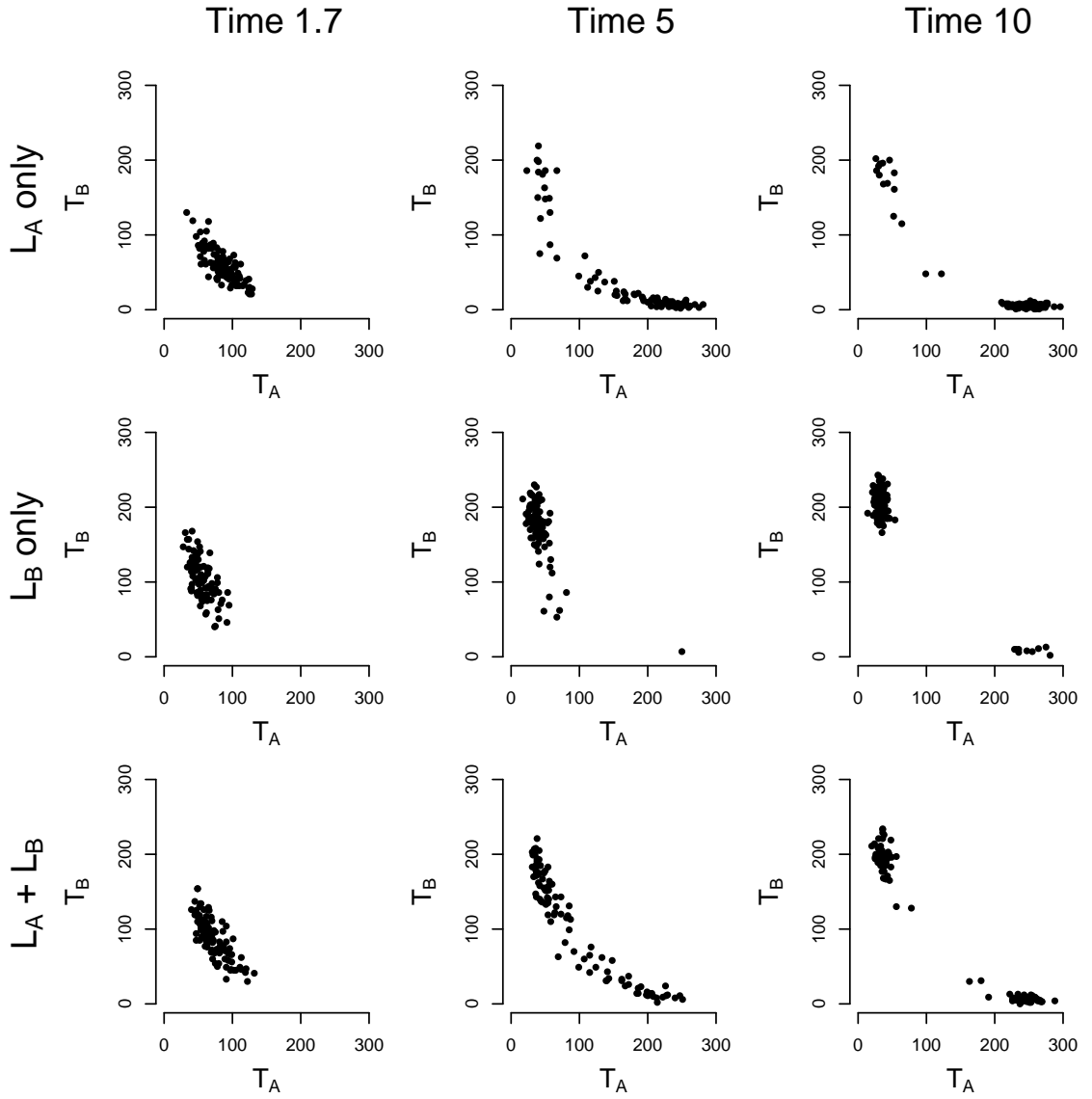
Figure 3-7: **Simulated master trancription factor dynamics under different treatment regimes.** The ECAA model was stochastically simulated under different ligand concentrations: L_A only ($L_A = 1000$, $L_B = 0$), L_B only ($L_A = 0$, $L_B = 1000$), $T_A + T_B$ ($L_A = 1000$, $L_B = 1000$). For each treatment condition, 100 separate simulations were performed, yielding 100 trajectories. Plots show snapshots of trajectories at different times. Each simulation was started in the bipotent state, at $T_A = 60$, $T_B = 70$. Simulation results show that under the given parameters, the ECAA model yields discrete decisions, which can be influenced by extrinsic cues.
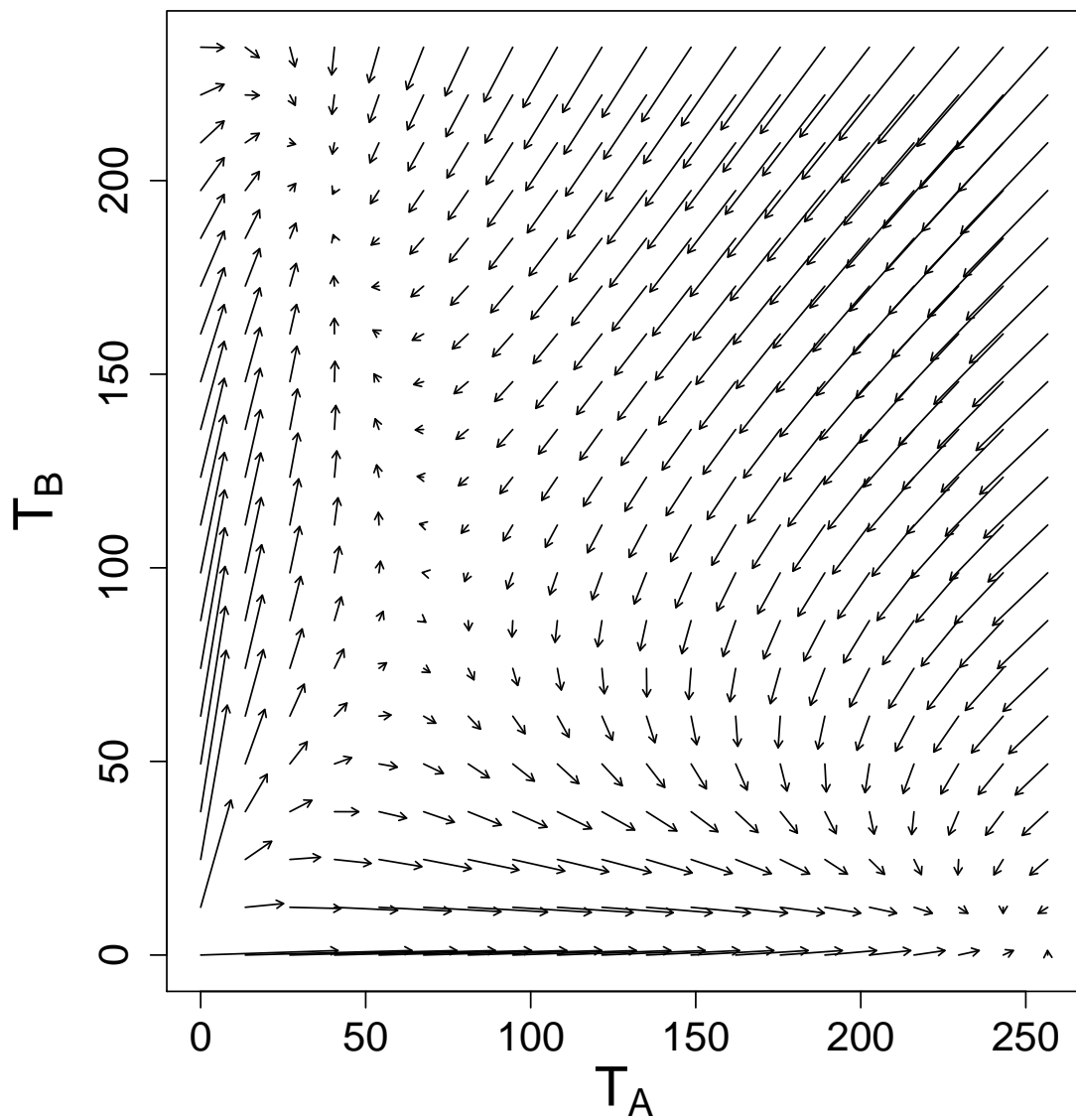
Figure 3-8: **Vector field for the ECAA model under high L_A, low L_B.** Vector field for the ECAA model under $L_A = 1000, L_B = 0.001$.
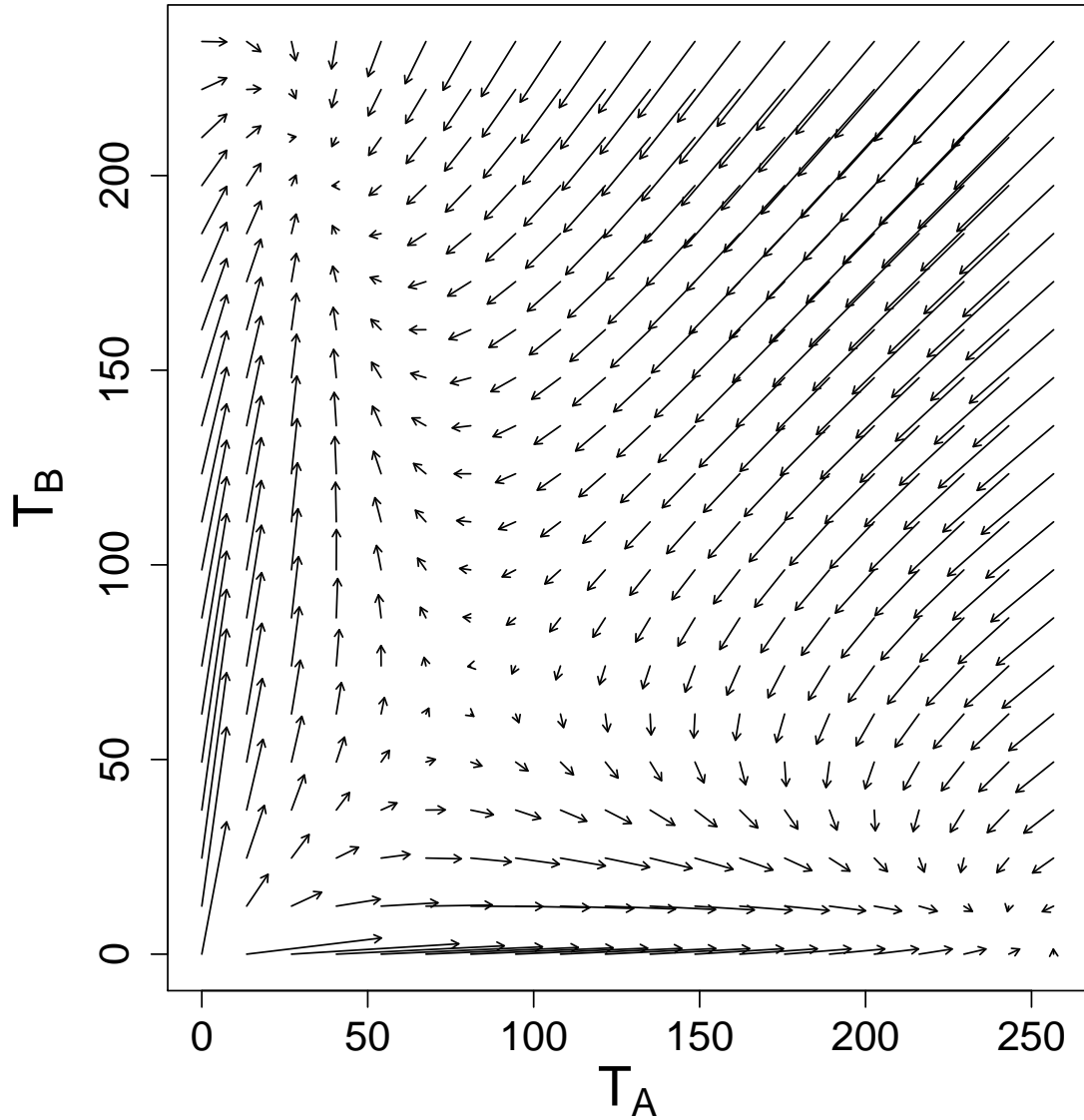
Figure 3-9: **Vector field for the ECAA model under low L_A, high L_B.** Vector field for the ECAA model under $L_A = 0.001, L_B = 1000$.
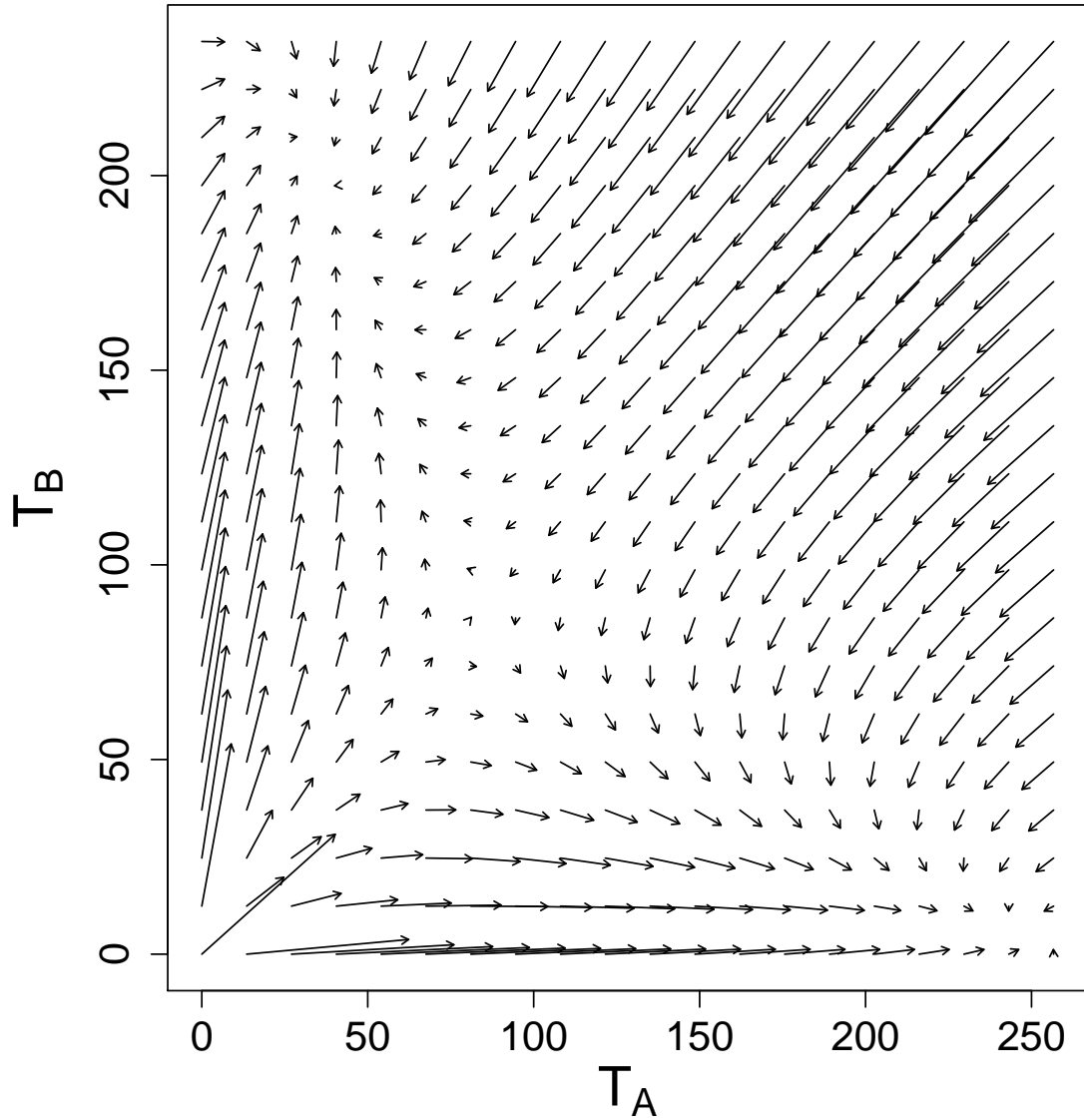
Figure 3-10: **Vector field for the ECAA model under high L_A, high L_B.** Vector field for the ECAA model under $L_A = 1000, L_B = 1000$.
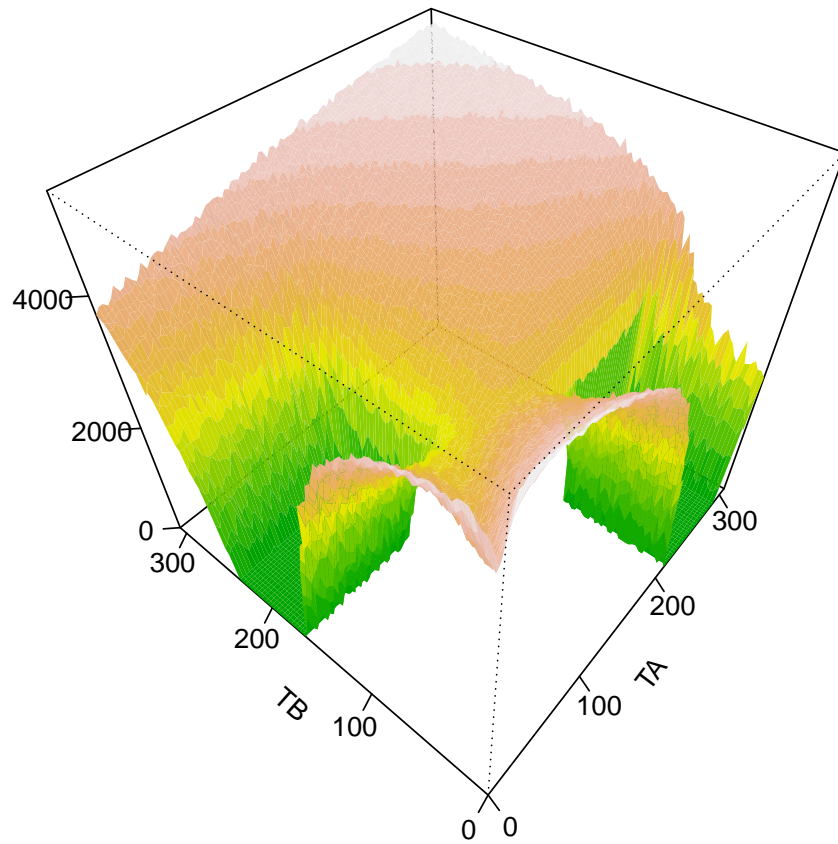
Figure 3-11: **Estimation of energy potential for the augmented CAA model under high L_A, high L_B.** Energy potential was estimated for the augmented CAA model under $L_A = 1000, L_B = 1000$. Simulations were started from a 100 x 100 grid of 10,000 points in (T_A, T_B) space. Each simulation was repeated 100 times. Algorithm details are described in the Methods section.
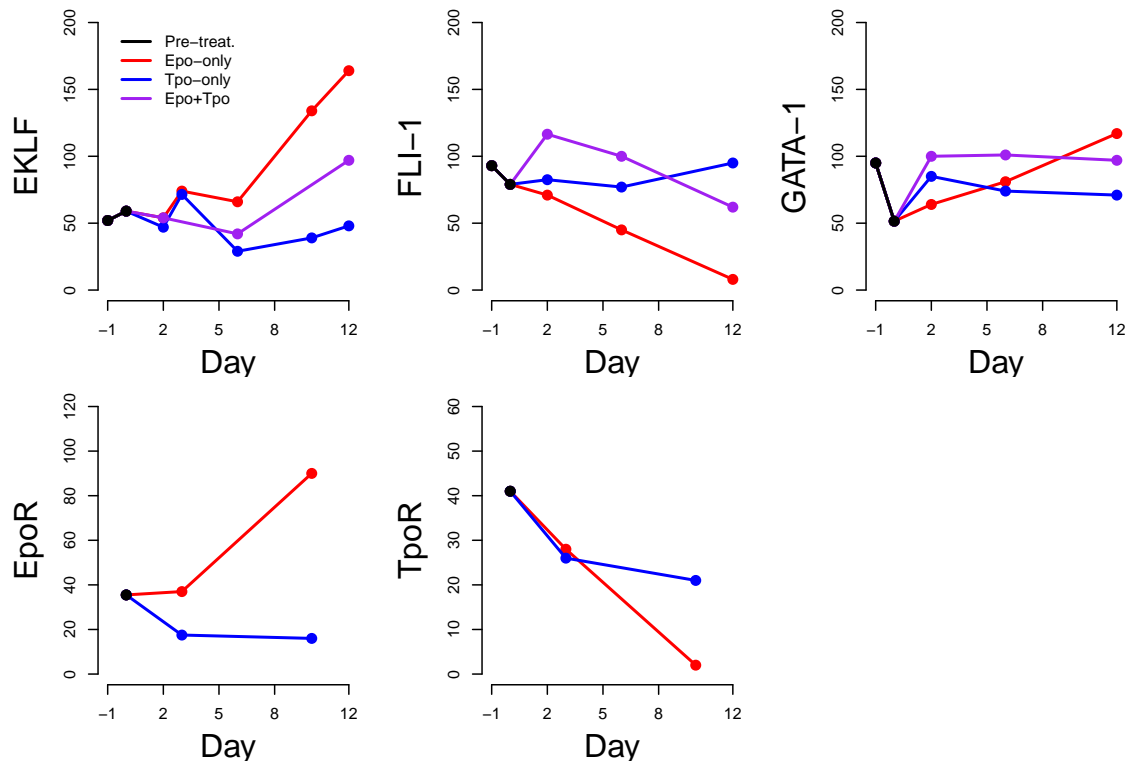
Figure 3-12: **Median transcript counts under different treatment regimes.**
UT-7/GM cells maintained in GMCSF medium were growth-factor starved for 18
hours and subsequently passaged into three different conditions: Epo-only (1 U/mL
Epo), Tpo-only (5 ng/mL Tpo), and Epo+Tpo (1 U/mL Epo, 5 ng/mL Tpo). At
regular timepoints, cells were fixed, and the mRNA FISH method was used to obtain
transcript counts in individual cells. Plots represent the median for each transcript.
Day 0 refers to the day when cells were first treated with cytokines Epo or Tpo.
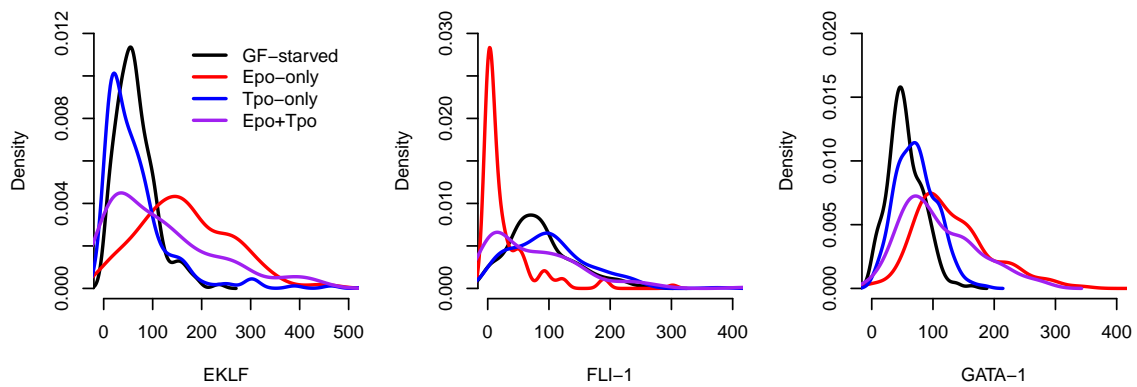
Figure 3-13: **Transcription factor distributions under different treatment regimes on Day 12.** UT-7/GM cells maintained in GMCSF medium were growth-factor starved for 18 hours and subsequently passaged into three different conditions: Epo-only (1 U/mL Epo), Tpo-only (5 ng/mL Tpo), and Epo+Tpo (1 U/mL Epo, 5 ng/mL Tpo). At regular timepoints, cells were fixed, and the mRNA FISH method was used to obtain transcript counts in individual cells. Probability density plots show the distributions of EKLF, FLI-1, and GATA-1 transcripts in individual cells on Day 12; for comparison, transcript counts from Day 0 (GF-starved) are also shown.
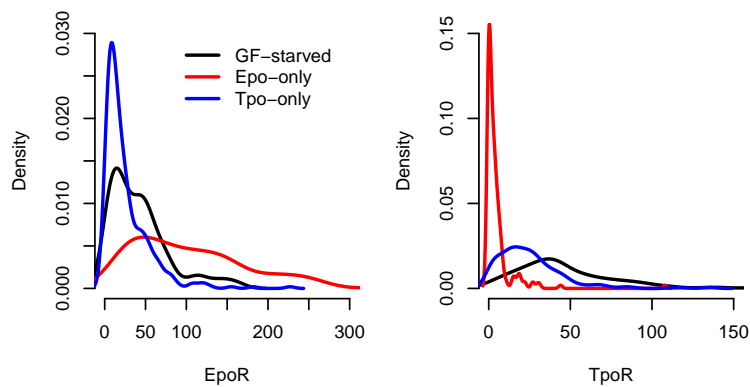
Figure 3-14: **Receptor distributions under different treatment regimes on Day 12.** UT-7/GM cells maintained in GMCSF medium were growth-factor starved for 18 hours and subsequently passaged into two different conditions: Epo-only (1 U/mL Epo) or Tpo-only (5 ng/mL Tpo). At regular timepoints, cells were fixed, and the mRNA FISH method was used to obtain transcript counts in individual cells. Probability density plots show the distributions of EpoR and TpoR transcripts in individual cells on Day 10; for comparison, transcript counts from Day 0 (GF-starved) are also shown.
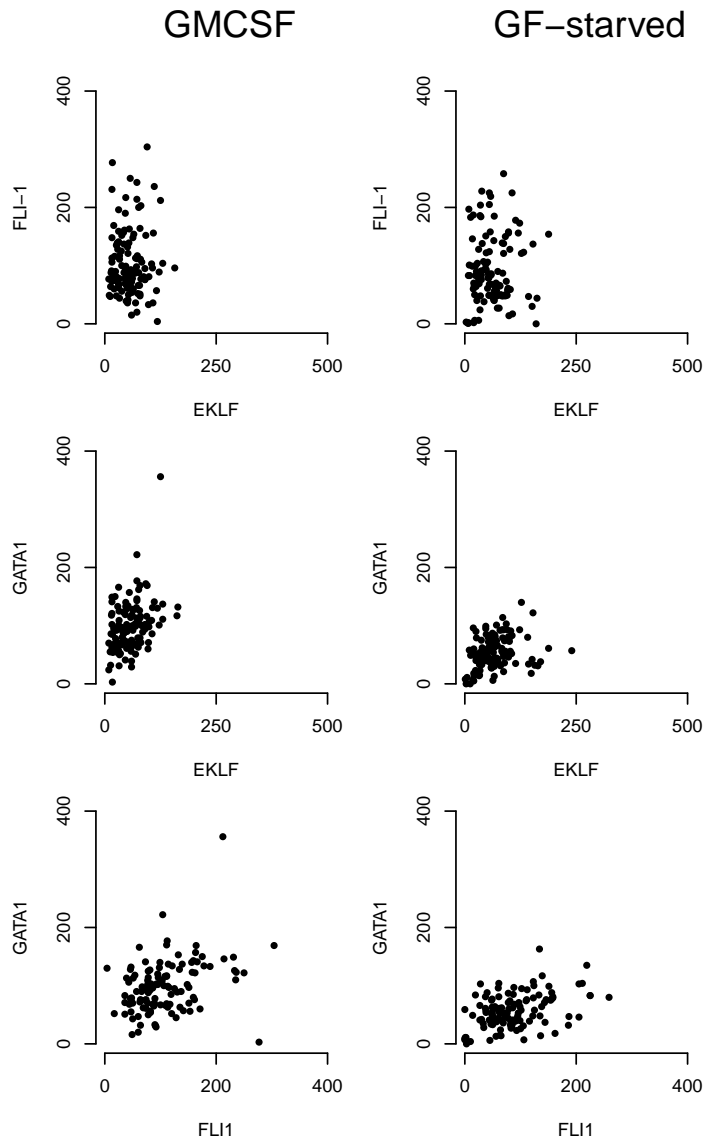
Figure 3-15: **Correlations between EKLF, FLI-1, and GATA-1 levels before treatment.** UT-7/GM cells maintained in GMCSF were growth-factor starved for 18 hours. Cells were fixed before and after starvation and the mRNA FISH method was used to obtain transcript counts for EKLF, FLI-1, and GATA-1 in individual cells. Phase plots show the degree of correlation between transcription factors before treatment with cytokines.
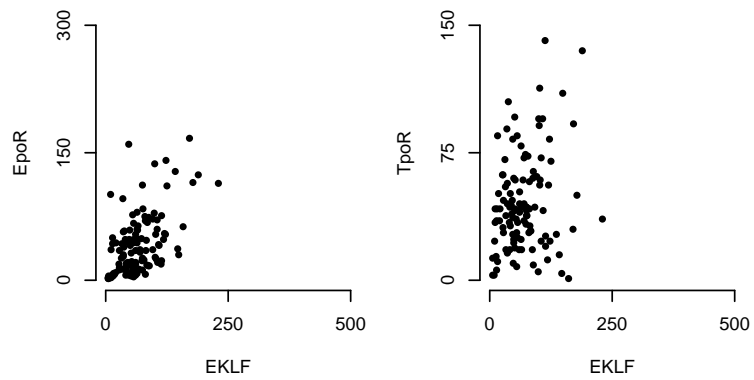
Figure 3-16: **Correlations between EKLF, EpoR, and TpoR levels before treatment.** UT-7/GM cells maintained in GMCSF were growth-factor starved for 18 hours. Cells were fixed before and after starvation and the mRNA FISH method was used to obtain transcript counts for EKLF, EpoR, and TpoR in individual cells. Phase plots show the degree of correlation between EKLF and either EpoR, or TpoR, before treatment with cytokines.
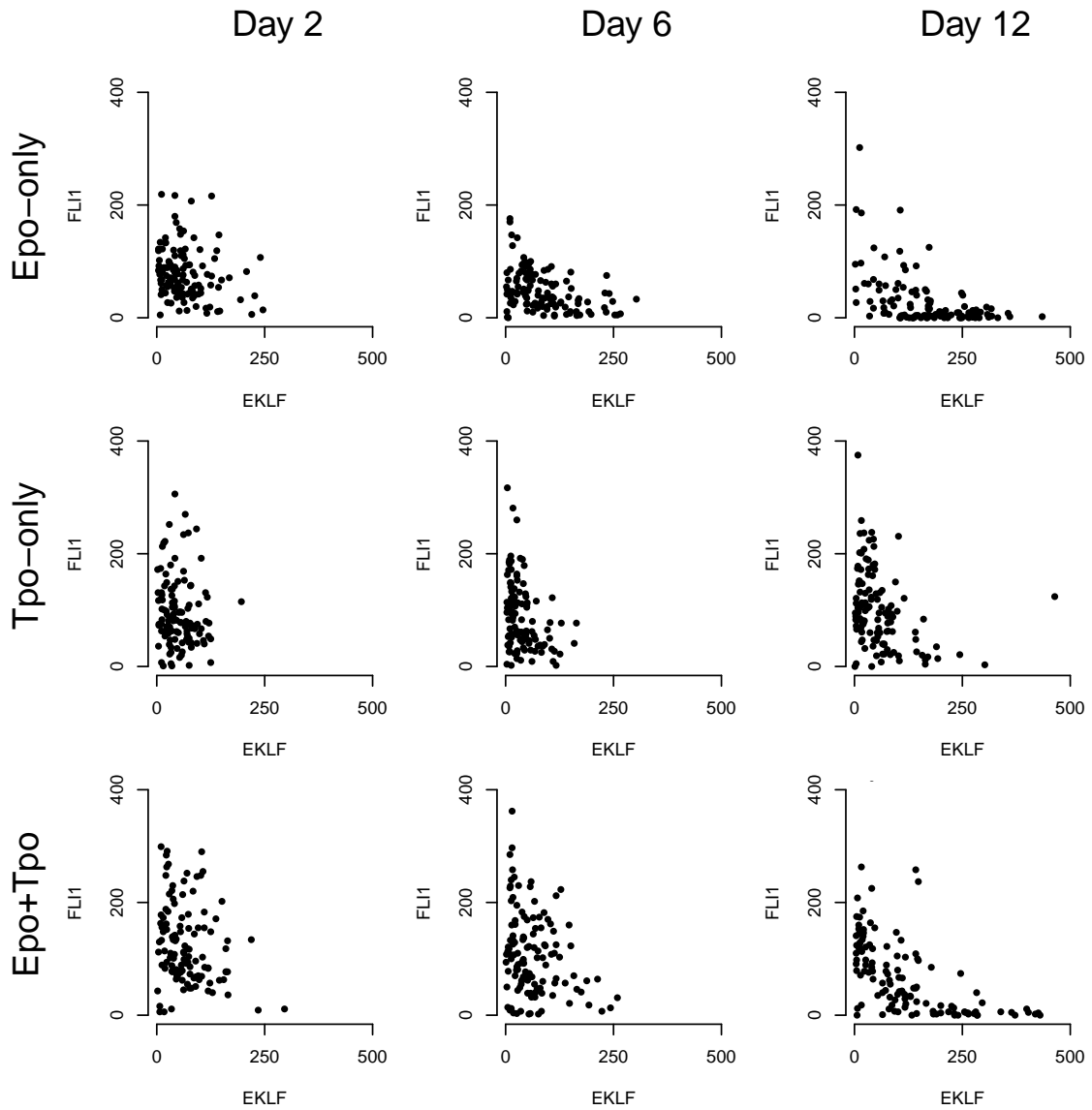
Figure 3-17: **Correlations between EKLF and FLI-1 levels during treatment with Epo, Tpo.** UT-7/GM cells were treated with Epo (1 U/mL), Tpo (5 ng/mL), or both, and the mRNA FISH method was used to obtain transcript counts for EKLF and FLI-1 in individual cells. Phase plots show the degree of correlation between transcription factors at different timepoints during differentiation.
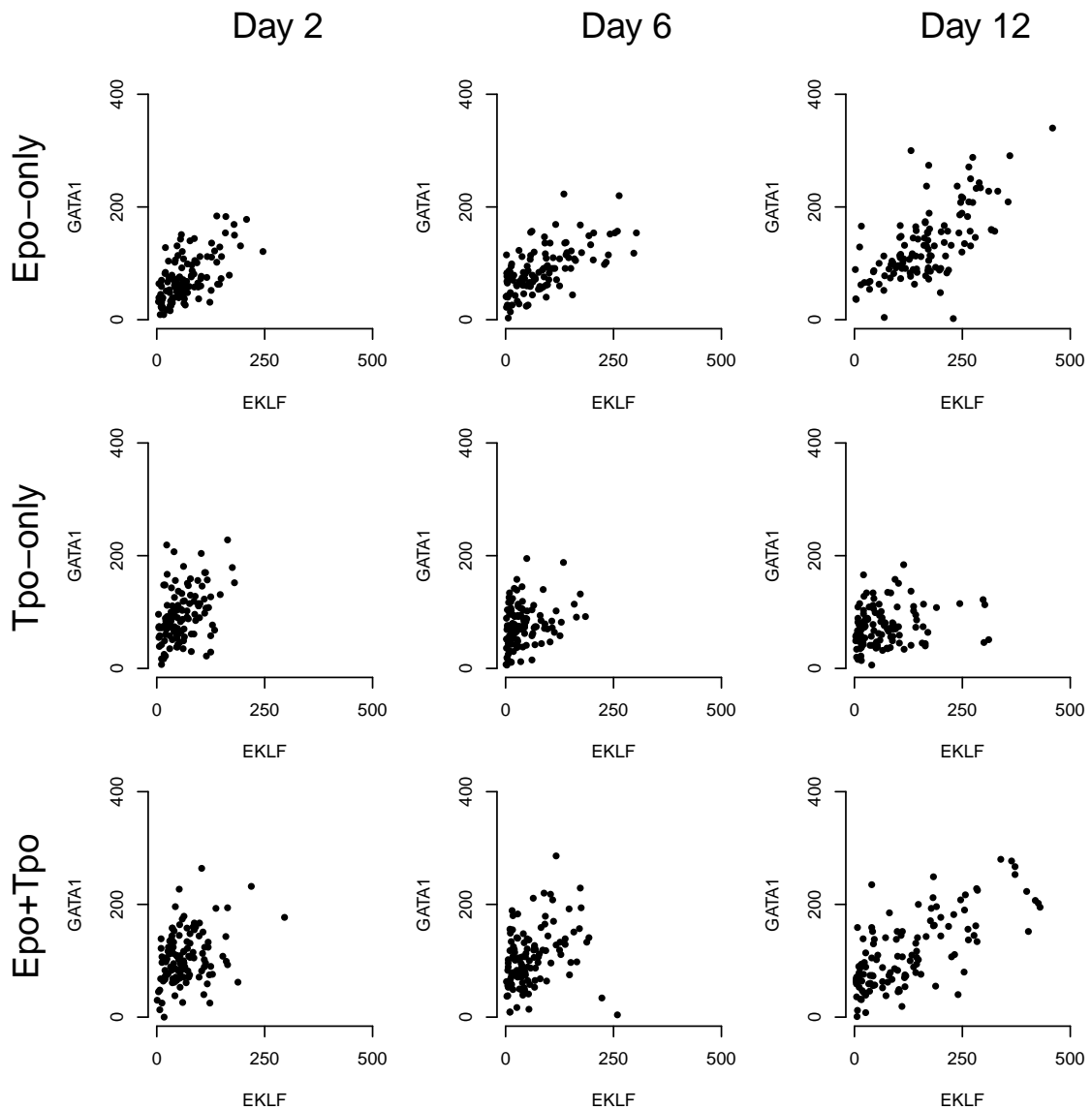
Figure 3-18: **Correlations between EKLF and GATA-1 levels during treatment with Epo, Tpo.** UT-7/GM cells were treated with Epo (1 U/mL), Tpo (5 ng/mL), or both, and the mRNA FISH method was used to obtain transcript counts for EKLF and GATA-1 in individual cells. Phase plots show the degree of correlation between transcription factors at different timepoints during differentiation.
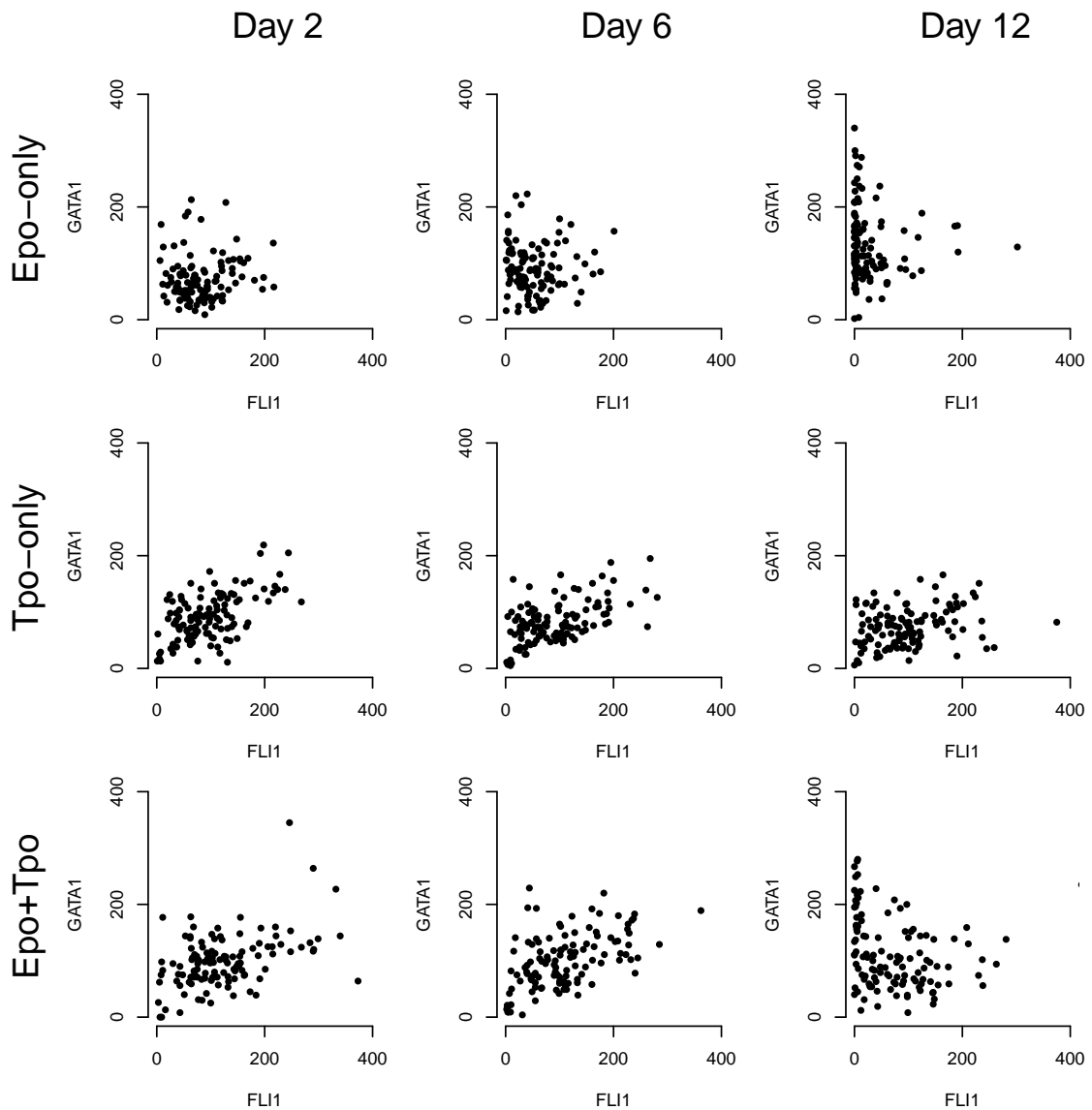
Figure 3-19: **Correlations between FLI-1 and GATA-1 levels during treatment with Epo, Tpo.** UT-7/GM cells were treated with Epo (1 U/mL), Tpo (5 ng/mL), or both, and the mRNA FISH method was used to obtain transcript counts for FLI-1 and GATA-1 in individual cells. Phase plots show the degree of correlation between transcription factors at different timepoints during differentiation.
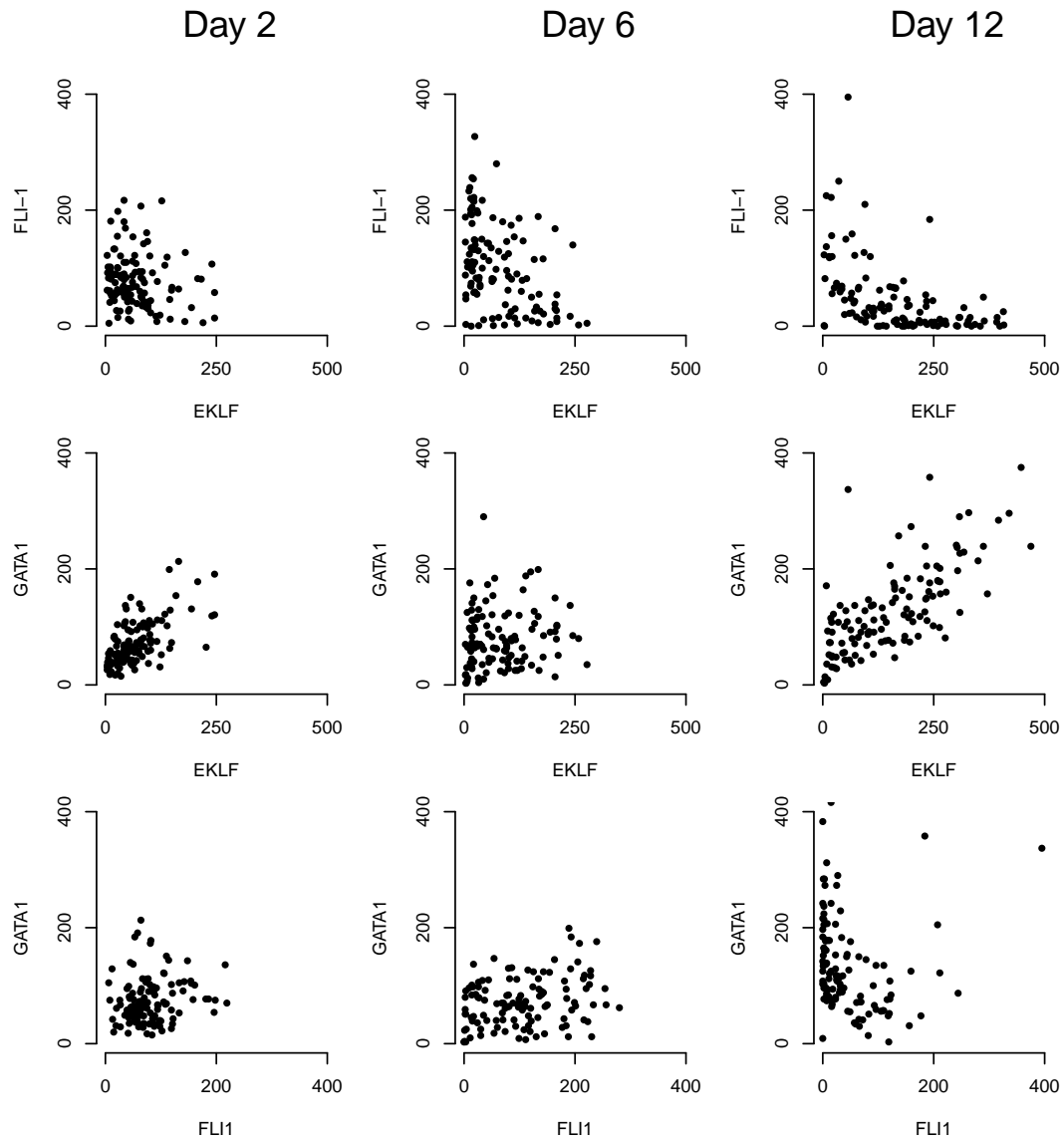
Figure 3-20: **Impact on trancription factor correlations from introduction of Tpo to an Epo-induced culture.** UT-7/GM cells were maintained in Epo (1 U/mL) medium. On Day 3, Tpo (5 ng/mL) was added to the medium. The mRNA FISH method was used to obtain transcript counts for EKLF, FLI-1, and GATA-1 in individual cells. When compared with Epo-only, Tpo-only, and Epo+Tpo treatments, phase plots show that introduction of Tpo leads to an increase in FLI-1 levels on Day 6; however, the distribution on Day 10 suggests that this effect appears to be overcome by cells with above-threshold levels of EKLF.
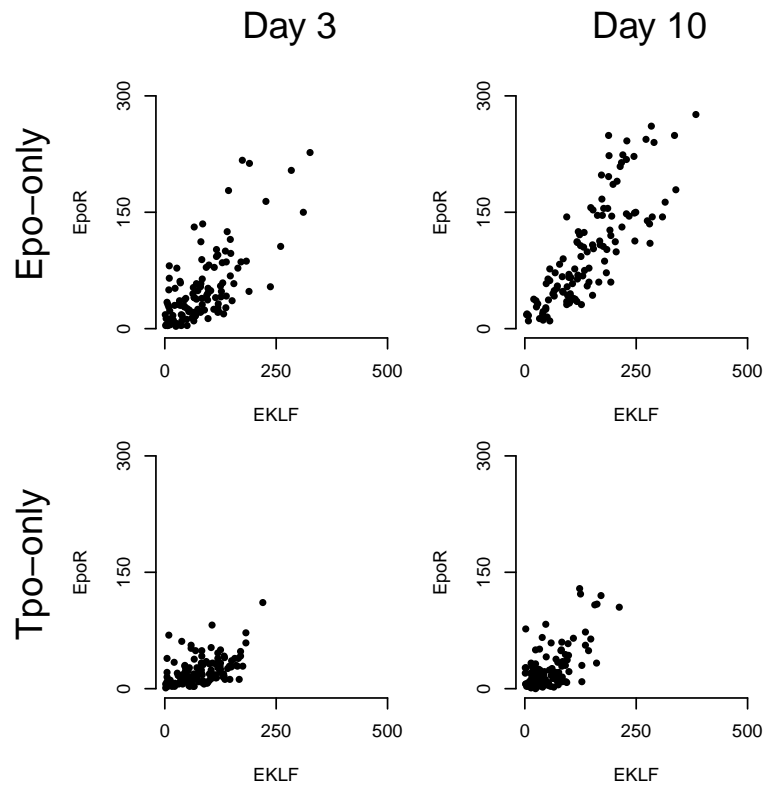
Figure 3-21: **Correlations between EKLF and EpoR levels during treatment with Epo, Tpo.** UT-7/GM cells were treated with Epo (1 U/mL), Tpo (5 ng/mL), or both, and the mRNA FISH method was used to obtain transcript counts for EKLF and EpoR in individual cells. Phase plots show the degree of correlation between EKLF and EpoR at different timepoints during differentiation.
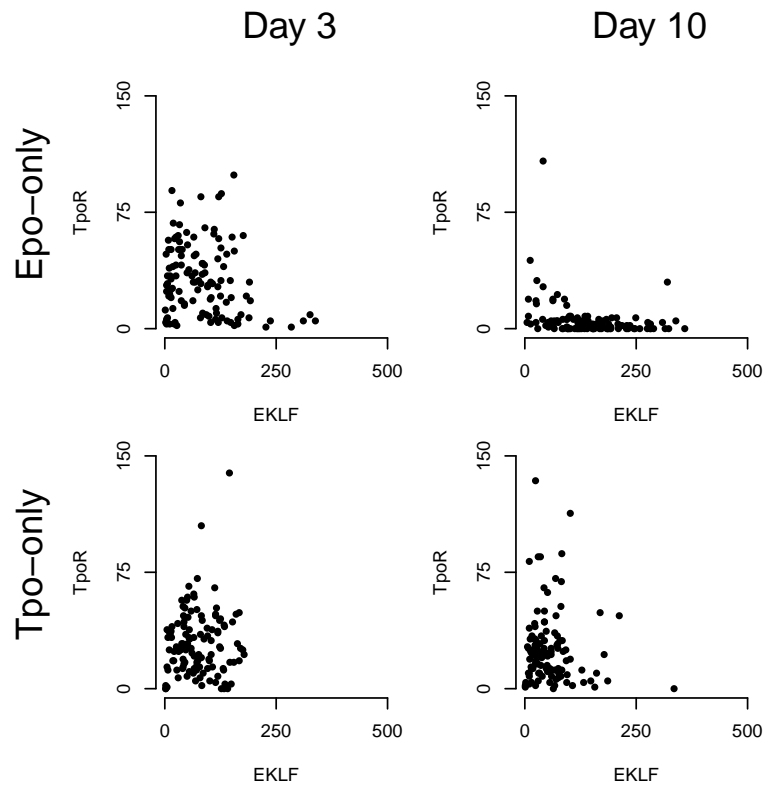
Figure 3-22: **Correlations between EKLF and TpoR levels during treatment with Epo, Tpo.** UT-7/GM cells were treated with Epo (1 U/mL), Tpo (5 ng/mL), or both, and the mRNA FISH method was used to obtain transcript counts for EKLF and TpoR in individual cells. Phase plots show the degree of correlation between EKLF and TpoR at different timepoints during differentiation.

| Genes	Treatment	Day	Test	P-value	ρ 95% C.I.
EKLF, FLI-1	GMCSF	-1	$-$	0.40	(-0.16,+0.23)
EKLF, FLI-1	GF-starved	0	$-$	0.25	(-0.28,+0.16)
EKLF, FLI-1	Epo-only	2	$-$	0.004	(-0.03,+0.05)
EKLF, FLI-1	Epo-only	6	$-$	0.001	(-0.65,-0.33)
EKLF, FLI-1	Epo-only	12	$-$	0.001	(-0.59,-0.27)
EKLF, FLI-1	Tpo-only	2	$-$	0.001	(-0.45,-0.09)
EKLF, FLI-1	Tpo-only	6	$-$	0.001	(-0.38,+0.01)
EKLF, FLI-1	Tpo-only	12	$-$	0.001	(-0.43,-0.06)
EKLF, FLI-1	Epo+Tpo	2	$-$	0.001	(-0.41,-0.01)
EKLF, FLI-1	Epo+Tpo	6	$-$	0.001	(-0.44, -0.04)
EKLF, FLI-1	Epo+Tpo	12	$-$	0.001	(-0.84,-0.65)
EKLF, GATA-1	GMCSF	-1	$+$	0.001	(+0.30,+0.62)
EKLF, GATA-1	GF-starved	0	$+$	0.001	(+0.07,+0.47)
EKLF, GATA-1	Epo-only	2	$+$	0.001	(+0.56,+0.79)
EKLF, GATA-1	Epo-only	6	$+$	0.001	(+0.49,+0.72)
EKLF, GATA-1	Epo-only	12	$+$	0.001	(+0.47,+0.72)
EKLF, GATA-1	Tpo-only	2	$+$	0.001	(+0.12,+0.49)
EKLF, GATA-1	Tpo-only	6	$+$	0.001	(+0.14,+0.51)
EKLF, GATA-1	Tpo-only	12	$+$	0.001	(-0.04,+0.36)
EKLF, GATA-1	Epo+Tpo	2	$+$	0.001	(+0.18,+0.53)
EKLF, GATA-1	Epo+Tpo	6	$+$	0.001	(+0.11,+0.51)
EKLF, GATA-1	Epo+Tpo	12	$+$	0.001	(+0.5,+0.78)
FLI-1, GATA-1	GMCSF	-1	$+$	0.001	(+0.16,+0.56)
FLI-1, GATA-1	GF-starved	0	$+$	0.001	(+0.19,+0.56)
FLI-1, GATA-1	Epo-only	2	$+$	0.008	(-0.15,+0.26)
FLI-1, GATA-1	Epo-only	6	$+$	0.878	(-0.3,+0.09)
FLI-1, GATA-1	Epo-only	12	$+$	0.990	(-0.29,+0.06)
FLI-1, GATA-1	Tpo-only	2	$+$	0.001	(+0.13,+0.54)
FLI-1, GATA-1	Tpo-only	6	$+$	0.001	(+0.21,+0.59)
FLI-1, GATA-1	Tpo-only	12	$+$	0.001	(+0.17,+0.54)
FLI-1, GATA-1	Epo+Tpo	2	$+$	0.001	(+0.21,+0.54)
FLI-1, GATA-1	Epo+Tpo	6	$+$	0.001	(+0.30,+0.64)
FLI-1, GATA-1	Epo+Tpo	12	$+$	0.999	(-0.54,-0.15)

Table 3.1: **Statistical tests for correlations between pairs of transcription factors.** Under Test, '-' indicates a test for anti-correlation, while '+' indicates a test for positive correlation. P-value denotes the significance of the corrleation, and ρ confidence interval indicates the estimated magnitude of ther correlation. See Methods for details.

Genes	Treatment	Day	Test	P-value	ρ 95% C.I.
EKLF, EpoR	GF-starved	0	+	0.001	(+0.28,+0.62)
EKLF, EpoR	Epo-only	3	+	0.001	(+0.45,+0.72)
EKLF, EpoR	Epo-only	10	+	0.001	(+0.75,+0.88)
EKLF, EpoR	Tpo-only	3	+	0.001	(+0.31,+0.66)
EKLF, EpoR	Tpo-only	10	+	0.001	(+0.52,+0.76)
EKLF, TpoR	GF-starved	-1	−	0.988	(-0.09,+0.35)
EKLF, TpoR	Epo-only	3	−	0.029	(-0.31,+0.11)
EKLF, TpoR	Epo-only	10	−	0.001	(-0.51,-0.13)
EKLF, TpoR	Tpo-only	3	−	0.477	(-0.22,+0.17)
EKLF, TpoR	Tpo-only	10	−	0.730	(-0.34,+0.07)

Table 3.2: **Statistical tests for correlations between EKLF, EpoR, and TpoR.** Under Test, '-' indicates a test for anti-correlation, while '+' indicates a test for positive correlation. P-value denotes the significance of the corrleation, and ρ confidence interval indicates the estimated magnitude of ther correlation. See Methods for details. See Methods for details.

Gene	Probe
EKLF	ggctgcctcgtgaactctga
EKLF	tacccggacagtagcccgta
EKLF	tactgaggcgccgggtacat
EKLF	gaagagctggaagtgccctt
EKLF	caaacaactcaggaaggagg
EKLF	acacctggatcctctgcagt
EKLF	acgaacgtcggcctcgcttg
EKLF	agctcttggtgtagctcttg
EKLF	cgcagatgcgccttcaggtg
EKLF	cttcccacgtgcaggcgtat
EKLF	cccgtgtgtttccggtagtg
EKLF	caccacttgaggaagtcatc
EKLF	cttcatgtgcaaggccaggt
EKLF	tcactaggagagtccaagtg
EKLF	tccattcgtgggaaaaccac
EKLF	gatctttgggaacgcgagtc
EKLF	tctatgggtccgtgtttgat
EKLF	gattttccgtaagaggctcc
EKLF	tttgcacgacagtttggaca
EKLF	tttggcggtctgtctcactg
EKLF	gagtgtccactgagtccgtt
EKLF	ggtctctgggaagcctcatc
EKLF	atgtcctgcgcctcttcgga
EKLF	ttcaggagccgctttctaga
EKLF	atatcagccacaataaggga
EKLF	ggacccataaccattgacag
EKLF	tcctcagacttcacgtggag
EKLF	catatgcgcccagagtctcg
EKLF	tgatcctccgaacccaaaag
EKLF	cccgggtacaccggttgcag
EKLF	gtccgcgggaagtagccacc
EKLF	gacgccgcaggcactgaaag

Table 3.3: Oligonucleotides for mRNA FISH targeting the human EKLF transcript.

Gene	Probe
EpoR	ccagcgagcaggagacaaag
EpoR	cgtctaggagcactacttca
EpoR	tagcggatgtgagacgtcat
EpoR	aggttgctcagcacacactc
EpoR	gaagctcggctcagccatac
EpoR	ctaggcgtcagcagcgacac
EpoR	cagatcttctgcttcagagc
EpoR	gtgaagaggccttcaaactc
EpoR	cagctggaagttacccttgt
EpoR	cagccatcattctggtacag
EpoR	agaggacttccagggaagca
EpoR	ctttgctctcgaacttgggg
EpoR	agataggtatcctgggcatg
EpoR	cagcaaccatttgtccagca
EpoR	ttcatccatggccactatgt
EpoR	cacaaggtacaggtactta
EpoR	ttgagatgccagagtcagat
EpoR	gagtcccctgagctgtagtc
EpoR	taggggccatcggataagcc
EpoR	ggctgttctcataagggttg
EpoR	taagagcaagccacatagct
EpoR	gtgaagcacagaagctcttc
EpoR	taggagaagctgtagttgcc
EpoR	ttccatggctcatcctcgag
EpoR	taggcagcgaacaccagaag
EpoR	cacgaagctcgacgtgtcgg
EpoR	ctgtgacgcgcaactctagg
EpoR	gtggatgacacggtgatatc

Table 3.4: Oligonucleotides for mRNA FISH targeting the human EpoR transcript.

Gene	Probe
FLI-1	attataggccagcagtgaac
FLI-1	atgttgggcttgcttttccg
FLI-1	ccggctcagcttgtcgtaat
FLI-1	cactttggtcataatgtttt
FLI-1	tgtaagcatatcttttgccg
FLI-1	ggtacttgtacatggacgac
FLI-1	acaaagttcaccttctgctg
FLI-1	gaggtccagtattgtgatgc
FLI-1	agtagtagctgcctaagtgt
FLI-1	ctcaatcgtgaggattggtc
FLI-1	actgagtcataagaagggtc
FLI-1	catgttattgccccaagctc
FLI-1	gactttgttgaggccagaa
FLI-1	tcttactgatcgtttgtgcc
FLI-1	aggatctgatacggatctgg
FLI-1	ctcccaggtgatacagctgg
FLI-1	gtccgtcattttgaactccc

Table 3.5: Oligonucleotides for mRNA FISH targeting the human FLI-1 transcript.

Gene	Probe
GATA-1	caagctcagtgtgatcccag
GATA-1	tgggacacacagttgaggca
GATA-1	cagatcttccacggcctggg
GATA-1	atgaaggcagtgcaggtccc
GATA-1	cataagcactattggggaca
GATA-1	gtactggaaaagtcagggcc
GATA-1	gagagttccacgaagcttgg
GATA-1	ctgttgctccgcagttcaca
GATA-1	ataggtagtggcctgtcctg
GATA-1	ctgcccattcatcttgtgat
GATA-1	cccgtttactgacaatcagg
GATA-1	atggagcctctggggattaa
GATA-1	cagttggtgcactgagtacc
GATA-1	agcttgtagtagaggccgca
GATA-1	ctcatgagctgagcggagcc
GATA-1	ctctggaggccatgctctgt
GATA-1	ctttgaaggttcaagccagg
GATA-1	gaggacaccagagcaggatc
GATA-1	aagaaaacccctgattctgg
GATA-1	gaggaagctgctgcatccaa
GATA-1	ctcagcgtccctgtagtagg
GATA-1	gtacacctgaaagactgggg
GATA-1	cctccatacagttgagcaat
GATA-1	cagccggcatatggtgagcc

Table 3.6: Oligonucleotides for mRNA FISH targeting the human EpoR transcript.

Gene	Probe
TpoR	ctgacttgggccaggttttg
TpoR	tacactgtccacaaagagga
TpoR	catggccttgatgatactgg
TpoR	ctgatctgaagttcccctgg
TpoR	ctgatttctggagctggctc
TpoR	ggagttcgtacctcaggaaa
TpoR	caatcagctgtatgaccgtg
TpoR	acatggagactggtccagag
TpoR	gtctgctttggtccatcttg
TpoR	agagctgaagcttctctact
TpoR	agggagatcccatcaggttc
TpoR	ccagcaaggagacatcttgg
TpoR	taaagcattgcagtccaagt
TpoR	ggtaacattcttcaggtcca
TpoR	ctagcatggtcctgttgctg
TpoR	ctgtggtagaagaagccttg
TpoR	ttcgtgacttgaagtggcag
TpoR	ccaggtagctgtgaacagta
TpoR	caattccagatgcccactgg
TpoR	ataacaggtctcttgggctg
TpoR	cttctcctgtgtatcggagt
TpoR	agcaccttccagtcctgatg
TpoR	cttcaggggctctgagtctg
TpoR	ctgtaaacggtagcgagatc
TpoR	ggtcaccaaggagatccagg
TpoR	ctgaggcccagcactagatg
TpoR	cctcaaatgttcgggagaaa
TpoR	tcatcccagaagcaagtgag
TpoR	catacagcagctggtatgtc
TpoR	catgctctgggaactcaggg
TpoR	acgtatcgggttccaaagtg
TpoR	ggaaagaagagacgcacttc

Table 3.7: Oligonucleotides for mRNA FISH targeting the human TpoR transcript.

Chapter 4

Engineered decision-making in a modular synthetic circuit

What I cannot create, I do not understand.

– Richard Feynman

The cross-antagonism auto-regulation (CAA) network motif, presented in Chapter 3 as a systems model for bipotent progenitor commitment, is believed to be the core driver of decision-making behavior in diverse cellular contexts. A detailed mechanistic understanding of how CAA circuits inform cellular decisions requires the ability to systematically perform uncoupled perturbations, which is impractical in natural systems. Using the engineering design strategy of re-use, which advocates assembling complex systems by using multiple instances of robust building blocks, we constructed a CAA circuit in the yeast *Saccharomyces cerevisiae* composed of fusions of a single core protein. This circuit can provide a thorough understanding of cellular decision-making in an insulated, highly tunable system, and further enables rational design of gene networks to engender specific decision-making behaviors.

4.1 Synthetic biology as a means to study cellular behaviors

The fundamental premise behind the systems approach to understanding biological phenomena is that complex, emergent behaviors at the cellular, organismal, and even ecological levels can be explained by relatively simple interactions at the molecular level. From this perspective, much of the research in the field of systems biology is aimed at building models that not only succinctly capture our understanding of biological processes, but more importantly, can be used to predict system behavior for a different configuration of inputs. However, building a representative model of even a very simple, fundamental process usually requires many iterations of the model formulation-prediction-testing cycle.

Mathematical modeling, the first step in the cycle, can by itself be quite illuminating to our understanding of a particular biological system; the process of compiling observations about the system, and rigorously translating them into a set of components and well-defined events can often reveal gaps and inconsistencies in our knowledge. At the prediction step, simulation of a mathematical model can yield specific, quantitative hypotheses about the behavior of the biological system that can be difficult to obtain intuitively, especially if a large number of components, or non-linear phenomena are involved. The third and final step involves performing experiments that either support or invalidate the model.

It is widely acknowledged that the hypothesis-testing step is typically far more challenging than the preceding steps. In most systems biology studies, this experimental step boils down to two sub-steps: first, making a precise perturbation to the system, and second, quantitatively measuring the response of the system. Recent advances in traditional assay technologies, such as enhanced quantitation in Western blotting and PCR, the broader accessibility of single-cell assays, such as flow cytometry and mRNA FISH, and high-throughput data-collection enabled by dramatic

scale-ups in sequencing technologies as well as scale-downs through miniaturization (e.g., n-well plate technologies), all facilitate quantitative interrogation of biological systems at many levels, and have hence rendered the measurement step significantly easier.

However, the perturbation step remains hindered with two types of challenges. First, despite advances in genome editing technologies, the ability to make precise genetic changes in cells of higher organisms remains difficult. For example, in the context of hematopoietic lineage commitment, a thorough study of the strength of repression exerted by a master transcription factor could entail replacing the repressive domain in the relevant gene's sequence with a stronger or weaker repression domain. Second, and more important, the strong coupling between pathways in natural systems makes it virtually impossible to perturb a specific component without eliciting unintended effects. For example, it would be difficult to modify a receptor to specifically study cytokine signal processing in hematopoietic lineage commitment, since cytokines also regulate cell viability.

Even if precise genetic manipulations are feasible, and reliable, quantitative, and single-cell assays are available, hypothesis-driven experiments aimed at studying a particular biological system can at best demonstrate that experiment results are consistent with the model; with this approach, one cannot definitively show that a given network can in fact recapitulate the dynamic behavior ascribed to it by the model.

The goal of demonstrating that disparate biological components assembled in a user-defined network topology could yield a desired dynamic behavior in living cells was the major impetus that led to the creation of the nascent field of synthetic biology. Ignited by the implementation of a genetic toggle switch [Gardner et al., 2000] and a simple oscillator [Elowitz and Leibler, 2000], synthetic biology validates the systems approach to studying cellular function, and represents a complementary approach to studying biological systems. While traditional approaches tend to be centered on studying a specific protein, nucleic acid, or other component in detail, the systems approach emphasizes the interactions over the identity of the components; in other

words, the systems approach focuses on the 'wiring diagram.'

Synthetic biological approaches have in many instances enabled the detailed, systematic study of important processes including transcriptional, post-transcriptional, and post-translational regulation, receptor signaling [Hammer et al., 2006, Mukherji and van Oudenaarden, 2009, Levskaya et al., 2005], as well as dynamic behaviors such as oscillations, ultrasensitivity, and bistability [Elowitz et al., 2002, O'Shaughnessy et al., 2011, Hooshangi et al., 2005, Gardner et al., 2000].

4.2 Challenges in synthetic biology

In addition to enabling a new platform for studying biological processes, research in synthetic biology can lead to many potential applications in industry; indeed, its proponents argue that the field is poised to revolutionize energy production, agriculture, and the materials and drug industries, among others. However, the initial excitement in this area has not been borne out by very many tangible successes; in general, extending early strategies and tools used to build very simple circuits to the implementation of larger, more complex systems has proved to be difficult. Here we briefly review some of the important factors that must be overcome or addressed if synthetic biology is to evolve into an engineering discipline [Andrianantoandro et al., 2006, Endy, 2005].

4.2.1 Lack of well-characterized parts

Perhaps the most obvious obstacle to implementing complex networks in living systems is that there are very few well-characterized biological components available to the community [Pleiss, 2006, Serrano, 2007]. For example, to put a gene of interest under the control of a drug-inducible transcriptional activator, one would ideally have access to information on:

- the composition of the activator protein (i.e., its constituent domains and their

functions),

- the DNA sequence motif to which the transcription factor binds, to be able to control the intended target, and also to determine whether existing instances of this sequence in the host organism can create unintended targets,

- the preferred location of the sequence binding motif, with respect to its distance from the gene, and from the promoter and other regulatory regions,

- the rates of binding and unbinding from DNA, to estimate the level of transcription factor needed to actuate a response,

- the strength and mechanism of activation exerted, to estimate the fold-expression of the target, and determine whether this mechanism can be utilized in the host,

- the drug binding rates or the dose-response curve, to determine the concentration of drug needed to elicit or block the activation;

- and whether any of the above properties are modulated by temperature, pH, or other environmental conditions.

Because very few biological components have been characterized at this level of detail, the assembly of even simple circuits necessitates multiple rounds of optimization by trial and error. Furthermore the introduction of each additional component yields a series of new possible interactions, and makes combinatorial optimization substantially more difficult.

4.2.2 Modularity and composability

Modularity is a central theme in human-engineered systems. Complex tasks are divided into smaller, simpler ones that can be performed by individual modules of interacting components, and interactions between modules are allowed only through minimal, well-defined interfaces. This enables easier identification of failure points and

replacement of faulty modules. Importantly, strong modularity and weak coupling also allows engineered parts to be highly composable in that it is relatively easy to link modules together in a new network or structure. Although biological processes are believed to be highly modular [Hartwell et al., 1999], they tend to also be highly coupled in that many components participate in multiple processes. Furthermore, while whole modules have been demonstrated to function heterologously [O'Shaughnessy et al., 2011], connecting these modules to modules from other organisms via non-transcriptional links can at best entail optimization of binding interactions, and at worst necessitate the engineering of new interactions.

4.2.3   Host compatibility

Incompatibility with the host organism is a frequent reason for the failure of a part or module; indeed, heterologously expressed components almost always place a metabolic load, and yield a noticeable effect on the host's growth. For instance, a transcription factor that binds to its target DNA sequence with high specificity in its native environment may bind promiscuously in the context of a larger genome, and perhaps lead to toxic, spurious activation of genes. Host incompatibility can also arise if the foreign component activates a defense mechanism, leading to its inactivation by the host, or if the component functions only in the context of a specific environment; as in the case of a cell-surface receptor that requires a specific membrane architecture to function [Andrianantoandro et al., 2006].

4.2.4   Orthogonality

Components used in electronic systems are highly orthogonal in that multiple instances of the same component can exist within a system without interfering with each other. This property promotes the design and manufacturing of very well-characterized, simple components that can be extensively re-used within a single system. For example, insulating wires can allow multiple instances of an AND gate

chip to function properly and without interference. This property is difficult to achieve in the biological context, where, to a rough approximation, components constantly come into contact with each other in the intracellular milieu. For example, if a transcription factor is set up to activate a particular gene, another instance or copy of the same transcription factor cannot of course be set up to activate a second gene, without activating the first one [Rao, 2012].

New strategies that address these critical challenges can have a significant impact on our ability to construct complex, reliable systems, and can hence enable important new applications in science and industry.

4.3   Synthetic decision-making circuit

We designed and constructed a synthetic decision-making circuit in yeast, and this effort allowed us to address two important topics. First, we implemented the CAA topology to facilitate a detailed understanding of the mechanisms by which this network confers discrete decisions. Second, we adopted the engineering design strategy of component re-use, and in effect demonstrate that a four-protein, four-interaction network can be constructed using orthogonal variants of a single core protein.

Specifically, our circuit offers the following advantages:

- the circuit provides a direct test of whether the CAA architecture alone can yield discrete decisions in living systems,

- the core protein, TetR, is not native to yeast, and hence facilitates the study of the CAA circuit in an insulated environment,

- the modularity of the system enables focused genetic perturbations including modulation of repression strength, promoter architecture, and gene dosage via plasmid copy number, and

- the fusion of fluorescent reporters to the opposing transcription activators enables real-time tracking of 'cell fate' over time.

4.3.1 Design overview

The abstract CAA network topology consists of two transcription factors, T_A and T_B, each promoting its own synthesis, and inhibiting the synthesis of the other. To our knowledge, there are no well-characterized and heterologously expressible transcription factors that can simultaneously exert activation and repression; hence, we represented each of T_A and T_B with an activator-repressor pair. Therefore, the implemented network is composed of four proteins, and four interactions (Fig. 4-1):

- the activator protein of T_A up-regulates its own synthesis,

- the activator protein of T_B up-regulates its own synthesis,

- the activator protein of T_A up-regulates the synthesis of the T_A repressor,

- the activator protein of T_B up-regulates the synthesis of the T_B repressor,

- the repressor protein of T_A inhibits the synthesis of the T_B activator,

- the repressor protein of T_B inhibits the synthesis of the T_A activator,

- the repressor protein of T_A inhibits the synthesis of the T_B repressor,

- and the repressor protein of T_B inhibits the synthesis of the T_A repressor.

4.3.2 Part re-use

In constructing this circuit, we adopted the engineering strategy of re-use, which despite its clear advantages, is not widely used in synthetic biology. As described previously, the synthetic biology community is hindered by a dearth of well-characterized parts that can be used in a wide array of hosts and contexts. This leads to three common problems in the construction of synthetic circuits. First, because the available

parts are by-and-large not orthogonal, meaning that they cannot be assigned separate roles in genotypically identical cells, one can quickly exhaust well-characterized options that are compatible with the target host. Second, each additional part in the network has to be compatible with not only the host, but also with all of the existing parts. Third, and most important in the construction of non-trivial networks, optimizing the interaction between every pair of disparate parts to effectively yield the desired transfer-functions often creates a very challenging, if not entirely impractical, combinatorial optimization task.

Consider these issues in the context of the CAA circuit. From modeling and simulation work in the previous chapter, we learned that despite its robustness in generating discrete decisions, each activation and inhibition interaction has to be within a fairly narrow tolerance of the other. For instance, in the theoretical model, if T_A exerts sufficiently higher activation than its counterpart T_B, the system may be rendered monostable such that each cell settles in a high T_A, low T_B state. Similarly, if the repression exerted by T_A is significantly stronger than that exerted by T_B, the system would be biased towards the A lineage, and rendered monostable. Finally, the activation and inhibition interactions have to be in balance with each other to yield a meaningful contrast between on and off states.

These problems become even more challenging when one considers implementation details. For instance, even if the repressors for T_A, T_B are very well-matched in terms of their DNA-binding on- and off-rates, and their actuating domains, seemingly innocuous differences in localization, transcription, translation rates (e.g. due to the lengths of the genes) can introduce imbalances that are hard to overcome.

To address these challenges, we explored the engineering design strategy of re-use. Accordingly, in our design each of the four proteins is actually a variant and fusion of a single core protein with an actuator domain.

4.3.3 TetR modularity

The Tet repressor, TetR, is an exceptionally well-studied protein that plays a critical role in the conferring of tetracycline-resistance in Gram-negative bacteria. Briefly, in the absence of tetracycline, TetR binds to, and inhibits, an operon containing TetA, a resistance-conferring antiporter; in the presence of tetracycline, TetR un-binds, and allows transcription of TetA [Berens and Hillen, 2003].

The TetR protein is highly modular with three domains: a DNA-binding domain which allows it to bind with high-specificity to the tetO operator sequence, a drug-binding domain where tetracycline and some of its analogs can bind, and a dimerization domain which enables binding of two TetR monomers (Fig. 4-1). In the cellular context, two monomers bind to each other, forming dimers, which can in turn bind to the operator sequence. The addition of tetracycline in the environment leads to a titratable un-binding from the the operator sequence [Berens and Hillen, 2003].

4.3.4 Orthogonal re-use of TetR variants

Rational design, directed evolution, and genotyping experiments have revealed a series of TetR variants [Berens and Hillen, 2003]. In particular, a recent study identified specific mutations in the DNA-binding domain of TetR that change its DNA-binding specificity to a different operator sequence, tetO-4C5G. The new variant is different from the wild-type by only three amino-acids: V36F, E37A, P39K. Importantly, the newly identified DNA-binding domain and tetO-4C5G pair is orthogonal to the original DNA-binding domain and tetO, meaning that at least in bacteria, each pair interferes only minimally with the other [Krueger et al., 2007]. We explored the possibility of expressing both the original wild-type pair, and the new variant pair simultaneously for different functions within the same yeast cell.

An orthogonal DNA-binding specificity potentially allows us to use the two TetR proteins as transcription factors within the same cell, and hence moves us closer to

implementation of the CAA network. However, if expressed within the same cell, a significant proportion of the two TetR monomers would form heterodimers, and lead to an unpredictable combination of sequestration (since one variant can prevent the other from performing its function) and aberrant actuation (since only one monomer in the heterodimer would bind to half of the operator sequence, it is unclear what the overall DNA-binding on- and off-rates would be).

Hence, to prevent hetero-dimerization of tetO and tetO-4C5G variants, an orthogonal dimerization specificity is needed. To address this, we mutated the dimerization domain in the canonical TetR(B) variant to match that found in the TetR(D) variant in nature. This entailed the substitution of four amino-acids: F188H, L192S, I193L, L197F [Schnappinger et al., 1998]. Hence, compared to the canonical TetR, the mutant TetR has seven amino-acid substitutions: V36F, E37A, P39K, F188H, L192S, I193L, L197F. We will refer to the dimerization domain, operator-sequence, and DNA-binding domain combination for the canonical TetR as set 0 (denoted in subscript), and the variant combination as set 1 (Fig. 4-1).

4.3.5   Parts list

Having identified candidate mutations that could yield a variant orthogonal to the wild-type TetR, we proceeded to compile the parts list. The CAA circuit requires pairs of transcriptional activators and repressors. Hence, we fused the VP16 activation domain (from the Herpes simplex virus, [Cress and Triezenberg, 1991]) separately to both $TetR_0$ and $TetR_1$, yielding two potentially orthogonal transcriptional activators, tTA_0 and tTA_1. To enable tracking of transcription factor levels via microscopy and flow cytometry, tTA_0 is additionally fused to GFP, and tTA_1 is additionally fused to mCherry. Both GFP and mCherry were placed at the C-terminus of the original TetR, and at the N-terminus of VP16.

Since the TetR protein exerts transcriptional repression, the $TetR_0$ and $TetR_1$ can be used directly as the two repressors in the CAA circuit; in addition to these

repressors, we constructed another set by fusing the strong SSN6 repression domain from yeast [Bellí et al., 1998] to the C-terminii of $TetR_0$ and $TetR_1$, yielding tTS_0 and tTS_1. Throughout this chapter, we refer to repressor proteins comprising TetR alone, as well as those comprising TetR and SSN6, as tTS, using weak and strong to identify the specific protein.

The canonical method for placing a target gene under the control of TetR involves placement of the full tetO operator sequence an empirically-determined distance (in bases) before the gene. The full tetO operator sequence consists of spacer-separated tandem repeats of a defined, base 19-bp sequence, which itself consists of two palindromic repeats, to accommodate binding of a TetR dimer. In the literature, both 7 (tetO-7x) and 2 (tetO-2x) repeats of the base tetO sequence have been used [Wishart et al., 2006]. In the canonical use, at a given time the TetR system is used to either activate or suppress the transcription of a target gene; hence, tetO-7x is used when a stronger effect is desired. However, since our circuit involves simultaneous activation and repression, we constructed two different versions of both, yielding four different full promoter sequences: $tetO_0$-2x, $tetO_0$-7x, $tetO_1$-2x, and $tetO_1$-7x.

4.3.6   Gene-regulatory network

Having defined a complete parts list, we proceeded to translate the CAA wiring map into a gene-regulatory network (Fig. 4-2). The full circuit is arranged as follows. The transcriptional activator for set 0, tTA_0, is regulated by the $tetO_0$ promoter. Basal transcription and subsequent translation of the tTA_0 gene leads to formation of the tTA_0 protein; in the absence of tetracycline, this protein can bind to tTA_0, and hence promote its own synthesis in a positive feedback loop. Analogous placement of the $tetO_1$ promoter before the tTA_1 gene creates a positive feedback loop for the other variant.

Transcriptional repression, as defined by the CAA circuit is implemented as follows. The tTS_1 gene is placed under control of the $tetO_0$ promoter, and the tTS_0 gene

is placed under control of the $tetO_1$ promoter. This arrangement accomplishes two tasks. First, placement of the tTS genes after promoters in effect couples their expression to the relevant activators; i.e., tTS_1 is transcribed as part of the tTA_0-mediated feedback loop, and tTS_0 is transcribed as part of the tTA_1-mediated feedback loop. Second, after transcription and translation, the tTS_0 protein dimer can bind to the $tetO_0$ promoter, and suppress the activation exerted by the tTA_0-mediated feedback loop. Analogously, the tTS_1 protein dimer can bind to the $tetO_1$ promoter, and suppress the activation exerted by the tTA_1-mediated feedback loop.

Taken together, the circuit is composed of two opposing sides, representing T_A, T_B in the minimal CAA model. T_A consists of tTA_0, tTS_1, and $tetO_0$, while T_B consists of tTA_1, tTS_0, and $tetO_1$ (Fig. 4-2).

4.3.7 Promoter architecture

The delivery of four different proteins, each with accompanying regulatory components including the operator, terminator, etc, lead to unwieldy DNA payloads that have to be delivered to the host. Hence, to allow for convenient, modular testing of different component combinations, we opted to use a bi-directional promoter architecture. The full circuit is delivered on two plasmids, with each side on one plasmid. Starting with the tetO promoter, we placed a minimal CYC1 promoter (containing a TATA box) at the 3'-end. Next, we added the tTA gene to the 3'-end of the CYC1 promoter, and a CYC1 terminator at the 3'-end of the tTA gene. We then placed another copy of the CYC1 promoter to the 5'-end of the tetO promoter. Next, we added the tTS gene to the 5'-end of the second CYC1 promoter. Finally, another copy of the CYC1 terminator was added to the 5'-end of the tTS gene. To allow for proper transcription, all components to the 5'-end of the promoter were placed in a 3'-5' orientation. Sets 0 and 1 were cloned into separate plasmids with HIS, URA auxotrophic markers, respectively (see Fig. 4-15 for an overview; details described in Methods).

4.4 Results and Discussion

To understand how parameters such as repression strength and rates of synthesis affect decision-making in the CAA topology, and to determine the extent to which these perturbations are captured by the systems model, we constructed and transformed several sets of plasmids and analyzed the resulting clones by flow cytometry. The core circuit architecture across all of our experiments consists of two activator genes and two repressor genes arranged in a cross-antagonistic topology, as described above.

First, we constructed a set of plasmids with the 2x version for both $tetO_0$ and $tetO_1$ operators, and with the tTS_0, tTS_1 repressors lacking the SSN6 domain. Second, we modified the previous plasmid set by fusing SSN6 repression domains to the repressors. Third, starting with the second plasmid set, we replaced the 2x tetO operator sites with the 7x version for both $tetO_0$ and $tetO_1$ operators. Fourth, starting with the second plasmid set, we removed the activator genes, creating repressor-only plasmids, which were transformed in the same reaction with the second plasmid set (i.e. containing activator genes as well as repressor genes); this allowed us to further boost repression strengths. Additionally, across all of these experiments, we studied a fifth perturbation: the two plasmids in each experiment integrate in varying numbers, yielding clones with different combinations of plasmid copy-numbers; in the simplified CAA model, this equates to varying the synthesis rates. We attempted to further expand the range of this effect by performing a subset of transformations with multiple concentrations of the two plasmids.

For each of the experiments outlined above, we transformed yeast cells with two chromosomally integrating plasmids simultaneously, carrying Set 0 and Set 1. The transformation process yields colonies on selection plates containing medium lacking histidine and uracil. Cells within each colony are genotypically identical; however, the genotype of two colonies resulting from the same transformation reaction can be different, because each of the two plasmids can integrate in one or more copies. This difference in gene dosage can potentially yield a difference in the phenotype, or

dynamic behavior of the circuit.

4.4.1 CAA circuit can yield discrete decisions, and clones exhibit a diverse spectrum of behaviors

After obtaining sets of transformant clones for the different variations of the circuit outlined above, we performed a global survey of circuit behavior. For each circuit variant, several colonies were inoculated into selective liquid medium, supplemented with 5 μg/mL doxycycline. In the presence of high levels of doxycycline, the TetR proteins cannot bind to DNA, and hence their activation or repression activity is abrogated, resulting in complete inhibition of the circuit. After overnight growth in liquid medium, the cultures were diluted into fresh medium containing doxycycline, and incubated for another 6-8 hours to enable cells to leave the lag phase. Subsequently, the cultures were centrifuged, washed twice with PBS, and resuspended in fresh medium lacking doxycycline to allow expression of the circuit. Since the activator for Set 0 is fused to GFP, and the activator for Set 1 is fused to mCherry, the levels of the transcription factors can be tracked by assaying for GFP, mCherry expression.

Analysis of expression via flow cytometry reveals that for approximately the first 4 hours after the removal of doxycycline, there is no appreciable expression of GFP or mCherry over background. Subsequently, fluorescent signals continue to rise in most clones, and show very strong expression by the 16-hour timepoint.

Phase-plots of GFP, mCherry at the 16-hour timepoint reveal two important points (Figs. 4-3, 4-4, 4-7, 4-8, 4-9). First, across different circuit variants and clones, cells within most individual cultures are tightly clustered into 4 discrete regions in phase-space: (high-GFP, low-mCherry), (low-GFP, high-mCherry), (high-GFP, high-mCherry), and (low-GFP, low-mCherry). In the context of the CAA model, clustering of cells by their GFP, mCherry expression within a genotypically identical population indicates the presence of multiple steady-states, and demon-

strates that the CAA circuit can yield discrete decisions. Second, despite clustering of cells into discrete regions, clones from the same transformation reaction yield a spectrum of GFP, mCherry response profiles, that differ in two aspects: the placement of the four discrete states in GFP, mCherry space, and the distribution of the clonal population amongst these states. We hypothesized that the response-profile diversity among clones from the same transformation reaction is attributable to differences in plasmid copy-numbers.

Next, we examined sets of response profiles from the different circuit variants in detail to understand the effects of specific genetic perturbations.

4.4.2 Strong mutual repression is a requirement for exclusive states

Analysis and simulations of the CAA network topology indicate that the strength of repression exerted by T_A, T_B is a principal determinant of the placement of the bipotent state in (T_A, T_B)-space, as well as the relative proportions of cells committing to the different states. Furthermore, if repression strength is sufficiently weakened, the two exclusive states (high T_A, low T_B) and (low T_A, high T_B) are eliminated, rendering the overall system monostable (Fig. 4-10).

We evaluated these predictions by comparing two variants of our synthetic circuits: one with weak repression activity conferred by TetR's binding, and hence occlusion of the promoter, and another with strong repression achieved by the SSN6 domain. Unlike TetR alone, TetR fused to SSN6 can potentially exert multiple types of repressive activity at the promoter, by interfering with the mediator complex and halting transcription by RNA polymerase II, and by recruiting histone de-acetylation machinery to inactivate the promoter [Malavé and Dent, 2006]. Both circuit variants examined include the 2x version of the tetO promoters.

Comparison of GFP, mCherry expression profiles reveals the following. First, none of the 32 clones of the weak-repression circuit examined exhibit both the (high GFP,

low mCherry) and the (low GFP, high mCherry) exclusive states (Figs. 4-8, 4-9). In contrast, a large number of clones from the strong-repression circuit variant yield both exclusive discrete states. Second, in most profiles of the weak-repression set, the distribution of cells is heavily skewed to the (high GFP, high mCherry) state.

Taken together, our experimental results support the CAA model prediction that weakening of repression strength can lead to the population being biased towards the (high GFP, high mCherry) state, and even to the abolishment of the exclusive states as the system becomes monostable (Figs. 4-5, 4-6).

4.4.3 Multiple operator sites impede exclusivity in decision-making

In the context of stem-cell lineage commitment, the CAA topology models the behavior of a bipotent progenitor, with the (high T_A, low T_B) and (low T_A, high T_B) states representing committed, mature cell lineages. After commitment, the mature cell must express its own program, and suppress the other lineage's program to maintain its proper identity. Hence, the decision-making system must be parameterized such that the bipotent state, in which intermediate levels of both T_A, T_B are expressed, is much closer to the origin than to the (high T_A, high T_B) point on the phase-plot; in other words, decision-making should be exclusive.

Analysis of GFP, mCherry expression profiles across different circuit variants and clones reveals that in general, strong exclusivity is rare. Consistent with CAA model predictions, this implies that the balance between the two opposing sides is crucial, and is difficult to optimize in implemented circuits. How natural systems achieve nearly digital behavior through precise gene-regulation amidst the constant binding and un-binding of transcription factors and other components remains unclear.

To investigate this question further, we modified the operator sites in the strong repression, tetO-2x circuit and constructed a version with tetO-7x operators. At a genetic level, this circuit contains 5 additional tetO operator sites, which translates

into substantially increased opportunities for the repressor and activator proteins to bind and exert their effects. Comparison of response profiles from the two circuit variants reveals that at a global level, the contrasts between high, low states for both GFP and mCherry are higher in the tetO-7x set; and that the clustering of cells into discrete populations is more diffuse in the tetO-2x set (Fig. 4-13). Given these two points, the tetO-7x circuit variant would be expected to be more likely to yield exclusive decisions; however, comparison of individual response profiles from the two sets reveals 4 exclusive response profiles (from 48 clones) in the tetO-2x set, and none from the tetO-7x set (Figs. 4-3, 4-4, 4-7).

Our results demonstrate that promoter architecture can impact decision-making behavior, and suggest the following mechanism as the driver of the differences in behavior between the two circuit variants. In the context of only activation or only suppression, one would expect an increase in the number of operator sites to increase the magnitude of the effect. However, when activation and inhibition are exerted simultaneously at a locus, one would expect a distribution in a population. For instance, if the promoter contains only one operator site, then this site can be either unbound, bound by an activator, or bound by a repressor. An increase in the number of operator sites in the promoter leads to significantly more possible configurations, and gives rise to an averaging effect; in other words, the activity of the promoter as a function of activator and repressor levels would be expected to be more switch-like with 1-2 operator sites, and continuous with more operator sites. In the context of the CAA circuit, the tetO-7x operator may be subject to more fluctuations, and this behavior may prevent one side of the system from gaining an unassailable advantage over the other.

This analysis has broader implications for gene regulation: the number of operator sites can be an important variable in situations where contradictory actions are exerted at the same loci (which is likely true for a large number of genes in higher organisms).

4.4.4 Activator:repressor ratio modulates response profiles

The results described above suggest that the balance between activation and repression in the CAA model can be an important modulator of system behavior. To further investigate this, we constructed an additional set of repressor-only plasmids by removing the activator genes from the tetO-2x, strong repression (with SSN6 domains) set, while keeping the rest of the components intact. We then transformed these repressor-only plasmids together with the original plasmids (i.e., four different plasmids, two original, and additional two without activators, in the same transformation reaction) in a 1:1 ratio. This can allow us to modulate the relative copy-numbers for the activator and repressor genes and can hence enable the study of even stronger repression in the CAA circuit. A ratio of 1:1 between the two types of plasmids means that on average, a clone would have two copies of given side's repressor for each copy of its activator. For this analysis, we only included clones that exhibit expression of both GFP and mCherry; since a clone with only a repressor-only plasmid can survive selection, it is possible for some clones to have no copies of an activator gene (to which GFP, mCherry are fused).

A global comparison of response profiles from the 1:1 and 1:2 activator:repressor circuit sets (Fig. 4-14) reveals that, consistent with model predictions, among the 1:2 activator:repressor set clones, there is a marked decrease in the proportion of cells in the (high GFP, high mCherry) promiscuous state.

4.4.5 Gene dosage modulates dynamic behaviors

As mentioned above, clones from the same transformation reaction, and hence having the same set of plasmids, exhibit diverse response profiles, or phenotypes. We hypothesized that this diversity results from differences in gene dosage; i.e., different clones integrate the two plasmids in different copy-numbers, and thus modifying the dynamics of the overall system.

Despite this diversity, clones from across different circuit variants also exhibit

strong similarities in their response profiles. To further explore the existence of general patterns, we applied a clustering method to all response profiles. Briefly, each response profile was binned into an $n \times n$ grid, and a distance metric was computed for each pair of response profiles (see Methods). The distances were used as input into the PAM algorithm, which partitions data points into groups through iterative optimization [Reynolds et al., 2006], to obtain clusters of response profiles. Reasonable values for n, the number of bins for each of the two dimensions, and k, the number of clusters, were obtained empirically by varying both parameters, and analyzing silhouette scores of the resulting clusterings. Cluster analysis suggests (Fig. 4-12) that all response profiles can be partitioned into approximately 15 representative groups, a modest number, given the number of combinations of genetic perturbations and plasmid copy-numbers used (Figs. 4-9, 4-11, 4-12).

Given the relatively tight clustering of response profiles, we explored the possibility of using our results to obtain a general regression that can predict the response profile, given promoter-type, repression strength, and plasmid copy-number as inputs. Such an analysis can reveal the relative weights of these parameters in determining the phenotype, and hence point to the primary drivers of behavior in natural systems.

Towards this goal, we selected four clones with differing GFP, mCherry response profiles from the tetO-2x, strong repression set, and used digital PCR to quantify copy-numbers of the two plasmids. Our preliminary results demonstrate that gene dosage can indeed strongly modulate the dynamic behavior of circuits (Fig. 4-16).

Our analysis points to broader implications of gene dosage in complex regulatory networks. A change in gene dosage through deletion or amplification, as is common in cancer [Santarius et al., 2010], can not only affect targets directly downstream, but in the context of a dynamical system, can also deleteriously alter the energy landscape by introducing strong biases toward certain states, or by introducing corrupt intermediate states.

4.5 Methods

4.5.1 Parts construction

Transcriptional activator genes were constructed as follows. The tTA gene was amplified from the pUG6-tTA plasmid [Yen et al., 2003] (obtained from EUROSCARF, Accession P30385) using primers NASo65, NASo55. This version of tTA has dimerization and operator-site specificities of type 0, as defined above. The yEGFP3 gene was amplified from pSP001 [Palani and Sarkar, 2011] with NASo57, NASo58, and cloned into the tTA gene via BssHII digestion and ligation. This gene is referred to as tTA_0 in the text.

To construct tTA_1, the following PCR reactions were first performed, all with pUG6-tTA as template. First, with primers NASo55, NASo68. Second, with primers NASo67 and NASo70. Third, with NASo69 and NASo065. Overlap PCR was performed on purified products from these these three PCR reactions, and outer primers NASo65, NASo55 were used to amplify the resulting gene. The mCherry gene was amplified from plasmid eco062 [O'Shaughnessy et al., 2011] using primers NASo75 and NASo76, and cloned into the tTA_1 gene via BssHII digestion and ligation.

Two types of transcriptional repressors were constructed, the TetR protein alone, or fused with the SSN6 repression domain. Construction of both types in the two specificities yields four separate genes. To construct TetR-only genes, primers NASo077 andNASo078 were used to PCR reactions with tTA_0 and tTA_1 separately as templates. To construct TetR-SSN6 fusions, part of the SSN6 gene was amplified from the *Saccharomyces cerevisiae* genome using primers NASo080 and NASo081. Next TetR genes of the two specificities were separately amplified using primers NASo077 and NASo079. Purified products from the preceding two PCR reactions were used in an overlap PCR reaction, and further amplified with outer primers NASo077 and NASo081.

The TetR mutation strategies were designed by Pamela Barendt.

4.5.2 Plasmid construction

The bi-directional promoter system was constructed as follows. First, two separate PCR reactions were performed with eco008 [O'Shaughnessy et al., 2011] as template and primer pairs NASo001, NASo002 and NASo003, NASo004. Purified products from these reactions were used in an overlap PCR reaction and outer primers NASo001 and NASo004 were used to further amplify the product. This product was digested with XhoI and BamHI, and ligated into eco008a to construct NAS001.1.

Primers NASo036 and NASo037 were used to amplify the CMV promoter and the ADH1 terminator from NAS001.1. The product was digested with EcoRI and AvrII, and ligated into NAS001.1. Next, the CYC1 TATA minimal promoter was amplified from NAS001.1 with primers NASo59, NAS060 and the product was cloned into the new plasmid via AvrII and XhoI enzymes. This plasmid establishes the following promoter architecture. First, the space between AvrII and ClaI sites is used to conveniently insert the desired operator: canonical tetO-2x, tetO-4C5G-2x, canonical tetO-7x, or tetO-4C5G-7x. On both the 5' and 3' ends of the operator space are copies of the minimal CYC1 TATA promoters. Following the CYC1 TATA sequence on the 3' side are BamHI and NotI restriction sites, followed by the CYC1 terminator sequence. Similarly, following the CYC1 TATA sequence on the 5' side are AflII and XhoI restriction sites, followed by the ADH1 terminator. The resulting plasmid was transferred to the HIS3 backbone of pERT252, creating two plasmids with different auxotrophic selection markers (URA3 and HIS3).

The tTA_0, tTA_1 activator genes (together with the respective GFP or mCherry fusion) were cloned in via BamHI and NotI sites, while the repressor genes, tTS_0, tTS_1, were cloned in via AflII and XhoI sites. Different tetO operator sequences were either purchased or constructed via PCR. These sequences were cloned into the plasmids via AvrII, ClaI cloning.

The background G418-resistance plasmid was constructed as follows. NAS001.1 was digested with AvrII, ClaI and oligos with the tetO-2x were ligated into it to

construct plasmid NAS004.1. Similarly, oligos with the tetO-4C5G-2x sequence were ligated into the NAS001.1 plasmid (digested with AvrII, ClaI). Primers NASo099 and NASo0100 were used to amplify the Kanamycin resistance gene, KanR, from the pUG6-tTA plasmid, and ligated into the two plasmids. This yields two constructs with KanR driven by tetO-2x and tetO-4C5G-2x, respectively.

Primers NASo112 and NASo113 were used to amplify the KanR cassette from the plasmid containing the tetO-2x sequence, and the resulting product was cloned into eco-007 via XmaI and XhoI. This places the KanR, tetO-2x cassette in a LEU2 auxotrophic marker background. Primers NASo001 and NASo041 were used to amplify from the plasmid containing KanR driven by tetO-4C5G-2x sequence, and the resulting product was ligated into the new plasmid containing KanR driven by tetO-4C5G-2x sequence, and a LEU2 background, via XhoI and MluI cloning. The ligated plasmid was named NAS135.

The NAS135 plasmid consists of two copies of the KanR gene, driven by the tetO-2x and tetO-4C5G-2x operators. In the context of the circuit, the KanR protein is expressed when either or both of tTA_0, tTA_1 are present in sufficient quantities. Hence, addition of G418 to the culture medium will inhibit growth of cells not expressing significant levels of both tTA proteins. This mechanism facilitates additional perturbations.

4.5.3 Yeast transformation

The NASy001 strain was constructed by transforming BMA-64 yeast cells with the NAS135 plasmid, and PCR-screening to identify a single-integrand. All CAA circuit plasmids were transformed into NASy001.

Circuit plasmids for the two opposing sides in each circuit variant were transformed simultaneously into NASy001 via the LiAc/SS carrier DNA/PEG protocol [Gietz and Schiestl, 2007]. Transformation reaction mixes were plated onto agar plates containing synthetic selective medium, as well as 5 μg/mL anhydrotetracycline (atc, Sigma) to

inhibit activation of the circuit during post-transformation growth. Atc was used instead of doxycycline since it has a longer half-life (the post-transformation growth period lasts 48-72 hours).

4.5.4 Yeast culture for experiments

Individual colonies were picked from transformation plates and inoculated into selective 'Magic' medium supplemented with 2% galactose and 5 μg/mL doxycycline to inhibit activation of the circuit. Magic medium lacks ammonium sulfate, and is hence better suited for experiments in which the drug G418 is used. Magic medium contains 1.7 g nitrogen base (without amino acids and ammonium sulfate, Sigma), 1.4 g drop-out mix lacking histidine, uracil,leucine, and tryptophan, and 1 g sodium glutamate (Sigma) per liter.

After overnight growth, cultures were diluted into fresh medium containing doxycycline, and grown for 6-8 hours to enable entry into log-phase growth. Cultures were then washed twice with PBS to remove doxycycline, and re-suspended in fresh medium lacking doxycycline to induce expression of the circuit. During experiments, cultures were regularly diluted as needed to maintain log-phase growth.

Subsequent to overnight growth with doxycycline, selected cultures were stored in glycerol (at 25% effective concentration) at -80C.

4.5.5 Flow cytometry data analysis

GFP, mCherry expression was assessed by flow cytometry using the BD LSRFortessa instrument, equipped with 488-nm and 561-nm lasers. Data files from the instrument were imported into the FlowJo software, and the FSC-H, FSC-W, FSC-A, and SSC-A parameters were used to exclude events not corresponding to individual yeast cells.

After gating, custom R programs were used to pool flow cytometry data and trim outlier events. Subsequently, each response profile was binned into an $n \times n$ grid, with n bins each for GFP and mCherry. For plotting all figures, $n = 100$ was used,

while for clustering of response profiles, $n = 10$ was used.

For cluster analysis of flow cytometry data, a distance metric was computed between each pair of binned response profiles. Each binned response profile was treated as a vector in n^2-space, and the Manhattan distance metric [Black, 2006] was used to compute the dissimilarity.

Masks for aggregated flow cytometry response profiles were created by first processing and binning each response profile as described above, and then computing an average over all of the response profiles. The data were normalized before averaging such that each response profile carries the same weight.

4.5.6   Plasmid copy-number estimation via digital PCR

For selected clones, the numbers of integrated copies of the two plasmids were estimated using digital PCR on the OpenArray Real-Time PCR System (Life Technologies). Briefly, clones were inoculated and grown in selective medium containing high levels of doxycyline to suppress circuit expression. Saturated cultures were then processed to extract purified genomic DNA. The genomic DNA was digested overnight at 37°C with the SacI-HF restriction endonuclease (New England Biolabs). The digested genomic DNA was then appropriately diluted and used as template in separate digital PCR reactions along with TaqMan probes (Life Technologies) targeting either the GFP, mCherry, or ALG9 genes. For quantifying the number of integrated copies, both GFP and mCherry are considered proxies of their respective plasmids, with the native yeast gene ALG9 being an endogenous control.

4.5.7   Live-cell imaging

GFP, mCherry expression dynamics of individual cells were monitored using a protocol modified from [Young et al., 2012]. Cells from individual clones were grown in selective medium containing doxycycline to suppress circuit expression, subsequently washed to remove doxycycline, and then transferred to agar pads containing

SC medium. After a few minutes, the pads were inverted and placed in a microscopy dish. A Deltavision Deconvolution Station (consisting of an Olympus IX70 inverted microscope, a Photometrics CoolSnap HQ high-resolution CCD camera, and a motorized stage) was used to acquire images at 15-minute intervals.

NASo001	AAG CTC CTC GAG TAA TTC GCG C
NASo002	ATC GAT GGC GCC CCT AGG AAC TCG ATC GAG GAA TTG ATC TGC CGG TAG AGG TG
NASo003	CCT AGG GGC GCC ATC GAT AAA GTC GAG CTC GGT ACC CTA TGG CAT GCA TGT GC
NASo004	AGA CAT GGA TCC CCC GAA TTG
NASo016	CTA GGT CCC TAT CAG TGA TAG AGA AAA GTG AAA GTC GAG TTT ACC ACT CCC TAT CAG TGA TAG AGA AT
NASo017	CGA TTC TCT ATC ACT GAT AGG GAG TGG TAA ACT CGA CTT TCA CTT TTC TCTAT CAC TGA TAG GGA C
NASo036	AAT TCG GGG AAT TCG ATC TGC CGG TAG AGG TGT GG
NASo037	AAT TCG GGG GAT CCA TCG ATG GCG CCC CTA GGA TGA ATT AAT TCG GGC CGC G
NASo055	TAA TCA GTG GAT CCA TGT CTA GAT TAG ATA AAA GTA AAG TGA
NASo058	AAA CTA GTG CGC GCC TTT GTA CAA TTC ATC CAT ACC ATG G
NASo059	ACT GAT TAC CTA GGT ATG GCA TGC ATG TGC TCT GT
NASo060	CCC GAA TTC TCG AGA TCA GTC TTA AGA TCC CCC GAA TTG ATC CGG T
NASo065	ACT GAT TAG CGG CCG CCT ACC CAC CGT ACT CGT C
NASo067	GCC CAG AAG CTA GGT TTT GCG CAG AAA ACA TTG TAT TGG CAT

Table 4.1: Primers used for construction of synthetic circuits - 1

NASo069	CCA GCC TTC TTA CAC GGC CTT GAA TCG CTC ATA TGC GGA TTT GAA AAA CAA CTT
NASo070	AAG TTG TTT TTC AAA TCC GCA TAT GAG CGA TTC AAG GCC GTG TAA GAA GGC TGG
NASo075	CGA TAG GCG CGC TGG TTT CCA AGG GTG AAG AAG
NASo076	CGA TAG GCG CGC CCT TGT ATA ATT CGT CCA TAC C
NASo077	CGA TAG CTT AAG ATG TCT AGA TTA GAT AAA AGT AAA GTG
NASo078	CTA TCG CTC GAG TTA CCC ACT TTC ACA TTT AAG TTG T
NASo079	AGA TCC ACT TTC ACA TTT AAG TTG
NASo080	CAA CTT AAA TGT GAA AGT GGA TCT ATG AAT CCG GGC GGT GAA C
NASo081	AAT GTC GCT CGA GTT TAG TCG TCG TAG TTT TCA TC
NASo090	CTA GGT CCC CGT CAG TGA CGG AGA AAA GTG AAA GTC GAG TTT ACC ACT CCC CGT CAG TGA CGG AGA AT
NASo091	CGA TTC TCC GTC ACT GAC GGG GAG TGG TAA ACT CGA CTT TCA CTT TTC TCC GTC ACT GAC GGG GAC
NASo099	TGA GCT GGA TCC ATG GGT AAG GAA AAG ACT CAC GTT TCG
NASo100	TCT TCG CGG CCG CTT AGA AAA ACT CAT CGA GCA TCA AAT GAA A
NASo112	CAT GCC CCG GGC GCG CCA CTT CTA AAT AAG CGA A
NASo113	AGA CTC TCG AGT TAC ATG ATG CGG CCC TCC TGC A

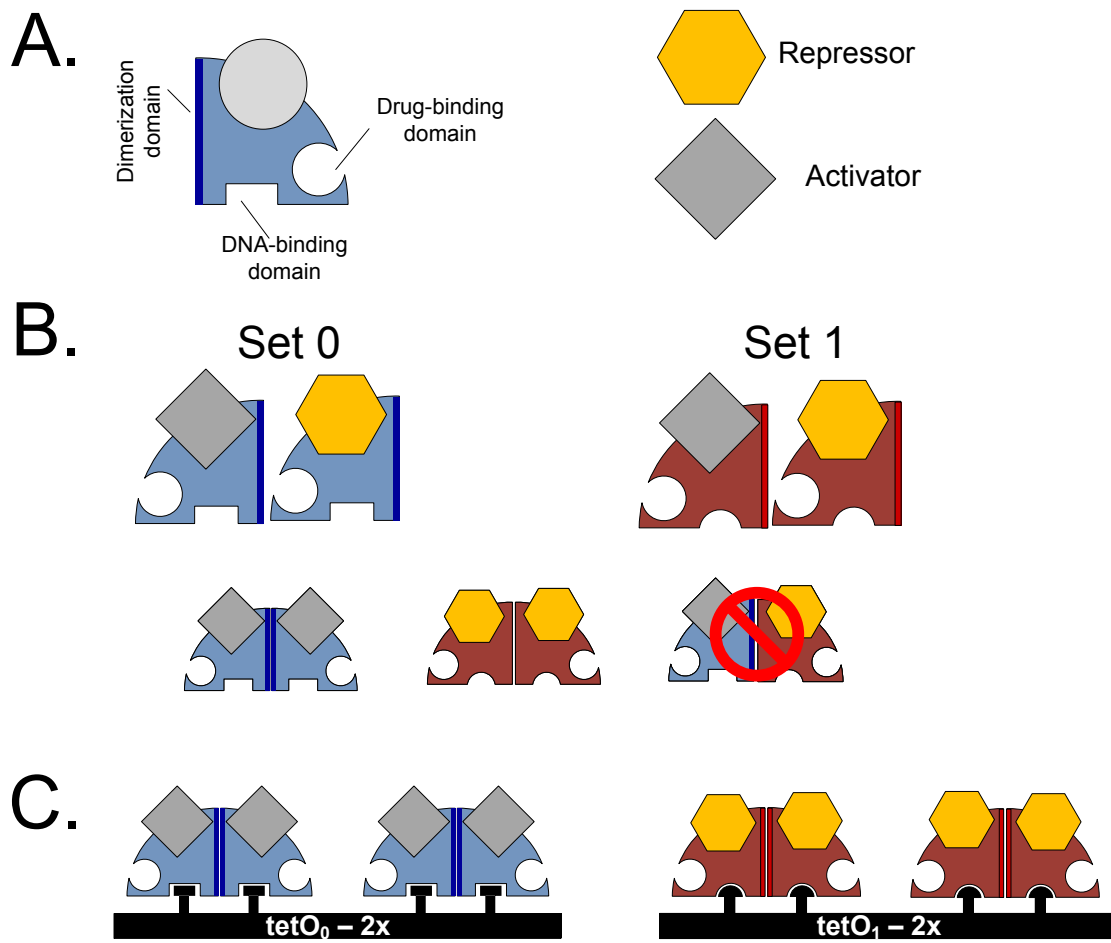Table 4.2: Primers used for construction of synthetic circuits - 2

Figure 4-1: **Re-use of orthogonal parts. A.** Schematic for the TetR monomer, which comprises a dimerization domain, a drug-binding domain, and a DNA-binding domain. TetR can be used alone as a weak repressor, or fused to an activation or stronger repression domain. **B.** Mutations in the DNA-binding and dimerization domains enable the simultaenous, non-interfering use of an orthogonal TetR (Set 1) along with the original TetR (Set 0). Dimerization is possible only between monomers of the same set. **C.** Set 0 TetR proteins bind to $tetO_0$ operators, while Set 1 TetR proteins bind to $tetO_1$ operators.
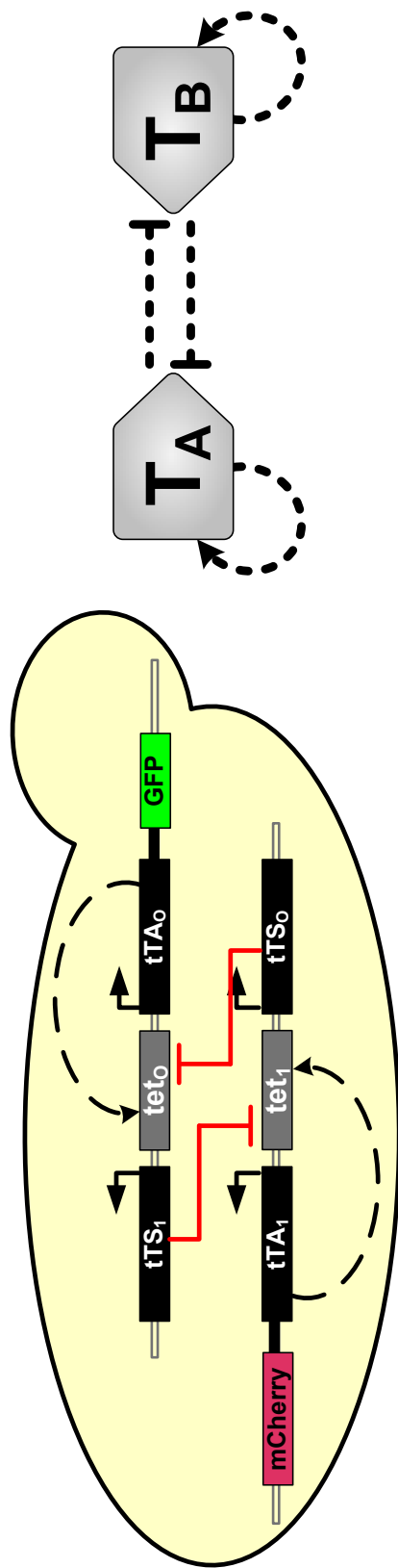
Figure 4-2: **Synthetic implementation of cross-antagonism autoregulation circuit** This synthetic circuit implements the CAA network topology. The CAA model comprises two opposing sides, T_A, T_B; here, T_A is represented by the transcriptional activator tTA_0, the repressor tTS_0, and the promoter $tetO_0$. Similarly, T_B is represented by the transcriptional activator tTA_1, the repressor tTS_0, and the promoter $tetO_1$. The activators are fused to either GFP or mCherry to allow for tracking via fluorescence measurements. The two sides are delivered on separate plasmids that integrate into yeast chromosomes.
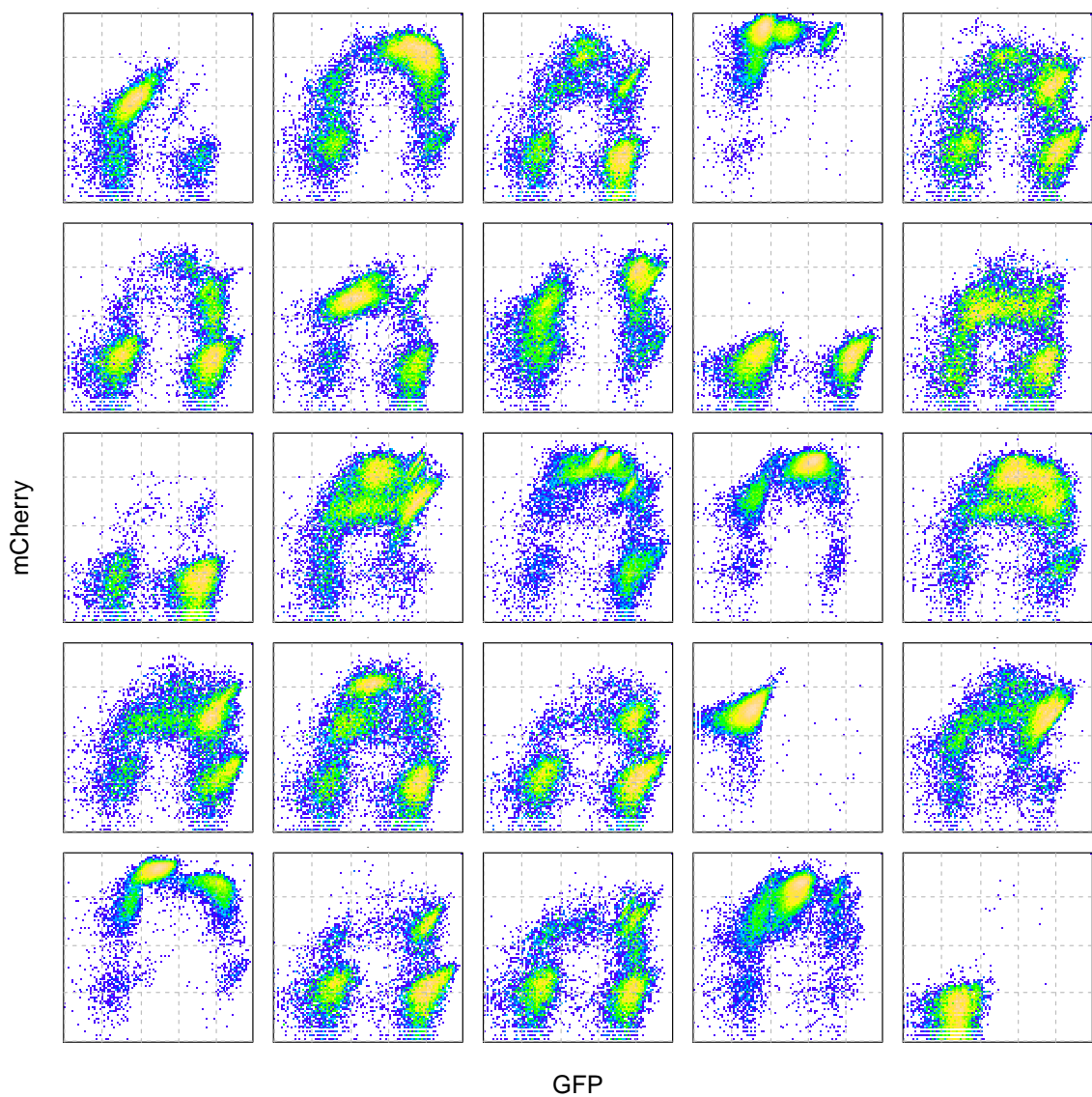
Figure 4-3: **Response profiles for the CAA circuit with tetO-2x, 2 μg set.** GFP, mCherry response profiles for different clones of CAA circuits with tetO-2x operators, and strong repression using SSN6 domains. High plasmid concentrations were used for this transformation set. Response profiles were processed and binned as described in the text. All profiles are on log-scales for GFP and mCherry. For each axis, the distance between two dashed lines denotes a ten-fold change in expression.

Figure 4-4: **Response profiles for the CAA circuit with tetO-2x, 500 ng set.** GFP, mCherry response profiles for different clones of CAA circuits with tetO-2x operators, and strong repression using SSN6 domains. Low plasmid concentrations were used for this transformation set. Response profiles were processed and binned as described in the text. All profiles are on log-scales for GFP and mCherry. For each axis, the distance between two dashed lines denotes a ten-fold change in expression.
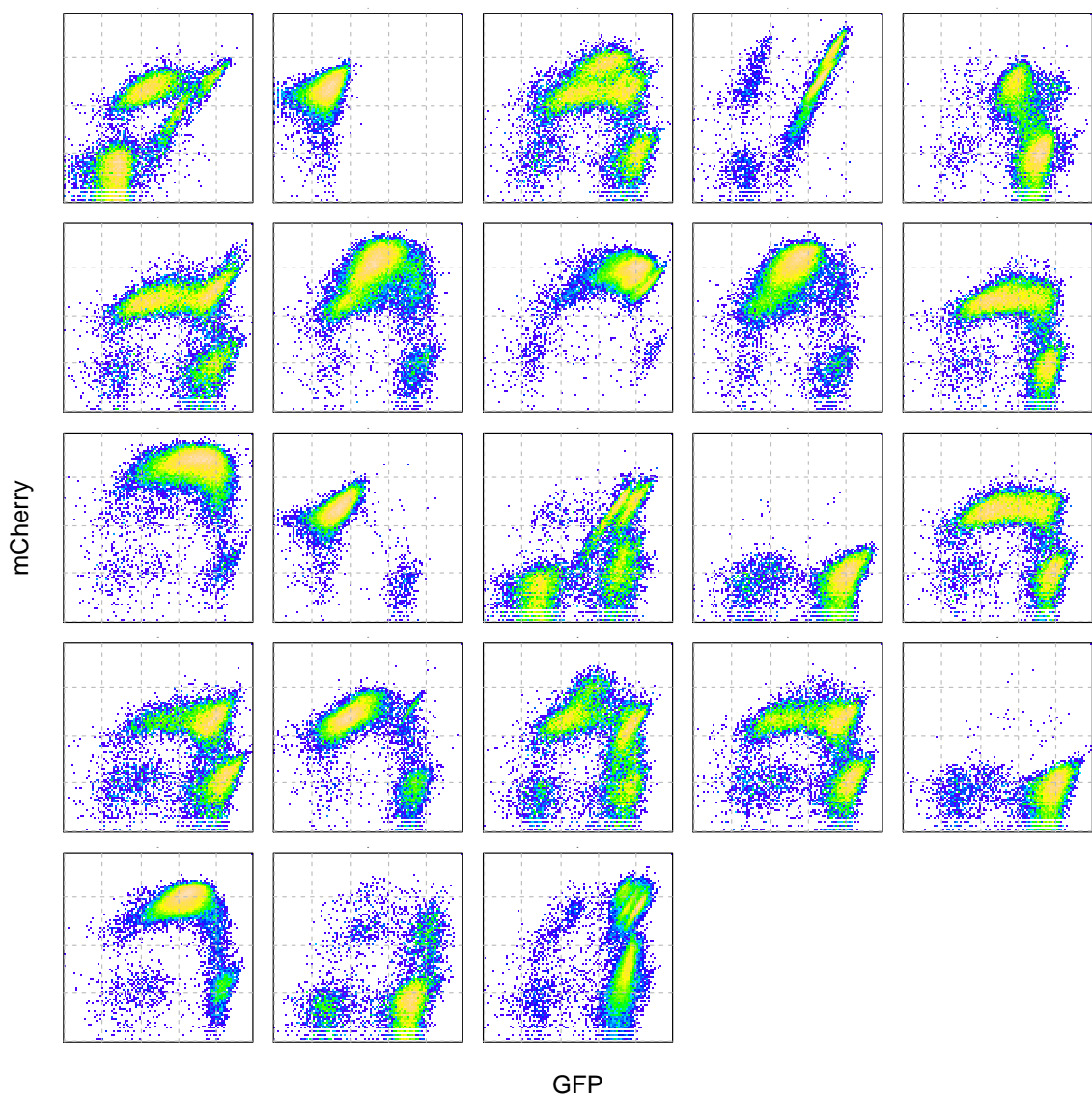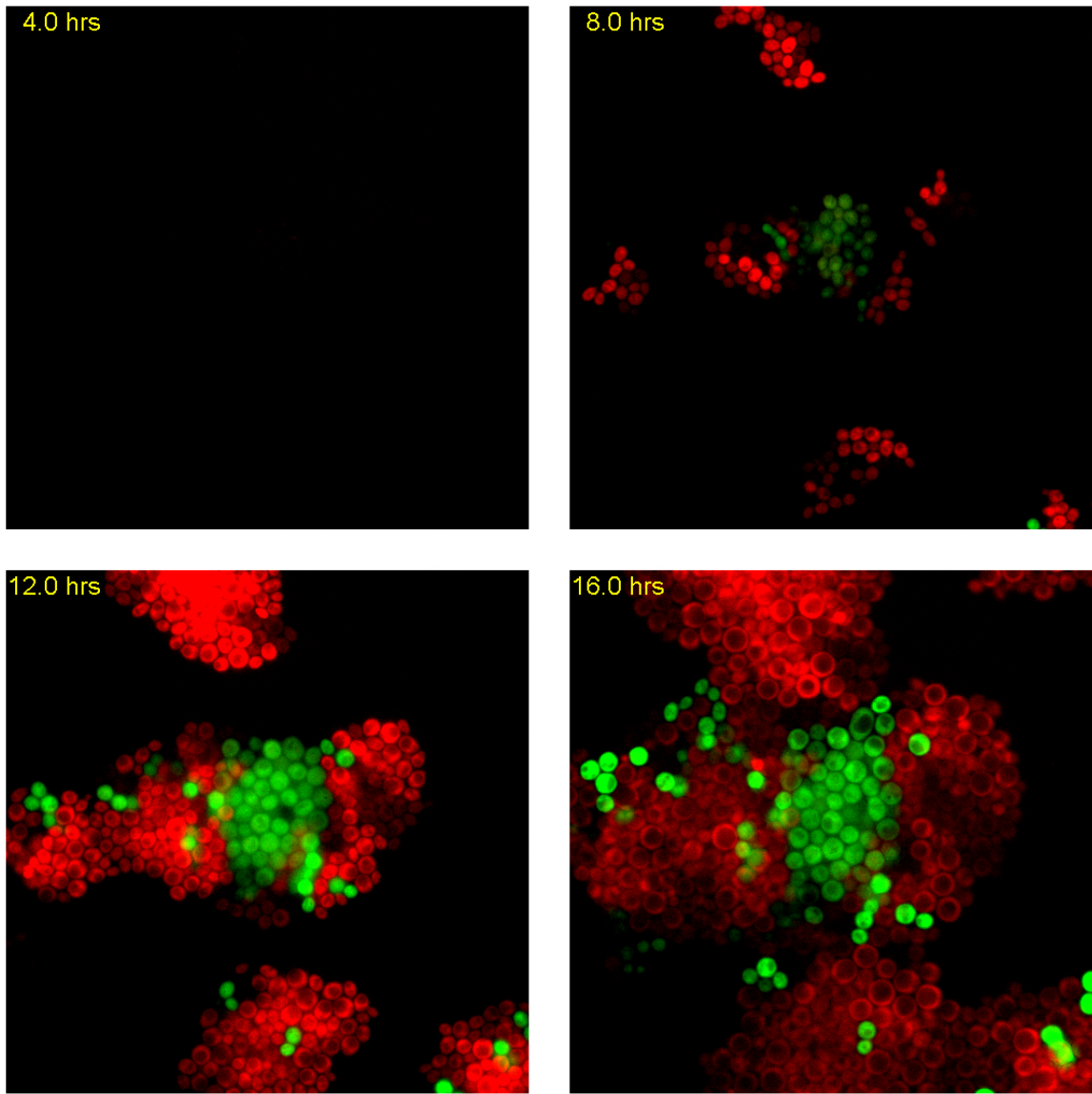
Figure 4-5: **Strong inhibition in the CAA circuit can yield exclusivity and memory.** Cells from a clone from the tetO-2x, strong repression set were grown in selective medium containing high doxycycline to suppress circuit expression, were subsequently washed to remove doxycycline, and then transferred to an agar pad for time-lapse microscopy. Images show that under strong mutual inhibition, the CAA circuit can yield exclusive decisions with memory, in that cells that decide to commit to a state remain committed.
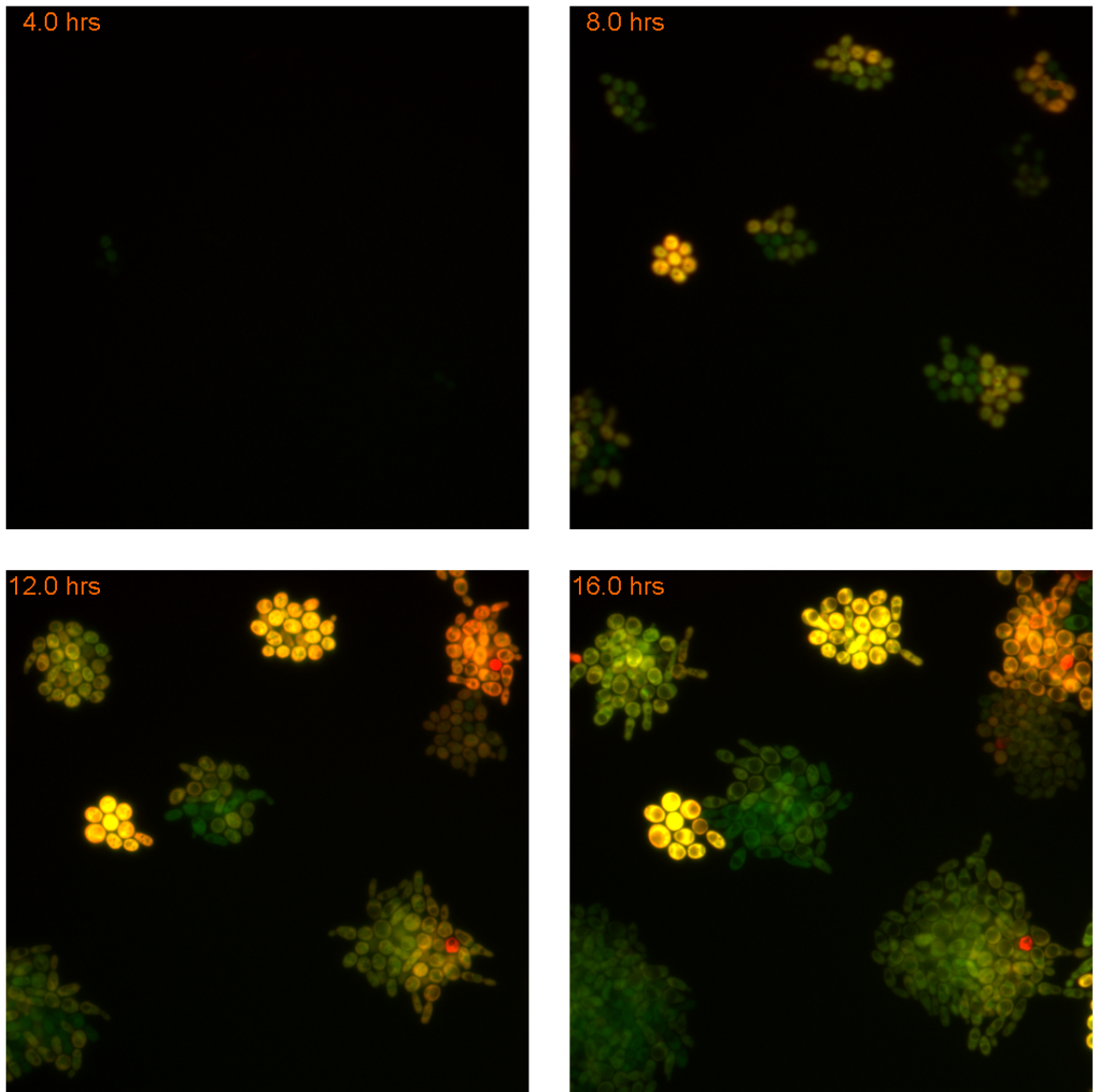
Figure 4-6: **Weak inhibition in the CAA circuit does not yield exclusivity.**
Cells from a clone from the tetO-2x, weak repression set were grown in selective
medium containing high doxycycline to suppress circuit expression, were subsequently
washed to remove doxycycline, and then transferred to an agar pad for time-lapse
microscopy. Images show that under weak mutual inhibition, the CAA circuit does
not yield exclusive decisions, in that most cells express the transcription factors from
both states at high levels.
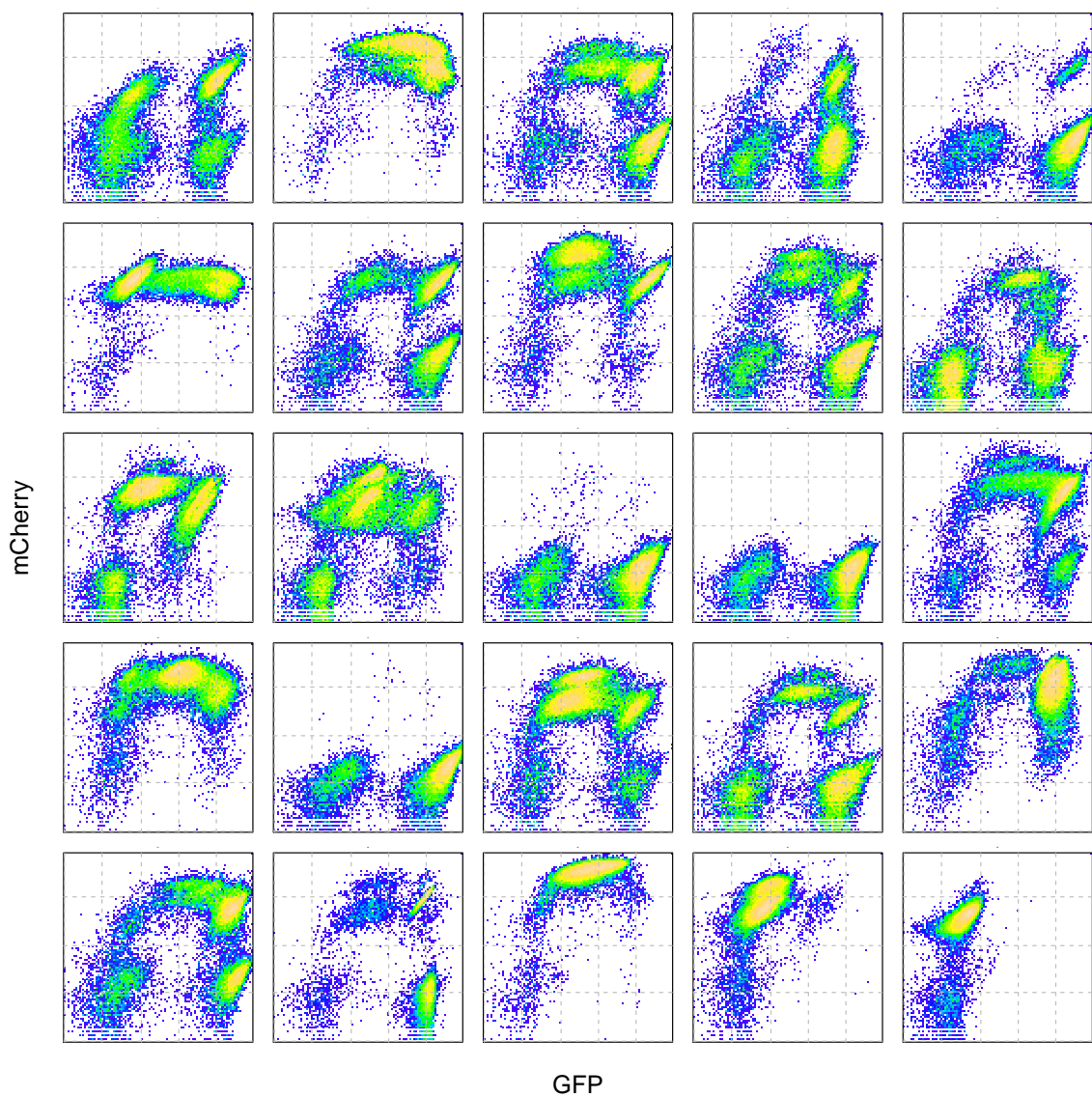
Figure 4-7: **Response profiles for the CAA circuit with tetO-7x.** GFP, mCherry response profiles for different clones of CAA circuits with tetO-7x operators, and strong repression using SSN6 domains. High plasmid concentrations were used for this transformation set. Response profiles were processed and binned as described in the text. All profiles are on log-scales for GFP and mCherry. For each axis, the distance between two dashed lines denotes a ten-fold change in expression.
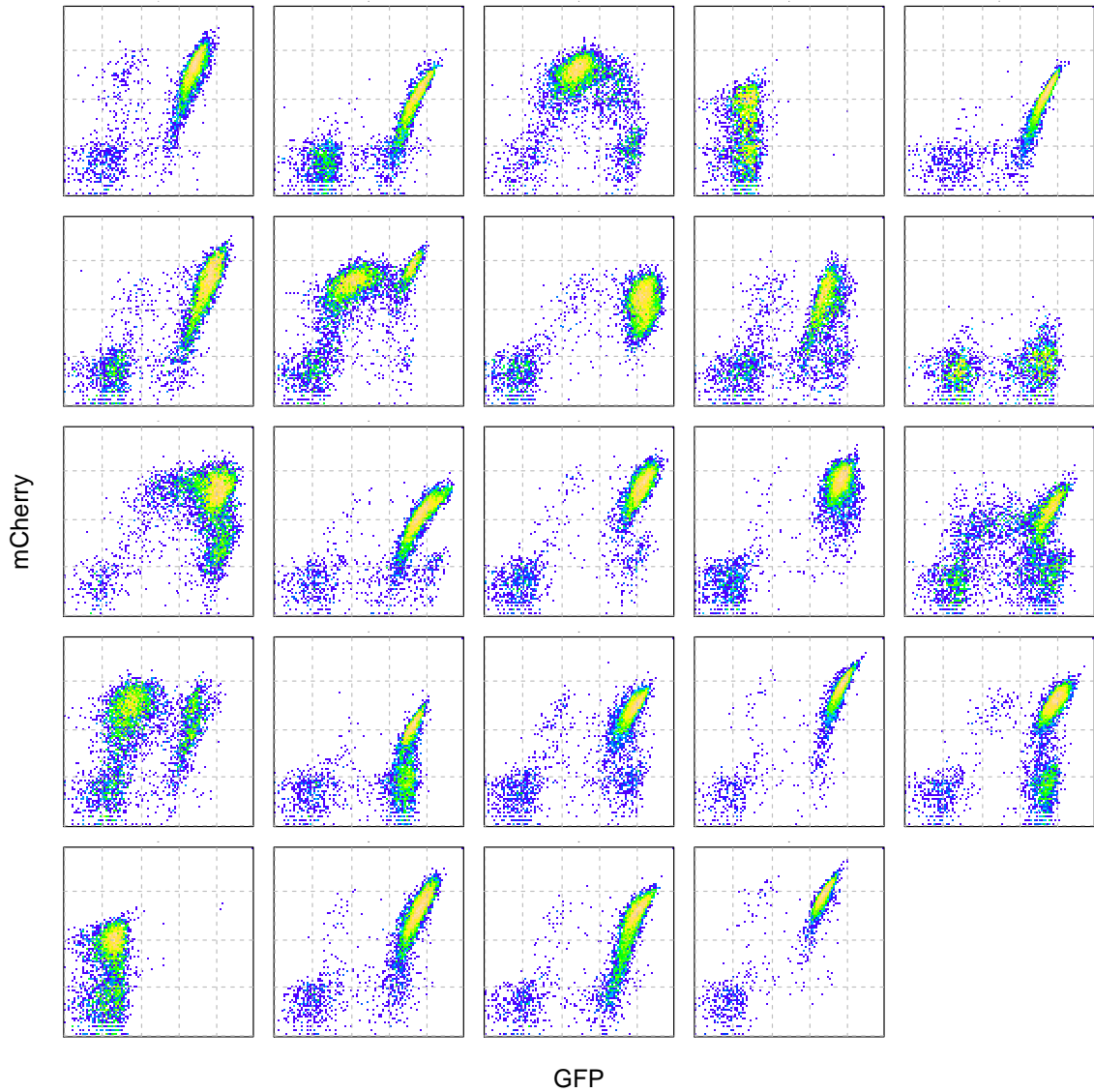
Figure 4-8: **Response profiles for the CAA circuit with tetO-2x and weak repression, 2 μg set.** GFP, mCherry response profiles for different clones of CAA circuits with tetO-2x operators, and weak repression (without SSN6 domains). High plasmid concentrations were used for this transformation set. Response profiles were processed and binned as described in the text. All profiles are on log-scales for GFP and mCherry. For each axis, the distance between two dashed lines denotes a ten-fold change in expression.
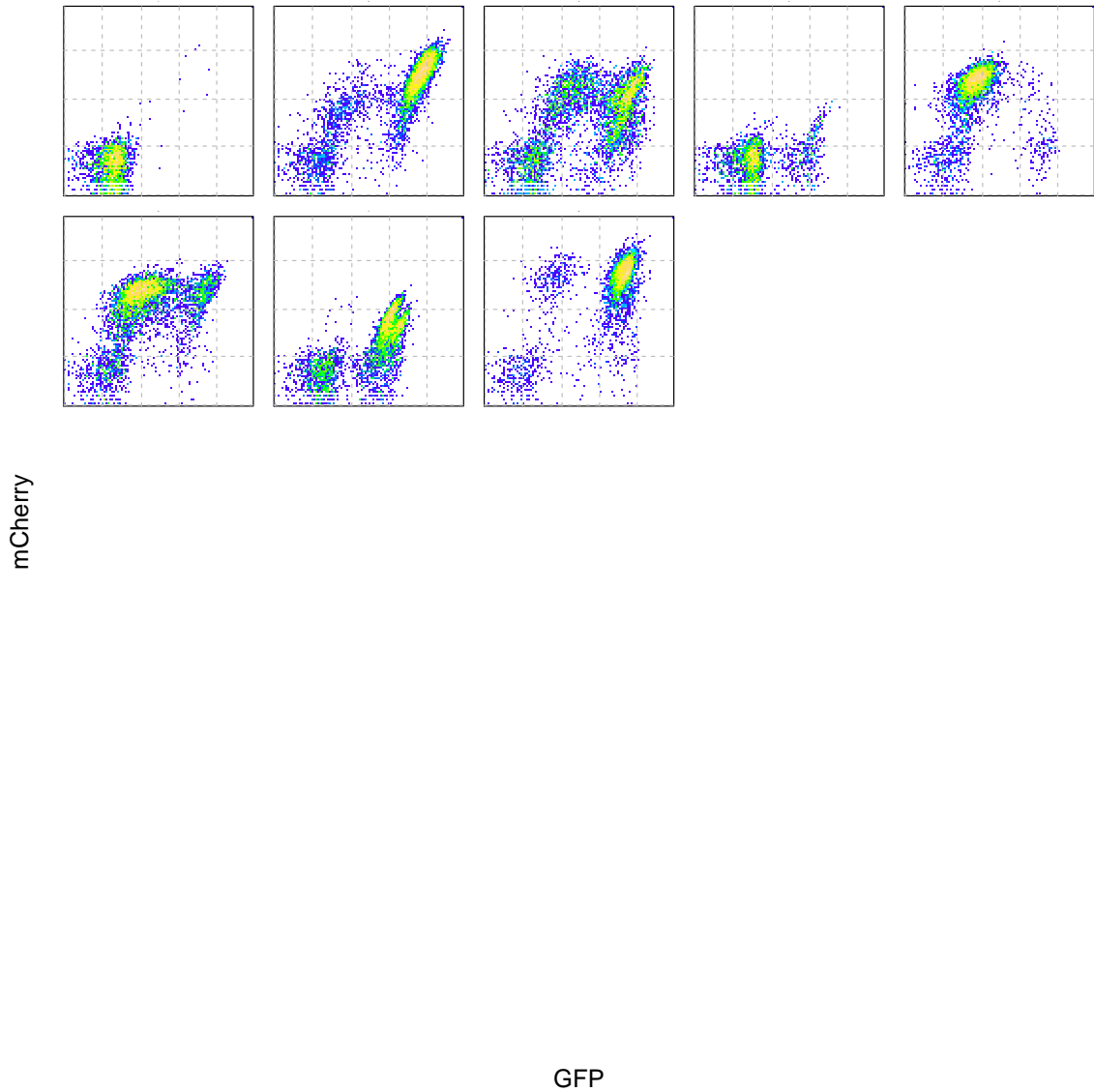
Figure 4-9: **Response profiles for the CAA circuit with tetO-2x and weak repression, 500 ng set.** GFP, mCherry response profiles for different clones of CAA circuits with tetO-2x operators, and weak repression (without SSN6 domains). Low plasmid concentrations were used for this transformation set. Response profiles were processed and binned as described in the text. All profiles are on log-scales for GFP and mCherry. For each axis, the distance between two dashed lines denotes a ten-fold change in expression.
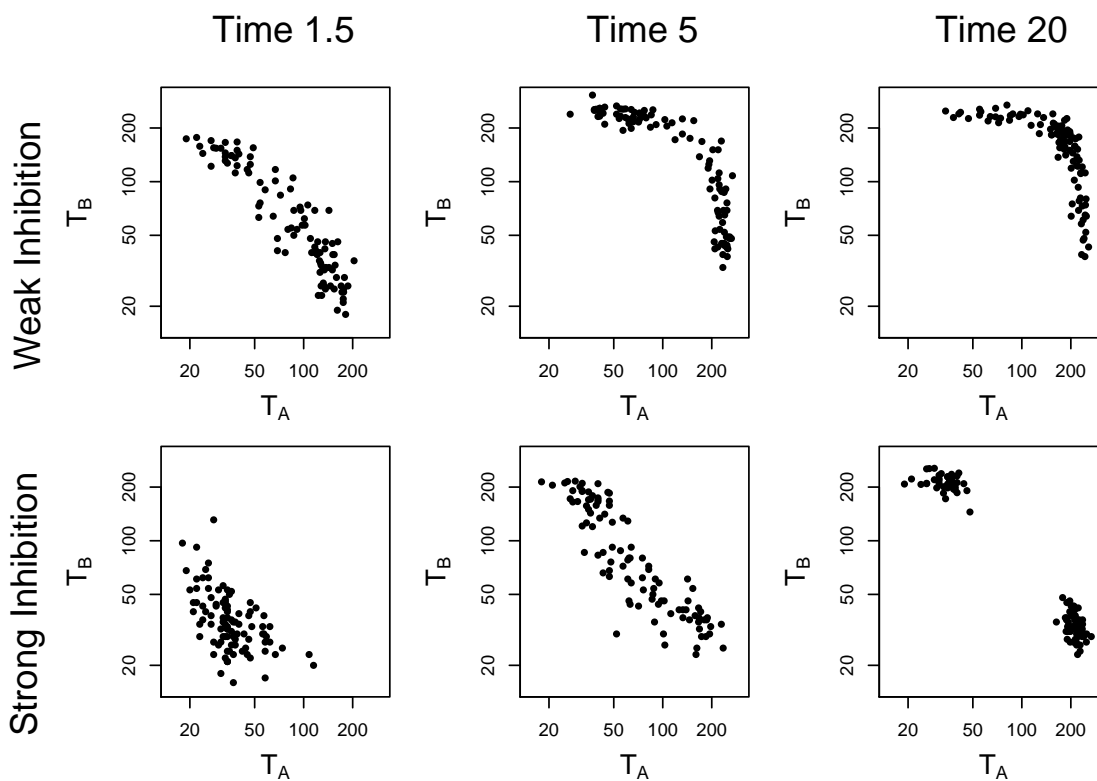
Figure 4-10: **Simulated trancription factor dynamics under different mutual inhibition strengths.** The CAA model was stochastically simulated under weak and strong mutual inhibition strengths. For each regime, 100 separate simulations were performed, yielding 100 trajectories. Plots show snapshots of trajectories at different times. Each simulation was started in the basal state, at $T_A = 1$, $T_B = 1$. Simulation results show that strong mutual inhibition is necessary to yield the two exclusive states (high T_A, low T_B) and (low T_A, high T_B).
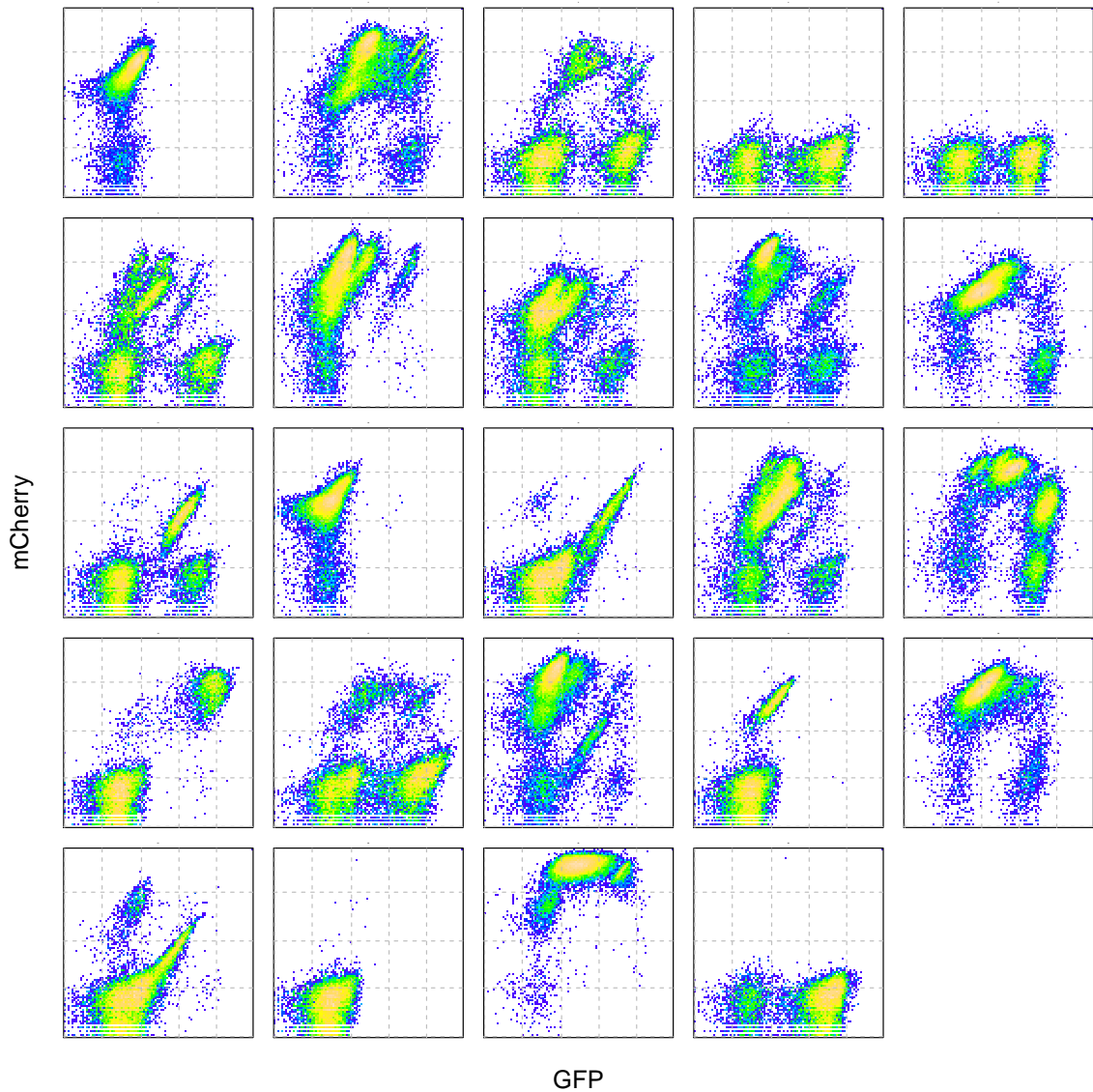
Figure 4-11: **Response profiles for the CAA circuit with tetO-2x, strong repression, with additional repressor copies.** GFP, mCherry response profiles for different clones of CAA circuits with tetO-2x operators, strong repression with SSN6 domains. A further increase in repression strength was attempted by including repressor-only plasmids alongside the regular plasmids in the transformation, as described in the text. Response profiles were processed and binned as described in the text. All profiles are on log-scales for GFP and mCherry. For each axis, the distance between two dashed lines denotes a ten-fold change in expression.
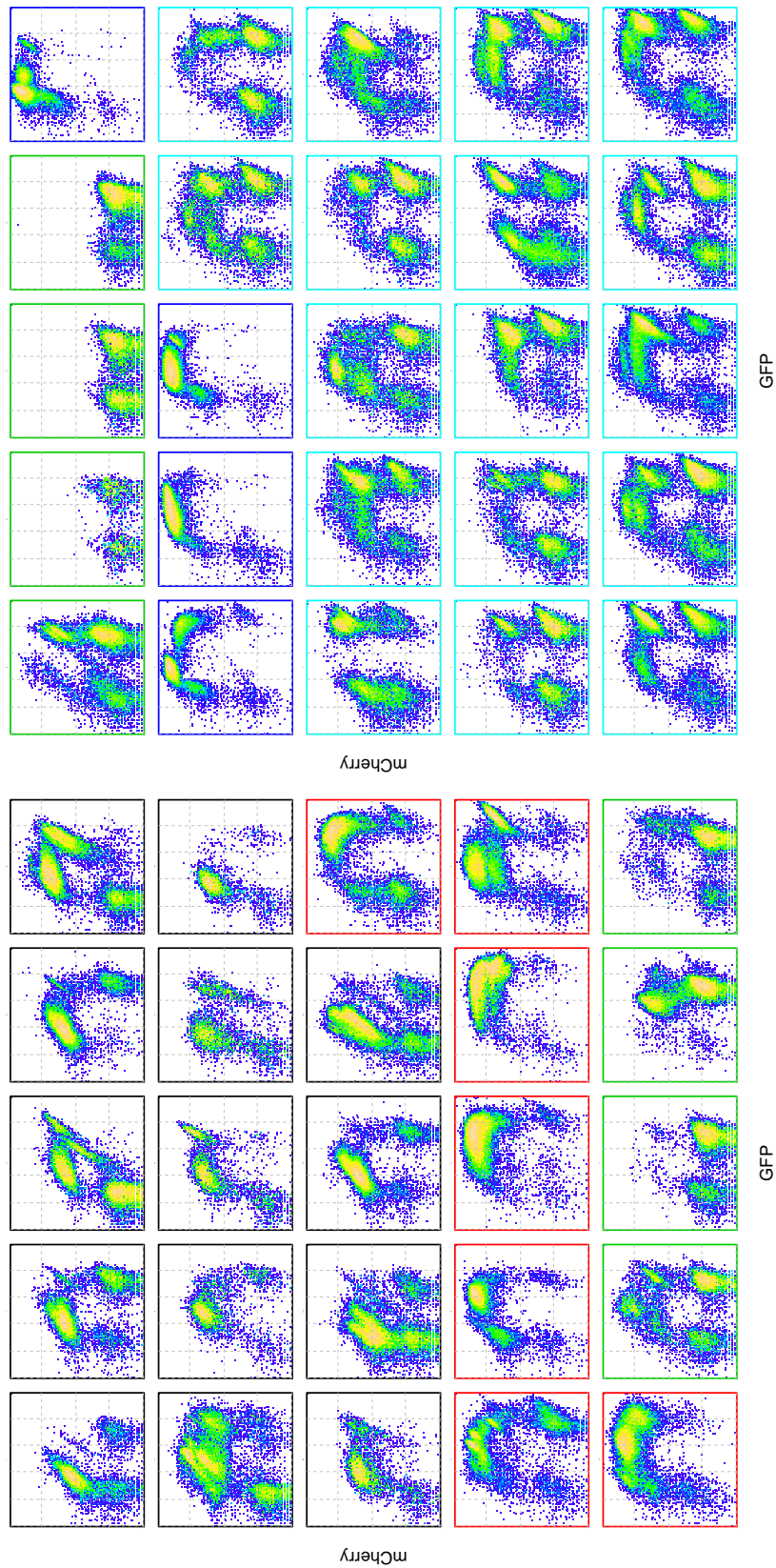
Figure 4-12: **Clustering of all response profiles - 1.** All response profiles obtained in the experiment were processed, binned ($n = 10$), and clustered using the Manhattan distance metric and the PAM algorithm with $k = 16$. Different outline colors for individual response profiles denote different clusters. All profiles are on log-scales for GFP and mCherry. For each axis, the distance between two dashed lines denotes a ten-fold change in expression.

Figure 4-12: **Clustering of all response profiles - 2.** All response profiles obtained in the experiment were processed, binned ($n = 10$), and clustered using the Manhattan distance metric and the PAM algorithm with $k = 16$. Different outline colors for individual response profiles denote different clusters. All profiles are on log-scales for GFP and mCherry. For each axis, the distance between two dashed lines denotes a ten-fold change in expression.

Figure 4-12: **Clustering of all response profiles - 3.** All response profiles obtained in the experiment were processed, binned ($n = 10$), and clustered using the Manhattan distance metric and the PAM algorithm with $k = 16$. Different outline colors for individual response profiles denote different clusters. All Profiles are on log-scales for GFP and mCherry. For each axis, the distance between two dashed lines denotes a ten-fold change in expression.
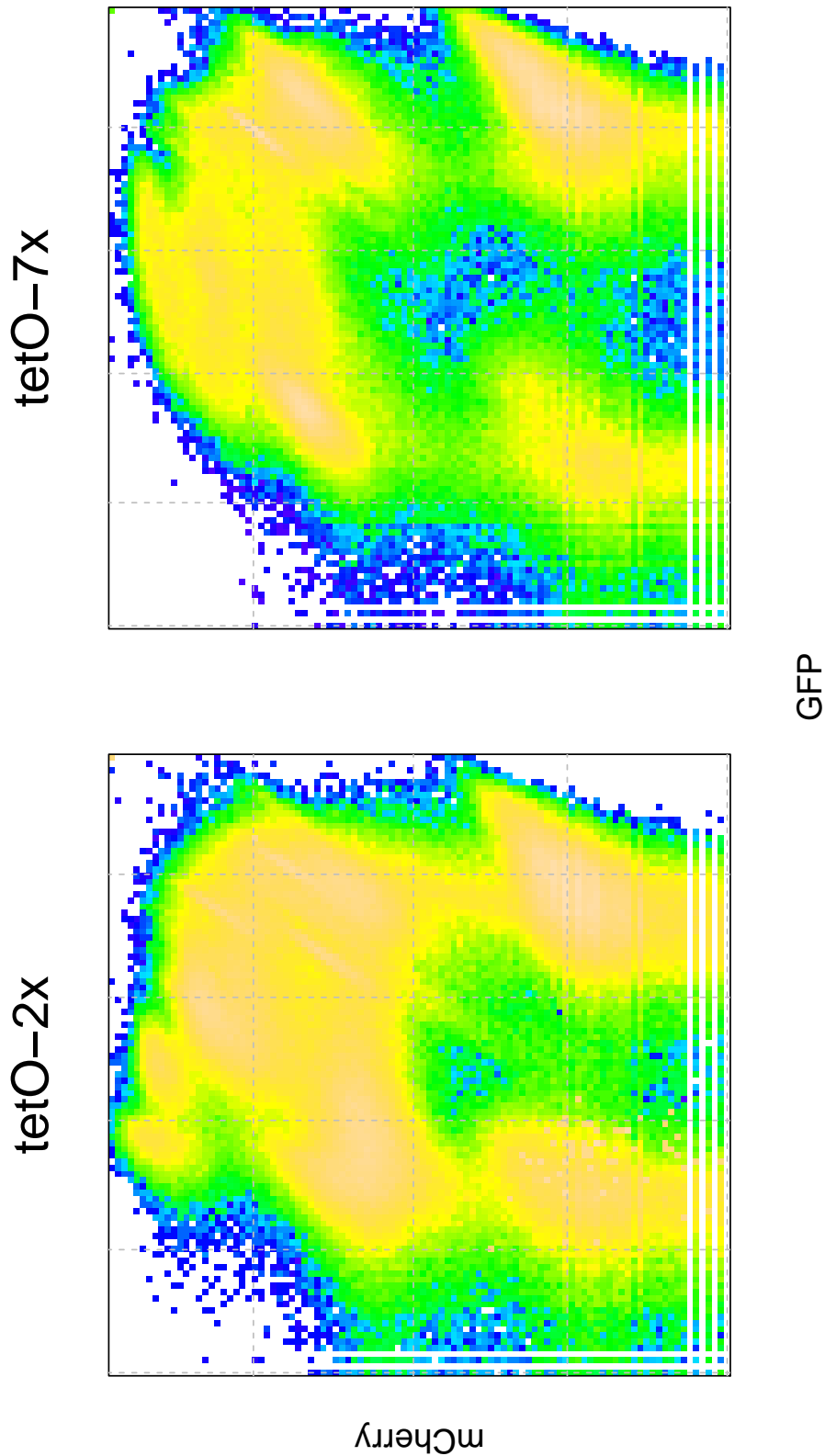
Figure 4-13: **Comparison of response profiles under tetO-2x and tetO-7x.** Response profiles from the clones of tetO-2x and tetO-7x circuit variants were merged to create masks, allowing for comparison of broad, global differences. Response profiles were processed and binned as described in the text. Both mask profiles are on log-scales for GFP and mCherry. For each axis, the distance between two dashed lines denotes a ten-fold change in expression.
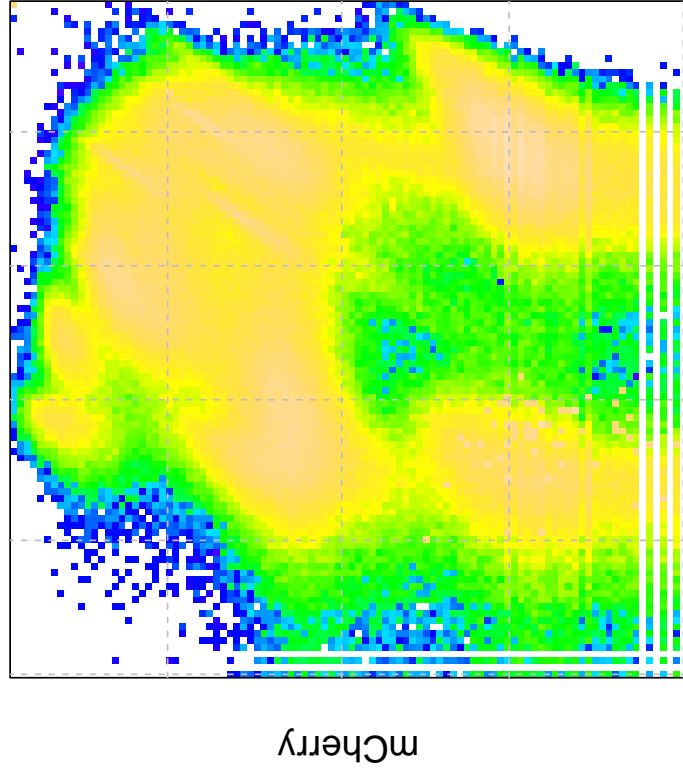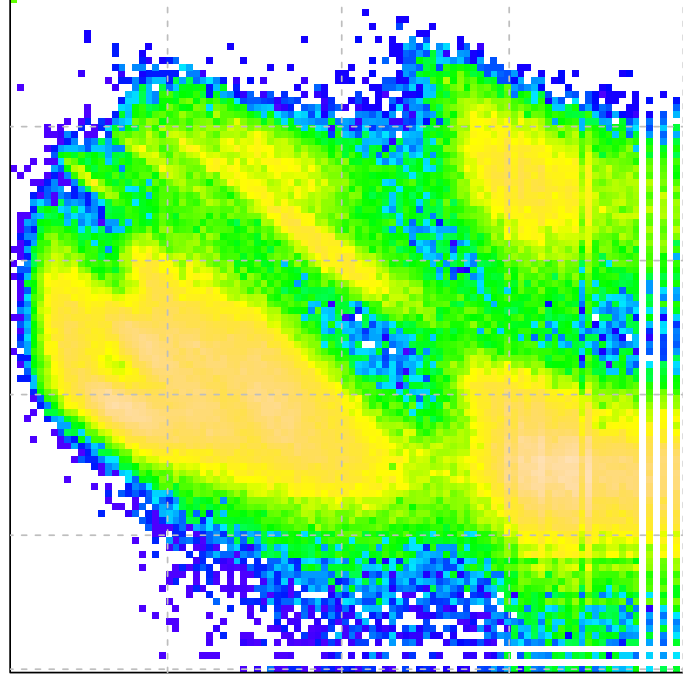
Figure 4-14: **Comparison of response profiles under different Activator:Repressor ratios.** Response profiles from the clones of 1:1 Activator:Repressor and 1:2 Activator:Repressor transformants were merged to create masks, allowing for comparison of broad, global differences. Response profiles were processed and binned as described in the text. Both mask profiles are on log-scales for GFP and mCherry. For each axis, the distance between two dashed lines denotes a ten-fold change in expression.
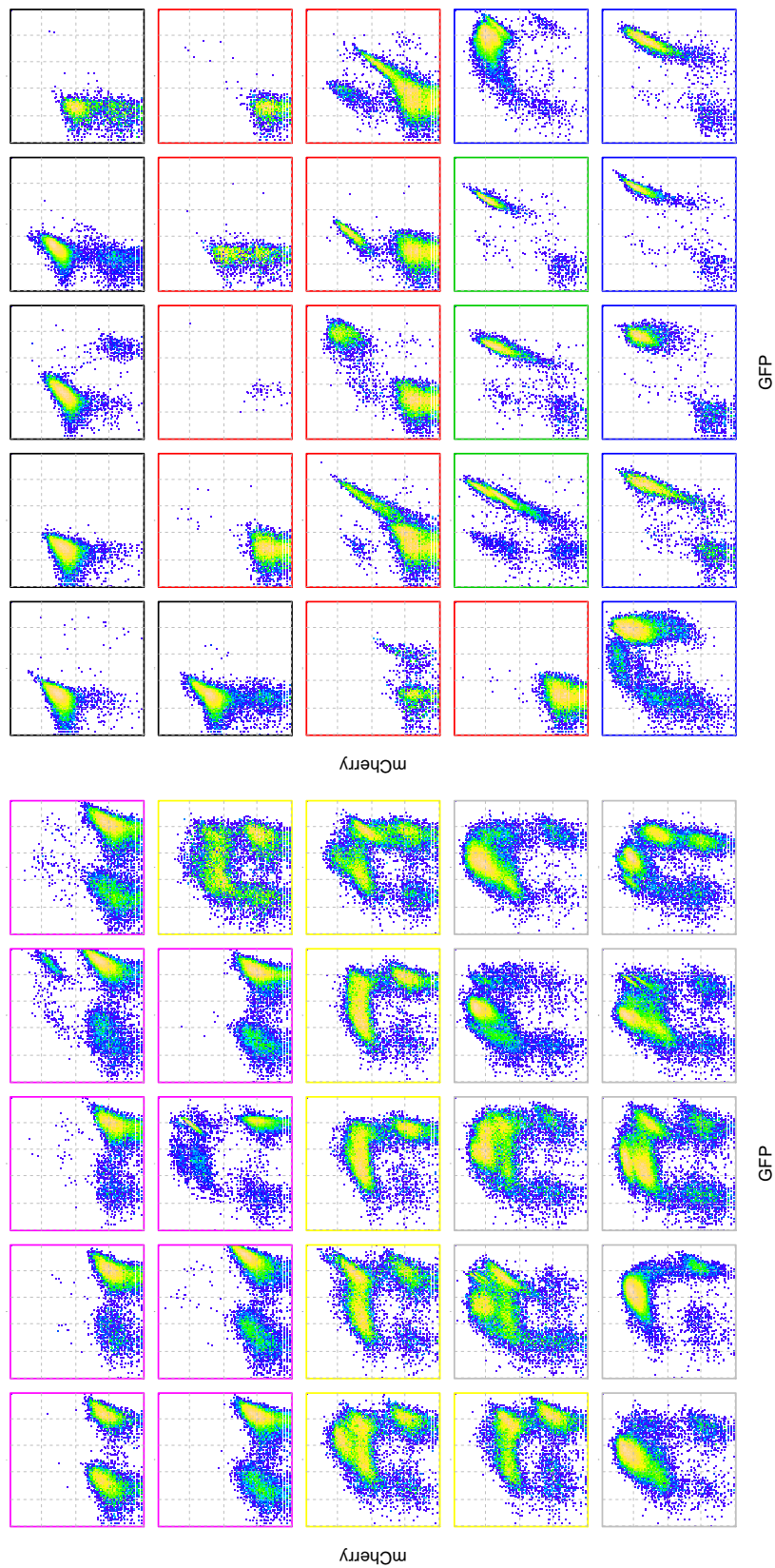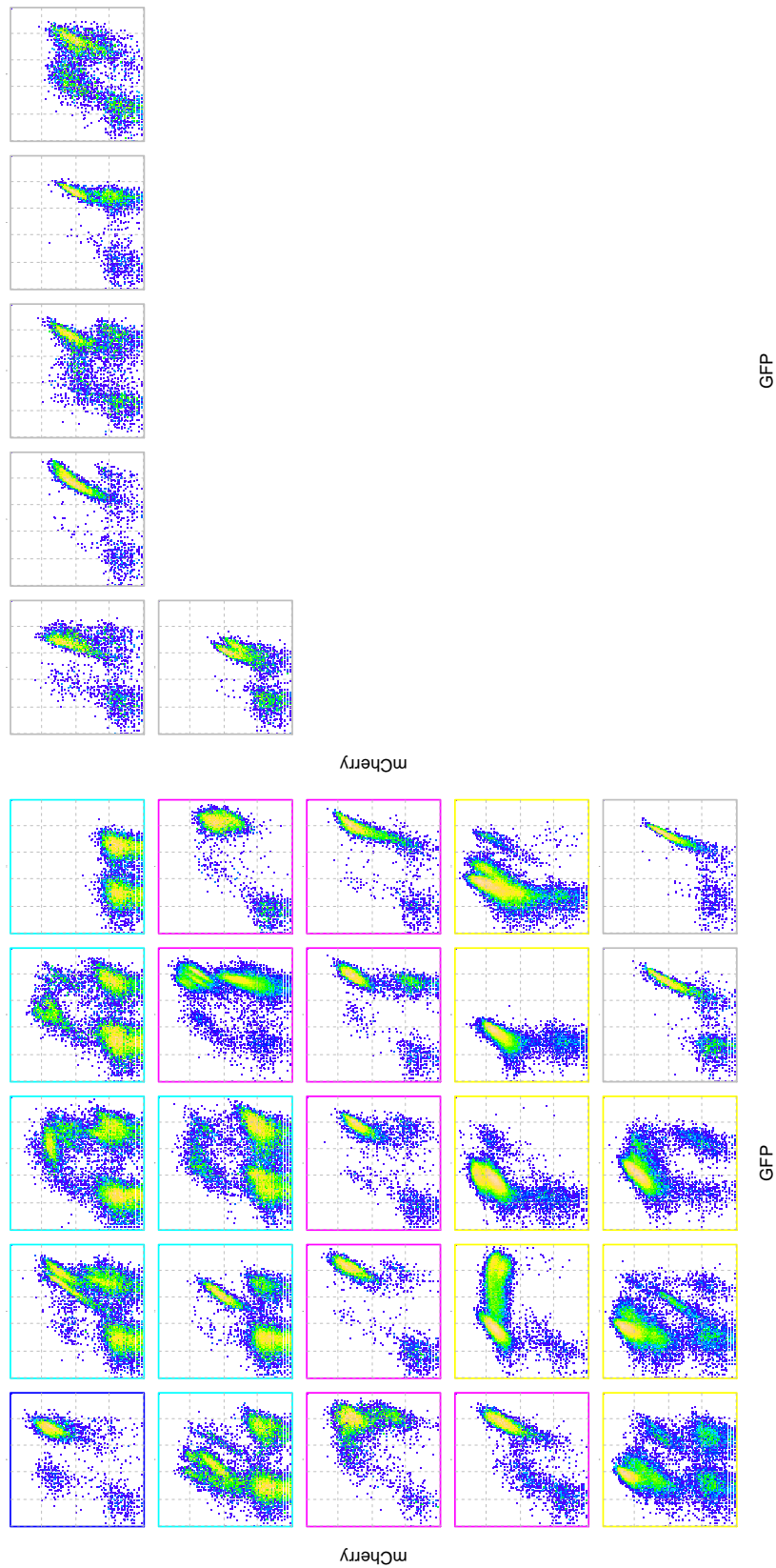
140
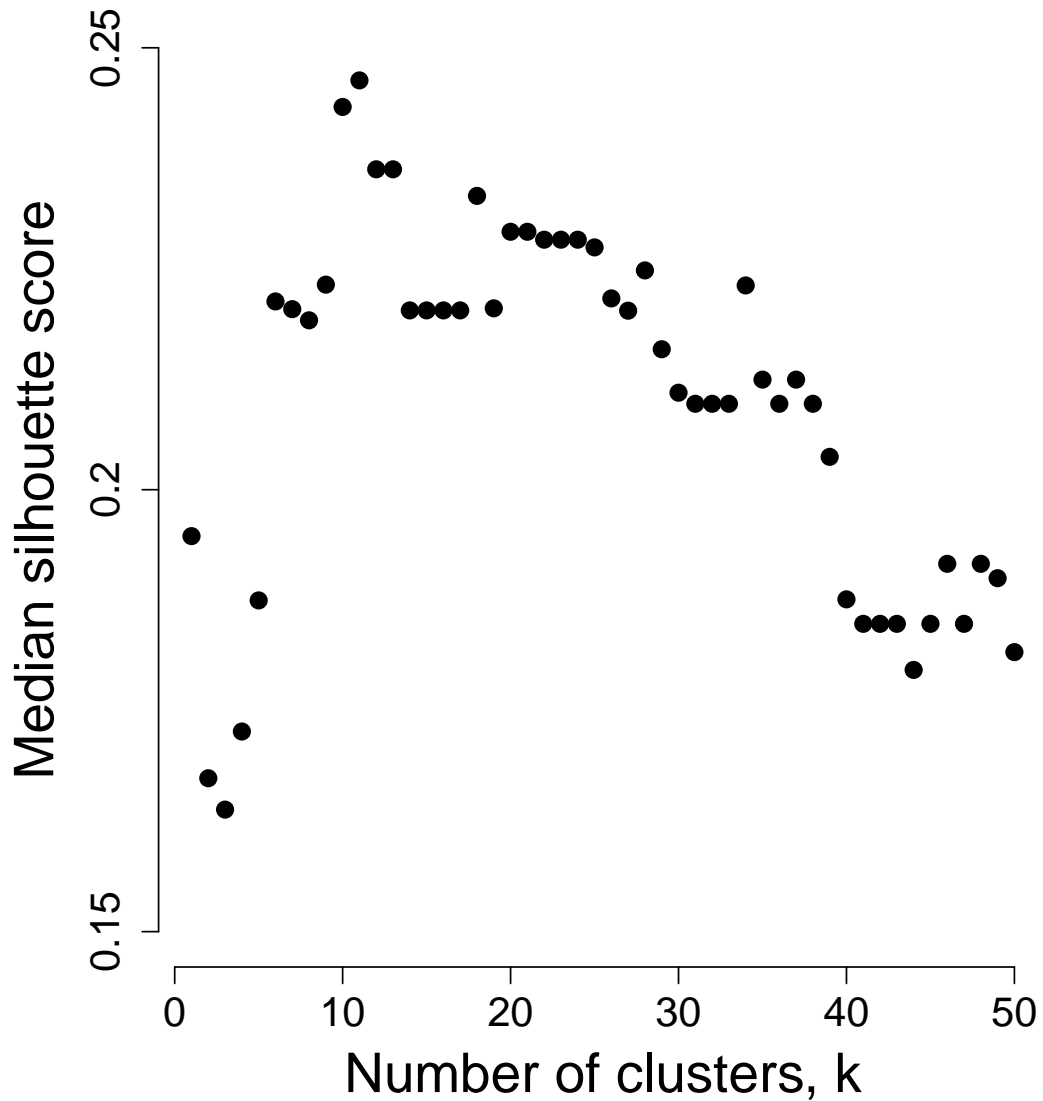
Figure 4-15: **Clustering of response profiles - silhouette values.** All response profiles obtained in the experiment were processed, binned ($n = 10$), and clustered using the Manhattan distance metric and the PAM algorithm with varying k. Median silhouette values were computed for clustering runs with varying k.
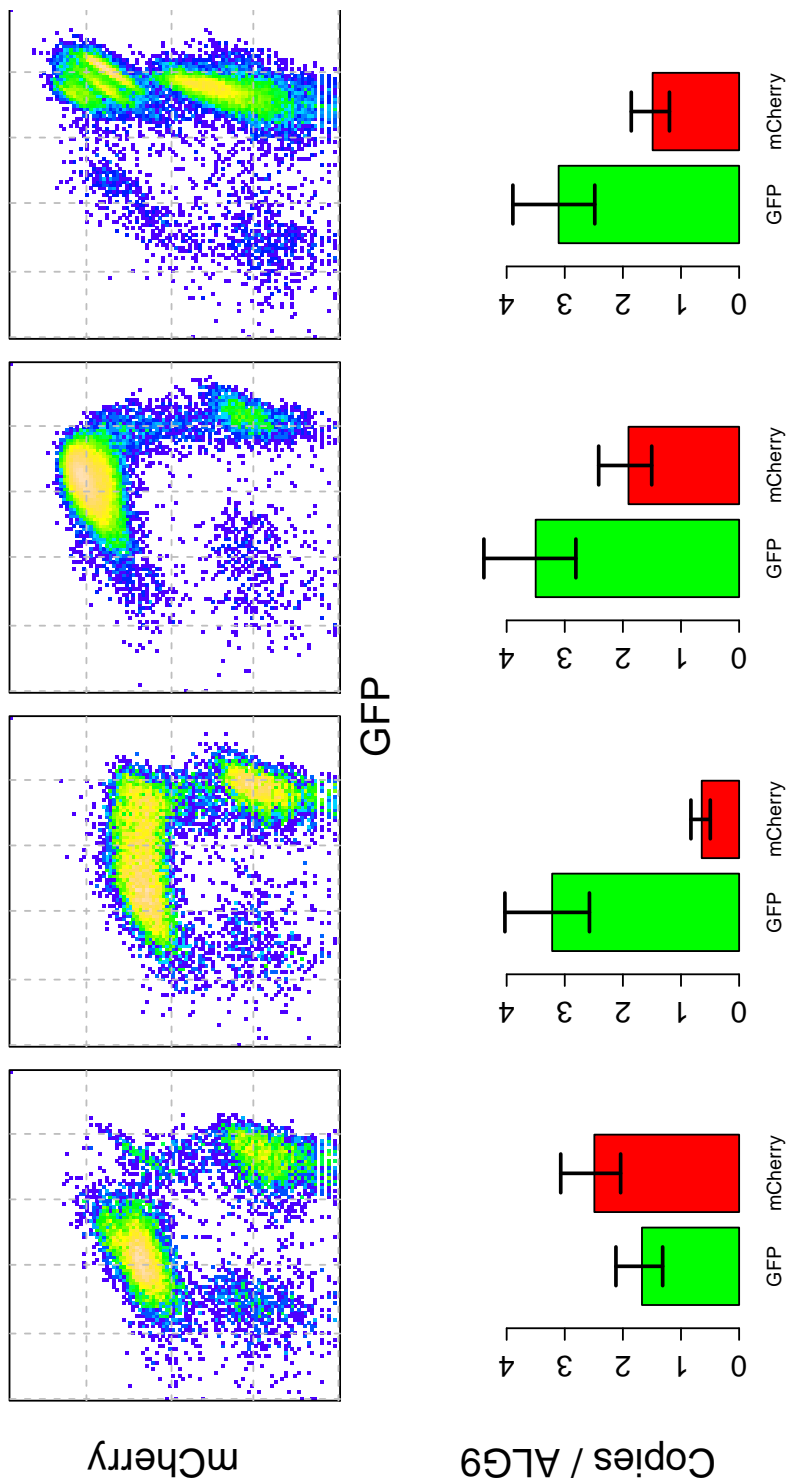
Figure 4-16: **Plasmid copy-number can modulate CAA response profile.** Selected clones from the tetO-2x, strong repression set were analyzed by flow cytometry to assess circuit expression, and by digital PCR to quantify the number of integrated copies of the two circuit plasmids. The GFP and mCherry genes were used as proxies for their respective plasmids, as detailed in Methods. The native yeast gene ALG9 was used as an endogenous control. Error-bars represent 95% confidence intervals.

Promoter architecture

```
  1  AATTCGATCTGCCGGTAGAGGTGTGGTCAATAAGAGGACCTCATACTATACCTG
 56  AGAAAGCAACCTGACCTACAGGAAAGAGTTACTCAAGAATAAGAATTTTCGTTTT
111  AAAACCTAAGAGTCACTTAAAATTTGTATACACTTATTTTTTTATAACTTATT
166  TAATAATAAAATCATAAATCATAAGAAATTCGGCTTATTTAGAAGTGGGCGAAT
221  TACTCGAGT..tTS..TTCTTAAGATCCCCGAATTGATCCGGTAATTTAGTG
276  TGTGTATTTGTGTTTGCGTGTCTATAGAAGTATAGTAATTTATGCTACAAAGGAC
331  CTAATGTATAAGGAAAGAATATTTAGAGAAAGAAGAAACAAGAGTTTTATATA
386  CATACAGAGACATGCATGCCATACCTAGG...TETO.OPERATOR...GGGCC
441  ATCGATAAAGTCGAGCTCGGTACCCTATGGCATCGATGTGCTCTGTATGTATATA
496  AAACTCTTGTTTTCTTTCTTTCTCTAAATATTCTTTCCTTATACATTAGGTCCTT
551  TGTAGCATAAATTACTATACTTCTATAGACACGCAAACACAAATACACACTAA
606  ATTACCGGATCAATTCGGGGATCC..tTA..AGGCGGCCGCTAGGGCCCTGC
661  AGGAGGGGCCGCATCATGTAATTAGTTATGTCACGCTTACATTCAGCGCCCTCCCCC
716  CACATCCGCTCAACCGAAAAGGAAGGAGTTAGACAACCTGAAGTCTAGGTCCCT
771  ATTTATTTTTTTATAGTTATGTTAGTATTAAGAAACGTTATTTATATTTCAAATTT
826  TTCTTTTTTCTGTACAGACGGCGGTGACGCATGTAACATTATACTGAAAACCTT
881  GCTTGAGAAGGTTTTGGGACGGTCGAAGGCTTTAATTTGCGGCC
```

$tetO_0$-2x

```
  1  CTAGGTCCCTATCAGTGATATAGAGAAAAGTGAAAGTCGAGTTTACCACTCCCTATC
 56  AGTGATAGAGAATCG
```

$tetO_1$-2x

```
  1  CTAGGTCCCGCTCAGTGACGGAGAAAAGTGAAAGTCGAGTTTACCACTCCCGTC
 56  AGTGACGGAGAATCG
```

$tetO_0$-7x

```
  1  CTAGGTCCCTATCAGTGATATAGAGAAAAGTGAAAGTCGAGTTTACCACTCCCTATC
 56  AGTGATAGAGAAAAGTGAAAGTCGAGTTTACCACTCCCTATCAGTGATAGAGAAA
111  AGTGAAAGTCGAGTTTACCACTCCCTATCAGTGATAGAGAAAAGTGAAAGTCGAG
166  TTTACCACTCCCTATCAGTGATAGAGAAAAGTGAAAGTCGAGTTTACCACTCCCT
221  ATCAGTGATAGAGAAAAGTGAAAGTCGAGTTTACCACTCCCTATCAGTGATAGAG
276  AATCG
```

$tetO_1$-7x

```
  1  CTAGGTCCCGCTCAGTGACGGAGAAAAGTGAAAGTCGAGTTTACCACTCCCGTC
 56  AGTGACGGAGAAAAGTGAAAGTCGAGTTTACCACTCCCGTCAGTGACGGAGAAA
111  AGTGAAAGTCGAGTTTACCACTCCCGTCAGTGACGGAGAAAAGTGAAAGTCGAG
166  TTTACCACTCCCGTCAGTGACGGAGAAAAGTGAAAGTCGAGTTTACCACTCCCC
221  GTCAGTGACGGAGAAAAGTGAAAGTCGAGTTTACCACTCCCGTCAGTGACGGAG
276  AATCG
```

Figure 4-17: **Promoter architecture for synthetic circuits.** A bi-directional promoter system was used for all synthetic circuits. Starting with the tetO operator in the middle (green), CYC1 TATA sequences were added in appropriate orientation on either side (yellow), followed by activator or repressor genes (red), and terminator sequences (gray). A total of 4 tetO operator sequence variants were used; binding sites are highlighted in green.

Chapter 5

Future Directions

Our theoretical work in switch-like behavior and experimental work in hematopoiesis and synthetic circuit engineering raise important questions about biological decision-making, and suggest a number of avenues for future exploration. Here, we describe our ongoing efforts, and outline strategies for further research.

5.1 Topology search

Our exhaustive study of all three-component network topologies exhibiting switch-like behavior demonstrates that while a large number of network topologies can yield ultrasensitive or bistable responses, only a small fraction of these do so robustly. Furthermore, robust network topologies tend to cluster tightly into topology-families, and analysis of these families can be useful for understanding the general strategies that have been evolved and selected throughout evolution.

5.1.1 Extension of framework to study additional behaviors

Just as exhaustive enumeration of network topologies has illuminated our understanding of adaptation [Ma et al., 2009] and switch-like behavior [Shah and Sarkar, 2011], it can also be readily applied to other dynamic and naturally ubiquitous behaviors

such as oscillations, with some modifications, particularly the number of parameter sets applied to each network topology, which should be determined empirically depending on the probability in uniform parameter-space with which the behavior of interest arises.

5.1.2 Extension of framework to study non-deterministic phenomena

Our approach to analyzing decision-making behavior involved deterministic simulation of network models, and hence assumed that all modeled species are present in large numbers of molecules. However, increasing evidence suggests that bistability is not a firm requirement for decision-making behavior, and specifically that bi- and multi-modality induced by stochastic phenomena can also effectively make decisions [Razooky and Weinberger, 2011, To and Maheshri, 2010, Golding, 2011]. Accordingly, a natural extension of our approach would involve re-simulation of all network topologies with the Gillespie algorithm [Gillespie, 1977], in place of numerical integration methods. Admittedly, such an application requires important considerations.

First, by its very nature, a stochastic simulation of a particular network under given parameters does not necessarily yield the same answer every time; hence, each simulation would have to be repeated enough times to allow for robust estimation. Second, in general, stochastic simulations begin to yield substantial differences from deterministic approaches as the number of molecules for species is reduced. Hence, it is possible that a stochastic simulation that does not yield bimodality, for example, does yield the behavior after species are scaled down in terms of numbers of molecules. Hence, a truly exhaustive approach would involve repeating each simulation under various biologically-realistic assumptions for numbers of molecules. Third, stochastic simulation is in general significantly more CPU-intensive than numerical integration (even when ignoring the previous point that simulations would need to be repeated many times), and can possibly render some questions impractical. However, the

general topology search approach is embarrassingly parallel, and the implementation can easily be modified to run on a cluster, as was partially done for the simulations presented in [Shah and Sarkar, 2011].

5.1.3 Identification of decision-making behavior in curated networks

While the topology search approach is useful for understanding general mechanisms driving a particular behavior, the results of such studies can also be applied to generate hypotheses regarding the existence of certain interesting behaviors. Specifically, high-ranking topologies can be used as queries into large network graphs of biological pathway databases such as WikiPathways [Pico et al., 2008], KEGG [Kanehisa et al., 2006], and BioGRID [Stark et al., 2011], and resulting hits can be analyzed further and validated experimentally. This approach could also serve as a test of the topology search strategy by comparing the graph search results to network rankings; for instance, one would expect robust topologies to appear more frequently. Computational complexity arising from the graph isomorphism problem should be an important practical consideration for such a study.

5.1.4 Refining synthetic network design strategies

As suggested earlier, topology search can be a useful tool in identifying robust strategies for programming desired behaviors in synthetic circuits. A validation of this idea has already been provided in work by our group; the top-ranking topology for generating switch-like behavior in our entire analysis was implemented synthetically in yeast using parts from multiple organisms, and demonstrated to exhibit the behavior with virtually no optimization [Palani and Sarkar, 2011]; examples of successful implementation of dynamic behaviors without labor-intensive cycles of optimization and testing are decidedly rare in the field of synthetic biology.

5.2 Hematopoiesis

Results and analysis from our work in Megakaryocyte Erythroid Progenitor commitment suggest specific hypotheses as well as general questions worth further study.

5.2.1 Hierarchical model for multi-lineage commitment

Single-cell measurements of master regulatory transcription factors provide support for the Cross Antagonism Autoregulation (CAA) network motif as a general mechanism for bipotent progenitor commitment. Recently, theoretical work has suggested that a larger network constructed from multiple instances of the CAA network layered in a hierarchy can yield discrete committed states, in a manner generally consistent with how the hematopoietic stem cell (HSC) is thought to yield mature, differentiated blood cells [Foster et al., 2009b]. For example, three CAA topologies can be layered to produce discrete decisions for four committed states (as well as three progenitor states): a CAA instance at the top of the hierarchy chooses a fate, activating the sub-tree connected to that fate, while repressing the other sub-tree, and hence eliminating the other sub-tree.

While elegant, this model does not address an important question: what if, in an uncommitted multi-potent progenitor, a master transcription factor at a lower level is up-regulated by noise, and hence commits to the relevant lineage? Under this scenario, a 'cousin' of the up-regulated master transcription can also be be up-regulated and can induce commitment to its lineage. Hence, this situation would lead to a corrupted state; i.e., a state that should not be allowed. In the context of hematopoiesis, the question is somewhat akin to asking what happens when, in an HSC, EKLF is forcibly up-regulated. A simple solution to this problem is to add positive feedback from each child master transcription factor to its parent. This approach can ensure that the overall system is consistent, regardless of the level in the hierarchy at which commitment is induced.

Clearly, simple models such as the CAA network topology cannot be realistically

assumed to be all-encompassing; however, analysis of experimental observations in the light of these models can be quite beneficial in unraveling the complexity of stem-cell lineage commitment.

5.2.2   EKLF, FLI-1 dynamics in multi-potent progenitors

CAA model simulations and erythrocyte and megakaryocyte experiments in the UT-7/GM cell-line demonstrate that a strong negative correlation between consistent with mutual antagonism develops between EKLF and FLI-1 during commitment of the bipotent progenitor. However, modeling work (and the hierarchical model described above) yields the hypothesis that the master transcription factors EKLF, FLI-1 are positively correlated en-route to the bipotent progenitor state from the multipotent progenitor state. Since UT-7/GM cells are not multi-potent, we were not able to test this hypothesis. However, a multipotent hematopoietic cell-line such as the murine FDCPmix [Spooncer et al., 1986] line can provide a platform for addressing this specific question, as well as the more general hierarchical model for multipotent progenitor commitment.

5.2.3   Validation in primary hematopoietic cells

Compared to primary cells, cell-lines offer a cleaner, more convenient platform for studying lineage commitment and allow researchers to more easily connect perturbations to responses, given the absence of myriad additional sources of variation in primary cells. However, cell-lines can potentially mask or modify important dynamics and behaviors that hematopoietic progenitors exhibit in the physiological context. Hence, validation in progenitors isolated from primary cells, of our findings from experiments in the UT-7/GM cell line can serve as an additional test of the CAA model, and our understanding of hematopoietic lineage commitment.

5.2.4 Bypassing of progenitor states

The general model of interpreting lineage commitment as a dynamical system with stable progenitor and mature cells represented as attractors points to important implications. In this view, most progenitor cells sequentially pass through defined, more restricted progenitor states on their way to becoming mature, committed cells; however, this view also leaves open the possibility of small proportions of the progenitor populations committing via alternative routes, particularly by avoiding intermediate, restricted progenitors. Furthermore, under the CAA multi-lineage commitment model, a progenitor cell's predilection for a mature, committed state would be reflected in its levels of master transcription factors.

An experimental strategy for studying this effect could involve identifying multi-potent progenitors poised towards an alternative route, and culturing them in an enriched population to study their trajectories through gene-expression-space. Although robust quantification of transcription factor levels in live, individual cells can be quite challenging, correlations between these proteins and cell-surface markers can perhaps be exploited.

5.2.5 Receptor feedback as a key mediator of decision-making

Results from our experiments in the UT-7/GM cell line demonstrate a strong, positive correlation between the expression of the pro-erythroid transcription factor EKLF, and the cognate receptor, EpoR, throughout the lineage commitment process; furthermore, our results also reveal a negative correlation between EKLF and TpoR, the cognate receptor for the opposing megakaryoid lineage. Previous systems modeling work [Palani and Sarkar, 2008, Palani and Sarkar, 2009] has suggested that more than a consequence, EpoR can in fact be a key driver of pro-erythroid decisions, through up-regulation of the EpoR gene by pro-erythroid transcription factors such as EKLF or GATA-1.

An ideal test of this mechanism would involve specifically blocking GATA-1- or

EKLF-mediated up-regulation of EpoR; however, it is unclear whether there is a practical strategy for conducting such an experiment. Alternatively, one could investigate whether aberrant expression of EpoR or TpoR can disrupt commitment to the opposing lineage. For instance, UT-7/GM cells could be stably transfected with a simple circuit that allows for expression of EpoR and TpoR up-regulation via separate drugs or other extrinsic factors. During erythrocyte lineage-commitment, EpoR is up-regulated, and TpoR is down-regulated; the effects of drug-induced over-expression of TpoR during the commitment process can be illuminating. According to the full receptor-feedback model [Palani and Sarkar, 2008, Palani and Sarkar, 2009], sufficiently strong over-expression of TpoR would activate the opposing pro-megakaryoid transcription factors, and suppress expression to the erythrocyte lineage. In summary, a population cultured in the presence of both Erythropoietin and Thrombopoietin and induced for TpoR over-expression would be expected to yield fewer, if any, erythrocytes. Analogously, forced expression of EpoR alone, or in conjunction with TpoR could also be informative.

5.3   Synthetic biology

5.3.1   Cross-species synthetic biology

As discussed earlier, a key aspect that makes standardization difficult and circuit construction ad-hoc, is that, unlike electronic systems, synthetic biological circuits are deployed in the context of living hosts. While yeast and bacteria have understandably been the workhorses of synthetic biology research efforts, these hosts are not the ideal vehicles for delivering drug payloads, producing chemicals, and performing the myriad other tasks that synthetic biologists hope to implement; hence, it is crucial for the field to develop general tools and strategies to facilitate synthetic biology in a variety of hosts.

While a sizable number of individual genes have been heterologously expressed, to

our knowledge, there has not been a demonstration of a circuit exhibiting a non-trivial, dynamic behavior in more than one host. Such an example would be a testament to robust design. Toward this end, we are attempting to port the synthetic CAA circuit from yeast to mammalian cells (specifically, HeLa cells). The circuit design and implementation is largely the same, except for two minor details: the yeast-specific SSN6 repression domain has been replaced with the kid the KRAB repression domain, and the yeast-specific CYC1 minimal promoter has been replaced with a minimal mammalian CMV promoter. Further work on this project can lead to important insights about decision-making as well as synthetic circuit design and implementation.

5.3.2 Orthogonal strategy for making multiple decisions

At least theoretically, the multiple instances of the CAA circuit can be linked hierarchically to decide between 2^n choices, where n is the number of levels in the hierarchy. Extending the synthetic CAA circuit in this manner could be informative for understanding decision-making. Of course, a three CAA-instance circuit necessitates six orthogonal master transcription factors, further highlighting the importance of this issue in synthetic biology.

While engineering additional orthogonal specificities for tetR may be possible, it would be prudent to move to a more accommodating platform. In particular, the recent push in the engineering of TALENs (Transcription activator-like effectors) has generated a number of orthogonal DNA-binding proteins as well as computational methods that claim to design facilitate the design of such parts [Doyle et al., 2012].
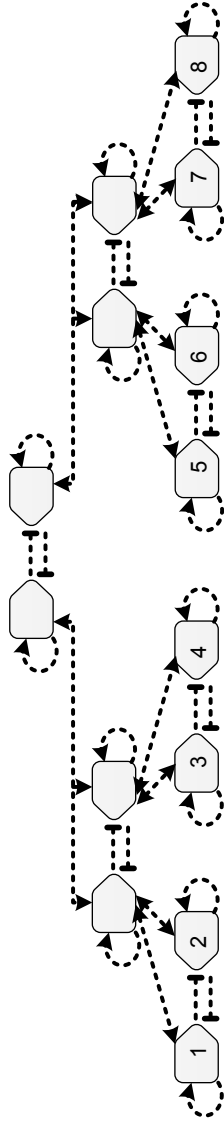
Figure 5-1: **Proposed hierarchical model for multi-lineage commitment.** Multiple instances of the CAA topology can be layered hierarchically to construct a network that can decide between multiple lineages. The complete network in this figure can decide between eight lineages. The basic model was introduced in [Foster et al., 2009b]. We propose introducing additional positive feedback interactions from each node to its immediate progenitor (parent) node; this prevents corrupt network states, as described in the text.

6

Bibliography

[Abkowitz et al., 1996] Abkowitz, J. L., Catlin, S. N., and Guttorp, P. (1996). Evidence that hematopoiesis may be a stochastic process in vivo. *Nature medicine*, 2(2):190–7.

[Adolfsson et al., 2005] Adolfsson, J., Må nsson, R., Buza-Vidas, N., Hultquist, A., Liuba, K., Jensen, C. T., Bryder, D., Yang, L., Borge, O.-J., Thoren, L. A. M., Anderson, K., Sitnicka, E., Sasaki, Y., Sigvardsson, M., and Jacobsen, S. E. W. (2005). Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment. *Cell*, 121(2):295–306.

[Akashi et al., 2000] Akashi, K., Traver, D., Miyamoto, T., and Weissman, I. L. (2000). A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature*, 404(6774):193–7.

[Alon, 2007] Alon, U. (2007). Network motifs: theory and experimental approaches. *Nature reviews. Genetics*, 8(6):450–61.

[Andrianantoandro et al., 2006] Andrianantoandro, E., Basu, S., Karig, D. K., and Weiss, R. (2006). Synthetic biology: new engineering rules for an emerging discipline. *Molecular systems biology*, 2:2006.0028.

[Angeli et al., 2004] Angeli, D., Ferrell, J. E., and Sontag, E. D. (2004). Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems. *Proceedings of the National Academy of Sciences of the United States of America*, 101(7):1822–7.

[Arkin et al., 1998] Arkin, A., Ross, J., and McAdams, H. H. (1998). Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected Escherichia coli cells. *Genetics*, 149(4):1633–48.

[Bagci et al., 2006] Bagci, E. Z., Vodovotz, Y., Billiar, T. R., Ermentrout, G. B., and Bahar, I. (2006). Bistability in apoptosis: roles of bax, bcl-2, and mitochondrial permeability transition pores. *Biophysical journal*, 90(5):1546–59.

[Bagowski et al., 2003] Bagowski, C. P., Besser, J., Frey, C. R., and Ferrell Jr., J. E. (2003). The JNK Cascade as a Biochemical Switch in Mammalian CellsUltrasensitive and All-or-None Responses. *Current Biology*, 13(4):315–320.

[Balázsi et al., 2011] Balázsi, G., van Oudenaarden, A., and Collins, J. J. (2011). Cellular decision making and biological noise: from microbes to mammals. *Cell*, 144(6):910–25.

[Bastian et al., 1999] Bastian, L. S., Kwiatkowski, B. A., Breininger, J., Danner, S., and Roth, G. (1999). Regulation of the Megakaryocytic Glycoprotein IX Promoter by the Oncogenic Ets Transcription Factor Fli-1. *Blood*, 93(8):2637–2644.

[Batish et al., 2011] Batish, M., Raj, A., and Tyagi, S. (2011). Single molecule imaging of RNA in situ. *Methods in molecular biology (Clifton, N.J.)*, 714:3–13.

[Becskei et al., 2001] Becskei, A., Séraphin, B., and Serrano, L. (2001). Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *The EMBO journal*, 20(10):2528–35.

[Bellí et al., 1998] Bellí, G., Garí, E., Piedrafita, L., Aldea, M., and Herrero, E. (1998). An activator/repressor dual system allows tight tetracycline-regulated gene expression in budding yeast. *Nucleic acids research*, 26(4):942–7.

[Berens and Hillen, 2003] Berens, C. and Hillen, W. (2003). Gene regulation by tetracyclines. Constraints of resistance regulation in bacteria shape TetR for application in eukaryotes. *European Journal of Biochemistry*, 270(15):3109–3121.

[Black, 2006] Black, P. (2006). *Dictionary of Algorithms and Data Structures*.

[Borg et al., 2010] Borg, J., Papadopoulos, P., Georgitsi, M., Gutiérrez, L., Grech, G., Fanis, P., Phylactides, M., Verkerk, A. J. M. H., van der Spek, P. J., Scerri, C. A., Cassar, W., Galdies, R., van Ijcken, W., Ozgür, Z., Gillemans, N., Hou, J., Bugeja, M., Grosveld, F. G., von Lindern, M., Felice, A. E., Patrinos, G. P., and Philipsen, S. (2010). Haploinsufficiency for the erythroid transcription factor KLF1 causes hereditary persistence of fetal hemoglobin. *Nature genetics*, 42(9):801–5.

[Bouilloux et al., 2008] Bouilloux, F., Juban, G., Cohet, N., Buet, D., Guyot, B., Vainchenker, W., Louache, F., and Morlé, F. (2008). EKLF restricts megakaryocytic differentiation at the benefit of erythrocytic differentiation. *Blood*, 112(3):576–84.

[Breitkreutz et al., 2010] Breitkreutz, A., Choi, H., Sharom, J. R., Boucher, L., Neduva, V., Larsen, B., Lin, Z. Y., Breitkreutz, B. J., Stark, C., Liu, G., Ahn, J., Dewar-Darch, D., Reguly, T., Tang, X., Almeida, R., Qin, Z. S., Pawson, T., Gingras, A. C., Nesvizhskii, A. I., and Tyers, M. (2010). A Global Protein Kinase and Phosphatase Interaction Network in Yeast. *Science*, 328(5981):1043–1046.

[Broudy et al., 1991] Broudy, V., Lin, N., Brice, M., Nakamoto, B., and Papayannopoulou, T. (1991). Erythropoietin receptor characteristics on primary human erythroid cells. *Blood*, 77(12):2583–2590.

[Brown et al., 1997] Brown, G. C., Hoek, J. B., and Kholodenko, B. N. (1997). Why do protein kinase cascades have more than one level? *Trends in biochemical sciences*, 22(8):288.

[Burrill and Silver, 2010] Burrill, D. R. and Silver, P. A. (2010). Making cellular memories. *Cell*, 140(1):13–8.

[Cantor and Orkin, 2001] Cantor, A. B. and Orkin, S. H. (2001). Hematopoietic development: a balancing act. *Current opinion in genetics & development*, 11(5):513–9.

[Cantor and Orkin, 2002] Cantor, A. B. and Orkin, S. H. (2002). Transcriptional regulation of erythropoiesis: an affair involving multiple partners. *Oncogene*, 21(21):3368–76.

[Chang and Karin, 2001] Chang, L. and Karin, M. (2001). Mammalian MAP kinase signalling cascades. *Nature*, 410(6824):37–40.

[Chen et al., 1995] Chen, H., Ray-Gallet, D., Zhang, P., Hetherington, C. J., Gonzalez, D. A., Zhang, D. E., Moreau-Gachelin, F., and Tenen, D. G. (1995). PU.1 (Spi-1) autoregulates its expression in myeloid cells. *Oncogene*, 11(8):1549–60.

[Chen and Weiss, 2005] Chen, M.-T. and Weiss, R. (2005). Artificial cell-cell communication in yeast Saccharomyces cerevisiae using signaling elements from Arabidopsis thaliana. *Nature biotechnology*, 23(12):1551–5.

[Cherry and Adler, 2000] Cherry, J. L. and Adler, F. R. (2000). How to make a biological switch. *Journal of theoretical biology*, 203(2):117–33.

[Chiba et al., 1991] Chiba, T., Ikawa, Y., and Todokoro, K. (1991). GATA-1 transactivates erythropoietin receptor gene, and erythropoietin receptor-mediated signals enhance GATA-1 gene expression. *Nucleic acids research*, 19(14):3843–8.

[Chickarmane and Peterson, 2008] Chickarmane, V. and Peterson, C. (2008). A computational model for understanding stem cell, trophectoderm and endoderm lineage determination. *PloS one*, 3(10):e3478.

155

[Constantinescu et al., 1999] Constantinescu, S., Ghaffari, S., and Lodish, H. (1999). The Erythropoietin Receptor: Structure, Activation and Intracellular Signal Transduction. *Trends in endocrinology and metabolism: TEM*, 10(1):18–23.

[Cress and Triezenberg, 1991] Cress, W. and Triezenberg, S. (1991). Critical structural elements of the VP16 transcriptional activation domain. *Science*, 251(4989):87–90.

[Cui et al., 2009] Cui, J.-W., Vecchiarelli-Federico, L. M., Li, Y.-J., Wang, G.-J., and Ben-David, Y. (2009). Continuous Fli-1 expression plays an essential role in the proliferation and survival of F-MuLV-induced erythroleukemia and human erythroleukemia. *Leukemia*, 23(7):1311–9.

[DeGracia et al., 2012] DeGracia, D. J., Huang, Z.-F., and Huang, S. (2012). A nonlinear dynamical theory of cell injury. *Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism*, 32(6):1000–13.

[Doyle et al., 2012] Doyle, E. L., Booher, N. J., Standage, D. S., Voytas, D. F., Brendel, V. P., Vandyk, J. K., and Bogdanove, A. J. (2012). TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic acids research*, 40(Web Server issue):W117–22.

[Eckfeldt et al., 2005] Eckfeldt, C. E., Mendenhall, E. M., and Verfaillie, C. M. (2005). The molecular repertoire of the 'almighty' stem cell. *Nature reviews. Molecular cell biology*, 6(9):726–37.

[Eissing et al., 2004] Eissing, T., Conzelmann, H., Gilles, E. D., Allgöwer, F., Bullinger, E., and Scheurich, P. (2004). Bistability analyses of a caspase activation model for receptor-induced apoptosis. *The Journal of biological chemistry*, 279(35):36892–7.

[Elowitz and Leibler, 2000] Elowitz, M. B. and Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335–8.

[Elowitz et al., 2002] Elowitz, M. B., Levine, A. J., Siggia, E. D., and Swain, P. S. (2002). Stochastic gene expression in a single cell. *Science (New York, N.Y.)*, 297(5584):1183–6.

[Endy, 2005] Endy, D. (2005). Foundations for engineering biology. *Nature*, 438(7067):449–53.

[Enver et al., 1998] Enver, T., Heyworth, C. M., and Dexter, T. M. (1998). Do stem cells play dice? *Blood*, 92(2):348–51; discussion 352.

[Enver et al., 2009] Enver, T., Pera, M., Peterson, C., and Andrews, P. W. (2009). Stem cell states, fates, and the rules of attraction. *Cell stem cell*, 4(5):387–97.

[Evans and Felsenfeld, 1989] Evans, T. and Felsenfeld, G. (1989). The erythroid-specific transcription factor Eryf1: a new finger protein. *Cell*, 58(5):877–85.

[Ferrell, 2002] Ferrell, J. E. (2002). Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Current opinion in cell biology*, 14(2):140–8.

[Ferrell, 2008] Ferrell, J. E. (2008). Feedback regulation of opposing enzymes generates robust, all-or-none bistable responses. *Current biology : CB*, 18(6):R244–5.

[Ferrell and Xiong, 2001] Ferrell, J. E. and Xiong, W. (2001). Bistability in cell signaling: How to make continuous processes discontinuous, and reversible processes irreversible. *Chaos (Woodbury, N.Y.)*, 11(1):227–236.

[Ferrell Jr., 1998] Ferrell Jr., J. E. (1998). The Biochemical Basis of an All-or-None Cell Fate Switch in Xenopus Oocytes. *Science*, 280(5365):895–898.

[Foster et al., 2009a] Foster, D. V., Foster, J. G., Huang, S., and Kauffman, S. A. (2009a). A model of sequential branching in hierarchical cell fate determination. *Journal of theoretical biology*, 260(4):589–97.

[Foster et al., 2009b] Foster, D. V., Foster, J. G., Huang, S., and Kauffman, S. A. (2009b). A model of sequential branching in hierarchical cell fate determination. *Journal of theoretical biology*, 260(4):589–97.

[Frontelo et al., 2007] Frontelo, P., Manwani, D., Galdass, M., Karsunky, H., Lohmann, F., Gallagher, P. G., and Bieker, J. J. (2007). Novel role for EKLF in megakaryocyte lineage commitment. *Blood*, 110(12):3871–3880.

[Gardner et al., 2000] Gardner, T. S., Cantor, C. R., and Collins, J. J. (2000). Construction of a genetic toggle switch in Escherichia coli. *Nature*, 403(6767):339–42.

[Ghaffari et al., 2001] Ghaffari, S., Kitidis, C., Fleming, M. D., Neubauer, H., Pfeffer, K., and Lodish, H. F. (2001). Erythropoiesis in the absence of janus-kinase 2: BCR-ABL induces red cell formation in JAK2(-/-) hematopoietic progenitors. *Blood*, 98(10):2948–57.

[Gietz and Schiestl, 2007] Gietz, R. D. and Schiestl, R. H. (2007). High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nature protocols*, 2(1):31–4.

[Gillespie, 1977] Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361.

[Goldbeter, 2005] Goldbeter, A. (2005). Zero-order switches and developmental thresholds. *Molecular systems biology*, 1:2005.0031.

[Goldbeter and Koshland, 1981] Goldbeter, A. and Koshland, D. E. (1981). An Amplified Sensitivity Arising from Covalent Modification in Biological Systems. *Proceedings of the National Academy of Sciences*, 78(11):6840–6844.

[Golding, 2011] Golding, I. (2011). Decision making in living cells: lessons from a simple system. *Annual review of biophysics*, 40:63–80.

[Graf and Enver, 2009] Graf, T. and Enver, T. (2009). Forcing cells to change lineages. *Nature*, 462(7273):587–94.

[Grass et al., 2003] Grass, J. A., Boyer, M. E., Pal, S., Wu, J., Weiss, M. J., and Bresnick, E. H. (2003). GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling. *Proceedings of the National Academy of Sciences of the United States of America*, 100(15):8811–6.

[Greer and Wintrobe, 2008] Greer, J. P. and Wintrobe, M. M. (2008). *Wintrobe's clinical hematology*. Lippincott Williams & Wilkins.

[Hammer et al., 2006] Hammer, K., Mijakovic, I., and Jensen, P. R. (2006). Synthetic promoter libraries–tuning of gene expression. *Trends in biotechnology*, 24(2):53–5.

[Hannon et al., 1991] Hannon, R., Evans, T., Felsenfeld, G., and Gould, H. (1991). Structure and promoter activity of the gene for the erythroid transcription factor GATA-1. *Proceedings of the National Academy of Sciences of the United States of America*, 88(8):3004–8.

[Hartwell et al., 1999] Hartwell, L. H., Hopfield, J. J., Leibler, S., and Murray, A. W. (1999). From molecular to modular cell biology. *Nature*, 402(6761 Suppl):C47–52.

[Hazzalin and Mahadevan, 2002] Hazzalin, C. A. and Mahadevan, L. C. (2002). MAPK-regulated transcription: a continuously variable gene switch? *Nature reviews. Molecular cell biology*, 3(1):30–40.

[Helias et al., 2013] Helias, V., Saison, C., Peyrard, T., Vera, E., Prehu, C., Cartron, J.-P., and Arnaud, L. (2013). Molecular analysis of the rare in(Lu) blood type: toward decoding the phenotypic outcome of haploinsufficiency for the transcription factor KLF1. *Human mutation*, 34(1):221–8.

[Hohaus et al., 1995] Hohaus, S., Petrovick, M. S., Voso, M. T., Sun, Z., Zhang, D. E., and Tenen, D. G. (1995). PU.1 (Spi-1) and C/EBP alpha regulate expression of the granulocyte-macrophage colony-stimulating factor receptor alpha gene. *Molecular and cellular biology*, 15(10):5830–45.

[Hooshangi et al., 2005] Hooshangi, S., Thiberge, S., and Weiss, R. (2005). Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proceedings of the National Academy of Sciences of the United States of America*, 102(10):3581–6.

[Huang and Ferrell, 1996] Huang, C. Y. and Ferrell, J. E. (1996). Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proceedings of the National Academy of Sciences of the United States of America*, 93(19):10078–83.

[Huang, 2009] Huang, S. (2009). Reprogramming cell fates: reconciling rarity with robustness. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 31(5):546–60.

[Huang et al., 2005] Huang, S., Eichler, G., Bar-Yam, Y., and Ingber, D. E. (2005). Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Physical review letters*, 94(12):128701.

[Huang et al., 2007] Huang, S., Guo, Y.-P., May, G., and Enver, T. (2007). Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Developmental biology*, 305(2):695–713.

[Iman et al., 1980] Iman, R., Davenport, J., and Zeigler, D. (1980). Latin hypercube sampling (program user's guide). Technical report, Sandia Labs, Albuquerque.

[Ingham and McMahon, 2001] Ingham, P. W. and McMahon, A. P. (2001). Hedgehog signaling in animal development: paradigms and principles. *Genes & development*, 15(23):3059–87.

[Inoue et al., 2001] Inoue, T., Higuchi, M., Hashimoto, Y., Seki, M., Kobayashi, M., Kato, T., Tabata, S., Shinozaki, K., and Kakimoto, T. (2001). Identification of CRE1 as a cytokinin receptor from Arabidopsis. *Nature*, 409(6823):1060–3.

[Iwasaki and Akashi, 2007] Iwasaki, H. and Akashi, K. (2007). Myeloid lineage commitment from the hematopoietic stem cell. *Immunity*, 26(6):726–40.

[Iwasaki et al., 2003] Iwasaki, H., Mizuno, S.-i., Wells, R. A., Cantor, A. B., Watanabe, S., and Akashi, K. (2003). GATA-1 converts lymphoid and myelomonocytic progenitors into the megakaryocyte/erythrocyte lineages. *Immunity*, 19(3):451–62.

[Kanehisa et al., 2006] Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K. F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., and Hirakawa, M. (2006). From genomics to chemical genomics: new developments in KEGG. *Nucleic acids research*, 34(Database issue):D354–7.

[Kaushansky, 2005] Kaushansky, K. (2005). The molecular mechanisms that control thrombopoiesis. *The Journal of clinical investigation*, 115(12):3339–47.

[Kaushansky, 2008] Kaushansky, K. (2008). Historical review: megakaryopoiesis and thrombopoiesis. *Blood*, 111(3):981–6.

[Kawada et al., 2001] Kawada, H., Ito, T., Pharr, P. N., Spyropoulos, D. D., Watson, D. K., and Ogawa, M. (2001). Defective megakaryopoiesis and abnormal erythroid development in Fli-1 gene-targeted mice. *International journal of hematology*, 73(4):463–8.

[Kholodenko et al., 2010] Kholodenko, B. N., Hancock, J. F., and Kolch, W. (2010). Signalling ballet in space and time. *Nature reviews. Molecular cell biology*, 11(6):414–26.

[Kholodenko et al., 1997] Kholodenko, B. N., Hoek, J. B., Westerhoff, H. V., and Brown, G. C. (1997). Quantification of information transfer via cellular signal transduction pathways. *FEBS letters*, 414(2):430–4.

[Kisseleva et al., 2002] Kisseleva, T., Bhattacharya, S., Braunstein, J., and Schindler, C. W. (2002). Signaling through the JAK/STAT pathway, recent advances and future challenges. *Gene*, 285(1-2):1–24.

[Kitano, 2002] Kitano, H. (2002). Systems biology: a brief overview. *Science (New York, N.Y.)*, 295(5560):1662–4.

[Komatsu et al., 1997] Komatsu, N., Kirito, K., Shimizu, R., Kunitama, M., Yamada, M., Uchida, M., Takatoku, M., Eguchi, M., and Miura, Y. (1997). In vitro development of erythroid and megakaryocytic cells from a UT-7 subline, UT-7/GM. *Blood*, 89(11):4021–33.

[Kondo et al., 2010] Kondo, M., Moriguchi, T., Suzuki, M., Engel, J. D., and Yamamoto, M. (2010). *Hematopoietic Stem Cell Biology*. Humana Press, Totowa, NJ.

[Kondo et al., 2003] Kondo, M., Wagers, A. J., Manz, M. G., Prohaska, S. S., Scherer, D. C., Beilhack, G. F., Shizuru, J. A., and Weissman, I. L. (2003). Biology of hematopoietic stem cells and progenitors: implications for clinical application. *Annual review of immunology*, 21:759–806.

[Kondo et al., 1997] Kondo, M., Weissman, I. L., and Akashi, K. (1997). Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell*, 91(5):661–72.

[Koshland et al., 1982] Koshland, D. E., Goldbeter, A., and Stock, J. B. (1982). Amplification and adaptation in regulatory and sensory systems. *Science (New York, N.Y.)*, 217(4556):220–5.

[Krantz, 1991] Krantz, S. B. (1991). Erythropoietin. *Blood*, 77(3):419–34.

[Krueger et al., 2007] Krueger, M., Scholz, O., Wisshak, S., and Hillen, W. (2007). Engineered Tet repressors with recognition specificity for the tetO-4C5G operator variant. *Gene*, 404(1-2):93–100.

[Kuramochi et al., 1990] Kuramochi, S., Ikawa, Y., and Todokoro, K. (1990). Characterization of murine erythropoietin receptor genes. *Journal of molecular biology*, 216(3):567–75.

[Lai, 1992] Lai, Z. (1992). Negative control of photoreceptor development in Drosophila by the product of the yan gene, an ETS domain protein. *Cell*, 70(4):609–620.

[Laslo et al., 2006] Laslo, P., Spooner, C. J., Warmflash, A., Lancki, D. W., Lee, H.-J., Sciammas, R., Gantner, B. N., Dinner, A. R., and Singh, H. (2006). Multilineage transcriptional priming and determination of alternate hematopoietic cell fates. *Cell*, 126(4):755–66.

[Lee and Maheshri, 2012] Lee, T.-H. and Maheshri, N. (2012). A regulatory role for repeated decoy transcription factor binding sites in target gene expression. *Molecular Systems Biology*, 8.

[Legewie et al., 2006] Legewie, S., Blüthgen, N., and Herzel, H. (2006). Mathematical modeling identifies inhibitors of apoptosis as mediators of positive feedback and bistability. *PLoS computational biology*, 2(9):e120.

[Levskaya et al., 2005] Levskaya, A., Chevalier, A. A., Tabor, J. J., Simpson, Z. B., Lavery, L. A., Levy, M., Davidson, E. A., Scouras, A., Ellington, A. D., Marcotte, E. M., and Voigt, C. A. (2005). Synthetic biology: engineering Escherichia coli to see light. *Nature*, 438(7067):441–2.

[Lichtman et al., 2010] Lichtman, M. A., Kipps, T. J., Seligsohn, U., Kaushansky, K., and Prchal, J. T. (2010). *Williams Hematology*. McGraw-Hill, 8 edition.

[Liew et al., 2006] Liew, C. W., Rand, K. D., Simpson, R. J. Y., Yung, W. W., Mansfield, R. E., Crossley, M., Proetorius-Ibba, M., Nerlov, C., Poulsen, F. M., and Mackay, J. P. (2006). Molecular analysis of the interaction between the hematopoietic master transcription factors GATA-1 and PU.1. *The Journal of biological chemistry*, 281(38):28296–306.

[Lisman, 1985] Lisman, J. E. (1985). A Mechanism for Memory Storage Insensitive to Molecular Turnover: A Bistable Autophosphorylating Kinase. *Proceedings of the National Academy of Sciences*, 82(9):3055–3057.

[Losick and Desplan, 2008] Losick, R. and Desplan, C. (2008). Stochasticity and cell fate. *Science (New York, N.Y.)*, 320(5872):65–8.

[Ma et al., 2009] Ma, W., Trusina, A., El-Samad, H., Lim, W. a., and Tang, C. (2009). Defining network topologies that can achieve biochemical adaptation. *Cell*, 138(4):760–73.

[Maamar et al., 2007] Maamar, H., Raj, A., and Dubnau, D. (2007). Noise in gene expression determines cell fate in Bacillus subtilis. *Science (New York, N.Y.)*, 317(5837):526–9.

[Malavé and Dent, 2006] Malavé, T. M. and Dent, S. Y. R. (2006). Transcriptional repression by Tup1-Ssn6. *Biochemistry and cell biology = Biochimie et biologie cellulaire*, 84(4):437–43.

[Malleshaiah et al., 2010] Malleshaiah, M. K., Shahrezaei, V., Swain, P. S., and Michnick, S. W. (2010). The scaffold protein Ste5 directly controls a switch-like mating decision in yeast. *Nature*, 465(7294):101–5.

[Malumbres and Barbacid, 2001] Malumbres, M. and Barbacid, M. (2001). To cycle or not to cycle: a critical decision in cancer. *Nature reviews. Cancer*, 1(3):222–31.

[Mao et al., 1994] Mao, X., Miesfeldt, S., Yang, H., Leiden, J., and Thompson, C. (1994). The FLI-1 and chimeric EWS-FLI-1 oncoproteins display similar DNA binding specificities. *J. Biol. Chem.*, 269(27):18216–18222.

[Markevich et al., 2004] Markevich, N. I., Hoek, J. B., and Kholodenko, B. N. (2004). Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades. *The Journal of cell biology*, 164(3):353–9.

[Martin and Orkin, 1990] Martin, D. I. and Orkin, S. H. (1990). Transcriptional activation and DNA binding by the erythroid factor GF-1/NF-E1/Eryf 1. *Genes & Development*, 4(11):1886–1898.

[Melen et al., 2005] Melen, G. J., Levy, S., Barkai, N., and Shilo, B.-Z. (2005). Threshold responses to morphogen gradients by zero-order ultrasensitivity. *Molecular systems biology*, 1:2005.0028.

[Metcalf, 1998] Metcalf, D. (1998). Lineage commitment and maturation in hematopoietic cells: the case for extrinsic regulation. *Blood*, 92(2):345–7; discussion 352.

[Metcalf, 2008] Metcalf, D. (2008). Hematopoietic cytokines. *Blood*, 111(2):485–91.

[Michelson, 2007] Michelson, A. D. (2007). *Platelets*. Academic Press.

[Mossadegh-Keller et al., 2013] Mossadegh-Keller, N., Sarrazin, S., Kandalla, P. K., Espinosa, L., Stanley, E. R., Nutt, S. L., Moore, J., and Sieweke, M. H. (2013). M-CSF instructs myeloid lineage fate in single haematopoietic stem cells. *Nature*.

[Mukherji and van Oudenaarden, 2009] Mukherji, S. and van Oudenaarden, A. (2009). Synthetic biology: understanding biological design from synthetic circuits. *Nature reviews. Genetics*, 10(12):859–71.

[Murphy, 2005] Murphy, K. M. (2005). Fate vs Choice: The Immune System Reloaded. *Immunologic Research*, 32(1-3):193–200.

[Murray, 1992] Murray, A. W. (1992). Creative blocks: cell-cycle checkpoints and feedback controls. *Nature*, 359(6396):599–604.

[Ninfa and Mayo, 2004] Ninfa, A. J. and Mayo, A. E. (2004). Hysteresis vs. graded responses: the connections make all the difference. *Science's STKE : signal transduction knowledge environment*, 2004(232):pe20.

[Novick and Weiner, 1957] Novick, A. and Weiner, M. (1957). ENZYME INDUCTION AS AN ALL-OR-NONE PHENOMENON. *Proceedings of the National Academy of Sciences of the United States of America*, 43(7):553–66.

[Nuez et al., 1995] Nuez, B., Michalovich, D., Bygrave, A., Ploemacher, R., and Grosveld, F. (1995). Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene. *Nature*, 375(6529):316–8.

[Okada et al., 2011] Okada, Y., Nobori, H., Shimizu, M., Watanabe, M., Yonekura, M., Nakai, T., Kamikawa, Y., Wakimura, A., Funahashi, N., Naruse, H., Watanabe, A., Yamasaki, D., Fukada, S.-i., Yasui, K., Matsumoto, K., Sato, T., Kitajima, K., Nakano, T., Aird, W. C., and Doi, T. (2011). Multiple ETS family proteins regulate PF4 gene expression by binding to the same ETS binding site. *PloS one*, 6(9):e24837.

[O'Neill et al., 1994] O'Neill, E. M., Rebay, I., Tjian, R., and Rubin, G. M. (1994). The activities of two Ets-related transcription factors required for Drosophila eye development are modulated by the Ras/MAPK pathway. *Cell*, 78(1):137–47.

[Orkin, 2000] Orkin, S. H. (2000). Diversification of haematopoietic stem cells to specific lineages. *Nature reviews. Genetics*, 1(1):57–64.

[O'Shaughnessy et al., 2011] O'Shaughnessy, E. C., Palani, S., Collins, J. J., and Sarkar, C. A. (2011). Tunable Signal Processing in Synthetic MAP Kinase Cascades. *Cell*.

[Ozbudak et al., 2004] Ozbudak, E. M., Thattai, M., Lim, H. N., Shraiman, B. I., and van Oudenaarden, A. (2004). Multistability in the lactose utilization network of Escherichia coli. *Nature*, 427(6976):737–40.

[Palani and Sarkar, 2008] Palani, S. and Sarkar, C. A. (2008). Positive receptor feedback during lineage commitment can generate ultrasensitivity to ligand and confer robustness to a bistable switch. *Biophysical journal*, 95(4):1575–89.

[Palani and Sarkar, 2009] Palani, S. and Sarkar, C. A. (2009). Integrating extrinsic and intrinsic cues into a minimal model of lineage commitment for hematopoietic progenitors. *PLoS computational biology*, 5(9):e1000518.

[Palani and Sarkar, 2011] Palani, S. and Sarkar, C. A. (2011). Synthetic conversion of a graded receptor signal into a tunable, reversible switch. *Molecular systems biology*, 7:480.

[Palani and Sarkar, 2012] Palani, S. and Sarkar, C. A. (2012). Transient noise amplification and gene expression synchronization in a bistable mammalian cell-fate switch. *Cell reports*, 1(3):215–24.

[Pardee, 1974] Pardee, A. B. (1974). A restriction point for control of normal animal cell proliferation. *Proceedings of the National Academy of Sciences of the United States of America*, 71(4):1286–90.

[Park et al., 1999] Park, J. K., Williams, B. P., Alberta, J. A., and Stiles, C. D. (1999). Bipotent Cortical Progenitor Cells Process Conflicting Cues for Neurons and Glia in a Hierarchical Manner. *J. Neurosci.*, 19(23):10383–10389.

[Perkins et al., 1995] Perkins, A. C., Sharpe, A. H., and Orkin, S. H. (1995). Lethal beta-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF. *Nature*, 375(6529):318–22.

[Pevny et al., 1991] Pevny, L., Simon, M. C., Robertson, E., Klein, W. H., Tsai, S. F., D'Agati, V., Orkin, S. H., and Costantini, F. (1991). Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature*, 349(6306):257–60.

[Pico et al., 2008] Pico, A. R., Kelder, T., van Iersel, M. P., Hanspers, K., Conklin, B. R., and Evelo, C. (2008). WikiPathways: pathway editing for the people. *PLoS biology*, 6(7):e184.

[Pleiss, 2006] Pleiss, J. (2006). The promise of synthetic biology. *Applied microbiology and biotechnology*, 73(4):735–9.

[Pomerening, 2008] Pomerening, J. R. (2008). Uncovering mechanisms of bistability in biological systems. *Current opinion in biotechnology*, 19(4):381–8.

[Radonić et al., 2004] Radonić, A., Thulke, S., Mackay, I. M., Landt, O., Siegert, W., and Nitsche, A. (2004). Guideline to reference gene selection for quantitative real-time PCR. *Biochemical and biophysical research communications*, 313(4):856–62.

[Raj et al., 2006] Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y., and Tyagi, S. (2006). Stochastic mRNA synthesis in mammalian cells. *PLoS biology*, 4(10):e309.

[Raj and Tyagi, 2010] Raj, A. and Tyagi, S. (2010). Detection of individual endogenous RNA transcripts in situ using multiple singly labeled probes. *Methods in enzymology*, 472:365–86.

[Raj et al., 2008] Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A., and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. *Nature methods*, 5(10):877–9.

[Raj and van Oudenaarden, 2008] Raj, A. and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*, 135(2):216–26.

[Ramakrishnan and Bhalla, 2008] Ramakrishnan, N. and Bhalla, U. S. (2008). Memory switches in chemical reaction space. *PLoS computational biology*, 4(7):e1000122.

[Rao, 2012] Rao, C. V. (2012). Expanding the synthetic biology toolbox: engineering orthogonal regulators of gene expression. *Current opinion in biotechnology*, 23(5):689–94.

[Razooky and Weinberger, 2011] Razooky, B. S. and Weinberger, L. S. (2011). Mapping the architecture of the HIV-1 Tat circuit: A decision-making circuit that lacks bistability and exploits stochastic noise. *Methods (San Diego, Calif.)*, 53(1):68–77.

[Reynolds et al., 2006] Reynolds, A. P., Richards, G., Iglesia, B., and Rayward-Smith, V. J. (2006). Clustering Rules: A Comparison of Partitioning and Hierarchical Clustering Algorithms. *Journal of Mathematical Modelling and Algorithms*, 5(4):475–504.

[Rieger et al., 2009] Rieger, M. A., Hoppe, P. S., Smejkal, B. M., Eitelhuber, A. C., and Schroeder, T. (2009). Hematopoietic cytokines can instruct lineage choice. *Science (New York, N.Y.)*, 325(5937):217–8.

[Robb, 2007] Robb, L. (2007). Cytokine receptors and hematopoietic differentiation. *Oncogene*, 26(47):6715–23.

[Rosenbauer and Tenen, 2007] Rosenbauer, F. and Tenen, D. G. (2007). Transcription factors in myeloid development: balancing differentiation with transformation. *Nature reviews. Immunology*, 7(2):105–17.

[Santarius et al., 2010] Santarius, T., Shipley, J., Brewer, D., Stratton, M. R., and Cooper, C. S. (2010). A census of amplified and overexpressed human cancer genes. *Nature reviews. Cancer*, 10(1):59–64.

[Santillán et al., 2007] Santillán, M., Mackey, M. C., and Zeron, E. S. (2007). Origin of bistability in the lac Operon. *Biophysical journal*, 92(11):3830–42.

[Schnappinger et al., 1998] Schnappinger, D., Schubert, P., Pfleiderer, K., and Hillen, W. (1998). Determinants of protein-protein recognition by four helix bundles: changing the dimerization specificity of Tet repressor. *The EMBO journal*, 17(2):535–43.

[Serrano, 2007] Serrano, L. (2007). Synthetic biology: promises and challenges. *Molecular systems biology*, 3:158.

[Shah and Sarkar, 2011] Shah, N. A. and Sarkar, C. A. (2011). Robust Network Topologies for Generating Switch-Like Cellular Responses. *PLoS Computational Biology*, 7(6):e1002085.

[Shen-Orr et al., 2002] Shen-Orr, S. S., Milo, R., Mangan, S., and Alon, U. (2002). Network motifs in the transcriptional regulation network of Escherichia coli. *Nature genetics*, 31(1):64–8.

[Shivdasani, 2001] Shivdasani, R. A. (2001). Molecular and transcriptional regulation of megakaryocyte differentiation. *Stem cells (Dayton, Ohio)*, 19(5):397–407.

[Siatecka and Bieker, 2011] Siatecka, M. and Bieker, J. J. (2011). The multifunctional role of EKLF/KLF1 during erythropoiesis. *Blood*, 118(8):2044–54.

[Siatecka et al., 2007] Siatecka, M., Xue, L., and Bieker, J. J. (2007). Sumoylation of EKLF promotes transcriptional repression and is involved in inhibition of megakaryopoiesis. *Molecular and cellular biology*, 27(24):8547–60.

[Singleton et al., 2008] Singleton, B. K., Burton, N. M., Green, C., Brady, R. L., and Anstee, D. J. (2008). Mutations in EKLF/KLF1 form the molecular basis of the rare blood group In(Lu) phenotype. *Blood*, 112(5):2081–8.

[Smith et al., 1996] Smith, L. T., Hohaus, S., Gonzalez, D. A., Dziennis, S. E., and Tenen, D. G. (1996). PU.1 (Spi-1) and C/EBP alpha regulate the granulocyte colony-stimulating factor receptor promoter in myeloid cells. *Blood*, 88(4):1234–47.

[Sneppen et al., 2008] Sneppen, K., Micheelsen, M. A., and Dodd, I. B. (2008). Ultrasensitive gene regulation by positive feedback loops in nucleosome modification. *Molecular systems biology*, 4:182.

[Soneji et al., 2007] Soneji, S., Huang, S., Loose, M., Donaldson, I. J., Patient, R., Göttgens, B., Enver, T., and May, G. (2007). Inference, validation, and dynamic modeling of transcription networks in multipotent hematopoietic cells. *Annals of the New York Academy of Sciences*, 1106:30–40.

[Spooncer et al., 1986] Spooncer, E., Heyworth, C. M., Dunn, A., and Dexter, T. M. (1986). Self-renewal and differentiation of interleukin-3-dependent multipotent stem cells are modulated by stromal cells and serum factors. *Differentiation; research in biological diversity*, 31(2):111–8.

[Stark et al., 2011] Stark, C., Breitkreutz, B.-J., Chatr-Aryamontri, A., Boucher, L., Oughtred, R., Livstone, M. S., Nixon, J., Van Auken, K., Wang, X., Shi, X., Reguly, T., Rust, J. M., Winter, A., Dolinski, K., and Tyers, M. (2011). The

BioGRID Interaction Database: 2011 update. *Nucleic acids research*, 39(Database issue):D698–704.

[Tallack et al., 2012] Tallack, M. R., Magor, G. W., Dartigues, B., Sun, L., Huang, S., Fittock, J. M., Fry, S. V., Glazov, E. A., Bailey, T. L., and Perkins, A. C. (2012). Novel roles for KLF1 in erythropoiesis revealed by mRNA-seq. *Genome research*, 22(12):2385–98.

[Tan et al., 2009] Tan, C., Marguet, P., and You, L. (2009). Emergent bistability by a growth-modulating positive feedback circuit. *Nature chemical biology*, 5(11):842–8.

[To and Maheshri, 2010] To, T.-L. and Maheshri, N. (2010). Noise can induce bimodality in positive transcriptional feedback loops without bistability. *Science (New York, N.Y.)*, 327(5969):1142–5.

[Tsai et al., 1991] Tsai, S. F., Strauss, E., and Orkin, S. H. (1991). Functional analysis and in vivo footprinting implicate the erythroid transcription factor GATA-1 as a positive regulator of its own promoter. *Genes & Development*, 5(6):919–931.

[Voigt et al., 2005] Voigt, C. A., Wolf, D. M., and Arkin, A. P. (2005). The Bacillus subtilis sin operon: an evolvable network motif. *Genetics*, 169(3):1187–202.

[Wang et al., 2002] Wang, X., Crispino, J. D., Letting, D. L., Nakazawa, M., Poncz, M., and Blobel, G. a. (2002). Control of megakaryocyte-specific gene expression by GATA-1 and FOG-1: role of Ets transcription factors. *The EMBO journal*, 21(19):5225–34.

[Welch et al., 2004] Welch, J. J., Watts, J. A., Vakoc, C. R., Yao, Y., Wang, H., Hardison, R. C., Blobel, G. A., Chodosh, L. A., and Weiss, M. J. (2004). Global regulation of erythroid gene expression by transcription factor GATA-1. *Blood*, 104(10):3136–47.

[Wishart et al., 2006] Wishart, J. a., Osborn, M., Gent, M. E., Yen, K., Vujovic, Z., Gitsham, P., Zhang, N., Ross Miller, J., and Oliver, S. G. (2006). The relative merits of the tetO2 and tetO7 promoter systems for the functional analysis of heterologous genes in yeast and a compilation of essential yeast genes with tetO2 promoter substitutions. *Yeast (Chichester, England)*, 23(4):325–31.

[Wolf, 2003] Wolf, D. (2003). Motifs, modules and games in bacteria. *Current Opinion in Microbiology*, 6(2):125–134.

[Wu et al., 1995] Wu, H., Liu, X., Jaenisch, R., and Lodish, H. F. (1995). Generation of committed erythroid BFU-E and CFU-E progenitors does not require erythropoietin or the erythropoietin receptor. *Cell*, 83(1):59–67.

[Xiong and Ferrell, 2003] Xiong, W. and Ferrell, J. E. (2003). A positive-feedback-based bistable 'memory module' that governs a cell fate decision. *Nature*, 426(6965):460–5.

[Yao et al., 2008] Yao, G., Lee, T. J., Mori, S., Nevins, J. R., and You, L. (2008). A bistable Rb-E2F switch underlies the restriction point. *Nature cell biology*, 10(4):476–82.

[Yen et al., 2003] Yen, K., Gitsham, P., Wishart, J., Oliver, S. G., and Zhang, N. (2003). An improved tetO promoter replacement system for regulating the expression of yeast genes. *Yeast (Chichester, England)*, 20(15):1255–62.

[Young et al., 2012] Young, J. W., Locke, J. C. W., Altinok, A., Rosenfeld, N., Bacarian, T., Swain, P. S., Mjolsness, E., and Elowitz, M. B. (2012). Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy. *Nature protocols*, 7(1):80–8.

[Zhang et al., 1994] Zhang, D. E., Hetherington, C. J., Chen, H. M., and Tenen, D. G. (1994). The macrophage transcription factor PU.1 directs tissue-specific expression of the macrophage colony-stimulating factor receptor. *Molecular and cellular biology*, 14(1):373–81.

[Zon et al., 1991] Zon, L. I., Youssoufian, H., Mather, C., Lodish, H. F., and Orkin, S. H. (1991). Activation of the erythropoietin receptor promoter by transcription factor GATA-1. *Proceedings of the National Academy of Sciences of the United States of America*, 88(23):10638–41.