Dynamics and Heterogeneity of Gene Expression and Epigenetic Regulation at the Single-Cell Level

Thesis by

John Yong

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2015

(Defended May 26, 2015)

To Mom, Dad, Vivian, and Olivia.

Acknowledgements

Through reading other's theses and on the journey of writing my own, I have developed a fondness for the Acknowledgements section. Surely, the rest of a thesis past these first pages encompasses the culmination of many years of work. These opening words, however, reflect the personal, human side of scientific research. They put on display the joy, sweat, excitement, camaraderie, and humility that fill the days and nights behind the work presented. They remind us that the work of a thesis is vary rarely the product of a single person, but of collaborations, interactions, formal and fortuitous mentorships, influences and inspirations, mental supports, and actual helping hands, inside and outside the laboratory. They tell us the story behind the story.

My story was shaped on many different levels by my advisor, Michael Elowitz. He attracted people from diverse intellectual, national, and cultural backgrounds into an amalgam of Elowitzians and created this extraordinarily stimulating environment, where no question is unasked and, just as importantly, no question asked is disparaged. Through Michael, I learned the importance of story. I learned that true understanding of a system does not come from merely having a map of all interactions, but from the ability to intuit and distill the essence and principles behind its working. Michael embodies the highest standards of both verbal and visual scientific communication and, despite encountering the occasional bemoaning, he persistently encouraged me to sharpen my skills at every opportunity. He

has forever changed the way I see a seminar or a paper, or prepare my own. For all of these, I am extremely grateful.

The work presented in this thesis is made possible by two major collaborations. Chapter 2 and Chapter 3 are born out of work shared equally between Zak Singer and me, and Lacra Bintu and me, respectively. These two colleagues and friends helped me discover the joy of collaborations, and brought out the best in me as a scientist. Many thanks, also, to other collaborators involved in these projects: Julia Tischler, Alphan Altinok, Yaron Antebi, Kayla McCue, and the laboratories of Long Cai, Azim Surani, and Mitsuo Oshimura. My other thesis committee members – Ellen Rothenberg, Mitch Guttman, Thanos Siapas, and Kathrin Plath – provided helpful feedback and advice, and the encouragement that kept me marching forward, to which I am very grateful. To the guest cheerleader from Barcelona, Jordi Garcia-Ojalvo, who made occasional but always appreciated appearances. I would also like to thank Rochelle Diamond, Diana Perez, Josh Verceles, Janice Grimm, and Keith Beadle at the Caltech Flow Cytometry facility, as well as David Jackson and Mark Eastman from Olympus America Inc., for their technical support on equipment and my experiments.

In addition to the Elowitzians mentioned above, I want to extend my gratitude to the rest of the Elowitz Lab for teaching me so many things and making my journey at Caltech as enjoyable as it could be. In particular: To David Sprinzak, who mentored my rotation projects and taught me how to make and analyze movies. To Michelle Fontes, Leah Santat, James Linton, Jo Leonardo, and Tara Orr, for maintaining a well-oiled lab. To the SURF and rotation students who worked with me – Jennifer Ky, Kayla McCue, and Junyue Cao. To the various occupants of the 'J room' over the years – Joe Levine, Joe Markson, Jin Park, and James Locke – for a great office and snack 'sharing'. To the CrossFit bunch – Joe Levine, Sandy Nandagopal, Fred Tan, Jin Park, Jonathan Young, and Chiraj Dalal

– for inspiring me to push my body to the limit, and showing me by example when one has pushed too far. To the late night crew – Mark Budde, Adam Rosenthal, Yihan Lin, Amit Lakhanpal, Fangyuan Ding, Yutao Qi, and Kirsten Frieda for livening up the hallways around the clock. Also to Hao Yuan Kueh, Lauren LeBon, Pierre Neveu, Emily Capra, and Pulin Li for their friendship and advice at various points over the years. Most importantly, to my buddies, Joe Levine and Sandy Nandagopal, for sharing the full range of grad school moments and emotions, and for enduring my *futile* and *irrelevant* jokes and puns, and quotes from a 20-some-year-old TV series.

And then there are my dearest friends and organizations outside the lab that made LA home for this foreigner. Mikey, Arjuna, Amy, Lorelei, Andreas, Nick, Alex, Kawai, Charlie, Laura, Eric, Harry, Ryan, Johnathan, Kenny, Kyle, Keith, LAFR, Caltech PRISM, Caltech Glee Club, GMCLA, and Westside Comedy. You have touched my life, helped me grow, and changed me – mostly for the better, but definitely for good.

Finally, to my family, for your unfaltering love and support, without which none of this would have been possible.

Abstract

The ability of cells to establish and remember their gene expression states is a cornerstone of multicellular life. This thesis explores how gene expression states are regulated dynamically, and how these regulations differ in individual cells even under the same conditions. These properties underlie cellular state decisions and often determine the balance between different cell types in a multicellular system, but are typically inaccessible to conventional techniques that rely on static snapshots and population averaging. We address these issues in two separate contexts, one natural and one synthetic, using time-lapse imaging and other single-cell techniques.

In the first context, we use embryonic stem (ES) cells, which were shown to exist in a mixed population of at least two cellular states with distinct differentiation propensities, as a model to study natural dynamics of cellular states. These cells display rare, stochastic, and spontaneous transitions between the two states, as well as more frequent fluctuations in gene expression levels within each state. Our system enables us to further investigate how these dynamics are modulated under a cell signaling environment that enhances pluripotency, and the role DNA methylation plays in maintaining these states.

In the second context, we investigate how chromatin regulators (CRs), part of a complex system that enables cells to modulate gene expression and epigenetic memory, operate dynamically in individual cells. We build a synthetic platform to measure the isolated

effect of recruitment and de-recruitment of four individual CRs. In contrast to conventional transcription factor control, all CRs tested regulate gene expression in all-or-none events, controlling the probability of stochastic transitions between fully active and silent states rather than the strength of gene expression. The qualitative and quantitative responses of a cell population are determined by the set of event rates associated with each CR, as well as the duration of CR recruitment. These results provide a framework for understanding and engineering chromatin-based cellular states and their dynamics.

Contents

List of Figures

Chapter 1

Introduction

The development and operation of a multicellular organism is an intricate business. Development begins with a single cell, the fertilized egg, and spontaneously unfolds over many cell divisions and cellular differentiation, eventually giving rise to an adult body. Take humans as an example: An adult human consists of about 30 trillion cells [1], in over hundreds of different cell types [2]. Some of these cell types can be incredibly stable – for instance, a neuron can maintain its identity for decades. On the other hand, some cells can switch cell types when necessary, a property that is especially critical during development and in response to environmental signals. With a few exceptions, most notably in the adaptive immune system, where genetic recombination and mutation are harnessed to generate diversity, all of these cells rely on essentially the same genome [3].

How do cells sharing the same genetic material give rise to different cell types and provide the remarkably diverse range of functions? This is possible because cells can inhabit different cellular states, defined as a collection of expression states across the various genes within a genome (Figure 1.1A). Within a cellular state, specific genes that are required for a particular cell type are set in active expression, while those that are unnecessary are repressed. Just as importantly, cells need to remember who they are by maintaining these gene expression states over time, and be able to switch to different states when needed.

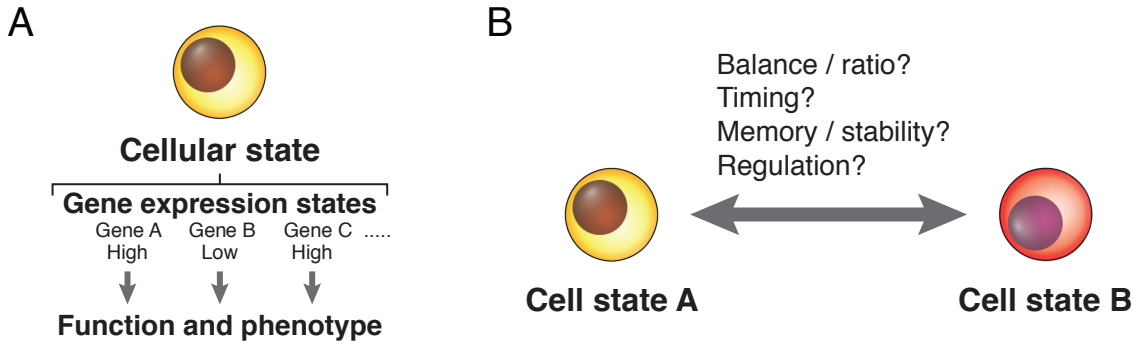


Figure 1.1: Cellular states and their dynamics
(A) A cellular state is defined by the collection of gene expression states, which determine the function and phenotype of the cell. See Section 1.1 for detailed discussions.
(B) Fundamental questions about cellular state dynamics and regulation.

Many genome-wide studies have revealed profiles of gene expression levels of different cell types, as well as specific transcription factors whose expressions are required for these cells [4–6]. Nevertheless, fundamental questions about gene expression states remain to be answered (Figure 1.1B). For instance, how are gene expression states regulated? What is the expression dynamics of a gene within a gene expression state? What happens when a gene switches from one state to another? How fast and frequently do switches occur, and how long do the new states last? Understanding these processes could eventually enable us to control cellular identity in synthetic systems and correct aberrant gene expression states in diseases.

A relevant issue in the study of gene expression states is the inherent heterogeneity in biological systems. This heterogeneity can manifest in the variability in gene expression levels among individual cells as well as the way these cells behave in under the same condition (reviewed in [7], see also [8, 9]). As a result, the study of cellular states may be obscured. Nevertheless, phenotypic variation may be a desirable feature in a system in certain contexts, such as to brace the system against fluctuating environmental conditions (bet-hedging) [7,

10–12]. Using single cell techniques, we can directly characterize the extent of heterogeneity in a system and separate that from the underlying changes in cellular states.

This thesis explores the properties of cellular states and seeks to understand how cells set, maintain, and alter their memories in gene expression. We employ single-cell techniques to investigate these questions in two separate contexts. Embryonic stem (ES) cells exist in mixed populations of at least two cellular states. Chromatin regulators (CRs) are members of a complex system that is capable of establishing gene expression states and memory. These two contexts are used as a natural and a synthetic model system, respectively, for the the study of gene expression states. In the remainder of this chapter, we further our discussions on cellular states and heterogeneity, including a review on the molecular mechanisms behind these phenomena. We also provide the background for ES cells and chromatin modifications, and present the specific questions related to these systems. In Chapter 2, we measure natural dynamics of genes that are critical to the identity of ES cells. These naturally occurring dynamics allow us to distinguish the fluctuations in expression levels inherent to a gene expression state from those that arise from cells switching between different states. In Chapter 3, we build a synthetic system to study how the expression of a gene responds to active recruitment and de-recruitment of different CRs. This system of artificial perturbations using CRs enables us to isolate the different dynamic responses in gene expression and types of memory that each of these regulators can produce. Furthermore, these results form the foundation of a unified quantitative model, which allows us to formalize our understanding of the regulation of cellular states, and engineer new synthetic genetic circuits that take advantage of these regulations. More generally, the results reveal a recurring theme: in both contexts studied, transition of a gene from one state to another occurs through stochastic and mostly digital, switch-like events. In both cases, transition

rates are tunable by either changing the signaling environment (in ES cells), or switching the regulators recruited (in the synthetic recruitment system). Together, we propose that these dynamic properties enable multicellular systems to do two things: to establish stable, well-defined gene expression states, and to tune the fraction of cells in each of these states. This is discussed in detail in Chapter 3.

1.1 Cellular and Gene Expression States

1.1.1 What are cellular states?

One way to define a cell type or cellular state is at the phenotypic level – what morphology a cell adopts and what function it is capable of performing. For instance, a neuron can be excited electrically and communicate with other neurons through synaptic connections; a naïve T lymphocyte can be activated into different types of mature T cells to carry out cell-mediated immunity. Another approach to define a cell type is by the cell's potential to differentiate into other cell types. In mouse blastocysts, for instance, trophoblast cells can give rise to the placenta, while cells in the inner cell mass can give rise to the embryo proper. However, these broad definitions of cell types can be problematic, since many of the functions and differentiation potentials of cells are not static but can fluctuate over time. For example, when nutrients are limited, the soil bacteria *Bacillus subtilis* can transiently become genetically competent, a differentiated state that allows DNA uptake from the environment [13]. Additionally, even cells that are grown under the same condition and have similar morphologies may not share the same differentiation potential. For *B. subtilis*, only about 10-20% of all cells will become competent, even when conditions are optimal and applied across the entire culture [14]. The study of cellular states therefore critically requires

indicators that can track cellular functions and differentiation potentials dynamically, and in individual cells. This role could be served by the expression of key regulatory genes.

Both the functions and differentiation potentials of a cell ultimately originate from the activities of and interactions between different biomolecules in a cell, including DNA, RNA and proteins. The transcription of genes from DNA, partly controlled by the levels of transcription factors in the cell, governs the levels of RNA and proteins, and through their activities, the rate of biochemical reactions inside the cell. Furthermore, the level of some proteins and RNA (e.g. transcription factors and miRNA) can feed back into regulating gene expression and, collectively, these interactions form the genetic circuits. Thus, the expression levels of different genes reflect a cell's function as well as the active interactions within the genetic circuit at a given time. Here, we define cellular state as the collection of gene expression states in a cell. Since gene expression is an intrinsically stochastic process, a gene expression state does not deterministically specify a single expression level of a gene in the cell. Instead, it is associated with a probability distribution of gene expression levels [15, 16]. This inherent variability in gene expression levels within a single cellular state can obscure the identification of cellular states, as we will discuss in detail in Section 1.2.1.

1.1.2 How do cellular states arise?

The implementation of a cellular state requires a coordinated program of gene expression – selective expression of genes required for the cellular function and repression of those that are unnecessary. This program needs to persist for as long as it is necessary for the cellular state, the timescale of which corresponds to the level of gene expression memory. Regulation of gene expression and memory can be achieved by at least two mechanisms at the molecular level – chromatin modifications and transcription feedback. Broadly speaking, chromatin

modifications can control the accessibility of a gene locus to transcriptional activity, while transcription factors can interact with each other to form feedback circuits that lock in expression levels of the genes involved. These modes of action are discussed in detail below. Nevertheless, these two systems are not mutually exclusive, as some transcription factors are known to alter chromatin modifications, and factors that primarily alter chromatin modifications can regulate each other and form a circuit.

Chromatin modifications and regulators

In eukaryotes, DNA is wrapped around histone proteins to form nucleosomes, the basic structural units of chromatin. Without altering the sequence of the DNA, and therefore the identity of the RNA and protein encoded, chromatin can be covalently modified to regulate the expression of genes nearby [17]. The first chromatin modification discovered was cytosine methylation on DNA at CpG dinucleotides, which was associated with repressed gene expression in vertebrates [18–20]. In the two decades that followed, many more chromatin modifications were identified, including cytosine hydroxymethylation on DNA, and methylation, acetylation, phosphorylation, ubiquitination, as well as other modifications on various amino acid residues on the N-terminal of histone proteins [17, 21, 22]. Methylation and hydroxymethylation of DNA in vertebrate genomes mostly occur at CpG sites. In contrast, modifications on histones are more complicated, since modifications can be added to histones at different amino acid residues and for different number of times. For instance, histone 3 (H3) can be methylated at different lysine residues (e.g. H3K4, H3K9, H3K27 etc.), and for one to three times (e.g. H3K4me, H3K4me2, H3K4me3, H3K9me etc.).

Many proteins, known as chromatin regulators (CRs), have been identified to alter chromatin modifications and mediate their effects on gene regulation. These regulators include

factors that can specifically add or remove modifications (*writers* and *erasers*, respectively), as well as factors that specifically bind to the modifications (*readers*). *Readers* can, in turn, recruit other factors, including *writers* and *erasers*. Take DNA methylation, for example; methyl groups are specifically added to CpG dinucleotides by DNA methyltransferases (Dnmts, *writers*), and removed through a partially mapped mechanism that involves conversion of methylcytosine to hydroxymethylcytosine by Tet proteins (*erasers*). *Writers* and *erasers* are often part of larger chromatin-modifying complexes that include other subunits with context-dependent DNA-binding properties. Finally, methylated CpG can be specifically bound by a family of methyl-CpG-binding proteins, including MeCP2 (*readers*).

Many chromatin modifications and regulators have been studied extensively and associated with gene activation and repression, as well as different degrees of gene expression memory [17, 23]. For example, DNA methylation and H3K9me3 at a gene promoter are associated with stable and inheritable gene silencing [24, 25]. In contrast, histone acetylation typically has a high turnover rate and its presence at a gene promoter is associated with genes that are being actively transcribed [26, 27]. A change in gene expression pattern without altering the DNA sequence is one of the definitions of epigenetics, though the use of this term to describe non-heritable changes is a subject of debate [28–30]. Here we adopt the broader definition of epigenetics and will refer to chromatin modifications and epigenetic modifications interchangeably in the remainder of this thesis. The mechanisms by which chromatin modifications regulate gene expression and memory will be explored in greater detail in Section 1.4.

Transcription feedback

Transcription factors are proteins that bind to specific DNA sequences and alter the level of transcription of adjacent genes, either up-regulating (activators) or down-regulating (repressors) them. In eukaryotes, the mechanisms of transcription regulation fall into two conceptual categories [31]. In the first category, transcription factors can control recruitment and activity of the transcription apparatus. Activators can physically bind to the RNA polymerase II-containing transcription initiation apparatus and recruit it to the target gene, or promote transcription elongation by the polymerase. In contrast, repressors can interfere with the binding of the transcription initiation apparatus to DNA, or compete for the activator binding sites. In the second category, transcription factors can recruit CRs that write or erase chromatin modifications. Although transcription factors in this second category may regulate gene expression and memory through chromatin modifications, non-chromatin-modifying transcription factors are themselves sufficient to generate gene expression memory through the formation of transcription factor circuits. We will focus our discussion on this emergent property of transcription factor circuits.

Because a transcription factor binds to DNA by recognizing a specific binding site, it can be recruited to an arbitrary number of genes, including the transcription factor itself, simply by having its binding site inserted near the promoter of these genes. This allows the transcription factor to regulate the expression of its target genes in a concerted manner. These target genes may code for proteins that carry out various biological functions, or other transcription factors. Different transcription factors can therefore connect to form a regulatory network, i.e. a transcription factor circuit.

The dynamic and steady-state behaviors of a transcription factor circuit depend on the wiring between different factors, and have been the subject of active research, combining

experiments and mathematical modeling, since the 2000s (reviewed in [32, 33], also see [34]). Genetic circuits that display oscillation [35], multi-stability [36], reduced variability [37], switch-ability [38], or excitability [13] were synthesized and tested in living cells. Many of these circuits contain motifs where a transcription factor regulates its own expression, either directly (auto-regulatory) or indirectly (through other factors). These motifs are shown to produce well-defined gene expression levels and/or result in gene expression memory [39]. For instance, a single transcription factor that down-regulates its own expression creates a simple negative feedback loop, which produces stable expression of the factor with a reduced range of fluctuation [37]. In contrast, positive feedback loops, in their simplest forms as one-node (auto-regulatory) [36] or two-node (two factors that mutually repress each other) [38] systems, can generate bi-stability. Whereas a one-node system can either be fully OFF, or locked into an ON state, a two-node system can be locked into expressing either one, but not both, of the two factors. Effectively, a system with a positive feedback loop can generate more than one gene expression state. Since transcription factor molecules are partitioned more or less evenly into daughter cells over cell division events, gene expression regulation can be maintained over these events. As a result, the expression state in a cell can persist through cell generations. Transcription feedback can therefore be used, and is used, in both natural and synthetic systems to establish and maintain gene expression states.

1.2 Heterogeneity in Multicellular Systems

The study of cellular and gene expression states cannot be complete without considering heterogeneity in biological systems. Individual cells within an isogenic population and under the same growth conditions can exhibit variation in the level of cellular molecules and their phenotypes [8, 9, 16, 40]. As a result, cells that inhabit the same state can exhibit

different expression levels of one gene (Figure 1.2). On the contrary, distinct states that are close to each other in their mean expression levels can overlap in their expression level distributions at the population level, making it difficult to discern the state of a cell with just one measurement of the expression of one gene. In this section, we will explore the origins of heterogeneity in multicellular systems, and the experimental techniques used to study this phenomenon.

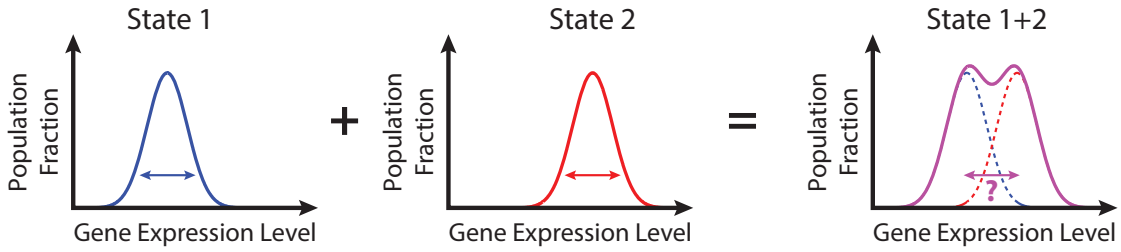


Figure 1.2: Cell-cell variability in gene expression levels Cells within a single state can exhibit variation in gene expression levels due to gene expression noise (left hand side, blue and red arrows indicate variance in each distribution). Mixture of cells in different states can convolute the estimation of the number of states within the system, and the assigning of state to cells found within the overlapping range of expression levels (right hand side, magenta arrow).

1.2.1 How does heterogeneity in gene expression levels arise?

In theory, variability in gene expression can arise through at least two mechanisms – the co-existence of multiple cellular states and the inherent stochasticity in the gene expression process. On one hand, a cell population can exist as a mixture of cells that are in different states. As outlined in Section 1.1.2, these states can be established either by chromatin modifications or through multi-stability generated by feedbacks in the transcription factor circuit. Both of these processes can lead to memory in gene expression levels associated with an individual cell. As a result, individual cells in a cell population can express a gene at different levels, depending on what states they are in.

On the other hand, the gene expression process itself has long been proposed to contribute to significant variability in gene expression levels. This phenomenon, termed gene expression noise, arises due to the inherent stochastic nature of the gene expression process, which involves random interactions of small numbers of molecules and stochastic switching between different chromatin configurations of the promoter. These models have been reviewed extensively [7, 16], and are supported by an abundance of recent experimental studies [9, 40–42]. As a result of gene expression noise, even cells that inhabit a single state can display a range of expression levels of a gene.

As discussed above, variability in gene expression can confound the investigation of cellular states. Nevertheless, gene expression noise can also be important to cell fate decision. In some cases, cells have been shown to take advantage of this noise in gene expression and augment it through positive feedback to generate phenotypic variation to the benefit of the multicellular system as a whole [7, 13, 43, 44]. The quest for quantifying heterogeneity is therefore twofold – to distinguish the underlying cellular states from noisy gene expression, and to characterize the extent of noise in different genes, which may help us understand how noise contributes to cellular behaviors [45–47].

1.2.2 Techniques to study cellular states and heterogeneity

Until recently, studies of cellular states, using genome-wide techniques such as microarray and sequencing, have relied on static snapshots to provide average measurements over large numbers of cells, typically due to the minimal initial material required by these techniques [4–6, 48]. These assays provided a vast reserve of knowledge on the average levels of gene expression associated with different tissues and cell types, but were unable to access the variability in gene expression among individual cells. Except in exceptional cases where

gene expression heterogeneity results in observable phenotypic differences, such as in fur color of calico cats due to random X chromosome inactivation [49, 50] and eye color of mutant *Drosophila* due to position-effect variegation [51, 52], this heterogeneity among otherwise identical cells can only be visualized using *in situ* hybridization or other imaging techniques [53, 54]. Alternatively, cell-cell variability may be detected using non-imaging based methods at the single-cell level when the limitation of initial sample size is overcome [55, 56]. These techniques, however, remain limited to static measurements due to their reliance on fixed or lysed cells.

With the advent of fluorescent proteins [57], it has recently become possible to measure and follow the expression levels of specific genes in living cells. Combining fluorescent reporter transgenes with techniques that enable signal quantification of individual cells, such as flow cytometry and fluorescence time-lapse microscopy, now gives us the tools to separate the dynamic behaviors of individual cells from the averaged properties of a cell population. These tools also allow us to investigate the dynamic nature of transcription circuits [13, 35, 58, 59].

Together, quantitative analysis of single-cell dynamics can help us identify cellular states and reveal fundamental properties of the processes that establish and regulate these states. This thesis explores the properties of cellular states and their regulatory dynamics at the single-cell level. We use ES cells as a model system to study naturally occurring multi-state dynamics, and a inducible CR recruitment platform as a synthetic model system to study the dynamics of cellular state regulation. In the next two sections, we will provide further background for these systems.

1.3 ES Cells, Cellular States and Dynamic Heterogeneity

1.3.1 What are ES cells?

After fertilization, a mammalian egg undergoes several rounds of cell cleavage to give rise to a spherical mass of cells known as the morula, followed by compaction of these cells and further cell divisions to generate a blastocyst. The early blastocyst (∼3.5 days post fertilization, or dpf, in mice; ∼5 dpf in humans) consists of a cavity and chiefly two distinct lineages of cells (Figure 1.3): the trophectoderm (TE), which forms the outer layer of the blastocyst, and the inner cell mass (ICM), which resides in one side of the cavity. It is from the ICM that ES cells are derived, by explanting and culturing in a supporting medium [60–62]. The inner cell mass (ICM) and ES cells are both pluripotent, in that they can contribute to all cell types within the body when they are re-introduced into the embryo. Additionally, ES cells can be differentiated into various cell types *in vitro*. Furthermore, ES cells can be propagated extensively in culture while largely maintaining pluripotency over many generations, a characteristic called self-renewal. These properties make ES cells a very attractive candidate for therapeutic research, as they may be manipulated for a potentially unlimited supply of different cell types that can be used for regenerative treatment. Furthermore, they also serve as an *in vitro* model for the study of differentiation.

25 years after ES cells were first derived from mouse blastocysts, a group led by Shinya Yamanaka described a new method for generating pluripotent stem cells, which catapulted the field of stem cell research and therapeutics to a whole different paradigm [63]. In this method, adult fibroblast cells can be reprogrammed to gain pluripotency by forced expression of only four transcription factors – Oct4, Klf4, Sox2, and c-Myc. These reprogrammed cells, coined induced pluripotent stem (iPS) cells by the scientists, are similar to ES cells in

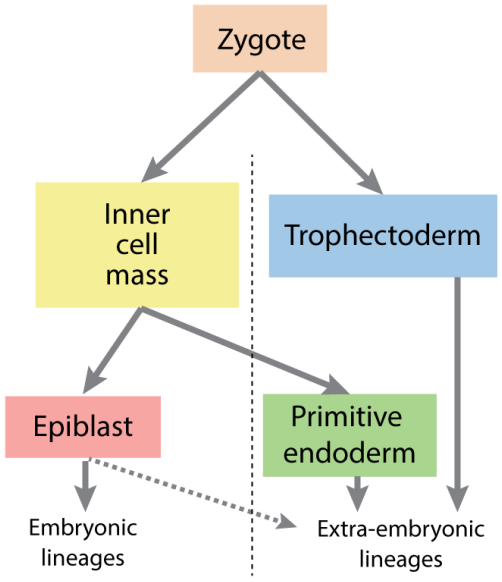


Figure 1.3: Lineage specification in early embryonic development

their ability to contribute to all somatic cell types. Unlike ES cells, however, iPS cells can be derived directly from a patient and therefore share the same genotype. This property provides iPS cells with the distinct advantage of obviating immune rejection when these cells are treated and returned into the patient. The invention of iPS cells represents a big stride towards personalized medicine, and earned Yamanaka a Nobel Prize in 2012 – a prize he shared with John Gurdon for the latter's research on somatic cell nuclear transfer in 1962 [64] – with simultaneously one of the shortest *and* longest delays between discovery and award in the history of the prize.

1.3.2 Heterogeneity in stem cells

In the late blastocyst stage (∼4.5 dpf in mice, ∼6-7 dpf in humans), the ICM from early blastocyst further segregates into two immediate lineages with distinct differentiation propensities – epiblast, from which the three germ layers are derived, and primitive endoderm, which lines the boundary between epiblast and the blastocyst cavity, and from which the

yolk sac is derived. Similarly, ES cells are able to contribute to both epiblast and, at a lower frequency, primitive endoderm when introduced into an early blastocyst to form a chimera [65]. The transcription factors Nanog and Gata6 were shown to be necessary for the establishment of epiblast and primitive endoderm, respectively [66, 67].

Surprisingly, it was later discovered that Nanog and Gata6 are expressed in opposing salt-and-pepper patterns in the ICM of early blastocysts [68], indicating at least two cellular states within the population. A similar pattern of Nanog expression was later found in ES cells (Gata6 expression was also detected, albeit at a lower level) cultured under the standard pluripotency supporting condition [69, 70]. These results suggest that the ICM and ES cells inhabit at least two cellular states. Nevertheless, the gene expression patterns in these cells are not deterministic, as application of inhibitors along the FGF/MEK/ERK signaling pathway can force all cells into the higher Nanog expressing state in both contexts [71, 72]. Finally, Oct4, a pluripotency-associated transcription factor that is expressed in epiblast, but not primitive endoderm, is expressed uniformly in both the ICM and ES cells, indicating that these cells are not merely mixtures of differentiated epiblast and primitive endoderm cells. Together, these findings shattered earlier assumptions that pluripotent cell types exist as homogeneous populations, and heralded the idea that spontaneously generated gene expression heterogeneity may serve to prime cells into distinct lineages once differentiation commences.

Similar discoveries of gene expression heterogeneity have since been made in hematopoietic progenitor cells [47] and human breast cancer cell lines [73], as well as with other transcription factors in ES cells [74, 75]. These observations indicate that heterogeneity may be a common phenomenon among multipotent systems and of clinical significance. In many of these examples, the heterogeneity is in a dynamic equilibrium, such that individ-

ual cells switch between different cellular states. This was indicated by experiments sorting cells into subpopulations using the level of surface markers or fluorescence reporter, and following these subpopulations at subsequent time points. In ES cells, for instance, sorted Nanog-high and Nanog-low subpopulations can slowly give rise to cells of the opposite state on the timescale of days [69]. In a separate context, human breast cancer cells can be sorted into stem-like, basal, or luminal states, with the stem-like cancer cells having higher tumor-seeding ability and drug resistance [73]. Critically, each of these cell types is capable of regenerating all three cell types, suggesting that a treatment regime that only targets stem-like cancer cells is unlikely to completely ablate tumor malignance, an implication that is of obvious clinical relevance.

1.3.3 Transcription factor circuit in ES cells

What is the origin of the heterogeneous yet dynamic system in ES cells? A candidate mechanism, as we have introduced in the previous section, is transcription feedback. An excitatory [13] or oscillatory [35] transcription factor circuit, for example, can generate gene expression patterns that are both heterogeneous and dynamic. In order to obtain a picture of the effective transcription factor circuit in ES cells – a minimal circuit that can describe the essence of the observed dynamics and degree of heterogeneity – one needs to first identify the nodes (transcription factors) involved in the circuit and the wiring between them.

Transcription factors connected in a circuit are likely to have their expression levels rise and fall together (or in opposite directions, if the connection is inhibitory). Using biochemical techniques, such as immunostaining, *in situ* hybridization, and cell sorting, followed by reverse transcription polymerase chain reaction (RT-PCR) or western blot, several more transcription factors were shown to have expression levels that correlate with that of Nanog

[69, 70, 74, 75]. These include Stella, Rex1, Sox2, Klf4, Tbx3, and Esrrb. Two groups of researchers then performed genome-wide assays to elucidate the regulatory connections among these transcription factors and other core pluripotency associated transcription factors, such as Nanog, Oct4 and c-Myc. These experiments either map the global binding patterns of these transcription factors [76], or the changes in global gene expression patterns in response to perturbations of the expression levels of these factors [77]. The results from these studies typically suggested that the transcription factor circuit regulating pluripotency is highly interconnected, with different transcription factors often up-regulating themselves and many other factors. At the core of these circuits are the interactions between Nanog, Oct4 and Sox2, where Oct4 and Sox2, either independently or as a dimer, activate each other as well as Nanog. Nanog, in turn, activates both Oct4 and Sox2. Finally, all three factors activate themselves. This core circuitry forms a strong positive feedback system, which is believed to generate multi-stability in gene expression, as we discussed in Section 1.1.2.

To see if this core circuitry can give rise to the dynamic and heterogeneous gene expression as indicated in Nanog by the sorting and flow cytometry experiments [69], Glauche et al. [59] and Kalmar et al. [58] independently created three mathematical models that included different versions of the circuitry. Stochastic simulations of these models demonstrated that the simple circuits, each with three or fewer components, are all capable of qualitatively replicating the experimental distribution of Nanog expression levels. Moreover, consistent with the dynamics implied by sorting experiments, individual cells can transition between different states in all three models. Nevertheless, even though these models only have minor differences from one another in their configurations, they give rise to qualitatively different dynamic behaviors. These behaviors range from a bistable system with stochastic transi-

tions, to an oscillatory system [59], to a noise-driven excitable system [58]. At the time of these models, there were no reported direct measurements of the dynamics of Nanog expression in individual cells to impose restrictions on these models. It is also important to note that all of these models were drawn from a subset of putative transcription interactions that are each supported by some experimental evidence. The divergent dynamic behaviors of these slightly differing models thus highlight the critical need for these theoretical approaches to the understanding of a complex system to be supplemented by direct dynamic measurements of the system.

1.4 Chromatin Modifications, Chromatin Regulators and Gene Expression States

1.4.1 What are chromatin modifications and chromatin regulators?

As introduced in Section 1.4, many chromatin modifications and regulators are associated with the regulation of gene expression states and epigenetic memory. In this section, we will focus on how these associations are uncovered, and the mechanisms through which the regulation is mediated.

Many chromatin modifications and regulators have been studied using biochemical and genome-wide techniques. To find out whether a chromatin mark is associated with gene activation or repression, chromatin immunoprecipitation (ChIP) is usually performed, followed by quantitative polymerase chain reaction (qPCR) or sequencing, which generate maps of the levels of a mark over many gene loci. These chromatin modification maps can be analyzed in conjunction with data from genome-wide assays of transcriptional activities, such as RNA sequencing (RNA-seq) and genomic run-on (GRO) or RNA polymerase II

immunoprecipitation (PolII RIP), followed by sequencing. These studies provided useful insights on the correlation between different modifications and the level of gene activation or repression they may provide.

1.4.2 How do chromatin modifications regulate gene expression?

There are two types of characterized mechanisms that can mediate gene regulation downstream of chromatin modifications. The first type involves disrupting the contacts between different histones in neighboring nucleosomes, or that between histones and DNA [17]. Among the various types of histone modification, acetylation is at the center of this hypothesis due to its ability to modulate the electric charge of histone tails. The addition of acetyl groups neutralizes the positive charge of lysine residues, and can disrupt binding between these lysine residues and the negatively charged phosphate backbone on DNA [78]. As for interactions between neighboring nucleosomes, the tail of nonacetylated histone H4 on one nucleosome is proposed to interact with a cluster of acidic amino acids on histones H2A/H2B on an adjacent nucleosome [79]. This interaction could cause a multi-nucleosomal array to condense into a compact 30nm-fiber, a secondary chromatin structure. *In vitro* experiments demonstrated that constitutive acetylation of histone H4 at residue K16 impedes the formation of this fiber [80]. Together, these results support a molecular mechanism in which histone acetylation promotes a more open chromatin structure, which in turn leads to higher accessibility to transcription machineries and increased gene expression. Removal of acetylation by histone deacetylases (HDACs) can therefore promote chromatin compaction and lead to gene silencing [81, 82] .

The second type of mechanism for chromatin regulation involves the differential binding of transcription factors and *readers* to a gene locus depending on the presence (or absence)

of different modifications [17]. These transcription factors and *readers* may provide transcriptional or regulatory activities to the gene locus, or they may recruit other factors that carry such activities. DNA methylation, for example, exhibits this type of regulation on various levels [22].

First, methylation can directly interfere with the binding of transcription factors. *In vitro* experiments showed that methylation at a viral promoter, AdMLP, blocks the binding of a mammalian transcription factor, MLTF, and inhibits transcription from the promoter [83]. The occlusion to binding depends on methylation at a specific CpG position within the transcription factor binding site at the promoter, and unmethylated promoter is actively transcribed.

Second, methylation can interfere with the binding of "insulator" proteins, which regulate the interactions between a promoter and its enhancers. This is shown in the imprinting of mouse Igf2 gene, where methylation of an intergenic region is associated with active expression of a nearby gene, as oppose to promoter methylation in the other examples. Here, binding of "insulator" proteins, CTCFs, at an unmethylated control region on the maternal allele prevents the action of a distal enhancer and blocks Igf2 expression [84]. On the paternal allele, however, the CpG sites within the control region are highly-methylated, thereby abolishing CTCF binding and enabling Igf2 to express.

Finally, methylation can also specifically recruit *readers* to effect further modification of the locus. This is illustrated by a family of methyl-CpG-binding proteins, including MeCP2, the most characterized member of this family. Mutations in this gene in human lead to a devastating neurological disorder called Rett syndrome [85]. MeCP2 contains two domains: a methyl-CpG-binding domain (MBD), which enables it to be recruited specifically to methylated chromosomal regions, and a transcriptional-repressor domain (TRD), which is

shown to associate with corepressor complex mSin3a and HDACs [86]. Binding of MeCP2 at methylated chromosomal loci therefore represses gene expression partly by removing histone acetylation and promoting chromatin compaction as described above. Altogether, these studies highlight the complex nature of chromatin regulation, with crosstalks between pathways of different modifications, and the role of these modifications in gene regulation and pathology.

1.4.3 Inheritance and propagation of chromatin modifications

The discovery of base-pairing in double stranded DNA pointed to an elegant solution to the faithful inheritance of genetic code through replication. Similarly, the discovery of Dnmt1 foretold a simple model of maintenance for CpG methylation [87]. Dnmt1 can bind hemi-methylated CpG, sites that are methylated on one strand only, and methylate the opposite strand. Unmethylated DNA can acquire CpG methylation on one strand through the action of *de novo* methyltransferases, Dnmt3a or Dnmt3b. The resulting hemi-methylated DNA is then recognized by Dnmt1, which leads to methylation on both sides of the CpG palindrome. During DNA replication, each progeny DNA will have one inherited, methylated strand and one newly-synthesized, unmethylated strand. Continuous presence of Dnmt1 ensures that all hemi-methylated progeny DNA will be fully methylated, and that the methylation pattern persists over cell generations.

Whether and how histone modifications are inherited over DNA replication is less well-established [88]. One proposed mechanism for maintaining the domain of histone marks stems from an assumption that histone octamers are randomly (but evenly, on average) distributed onto the two progeny DNA double-helixes, seeding each progeny with half the dosage of each histone modification [89]. Furthermore, a few histone marks are known

to have *readers* that can recruit *writers*/*erasers* for the same mark. These include Eed (*reader*) and Ezh2 (*writer*) for H3K27me3 [90], HP1 (*reader*) and SUV39H (*writer*) for H3K9me3 [88, 91–93], and Sir3 (*reader*) and Sir2 (*eraser*) for deacetylated H4K16 [94–96]. These *reader-writer* and *reader-eraser* pairs are proposed to recognize the inherited histones seeded with a particular modification and replicate the same modification to neighboring, newly incorporated histones, thereby conserving the original pattern of modifications.

This proposed mechanism, however, is challenged by recent findings that methylated histones (H3K27me3 and H3K4me3) are replaced by non-methylated histones during DNA replication [97]. Instead of existing histone modification seeding the progeny DNA molecules, it is indicated that complexes writing these marks (Polycomb and Trithorax groups, for H3K27me3 and H3K4me3, respectively) are continuously associated with the DNA during the replication process. These results suggest a contrasting model for histone mark inheritance, where *writer* complexes remain bound to specific gene loci and re-establish the modification patterns after replication. It will be interesting to see if this second mechanism represents a general strategy for all histone modifications and cell types, or if the two proposed mechanisms co-exist in a context-dependent manner. Regardless of the mechanism, however, it is well-established that some histone modifications (e.g. H3K9me3) can persist across cell divisions even in the absence of the signal that originally triggered the addition of the mark [98]. These results provide strong evidence that chromatin marks, together with the background machineries of CRs, are sufficient to impart gene expression regulation and memory.

Besides inheritance of chromatin modifications over DNA replication, the association of *readers* with *writers* or *erasers* for the same marks can also cause a mark to spread along a gene locus from an initial modification at a single site [89]. Using computer sim-

ulation of a simple stochastic model, Hathaway et al. [98] showed that modification at a single nucleation site can lead to a steady state domain through spreading. This domain has a finite size as a result of competition between propagation of marks to neighboring nucleosomes and turnover of marked nucleosomes with unmarked ones. The profile of this simulated modification domain resembles the ones obtained from chromatin immunoprecipitation (ChIP) assays at different sites around a single gene [98], or averaged genome-wide across different loci [99]. These results highlighted the dynamic nature of chromatin regulation, and supported a stochastic model for the underlying molecular process, suggesting that the response at the level of individual cells may also be stochastic.

1.5 Dynamics of Cellular State Regulation: Outstanding Questions

Despite much work on elucidating the molecular players and pathways involved in cellular states, fundamental questions on how these states are established, maintained over time, and altered remain unanswered. More specifically, we ask:

- **What are the dynamics of gene expression in individual cells within a cellular state and in cells transitioning between states?** It is now evident that many multicellular systems exist as a dynamic mixture of different cell types, even under constant conditions. Given that heterogeneity in gene expression levels exists even for a promoter at a constant state, how does one distinguish between heterogeneity that originates from the co-existence of multiple cellular states and from transcriptional noise? Once the dynamic behavior of cells within a cellular state is resolved, we can ask how fast and how frequently cells switch between these different

cell types. These questions can only be answered using single-cell techniques. The resulting experimental platform, in turn, enables further interrogation of the system under different perturbations.

- **How do CRs control gene expression in single cells?** From the perspective of establishing a new gene expression state, either under a natural or synthetic context, CRs represent a critical point of control connecting the signal that triggers CR recruitment to the alteration of chromatin and its downstream gene regulation and memory. What is the causal relationship between recruitment of different CRs and the dynamic response of the target gene? How strongly and how rapidly can each CR alter gene expression? What type of gene expression memory can they generate? Furthermore, given the evidence of stochastic response in some context of chromatin regulation, such as in X-inactivation and position-effect variegation, we would like to know how the response to CR recruitment varies between individual cells.

- **How can a multicellular system control the balance between different cellular states?** The dynamics of switching between cellular states can determine the fraction of cells in different cell types within a multicellular system. How frequently do cells transition between different states? How do the transition rates depend on parameters such as inter-cellular signaling, and the strength and duration of CR recruitment? If we understand how these dynamics are controlled in both natural and synthetic systems, we may gain quantitative insights into how a multicellular system obtains optimal balance between different cellular states, and may eventually be able to control cellular identity in synthetic systems and correct aberrant gene expression programs in diseases.

1.6 Contributions of This Thesis

The chapters in this thesis investigate the dynamics of cellular state decisions at two different levels and make several contributions to our understanding of this topic:

- **Natural dynamics in ES cells.** We created a platform for time-lapse imaging of ES cells and reported novel observations of single-cell gene expression dynamics in these cells. Combining our data from movies and single molecule mRNA fluorescence *in situ* hybridization (smFISH) revealed that these dynamics are the result of two processes: rare, abrupt transitions between two main cellular states, and frequent transcriptional bursts within each of these states. These results demonstrated unequivocally the existence of two broadly correlating cellular states in ES culture, even in wild-type cells, and transitions between them. Moreover, our platform enabled us to further investigate how signaling pathway perturbation alters pluripotency by modulating both the timescale of transcriptional bursts and the rates of transitions between the states, tipping the balance of the system in favor of a more pluripotent state. Finally, we presented evidence that DNA methylation is critically involved in maintaining bistability in the system.

- **Dynamics of chromatin regulation in a synthetic system.** We generated a targeted recruitment platform to measure the dynamic effects of different CRs on gene expression and memory. We directly observed, in individual cells, silencing events triggered by CR recruitment and subsequent reactivation events after de-recruitment. Remarkably, these events appear to be all-or-none – individual cells switch stochastically between fully active and fully silent states at rates that depend on which CR is recruited. The identity of the CR also dictates the types and timescales of memory of

the silent states. As a result, instead of controlling the level of expression in individual cells, CR recruitment controls the fraction of active or silent cells, and this control is dependent on the duration of the recruitment. These observations enabled us to build a quantitative dynamic framework, which could potentially be applied to the analysis of other CRs and the engineering of synthetic systems that involve chromatin regulation.

Chapter 2

Dynamic Heterogeneity and DNA Methylation in Embryonic Stem Cells

(This chapter was adapted, in part, from [100])

2.1 Abstract

Cell populations can be strikingly heterogeneous, composed of multiple cellular states, each exhibiting stochastic noise in its gene expression. A major challenge is to disentangle these two types of variability, and to understand the dynamic processes and mechanisms that control them. Embryonic stem cells (ESCs) provide an ideal model system to address this issue because they exhibit heterogeneous and dynamic expression of functionally important regulatory factors. We analyzed gene expression in individual ESCs using single-molecule RNA-FISH and quantitative time-lapse movies. These data discriminated stochastic switching between two coherent (correlated) gene expression states and burst-like transcriptional noise. We further showed that the "2i" signaling pathway inhibitors modulate both types of variation. Finally, we found that DNA methylation plays a key role in maintaining these metastable states. Together, these results show how ESC gene expression states and dy-

namics arise from a combination of intrinsic noise, coherent cellular states and epigenetic regulation.

2.2 Introduction

Many cell populations appear to consist of mixtures of cells in distinct cellular states. In fact, interconversion between states has been shown to underlie processes ranging from adult stem cell niche control [101, 102], to bacterial fitness [13], to cancer development [73]. A central challenge is to identify transcriptional states, along with the mechanisms that control their stability and generate transitions among them.

Single-cell transcriptional studies have revealed substantial gene expression heterogeneity in stem cells [47, 56, 69, 71, 103]. Moreover, subpopulations expressing different levels of Nanog, Rex1, Dppa3, or Prdm14, show functional biases in their differentiation propensities [70, 74, 75, 104]. This heterogeneity could in principle arise from stochastic fluctuations, or 'noise', in gene expression [7, 105, 106]. Alternatively, it could reflect the coexistence of multiple cellular states, each with a distinct gene expression pattern showing correlation between a set of genes [56, 73, 107, 108]. Disentangling these two sources of variation is important for interpreting the transcriptional states of individual cells and understanding stem cell dynamics.

A related challenge is to understand the mechanisms that stabilize cellular states despite noise. DNA methylation has been shown to be heritable over many generations, is critical for normal development [109], and may help stabilize irreversible cell fate transitions [110–113]. However, the role of DNA methylation in the reversible cell state transitions that underlie equilibrium population heterogeneity has been much less studied [114, 115]. Recently, it was reported that exposing ES cells to inhibitors of MEK and GSK3 (called 2i) abolishes

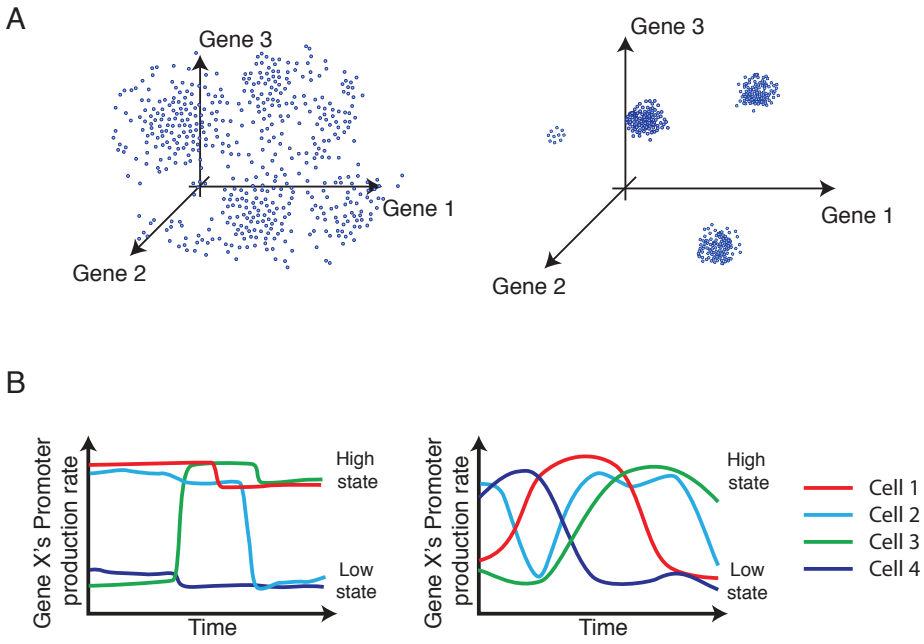


Figure 2.1: Different types of gene expression heterogeneity
(A) Intrinsic noise in gene expression can lead to uncorrelated variation (left), while the coexistence of distinct cellular states can produce correlated variability in gene expression (right). Both panels depict schematic static 'snapshots' of gene expression.
(B) Dynamically, gene expression levels could vary infrequently and abruptly (left) or more frequently and gradually (right) both within and between cellular states (schematic).

heterogeneity and induces a 'naïve' pluripotent state [116, 117] with reduced methylation [118–120]. However, a causal role linking methylation, heterogeneity, and 2i remains to be elucidated.

Together, these observations provoke several fundamental questions: First, how do noise and states together determine the distribution of expression levels of individual regulatory genes (Figure 2.1A)? Second, how do gene expression levels vary dynamically in individual cells, both within a state and during transitions between states (Figure 2.1B)? Finally, how do cells stabilize metastable gene expression states, and what role does DNA methylation play in this process?

Using single-molecule RNA-FISH (smFISH), we analyzed the structure of heterogeneity in the expression of key cell fate regulators, finding that distinct cell states account for

most variation in some genes, while others are dominated by stochastic bursts. Using time-lapse movies of individual cells, we observed abrupt, step-like dynamics due to cell state transitions and transcriptional bursts. Finally, using perturbations, we observed that DNA methylation modulates the population fraction of cells in the two states, consistent with reciprocal expression of the methyltransferase Dnmt3b and the hydroxymethylase Tet1. Together, these results suggest how noise, dynamics, and epigenetic regulatory mechanisms contribute to the overall distribution of gene expression states in stem cell populations.

2.3 Results

Mouse ESCs show three distinct types of gene expression distributions

The process of mRNA transcription is inherently stochastic. As a result, even a clonal cell population in a single state is expected to display variability in the copy number of each mRNA [9, 15, 40, 41, 121–125], potentially leading to phenotypic differences between otherwise identical cells [13, 44, 126–128].

In order to accurately measure mRNA copy numbers in large numbers of individual ESCs, we developed an automated platform for smFISH (Supp. Info.). This system enables rapid analysis of four genes per cell across ∼400 cells per sample (Figure A.1A-D). We validated the system by comparing three measures of expression of the same gene in the same cells using a Rex1-dGFP reporter line [129] (Figure A.1E).

Using this platform, we analyzed 36 pluripotency associated regulators that play critical roles in ESCs or are heterogeneously expressed, as well as several markers of early cell fates and housekeeping genes. The resulting mRNA distributions exhibited a range of distribution shapes and degrees of heterogeneity (Figure 2.2A). We analyzed these distributions

within the framework of bursty transcriptional dynamics. In this model, mRNA production occurs in stochastic bursts that are brief compared to the mean inter-burst interval, and are exponentially distributed in size. Bursty dynamics produce negative binomial (NB) mRNA distributions [15, 130], whose shape is determined by the frequency and mean size of bursts.

Genes exhibited three qualitatively distinct types of mRNA distributions. First, most genes were unimodal and well-fit by a single NB distribution (Figures 2.2B, A.2A, maximum-likelihood estimation (MLE), χ^2 Goodness of Fit (GOF) test $p>0.05$). This class included Oct4, Rest, Tcf3, Smarcc1, Sall4, and Zfp281. Coefficients of variation (CV) were typically $\sim$0.5 for the most homogeneous genes (Figure 2.2A).

Second, a subset of unimodal genes exhibited long-tailed distributions, in which most cells had few, if any, transcripts, while a small number of cells displayed many transcripts. These distributions were also well fit by a single NB distribution, but with resulting distributions that generally decreased monotonically with increasing mRNA concentration (Figures 2.2B, A.2A, χ^2 GOF. $p>0.05$). The most heterogeneous long-tailed genes had burst frequencies of less than one burst per mRNA half-life. These included Tbx3 (CV=2.130.23, means.e.m.), Dppa3 (CV=1.760.31), and Prdm14 (CV=1.5990.20). Other long-tailed genes such as Pecam1, Klf4, Blimp1, Socs3, Nr0b1, and Fgfr2 had higher burst frequencies and less skew. Long-tailed genes arising from rare bursts could provide a source of stochastic variation that could propagate to downstream genes.

Third, there were some genes whose mRNA distributions were significantly better fit by a linear combination of two NB distributions than by one (Supp. Info., Akaike's Information Criteria (AIC) and log-likelihood ratio test, $p<0.05$). These genes included Rex1, Nanog, Esrrb, Tet1, Fgf4, Sox2, Tcl1, and Lifr (Figures 2.2B, A.2A). In some cases, the two components of these distributions were well separated from one another (e.g. Rex1

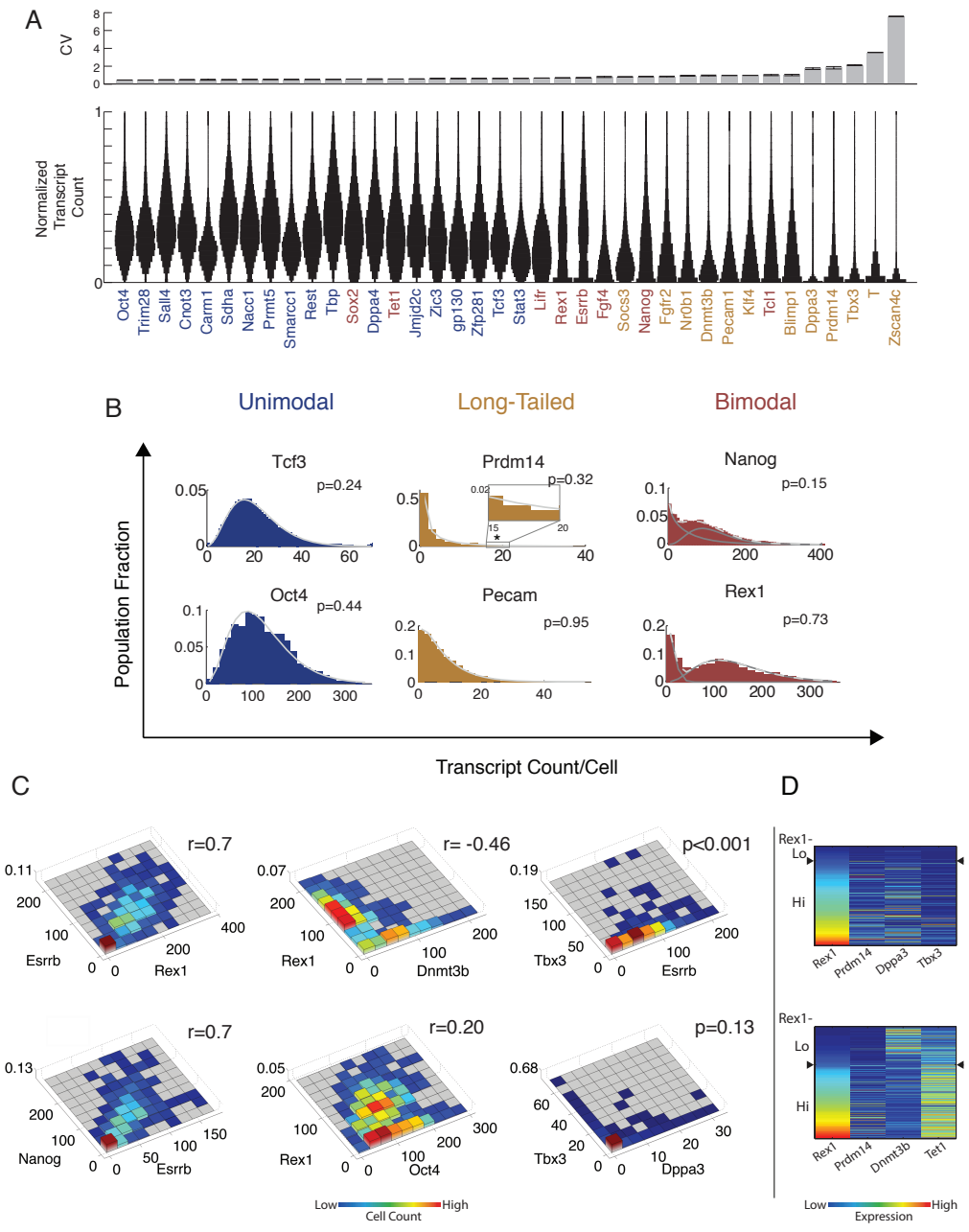


Figure 2.2: smFISH reveals gene expression heterogeneity and correlation
(A) Top: coefficients of variation (CV, meanSEM) for ESC-associated regulators and housekeeping genes. Bottom: Distributions (violin plots) normalized by maximum expression level reveal qualitatively distinct gene expression distributions. Genes are sorted by increasing CV.
(B) Smoothed histograms for mRNA distributions overlaid with NB fits. Solid lines show individual NB distributions. Dashed gray lines show their sum (for bimodal genes). * denotes 95th percentile for Prdm14. p-value: 2 goodness of fit test.
(C) Pairwise relationships between genes, analyzed by smFISH (r, Pearson correlation coefficient; p-value by 2D K-S test (methods, Figure A.2A,B)).
(D) Heat maps show examples of 4-dimensional data sets.

and Esrrb), while in other cases they overlapped strongly (e.g. Nanog and Lifr), such that the absolute number of transcripts for a single gene did not accurately indicate to which state the cell belonged. These bimodal distributions suggested the existence of multiple cell states (see below).

Markers of most differentiated fates including Pax6 (neuroectoderm), Fgf5 (epiblast), Sox17 and FoxA2 (definitive endoderm), and Gata6 (primitive endoderm) showed no detectable expression (data not shown). However, the mesendodermal regulator Brachyury (T) was expressed at a level of ∼5-20 transcripts in 6% of Rex1-low cells. Similarly, the two-cell-like state marker Zscan4c [131] showed ∼3-60 transcripts in 3% of cells (Figure A.2A). These genes did not fit well to NB distributions, suggesting that processes other than transcriptional bursting impact their expression in this small fraction of cells.

Bimodal genes vary coherently

We next used the smFISH data to determine whether the bimodal genes were correlated, which would suggest their control by a single pair of distinct cell states, or varied independently, which would suggest a multiplicity of states. The data revealed a cluster of bimodal genes that correlated with one another. Rex1, Nanog, and Esrrb displayed the strongest correlations (r≈0.7, Figures 2.2, A.2B), while genes with strong overlap between modes, such as Tcl1, Lifr, Sox2, and Tet1, displayed somewhat weaker, but still significant, correlations (r≈0.5, Figures 2.2C, A.2B), beyond those observed between bimodal and non-bimodal genes (e.g. r=0.2 for Rex1 and Oct4). Thus, a cell in the high or low expression state of one bimodal gene is likely to be in the corresponding expression state of others. Some correlations were negative: expression of the *de novo* methyltransferase Dnmt3b was reduced in the Rex1-high state (r=-0.46, Figure 2.2C). Note that cell cycle effects did not

explain these correlations (Figure A.2C). Together, these data suggest that bimodal genes appear to be broadly co-regulated in two distinct states.

Long-tailed genes exhibited more complex relationships. Those with very large variation (CV>1.5) were correlated with the expression state of the bimodal gene cluster, but not with one another (Figures 2.2C & D, A.2B). For example, genes like Dppa3, Tbx3, and Prdm14 burst predominantly in the Rex1-high state, but even in this state, most cells showed no transcripts of these genes ($p<0.001$, see Supp. Info. for statistical analysis). Thus, it appears that these genes are expressed in infrequent, stochastic bursts that occur mainly in one of the two cellular states.

Interestingly, expression of Socs3, a negative regulator and direct target of Stat signaling [132], appeared conditional on expression of its bimodally expressed receptor Lifr (note absence of Socs3 expression in low Lifr cells in Figure A.2B). Analysis of additional regulators not measured here could in principle reveal additional states or more complex distributions. Overall, however, the multi-dimensional mRNA distributions measured here are consistent with a simple picture based on two primary states and stochastic bursting.

The two primary states exhibit distinct DNA methylation profiles

These data contained an intriguing relationship between three factors involved in DNA methylation: the *de novo* methyltransferase Dnmt3b; the hydroxylase Tet1, which has been implicated in removing methylation [133–136]; and Prdm14, which represses expression of Dnmt3b [104, 120, 137, 138]. While Rex1 was anticorrelated with Dnmt3b expression, and positively correlated with Tet1 (Figure 2.3A), Prdm14 showed a long-tailed distribution conditioned on the Rex1-high state (Figure 2.2D). Based on these relationships and the observation that methylation increases during early development [139], we hypothesized

that the Rex1-low state might exhibit increased methylation compared to the Rex1-high state.

To test this hypothesis, we sorted Rex1-high and -low cells using the Rex1-dGFP reporter line, and performed locus-specific bisulfite sequencing at known targets of methylation Dazl, Mael, and Sycp3 [140] (Figures 2.3B, A.3A & B). These promoters exhibited 2-3 times greater methylation in Rex1-low cells compared to Rex1-high cells, indicating the two states are differentially methylated in at least some genes. In contrast, Rex1-low cells that subsequently reverted to the Rex1-high state recovered the methylation levels of Rex1-high cells, indicating that methylation is reversible. Similarly, quantitative enzyme-linked immunosorbent assay (ELISA) analysis demonstrated both differential methylation and reversibility in global methylation levels (Figure 2.3C).

We next asked more generally which genes exhibited differential promoter methylation. We again sorted Rex1-high and -low cells and assayed DNA methylation by Reduced Representation Bisulfite Sequencing analysis (RRBS), analysing regions 2kb upstream to 500bp downstream of each ESC-expressed mRNA transcriptional start site [116, 139]. The distributions of methylation levels across genes were bimodal in both Rex1 states, with the more highly methylated peak shifted to even higher methylation levels in Rex1-low cells (Figure 2.3D). By analyzing the shift in methylation on a gene-by-gene basis, we found that the increase in methylation in Rex1-low cells occurred predominantly through increased methylation of the promoters that were more highly methylated in Rex1-high cells (Figures 2.3E, A.3C). Thus, the change in promoter methylation occurs in a specific subset of promoters. Furthermore, the overall methylation level of a gene was related to the number of CpGs in its promoter, such that the larger the CpG content of a promoter, the lower its methylation in both states. Not all gene promoters were well covered by RRBS. However, among those

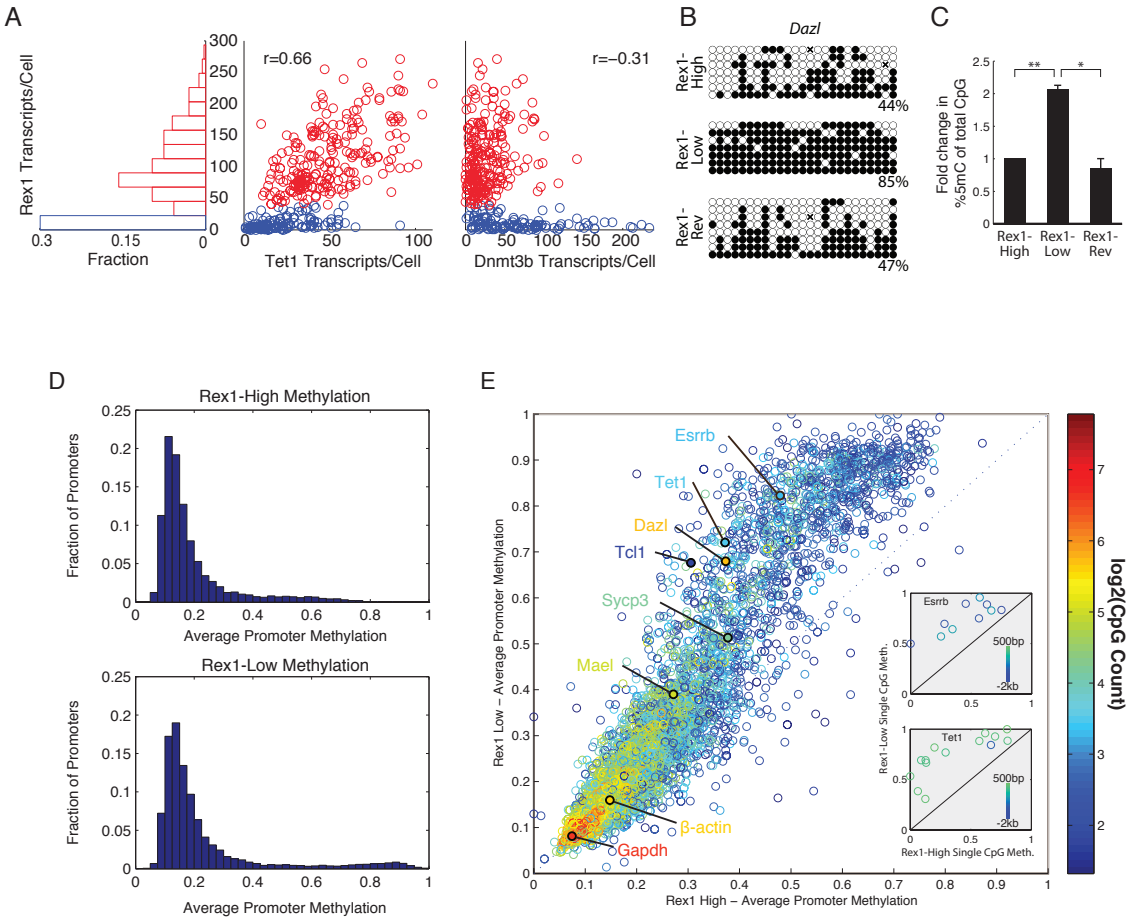


Figure 2.3: The two Rex1 states are differentially methylated
(A) smFISH measurements show Rex1 bimodality is correlated with Tet1, and anticorrelated with Dnmt3b expression.
(B) Locus-specific bisulfite sequencing of the Dazl promoter. Methylation levels shown are in the Rex1-high (top), Rex1-low (middle), and Rex1-low-to-high reverting (bottom) populations.
(C) Global levels of 5mC measured by quantitative ELISA in the Rex1-high, -low, and -low-to-high reverting cells.
(D) Histogram of promoter methylation shows bimodality in the Rex1-high (top) and -low (bottom) states, as quantified by RRBS.
(E) Scatter plot of promoter methylation between Rex1-high and -low states. Each point is the methylation fraction of a single gene promoter, color-coded by the number of CpGs in that promoter. Divergence from the diagonal implies differential methylation between states. Inset) Single CpGs in the promoter of the specific gene labeled, color coded by distance from TSS; see Figure A.3C for additional genes.

that were, several key ES regulators including Esrrb, Tet1, and Tcl1 all showed increased levels of methylation in the Rex1 -low state. Figures 2.3E (inset) and A.3C show methylation levels of individual CpGs for 17 gene promoters. These results provide a view of the change in promoter methylation that occurs during transitions between the Rex1-high and -low states.

Bursty transcription generates dynamic fluctuations in individual cells

Evidently, cells populate two transcriptional states, each characterized by distinct methylation profiles. To understand the dynamic changes in gene expression that occur as individual cells switch between these states, we turned to time-lapse microscopy. We analyzed transcriptional reporter cell lines for Nanog and Oct4, each containing a histone 2B (H2B)-tagged fluorophore expressed under the control of the corresponding promoter (Figure A.4A & B; see also Supp. Info.). We imaged the reporter cell lines for ∼50-hour periods with 15-minute intervals between frames, and segmented and tracked individual cells over time in the resulting image sequences. For each cell lineage, we quantified the instantaneous reporter production rate, defined as the rate of increase of total fluorescent protein in the cell, corrected for the partitioning of fluorescent protein into daughter cells during cell division (Supp. Info.). The H2B-fluorescent protein degradation rate is negligible under these conditions (Figure A.4C), enabling us to use the reporter production rate as a measure of instantaneous mRNA level. An advantage of this approach is that it provides relatively strong fluorescence signals per cell, but still enables high time resolution analysis [141]. Consistent with static smFISH distributions, the production rate distributions of the Nanog and Oct4 fluorescent reporters were bimodal and unimodal, respectively (Figure 2.4A).

Dynamically, cells remained in either one of two distinct Nanog expression states for

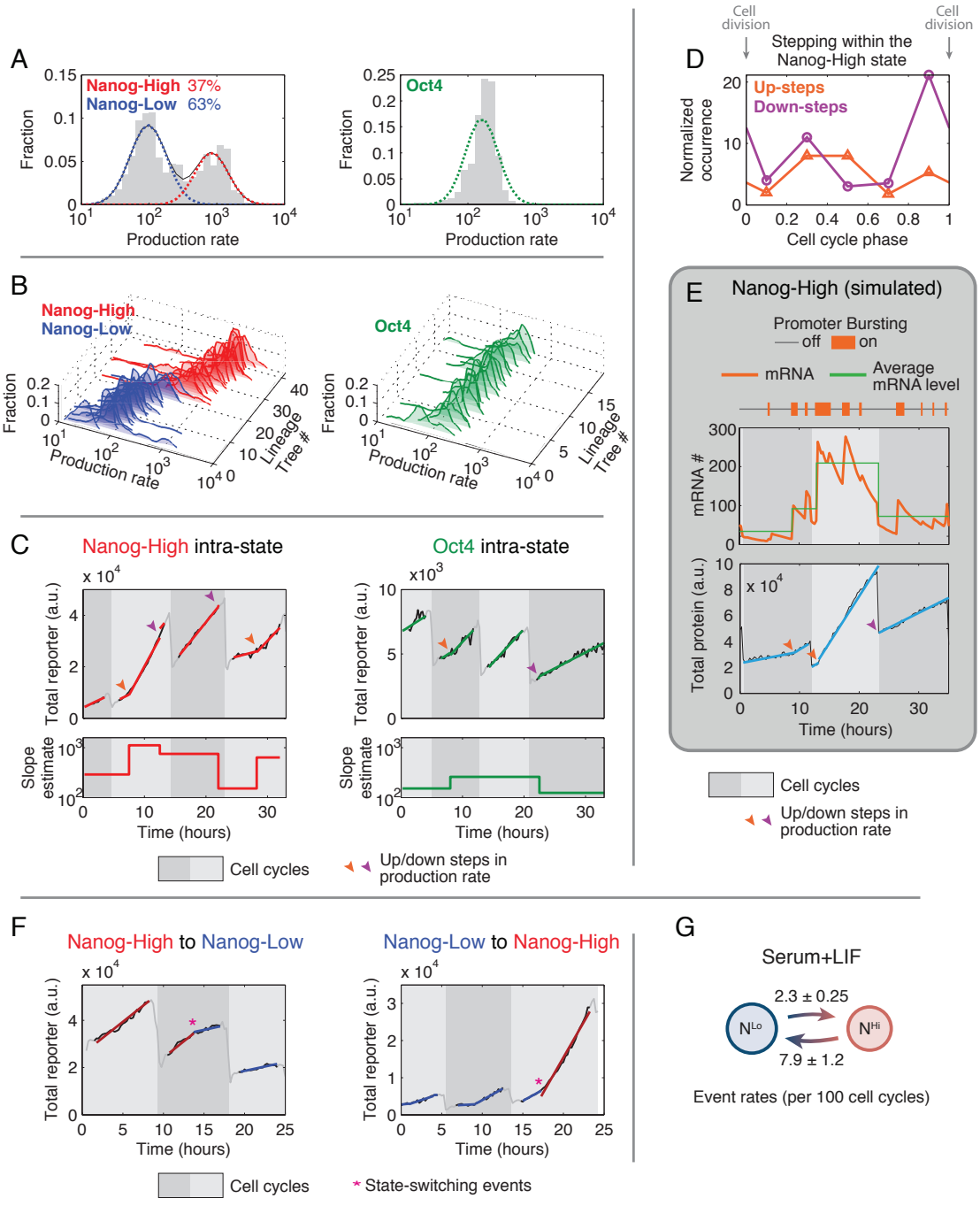


Figure 2.4: (Caption on the following page.)

Figure 2.4: Movies reveal transcriptional bursting and state-switching dynamics in individual cells
(A) Distribution of Nanog and Oct4 production rates from representative movies in serum+LIF, and Gaussian fits to the components. Production rates were extracted from a total of 376 and 103 tracked cell cycles for Nanog and Oct4, respectively.
(B) Production rate distributions of individual cell lineage trees, each consisting of closely related cells descending from a single cell. Lineage trees are color-coded by the state they spend the majority of time in.
(C) Example single lineage traces exhibiting step-like changes in production rates within a state.
(D) Cell cycle phase distribution of steps within the Nanog-high state. Step occurrences are normalized by the frequencies of each cell cycle phase observed in the tracked data.
(E) Representative trace showing apparent steps from simulations under the bursty transcription model, using parameters estimated from mRNA distribution for the Nanog-high state (see Supp. Info.; see Figure A.4E for simulation of Oct4 dynamics).
(F) Example traces of individual cells switching between Nanog-low and Nanog-high states.
(G) Empirical transition rates (mean SD) between the two Nanog states (NHi, Nanog-high; NLo, Nanog-low).

multiple cell cycles (Figure 2.4B). During these periods, expression levels varied over the full range of Nanog expression levels within each state, with no evidence for persistent sub-states. However, closer examination revealed fluctuations within a single state, which typically occurred in discrete steps separated by periods of steady expression (Figure 2.4C). Using a computational step detection algorithm (Figure A.4D, Supp. Info.), we found that Nanog and Oct4 reporters exhibited 0.38 and 0.29 steps per cell cycle, respectively. These steps occurred in a cell cycle phase-dependent manner (Figure 2.4D), with down-steps clustered around cell division events and up-steps more broadly distributed across cell cycle phases.

Could these step-like dynamics arise simply from transcriptional bursting? To address this question, we simulated single cell mRNA and protein traces using the bursty transcription model, with parameters determined from the NB fits of the static mRNA distributions (Supp. Info.). These simulations generated dynamic traces resembling those observed experimentally (Figures 2.4E, A.4E). In the simulations, mRNA half-life and burst frequency

determine the characteristics of detectable steps (Figure A.4F); in general, steps were most prominent at low burst frequencies and short mRNA half-lives, and became difficult to discriminate at high burst frequencies and long mRNA half-lives.

Step-like dynamics appear to be a natural consequence of stochastic expression, with up-steps reflecting burst-like production of mRNA, and down-steps resulting from ∼2-fold reduction in mRNA copy number at cell division (Figure A.4G). This interpretation is consistent with the observed clustering of down-steps around cell division events, and a more uniform cell cycle distribution of up-steps (Figure 2.4D). Because large bursts can effectively cancel out mRNA dilution at cell division events, they may appear under-represented near cell division events. Note that most cell cycles showed no up-steps, suggesting that they are not due to increased gene dosage after chromosome replication [142, 143].

Dynamic transitions between cellular states

We next asked how cells transition dynamically between states. Previous work has relied on cell sorting, which can distort the signaling environments. By contrast, movies enable direct observation of switching events within a mixed cell population. Since the Nanog reporter production rate fluctuates even within a single state, we used a Hidden Markov Model (HMM) to classify each cell into either Nanog-high or Nanog-low at every point in time (Supp. Info.). We trained the HMM using time-series data of Nanog reporter production rates, sampled at fixed intervals across all tracked cell cycles, and used it to identify switching events and estimate switching frequencies.

Transitions from the Nanog-low to the Nanog-high state, or vice versa, occurred at a rate of 2.3 ± 0.25, or 7.9 ± 1.2, transitions per 100 cell cycles (mean $\pm$ SD), respectively (Figure 2.4F & G). These events did not correlate between sister cells (Table A.11), consistent with

independent, stochastic events. Inter-state switching on average showed bigger and longer-lasting fold-changes than intra-state steps in production rates (Figure A.4H). Together, these results suggest that gene expression dynamics are dominated by a combination of step-like changes due to bursty transcription on shorter timescales, and abrupt, apparently stochastic, inter-state switching events on longer timescales.

'2i' inhibitors modulate bursty transcription and state-switching dynamics

We next asked how gene expression heterogeneity and dynamics change in response to key perturbations. Dual inhibition of MEK and GSK3, known as '2i', was shown to enhance pluripotency and reduce Rex1 and Nanog heterogeneity [116, 129]. However, it has remained unclear how 2i affects the distribution of other heterogeneously expressed regulatory genes, and what impact it has on dynamic fluctuations in gene expression.

We found that addition of 2i to serum+LIF media reduced variability in the mRNA levels of most genes (Figure 2.5A). In principle, this could reflect the elimination of one cellular state and/or changes in burst parameters. In 2i, the bimodal genes from Figure 2.2A became unimodal, suggesting that 2i suppresses one of the two cellular states (Figures 2.5A, A.5A). In the case of Nanog, the remaining state increased its mean transcript level by ∼1.5-fold, to what we term Nanog-SH (Super High). Tet1, Sox2, and Tcl1 also became unimodal, but displayed an overall decrease in absolute expression. With long-tailed genes, we found that mean Dppa3 expression decreased slightly, while Prdm14 and Tbx3 became more homogenously expressed, exhibiting an increase in mean expression by ∼300% and ∼1000%, respectively. These changes reflect the fact that nearly all cells were now observed to express Prdm14 and Tbx3. Thus, 2i appeared to reduce variability in most genes, either by eliminating bimodality or by increasing their burst frequency.

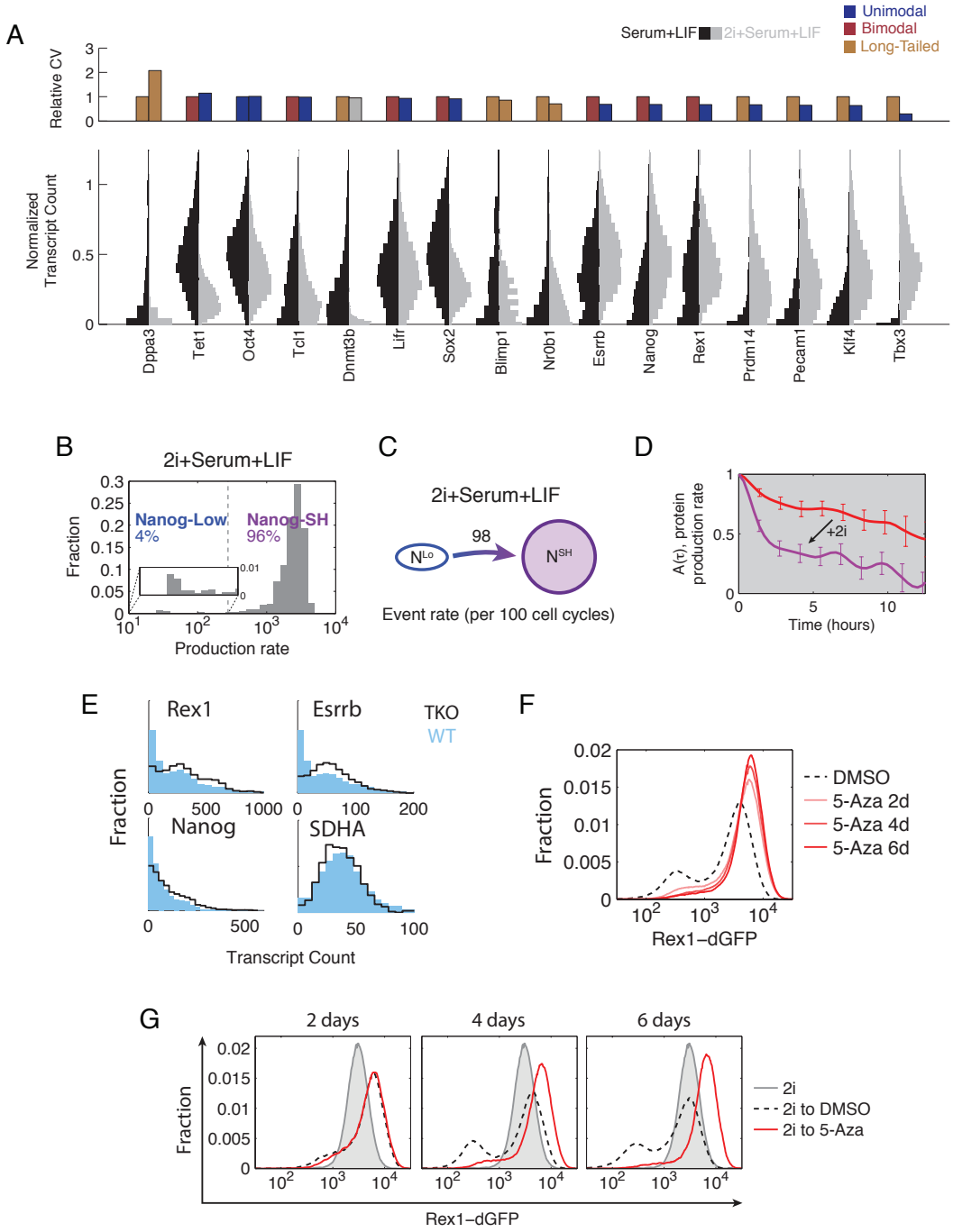


Figure 2.5: (Caption on the following page.)

Figure 2.5: 2i and DNA methylation modulate bursty transcription and state-switching dynamics
(A) Comparison of mRNA distributions and CV between cells grown in serum+LIF and 2i+serum+LIF. Top: For each gene, the CV in serum+LIF is plotted on the left, and the CV for 2i+serum+LIF is plotted on the right. Dnmt3b in 2i+serum+LIF is represented in gray to reflect its marginal case of poor quality of fit in both bimodal and long-tailed models. Bottom: The left half of each violin represents the mRNA distribution in serum+LIF, while the right represents 2i+serum+LIF. For each gene, both conditions are normalized by the same value that is the larger of the pair's 95th percentile expression level.
(B) Distribution of Nanog production rates from movies in 2i+serum+LIF.
(C) Empirical transition rates between the two Nanog states in the presence of 2i (NLo, Nanog-low; NSH, Nanog-SH).
(D) Mixing time in each condition is estimated from auto-correlation, A(τ), of production rate ranks shown in Figure A.5D, right panels. Red: serum+LIF; Purple: 2i+serum+LIF; Error bars: standard deviation, bootstrap method.
(E) Comparison of transcriptional heterogeneity between Dnmt TKO (black line) and the parental line (blue bars) as measured by smFISH for Rex1, Nanog, Esrrb, and SDHA. Note that for Rex1/Nanog/Essrb, there are fewer off cells in the leftmost bins for the TKO than WT.
(F) Rex1-dGFP distribution as measured by flow cytometry grown in serum+LIF with 5-aza or DMSO (carrier control). Time-points were taken after 2, 4, and 6 days.
(G) Cells were grown in 2i+serum+LIF, and subsequently re-plated into serum+LIF with 5-aza or DMSO (carrier control). Time-points were taken after 2, 4, and 6 days. GFP levels were measured by flow-cytometry.

Recently, it was shown that 2i-treated cells exhibit differentiation propensities similar to sorted Rex1-high subpopulations in embryoid body formation, suggesting they may represent similar functional states [116]. We used the time-lapse movie system to compare the dynamic behavior of 2i-treated cells to that of cells in the Rex1-high subpopulation. Consistent with mRNA measurements, 2i shifted most cells into a Nanog-SH state (96% of total), characterized by a $\sim$3-fold higher median production rate compared to the Nanog-high state in serum+LIF (Figure 2.5B). Only a small fraction of cells showed expression overlapping with the Nanog-low state in serum+LIF at the beginning of the movies (after 6 days in treatment). Moreover, these cells switched to the Nanog-SH state at a >40-fold higher rate than the Nanog-low to Nanog-high switching rate measured in serum+LIF, with no reverse transitions observed (Figures 2.5C, A.5B). These observations suggest that 2i in-

creases the Nanog-low to Nanog-high switching rate and reduces or eliminates Nanog-high to Nanog-low transitions (Figure 2.5C).

What effect, if any, does 2i have on the dynamics of gene expression noise? Static distributions suggested that 2i increased Nanog burst frequency by ∼45%, from 0.39 to 0.55 burst/hour, using Nanog mRNA half-life previously estimated (Table A.11 in [144]) and assuming no change between conditions. To analyze the effects on dynamics, we computed the mixing time, previously introduced to quantify the mean timescale over which a cell maintains a given expression level relative to the rest of the cell population [8]. Simulations of the bursty gene expression model showed that higher burst frequencies lead to faster mixing times, while burst size has little effect (Figure A.5C). Together with smFISH measurements, this model predicted that Nanog mixing times should be faster in 2i. Qualitatively consistent with this prediction, the mixing time of Nanog production rate was reduced from 8.5 hours in Nanog-high in serum+LIF media to 1.7 in Nanog-SH cells in 2i-containing media (Figures 2.5D, A.5D).

Together, these results indicate that 2i impacts ESC heterogeneity at three levels: First, it reduces gene expression variation in many, but not all, genes. Second, it eliminates one cell state by increasing the rate of transitions out of the Nanog-low state and inhibiting the reverse transition. Third, as shown for Nanog, 2i increases burst frequency and reduces mixing times, effectively speeding up the intra-state equilibration rate within the cell population.

DNA methylation modulates metastability

Previous work has shown that in addition to reducing heterogeneity, 2i also diminishes global levels of DNA methylation [119, 120, 133]. While the Rex1-high and -low states

appear differentially methylated (Figures 2.3B-E), it remains unclear whether methylation plays a functional role in stabilizing these states. To address this issue, we used a triple-knockout (TKO) cell line lacking the active DNA methyltransferases Dnmt1, Dnmt3A, and Dnmt3B [145]. We compared the expression distribution of Rex1, Nanog, and Esrrb in TKO cells to their parental line using smFISH. The TKO cell lines had 352% fewer cells in the Rex1-low state (Figure 2.5E), with similar results observed for Nanog and Esrrb. This change did not reflect global up-regulation of all genes, as expression of the houskeeping gene SDHA was indistinguishable between the two cell lines. These results suggest that DNA methyltransferases increase the population of the Rex1-low state.

To see if these results could be recapitulated with acute rather than chronic perturbations to methylation, we assayed changes in heterogeneity in Rex1-dGFP reporter cells exposed to 70nM 5-azacytidine (5-aza), an inhibitor of DNA methylation. Within six days, the number of cells in the Rex1-low state diminished by more than half from 29% to 13% of all cells (Figure 2.5F). Thus, acute as well as chronic methylation inhibition reduced the occupancy of the low state.

Finally, we asked whether methylation was similarly required for cells to return to the low state after removal of 2i from 2i+serum+LIF conditions. The Rex1-low population began to emerge within 48h of 2i removal (Figure 2.5G). However, when 2i was removed and replaced with 5-aza, the emergence of Rex1-low cells was severely delayed and diminished. After 6 days, 5-aza treated cells only showed 6% Rex1-low cells, compared to 25% in DMSO-treated cells. Together, these results suggest that methylation is required for normal exit from the 2i state. Reduced methylation in 2i thus contributes to the stability of the 2i 'ground state' [116, 119, 120, 133].

2.4 Discussion

Recent work on ESC biology from a systems perspective has highlighted the apparent complexity and strong interconnectivity of the circuit governing pluripotency [146, 147]. But it has been unclear how variably this circuit behaves in different cells, and to what extent population average measurement techniques may obscure its single-cell dynamics. Because gene expression is a stochastic process, levels of both mRNA and protein in each cell are effectively random variables, best characterized by their distributions. The framework of stochastic gene expression provides a tool to more rigorously and quantitatively separate stochastic fluctuations inherent to gene expression from variation due to multiple cell states specified by the underlying transcriptional and signaling circuit. While the simplified model of bursty transcription used here can explain the data, other models, including the telegraph model of transcription, may provide other insights [123, 148].

Our data suggest that heterogeneity emerges in three distinct ways: First, gene expression is inherently noisy, occurring in stochastic bursts, even in genes such as Oct4 that are distributed in a relatively uniform fashion. Second, cells switch stochastically between distinct states that impact the expression of many genes in a coordinated manner. Third, 'long-tailed' regulators such as Prdm14, Tbx3, and Dppa3 are uncorrelated with one another and show low burst frequencies and large burst sizes, leading to very high variability. Live cell imaging will be required to determine the absolute burst kinetics for these genes. However, an mRNA distribution in which only a small subpopulation of cells exhibit a large number of mRNA molecules for a particular gene need not, and most likely does not, indicate a distinct cellular state. Moreover, infrequent bursting may provide a potential mechanism for stochastic priming of cell fate decision-making [13, 127]. Further investiga-

tion of this possibility will require determining whether these bursts propagate to influence subsequent cell fate decision-making events [149, 150].

The data above implicate methylation as a key regulatory mechanism affecting stochastic switching between states. Methylation was previously explored in ES cells at the population level [114, 115, 119, 120, 133, 134], but it remained unclear whether methylation itself contributes to heterogeneity [69, 74, 75, 104]. smFISH data revealed a strong reciprocal relationship between the hydroxymethylase Tet1 and the DNA methyltransferase Dnmt3b, with Tet1 expressed more highly in the Rex1-high state, and Dnmt3b expressed more highly in the Rex1-low state. This difference in expression correlates with a differential global DNA methylation and in the methylation of the promoters of key pluripotency regulators. Finally, methylation appears to be functionally required for transitions, since either genetic deletion of DNA methyltransferases or pharmacological inhibition both impact the populations of the two cell states and the underlying dynamics of state-switching (Figure 2.5E-G). It will be interesting to see whether methylation plays similar functional roles in other stochastic state-switching systems.

These data provoke further questions about the molecular mechanisms through which methylation is regulated and through which it modulates metastability. For example, while known methyl binding proteins that aid in methylation-dependent chromatin compaction and silencing are expressed in ES cells [116], DNA methylation may also inhibit binding of transcription factors [151–153], and can control mRNA isoform selection via alternative splicing [151]. The Esrrb gene, whose activity is central to maintenance of pluripotency [154, 155], may provide a good model system to investigate the effects of methylation, since its methylation levels and expression levels are both strongly state-dependent. Regulation of this methylation likely involves Prdm14, which is known to inhibit Dnmt3b expression

[104, 118–120, 137, 138]. Given the long-tailed expression pattern of Prdm14 observed here in serum+LIF and its strong up-regulation in 2i, it will be interesting to see how much of the variation in Dnmt3b/Tet1, and methylation more generally, can be attributed to bursty expression of Prdm14.

Previous studies of ESC gene expression dynamics have focused on the equilibration of FACS-sorted subpopulations of high and low Nanog and Rex1 expression [69, 75]. Two groups explored transcriptional circuit models to explain the long timescales of state-switching dynamics [58, 59]. These included noise-induced bistable switches, oscillators, and noise-excitable circuits [156]. Our dynamic data demonstrate that both Nanog-high and Nanog-low states in serum+LIF conditions typically persist for 4 cell cycles, and that state-switching events are abrupt at the level of promoter activity. Depending on protein and mRNA half-lives, the timescale of resulting protein level changes may follow somewhat more slowly. State-switching events are also infrequent (<10% per cell cycle), and uncorrelated between sister cells. Together, these findings appear incompatible with oscillatory or excitable models, which predict deterministic state-switching or an unstable excited state, respectively, but are consistent with the stochastic bistable switch model previously proposed [58]. These properties could make this state-switching system a useful model for understanding the circuit level dynamics of spontaneous cell state transitions in single cells.

Several competing explanations were proposed for the apparent heterogeneity in Nanog expression in serum+LIF conditions. These models suggest heterogeneity is an artifact of knock-in reporters [157], or that it arises at least in part from monoallelic regulation [158] or is manifested biallelically [159, 160]. Our smFISH data support the existence of Nanog expression heterogeneity in wild-type cells in a standard feeder-free culture condition. Further, both static and dynamic measurements indicate that intra-state heterogeneity in

Nanog is consistent with bursty transcription, with a relatively low burst frequency of ∼0.39 burst/hour. Thus, active transcriptional loci analyzed by smFISH against nascent transcripts [158] would be expected to 'flicker' on and off stochastically due to bursting. Such bursting could also lead to the misleading appearance of weak correlations between alleles in static snapshots, and in measurements based on destabilized fluorescent reporters. On the other hand, the Nanog protein fusion reporters analyzed by [159] showed correlated static levels between alleles, likely because the longer lifetime of their reporters allowed integration of signal over many transcriptional bursts, and because transitions between cellular states are rare and affect both alleles in a correlated fashion. The results of Faddah et al. with dual transcriptional reporters similarly showed general correlations between the two alleles, consistent with the smFISH correlations shown here (Figures A.1E, A.4B). Taken together, these previous studies and the data presented here appear to converge on a relatively simple picture of heterogeneity based on two states and stochastic bursting.

The quantitative measurement and analysis platform described above should enable further investigation of the structure of static and dynamic heterogeneity in single ESCs. With the advent of higher dimensionality smFISH [161, 162], single-cell RNA-Seq, and microfluidic high-throughput qPCR approaches, as well as improved methods for rapidly and accurately constructing knock-in reporters [163], it will soon be possible to explore the dynamics of ESC components in higher dimensions in individual cells, both within metastable states and during cell state transitions [164]. Ultimately, this should provide the capability of better understanding the dynamic architecture of cell fate transition circuits.

2.5 Materials and Methods

Culture conditions and cell lines

E14 cells (E14Tg2a.4) obtained from Mutant Mouse Regional Resource Centers were used for smFISH studies. NKICit cells, created by Kathrin Plath, were generated by targeting the endogenous Nanog locus in V6.5 cells with H2B-Citrine-IRES-Neo-SV40pA (Figure A.4A). NKICit+Cer cells were generated by randomly integrating into NKICit cells a linearized PGK-H2B-Cerulean-BGHpA-SV40-Hygro-SV40pA vector. OBACCer cells were generated by randomly integrating into E14 cells (a kind gift from Bill Skarnes and Peri Tate) a linearized bacterial artificial chromosome (BAC) containing the Oct4 locus (BACPAC (CHORI)), in which H2B-Cerulean-SV40pA-PGK-Neo-BGHpA was inserted before the coding sequence (Figure A.4A). Rex1-dGFP cells were kindly provided by the Austin Smith lab [117]. All cells were maintained on gelatin-coated dishes without feeders.

smFISH hybridization, imaging, and analysis

The RNA FISH protocol from [105] was adapted for fixed cells in suspension. See supplemental experimental procedures for details. Semi-automated dot detection and segmentation were performed using custom Matlab software. A Laplacian-of-Gaussian (LoG) Kernel was used to score potential dots across all cells. The distribution of these scores across all potential dots was thresholded by Otsu's method to discriminate between true dots and background dots (see Figure A.1A-D).

mRNA distribution fitting

The Negative Binomial Distribution is defined as

$$P(n, r, p_o) = \binom{n + r - 1}{n} p_0^r (1 - p_o)^n,$$

where n=number of transcripts per cell, p_0 = probability of transition from the on promoter state to the off promoter state, and r = number of bursting events per mRNA half-life. The average burst size is computed as $b = (1 - p_0)/p_0$. Using this model, individual mRNA distributions were fit using maximum likelihood estimation. To discriminate between unimodal and bimodal fits, two tests were used to ensure that the improvement of the fit was counterbalanced by the additional degrees of freedom from the added parameters. To be considered bimodal, a distribution was required to pass both Akaike Information Criteria (AIC) and the log-likelihood ratio test ($p < 0.05$).

Fluorescence time-lapse microscopy and data analysis

Reporter cells were mixed with unlabeled parental cells at 1:9 ratio and plated at a total density of 20,000 cells/cm^2 on glass-bottom plates (MatTek) coated with human laminin-511 (BioLamina) >12 hours before imaging. Images were acquired every 15 minutes for ∼50 hours with daily medium change. Cells were segmented and tracked from the acquired images using our own Matlab code (see supplementary for image analysis methods).

2i perturbation and analysis

For 2i treatment we supplemented serum+LIF media with MEK inhibitor PD0325901 at 1uM and GSK3 inhibitor CH99021 at 3uM. Cells grown in serum+LIF media were treated

with 2i for 6 days before harvesting for smFISH assay and imaging for movies.

Methylation analysis and perturbation

RRBS preparation and high-throughput sequencing were performed by Zymo Research. Analysis was performed using Bismark and Galaxy, with a single CpG coverage threshold 5. 5-azacytidine (Sigma) was used at a final concentration of 70nM. 5mC ELISA was performed with ELISA 5mC kit (Zymo).

Accession information

Sequencing data has been deposted in NCBI's GEO under accession number GSE58396.

Chapter 3

Dynamics of Epigenetic Regulation at the Single-Cell Level

(This chapter was adapted, in part, from a manuscript under review [165])

3.1 Abstract

Chromatin regulation enables cells to establish heritable gene expression states. However, how chromatin regulators control gene expression in an individual cell, quantitatively and over time, has remained unclear. Here, we analyzed the dynamics of gene expression in individual cells before, during, and after recruitment of four diverse chromatin regulators (Dnmt3b, HDAC4, EED, and KRAB) to the promoter of a reporter gene. We found that gene silencing and subsequent reactivation both occurred through abrupt, all-or-none, and stochastic events. Remarkably, the effects of all factors could be quantitatively described by a unified model based on stochastic transitions among three discrete states: actively expressing, reversibly silent, and irreversibly silent, with each factor generating a distinct set of transition rates. These dynamics provide a simple mechanism for controlling the fraction of cells in a given gene expression state. They also represent a predictive, statistical framework for analyzing dynamic chromatin regulation and for engineering mammalian gene

circuits with epigenetic memory.

3.2 Introduction

Living cells can establish gene expression states that persist for many generations. Such epigenetic regulation plays critical roles in almost every aspect of multicellular development, physiology, aging, and disease [166–168]. It is implemented through a system of chromatin regulators (CRs) that read, write, and erase chemical modifications on specific histone residues or DNA bases to set and maintain gene expression states [17, 22]. Previous work suggests that CR-based regulation can provide key dynamic capabilities to cells and developing organisms, such as ensuring the irreversibility of cell fate decision-making events [169–171], integrating signals over very long timescales [172, 173], and maintaining long-term silencing of foreign or viral DNA elements (reviewed in [174]). The ability of CRs to implement these functions depends on the quantitative and dynamic manner in which they regulate expression of their target genes.

Recent work has begun to establish a framework for quantitative analysis of conventional gene regulation systems, in which both rates of expression and levels of fluctuation (noise) depend on the instantaneous concentrations of transcription factors in a continuous fashion, both in prokaryotic and eukaryotic cells [34, 175–177]. A similar framework for understanding how CRs dynamically control regulation of their target genes could provide a better understanding of many biological processes, and enable the engineering of new cellular capabilities based on chromatin regulation. Achieving such a framework will require the addressing of one central issue: how the recruitment or removal of a given CR at a particular gene influences the present and future expression of that gene. More specifically, we need to know: (1) what range of expression levels a CR can generate, and whether they are

continuous or discrete; (2) how rapidly and permanently a CR exerts its effects; and (3) how deterministic or stochastic the overall process of chromatin regulation is. Finally, we will need to understand how all of these characteristics depend on the type of CR considered.

Measuring these dynamic properties of CRs requires one to directly manipulate the binding of a CR at a given site and analyze its subsequent effects on gene expression in individual cells. To control binding, one must be able to reversibly recruit each CR. Ideally this should be done at the same locus across CRs to enable controlled comparisons. Such a controlled recruitment system was previously achieved by fusing specific factors to DNA binding domains [98, 178–184], or more recently by using the CRSPR/Cas9 system [185]. Controlled recruitment has also been used to systematically compare the activities of different CRs in yeast [186]. However, none of these studies analyzed the effects of CR recruitment dynamically in individual cells.

To follow the effects of CRs during and after recruitment requires quantitative single cell time-lapse analysis of gene expression. This analysis should extend over multiple cell cycles to capture the range of timescales over which chromatin regulation occurs. This approach has provided powerful insights into gene regulation and differentiation circuits in microbial and mammalian cells [141, 187–190]. Extending these techniques to CR recruitment can elucidate the effects of different CRs on gene expression states in individual cells.

Here, we measured the dynamics of gene expression in single mammalian cells in response to recruitment and de-recruitment of four CRs spanning a diverse set of the most studied chromatin modifications: EED and KRAB (histone methylation), Dnmt3b (DNA methylation), and HDAC4 (histone deacetylation). Each was previously shown to be sufficient to implement gene silencing when recruited to a target promoter [178, 180, 182, 184], making them good candidates for use in synthetic regulatory systems. EED functions as

part of the Polycomb Repressive Complex 2 (PRC2). This complex methylates histone H3 at lysine 27 (H3K27me3) and is essential in repression of cell fate regulators [191], and many other developmental processes [192]. KRAB (Krppel-associated box) is a silencing domain that appears in over 400 zinc finger transcription factors [193] and is popularly utilized in genetic engineering applications [185, 194, 195]. KRAB was previously shown to associate with the histone methylase SetDB1, a *writer* of histone methylation at lysine 9 (H3K9me3), as well with the chromatin compacting protein HP1 that binds to H3K9me3 [196]. Dnmt3b causes *de novo* methylation of CpGs to silence target genes during embryonic development [109]. Finally, the histone deacetylase HDAC4 silences gene expression by removing acetyl groups from histones H3 and H4 [197], and is recruited to specific genes involved in bone and muscle development by transcription factors [198].

We found that, despite their diverse mechanisms of action, all four CRs exhibited common features. Changes in gene expression occurred predominantly through stochastic, abrupt, all-or-none silencing and reactivation events, suggesting that chromatin regulation generates probabilistic, binary control over the rate of gene expression. Our results indicate that chromatin regulation can be operationally described as a stochastic three-state process: CR recruitment can switch actively expressing cells into a reversibly silenced state. If CR recruitment is sustained, cells can further switch to an irreversibly silent state. Upon removal of the CR, reactivation from the reversibly silent state occurs in a stochastic fashion.

While all factors generated dynamic behaviors consistent with this simple general scheme, the different CRs behaved quite differently from one another. They produced distinct molecular signatures during silencing and exhibited factor-specific rates of silencing and reactivation. As a result, different CRs produced different levels of epigenetic memory, ranging

from transient (days, HDAC4), to permanent (over a month, Dnmt3b). They also implemented unexpected hybrid behaviors based on a mixture of slowly reversible (weeks) and permanent memory states in the same cell population (EED and KRAB). These results suggest that even though all of the CRs analyzed here are capable of silencing, each generates qualitatively different types and timescales of regulation and epigenetic memory.

Finally, we discuss how the all-or-none, stochastic nature of epigenetic regulation enables population-level control of the fraction of cells in a particular state in response to signals. Taken together, these results provide a framework for operational analysis of chromatin regulation and, more generally, establish the beginnings of a 'design guide' for future use of CRs in synthetic mammalian gene circuits.

3.3 Results

A platform for analysis of silencing and reactivation dynamics in individual cells

In order to analyze the effects of CRs at a single, well-controlled locus in individual cells, we constructed a reporter cell line containing the histone 2B (H2B)-Citrine gene driven by the constitutive pEF promoter [199]. pEF naturally drives the expression of translation elongation factor 1 in mammalian cells, and, unlike some viral promoters, tends to resist spontaneous silencing [200]. In order to provide a controlled chromatin environment for the reporter gene, we flanked it with two tandem copies of the HS4 insulator [201], and stably integrated the resulting construct into a human artificial chromosome (HAC) in Chinese Hamster Ovary (CHO) cells (Figure 3.1A). The HAC provides an ideal platform for this analysis because it is independent of endogenous chromosomes, exists in a single copy in

the cell and can potentially be moved to other cell lines for future analyses in other cell contexts [202, 203].

To enable control over the timing and duration of recruitment, each CR was fused to reverse Tet repressor (rTetR), a domain that binds DNA only in the presence of doxycycline (dox) [204] (Figure 3.1A). This inducible domain allows CR recruitment at an array of Tet operator (TetO) binding sites near the reporter gene. Since these binding sites are upstream of the promoter region, recruitment of rTetR alone does not repress reporter expression (Figure B.1A). Using a separate cell line, we verified that rTetR binding is reversible, and that our system can detect responses to the addition and washout of dox within 82 hours and 2.52 hours, respectively (Figure B.1B). These results indicate that the rTetR system enables analysis of epigenetic silencing and reactivation dynamics over timescales as short as a few hours.

We next generated cell lines that constitutively express each of the rTetR-CR fusion proteins by randomly and stably integrating each rTetR-CR construct into the reporter cell line. These constructs also constitutively co-express H2B-mCherry to enable cell tracking even when the H2B-Citrine reporter is silenced (Figure 3.1A). Prior to dox treatment, the CR cell lines expressed H2B-Citrine at levels comparable to those in the parental reporter line (Figure B.1C). Addition of dox triggered binding of the rTetR-CR fusion proteins upstream of the reporter (Figure 3.1B, 'recruitment'). This recruitment initiated downstream regulatory processes that led to gene silencing (Figure 3.1C). Removal of dox released the CR from the reporter locus (Figure 3.1B, 'de-recruitment'), allowing us to test the heritability of silencing and the dynamics of reactivation (Figure 3.1D). Thus, the resulting cell lines enabled comparison of silencing and reactivation dynamics across the four CRs.

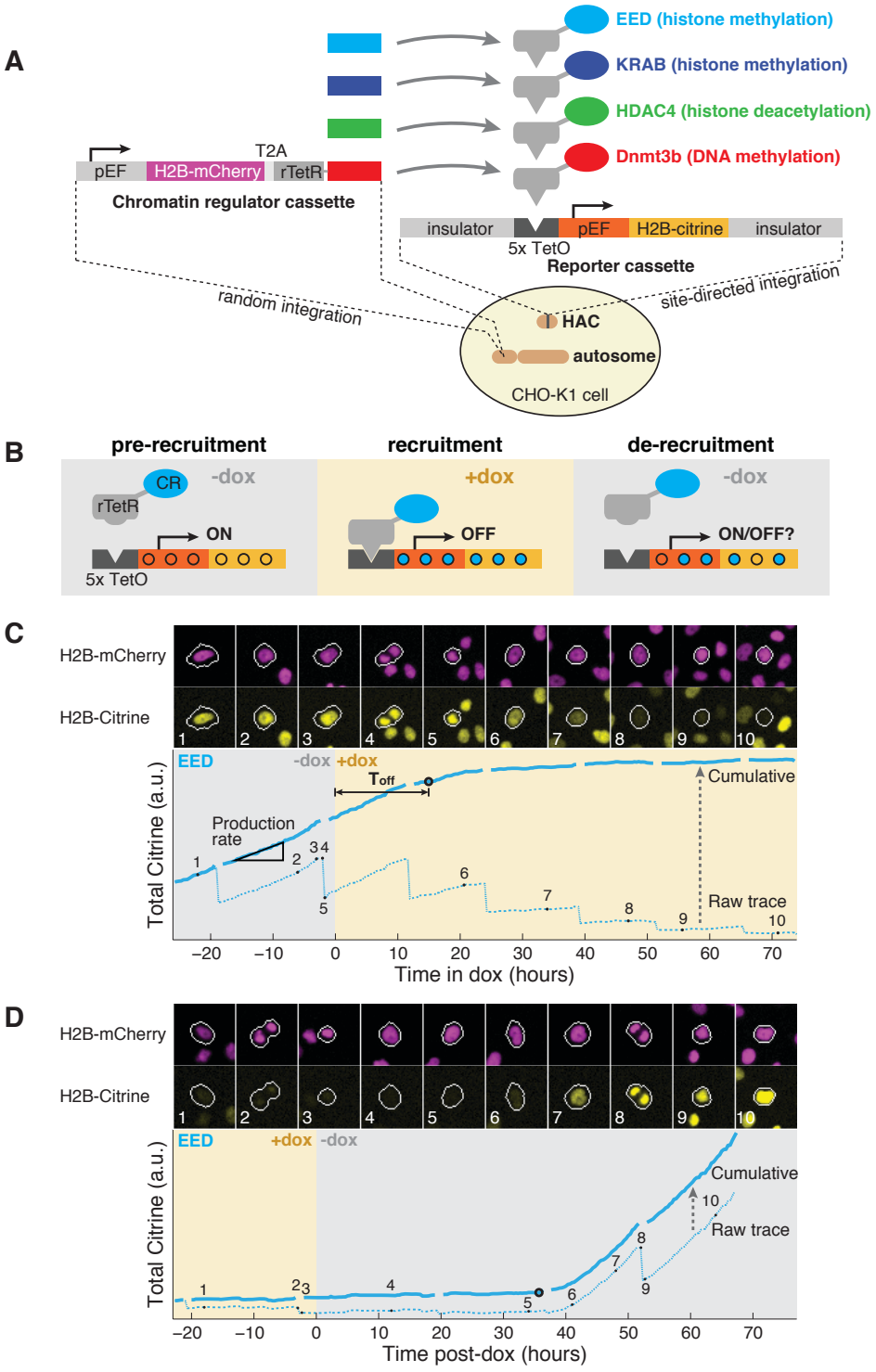


Figure 3.1: (Caption on the following page.)

Figure 3.1: Platform for studying the dynamic control of epigenetic regulation
(A) Schematic of cell line design. An H2B-Citrine reporter cassette was integrated at a specific site on a human artificial chromosome (HAC) previously introduced in CHO-K1 cells. Cassettes constitutively expressing chromatin regulators fused with rTetR (top four constructs), as well as H2B-mCherry, were added separately through random but stable integration to create four separate cell lines (one for each CR).
(B) rTetR is recruited to TetO binding sites only in the presence of dox. Experiments involved recruitment of regulators by addition of dox (second panel) and subsequent de-recruitment by removal of dox (third panel). Recruitment can add or remove modifications (blue dots) and cause changes in reporter expression.
(C & D) Filmstrips and corresponding fluorescence traces of single reporter cells responding to recruitment (C) and de-recruitment (D) of rTetR-EED. H2B-mCherry and H2B-Citrine are shown in magenta and yellow, respectively. H2B-Citrine reporter fluorescence is plotted as function of time for the cell lineages circled in white (raw trace, dotted line). Cumulative fluorescence traces (solid cyan) were obtained by computationally restoring the constant fluorescence level lost to the sister cell at each division. This procedure facilitates continuous quantification of reporter production rate (slope of cumulative trace). Black circles indicate inflection points identified during silencing or reactivation events. Time points indicated by numbers 1-10 in the graph refer to corresponding frames in the filmstrip above. Cell lineages in (C) and (D) are highlighted in movies S1 and S5, respectively.

Silencing dynamics at the single-cell level

We first set out to characterize basic aspects of silencing dynamics – speed, extent, and variability among cells – and compare them across CRs. We acquired time-lapse movies, beginning with cells not previously exposed to dox. In these cells the reporter gene was actively expressed. After ∼24 hours, we added dox to initiate recruitment of a CR, and continued to record movies for approximately three more days (Figure 3.1C, top). From these movies, we extracted total H2B-Citrine fluorescence per cell over time. These traces have a sawtooth appearance, reflecting periods of constant accumulation of fluorescent protein punctuated by cell division events at which ∼50% of fluorescence is lost to the sister cells (dashed lines, Figure 3.1C). In order to focus on the rate of H2B-Citrine production, we computed a cumulative trace, by adding back fluorescence levels lost to partitioning (solid line, Figure 3.1C). Using these cumulative traces, we calculated the reporter production rate, defined as the rate of increase of cumulative fluorescent protein in an individual cell

(slope of solid line, Figure 3.1C). This analysis allowed us to identify silencing events, defined as transitions from high to low production rate (circle, Figure 3.1C), and record their timing relative to dox addition, T_{off} (Figure B.1E & F, see also Experimental Procedures).

The timing of silencing events (T_{off}) varied widely between cells and depended on the CR used (Figure 3.2A & B). Silencing by EED and Dnmt3b occurred at a low rate, and with a broad distribution of T_{off} values, 41.4±17.2 and 50.6±19.0 hrs, respectively (mean±s.d. of all events pooled from independent experiments). With these two factors, we observed silencing in individual cells continually throughout the entire duration of the movie. In contrast, silencing mediated by KRAB and HDAC4 was much faster, with mean T_{off} values of 6.0±3.3 and 16.0±13.3 hrs, respectively. For these factors, most cells were silenced by 24 hrs (Figure 3.2B). The broad variability in silencing times observed here suggests that chromatin silencing is a stochastic process.

For factors with silencing times (T_{off}) extending beyond the first cell cycle (EED and Dnmt3b), we tested whether silencing between sister cells was correlated. Consistent with a stochastic mechanism, we found no strict pattern of silencing between sister cells. Nevertheless, we observed a higher frequency of concordant silencing between sister cells than expected if silencing were perfectly independent (Figure B.2A), indicating that a weak predisposition to silencing can be inherited from parent to daughter cells. This behavior could reflect an underlying multi-step process that extends across cell division events, or slow fluctuations in the concentrations of trans-acting factors.

Despite variability in silencing times, the profiles of individual silencing events were remarkably similar between cells and across different CRs. This similarity becomes more obvious when single-cell traces are aligned around the individual silencing events (Figure 3.2C). Averaging these aligned traces revealed that recruitment of all four factors led to complete

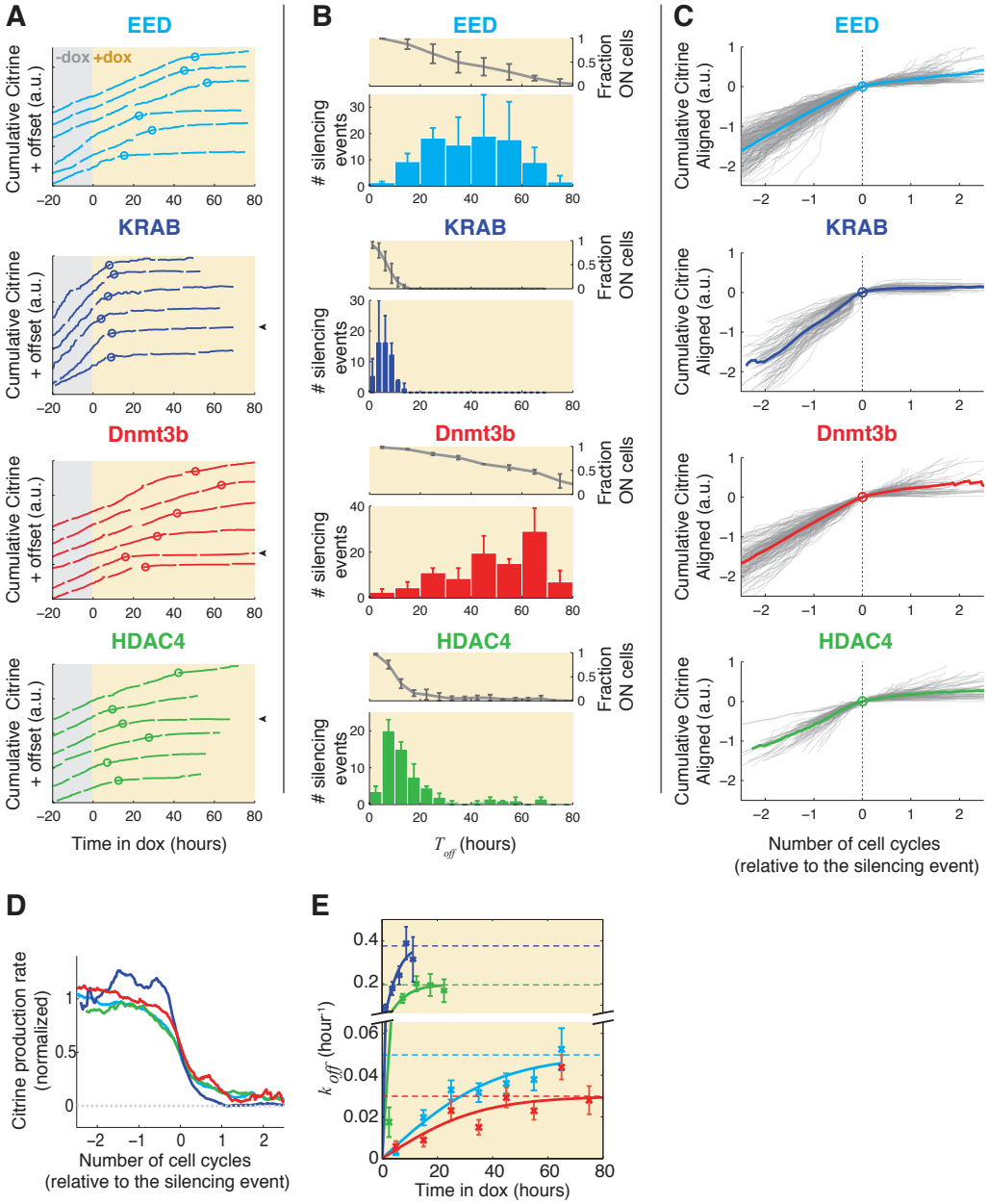


Figure 3.2: (Caption on the following page.)

Figure 3.2: Silencing occurs in a stochastic, abrupt, and all-or-none fashion
(A) Representative single-cell traces show examples of silencing events induced by recruitment of the indicated factor. Only cells undergoing silencing during the movie are shown. Circles indicate silencing event inflection points. To avoid overlap, traces are offset by arbitrary amounts on the y-axis. Arrowheads indicate lineages highlighted in movies S2 to S4.
(B) The fraction of cells remaining ON as a function of time (upper panels) and the distributions of silencing times, T_{off} (lower panels), are plotted for each factor (mean ± s.d. over two of more independent experiments).
(C) All single-cell cumulative fluorescence traces (gray lines) were aligned at the silencing event (0 on the x-axis) and superimposed. The mean trace for each factor is plotted as a colored line.
(D) Mean reporter production rates around the silencing event, obtained by averaging the slopes of the individual traces in (C).
(E) Silencing rates, $k_{off}(t)$, were estimated by dividing the number of silencing events in each time bin after dox addition by the number of active cells at that time. Trend-lines represent phenomenological fits to the positive portion of a logistic function (see Supplementary Experimental Procedures). Dotted lines represent the asymptotic silencing rates, as determined from the fit. In D and E, colors correspond to those in A-C. Error bars denote the estimated s.d. of silencing rates, assuming the number of events in each bin follows a Poisson distribution.

silencing, with the average reporter production rates after silencing close to zero (Figure 3.2D, see also Figure B.2B). Silencing events were also abrupt relative to the timescales over which we observed gene expression. Reporter production rates typically dropped from 80% to 20% of their initial values within ∼9 to 22 hours, faster than or comparable to the cell cycle time of ∼20 hours.

While the majority of silencing events were abrupt and all-or-none, we also observed deviations from all-or-none expression in a minority of cells. For example, in ∼8% of all silencing lineages with HDAC4, the silent state was not completely off, but rather reduced to a lower basal level of gene expression (Figure B.2C), suggesting that HDAC4 could be an intrinsically weaker silencing factor. Additionally, for HDAC4, EED, and Dnmt3b, we observed events in which promoter activity transiently decreased and then recovered (Figure B.2D). These dynamics occurred in ∼14% (EED), ∼22% (Dnmt3b), or ∼6% (HDAC4) of

all lineages displaying silencing events, but were not observed for KRAB. In many of these traces, complete silencing occurred in the next cell cycle, suggesting that silencing events could require a minimum time to be fully committed or stabilized.

The simplest way to account for stochastic, all-or-none silencing events is through a one-step stochastic process, in which active cells transition to a silent state at a constant rate. To test this hypothesis, we computed the silencing rate, $k_{off}(t)$, defined as the fraction of active cells that transition to the silent state per unit time, after a time interval t since the start of recruitment. For a simple one-step process $k_{off}(t)$ should be constant over time. For all factors, we observed a transient increase of $k_{off}(t)$, followed by a longer period of constant $k_{off}(t)$ (see Figure 3.2E, where trend-lines are fits to a phenomenological function with two parameters, which reflect the maximum $k_{off}(t)$ and the sharpness of its initial increase, respectively). For HDAC4 and KRAB, $k_{off}(t)$ reached half of its maximum value rapidly ($\sim$5hrs), comparable to the time resolution of our system (4 hrs, Figure B.1D). For EED and Dnmt3b, recruitment caused a more gradual initial increase in $k_{off}(t)$ with saturating levels an order of magnitude lower than those of HDAC4 and KRAB.

The transient increase in $k_{off}(t)$ suggests that the silencing process involves intermediate steps. These steps could correspond to the stochastic gain, loss, and spreading of chromatin marks. In fact, simple models based on such dynamics can recapitulate the dynamics of silencing observed here (Figure B.2E & F, see also Supplementary Experimental Procedures). Nevertheless, for each factor, most of the silencing events occurred after $k_{off}(t)$ reached half of its maximum value. Therefore, this plateau k_{off} value represents the dominant variable characterizing the silencing dynamics. In what follows, we focus on these first-order approximations of the silencing rates.

The molecular characteristics of silent states

All four CRs analyzed here silenced gene expression fully, but did so with quite distinct dynamics. This observation poses the question of whether these final silent states have distinct epigenetic signatures or whether they represent an identical, common state. In order to answer this question, we used ChIP-qPCR to analyze modifications associated with each of the four factors after different durations of factor recruitment (Figure 3.3A). More specifically, for all factors, we analyzed H3K27me3 (associated with EED), H3K9me3 (associated with KRAB), H3 acetylation (erased by HDAC4), and CpG methylation (written by Dnmt3b) [140, 191, 196, 198]. We also analyzed H3K4me3, which is associated with active transcription [205, 206]. In each case we analyzed two regions, one within the promoter, adjacent to the factor recruitment site, and the other in the reporter gene body.

After 5 days of recruitment, gene expression was silenced in nearly all cells, by all four factors. At this time point, each factor strongly altered the level of its corresponding modification, and did so in the expected direction (diagonal elements of the matrix in Figure 3.3B). In addition, all four CRs strongly reduced active marks: H3 acetylation and H3K4me3. Finally, each factor also produced some additional silencing modifications (off-diagonal elements, Figure 3.3B). For instance, in addition to appearing after KRAB recruitment, H3K9me3 was also enriched in the promoter region following recruitment of EED and Dnmt3b. However, while the molecular states produced by individual silencing factors partially overlapped, they retained clear signatures of the recruited factor, even after 10 days of recruitment (summarized schematically in Figure 3.3C). In particular, KRAB- and Dnmt3b-mediated silencing never produced significant enrichment of H3K27me3, and DNA methylation was only enriched in the Dnmt3b line (Figures 3.3B and B.3). These results suggest that each CR produces a silent state with a distinct molecular signature.

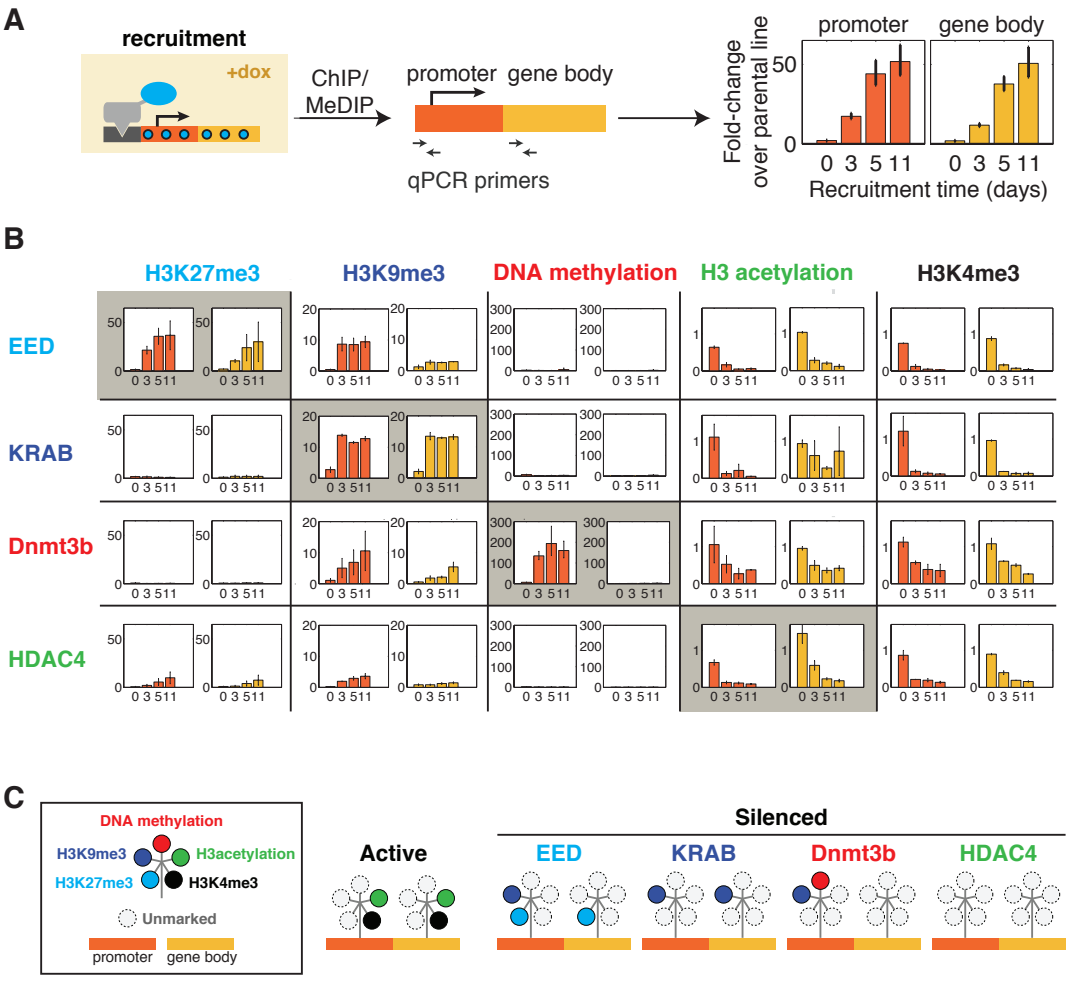


Figure 3.3: Molecular characteristics of the silent states

(A) ChIP-qPCR assays were performed on each cell line after treatment with dox for 0, 3, 5, or 11 days to assay changes in levels of five histone modifications at two positions, one in the promoter (orange) and one in the gene body (yellow). Parallel experiments were performed using MeDIP-qPCR to quantify changes in DNA methylation.

(B) qPCR signals from promoter (orange) and gene body (yellow) loci were normalized by -Actin (for active marks) or Igf2 (for repressive marks) and represented as fold-change (mean s.d. over duplicate experiments) relative to that of parental (reporter only) cell line. The diagonal set of plots highlights the changes seen in the levels of chromatin modifications expected for each factor. Off-diagonal plots represent potential cross-interactions.

(C) Summary of the dominant chromatin modifications associated with the active state and the silent states established by the different CRs. Each potential modification is indicated by the coloring of a corresponding circle at each locus. Note that HDAC4 erases, rather than writes, its modification.

Reactivation dynamics and epigenetic memory

Epigenetic memory enables gene expression states to remain stable even after removal of initiating factors. To understand how the silent states observed here differ in terms of epigenetic memory, we tracked gene expression in cells after terminating recruitment of CRs (Figure 3.1B, 'de-recruitment'). This allowed us to record how long individual cells remain silenced, and measure the gene expression dynamics if and when they reactivated.

We first analyzed the dynamics of gene expression after the end of a 5-day period of recruitment using time-lapse movies (Figure 3.1D). For Dnmt3b, the silenced state remained stable throughout the duration of the movie (Figure 3.4A). In contrast, for EED, KRAB, and HDAC4, we observed reactivation events that occurred with widely varying delays after de-recruitment. These reactivation events, like silencing, were abrupt, stochastic, and all-or-none (Figure 3.4A).

Nevertheless, a fraction of cells did not reactivate by the end of the movies, at which point cell densities became too high for accurate tracking. Therefore, we used flow cytometry to extend the analysis of reactivation dynamics to longer times (up to $\sim$1 month). Fluorescence distributions measured by flow cytometry were bimodal, and the fluorescence levels of reactivated cells were similar to cells that were never silenced (Figures 3.4B, B.4A-C), suggesting that reactivation remained an all-or-none process even after the end of time-lapse imaging. This bimodality permitted quantitative analysis of the fraction of cells in the silent state as a function of time (Figure 3.4C). These data revealed that each CR produced a qualitatively different mode of epigenetic memory. HDAC4 imparted short-term memory: upon its removal, silencing was lost in all cells within five days. In contrast, Dnmt3b produced stable silencing over timescales of $>$1 month. Finally, both EED and KRAB showed an unexpected hybrid behavior: a fraction of cells fully reactivated within $\sim$2 weeks, while

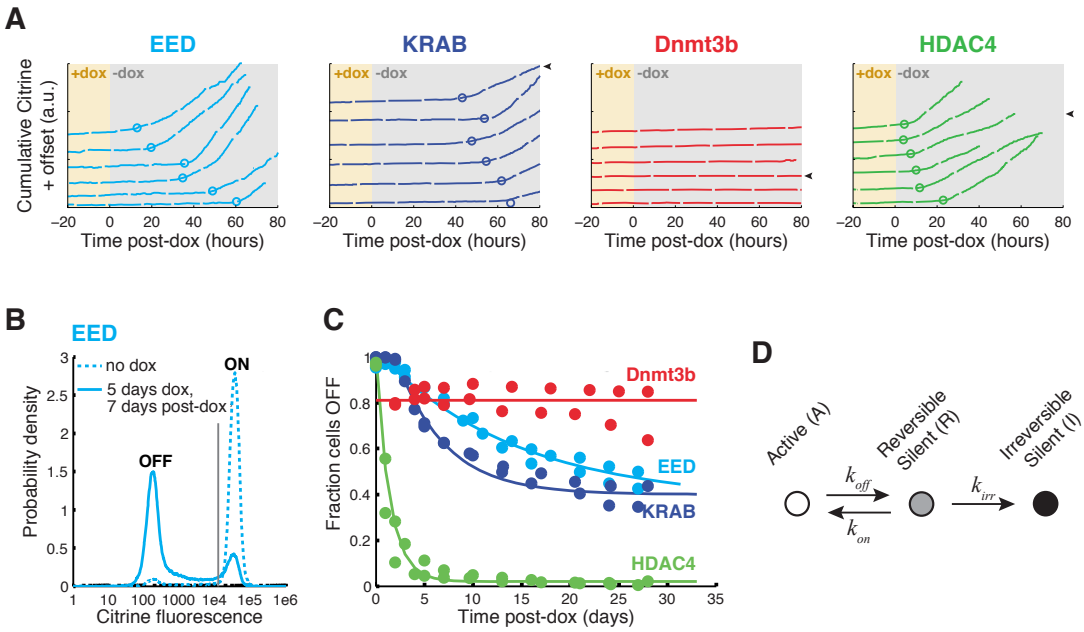


Figure 3.4: Dynamics of reactivation and epigenetic memory

(A) Cumulative single-cell fluorescence traces show individual reactivation events (circles). For EED, KRAB and HDAC4, only traces that reactivated during the movie are shown, but many cells did not reactivate over the duration of the movie. For Dnmt3b, no reactivation event was observed, and only silent cells are shown. Traces are vertically offset for clarity. Arrowheads indicate lineages highlighted in movies S6 to S8.

(B) Example of flow cytometry data showing the fluorescent distributions for EED-expressing cells treated with dox for 5 days (cyan solid, measured 7 days after dox removal), or untreated (cyan dashed). Cells were classified as OFF (low fluorescence peak), or ON (high fluorescence peak) with a threshold (grey line).

(C) The fraction of cells in the OFF state was measured at varying time points after the end of recruitment. Each dot represents one flow cytometry measurement of the type shown in (B). Spontaneous background silencing rates have been subtracted (Figure B.4A, Experimental Procedures). Solid lines are fits to the model described in the text and panel D.

(D) The progressive silencing model involves stochastic transitions between three functionally distinct epigenetic states, involving three first order rate constants, as indicated.

the remaining fraction remained completely silenced for at least a month.

The hybrid behavior of the EED and KRAB silenced cells suggests that cells can occupy at least two functionally distinct silent states, characterized by very different rates of reactivation. In this model, recruitment of a silencing CR causes cells to stochastically advance from the active state (A) to a reversibly silent state (R), and then to an irreversibly silent state (I) (Figure 3.4D). These dynamics can be summarized as $A \leftrightarrow R \rightarrow I$, with both forward silencing rates (k_{off} for $A \rightarrow R$, and k_{irr} for $R \rightarrow I$) dependent on continued recruitment of the silencing factor. After the end of recruitment, the forward silencing rates become negligible, allowing cells from the R state to reactivate ($A \leftarrow R$, rate k_{on}). Defining $C_{OFF}(t, \tau)$ as the fraction of cells silenced at time t after the end of a recruitment period of duration τ, we have

$$C_{OFF}(t, \tau) = C_R(\tau)e^{-k_{on}t} + C_I(\tau),$$

where $C_R(\tau)$ and $C_I(\tau)$ are the fraction of cells in the R and I states, respectively, at the end of recruitment.

One prediction of this model is that longer durations of recruitment should increase the fraction of irreversibly silenced cells. To test this hypothesis, we systematically varied the duration of recruitment and analyzed the subsequent reactivation dynamics (Figure 3.5A). For both EED and KRAB, the fraction of cells remaining silent after 30 days of de-recruitment increased with the duration of the recruitment phase, as shown by the progression from green to red data points in Figure 3.5B & C. We noted a small time lag between the end of recruitment and the onset of the reactivation phase, especially for durations of recruitment over 3 days. One possible mechanism to explain this lag is progressive

spreading of nucleosome modifications during continuous recruitment [98]. In this scenario, longer silencing times would result in more modifications, which could require more time to reduce sufficiently for reactivation. Remarkably, aside from this relatively small lag (1-2 days), once reactivation started, all data for a given factor could be fit with a single reactivation rate, k_{on}, across a range of different recruitment durations (continuous curves in Figure 3.5B-E).

While the rate constants thus appeared constant for each individual factor, they varied substantially across factors (Figure 3.5B-E). For example, even though KRAB and EED memory modes were qualitatively similar, they still showed significantly different reactivation rates, with KRAB reactivating about three times faster than EED ($k_{on} = 0.25d^{-1}$ versus $k_{on} = 0.07d^{-1}$). Moreover, HDAC4 and Dnmt3b could be fit by simplified forms of this model. For HDAC4, the majority of cells were reversibly silenced for recruitment times varying between 1 and 5 days ($C_I = 0$), making the irreversible state unnecessary ($A \leftrightarrow R$ only). In contrast, for Dnmt3b, once silenced, cells remained irreversibly off for at least a month ($C_R = 0$), suggesting that Dnmt3b-mediated silencing effectively bypasses the reversible state ($A \rightarrow I$ only). Taken together, these results provide a simple quantitative framework capable of describing a wide range of behaviors of epigenetic memory and reactivation dynamics across different CRs.

3.4 Discussion

Silencing and reactivation events are stochastic, abrupt and all-or-none

Epigenetic regulation has been studied extensively in diverse systems, but the dynamic process through which a single CR changes the expression level of a target gene within an

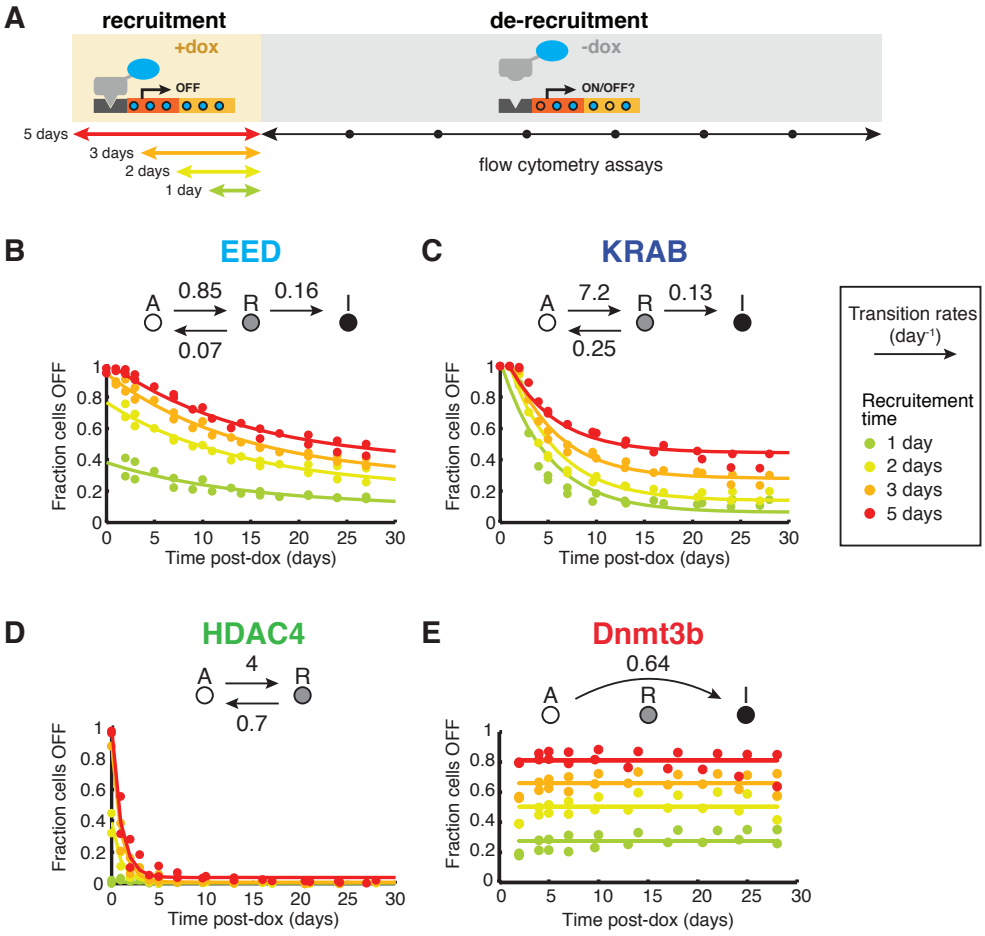


Figure 3.5: Each regulator implements distinct transition rates, independent of recruitment time

(A) Schematic of flow cytometry experiments following varying durations of recruitment (1 to 5 days, indicated by colored arrows). Cells were analyzed at multiple time-points between 0 and 30 days after de-recruitment (black circles).

(B-E) The fraction of cells in the OFF state is plotted as a function of time since the end of recruitment for each factor. Colors indicate the duration of recruitment, as in (A). Dots represent data from two independent experiments, and the solid lines represent fits to the models indicated schematically above each panel. For each factor, all data are fit with a single set of rate constants. Transition rates between active (A), reversibly silent (R) and irreversibly silent (I) states are in units of day^{-1}. A to R silencing rates (k_{off}) represent the asymptotic values measured from movies (Figure 3.2E), R to A reactivation rates (k_{on}) are extracted by fitting the data in panels B-E with exponential decays, and R to I rates (k_{irr}) are extracted from fitting the fraction of irreversibly silenced cells as a function of time with the full model shown above each plot.

individual cell has remained unclear. By combining an inducible recruitment platform with quantitative time-lapse microscopy, we were able to study silencing and reactivation events dynamically in single cells.

Silencing occurred through abrupt, all-or-none stochastic events whose timing varied among individual cells. Similarly, reactivation events observed after the termination of CR recruitment were also abrupt, all-or-none, and stochastic, resembling silencing events in reverse. These observations are consistent with the bimodal distributions of gene expression previously observed in response to HP1 recruitment [98]. The fact that CRs with very different molecular mechanisms and dynamic behaviors produced qualitatively similar silencing and reactivation events in our studies suggests that these are likely to be general features of epigenetic regulation. Extending this method to some of the many other known CRs, such as the 223 different CRs recently studied in yeast [186], could help to establish how general the binary chromatin response is in other cellular contexts.

Functionally, abrupt, all-or-none silencing and reactivation events could facilitate sharp transitions between discrete cell fates. In fact, such transitions, with accompanying epigenetic changes, have been shown to be sudden and/or stochastic in T cell activation [207, 208], ES cell differentiation [209], and the response of fibroblasts to stimulation by tumor-necrosis factor (TNF)-α [190]. Similarly, recent single-cell analysis in mouse ES cells revealed abrupt stochastic transitions between two states with different gene expression and DNA methylation profiles [100]. An abrupt, all-or-none switch in the expression of a key regulatory protein, for instance, could help to stochastically trigger a differentiation event, while avoiding potentially ambiguous intermediate gene expression levels of its targets.

Connecting silencing and reactivation events to epigenetic mark dynamics

In this study, we focused on characterizing the dynamics of silencing and reactivation events. Nevertheless, our results raise fundamental questions about the molecular mechanism underlying these events. In particular, what gives rise to the all-or-none, stochastic and abrupt nature of silencing and reactivation? Using the prevailing model of chromatin spreading, we can reproduce the stochastic nature of silencing in individual cells and the progression of silencing rates over time (Figures 3.2E, B.2E & F). However, it remains unclear what generates the abruptness of silencing and reactivation events. In one scenario, the chromatin spreading model can be extended to include cooperative propagation of the marks [210]. Under this regime, the domain size of chromatin marks can change suddenly, leading to abrupt gene regulation events. Recently developed single-cell techniques that detect histone modification at a specific locus [211] could be combined with the time-lapse approaches used here to map the extent of a particular histone mark before and immediately after silencing. The resulting distribution of histone domain sizes could reveal whether spreading is gradual or bimodal and whether a sharp domain size threshold is required to initiate silencing. Alternatively, sudden changes in gene expression could result from larger scale processes such as chromatin compaction [212] or movement to the nuclear lamina [213].

Operational, dynamic perspective on epigenetic states and memory

The effects of distinct CRs have previously been studied at two distinct levels. At the molecular level, each CR recruitment can produce a characteristic set of molecular modifications [214–217]. At the gene expression level, CR recruitment can lead to varying degrees of gene repression or activation in population average studies [179, 181, 183, 186]. The results obtained here suggest a complementary view, in which epigenetic states can be distinguished

both by the level of gene expression they produce (i.e on or off), and by their propensity for switching to other epigenetic states (i.e. reversible or irreversible) (Figure 3.4D).

Within this framework, the behavior of each of the four CRs can be quantitatively described with just three, and sometimes only two, states. Cells transition stochastically between active, reversibly silenced, and irreversibly silenced epigenetic states (Figure 3.4D). This model provides a coarse-grained view of the dynamics over the timescales of days to weeks investigated here. We did observe some evidence for possible additional sub-states not explicitly incorporated in the model (e.g. the delays before the onset of silencing and reactivation). Nevertheless, the 3-state model describes the major dynamic behavior of these diverse systems over the most relevant timescales, making it a useful framework for characterizing and designing CR regulatory systems.

Critically, while this model can describe much of the dynamics generated from all four factors, the states themselves differ at both functional and molecular levels depending on the CR used. Functionally, each CR generated silenced states with distinct propensities for reactivation or irreversible silencing. These transition rates explain the type of epigenetic memory achievable by each factor: permanent (Dnmt3b), transient (HDAC4), or a hybrid of the two (EED and KRAB) (Figure 3.6A). DNA methylation and histone deacetylation were previously suggested to enable permanent and transient regulation, respectively [24, 26]. The hybrid memory observed for EED and KRAB represents a distinct mode of regulation, in which all three states come into play, providing two timescales for epigenetic memory from the same factor. The distinct set of transition rates, k_{off}, k_{irr}, and k_{on}, encapsulates the functionality offered by each factor (Figure 3.6B).

Molecularly, each CR produced a distinct chromatin modification signature (Figure 3.3). For example, even after extended periods of EED and KRAB recruitment sufficient to ir-

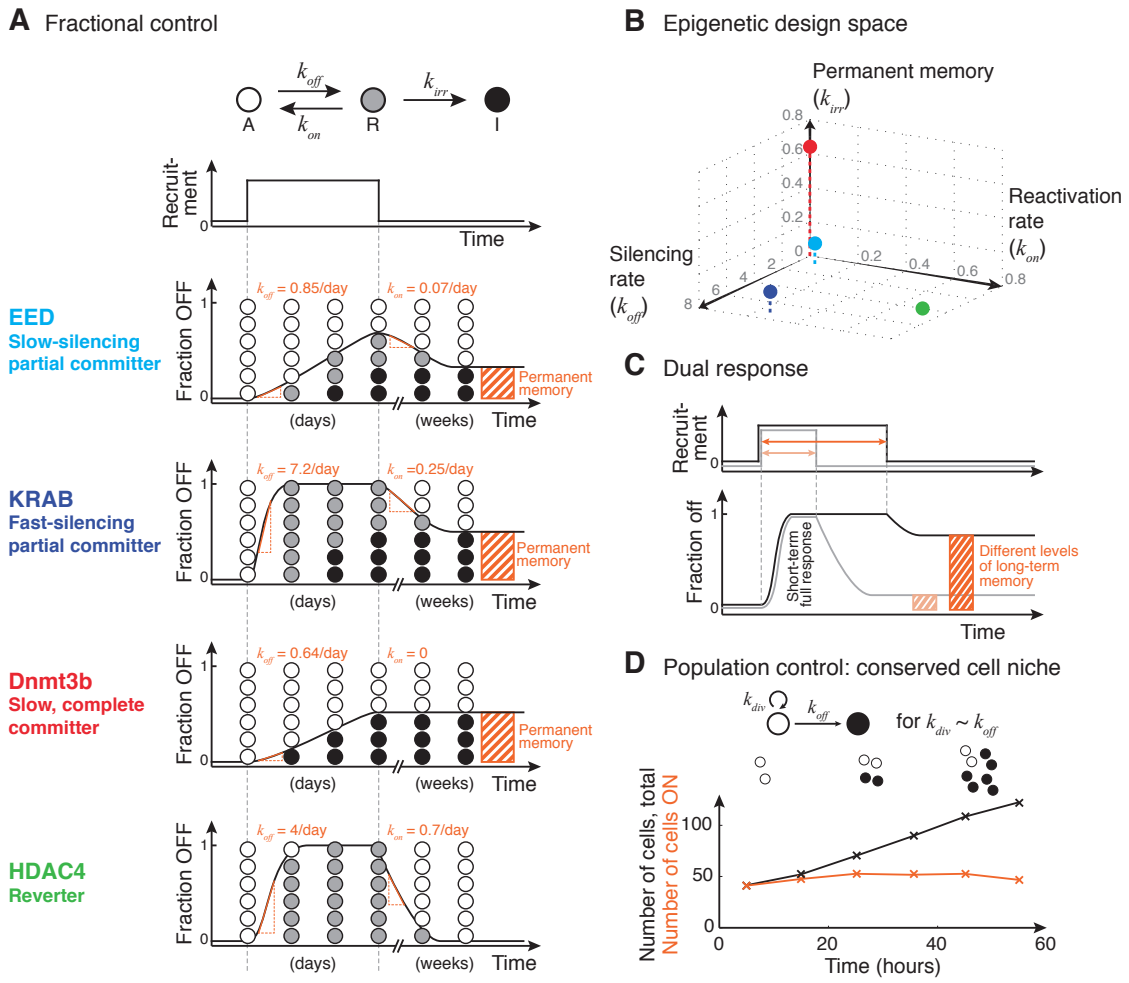


Figure 3.6: Operational framework and design space of epigenetic regulation
(A) The rates of stochastic switching between three functional states (top panel) determine the dynamic response of different chromatin regulators (bottom four panels) to an interval of chromatin recruitment (schematic). At the population level, the response of gene expression to factor recruitment is fractional. Individual cells within a population (circle stacks) are either fully active (white), reversibly silenced (gray), or irreversibly silenced (black) depending on the rates associated with each regulator and the duration of recruitment or de-recruitment.
(B) Each CR occupies a distinct operating point within the larger potential design space defined by the three transition rates between epigenetic states.
(C) The hybrid form of memory observed for histone methylases enables a dual response to signals. Increasing signal durations (top), while resulting in the same, full short-term population response, lead to increasing levels of long-term memory manifested as fractions of irreversibly silenced cells (bottom).
(D) When the cell division and silencing rates are similar ($k_{div} \approx k_{off}$), the number of cells remaining ON can be held constant (open circles, orange line), while the total number of cells grows exponentially (open + closed circles, black line). This regime occurred under continuous EED recruitment. X's indicate data points, averaged over three independent experiments.

reversibly silence substantial fractions of cells, these factors did not directly or indirectly cause DNA methylation (Figures 3.3, B.3). This contrasts with the observation that DNA methylation is necessary for persistent epigenetic memory generated by HP1 recruitment [98]. Conversely, Dnmt3b-dependent methylation did not lead indirectly to H3K27 methylation. These results suggest that complete silencing can be generated through distinct molecular chromatin states.

Epigenetic regulation enables fractional control at the population level

In stochastic fractional control systems, the magnitude and/or duration of input signals controls the fraction of cells that respond, with each individual cell responding in an all-or-none manner [55]. For example, in the immune system, the fraction of T cells that differentiate into effector cells appears to depend quantitatively on the duration of stimulation (reviewed in [218]). Similarly, recent work has suggested that adipocyte differentiation can also occur stochastically in a small fraction of cells [219]. Finally, in plant vernalization, the probability of flowering changes with the durations of exposure to cold, through a mechanism that involves progressive fractional silencing by PRC2 [172, 173].

Despite large quantitative differences in silencing and reactivation rate constants, as well as qualitative differences in the types of chromatin marks they write, all CRs studied here effectively implemented stochastic fractional control of gene expression, and did so in several distinct ways: First, during continuous recruitment of a silencing factor, the fraction of silenced cells (I and R states) grows monotonically with time. In this mode, the steepness of the input-output response is determined by the value of the silencing rate, k_{off} (Figure 3.6A, all factors). Because k_{off} differs among silencing factors, the choice of CR determines the sensitivity of the silenced fraction to the duration of the input signal. Second, for factors

that allow full reactivation, the fraction of cells that remain silent at a given time reflects the time elapsed since the end of recruitment (Figure 3.6A, HDAC4). Third, for factors that generate permanent memory, the fraction of cells that remain irreversibly silenced encodes the total duration of recruitment (Figure 3.6A, Dnmt3b, EED, and KRAB).

Furthermore, the hybrid memory observed for EED and KRAB separates the timescales associated with response and memory. For example, if KRAB recruitment is maintained for timescales ranging from 2-5 days, all cells in the population can be silenced, but the fraction of irreversibly silenced cells varies (Figure 3.6C). This enables a cell population to generate an identical initial system-wide response to a signal while encoding recruitment duration in the long-term memory.

Finally, stochastic binary gene regulation also enables a system to fix the absolute number of cells in a particular state, by matching rates of cell division and stochastic silencing. In this scenario, each cell division will result, on average, in one silenced and one active daughter. Since the silent state is heritable, its population will grow exponentially, while the total number of active cells remains constant. In fact, we observed this behavior with EED recruitment under the conditions investigated here (Figure 3.6D). This mechanism provides a very simple way to maintain a fixed population size, a feature of stem cell niches that balances self-renewal and differentiation [220]. However, exactly matching the silencing and division rates requires 'fine-tuning' [221]. Perhaps for this reason, naturally occurring systems appear to involve more complex mechanisms such as feedback loops and cell-to-cell signaling [222, 223]. It remains to be seen how these different population control strategies compare at the systems level.

Exploring the epigenetic regulation design space

Our approach here has been to take a design-oriented view of chromatin regulation, centered on understanding the quantitatively and qualitatively distinct gene regulation capabilities provided by each CR. For the four factors analyzed here, these capabilities can be captured by three key parameters: the rates of silencing and commitment to the irreversibly silent state in the presence of the CR, k_{off} and k_{irr}, respectively, and the rate of reactivation in the absence of the CR, k_{on}. Each of the four factors tested occupies a distinct location within this three-dimensional design space (Figure 3.6B), with a corresponding behavior (Figure 3.6A). Dnmt3b is a slow, but complete, committer: it has a low, stochastic silencing rate, but irreversibly commits cells to a silent state. HDAC4 is a reverter: it displays fast silencing with a relatively short memory. In contrast, EED and KRAB, both *writers* of histone methylation marks, are partial committers: they silence on one timescale, and commit on another, producing a mixture of reversibly and irreversibly silenced cells at intermediate timescales. While EED and KRAB both show similar transition rates to the permanently silenced state ($R \rightarrow I$), they induce the initial silencing step ($A \rightarrow R$) at quite different rates, thus occupying different points in the design space (Figure 3.6C). Because of these different characteristics each of these specific CRs could provide distinct capabilities for epigenetic applications in synthetic biology.

More generally, the CRs investigated here exist within a larger design space of potential regulatory modes. Other CRs not yet analyzed could differ from these quantitatively, in the values of their dynamic rate constants (Figure 3.6B), or qualitatively, for example by generating additional, functionally distinct epigenetic states. It will therefore be interesting to determine where other CRs fit within this space and how their behaviors depend on genomic context, cell type, and promoter architecture. Recent advances in targeted recruit-

ment techniques, such as CRISPRs and TALs, provide further opportunities to expand the analysis to additional genomic loci and to explore the effects of combinatorial recruitment of multiple CRs [181, 185]. Understanding the design space of chromatin regulation will both produce a better understanding of why specific CRs are employed in natural genetic circuits and enable the design of synthetic gene circuits that take advantage of the inherent temporal control and memory capabilities provided by epigenetics.

3.5 Materials and Methods

Cell lines construction

The reporter line was created by co-transfecting CHO-K1 cells carrying the MI-HAC artificial chromosome [203] with 600 ng PhiC31-Neo-ins-5xTetO-pEF-H2B-Citrine-ins reporter plasmid and 200 ng PhiC31 integrase plasmid. Transfection was done using Lipofectamine 2000 (Invitrogen). Cells were transferred to 6-well plates 24 hours later and selected with 400 ng/ul geneticin for 12 days, starting 40 hours after the transfection. Single clones were obtained by limiting dilution. The integration of the reporter in the HAC was verified by genomic PCR, and a single clone was chosen for further analysis. Each of the CR plasmids (pEF-H2B-mCherry-T2A-rTet-CR) was randomly integrated into this reporter line by transfection with Lipofectamine 2000 (Invitrogen). These cells were selected using 300 ng/ul zeocin starting 24 hours after transfection for a total of 12 days. Finally, single clones were selected for each CR by limiting dilution. See Supplemental Experimental Procedures for details on plasmids construction and culture conditions.

Fluorescence time-lapse movies

Reporter cells expressing each of the four CRs were plated around 12 hours before imaging, at low density (1,500 cells/cm^2), on glass-bottom plates (MatTek) coated with 5g/ml hamster fibronectin (Oxford Biomedical Research). Imaging was done using an inverted Olympus IX81 with Zero Drift Control (ZDC), a 20X objective, and an iKon-M CCD camera (Andor, Belfast, NIR). Fluorophores were excited using an X-Cite XLED1 light source (Lumen Dynamics). Images were automatically acquired every 20 minutes, using Metamorph software (Molecular Devices). The microscope was encased in a chamber kept at 37C and 5% CO2, and the imaging growth media (see Supplemental Experimental Procedures) was changed daily. Silencing movies began with reporter cells actively expressing the reporter gene. Dox (1g/ml) was added to the cells at ∼24 hours, after which imaging continued for at least 3 more days and until cell tracking became difficult due to high cell density. Cells were then replated similarly and at low density, in the presence of dox, for the subsequent acquisition of reactivation movies. Imaging began with these cells ∼12 hours after re-plating, and dox was washed-out at ∼24 hours into the movies (5 days since the beginning of dox addition). See Supplemental Experimental Procedures for details on movie analysis and event detection.

ChIP-qPCR and MeDIP-qPCR

Each cell line was treated with dox (1ug/ml) for 0, 3, 5, and 11 days before harvesting. ChIP and MeDIP were performed using LowCell# ChIP and MagMeDIP kits, respectively, with the Bioruptor sonicator (all from Diagenode). For ChIP, we used the following antibodies: anti-H3K27me3 (Milipore, 07-449), anti-H3K9me3 (Abcam, ab8898), anti-acetyl-H3 (Millipore, 06-599), anti-H3K4me3 (Abcam, ab8580). For MeDIP, we used the 5-methylcytidine

antibody from the MagMeDIP kit (Diagenode). qPCR was performed using SsoFast EvaGreen Supermix on a CFX96-C1000 Real-Time PCR System (both from Bio-Rad Laboratories). For qPCR primer sequences, see Table B.1. Reported fold-enrichment values from qPCR experiments were obtained by normalizing first against an internal positive control locus, and then against the parental cell line without any CR, i.e.,

$$fold-change = \frac{2^{\{C_t(control,CF)-C_t(locus,CF\}}}{2^{\{C_t(control,parental)-C_t(locus,parental\}}}.$$

The internal control loci account for variations in the amount of DNA and pull-down efficiency for each sample, and are chosen to be actin for the marks associated with active genes (H3K4me3) and Igf2 for the silencing marks (H3K27me3, H3K9me3, and 5mC). Igf2 was chosen for its lack of expression in mouse adult ovary cells (MGI Ref. ID J:46439) and elevated levels of H3K9me3 and H3K27me3 implicated in the imprinting of the locus [224].

Flow cytometry for epigenetic memory analysis

For each cell line, cells were plated in multiple wells at the same time, and treated with dox (1ug/ml) starting at different time intervals. Dox was removed simultaneously from all samples. At different time points following dox removal, cells were harvested using 0.25% Trypsin (Life Technologies). A fraction of the cells were re-plated for the next time point. The rest of the cells were resuspended in flow buffer (Hank's Balanced Salt Solution (Life Technology) and 2.5 mg/ml BSA), filtered through 40 m strainers (BD Falcon) and their fluorescent distributions were measured with a MACSQuant VYB machine (Miltenyi Biotec, Bergisch Gladbach, Germany). The resulting data were analyzed with a custom Matlab program called EasyFlow. Single cells were selected based on side and forward scatter properties. A manual gate was imposed on the Citrine fluorescence to determine

the percent of cells in the OFF state for each sample. The gate was selected to contain 1-2% of the positive Citrine peak of untreated cells. In all cell lines containing CRs, we noticed an increase in the percentage of silent cells over 30 days even in the absence of dox treatment (Figure B.4D). This background silencing is a combination of spontaneous silencing of the reporter locus (as seen in the parental line with no CRs, Figure B.4D), and non-specific recruitment of rTetR to the reporter. In order to correct for this background silencing, for each cell line, we subtracted the fraction of cells silenced in the untreated line and normalized by the fraction of untreated cells that were active at each time point, i.e. OFF=(OFF(treated)-OFF(untreated))/ON(untreated).

Concluding Remarks

In this thesis we examined the single-cell dynamics of gene expression, either during spontaneous cellular state transitions in ES cells or under artificial recruitment of CRs. In both contexts, individual cells transition between distinct states in stochastic events, without persisting in intermediate gene expression levels. These similarities provoke the question of what role chromatin regulation may have on state transitions in ES cells. More generally, the diverse timescales of silencing and reactivation, and types of epigenetic memory elicited by different CRs prompt us to wonder how these properties may be specifically utilized in different contexts during the development. Here we discuss these issues and suggest how the experimental and analytical framework presented in this thesis may be extended to address them.

In Chapter 2, we reported the correlation of the two ES cell states with differential expression of Dnmt3b and Tet1, as well as changes in the global level of DNA methylation over many genes. Furthermore, we demonstrated that DNA methylation is partially required for the establishment and maintenance of the Nanog/Rex1-low state. These results suggest a major role of DNA methylation in the regulation of the metastability in undifferentiated ES cells. Nevertheless, transition between these two states likely involves many other mechanisms, such as cell-cell signaling and transcription feedback. What is the hierarchy among these regulatory processes? In one potential scenario, state transition may be initiated by

fluctuations in the activities of various signaling pathways, such as Fgf and Wnt. These pathways can impinge on the expression of Nanog through post-transcriptional and post-translational mechanisms [71, 225–228]. Changes in Nanog expression can then propagate to downstream regulators through the transcription factor circuit, and finally results in the establishment of more stable states through changes in chromatin modifications. In an alternative scenario, spontaneous fluctuations in chromatin regulator expression may initiate changes in chromatin modifications in core regulators and their expression levels, which are then propagated to the rest of the transcription circuit.

To distinguish between these scenarios, we can follow the dynamics of multiple genes in the same cell during state transitions, and determine if these genes switch their expressions under a fixed sequence. This can be done by constructing multi-fluorescence reporter cell lines. Furthermore, we can identify a core set of genes whose changes in expression level are sufficient to trigger state transition. This can be achieved by transient over-expression or knock-down of one or more candidate genes, while monitoring the status of a cell state reporter such as Nanog or Rex1. One candidate gene of particular interest is Tet1, which we showed to be more highly expressed and less methylated in the Nanog/Rex1-high state, and is itself involved in the process of DNA demethylation through its 5-methylcytosine hydroxylase activity [135]. Interestingly, Tet1 and Tet2 were recently shown to associate with Nanog through protein-protein interaction, and this interaction is implicated in the process of reprogramming through fusing somatic cells with embryonic stem cells [229]. Furthermore, simultaneous ectopic expression of Nanog and Tet1 leads to increased levels of 5-hydroxymethylcytosine on their common target loci, including Esrrb and Oct4 [229]. Finally, Tet1 is also known to bind to its own promoter as well as that of Nanog [136]. Together, these results suggest a positive feedback loop that involves the chromatin reg-

ulatory activity of Tet1. It will be interesting to see if ectopic expression of Tet1 and/or targeted recruitment of Tet1 to the promoter of itself and other target genes is sufficient to trigger transition from Nanog/Rex1-low to Nanog/Rex1-high states. With the advent of CRISPR-mediated RNA-guided targeting system [185], our chromatin recruitment system in Chapter 3 can be conveniently adapted to achieve the latter goal.

In Chapter 3, we investigated the response of target gene expression to direct recruitment and de-recruitment of four different CRs. Remarkably, all four CRs demonstrated similar all-or-none event profiles during silencing and reactivation, while displaying a diverse range of silencing and reactivation rates, and types of epigenetic memory. How may these differences be utilized in developmental contexts? During early embryonic development, when cells in the epiblast differentiate into cells in the three germ layers, pluripotency-associated genes are down-regulated, while lineage-associated genes begin to express. In particular, genes that are associated with self-renewal and can induce dedifferentiation, such as Oct4, needs to be permanently silenced and safeguarded against re-expression in somatic cells. Recent research in ES cell differentiation, in response to retinoic acid, showed that this safeguard may be achieved with a multi-step mechanism [230]. This process involves a series of regulations in the following sequence: transcriptional repression, G9a-dependent methylation of H3K9, removal of H3K4 methylation and histone acetylations, and, finally, Dnmt3a/3b-dependent methylation of DNA. Interestingly, our results demonstrated that KRAB-mediated silencing (associated with H3K9 methylation) is rapid yet only partially permanent, while Dnmt3b-mediated silencing (associated with DNA methylation) is permanent but slow. Combining different chromatin modifications and regulators may therefore represent a unique strategy for rapid and permanent gene silencing.

It should be noted that our chromatin regulation reporter consists of a EF1α pro-

moter, which has two adjacent CpG islands, while there is no CpG island at the Oct4 promoter. CpG islands in mammalian genes are typically unmethylated, with the exception of imprinted genes and on the inactive X chromosome. Our data demonstrated that EF1α promoter can be methylated and permanently silenced by Dnmt3b recruitment, and suggests that CpG island methylation is sufficient to provide the stable silencing that is observed in imprinting and X chromosome inactivation. It will be interesting to see if the stability of silencing varies with CpG densities at the promoter. Furthermore, since we have demonstrated that ES cells are subject to fluctuating DNA methylation and demethylation activities, it will also be interesting to test Dnmt3b recruitment to endogenous genes in the ES cell context.

Together, the experimental and analytical platforms presented in this thesis provide novel tools for investigating the natural gene expression dynamics in more biological contexts. Finally, insights from the results presented here and from future studies using these platforms will enable us to design and construct novel synthetic systems with predictable cellular state dynamics.

Bibliography

[1] Bianconi, E., Piovesan, A., et al. "An estimation of the number of cells in the human body." *Annals of Human Biology* 40.6 (2013), pp. 463–471.

[2] Vickaryous, M. K. and Hall, B. K. "Human cell type diversity, evolution, development, and classification with special reference to cells derived from the neural crest." *Biological Reviews of the Cambridge Philosophical Society* 81.3 (2006), pp. 425–455.

[3] Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., and Walter, P. *Molecular Biology of the Cell*. Garland Science, 2014.

[4] Velculescu, V. E., Zhang, L., Vogelstein, B., and Kinzler, K. W. "Serial analysis of gene expression." *Science* 270.5235 (1995), pp. 484–487.

[5] Strausberg, R. L., Feingold, E. A., et al. "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences." *Proceedings of the National Academy of Sciences* 99.26 (2002), pp. 16899–16903.

[6] Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. "Mapping and quantifying mammalian transcriptomes by RNA-Seq." *Nature Methods* 5.7 (2008), pp. 621–628.

[7] Eldar, A. and Elowitz, M. B. "Functional roles for noise in genetic circuits." *Nature* 467.7312 (2010), pp. 167–173.

[8] Sigal, A., Milo, R., et al. "Variability and memory of protein levels in human cells". *Nature* 444.7119 (2006), pp. 643–646.

[9] Blake, W. J., KAErn, M., Cantor, C. R., and Collins, J. J. "Noise in eukaryotic gene expression." *Nature* 422.6932 (2003), pp. 633–637.

[10] Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L., and Leibler, S. "Bacterial persistence as a phenotypic switch." *Science* 305.5690 (2004), pp. 1622–1625.

[11] Kussell, E. and Leibler, S. "Phenotypic diversity, population growth, and information in fluctuating environments." *Science* 309.5743 (2005), pp. 2075–2078.

[12] Losick, R. and Desplan, C. "Stochasticity and cell fate." *Science* 320.5872 (2008), pp. 65–68.

[13] Süel, G. M., Garcia-Ojalvo, J., Liberman, L. M., and Elowitz, M. B. "An excitable gene regulatory circuit induces transient cellular differentiation." *Nature* 440.7083 (2006), pp. 545–550.

[14] Hamoen, L. W., Venema, G., and Kuipers, O. P. "Controlling competence in Bacillus subtilis: shared use of regulators." *Microbiology* 149.Pt 1 (2003), pp. 9–17.

[15] Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y., and Tyagi, S. "Stochastic mRNA Synthesis in Mammalian Cells". *PLoS Biology* 4.10 (2006), e309.

[16] KAErn, M., Elston, T. C., Blake, W. J., and Collins, J. J. "Stochasticity in gene expression: from theories to phenotypes." *Nature Reviews Genetics* 6.6 (2005), pp. 451–464.

[17] Kouzarides, T. "Chromatin modifications and their function." *Cell* 128.4 (2007), pp. 693–705.

[18] Stein, R., Razin, A., and Cedar, H. "In vitro methylation of the hamster adenine phosphoribosyltransferase gene inhibits its expression in mouse L cells." *Proceedings of the National Academy of Sciences* 79.11 (1982), pp. 3418–3422.

[19] Harbers, K., Schnieke, A., Stuhlmann, H., Jähner, D., and Jaenisch, R. "DNA methylation and gene expression: endogenous retroviral genome becomes infectious after molecular cloning." *Proceedings of the National Academy of Sciences* 78.12 (1981), pp. 7609–7613.

[20] Bird, A., Taggart, M., Frommer, M., Miller, O. J., and Macleod, D. "A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA." *Cell* 40.1 (1985), pp. 91–99.

[21] Lall, S. "Primers on chromatin." *Nature Structural & Molecular Biology* 14.11 (2007), pp. 1110–1115.

[22] Li, E. and Zhang, Y. "DNA methylation in mammals." *Cold Spring Harbor Perspectives in Biology* 6.5 (2014), a019133.

[23] Li, B., Carey, M., and Workman, J. L. "The role of chromatin during transcription." *Cell* 128.4 (2007), pp. 707–719.

[24] Bird, A. "DNA methylation patterns and epigenetic memory." *Genes & Development* 16.1 (2002), pp. 6–21.

[25] Jones, P. A. "Functions of DNA methylation: islands, start sites, gene bodies and beyond." *Nature Reviews Genetics* 13.7 (2012), pp. 484–492.

[26] Katan-Khaykovich, Y. and Struhl, K. "Dynamics of global histone acetylation and deacetylation in vivo: rapid restoration of normal histone acetylation status upon removal of activators and repressors." *Genes & Development* 16.6 (2002), pp. 743–752.

[27] Barth, T. K. and Imhof, A. "Fast signals and slow marks: the dynamics of histone modifications." *Trends in Biochemical Sciences* 35.11 (2010), pp. 618–626.

[28] Ptashne, M. "On the use of the word 'epigenetic'." *Current Biology : CB* 17.7 (2007), R233–6.

[29] Bird, A. "Perceptions of epigenetics." *Nature* 447.7143 (2007), pp. 396–398.

[30] Roadmap Epigenomics Consortium, Kundaje, A., et al. "Integrative analysis of 111 reference human epigenomes." *Nature* 518.7539 (2015), pp. 317–330.

[31] Lee, T. I. and Young, R. A. "Transcription of eukaryotic protein-coding genes." *Annual Review of Genetics* 34 (2000), pp. 77–137.

[32] Sprinzak, D. and Elowitz, M. B. "Reconstruction of genetic circuits." *Nature* 438.7067 (2005), pp. 443–448.

[33] Weber, W. and Fussenegger, M. "Synthetic gene networks in mammalian cells." *Current Opinion in Biotechnology* 21.5 (2010), pp. 690–696.

[34] Alon, U. *An Introduction to Systems Biology.* Design Principles of Biological Circuits. CRC Press, 2006.

[35] Elowitz, M. B. and Leibler, S. "A synthetic oscillatory network of transcriptional regulators". *Nature* 403.6767 (2000), pp. 335–338.

[36] Becskei, A., Séraphin, B., and Serrano, L. "Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion." *The EMBO Journal* 20.10 (2001), pp. 2528–2535.

[37] Becskei, A. and Serrano, L. "Engineering stability in gene networks by autoregulation." *Nature* 405.6786 (2000), pp. 590–593.

[38] Gardner, T. S., Cantor, C. R., and Collins, J. J. "Construction of a genetic toggle switch in Escherichia coli." *Nature* 403.6767 (2000), pp. 339–342.

[39] Alon, U. "Network motifs: theory and experimental approaches." *Nature Publishing Group* 8.6 (2007), pp. 450–461.

[40] Elowitz, M. B., Levine, A. J., Siggia, E. D., and Swain, P. S. "Stochastic gene expression in a single cell." *Science* 297.5584 (2002), pp. 1183–1186.

[41] Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D., and Oudenaarden, A. van. "Regulation of noise in the expression of a single gene." *Nature Genetics* 31.1 (2002), pp. 69–73.

[42] Raser, J. M. and O'Shea, E. K. "Control of stochasticity in eukaryotic gene expression." *Science* 304.5678 (2004), pp. 1811–1814.

[43] Süel, G. M., Kulkarni, R. P., Dworkin, J., Garcia-Ojalvo, J., and Elowitz, M. B. "Tunability and noise dependence in differentiation dynamics." *Science* 315.5819 (2007), pp. 1716–1719.

[44] Choi, P. J., Cai, L., Frieda, K., and Xie, X. S. "A stochastic single-molecule event triggers phenotype switching of a bacterial cell." *Science* 322.5900 (2008), pp. 442–446.

[45] Munsky, B., Neuert, G., and Oudenaarden, A. van. "Using gene expression noise to understand gene regulation." *Science* 336.6078 (2012), pp. 183–187.

[46] Raj, A. and Oudenaarden, A. van. "Single-molecule approaches to stochastic gene expression." *Annual Review of Biophysics* 38 (2009), pp. 255–270.

[47] Chang, H. H., Hemberg, M., Barahona, M., Ingber, D. E., and Huang, S. "Transcriptome-wide noise controls lineage choice in mammalian progenitor cells". *Nature* 453.7194 (2008), pp. 544–547.

[48] Lockhart, D. J. and Winzeler, E. A. "Genomics, gene expression and DNA arrays." *Nature* 405.6788 (2000), pp. 827–836.

[49] LYON, M. F. "Gene action in the X-chromosome of the mouse (Mus musculus L.)." *Nature* 190 (1961), pp. 372–373.

[50] Ahn, J. and Lee, J. "X Chromosome Inactivation — Learn Science at Scitable". *Nature Education* (2008).

[51] Muller, H. J. "Types of visible variations induced by X-rays inDrosophila". *Journal of Genetics* 22.3 (1930), pp. 299–334.

[52] Wakimoto, B. T. "Beyond the nucleosome: epigenetic aspects of position-effect variegation in Drosophila." *Cell* 93.3 (1998), pp. 321–324.

[53] Rubin, D. C., Ong, D. E., and Gordon, J. I. "Cellular differentiation in the emerging fetal rat small intestinal epithelium: mosaic patterns of gene expression." *Proceedings of the National Academy of Sciences* 86.4 (1989), pp. 1278–1282.

[54] Carter, M. G., Stagg, C. A., et al. "An in situ hybridization-based screen for heterogeneously expressed genes in mouse ES cells". *Gene Expression Patterns : GEP* 8.3 (2008), pp. 181–198.

[55] Ferrell, J. E. and Machleder, E. M. "The biochemical basis of an all-or-none cell fate switch in Xenopus oocytes." *Science* 280.5365 (1998), pp. 895–898.

[56] Guo, G., Huss, M., Tong, G. Q., Wang, C., Li Sun, L., Clarke, N. D., and Robson, P. "Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst." *Developmental Cell* 18.4 (2010), pp. 675–685.

[57] Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W., and Prasher, D. C. "Green fluorescent protein as a marker for gene expression." *Science* 263.5148 (1994), pp. 802–805.

[58] Kalmar, T., Lim, C., Hayward, P., Muñoz-Descalzo, S., Nichols, J., Garcia-Ojalvo, J., and Martinez Arias, A. "Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells." *PLoS Biology* 7.7 (2009), e1000149.

[59] Glauche, I., Herberg, M., and Roeder, I. "Nanog variability and pluripotency regulation of embryonic stem cells–insights from a mathematical model analysis." *PLoS ONE* 5.6 (2010), e11238.

[60] Evans, M. J. and Kaufman, M. H. "Establishment in culture of pluripotential cells from mouse embryos." *Nature* 292.5819 (1981), pp. 154–156.

[61] Martin, G. R. "Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells." *Proceedings of the National Academy of Sciences* 78.12 (1981), pp. 7634–7638.

[62] Thomson, J. A., Itskovitz-Eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel, J. J., Marshall, V. S., and Jones, J. M. "Embryonic stem cell lines derived from human blastocysts." *Science* 282.5391 (1998), pp. 1145–1147.

[63] Takahashi, K. and Yamanaka, S. "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors." *Cell* 126.4 (2006), pp. 663–676.

[64] Gurdon, J. B. "The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles." *Journal of Embryology and Experimental Morphology* 10 (1962), pp. 622–640.

[65] Beddington, R. S. and Robertson, E. J. "An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo." *Development* 105.4 (1989), pp. 733–737.

[66] Mitsui, K., Tokuzawa, Y., et al. "The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells." *Cell* 113.5 (2003), pp. 631–642.

[67] Koutsourakis, M., Langeveld, A., Patient, R., Beddington, R., and Grosveld, F. "The transcription factor GATA6 is essential for early extraembryonic development." *Development* 126.9 (1999), pp. 723–732.

[68] Chazaud, C., Yamanaka, Y., Pawson, T., and Rossant, J. "Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway". *Developmental Cell* 10.5 (2006), pp. 615–624.

[69] Chambers, I., Silva, J., et al. "Nanog safeguards pluripotency and mediates germline development." *Nature* 450.7173 (2007), pp. 1230–1234.

[70] Singh, A. M., Hamazaki, T., Hankowski, K. E., and Terada, N. "A heterogeneous expression pattern for Nanog in embryonic stem cells." *Stem Cells* 25.10 (2007), pp. 2534–2542.

[71] Yamanaka, Y., Lanner, F., and Rossant, J. "FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst." *Development* 137.5 (2010), pp. 715–724.

[72] Ying, Q.-L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. "The ground state of embryonic stem cell self-renewal". *Nature* 453.7194 (2008), pp. 519–523.

[73] Gupta, P. B., Fillmore, C. M., Jiang, G., Shapira, S. D., Tao, K., Kuperwasser, C., and Lander, E. S. "Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells." *Cell* 146.4 (2011), pp. 633–644.

[74] Hayashi, K., Lopes, S. M. C. d. S., Tang, F., and Surani, M. A. "Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states". *Cell Stem Cell* 3.4 (2008), pp. 391–401.

[75] Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K., and Niwa, H. "Identification and characterization of subpopulations in undifferentiated ES cell culture." *Development* 135.5 (2008), pp. 909–918.

[76] Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S. H. "An extended transcriptional network for pluripotency of embryonic stem cells". *Cell* 132.6 (2008), pp. 1049–1061.

[77] Ivanova, N., Dobrin, R., et al. "Dissecting self-renewal in stem cells with RNA interference". *Nature Cell Biology* 442.7102 (2006), pp. 533–538.

[78] Lee, D. Y., Hayes, J. J., Pruss, D., and Wolffe, A. P. "A positive role for histone acetylation in transcription factor access to nucleosomal DNA." *Cell* 72.1 (1993), pp. 73–84.

[79] Dorigo, B., Schalch, T., Kulangara, A., Duda, S., Schroeder, R. R., and Richmond, T. J. "Nucleosome arrays reveal the two-start organization of the chromatin fiber." *Science* 306.5701 (2004), pp. 1571–1573.

[80] Shogren-Knaak, M., Ishii, H., Sun, J.-M., Pazin, M. J., Davie, J. R., and Peterson, C. L. "Histone H4-K16 acetylation controls chromatin structure and protein interactions." *Science* 311.5762 (2006), pp. 844–847.

[81] Wang, X., He, C., Moore, S. C., and Ausio, J. "Effects of histone acetylation on the solubility and folding of the chromatin fiber." *The Journal of Biological Chemistry* 276.16 (2001), pp. 12764–12768.

[82] Delcuve, G. P., Khan, D. H., and Davie, J. R. "Roles of histone deacetylases in epigenetic regulation: emerging paradigms from studies with inhibitors." *Clinical Epigenetics* 4.1 (2012), p. 5.

[83] Watt, F. and Molloy, P. L. "Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter." *Genes & Development* 2.9 (1988), pp. 1136–1143.

[84] Bell, A. C. and Felsenfeld, G. "Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene." *Nature* 405.6785 (2000), pp. 482–485.

[85] Amir, R. E., Van den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U., and Zoghbi, H. Y. "Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2." *Nature Genetics* 23.2 (1999), pp. 185–188.

[86] Nan, X., Ng, H. H., Johnson, C. A., Laherty, C. D., Turner, B. M., Eisenman, R. N., and Bird, A. "Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex." *Nature* 393.6683 (1998), pp. 386–389.

[87] Holliday, R. and Pugh, J. E. "DNA modification mechanisms and gene activity during development." *Science* 187.4173 (1975), pp. 226–232.

[88] Margueron, R. and Reinberg, D. "Chromatin structure and the inheritance of epigenetic information." *Nature Reviews Genetics* 11.4 (2010), pp. 285–296.

[89] Zhu, B. and Reinberg, D. "Epigenetic inheritance: uncontested?" *Cell Research* 21.3 (2011), pp. 435–441.

[90] Xu, C., Bian, C., et al. "Binding of different histone marks differentially regulates the activity and specificity of polycomb repressive complex 2 (PRC2)." *Proceedings of the National Academy of Sciences of the United States of America* 107.45 (2010), pp. 19266–19271.

[91] Rea, S., Eisenhaber, F., et al. "Regulation of chromatin structure by site-specific histone H3 methyltransferases." *Nature* 406.6796 (2000), pp. 593–599.

[92] Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. "Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins." *Nature* 410.6824 (2001), pp. 116–120.

[93] Bannister, A. J., Zegerman, P., Partridge, J. F., Miska, E. A., Thomas, J. O., Allshire, R. C., and Kouzarides, T. "Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain." *Nature* 410.6824 (2001), pp. 120–124.

[94] Hecht, A., Laroche, T., Strahl-Bolsinger, S., Gasser, S. M., and Grunstein, M. "Histone H3 and H4 N-termini interact with SIR3 and SIR4 proteins: a molecular model for the formation of heterochromatin in yeast." *Cell* 80.4 (1995), pp. 583–592.

[95] Imai, S., Armstrong, C. M., Kaeberlein, M., and Guarente, L. "Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase." *Nature* 403.6771 (2000), pp. 795–800.

[96] Oppikofer, M., Kueng, S., Martino, F., Soeroes, S., Hancock, S. M., Chin, J. W., Fischle, W., and Gasser, S. M. "A dual role of H4K16 acetylation in the establishment of yeast silent chromatin." *The EMBO Journal* 30.13 (2011), pp. 2610–2621.

[97] Petruk, S., Sedkov, Y., et al. "TrxG and PcG proteins but not methylated histones remain associated with DNA through replication." *Cell* 150.5 (2012), pp. 922–933.

[98] Hathaway, N. A., Bell, O., Hodges, C., Miller, E. L., Neel, D. S., and Crabtree, G. R. "Dynamics and memory of heterochromatin in living cells." *Cell* 149.7 (2012), pp. 1447–1460.

[99] Young, M. D., Willson, T. A., Wakefield, M. J., Trounson, E., Hilton, D. J., Blewitt, M. E., Oshlack, A., and Majewski, I. J. "ChIP-seq analysis reveals distinct H3K27me3 profiles that correlate with transcriptional activity." *Nucleic Acids Research* 39.17 (2011), pp. 7415–7427.

[100] Singer, Z. S., Yong, J., Tischler, J., Hackett, J. A., Altinok, A., Surani, M. A., Cai, L., and Elowitz, M. B. "Dynamic heterogeneity and DNA methylation in embryonic stem cells." *Molecular Cell* 55.2 (2014), pp. 319–331.

[101] Lander, A. D., Gokoffski, K. K., Wan, F. Y. M., Nie, Q., and Calof, A. L. "Cell lineages and the logic of proliferative control". *PLoS Biology* 7.1 (2009), e15.

[102] Rompolas, P., Mesa, K. R., and Greco, V. "Spatial organization within a niche as a determinant of stem-cell fate." *Nature* 502.7472 (2013), pp. 513–518.

[103] Canham, M. A., Sharov, A. A., Ko, M. S. H., and Brickman, J. M. "Functional heterogeneity of embryonic stem cells revealed through translational amplification of an early endodermal transcript". *PLoS Biology* 8.5 (2010), e1000379.

[104] Yamaji, M., Ueda, J., et al. "PRDM14 ensures naive pluripotency through dual regulation of signaling and epigenetic pathways in mouse embryonic stem cells." *Cell Stem Cell* 12.3 (2013), pp. 368–382.

[105] Raj, A., Bogaard, P. van den, Rifkin, S. A., Oudenaarden, A. v., and Tyagi, S. "Imaging individual mRNA molecules using multiple singly labeled probes". *Nature Methods* 5.10 (2008), pp. 877–879.

[106] Zenklusen, D., Larson, D. R., and Singer, R. H. "Single-RNA counting reveals alternative modes of gene expression in yeast." *Nature Structural & Molecular Biology* 15.12 (2008), pp. 1263–1271.

[107] Jaitin, D. A., Kenigsberg, E., et al. "Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types." *Science* 343.6172 (2014), pp. 776–779.

[108] Shalek, A. K., Satija, R., et al. "Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells." *Nature* (2013).

[109] Okano, M., Bell, D. W., Haber, D. A., and Li, E. "DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development." *Cell* 99.3 (1999), pp. 247–257.

[110] Hackett, J. A., Sengupta, R., Zylicz, J. J., Murakami, K., Lee, C., Down, T. A., and Surani, M. A. "Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine." *Science* 339.6118 (2013), pp. 448–452.

[111] Reik, W. "Stability and flexibility of epigenetic gene regulation in mammalian development." *Nature* 447.7143 (2007), pp. 425–432.

[112] Schübeler, D., Lorincz, M. C., Cimbora, D. M., Telling, A., Feng, Y. Q., Bouhassira, E. E., and Groudine, M. "Genomic targeting of methylated DNA: influence of methylation on transcription, replication, chromatin structure, and histone acetylation." *Molecular and Cellular Biology* 20.24 (2000), pp. 9103–9112.

[113] Smith, Z. D., Chan, M. M., Mikkelsen, T. S., Gu, H., Gnirke, A., Regev, A., and Meissner, A. "A unique regulatory phase of DNA methylation in the early mammalian embryo." *Nature* 484.7394 (2012), pp. 339–344.

[114] Fouse, S. D., Shen, Y., Pellegrini, M., Cole, S., Meissner, A., Van Neste, L., Jaenisch, R., and Fan, G. "Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation." *Cell Stem Cell* 2.2 (2008), pp. 160–169.

[115] Mohn, F., Weber, M., Rebhan, M., Roloff, T. C., Richter, J., Stadler, M. B., Bibel, M., and Schübeler, D. "Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors." *Molecular Cell* 30.6 (2008), pp. 755–766.

[116] Marks, H., Kalkan, T., et al. "The transcriptional and epigenomic foundations of ground state pluripotency." *Cell* 149.3 (2012), pp. 590–604.

[117] Wray, J., Kalkan, T., Gomez-Lopez, S., Eckardt, D., Cook, A., Kemler, R., and Smith, A. "Inhibition of glycogen synthase kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation." *Nature Cell Biology* 13.7 (2011), pp. 838–845.

[118] Ficz, G., Hore, T. A., et al. "FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency." *Cell Stem Cell* 13.3 (2013), pp. 351–359.

[119] Habibi, E., Brinkman, A. B., et al. "Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells." *Cell Stem Cell* 13.3 (2013), pp. 360–369.

[120] Leitch, H. G., McEwen, K. R., et al. "Naive pluripotency is associated with global DNA hypomethylation." *Nature Structural & Molecular Biology* 20.3 (2013), pp. 311–316.

[121] Friedman, N., Cai, L., and Xie, X. S. "Linking stochastic dynamics to population distribution: an analytical framework of gene expression." *Physical Review Letters* 97.16 (2006), p. 168302.

[122] Paulsson, J. and Ehrenberg, M. "Random signal fluctuations can reduce random fluctuations in regulated components of chemical regulatory networks." *Physical Review Letters* 84.23 (2000), pp. 5447–5450.

[123] Peccoud, J. and Ycart, B. "Markovian modeling of gene-product synthesis". *Theoretical Population Biology* 48.2 (1995), pp. 222–234.

[124] Shahrezaei, V. and Swain, P. S. "Analytical distributions for stochastic gene expression." *Proceedings of the National Academy of Sciences of the United States of America* 105.45 (2008), pp. 17256–17261.

[125] Suter, D. M., Molina, N., Gatfield, D., Schneider, K., Schibler, U., and Naef, F. "Mammalian genes are transcribed with widely different bursting kinetics." *Science* 332.6028 (2011), pp. 472–474.

[126] Cai, L., Friedman, N., and Xie, X. S. "Stochastic protein expression in individual cells at the single molecule level." *Nature* 440.7082 (2006), pp. 358–362.

[127] Maamar, H., Raj, A., and Dubnau, D. "Noise in gene expression determines cell fate in Bacillus subtilis." *Science* 317.5837 (2007), pp. 526–529.

[128] Zong, C., So, L.-h., Sepúlveda, L. A., Skinner, S. O., and Golding, I. "Lysogen stability is determined by the frequency of activity bursts from the fate-determining gene." *Molecular Systems Biology* 6 (2010), p. 440.

[129] Wray, J., Kalkan, T., and Smith, A. G. "The ground state of pluripotency." *Biochemical Society Transactions* 38.4 (2010), pp. 1027–1032.

[130] Paulsson, J., Berg, O. G., and Ehrenberg, M. "Stochastic focusing: fluctuation-enhanced sensitivity of intracellular regulation." *Proceedings of the National Academy of Sciences* 97.13 (2000), pp. 7148–7153.

[131] Macfarlan, T. S., Gifford, W. D., et al. "Embryonic stem cell potency fluctuates with endogenous retrovirus activity." *Nature* 487.7405 (2012), pp. 57–63.

[132] Auernhammer, C. J., Bousquet, C., and Melmed, S. "Autoregulation of pituitary corticotroph SOCS-3 expression: characterization of the murine SOCS-3 promoter." *Proceedings of the National Academy of Sciences* 96.12 (1999), pp. 6964–6969.

[133] Ficz, G., Branco, M. R., et al. "Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation." *Nature* 473.7347 (2011), pp. 398–402.

[134] Ito, S., D'Alessio, A. C., Taranova, O. V., Hong, K., Sowers, L. C., and Zhang, Y. "Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification." *Nature* 466.7310 (2010), pp. 1129–1133.

[135] Koh, K. P., Yabuuchi, A., et al. "Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells." *Cell Stem Cell* 8.2 (2011), pp. 200–213.

[136] Wu, H., D'Alessio, A. C., et al. "Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells." *Nature* 473.7347 (2011), pp. 389–393.

[137] Grabole, N., Tischler, J., Hackett, J. A., Kim, S., Tang, F., Leitch, H. G., Magnúsdóttir, E., and Surani, M. A. "Prdm14 promotes germline fate and naive pluripotency by repressing FGF signalling and DNA methylation." *EMBO Reports* 14.7 (2013), pp. 629–637.

[138] Ma, Z., Swigut, T., Valouev, A., Rada-Iglesias, A., and Wysocka, J. "Sequence-specific regulator Prdm14 safeguards mouse ESCs from entering extraembryonic endoderm fates." *Nature Structural & Molecular Biology* 18.2 (2011), pp. 120–127.

[139] Meissner, A., Mikkelsen, T. S., et al. "Genome-scale DNA methylation maps of pluripotent and differentiated cells." *Nature* 454.7205 (2008), pp. 766–770.

[140] Borgel, J., Guibert, S., Li, Y., Chiba, H., Schübeler, D., Sasaki, H., Forné, T., and Weber, M. "Targets and dynamics of promoter DNA methylation during early mouse development." *Nature Genetics* 42.12 (2010), pp. 1093–1100.

[141] Locke, J. C. W. and Elowitz, M. B. "Using movies to analyse gene circuit dynamics in single cells." *Nature Reviews. Microbiology* 7.5 (2009), pp. 383–392.

[142] Brewster, R. C., Weinert, F. M., Garcia, H. G., Song, D., Rydenfelt, M., and Phillips, R. "The transcription factor titration effect dictates level of gene expression." *Cell* 156.6 (2014), pp. 1312–1323.

[143] Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S., and Elowitz, M. B. "Gene regulation at the single-cell level." *Science* 307.5717 (2005), pp. 1962–1965.

[144] Sharova, L. V., Sharov, A. A., Nedorezov, T., Piao, Y., Shaik, N., and Ko, M. S. H. "Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells." *DNA Research* 16.1 (2009), pp. 45–58.

[145] Tsumura, A., Hayakawa, T., et al. "Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b." *Genes to Cells : Devoted to Molecular & Cellular Mechanisms* 11.7 (2006), pp. 805–814.

[146] Chen, X., Xu, H., et al. "Integration of external signaling pathways with the core transcriptional network in embryonic stem cells". *Cell* 133.6 (2008), pp. 1106–1117.

[147] Wang, J., Rao, S., Chu, J., Shen, X., Levasseur, D. N., Theunissen, T. W., and Orkin, S. H. "A protein interaction network for pluripotency of embryonic stem cells". *Nature* 444.7117 (2006), pp. 364–368.

[148] Iyer-Biswas, S., Hayot, F., and Jayaprakash, C. "Stochasticity of gene products from transcriptional pulsing." *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics* 79.3 Pt 1 (2009), p. 031911.

[149] Dunlop, M. J., Cox, R. S., Levine, J. H., Murray, R. M., and Elowitz, M. B. "Regulatory activity revealed by dynamic correlations in gene expression noise". *Nature Genetics* 40.12 (2008), pp. 1493–1498.

[150] Pedraza, J. M. and Oudenaarden, A. van. "Noise propagation in gene networks." *Science* 307.5717 (2005), pp. 1965–1969.

[151] Shukla, S., Kavak, E., et al. "CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing." *Nature* 479.7371 (2011), pp. 74–79.

[152] Takizawa, T., Nakashima, K., et al. "DNA methylation is a critical cell-intrinsic determinant of astrocyte differentiation in the fetal brain." *Developmental Cell* 1.6 (2001), pp. 749–758.

[153] You, J. S., Kelly, T. K., De Carvalho, D. D., Taberlay, P. C., Liang, G., and Jones, P. A. "OCT4 establishes and maintains nucleosome-depleted regions that provide additional layers of epigenetic regulation of its target genes." *Proceedings of the National Academy of Sciences of the United States of America* 108.35 (2011), pp. 14497–14502.

[154] Festuccia, N., Osorno, R., et al. "Esrrb is a direct Nanog target gene that can substitute for Nanog function in pluripotent cells." *Cell Stem Cell* 11.4 (2012), pp. 477–490.

[155] Martello, G., Sugimoto, T., et al. "Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal." *Cell Stem Cell* 11.4 (2012), pp. 491–504.

[156] Furusawa, C. and Kaneko, K. "A dynamical-systems view of stem cell biology." *Science* 338.6104 (2012), pp. 215–217.

[157] Faddah, D. A., Wang, H., Cheng, A. W., Katz, Y., Buganim, Y., and Jaenisch, R. "Single-cell analysis reveals that expression of nanog is biallelic and equally variable as that of other pluripotency factors in mouse ESCs." *Cell Stem Cell* 13.1 (2013), pp. 23–29.

[158] Miyanari, Y. and Torres-Padilla, M.-E. "Control of ground-state pluripotency by allelic regulation of Nanog." *Nature* 483.7390 (2012), pp. 470–473.

[159] Filipczyk, A., Gkatzis, K., Fu, J., Hoppe, P. S., Lickert, H., Anastassiadis, K., and Schroeder, T. "Biallelic expression of nanog protein in mouse embryonic stem cells." *Cell Stem Cell* 13.1 (2013), pp. 12–13.

[160] Hansen, C. H. and Oudenaarden, A. van. "Allele-specific detection of single mRNA molecules in situ." *Nature Methods* 10.9 (2013), pp. 869–871.

[161] Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M., and Cai, L. "Single-cell in situ RNA profiling by sequential hybridization." *Nature Methods* 11.4 (2014), pp. 360–361.

[162] Lubeck, E. and Cai, L. "Single-cell systems biology by super-resolution imaging and combinatorial labeling." *Nature Methods* 9.7 (2012), pp. 743–748.

[163] Wang, H., Yang, H., Shivalila, C. S., Dawlaty, M. M., Cheng, A. W., Zhang, F., and Jaenisch, R. "One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering." *Cell* 153.4 (2013), pp. 910–918.

[164] Kueh, H. Y., Champhekar, A., Champhekhar, A., Nutt, S. L., Elowitz, M. B., and Rothenberg, E. V. "Positive feedback between PU.1 and the cell cycle controls myeloid differentiation." *Science* 341.6146 (2013), pp. 670–673.

[165] Bintu, L., Yong, J., Antebi, Y., McCue, K., Kazuki, Y., Uno, N., Oshimura, M., and Elowitz, M. B. "Dynamics of epigenetic regulation at the single-cell level". *(submitted)* (2015).

[166] Bibikova, M., Laurent, L. C., Ren, B., Loring, J. F., and Fan, J.-B. "Unraveling epigenetic regulation in embryonic stem cells." *Cell Stem Cell* 2.2 (2008), pp. 123–134.

[167] Egger, G., Liang, G., Aparicio, A., and Jones, P. A. "Epigenetics in human disease and prospects for epigenetic therapy." *Nature* 429.6990 (2004), pp. 457–463.

[168] Fraga, M. F. and Esteller, M. "Epigenetics and aging: the targets and the marks." *Trends in Genetics : TIG* 23.8 (2007), pp. 413–418.

[169] Gifford, C. A., Ziller, M. J., et al. "Transcriptional and epigenetic dynamics during specification of human embryonic stem cells." *Cell* 153.5 (2013), pp. 1149–1163.

[170] Hawkins, R. D., Hon, G. C., et al. "Distinct epigenomic landscapes of pluripotent and lineage-committed human cells." *Cell Stem Cell* 6.5 (2010), pp. 479–491.

[171] Zhu, J., Adli, M., et al. "Genome-wide chromatin state transitions associated with developmental and environmental cues." *Cell* 152.3 (2013), pp. 642–654.

[172] Angel, A., Song, J., Dean, C., and Howard, M. "A Polycomb-based switch underlying quantitative epigenetic memory." *Nature* 476.7358 (2011), pp. 105–108.

[173] Song, J., Angel, A., Howard, M., and Dean, C. "Vernalization - a cold-induced epigenetic switch." *Journal of Cell Science* 125.Pt 16 (2012), pp. 3723–3731.

[174] Rowe, H. M. and Trono, D. "Dynamic control of endogenous retroviruses during development." *Virology* 411.2 (2011), pp. 273–287.

[175] Bashor, C. J., Horwitz, A. A., Peisajovich, S. G., and Lim, W. A. "Rewiring cells: synthetic biology as a tool to interrogate the organizational principles of living systems." *Annual Review of Biophysics* 39 (2010), pp. 515–537.

[176] Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J., and Phillips, R. "Transcriptional regulation by the numbers: models". *Current Opinion in Genetics & Development* 15.2 (2005), pp. 116–124.

[177] Mukherji, S. and Oudenaarden, A. van. "Synthetic biology: understanding biological design from synthetic circuits." *Nature Reviews Genetics* 10.12 (2009), pp. 859–871.

[178] Hansen, K. H., Bracken, A. P., et al. "A model for transmission of the H3K27me3 epigenetic mark." *Nature Cell Biology* 10.11 (2008), pp. 1291–1300.

[179] Konermann, S., Brigham, M. D., et al. "Optical control of mammalian endogenous transcription and epigenetic states." *Nature* 500.7463 (2013), pp. 472–476.

[180] Li, F., Papworth, M., Minczuk, M., Rohde, C., Zhang, Y., Ragozin, S., and Jeltsch, A. "Chimeric DNA methyltransferases target DNA methylation to specific DNA sequences and repress expression of target genes." *Nucleic Acids Research* 35.1 (2007), pp. 100–112.

[181] Maeder, M. L., Angstman, J. F., et al. "Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins." *Nature Biotechnology* 31.12 (2013), pp. 1137–1142.

[182] Margolin, J. F., Friedman, J. R., Meyer, W. K., Vissing, H., Thiesen, H. J., and Rauscher, F. J. "Krüppel-associated boxes are potent transcriptional repression domains." *Proceedings of the National Academy of Sciences* 91.10 (1994), pp. 4509–4513.

[183] Mendenhall, E. M., Williamson, K. E., Reyon, D., Zou, J. Y., Ram, O., Joung, J. K., and Bernstein, B. E. "Locus-specific editing of histone modifications at endogenous enhancers." *Nature Biotechnology* 31.12 (2013), pp. 1133–1136.

[184] Wang, A. H., Bertos, N. R., et al. "HDAC4, a human histone deacetylase related to yeast HDA1, is a transcriptional corepressor." *Molecular and Cellular Biology* 19.11 (1999), pp. 7816–7827.

[185] Gilbert, L. A., Larson, M. H., et al. "CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes." *Cell* 154.2 (2013), pp. 442–451.

[186] Keung, A. J., Bashor, C. J., Kiriakov, S., Collins, J. J., and Khalil, A. S. "Using targeted chromatin regulators to engineer combinatorial and spatial transcriptional regulation." *Cell* 158.1 (2014), pp. 110–120.

[187] Lahav, G., Rosenfeld, N., Sigal, A., Geva-Zatorsky, N., Levine, A. J., Elowitz, M. B., and Alon, U. "Dynamics of the p53-Mdm2 feedback loop in individual cells". *Nature Genetics* 36.2 (2004), pp. 147–150.

[188] Levine, J. H., Lin, Y., and Elowitz, M. B. "Functional roles of pulsing in genetic circuits." *Science* 342.6163 (2013), pp. 1193–1200.

[189] Muzzey, D. and Oudenaarden, A. van. "Quantitative time-lapse fluorescence microscopy in single cells." *Annual Review of Cell and Developmental Biology* 25 (2009), pp. 301–327.

[190] Tay, S., Hughey, J. J., Lee, T. K., Lipniacki, T., Quake, S. R., and Covert, M. W. "Single-cell NF-kappaB dynamics reveal digital activation and analogue information processing." *Nature* 466.7303 (2010), pp. 267–271.

[191] Boyer, L. A., Plath, K., et al. "Polycomb complexes repress developmental regulators in murine embryonic stem cells." *Nature* 441.7091 (2006), pp. 349–353.

[192] Margueron, R. and Reinberg, D. "The Polycomb complex PRC2 and its mark in life". *Nature* 469.7330 (2011), pp. 343–349.

[193] Urrutia, R. "KRAB-containing zinc-finger repressor proteins." *Genome Biology* 4.10 (2003), p. 231.

[194] Fussenegger, M., Morris, R. P., Fux, C., Rimann, M., Stockar, B. von, Thompson, C. J., and Bailey, J. E. "Streptogramin-based gene regulation systems for mammalian cells." *Nature Biotechnology* 18.11 (2000), pp. 1203–1208.

[195] Weber, W., Fux, C., et al. "Macrolide-based transgene control in mammalian cells and mice." *Nature Biotechnology* 20.9 (2002), pp. 901–907.

[196] Ayyanathan, K., Lechner, M. S., et al. "Regulated recruitment of HP1 to a euchromatic gene induces mitotically heritable, epigenetic gene silencing: a mammalian cell culture model of gene variegation." *Genes & Development* 17.15 (2003), pp. 1855–1869.

[197] Petrie, K., Guidez, F., Howell, L., Healy, L., Waxman, S., Greaves, M., and Zelent, A. "The histone deacetylase 9 gene encodes multiple protein isoforms." *The Journal of Biological Chemistry* 278.18 (2003), pp. 16059–16072.

[198] Miska, E. A., Karlsson, C., Langley, E., Nielsen, S. J., Pines, J., and Kouzarides, T. "HDAC4 deacetylase associates with and represses the MEF2 transcription factor." *The EMBO Journal* 18.18 (1999), pp. 5099–5107.

[199] Wakabayashi-Ito, N. and Nagata, S. "Characterization of the regulatory elements in the promoter of the human elongation factor-1 alpha gene." *The Journal of Biological Chemistry* 269.47 (1994), pp. 29831–29837.

[200] Wang, R., Liang, J., Jiang, H., Qin, L.-J., and Yang, H.-T. "Promoter-dependent EGFP expression during embryonic stem cell propagation and differentiation." *Stem Cells and Development* 17.2 (2008), pp. 279–289.

[201] Yusufzai, T. M. and Felsenfeld, G. "The 5'-HS4 chicken beta-globin insulator is a CTCF-dependent nuclear matrix-associated element." *Proceedings of the National Academy of Sciences* 101.23 (2004), pp. 8620–8624.

[202] Kazuki, Y. and Oshimura, M. "Human artificial chromosomes for gene delivery and the development of animal models." *Molecular Therapy : the Journal of the American Society of Gene Therapy* 19.9 (2011), pp. 1591–1601.

[203] Yamaguchi, S., Kazuki, Y., Nakayama, Y., Nanba, E., Oshimura, M., and Ohbayashi, T. "A method for producing transgenic cells using a multi-integrase system on a human artificial chromosome vector." *PLoS ONE* 6.2 (2011), e17267.

[204] Urlinger, S., Baron, U., Thellmann, M., Hasan, M. T., Bujard, H., and Hillen, W. "Exploring the sequence space for tetracycline-dependent transcriptional activators: novel mutations yield expanded range and sensitivity." *Proceedings of the National Academy of Sciences of the United States of America* 97.14 (2000), pp. 7963–7968.

[205] Greer, E. L. and Shi, Y. "Histone methylation: a dynamic mark in health, disease and inheritance." *Nature Reviews Genetics* 13.5 (2012), pp. 343–357.

[206] Schuettengruber, B., Martinez, A.-M., Iovino, N., and Cavalli, G. "Trithorax group proteins: switching genes on and keeping them active." *Nature Reviews Molecular Cell Biology* 12.12 (2011), pp. 799–814.

[207] Bruniquel, D. and Schwartz, R. H. "Selective, stable demethylation of the interleukin-2 gene enhances transcription by an active process." *Nature Immunology* 4.3 (2003), pp. 235–240.

[208] Busslinger, M. and Tarakhovsky, A. "Epigenetic control of immunity." *Cold Spring Harbor Perspectives in Biology* 6.6 (2014).

[209] Lin, T., Chao, C., Saito, S., Mazur, S. J., Murphy, M. E., Appella, E., and Xu, Y. "p53 induces differentiation of mouse embryonic stem cells by suppressing Nanog expression." *Nature Cell Biology* 7.2 (2005), pp. 165–171.

[210] Dodd, I. B., Micheelsen, M. A., Sneppen, K., and Thon, G. "Theoretical analysis of epigenetic cell memory by nucleosome modification." *Cell* 129.4 (2007), pp. 813–822.

[211] Gomez, D., Shankman, L. S., Nguyen, A. T., and Owens, G. K. "Detection of histone modifications at specific gene loci in single cells in histological sections." *Nature Methods* 10.2 (2013), pp. 171–177.

[212] Simon, J. A. and Kingston, R. E. "Mechanisms of polycomb gene silencing: knowns and unknowns." *Nature Reviews Molecular Cell Biology* 10.10 (2009), pp. 697–708.

[213] Deniaud, E. and Bickmore, W. A. "Transcription and the nuclear periphery: edge of darkness?" *Current Opinion in Genetics & Development* 19.2 (2009), pp. 187–191.

[214] Ram, O., Goren, A., et al. "Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells." *Cell* 147.7 (2011), pp. 1628–1639.

[215] Suganuma, T. and Workman, J. L. "Signals and combinatorial functions of histone modifications." *Annual Review of Biochemistry* 80 (2011), pp. 473–499.

[216] Venters, B. J., Wachi, S., et al. "A Comprehensive Genomic Binding Map of Gene and Chromatin Regulatory Proteins in Saccharomyces". *Molecular Cell* 41.4 (2011), pp. 480–492.

[217] Zhou, V. W., Goren, A., and Bernstein, B. E. "Charting histone modifications and the functional organization of mammalian genomes." *Nature Reviews Genetics* 12.1 (2011), pp. 7–18.

[218] Lanzavecchia, A. and Sallusto, F. "Dynamics of T lymphocyte responses: intermediates, effectors, and memory cells." *Science* 290.5489 (2000), pp. 92–97.

[219] Ahrends, R., Ota, A., Kovary, K. M., Kudo, T., Park, B. O., and Teruel, M. N. "Controlling low rates of cell differentiation through noise and ultrahigh feedback." *Science* 344.6190 (2014), pp. 1384–1389.

[220] Voog, J. and Jones, D. L. "Stem cells and the niche: a dynamic duo." *Cell Stem Cell* 6.2 (2010), pp. 103–115.

[221] Dekel, E. and Alon, U. "Optimality and evolutionary tuning of the expression level of a protein." *Nature* 436.7050 (2005), pp. 588–592.

[222] Mendelson, A. and Frenette, P. S. "Hematopoietic stem cell niche maintenance during homeostasis and regeneration." *Nature Medicine* 20.8 (2014), pp. 833–846.

[223] Yeung, T. M., Chia, L. A., Kosinski, C. M., and Kuo, C. J. "Regulation of self-renewal and differentiation by the intestinal stem cell niche." *Cellular and Molecular Life Sciences : CMLS* 68.15 (2011), pp. 2513–2523.

[224] Li, T., Hu, J.-F., et al. "CTCF regulates allelic expression of Igf2 by orchestrating a promoter-polycomb repressive complex 2 intrachromosomal loop." *Molecular and Cellular Biology* 28.20 (2008), pp. 6473–6482.

[225] Lanner, F. and Rossant, J. "The role of FGF/Erk signaling in pluripotent cells." *Development* 137.20 (2010), pp. 3351–3360.

[226] Tan, F. E. and Elowitz, M. B. "Brf1 posttranscriptionally regulates pluripotency and differentiation responses downstream of Erk MAP kinase." *Proceedings of the National Academy of Sciences of the United States of America* 111.17 (2014), E1740–8.

[227] Brumbaugh, J., Russell, J. D., Yu, P., Westphall, M. S., Coon, J. J., and Thomson, J. A. "NANOG Is Multiply Phosphorylated and Directly Modified by ERK2 and CDK1 In Vitro". *Stem Cell Reports* 2.1 (2014), pp. 18–25.

[228] Yi, F., Pereira, L., Hoffman, J. A., Shy, B. R., Yuen, C. M., Liu, D. R., and Merrill, B. J. "Opposing effects of Tcf3 and Tcf1 control Wnt stimulation of embryonic stem cell self-renewal." *Nature Cell Biology* 13.7 (2011), pp. 762–770.

[229] Costa, Y., Ding, J., et al. "NANOG-dependent function of TET1 and TET2 in establishment of pluripotency." *Nature* 495.7441 (2013), pp. 370–374.

[230] Feldman, N., Gerson, A., Fang, J., Li, E., Zhang, Y., Shinkai, Y., Cedar, H., and Bergman, Y. "G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis". *Nature Cell Biology* 8.2 (2006), pp. 188–194.

Appendix A

Supplemental Information for Chapter 2

Supplemental Experimental Procedures

Detailed Culture Conditions

All cells were maintained in humidity-controlled chambers at 37°C, with 5% CO2 in serum+LIF media [Glasgow Minimum Essential Medium (GMEM) supplemented with 10% FBS (HyClone, Thermo Scientific), 2 mM glutamine, 100 units/ml penicillin, 100 ug/ml streptomycin, 1 mM sodium pyruvate, 1000 units/ml Leukemia Inhibitory Factor (LIF, Millipore), 1X Minimum Essential Medium Non-Essential Amino Acids (MEM NEAA, Invitrogen) and 50 uM β-Mercaptoethanol].

Correlation between Citrine and Nanog transcripts in Nanog knock-in reporter cells (NKICit)

We validated the Nanog knock-in reporter by performing smFISH for correlation between Nanog (unmodified allele) and Citrine (knock-in reporter on second allele) (Figure A.4B). We observed that when grown in serum+LIF conditions, $\sim$10% of cells contained Nanog but no Citrine transcripts, likely due to silenced expression of their reporter cassettes during

prolonged propagation without antibiotics. The remaining cell population showed even stronger correlation between Nanog and Citrine transcripts (r = 0.78). We corrected for the potential systematic error that may result in the calculation of low-to-high switching rate such that an observed rate of 1.9 0.29 transitions per 100 cell cycles was adjusted to the reported 2.3 0.25 (mean SD). We note that the magnitude of this error does not alter key conclusions, including those about the relative stabilities of the two states. Furthermore, the asymmetry of this silencing behavior – we did not find a corresponding fraction of cells expressing Citrine but no Nanog transcripts – suggests that this is not likely a result of mono-allelic regulation.

smFISH procedure and imaging system

Up to 48 20mer DNA probes per target mRNA were synthesized and conjugated to Alexa fluorophore 488, 555, 594, or 647 (Life Technologies) and then purified by HPLC. Cells for smFISH experiments were plated at 40,000/cm2 and harvested after 48 hours. Trypsinized cells were washed in PBS and fixed in 4% formaldehyde at room temperature for 5 mins. Fixed cells were resuspended in 70% ethanol and stored at -20°C overnight. The next day, cells were hybridized with 4nM probe per target species at 30°C, in 20% Formamide, 2X SSC, 0.1g/ml Dextran Sulfate, 1mg/ml E.coli tRNA, 2mM Vanadyl ribonucleoside complex, and 0.1% Tween 20 in nuclease free water. The following morning, cells were washed in 20% Formamide, 2x SSC, and 0.1% Tween 20 at 30°C, followed by two washes in 2x SSC + 0.1% Tween 20 at room temperature. Hybridized cells were placed between #1 coverslips and flattened by applying pressure evenly across the glass.

After flattening cells between coverslips, dots typically span two distinct focal planes. However, to maximize the number of cells imaged in a given acquisition time, only one

of these focal planes was captured. This results in approximately ∼60% of each cell's transcripts being captured in a single slice, as compared to taking a stack of images across the entire volume of each cell (Figure A.1A-D).

Imaging was performed on a Nikon Ti-E with Perfect Focus, Semrock FISH filtersets, Lumencor Sola illumination, 60x 1.4NA oil objective, and a Coolsnap HQ2 camera. Snapshots were taken using an automated grid-based acquisition system on a motorized ASI MS-2000 stage.

Monte-Carlo bivariate Kolmogorov-Smirnov test

The 1D Kolmogorov-Smirnov test was extended to two dimensions [1] to determine whether an empirical bivariate distribution showed any dependence between variables; the 2D Cumulative Distribution Function (CDF) is computed in each possible quadrant of the 2D plane P(x<x0), P(y<y0); P(x>x0), P(y<y0); P(x<x0), P(y>y0); and P(x>x0), P(y>y0). The 2D KS test statistic is thus defined as the largest difference between empirical and theoretical distributions across each of these possible regions. In order to generate a test-statistic distribution under the null hypothesis, we performed a Monte-Carlo simulation where sets of random pairs of data points are sampled from the PDF formed by the product of the marginal distributions. The resulting bivariate CDF is compared to the theoretical CDF and the maximal difference is taken. This is performed repeatedly in order to generate a distribution of maximal differences that would occur by chance. Finally, the test statistic is computed from the empirical distribution, and compared to this distribution at a 95% confidence level.

Movie acquisition system

Images were acquired on the IX81 inverted microscope system (Olympus) using the Metamorph acquisition software (Molecular Devices) with the iKon Charge Coupled Device (CCD) camera (Andor). Fluorophores were excited using X-Cite XLED1 light source (Lumen Dynamics) equipped with the BLX, BGX, and GYX modules.

Movie data analysis: Segmentation and tracking

The Schnitzcells script package [2] was used to segment and track cells from the acquired images. This package performs a number of procedures as described below. Briefly, cells were segmented with Matlab built-in edge detection script, using the Laplacian of Gaussian method. Segmented cells in individual frames were then tracked across all time points by performing a point-matching algorithm on successive pairs of frames to generate a cell lineage data structure. To obtain the total fluorescence level of each cell, the images were flattened by correcting for the nonuniformity of illumination, followed by local background correction that takes into account the camera acquisition background, autofluorescence from the medium and fluorescence contribution from neighboring cells.

Movie data analysis: Production rate estimation and step detection

To enable the continuous estimation of production rates (slopes), frames around cell divisions are removed and fluorescence lost during divisions (to sister cells) is added back to the trace of interest to create a continuous total fluorescence trace for each lineage. Instantaneous fluorescence production rates were estimated by fitting the continuous total fluorescence of a 6-hour window to a linear section using the linear least squares method. Distributions of reporter production rates (Figures A.4A & B) were obtained by sampling

the instantaneous fluorescence production rates of all cell lineages at 1-hour intervals. To characterize abrupt changes in production rates, we identified sharp inflections of the continuous total fluorescence traces by applying a custom-built step detector on overlapping and consecutive 6-hour windows 15 minutes apart (Figure A.4D). For each window, we obtained fits to a linear polynomial model and a continuous two-piece linear polynomial model with a joint at midpoint using the linear least squares and non-linear least squares methods, respectively. The continuous two-piece linear model can be represented as follows:

$$y = \begin{cases} m_a x + c & , \quad x < x_{mid}, \\ m_a x_{mid} + c + m_b(x - x_{mid}) & , \quad x \geq x_{mid}, \end{cases}$$

where x_{mid} is the midpoint of the window.

We used two criteria to determine whether a given window fits better to the one-piece or two-piece linear fits: (1) whether the noisiness of the trace can explain the deviation of the data from the one-piece fit (mean sum of squared errors, $M.S.E.$), and (2) whether the two slopes obtained from the two-piece fit are significantly different from each other. For (1), we define the noisiness of the trace as the variance of the distribution of frame-to-frame fluctuations in total fluorescence, i.e. $var(Y_{t+1} - Y_t)$. For a perfectly linear trace without noise, the mean of $Y_{t+1} - Y_t$ equals the slope of the trace. As the observation noise increases, the sum of squared errors ($S.S.E.$) of one-piece fit increases even if the underlying trace has a constant slope. We therefore estimated the portion of $S.S.E.$ of one-piece fit unexplained by the noisiness of the trace as the residual noise, defined as $M.S.E._{\cdot 1pc}/var(Y_{t+1} - Y_t)$, where n is the number of frames within a window. For (2) we obtained the 95% confidence bounds of the two slopes in the two-piece fit and determined if they overlap. Using (1) and (2), we identified stretches of frames where two-piece fit is significantly better than

one-piece. The frame with the highest residual noise among each of these stretches was designated as the point of inflection and the rest of the trace was approximated by linear segments between these points.

Movie data analysis: Hidden Markov Model and Viterbi Algorithm

We set up a two-state HMM to estimate the frequency of state-switching events between the higher and lower Nanog states. We assume each of the two states can produce an independent Gaussian distribution of production rates, with specified mean and variance, including potential overlap between the two states. Over each unit time, a cell can either stay at its current state or switch to the other state with specified probabilities. Thus, given a specific parameter set, there exists for the production rate time-series of each cell a corresponding series of underlying states that has the maximum likelihood. This likelihood is a balance between the probability of observing a production rate at the corresponding state and that of switching to another state, such that a cell that transiently exhibits a production rate far from the mean of its current state is more likely to be fluctuating rapidly within a state than switching away and back. The Baum-Welch algorithm [3] maximizes the sum of this likelihood over all cells by iteratively changing the parameters in small increments, improving the total likelihood each time.

Prior to training the model with data, initial transition rates between the states in both directions were set at 0.0001/hour. Initial parameters for each state were set with the mean value drawn from the range of observed production rates and variance. Re-initializing the random parameters in the model yielded similar results. We employed the HMM toolbox for Matlab [4], which generated a maximum likelihood estimate of the model parameters using the Baum-Welch algorithm. Since the production rate sequences used to train the

HMM contained repeated time-series when multiple lineages shared the same ancestor, the state-transition rates generated directly from HMM could be an overestimation. We applied the Viterbi algorithm [5] to combine the model parameter estimates and the observed data to infer the most likely state sequence for each cell lineage. From this we reported the empirical state-transition rates, normalized to the average length of a cell cycle.

Bursty transcription simulation and mixing time analysis

Bursty transcription was simulated using the model previously described [6]. In this model, a promoter can transit stochastically between an active and an inactive form. This is not to be confused with a cellular state, which is usually maintained over a longer timescale and within which a gene bursts with a characteristic burst size and frequency. Transcription occurs only when the promoter is in its active form, producing a burst of mRNA molecules, which decay exponentially. To aid comparison between the simulated transcription dynamics and our experimental observations, we added protein production to the simulation. Further, since our fluorescence protein is stable, and to restrict the source of heterogeneity in our simulation to stochastic transcription, we assumed zero protein decay rate and deterministic protein production at a constant rate. Lastly, both mRNA and protein are partitioned when cells divide, which were set to have division rates similar to experimental data. This model can be described by the following reactions:

$$
\begin{aligned}
I &\xrightarrow{k_{on}} A,\\
A &\xrightarrow{k_{off}} I,\\
A &\xrightarrow{\beta_m} A + m,\\
m &\xrightarrow{\alpha_m} \emptyset,\\
m &\xrightarrow{\beta_p} m + p.
\end{aligned}
$$

Here, A and I denote the promoter in its active and inactive forms, respectively; m – mRNA level; p – protein level; k_{on} and k_{off} – activating and inactivating rates of the promoter, respectively; α_m – mRNA degradation rate; β_m – mRNA production rate; $\emptyset$ – mRNA degradation; β_p – protein production rate.

A cellular state is thus characterized by the frequency of mRNA bursts and the mean number of mRNA molecules produced per burst. Here, we considered one limiting case of this model, where k_{off} is significantly larger than k_{on} and somewhat larger than α_m. This assumption can be related physically to a scenario where bursts are relatively infrequent and have short durations, and the distribution of mRNA levels produced under these assumptions can be described with a single gamma [7] or NB function [8]. A cell changes state in a gene when one or more of the parameters k_{on}, k_{off}, or β_m for that gene is changed, thus resulting in different burst frequencies and sizes.

To simulate mRNA and protein dynamics for the Nanog-high state in serum+LIF condition (shown in Figures A.4E, A.4E), we used the following parameters estimated from mRNA distributions in smFISH: for Nanog – burst size = 33 mRNA/hour, burst frequency = 0.39 bursts/hour; for Oct4 – burst size = 87 mRNA/hour, burst frequency = 0.52 bursts/hour. These assume that mRNA half-lives of Nanog and Oct4 are 5.85 and 7.4 hours, respectively [9].

We utilized computer simulations of this model to explore whether changes in state affect the intra-state dynamics of heterogeneity. Varying burst frequency and burst size results in traces with various frequencies of apparent steps when analyzed using the same step detector, which identified regimes in the bursty transcription parameter space where steps of similar quality to the ones observed can be generated (Figures A.4F, A.5C). Furthermore, the resulting dynamics also display a wide range of shapes of fluctuation and levels of expression. We quantified these variations with an objective measurement, the mixing time, a population metric adapted from Sigal et al. (2006). For each simulated population (n = 200 traces) using a single parameter set, we ranked all traces by their production rate at each time point. Thus a cell starting with the lowest production rate among the population may change in this rank when its production rate changes over time. We computed the autocorrelation function $A(\tau)$ of this rank for each population and the mixing time is defined as the time lag τ at which $A(\tau)$ decayed to 0.5. We opted to calculate the mixing time using production rate but not total fluorescence level because the stable fluorescent reporter facilitates accurate production rate estimate but may not reflect the physiological level of endogenous proteins. Additionally, for more direct comparison between the mixing times calculated from simulated and observed data, the production rates in simulation were computed using the simulated protein traces after Gaussian noise similar to the level observed was added.

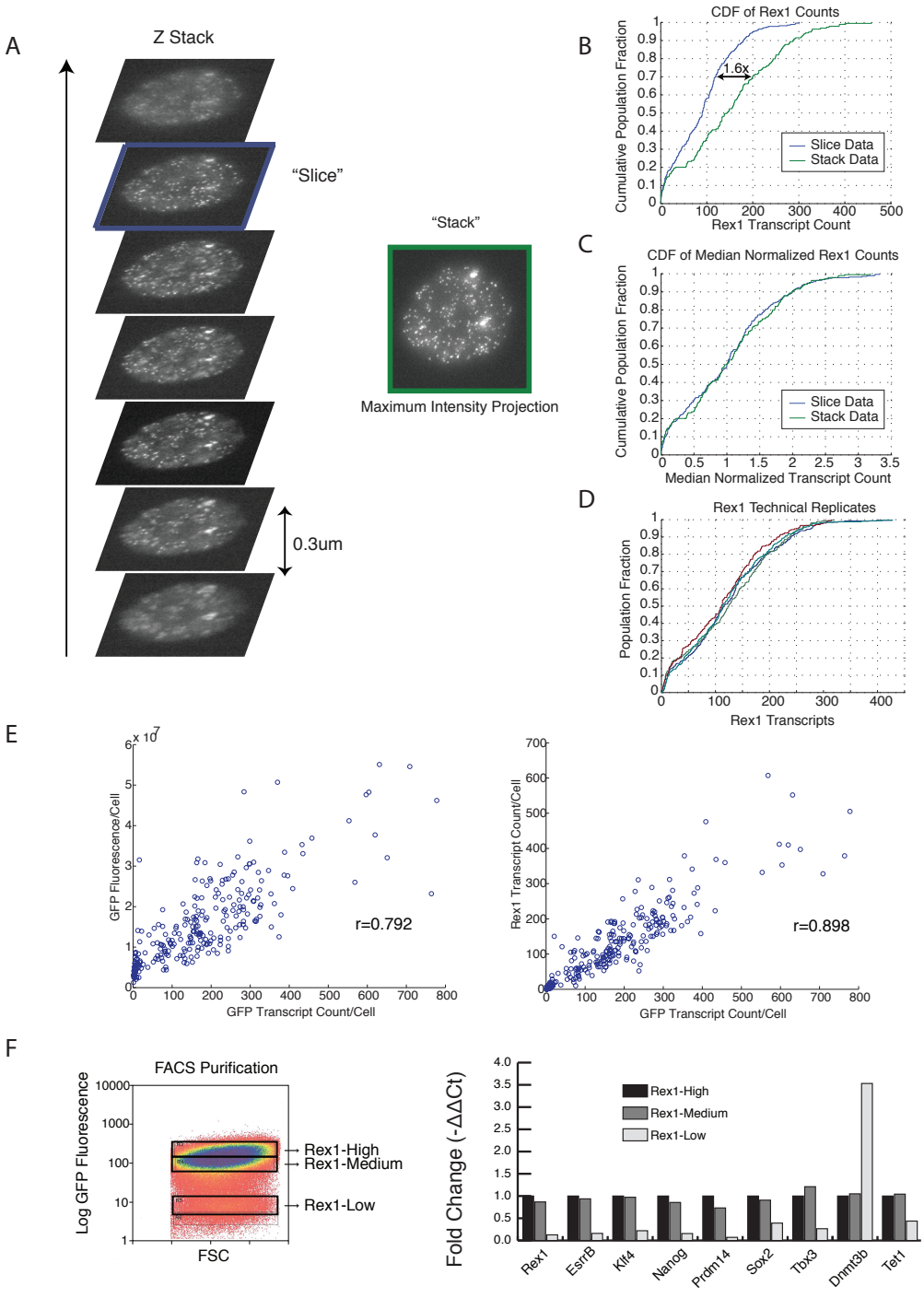


Figure A.1: (Caption on the following page.)

Figure A.1: Validation of smFISH
(A) A stack of snapshots taken through the whole volume of a single cell; the resulting maximum-intensity projection (green box), and a single slice (blue box) are fed into the image-processing algorithm for dot-detection.
(B) Cumulative Distributions of dot counts for each of the two imaging approaches is shown across a population of cells.
(C) Same distributions as in B, but normalized by the sample median.
(D) Technical replicates for the single-slice approach.
(E) Correlation between dGFP protein fluorescence as measured simultaneously with dGFP transcripts (left), and correlation between Rex1 (unmodified allele) and dGFP (knock-in reporter on second allele) transcripts (right). r is the Pearson correlation coefficient.
(F) (Left) Sorted subpopulations of the bimodal Rex1-dGFP knock-in reporter. (Right) qPCR results on these subpopulations for a subset of target genes also examined by sm-FISH. Values were normalized to expression levels of the housekeeping gene Gapdh, and are represented as $2^{-\Delta\Delta Ct}$ with respect to the 'Rex1-high' subpopulation

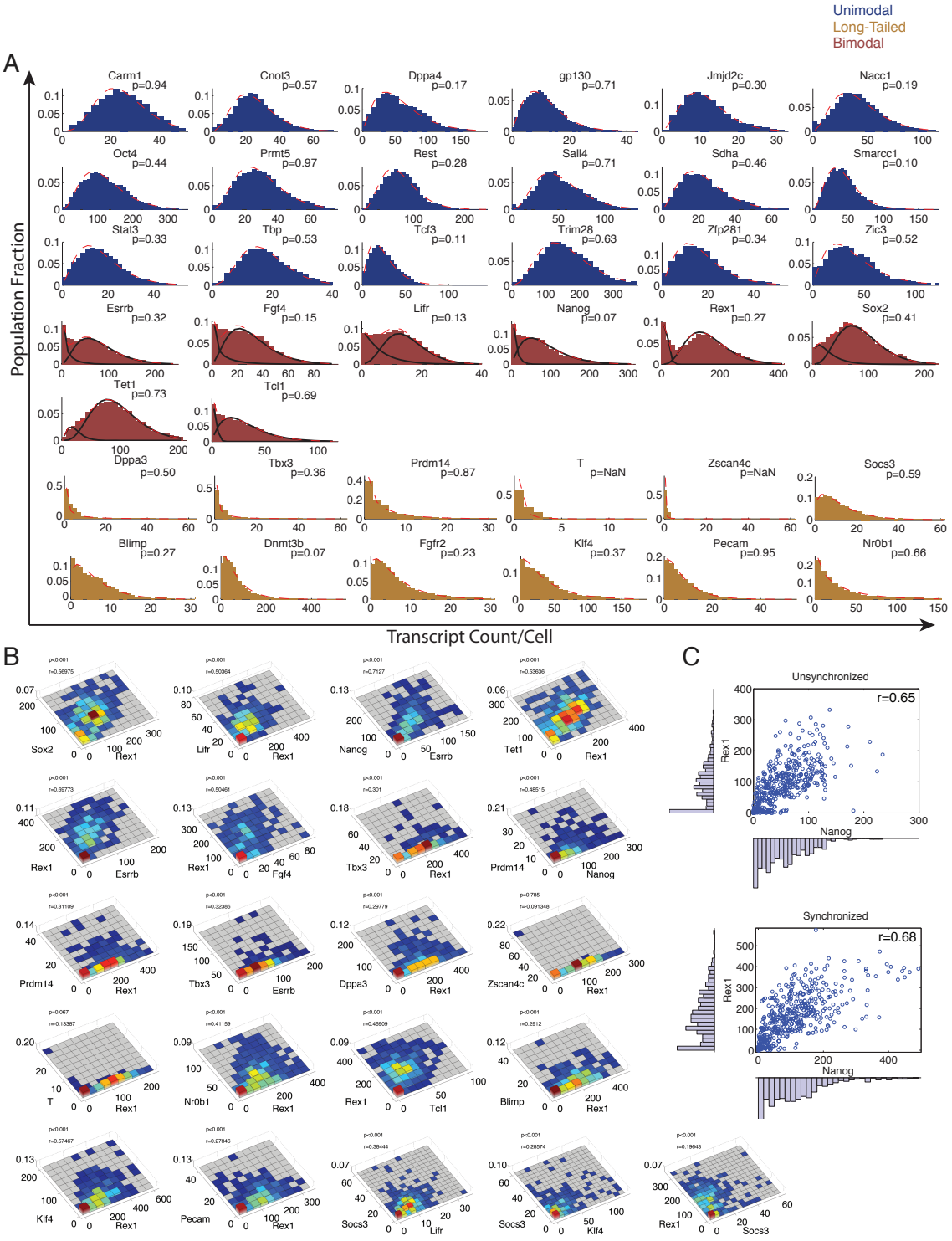


Figure A.2: (Caption on the following page.)

Figure A.2: mRNA distributions and correlations by smFISH
(A) Empirical distributions and MLE fits for unimodal, bimodal, and long-tailed genes. p-values are for χ^2 GOF tests. p>0.05 indicates that the fit to the distribution is indistinguishable from the empirically measured distribution. Where present, solid lines represent components of the fit. Dashed line represents the overall fit to the distribution.
(B) Pairwise relationships between heterogeneously expressed genes. p-values are from the 2D KS-test. r is the Pearson correlation coefficient.
(C) Correlation and marginal distributions of Rex1 and Nanog in a control population (top) and population synchronized by a double thymidine block fixed immediately following the block (bottom). r is the Pearson correlation coefficient.

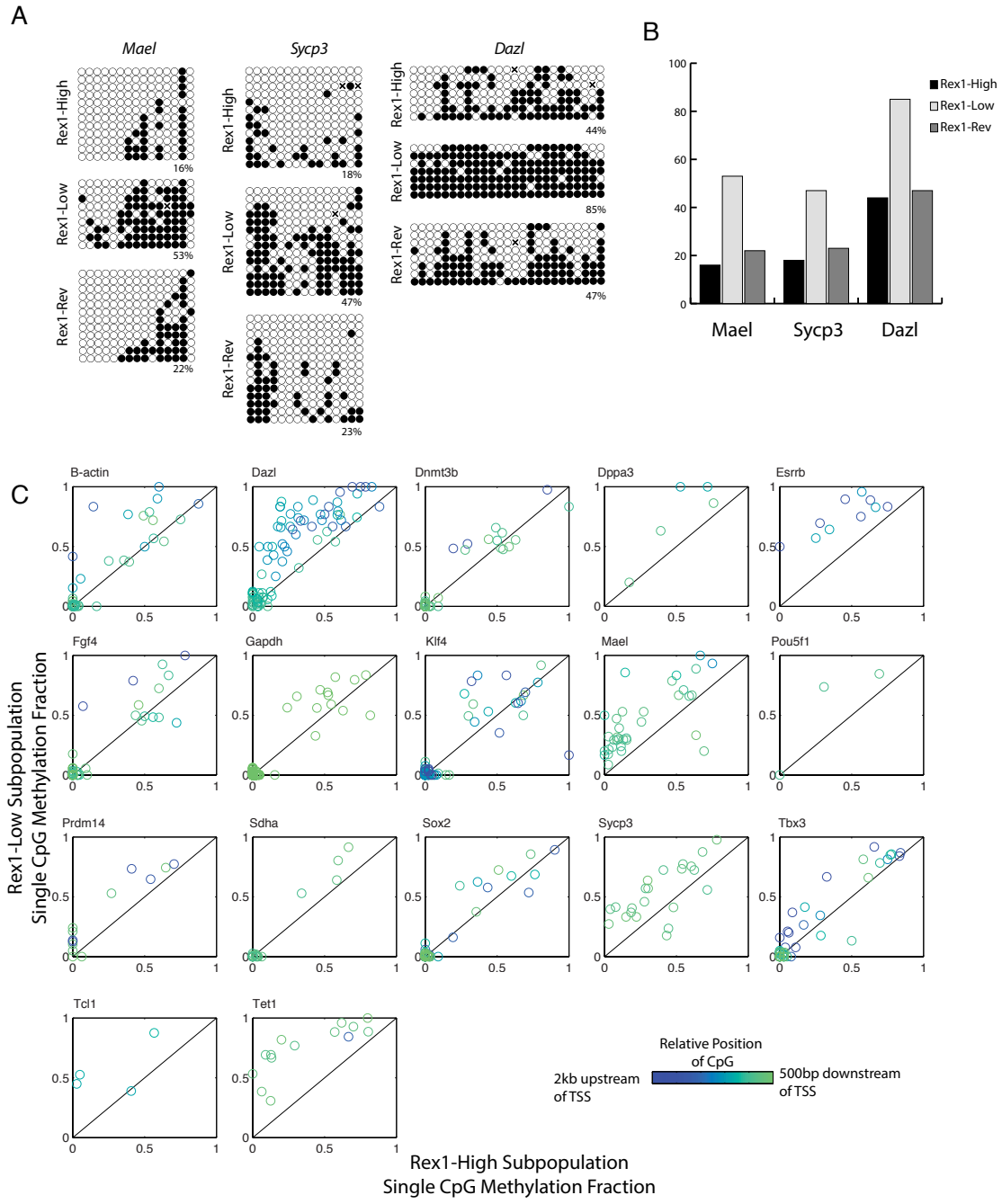


Figure A.3: Differential methylation between Rex1 states
(A) Locus specific bisulfite sequencing plots between Rex1-high, -low, and -low-to-high-reverting cells at three targets of methylation. Open circles are unmethylated, filled circles are methylated, and x's are unknown.
(B) Measurements from A are plotted as bar graphs for comparison.
(C) Scatter plots showing how single CpGs in the promoters of a given gene change between Rex1-high and -low states. Color coding represents the position of a base relative to the transcriptional state site.

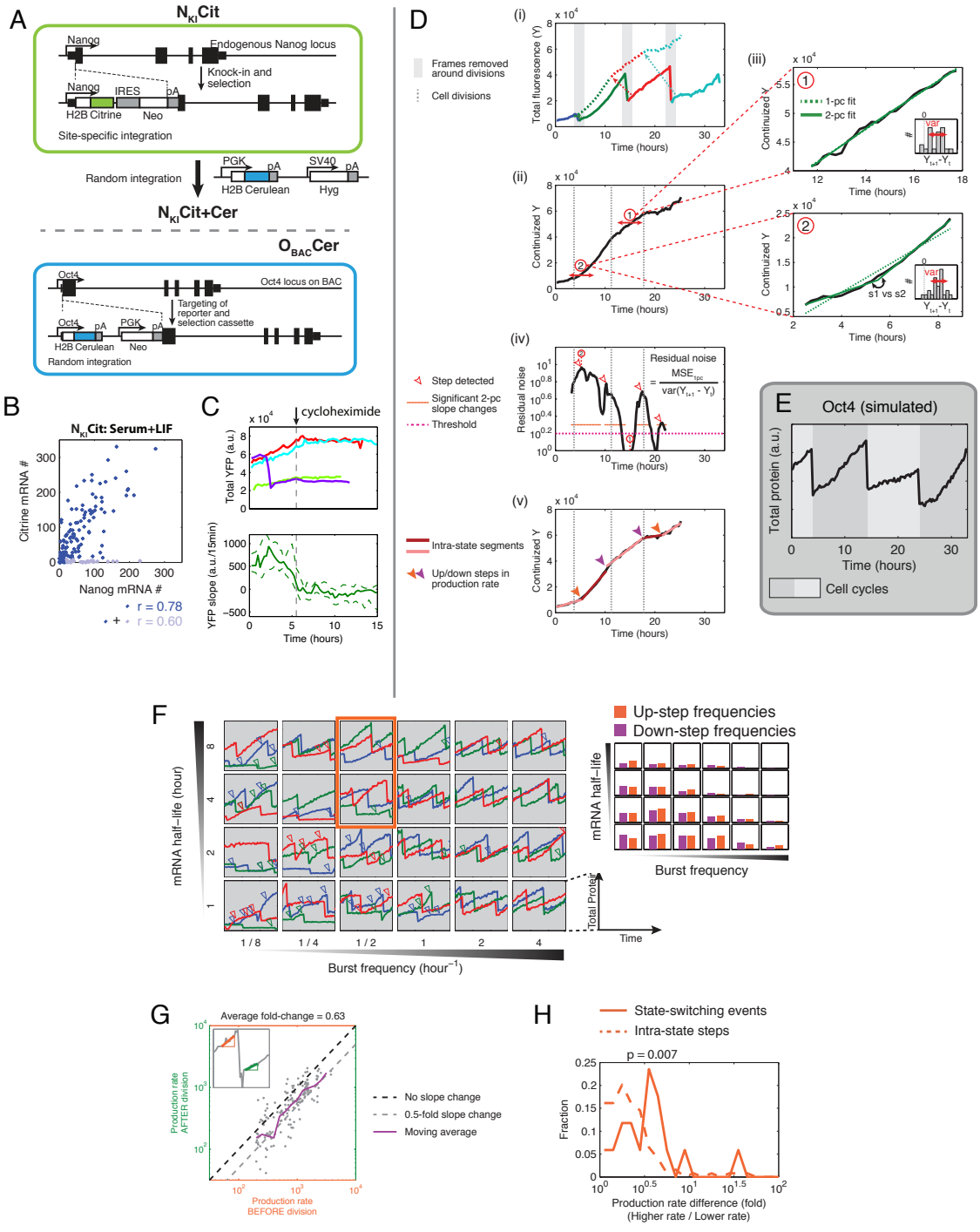


Figure A.4: (Caption on the following page.)

Figure A.4: Construction and analysis of live cell reporters, and simulations based on observed kinetics
(A) Schematic of Nanog reporter (top) and Oct4 reporter (bottom) construction.
(B) Correlation between Nanog (unmodified allele) and Citrine (knock-in reporter on second allele) transcripts in NKICit cell line. r, Pearson correlation coefficient. Light blue, presumed fraction of cells with silenced reporter cassettes (∼10% of all cells; see Supp. Info. for discussion); dark blue, remaining cell population.
(C) H2B-Citrine protein degradation rate assayed by blocking translation during movie at time indicated. Total YFP became flat (top) with negligible slope (bottom) shortly after cycloheximide treatment.
(D) Identification of sharp inflections in total fluorescence traces. (i) First, frames around cell divisions are removed and fluorescence lost during divisions is added back to the daughter trace to create a continuous trace for each lineage (ii), where a step detector spanning a 6-hour window is applied across consecutive frames. (iii) For each window, a one-piece linear fit is compared with a two-piece fit that is flexible at the midpoint. A two-piece fit is considered better than a one-piece fit when two criteria are met: 1) residual noise of the one-piece fit is higher than a threshold (see Supp. Info.), and 2) the slopes of the two-piece fit are significantly different between the two pieces. (iv) For each stretch of frames meeting both criteria 1 (magenta line indicates threshold) and 2 (orange line indicates where two-piece fit yields significantly different slopes), the window with the highest residual noise is assigned to be the inflection. (v) Continuized trace approximated into linear segments between identified points of inflection.
(E) Apparent steps from simulated Oct4 expression under the bursty transcription model using parameters estimated from smFISH.
(F) Protein traces were simulated under the bursty transcription model over various mRNA half-life and burst frequency combinations; mean burst size was kept constant at 35 mRNA/burst. Gaussian noise proportional to the total protein level and equivalent to the magnitude of frame-to-frame variation empirically observed was added to the simulated traces for comparability. Arrowheads indicate detected steps on simulated trace of the corresponding color. Note that changes in production rate around cell division events can be identified as steps either before or after the division. Red box: estimated regime for Nanog-Hi in serum+LIF. Right: variation in the frequency of detected steps over the same parameter space.
(G) Production rates decrease by an average of 0.63-fold across cell divisions. Each point represents a division event. Average production rates of the 4-hour windows before and after each cell division are compared. Black dotted line: zero change; grey dotted line: 0.5-fold change; purple line: average trend; Inset: example trace indicating slope before and after division.
(H) Changes in production rate over state-switching events or intra-state steps. Higher rate-to-lower rate ratios are plotted for all steps and events, i.e. down-steps and Nanog-high-to-Nanog-low switching events are represented by the reciprocals of rate change (p-value, KS test).

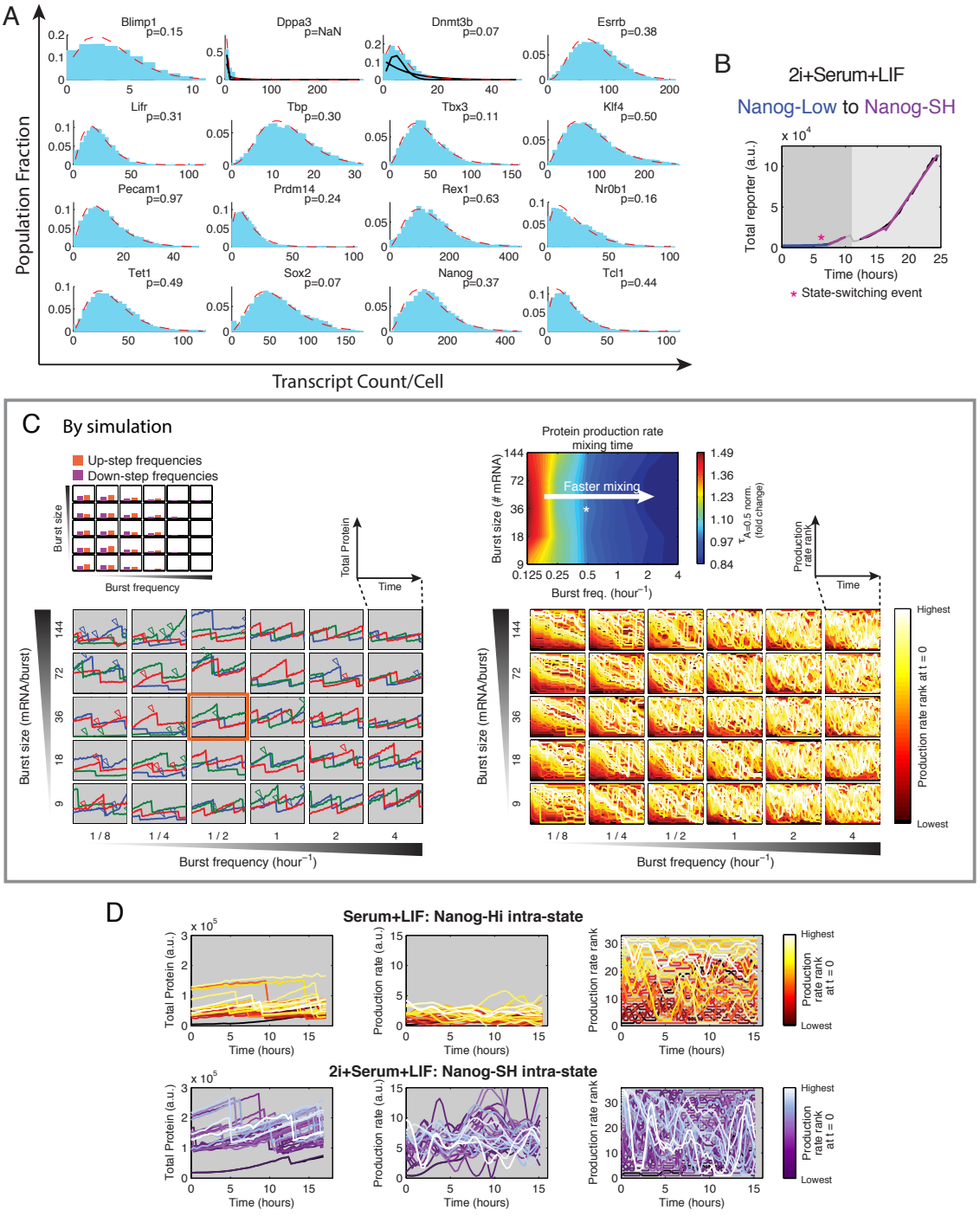


Figure A.5: (Caption on the following page.)

Figure A.5: Quantitative analysis of how 2i+serum+LIF affect static distributions and dynamics of gene expression for pluripotency regulators
(A) smFISH transcript count distribution of factors in 2i+serum+LIF with MLE fits overlaid. p-values are for χ^2 GOF tests. p>0.05 indicates that the fit to the distribution is indistinguishable from the empirically measured distribution. Where present, solid lines represent components of the fit. Dashed line represents the overall fit to the distribution.
(B) Example trace of cells switching from Nanog-low to Nanog-SH in 2i+serum+LIF.
(C) Left: simulated traces similar to Figure A.4F, except over various combinations of burst size and burst frequency; mRNA half-life was kept constant at 4 hours. Bottom right: rank of production rate of 30 randomly selected traces (out of a total of 200) in each simulation under the corresponding parameter combination. Traces are color-coded by the initial rank at t = 0 as in D. Top right: mixing time of protein production rate, defined as the time where auto-correlation of rank drops below 0.5.
(D) Nanog expression dynamics of cells in serum/LIF with or without 2i. Each trace represents one cell randomly picked from a tracked lineage tree. Production rates are normalized by cell size and ranked within the group for each time point. Traces are color-coded by the initial rank at t = 0.

# of sister pairs Direction of switch	Neither sister switched	Only one sister switched	Both sisters switched	Expected number of sister pairs that both switched [95% C.I.]**
N^{Lo} to N^{Hi}	169	7	0	**[0 - 1]**
N^{Hi} to N^{Lo}	139	15	2	**[0 - 2]**

Table A.1: State-switching events show no correlation between sister cells
Data shown are combined results from two independent experiments. Analysis of individual data sets yields the same conclusion.
* Data points are discarded if one of the cells in a sister pair was lost or not traceable in the movie
** Confidence interval obtained by random permutation test with 100,000 trials. Green indicates observed frequency of sister pairs in which both cells switched falls within the 95% C.I.

Supplemental Bibliography – Appendix A

[1] Peacock, J. A. "Two-dimensional goodness-of-fit testing in astronomy". *Monthly Notices of the Royal Astronomical Society* 202.3 (1983), pp. 615–627.

[2] Young, J. W., Locke, J. C. W., Altinok, A., Rosenfeld, N., Bacarian, T., Swain, P. S., Mjolsness, E., and Elowitz, M. B. "Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy." *Nature Protocols* 7.1 (2012), pp. 80–88.

[3] Do, C. B. and Batzoglou, S. "What is the expectation maximization algorithm?" *Nature Biotechnology* 26.8 (2008), pp. 897–899.

[4] Murphy, K. P. "Hidden Markov Model (HMM) Toolbox for Matlab" (1998).

[5] Rabiner, L. R. "A tutorial on hidden Markov models and selected applications in speech recognition". *Proceedings of the IEEE* 77.2 (1989), pp. 257–286.

[6] Peccoud, J. and Ycart, B. "Markovian modeling of gene-product synthesis". *Theoretical Population Biology* 48.2 (1995), pp. 222–234.

[7] Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y., and Tyagi, S. "Stochastic mRNA Synthesis in Mammalian Cells". *PLoS Biology* 4.10 (2006), e309.

[8] Paulsson, J., Berg, O. G., and Ehrenberg, M. "Stochastic focusing: fluctuation-enhanced sensitivity of intracellular regulation." *Proceedings of the National Academy of Sciences* 97.13 (2000), pp. 7148–7153.

[9] Sharova, L. V., Sharov, A. A., Nedorezov, T., Piao, Y., Shaik, N., and Ko, M. S. H. "Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells." *DNA Research* 16.1 (2009), pp. 45–58.

Appendix B

Supplemental Information for Chapter 3

Supplemental Experimental Procedures

Plasmid construction

The PhiC31-Neo-ins-5xTetO-pEF-H2B-Citrine-ins reporter plasmid was assembled using a backbone containing the PhiC31 attB site, a neomycin resistance gene, and a multiple cloning site flanked by two 1.2kb chicken HS4 insulators on each side (phiC31-Neomycin-2xcHS4ins-MCS-2xcHS4ins)[1]. Individual elements of the reporter were PCR amplified as follows: five Tet binding sites from the TRE-tight plasmid (Clontech), pEF from pEF/FRT/V5-Dest (Life Technologies), and H2B-citrine from pEV2-12xCSL-H2B-Citrine [2]. These components were first sequentially cloned into the pExchange1 backbone using standard molecular biology techniques. The entire TRE-pEF-H2B-citrine was then PCR-amplified and combined by Gibson assembly with the phiC31-Neomycin-2xcHS4ins-MCS-2xcHS4ins backbone cut by AvrII. This construct was designed such that after integration, the neomycin gene would be expressed from a PGK promoter situated upstream of the phiC31 site in the HAC [1]. The PhiC31 integrase was a gift from the Oshimura Lab [1].

The plasmids containing the rTetR-CR fusions were built using Gibson assembly of

the pExchange1 backbone containing the pEF promoter (cut with BamHI and KpnI), H2B-mCherry (PCR-amplified from a derivative of pEV-12xCSL-H2B-mCherry [2]), rTetR (PCR-amplified from rtTA3 system, Clontech), and a PCR product for each CR. The plasmids encoding for the CRs were as follows: pCMV-HA-EED (Addgene 24231), HDAC4 Flag (Addgene 13821), Dnmt3b cDNA (isoform 5, OpenBiosystems MMM1013-99827219), and PSV40-E-KRAB-pA (pWW43 [3], a gift from Martin Fussenegger).

Culture conditions

Cells were cultured at 37C, in a humidified atmosphere with 5% CO2. For all experiments, except movies, the growth media consisted of Alpha MEM Earle's Salts (9144, Irvine Scientific) with 10% Tet Approved FBS (Clontech Laboratories) and 1X Penicillin/Streptomycin/L-glutamine (Life Technologies) added. Media containing the appropriate antibiotics (300 g/ml neomycin and 300 g/ml zeocin) was changed every 2-3 days during maintenance. During movies, cells were grown in low-fluorescence imaging media [2], which consisted of Alpha-MEM without phenol red, riboflavin, folic acid, and vitamin B12 (Life Technologies, custom made), supplemented with 10% FBS and 1X Pen/Strep/L-glutamine. During all recruitment and de-recruitment experiments, media did not contain neomycin or zeocin, and was changed every 24 hours in all wells. Cells were harvested by rinsing with Dulbecco's Phosphate-Buffered Saline (DPBS, Life Technologies), and incubating at room temperature with 0.25% Trypsin (Life Technologies). For long-term storage, cells were frozen in growth media with 10% DMSO, placed at -80C (for up to a month), and eventually transferred in liquid nitrogen.

Time-lapse movie analysis, event detection and silencing rate analysis

Cells were segmented and tracked using the H2B-mcherry fluorescence with in-house Matlab code. Using the contours obtained from this algorithm, total citrine fluorescence levels were extracted for each of the cell tracks. Since H2B-fluorophore fusion results in a stable protein reporter, total fluorescence level increases at a steady rate when the reporter is expressed at a constant level (Figure 3.1C, pre-recruitment), but flat-lines when expression stops completely. At each cell division, the total fluorescence signal is approximately halved as fluorescent protein molecules partition between daughter cells [4]. We corrected for this partitioning of fluorescence and obtained cumulative total fluorescence traces, from which reporter production rate can be assessed continuously across cell divisions.

To detect silencing events, a threshold on reporter production rate is set for each individual lineage at 50% of its median reporter production rate before dox addition (Figure B.1E). A cell is marked silenced when its reporter production rate drops and remains below this threshold for 12 hours, to avoid misidentification due to intrinsic fluctuations in gene expression levels. Similarly, reactivation events are identified when an OFF cell's reporter production rate rises and stays above a global threshold, set as 50% of the median reporter production rate of all cells before the initial dox addition (Figure B.1F). A global threshold is used for reactivation detection since cells in reactivation movies begin with reporter production rate close to zero.

To estimate the timescale of the transient increase in silencing rates, $k_{off}(t)$, and its maximum value for each factor, we fitted the experimental silencing rates over time with the positive portion of a logistic function,

$$k_{off}(t) = \frac{2a}{1 + e^{-bt}} - a,$$

where a and b are constants. In this two-parameter fit, the maximum k_{off} is equal to a, where the time t to reach half of the maximum k_{off} is given by $-log(1/3)/b$.

Chromatin mark spreading model

For the analysis in Figure B.2E & F, we modeled silencing by simulating chromatin mark gain, loss, and spreading over an array of 10 nucleosomes. Cells are considered silent if 9 or 10 nucleosomes are modified and active otherwise. We use a Gillespie algorithm to simulate the time evolution of each array (cell). Briefly, at each nucleosome, the probability of writing or erasing a modification is proportional to the corresponding rates of mark gain and spreading versus loss. The time increments between simulation steps are chosen by randomly sampling from an exponential distribution whose mean value equals the inverse of the sum of the rates associated with the nucleosome modified. Each simulation is run 100 times for 1000 cells (arrays). For each simulation, we calculate the effective rate of silencing (k_{off}) as a function of time in a manner similar to our experiments: we count the number of cells that move to the off state (with 9 or more nucleosomes modified) and normalize by the total number of cells still on at that time. The silencing rates presented in Figure 3.2 are obtained by averaging over the 100 simulations.

Targeted bisulfite sequencing

Targeted bisulfite sequencing was performed by Zymo Research using genomic DNA extracted from cell samples (with DNeasy kit, Qiagen) obtained concurrently with ChIP-qPCR and MeDIP-qPCR experiments. Each CpG had to be contained in at least 50 reads to be included in the analysis.

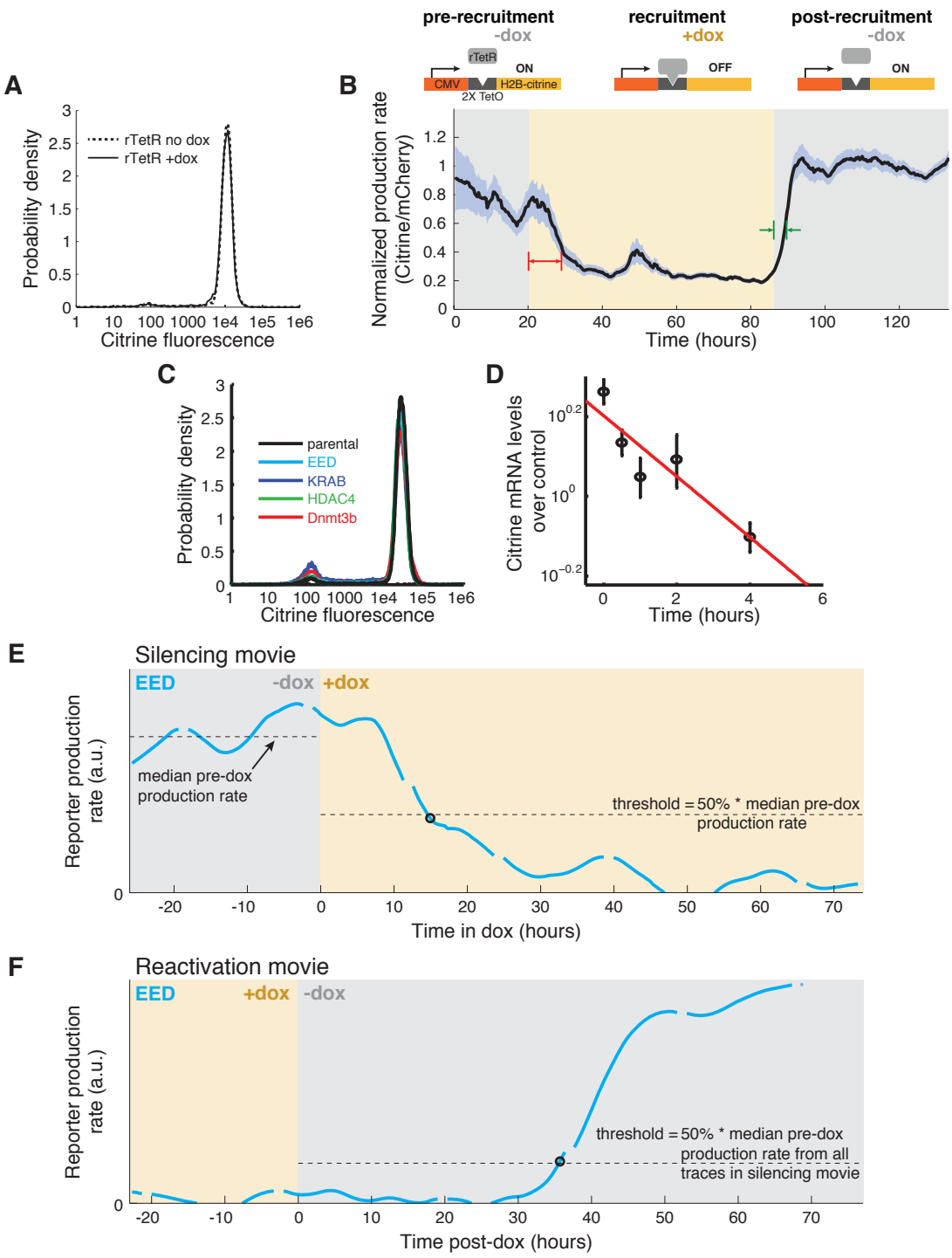


Figure B.1: (Caption on the following page.)

Figure B.1: Specificity and speed of recruitment and de-recruitment, and event detection

(A) To test if recruitment of rTetR alone affects reporter expression, we measured citrine fluorescence levels after transiently transfecting pEF-H2B-mCherry-T2A-rTetR into the parental reporter cell line, followed by dox treatment. The fluorescence distributions are identical in the absence of dox (dashed line) and after 33 hours of dox induction (continuous line), indicating that rTetR recruitment alone does not repress gene expression.

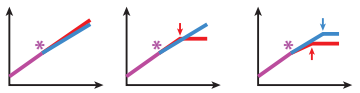
(B) In order to estimate the speed of rTetR binding and unbinding, we built a separate cell line containing the rTetR domain alone, recruited to a CMV promoter followed by 2xTetO sites (top cartoons). In this case, binding of rTetR directly represses expression by acting as a transcription roadblock. We measured changes in reporter production rates using time-lapse microscopy. By averaging over all single-cell traces, we obtained the mean production rate as a function of time (black curve). The gray shaded curves represent SEM. Addition of dox resulted in a reduction in gene expression within 82 hours (red arrows). Removal of dox relieved this repression within 2.52 hours (green arrows).

(C) Citrine fluorescence distributions were measured for all cell lines in the absence of dox, after 11 days of growth without selection antibiotics. The fluorescence of the cell lines containing rTetR-CR fusions (colored lines) are very similar to the parental reporter line (black), indicating that there is relatively little non-specific binding of rTetR at the reporter gene in the absence of dox.

(D) The half-life of citrine mRNA was measured by inhibiting transcription with actinomycin D (5g/ml) and following the levels of mRNA as a function of time. These values were measured by qPCR and normalized against a constant amount of mCherry mRNA spiked in as internal control. By fitting these data to an exponential decay (red line), we determined the half-life of the reporter mRNA to be 3.9 hours (with a 95% confidence interval of [0.9, 4.3] hours).

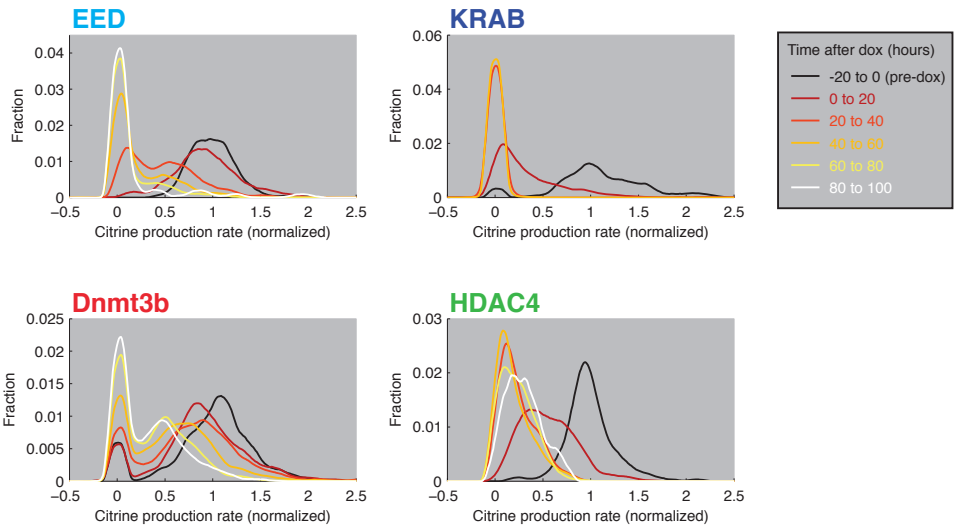
(E & F) Silencing (E) and reactivation (F) event detection. The cyan curves are smoothed time derivatives of the cumulative fluorescence traces for the same lineages shown in Figure 3.1C & D, respectively. Frames immediately around cell divisions were removed when the raw total fluorescence traces were processed to give cumulative traces (see Figure 3.1C & D). For silencing, we calculated the median reporter production rate before dox addition, and set a threshold at 50% of this value (dashed line in dox region). When the reporter production rate of this lineage crossed this threshold and remained under it for 12 hours, we assigned this event as silencing (circle). Reactivation events were detected similarly, but in the reverse direction (F). However, because reactivation involved cell lineages silenced prior to the start of the movie, it was impossible to obtain the median reporter production rates for each of the lineages when the gene was active. Therefore, the median pre-dox production rate from all lineage traces in the silencing movie immediately preceding the reactivation movie was used to set a unifying threshold to detect reactivation events.

A

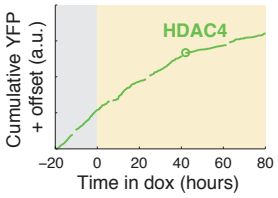


Factor	Number of sister cells pairs (Observed)				Expected OFF/OFF; [95% C.I.]
	Total	ON/ON	ON/OFF	OFF/OFF	
EED	288	161	82	45	25; [19 - 32]
KRAB	34	0	4	30	Most cells silenced within one cell cycle
Dnmt3b	207	105	75	27	20; [14 - 26]
HDAC4	43	5	8	30	Most cells silenced within one cell cycle

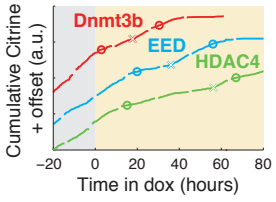
B



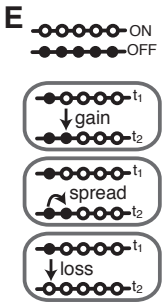
C



D



E



F

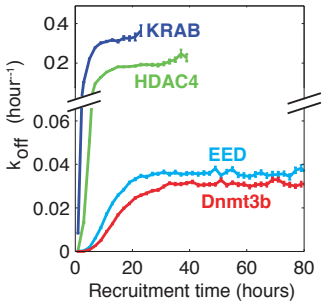


Figure B.2: (Caption on the following page.)

Figure B.2: Sister cell correlation, completeness of silencing, and deviations from the stereotypical silencing events
(A) Complete sister cell pairs that were tracked fully and originated from an ON parent cell were pooled from independent experiments and tabulated according to their silencing behaviors during this subsequent cell cycle. For each factor, the expected number and 95% confidence interval of sister cell pairs that both turned off were computed by a random permutation test with 100,000 trials. Red indicates cases in which it is possible to reject the null hypothesis of complete independence between sister cells. We did not apply the same analysis to KRAB and HDAC4, since most cells were silenced within the first cell cycle after dox addition in these cases.
(B) Distributions of H2B-citrine production rates in various time bins relative to dox addition (normalized to the median pre-dox production rate).
(C and D) Example of non-stereotypical silencing dynamics. Circles indicate detected silencing events, while x's indicate detected reactivation events.
(E) Stochastic silencing was modeled by allowing an array of nucleosomes to transition from unmarked (ON) to marked (OFF). The three rectangles show the types of possible transitions at each time step (from t1 to t2): random gain, spread, or loss of epigenetic marks at any nucleosome.
(F) The rates of silencing, $k_{off}(t)$, obtained from simulations are plotted against time. The loss rate is set to 0.01 (hour^{-1}) for all factors, while the other two rates are chosen to produce profiles similar to the experimental data recorded for each of the four factors. Parameters used in these simulations for each factor: gain=0.025, spread=1.6 (KRAB); gain=0.016, spread=0.8 (HDAC4); gain=0.0035, spread=0.5 (EED); and gain=0.0031, spread=0.4 (Dnmt3b)

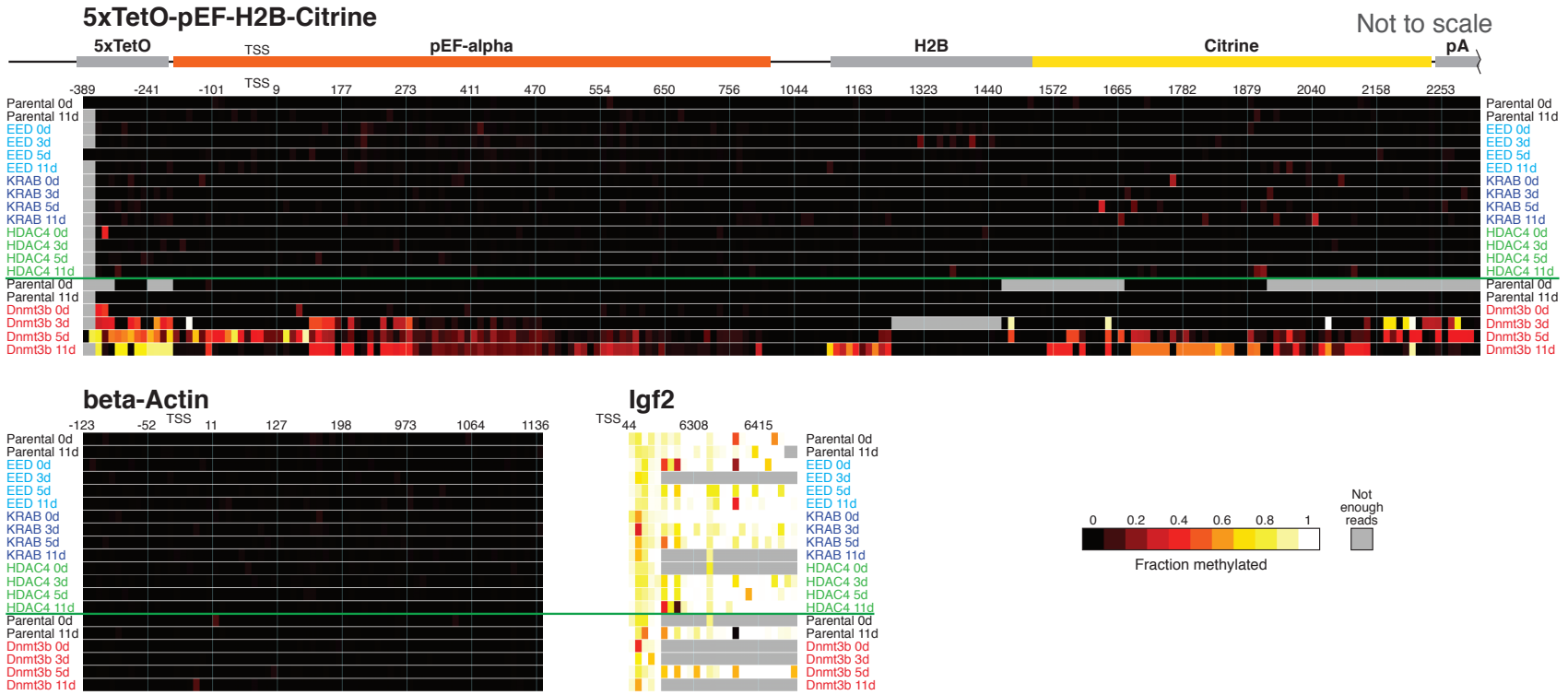


Figure B.3: Targeted bisulfite sequencing

Each of the factors was recruited for 0, 3, 5 and 11 days by dox treatment, followed by bisulfite sequencing targeting three different genomic regions: reporter (5xTetO-pEF-H2B-Citrine), negative control for DNA methylation (beta-Actin), and positive control for DNA methylation (Igf2). Each box represents the fraction of methylated reads at a particular CpG position for a given sample (averaged from two technical repeats). Grey indicates CpG sites with lower than 50 reads, which were excluded from the analysis.

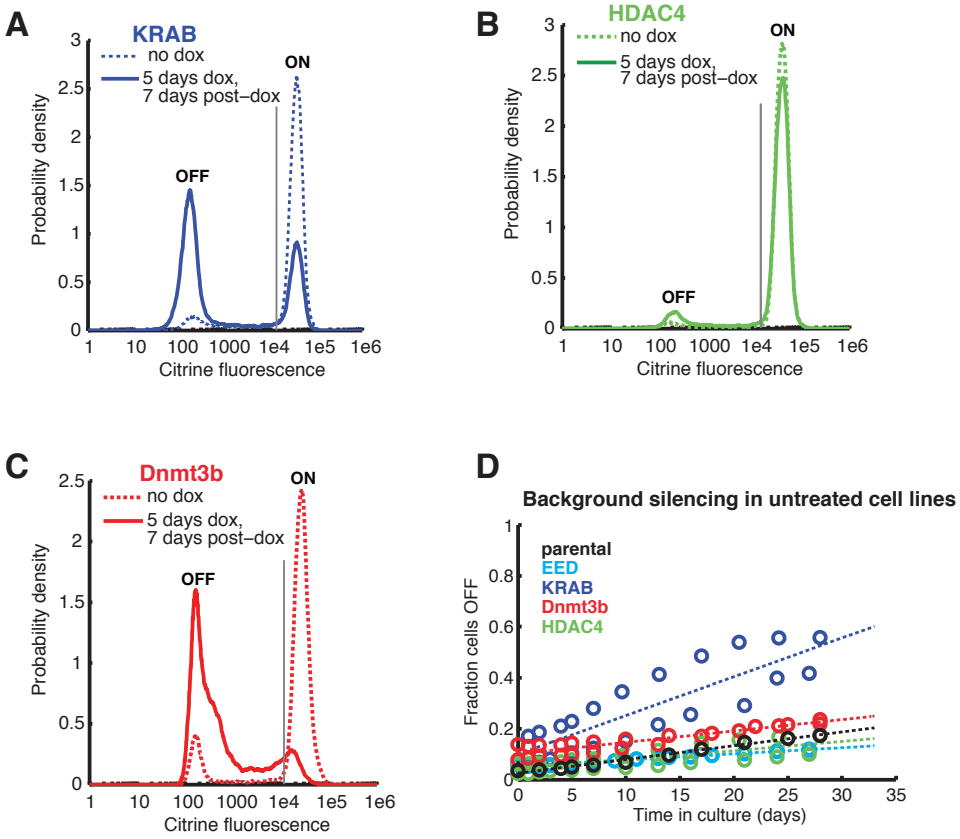


Figure B.4: Reactivation after silencing

Citrine fluorescence distributions were measured by flow cytometry for cell lines expressing rTetR fused to (A) KRAB, (B) HDAC4, or (C) Dnmt3b. Distributions are shown for cells treated with dox for 5 days, followed by 7 days of culture without dox (solid lines), and for cells cultured in parallel for 12 days with no dox (dashed lines). The vertical gray line in each panel represents the threshold used to determine the fraction of cells OFF.

(D) The fractions of cells OFF due to background silencing in the absence of dox were measured for all cell lines as a function of time. For each cell line, these background silencing values were subtracted from the total fractions of cells OFF to obtain the curves presented in Figures 3.4 & 3.5.

Locus	**Name**	**Sequence**
pEF (promoter)	pEF_F_1013ChIP	ACGTATGTCGAGGTAGGCGT
	pEF_R_1013ChIP	CTAGGCACCGGTTCAATTGC
citrine (gene body)	F_cit_Set2	CGGCGACGTAAACGGCCACAAGTTCAG
	R_cit_Set2	CTTGCCGGTGGTGCAGATGAA
actin (control)	bActin F	ACTGGGACGATATGGAGAAG
	bActin R	GGTCATCTTTTCACGGTTGG
Igf2 (control)	5-42_Igf2_F	CTGTGGCCTGTAGGTCCTTG
	5-43_Igf2_R	CCTCTGCCTTTCCCTCTTGG

Table B.1: Primers used for ChIP-qPCR and MeDIP-qPCR

Supplemental Bibliography – Appendix B

[1] Yamaguchi, S., Kazuki, Y., Nakayama, Y., Nanba, E., Oshimura, M., and Ohbayashi, T. "A method for producing transgenic cells using a multi-integrase system on a human artificial chromosome vector." *PLoS ONE* 6.2 (2011), e17267.

[2] Sprinzak, D., Lakhanpal, A., Lebon, L., Santat, L. A., Fontes, M. E., Anderson, G. A., Garcia-Ojalvo, J., and Elowitz, M. B. "Cis-interactions between Notch and Delta generate mutually exclusive signalling states." *Nature* 465.7294 (2010), pp. 86–90.

[3] Weber, W., Fux, C., et al. "Macrolide-based transgene control in mammalian cells and mice." *Nature Biotechnology* 20.9 (2002), pp. 901–907.

[4] Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S., and Elowitz, M. B. "Gene regulation at the single-cell level." *Science* 307.5717 (2005), pp. 1962–1965.