Understanding Cell State Transitions in Development and Disease

Phuc Huu Ba Nguyen

A dissertation

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Hao Kueh, Chair

Raymond Monnat

Jesse Zalatan

Program Authorized to Offer Degree:

Molecular Engineering and Sciences

University of Washington

Abstract

Understanding Cell State Transitions in Development and Disease

Phuc Huu Ba Nguyen

Chair of Supervisor Committee:

Hao Kueh

Department of Bioengineering

During development, eukaryotic cells undergo a series of state transitions that transform a stem cell into a diverse set of differentiated progenitors with distinctive biological functions. Disruption of these processes can drive cancer and autoimmune diseases. Yet it is incompletely understood what molecular mechanisms lead to maintenance and transition of cellular states and how scientists can effectively identify these states in novel biological systems. This thesis aims to address these challenges using hematopoietic cells as a model system. I start by investigating an epigenetic mechanism controlling activation timing of the fate determining gene Bcl11b during early T-cell development (Chapter I). I then explore the potential of such mechanism to generate diverse temporal schedules and patterns of population size control during development (Chapter II). Lastly, I build a machine learning workflow for the discovery and exploration of phenotypic states and their dynamics from brightfield movies of unmanipulated cells. This technique is applied to identify phenotypic states in primary patient acute myeloid leukemia stem cells during differentiation (Chapter III). This work lays the foundation for further understanding of how the timing of mammalian gene expression is regulated. Finally, it provides a robust method for recognizing important cellular phenotypes associated with development and disease.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENT

First, I would like to thank my mentor **Dr. Hao Yuan Kueh**. You introduced me to the world of quantitative biology, and your passion for understanding biological systems inspires me to work hard every day. Thank you for giving me the liberty to explore and excel at different machine learning techniques and turn them into useful tools for my research. Your feedbacks on my work and my academic writing style have helped me tremendously over the years. For all your help, I am ever more confident in myself as a scientist and forever more curious and excited about biological research.

To my co-mentors, **Dr. Pamela A Becker** and **Dr. Raymond Monnat**, for your invaluable insights and feedbacks into biological systems I am studying. You always remind me of how complex and intriguing cells can be and thank you very much for taking me on the research effort to understand and fight cancer, the insidious disease that claimed the lives of multiple family members of mine. I am humbled, and honored, to know that my research effort today may one day be translated into improving the lives of people suffering from it.

To the rest of my Supervisory Committee Members, **Dr. Jesse G Zalatan** and **Dr. Sreeram Kannan**, for providing my with invaluable insights and thoughts on my research throughout my graduate school journey. Your feedbacks have been instrumental for making my thesis the best version that it could be.

To my collaborators **Nick Pease**, **Kenneth Ng**, **Syliva Chien**, and **Jin Dai**, for your tireless work beside me, providing me with assistance when needed, regardless of the time of the day or week. My work would not be where it is without your help and support. Similarly, to my undergraduate assistant **Blythe Irwin**, for your excitement and passion for doing scientific research. Your optimism is always so uplifting and working with you is always so much fun.

To the rest of the Kueh lab members, **Matthew Wither**, **Kathleen Abadie**, **Jon Chu**, **Elisa Clark**, and **Will White**, for sharing feedbacks about my work, interesting and exciting scientific discussions, and many fun lab hangout. You are my great friends, and I am honored to work alongside you over the years.

And finally to my parents, **Trung Huu Nguyen** and **Mai Thi Tuyet Dang**, and my other family members for believing in me and supporting me throughout my graduate and undergraduate careers.

DEDICATION

This thesis is dedicated to my wonderful wife and life partner, **Jennifer Rose Meeker Dixon**, for always being by my side and encouraging me to pursue my dream despite all the stresses and challenges of graduate school. I could not have done this without you.

INTRODUCTION

During development, eukaryotic cells undergo a series of state transitions that transform a stem cell into a diverse set of differentiated progenitors with distinctive functions. Disruption of this process underlies various disorders such as cancer and autoimmune diseases. Yet it is incompletely understood which molecular mechanisms are responsible for the maintenance and transition of cellular states, and what the implications of such mechanisms are for multicellular development and function. Furthermore, it remains challenging to identify and monitor cell state transitions. This is especially true for primary cell populations, as the dynamics can be highly variable from cell-to-cell.

My thesis aims to address these challenges using hematopoietic cells as a model system. First, I dissect the molecular mechanisms behind an epigenetic timer controlling the activation timing of a lineage-specifying gene during early T-cell development. By integrating experiments and modeling, I describe a switching mechanism which involves active removal of repressive histone H3K27 trimethylation. This weakens inter-nucleosomal contacts at the Bcl11b locus, leading to gene decompaction and activation. The mechanism generates gene activation delays that are independent of cell cycle length and are tunable by both chromatin regulators and transcription factors. I propose that this division-independent epigenetic switch, with its robust operation over long timescales and regulatory flexibility, is broadly utilized to control differentiation timing in multicellular eukaryotes.

I then investigate how this mechanism can be employed to enable developmental population size control while preserving proportions for different cell types. I show that networks of timed epigenetic switches robustly uphold complex timetables for gene activation and lineage specification. Through a series of evolutionary case studies, I present evidence that mutations in epigenetic switching network can adjust developmental schedules to engender variation in organism sizes and proportions across species. These findings suggest that epigenetic switching networks represent central targets through which evolution manufactures diversity in organism size and form.

Finally, I designed a method for systematic characterization of phenotypic states and their dynamics from timelapse imaging of unmanipulated cells. This technique is applied profile morphological state changes in primary leukemic stem cells (LSCs) in acute myeloid leukemia (AML) during differentiation and drug tolerance. I identified, without human supervision, distinct morphological states associated with different blood cell types and elucidated important morphological features underlying these states. I used this technique to analyze the self-renewal dynamics of leukemic stem cells (LSCs) derived from an acute myeloid leukemia (AML) patient. Through our analysis, I identified key interpretable morphological features associating with stemness and differentiation, clustered them into distinctive phenotypic states, and elucidated the rates of interconversion between these states. These results demonstrate the utility of this method as a tool for unbiased exploration of cellular morphologies and their dynamics from large, time-resolved imaging datasets.

CHAPTER I. TUNABLE, DIVISION-INDEPENDENT CONTROL OF DEVELOPMENTAL TIMING BY AN EPIGENETIC SWITCH

1.1 INTRODUCTION

During multicellular development, stem and progenitor cells often differentiate many days and cell divisions after receiving instructive signals. These differentiation delays must be robust, yet tunable over timescales spanning multiple cell generations, for precise control over differentiated population sizes. In diverse contexts, cell differentiation delays are generated by timing mechanisms operating autonomously in single cells. In classic studies of oligodendrocyte differentiation, precursor cells exposed to signals delayed their differentiation by up to eight cell divisions, due to a cell-autonomous timing mechanism (Gao et al., 1997; Temple and Raff, 1986). Such autonomous timing control is seen in diverse systems, from brain and muscle development to adaptive immunity (Burton et al., 1999; Heinzel et al., 2017; Otani et al., 2016).

While timing delays in some embryonic systems are generated by mechanisms that count cell divisions (Amodeo et al., 2015; Newport and Kirschner, 1982), differentiation delays during later vertebrate development or in adult stem cells are often uncoupled from cell cycle progression, such that changes to rates of cell division do not affect delay duration (Burton et al., 1999; Gao et al., 1997; Heinzel et al., 2017; Li et al., 2019; Okamoto et al., 2016; Osmond, 1991; Otani et al., 2016). A mechanism for setting the elapsed time to differentiation apart from cell division could provide functional advantages to cells, including operation in non-dividing cells and an ability to modulate cell expansion while maintaining a constant temporal schedule for differentiation. However, it is unknown how division-independent timing control is implemented on a molecular level.

Polycomb repressive mechanisms, involving histone H3 lysine-27 trimethylation (H3K27me3), are important for differentiation timing control. H3K27me3 modifications are often found at the loci of lineage-specifying genes in stem cells, where they restrain the expression of these genes and resultant differentiation (Boyer et al., 2006; Lee et al., 2006). During differentiation, instructive signals activate transcription factors that bind to lineage-specifying genes and initiate H3K27me3 removal. However, while transcription factors usually bind rapidly upon signal exposure within minutes, H3K27me3 loss and gene activation often occur much more slowly, such that gene loci can heritably maintain a silent state over multiple cell generations prior to activation (Berry et al., 2017; Kaikkonen et al., 2013; Mayran et al., 2018). In a prevailing view, this epigenetic maintenance of the silent state before activation results from the passive dilution of H3K27me3 modified histones through serial cell division (Gaydos et al., 2014; Jadhav et al., 2020; Sun et al., 2014). However, while passive dilution mechanisms can delay differentiation over multiple cell divisions, these delays would depend on cell cycle duration and be inconsistent with the division-independent timers found in vertebrates. Alternatively, H3K27me3 loss may be actively controlled by the opposing PRC2 methyltransferase and Kdm6a/b demethylase activities (Li et al., 2014; Park et al., 2014; Seenundun et al., 2010). However, it is unclear whether such active mechanisms could generate differentiation delays spanning many cell divisions and whether these delays could be both tunable and cell-division independent.

To address these questions, we investigated the mechanism of a time-delayed epigenetic switch controlling the activation of Bcl11b, a transcription factor essential for T cell lineage commitment and identity (Hosokawa et al., 2018; Ikawa et al., 2010; Li et al., 2010) (Figure 1A). In early T cell progenitors, Notch signals activate Bcl11b, both directly (Ikawa et al., 2010; Li et al., 2010) and indirectly, by activating its upstream regulators Gata3 and TCF-1 (García-Ojeda et al., 2013; Germar et al., 2011; Zhou et al., 2019). However, while these upstream regulators become active shortly after thymic entry, Bcl11b activation and T cell lineage commitment occur ~5-10 days later, during which progenitors proliferate a thousand-fold (Manesso et al., 2012; Porritt et al., 2003; Zhou et al., 2019). Using a dual-allele Bcl11b reporter strain, where each endogenous gene copy is tagged with distinguishable fluorescent protein reporters (Ng et al., 2018), we found that this long delay in Bcl11b activation arises partly because of an epigenetic switch acting at individual Bcl11b loci, in cis. This epigenetic switch activates probabilistically, with a multi-day time constant that is controlled by Gata3 and TCF-1 (Kueh et al., 2016), together with a distal enhancer to which these factors bind (Ng et al., 2018). However, our previous studies did not clarify the mechanism of the timed epigenetic switch itself or its basis for generating controllable timing delays in gene activation. We address these outstanding questions here.

1.2 A TIMED EPIGENETIC SWITCH DELAYS BCL11B ACTIVATION AND T CELL LINEAGE COMMITMENT

The dual-allelic Bcl11b reporter system we previously described provides a unique tool to resolve epigenetic switching dynamics in living cells (Ng et al., 2018). To study the cis- epigenetic event controlling Bcl11b activation timing in isolation from other events occurring in trans, we analyzed Bcl11b locus activation dynamics in progenitors that already have one Bcl11b allele active and must therefore contain all trans-factors necessary for expression (Figure 1.1B). Using fluorescence-activated cell sorting (FACS), we purified monoallelic expressing Bcl11b DN2 progenitors from dual-allelic reporter mice (RFP+/YFP- or RFP-/YFP+) and analyzed the activation of the silent allele by co-culture with OP9-DL1 cells, an in vitro system that recapitulates early transitions in T-cell development (Holmes and Zuniga-Pflucker, 2009). In agreement with previous work (Ng et al., 2018), inactive Bcl11b alleles turned on after a long time delay, such that the fraction of expressing cells increased progressively over the course of five days (Figure 1.1C-D). Activation kinetics were similar for both YFP and RFP alleles ($k = 0.025 \pm 0.005$ hrs^{-1} for YFP allele and $k = 0.034 \pm 0.009$ hrs^{-1} for RFP allele) and were well-described by a single exponential curve, consistent with activation being controlled by a single stochastic event occurring with equal likelihood at each allele.

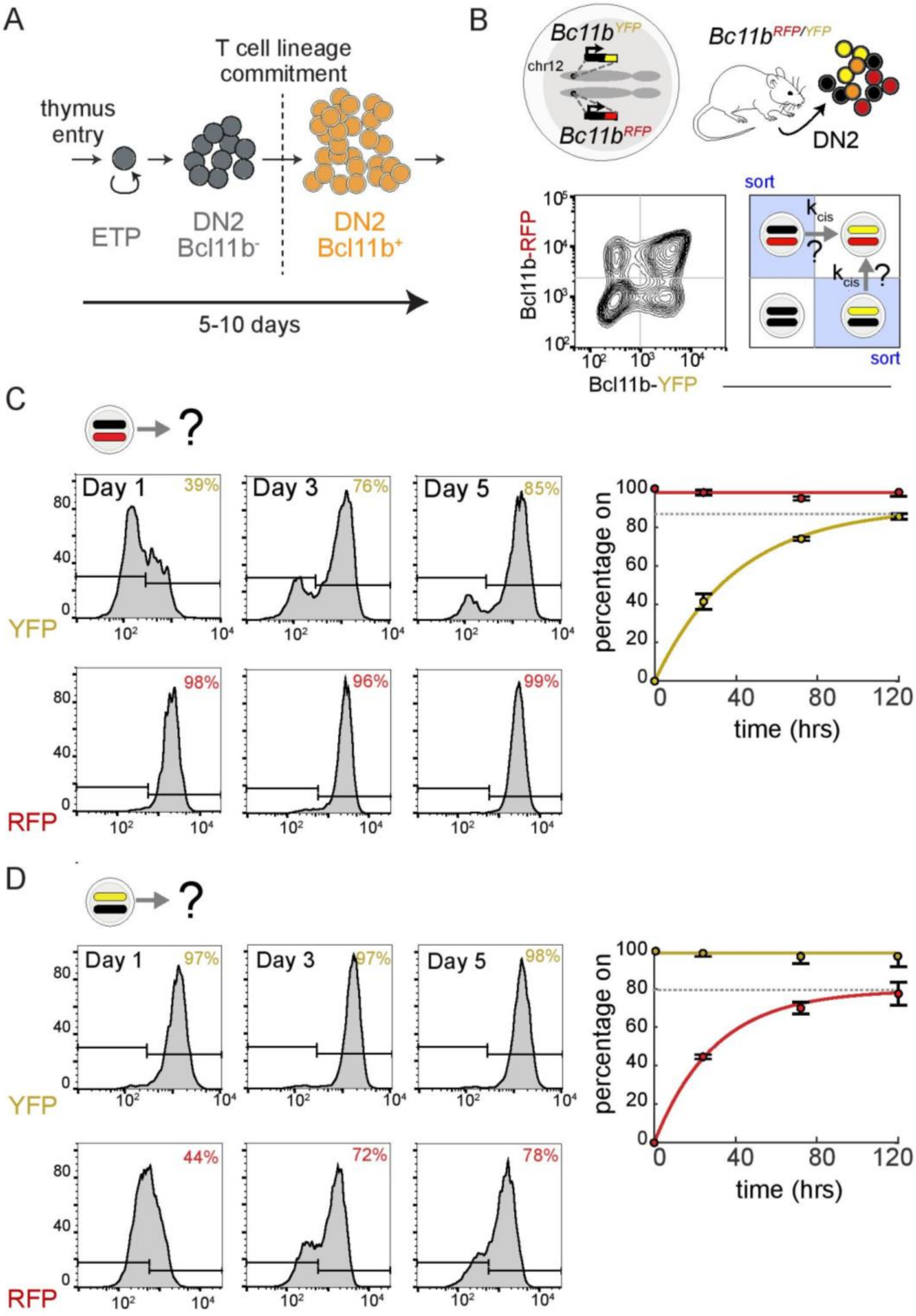


Figure 1.1. A timed epigenetic switch delays Bcl11b activation and T cell lineage commitment. (A) Schematic of early T cell development. Delayed activation of the T

cell commitment transription factor, Bcl11b, enables early progenitor expansion. (B) Dual-allelic Bcl11b reporter mouse (top), along with flow cytometry plot showing levels of each Bcl11b allele in DN2 progenitors (bottom left), along with strategy to purify monoallelic progenitors for live-cell analysis of timed epigenetic switch timing, k_{cis}. (C-D) DN2 monoallelic progenitors were purified by cell sorting, cultured on OP9-DL1 feeders with 5 ng/mL IL-7 and Flt3L, and analyzed by flow cytometry at the indicated timepoints. Mean values and 95% confidence intervals are plotted for n = 3 independent experiments. Curves represent fits to the equation $y = F(1-e^{-kt})$, where F is the final percentage of cells positive for assayed allele (represented by the dotted grey lines); $k = 0.025\ hrs^{-1}$ +/- 0.005 for Bcl11b-YFP and $k = 0.034\ hrs^{-1}$ +/- 0.009 for Bcl11b-RFP.

1.3 H3K27ME3 MODIFICATIONS TUNE BCL11B ACTIVATION TIMING

The repressive histone modification H3K27me3 is highly enriched at silent Bcl11b loci in hematopoietic progenitor cells but not in committed T-cells, where Bcl11b is expressed (Zhang et al., 2012); thus, its removal could regulate the epigenetic event controlling Bcl11b activation timing. To test this possibility, we first determined whether H3K27me3 marks are removed from the Bcl11b locus at the same time it turns on. To pinpoint when H3K27me3 loss occurs relative to locus activation, we measured H3K27me3 levels in three populations having different numbers of active Bcl11b alleles: hematopoietic progenitors from bone marrow, which have both Bcl11b alleles inactive; monoallelic Bcl11b expressing DN2 progenitors; and biallelic Bcl11b expressing DN2 progenitors. We employed CUT&RUN, a nuclease-based method for mapping DNA-protein complexes that can be combined with spike-in controls to provide quantitative readouts of H3K27me3 genomic abundance comparable across samples (Skene et al., 2018). If H3K27me3 marks are removed concurrently with Bcl11b activation, but not any sooner or later, we would expect H3K27me3 levels in monoallelic DN2 progenitors to fall to approximately half of the initial levels found in HSPCs and to approximately zero in biallelic DN2 progenitors. In bone marrow progenitors, where both Bcl11b alleles are inactive, there was an abundance of H3K27me3 across the 5’ end of Bcl11b (Figure 1.2A, yellow shaded region). These broad H3K27me3 peaks were roughly halved in monoallelic Bcl11b expressing cells and were almost completely absent in Bcl11b biallelic cells (Figure 1.2A). While these results suggest that removal of the repressive H3K27me3 modification from the Bcl11b locus occurs concurrently with gene activation, it remains unclear if H3K27me3 removal is the cause of activation.

H3K27me3 loss may control the timing of Bcl11b activation; alternatively, it may simply be a consequence of gene activation, due to clearance of methylated nucleosomes by active transcription (Hosogane et al., 2016; Kraushaar et al., 2013). To determine whether H3K27me3 modifications play a causal role in controlling Bcl11b activation timing, we cultured monoallelic Bcl11b expressing DN2 progenitors with small molecule inhibitors targeting H3K27me3-modifying enzymes, and analyzed the effects on activation of the silent Bcl11b allele. These inhibitors, which target either the PRC2 methyltransferase subunit Ezh2 (UNC1999, 3 μM) or the H3K27 demethylases Kdm6a/b (GSK-J4, 3 μM) (Figure 1.2B), resulted in a ~40% increase and ~60% decrease, respectively, in H3K27me3 abundance at the Bcl11b promoter in Bcl11b monoallelic cells (Figure 1.2C), indicating that they indeed modulate H3K27me3 levels at inactive Bcl11b loci.

To determine whether H3K27me3 levels regulate Bcl11b activation timing, we assayed

Bcl11b reporter expression by flow cytometry over the course of four days. In the absence of any inhibitors, the silent Bcl11b alleles in monoallelic progenitors activated at an average rate of 0.022 hrs^{-1} and 0.019 hrs^{-1} for YFP and RFP alleles, respectively (Figure 1.2D-F). Kdm6a/b demethylase inhibition decreased the activation rate for each silent allele in graded manner (27% reduction in YFP activation rate, $p = 5.9 \times 10^{-4}$; 32% reduction in RFP activation rate, $p = 5.9 \times 10^{-3}$). Conversely, Ezh2 inhibition increased the activation rate for each silent allele (41% increase in YFP activation rate, p = 3.4x10-5; 58% increase in RFP activation rate $p = 1.5 \times 10^{-3}$). Similar graded decreases or increases in Bcl11b activation probabilities were observed for another structurally unrelated Kdm6a/b inhibitor, IOX-1, and other Ezh2 inhibitors GSK-126 and GSK-343 (Figure S1.1A), arguing against off-target effects of the inhibitors. Furthermore, short hairpin RNA mediated knock-down of another essential PRC2 subunit, Eed, significantly increased the activation rate of silent Bcl11b alleles, arguing against non-specific pharmacological effects. Notably, all H3K27me3 perturbations tested affected Bcl11b in a graded manner, indicating that these modifications exert tunable control over activation rates. Furthermore, none of these perturbations affected Bcl11b expression levels after activation (Figure S1.1A) nor did they have any differential effect on apoptosis frequency among different populations (Figure S1.1B), indicating a specific role of these modifications in controlling gene activation timing. Taken together, these results show that the time constant for Bcl11b activation is tunably set by the balance between PRC2 and Kdm6a/b demethylase activities at the gene locus.

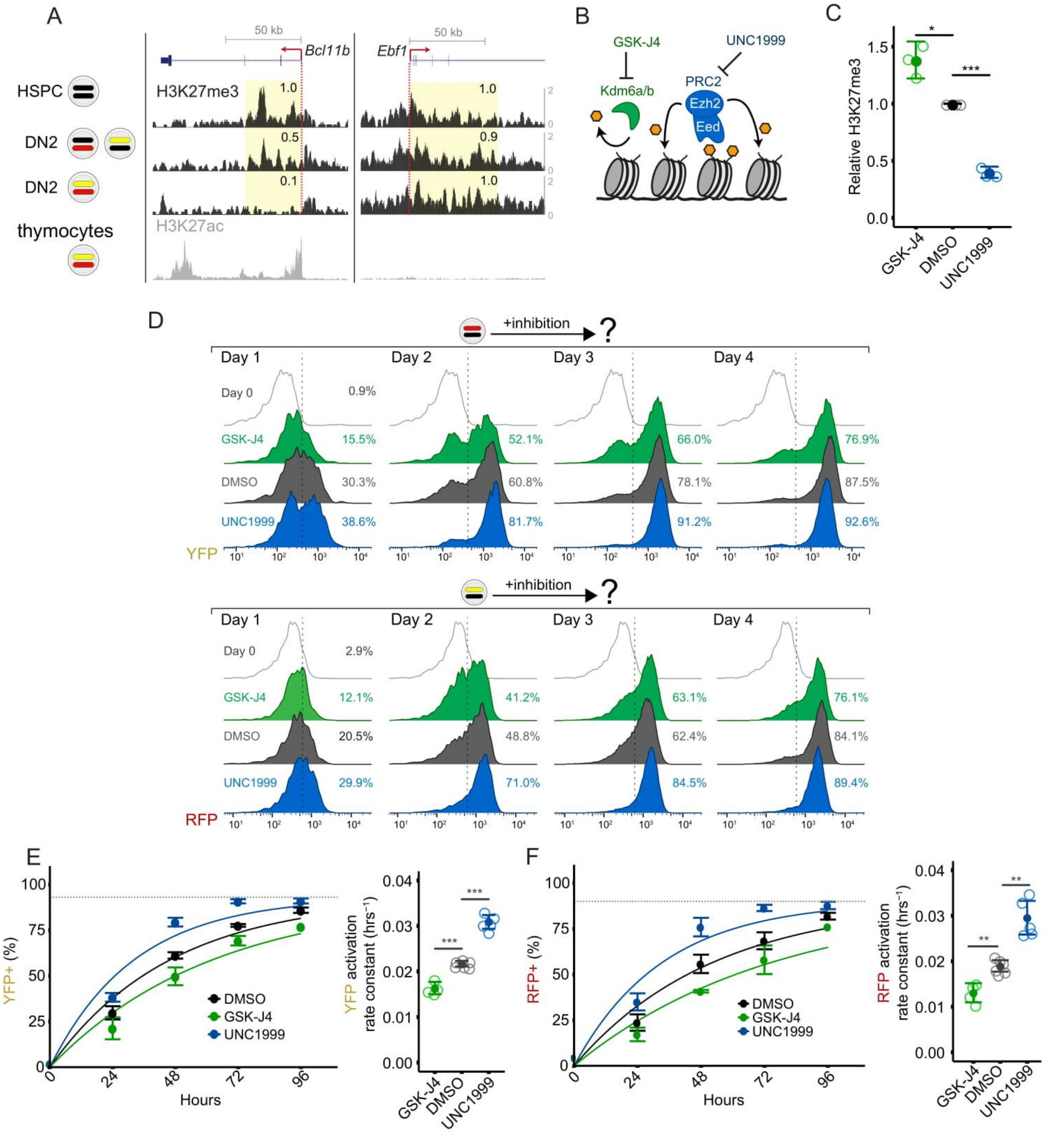
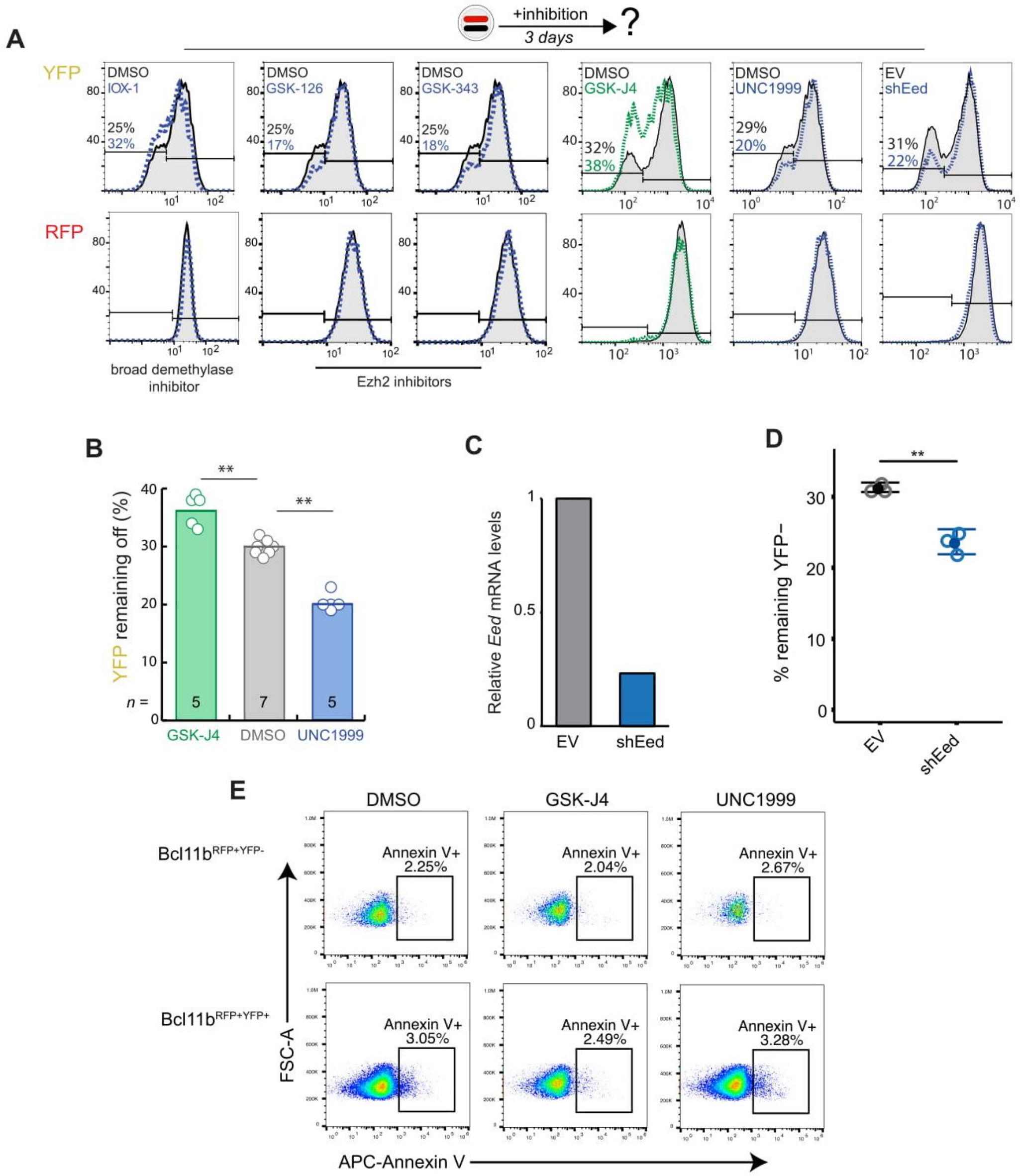


Figure 1.2: Histone H3K27me3 modifications tunably control epigenetic switch timing. (A) H3K27me3 distributions were profiled by CUT&RUN in Lin- bone marrow progenitors (HSPCs), as well as purified DN2 monoallelic and biallelic *Bcl11b* expressing cells with UCSC Genome Browser tracks showing H3K27me3 densities at *Bcl11b*, as well as at *Ebf1*, a B cell regulator that is repressed during T cell development. Relative read densities of shaded areas are shown. H3K27ac levels in thymocytes, obtained from a previ-ous study (Davis et al., 2018; Accession: ENCSR000CCH), are shown. Data are representative of two independent experiments. (B) Schematic depicting inhibiton of H3K27 demethylases Kdm6a/b or H3K27 methyltransferase PRC2. (C) DN2 monoallelic progenitors treated with the indicated inhibitors were sorted for anti-

H3K27me3 CUT&RUN followed by qPCR at the *Bcl11b* promoter. Mean values are presented with $n = 3$ experimental replicates (two-sample t-test, one-tailed: * $p < 0.05$, *** $p < 0.001$). (D) Purified DN2 monoallelic cells were recultured with the indicated inhibitors over the course of 4 days and analyzed by flow cytometry. Histograms show results from one representative experiment. (E-F, left) Mean activation percentages and 95% confidence intervals are plotted with curves representing fits to the equation $y = F(1-e^{-kt})$, where F = maximum percentage of cells positive for assayed allele (represented by the dotted grey lines). (E-F, right) Mean rate constants, k, with 95% confidence intervals (two-sample t-test, one-tailed: ** $p < 0.01$, *** $p<0.001$; n = 4-6 independent experiments).



Supplementary Figure S1.1. Perturbation of activation timing of Bcl11b by small molecules affecting epigenetics modifiers. (A) DN2 Bcl11bRFP+YFP- monoallelic cells were sorted, re-cultured in the presence of different small molecule inhibitors and analyzed by flow cytometry 3 days later. Structurally unrelated inhibitors show similar effects of decreaseing or increasing *Bcl11b* activation probabilities as observed in Figure 2E-F (*left*). All histograms show that while the perturba-tions affect the all-or-none activation probability for the initially inactive alleles (*top*), the pertubations have no effect on the expression maintenance nor magnitude of the initially active alleles (*bottom*). (B) Mean percentage of cells remaining Bcl11b YFP-negative (*n* = independent experiments, *p < 0.05, **p < 0.01, two-sample t-test, two-tailed). (C) Relative mRNA levels of *Eed* were measured by qPCR. (D) Mean percentage of cells remaining YFP-negative after DN2 Bcl11bRFP+/YFP- monoallelic cells were transduced with retroviral constructs and recultured for 3 days (two-sample t-test, one-tailed, $**p < 0.01$, $n = 3$ independent experiments, error bars = 95% confidence interval). (E) Small molecule inhibitors have no effect on the frequency of apoptotic cells as measured by Annexin V staining.

1.4 BCL11B ACTIVATION TIMING IS REGULATED INDEPENDENTLY OF CELL DIVISION

Activation delays of polycomb-repressed genes have been proposed to result from the passive dilution of H3K27me3 modified histones with cell division (Coleman and Struhl, 2017; Jiang and Berger, 2017; Strome et al., 2014). However, the observed regulation of Bcll1b activation timing by H3K27 demethylases that counteract PRC2-mediated methylation suggests an alternative mechanism, where gene activation timing may be regulated independently from cell division (Figure 1.2E). To determine whether Bcl11b activation timing depends on cell division or not, we accelerated the cell division rate in DN2 RFP+/YFP- progenitors by transducing them with the proto-oncogene c-Myc, then used quantitative live-cell imaging to measure the activation kinetics of the silent YFP allele (Figure 1.3A-B). While a passive dilution mechanism would result in activation times that increase with the cell division rates, an active mechanism could potentially allow for activation times to be uncoupled from cell division. As expected, c-Myc overexpression resulted in a ~two-fold increase in the cell division rate (Figure 1.3C, left; see also Figures S1.2, Supplementary Table 1.1 and Quantitative and Statistical Analysis). However, despite accelerating cell division, c-Myc transduction did not change Bcl11b activation timing, with control and accelerated progenitors activating the silent Bcl11b allele with indistinguishable dynamics, and with the same exponential time constant (~136 hrs, Figure 1.3C, right). Therefore, in contrast to the passive dilution paradigm, these results show that the epigenetic switch controlling Bcl11b activation generates time delays in gene activation that are independent of cell division.

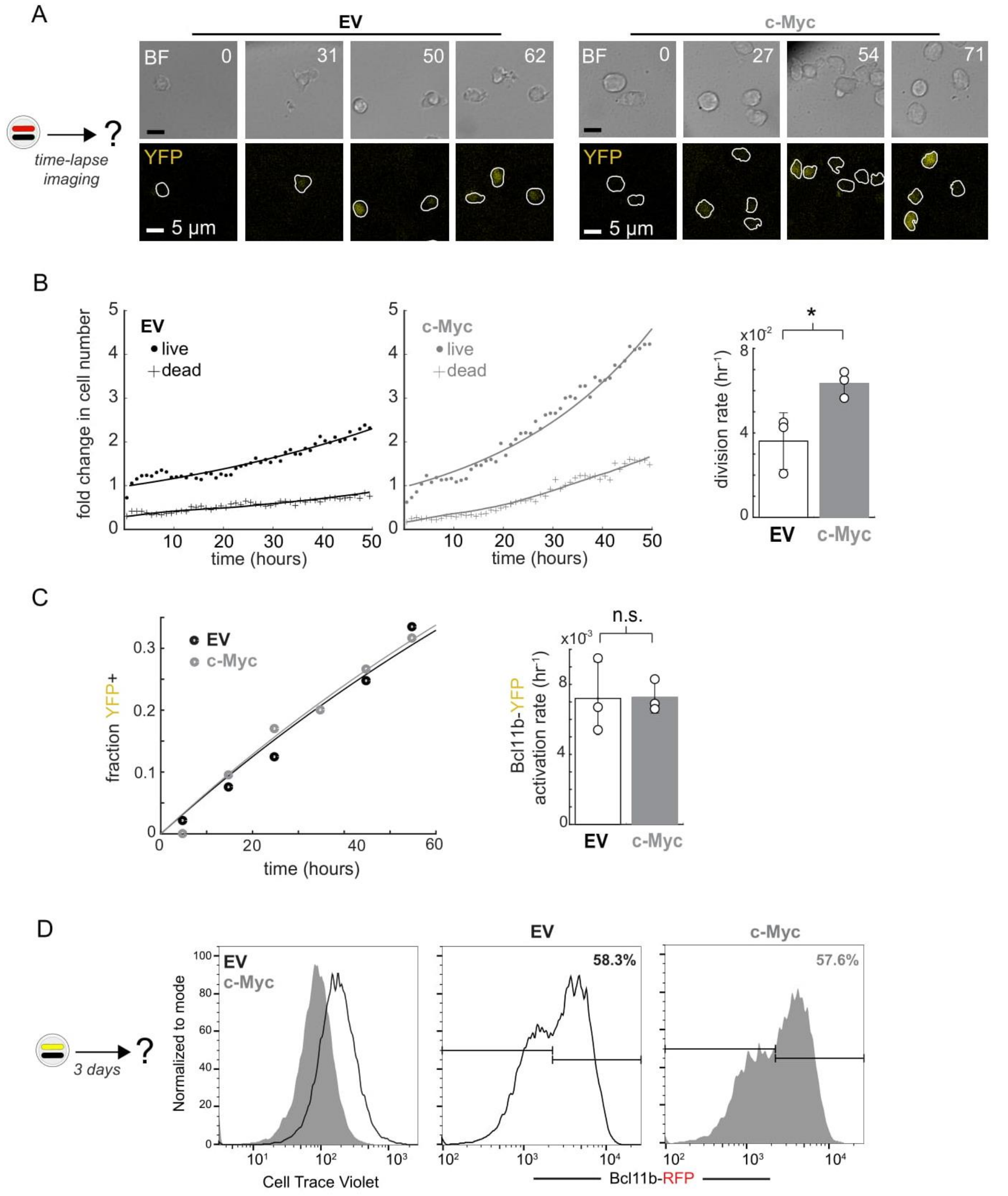
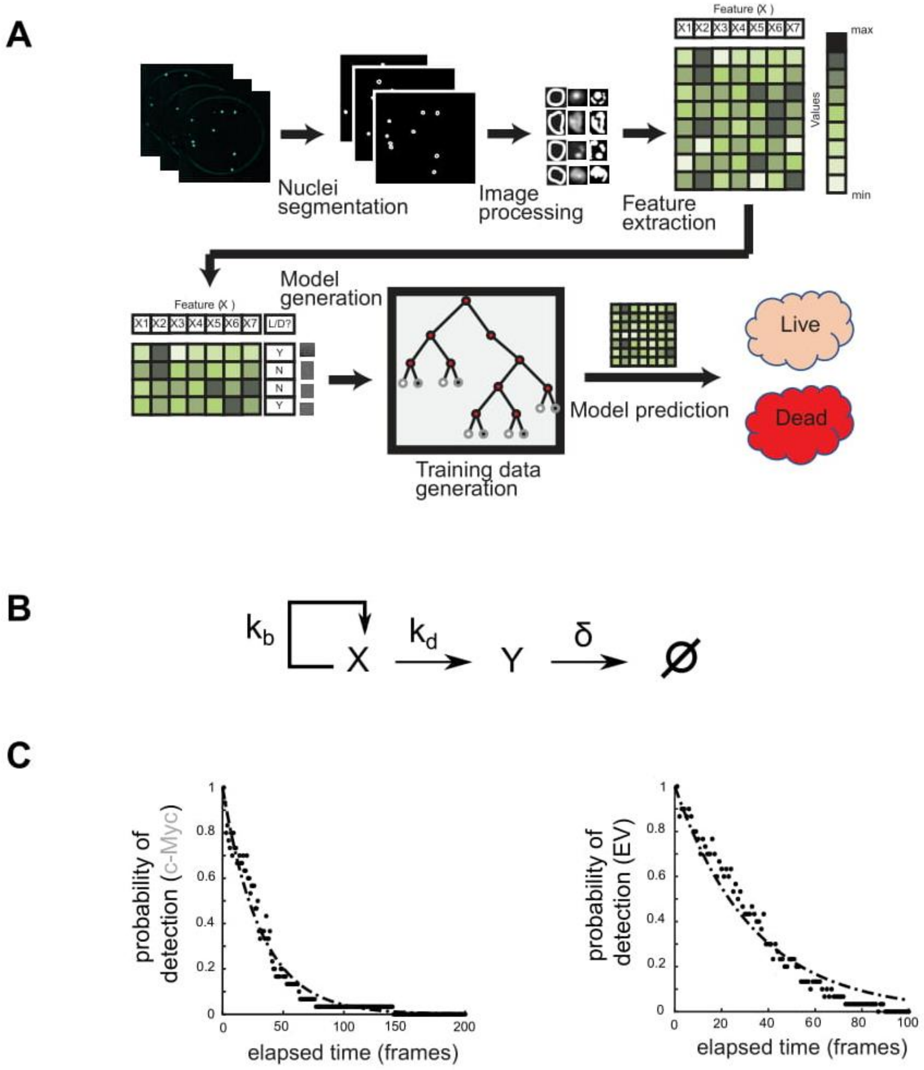


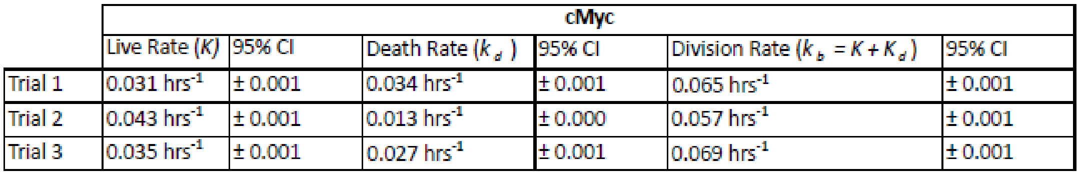
Figure 1.3. Epigenetic switch timing is regulated independently from cell division. Bcl11bYFP-/RFP+ DN2 progenitors transduced with either a MSCV empty vector (EV)

or MSCV c-Myc overexpression vector were purified and re-cultured on DL1-retronectin coated plates and monitored by timelapse imaging. (A) Timelapse images show brightfield (BF) and YFP fluorescence of progenitors. White boundaries show automated cell segmentation. Numbers show elapsed time in hours. (B, left) Time evolution of live and dead cell numbers from one representative imaging experiment. Data was fitted to population dynamics as modeled in Figure S2. (B, right) Mean and standard deviation of cell division rates (n = 3 independent experiments, two-sample t-test, one-tailed, * $p < 0.025$). (C, left) Fraction of cells Bc11b YFP+ over time for one representative experiment. (C, right) Mean and standard deviation of Bcl11b YFP activation rates (n = 3 independent experiments, two-sample t-test, one-tailed, n.s. not significant).

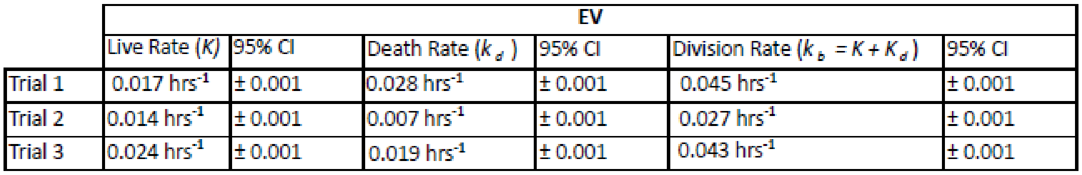


Supplementary Figure S1.2. Computational method for analyzing Bcl11b activation

timing at a single cell level using live imaging. (A) Cyan fluorescent protein (CFP) signals from transduced cells are used for segmentation of individual cells. Segmented populations were analyzed using imaging processing tools in MATLAB, and specific features such as object area, perimeter, and CFP intensity were collected. Approximately 10% of the total individual cells were then manually labeled as live or dead and fed into classification tree machine learning algorithm to generate a classification model. The rest of segmented cells are classified subsequently via the trained model. (B) Mathematical model describing the population dynamics of Bcl11bRFP+/YFP- cells transduced with c-Myc and empty vector (EV). The model includes a population of live cells (X) with a birth rate kb and a death rate kd that generates the dead cell population (Y). This population then has a permanent clearance rate of δ, indicating the process in which CFP degrades in these cells and the dead cells become undetected. (C) Experimentally determined decay of dead cells' detectability in time-lapsed movies. Individual dead cells were followed until their CFP level completely diminished and was undetectable by the segmentation algorithm, and the elapsed time was recorded. Thirty different individuals cells were recorded for each c-Myc and EV populations. Data was fit to an exponential decay function $P(\tau) = e^{-\delta\tau}$ with $\delta = 0.032$ per frame for c-Myc and $\delta = 0.030$ for EV for EV population. (D) Bcl11bRFP+/YFP- transduced with either EV or c-Myc were recultured on OP9-DL1 stromal monolayers for 3 days before analyzing by flow cytometry.



	cMyc					
	Live Rate (K)	95% CI	Death Rate (k_d)	95% CI	Division Rate ($k_b = K + K_d$)	95% CI
Trial 1	0.031 hrs^{-1}	± 0.001	0.034 hrs^{-1}	± 0.001	0.065 hrs^{-1}	± 0.001
Trial 2	0.043 hrs^{-1}	± 0.001	0.013 hrs^{-1}	± 0.000	0.057 hrs^{-1}	± 0.001
Trial 3	0.035 hrs^{-1}	± 0.001	0.027 hrs^{-1}	± 0.001	0.069 hrs^{-1}	± 0.001



	EV					
	Live Rate (K)	95% CI	Death Rate (k_d)	95% CI	Division Rate ($k_b = K + K_d$)	95% CI
Trial 1	0.017 hrs^{-1}	± 0.001	0.028 hrs^{-1}	± 0.001	0.045 hrs^{-1}	± 0.001
Trial 2	0.014 hrs^{-1}	± 0.001	0.007 hrs^{-1}	± 0.001	0.027 hrs^{-1}	± 0.001
Trial 3	0.024 hrs^{-1}	± 0.001	0.019 hrs^{-1}	± 0.001	0.043 hrs^{-1}	± 0.001

Supplementary Table S1.1. Tabulated doubling (K) and death (k_d) rates calculated from data fitting of live and dead populations from three independent imaging experiments. Data was fitted to population dynamics model described in Statistical and Quantitative Analysis section.

1.5 A METHYLATION-COMPACTION MODEL FOR TUNABLE, DIVISION-INDEPENDENT TIMING CONTROL

The molecular mechanism underlying the timed Bcl11b epigenetic switch must account for its observed emergent properties, namely: 1) its ability to robustly set time delays that span multiple cell generations; 2) its stochastic nature; 3) its tunability by histone-modifying enzyme activities; and 4) its cell division independence. To identify mechanisms with these emergent properties, we analyzed a series of candidate mechanisms using mathematical modeling. H3K27me3 can bind PRC2 at an allosteric site to stimulate its methyltransferase activity (Margueron et al., 2009), a cooperative mechanism that is thought to maintain repressive H3K27me3 marks across cell division. Therefore, we first considered a simple model in which individual nucleosomes in a one-dimensional array undergo H3K27 methylation (catalyzed by the presence of nearby methylated nucleosomes), demethylation, as well as H3K27me3 loss due to random nucleosome segregation during DNA replication (Figure 1.4A, left; see Mathematical Appendix) (Coleman and Struhl, 2017). Similar models have been shown to support multi-stable histone modification states that are heritable across cell division (Dodd et al., 2007; Zhang et al., 2014).

From simulations, we found that single loci could switch from a repressed H3K27 methylated state to a demethylated state with stochastic time delays spanning multiple cell divisions (Figure 1.4A, center). However, in our simulations, the average activation timing was extremely sensitive to H3K27 methylation levels in the silent state (Figure 1.4B). The sensitivity coefficients far exceeded those derived from experimental data (methylation-only model s ~14 vs. experimental results s ~ 0.5-1.0 derived from Figures

1.2C and 1.2F) such that minor changes in methylation levels (~10%) caused dramatic changes activation timing (~300 fold) (Figure 1.4B). This extreme sensitivity was found across different length scales of methylation enzyme L action (Figure S1.3A-B), and was also generally observed in other studies (Dodd et al., 2007; Zhang et al., 2014), indicating that it represents a general feature of such switching models. By analyzing this system using a transition state theory framework (see Mathematical Appendix and Figure S1.3B-C), we found that switching times scale exponentially with methylation or demethylation rates, thus explaining the observed extreme sensitivity. Thus, models that consider histone modification dynamics alone are inconsistent with the tunable control of activation by H3K27me3-modifying enzymes observed experimentally (Figure 1.2C-F).

H3K27me3 modifications repress gene expression by promoting the association of nucleosomes to form compacted assemblies that are inaccessible to the transcription machinery. For instance, H3K27me3 recruits the Polycomb Repressive Complex (PRC1), which can self-interact to form compacted, phase-separated chromatin domains (Plys et al., 2018; Tatavosian et al., 2018). Therefore, we considered a second model, where H3K27me3 levels do not directly repress gene transcription per se but promote the self-association of nucleosomes at the Bcl11b promoter to maintain a compacted, silent state (Figure 1.4C, left). In addition to methylated and demethylated nucleosomes, this methylation-compaction (MC) model also incorporates the association and dissociation of nucleosomes in a compacted assembly. Here, the methylation status of nucleosomes influences their affinity for each other, with methylated nucleosomes self-associating with stronger affinity compared to demethylated nucleosomes. We do not explicitly model the spatial conformation of the compacted assembly; instead, following a chemical kinetics framework, we approximate the assembly to be sphere with a minimum nucleus size, using established approaches to describe self-assembly of cytoskeletal polymers (Howard, 2001; Mitchison, 1992) (see Mathematical Appendix for details).

From simulations of the methylation-compaction mechanism, we found that the gene locus can maintain a H3K27 methylated and compacted state for multiple cell divisions before switching in an all-or-none manner to a decompacted, low-methylation state (Figure 1.4C, center). As with the methylation-only model, the time delay in switching is well described by a first-order stochastic process, with a constant probability of activation per unit time (Figure 1.4D). However, in contrast with the methylation-only model, but in concordance with our experimental results (Figure 1.2C-F), changing H3K27me3 levels by varying methylation or demethylation rates changed gene activation timing in a much more graded manner (Figure 1.4B, right; sensitivity coefficient = ~0.6), though it caused marked changes in H3K27me3 levels prior to activation (Figure 1.4E). This tunability was robust over different parameter ranges (Figure 1.4B, right), different degrees of assembly disruption after DNA replication (Figure S1.5), and different degrees of cooperativity for H3K27 methylation (Figure S1.4). From a transition state theory analysis (see Mathematical Appendix and Figure S1.3D-E), we identified a critical requirement for this tunability in activation timing: specifically, nucleosomes must be able to associate with each other even without H3K27me3 modifications, such that methylated and de-methylated nucleosomes can associate with each other with comparable affinities. Consistent with this idea, there are multiple mechanisms for nucleosomal interactions that work independently of H3K27me3 modifications: PRC1 and other heterochromatin-associated proteins can bind nucleosomes independently of H3K27me3 (Francis et al., 2004; Gibson et al., 2019; Larson et al., 2017; Sanulli et al., 2019; Strom et al., 2017); furthermore, nucleosomes themselves can also bind directly with each other through histone-tail dependent interactions (Gibson et al., 2019).

As the Bcl11b epigenetic switch activated with timing delays that were unaffected by cell cycle speed changes (Figure 1.3), we tested whether the methylation-compaction mechanism also generates activation delays that were cell division-independent. As a comparison, we analyzed a passive dilution model for H3K27me3 loss and gene activation, and in both cases varied cell cycle rates and measured effects on activation times (Figure 1.4F). In agreement with our experimental data, we found that the methylation-compaction model generates gene activation times that remain relatively constant amid variable cell cycle lengths ranging from 10-30 hours (Figure 1.4G). This independence between activation timing and cell cycle duration held even when we adjusted the model such that DNA replication led to partial disruption of the compacted nucleosome assembly (Figure S1.5A-E). In contrast, when H3K27me3 is removed due to passive dilution by DNA replication, activation time delays increase with increasing cell cycle lengths (Figure 1.4F-G), effectively implementing a cell cycle counting mechanism for control of activation. Histone methylation and nucleosome compaction dynamics occur with reported timescales of minutes and seconds (Kristensen et al., 2011; Larson et al., 2017; Sneeringer et al., 2010), respectively, which are far faster than the typical cell division rates which range from hours to days. This explains why the epigenetic state recovers rapidly after DNA replication as observed in our simulations (Figure 1.4A-B, right). In agreement with this prediction, we found that the activation timing can be coupled to cell cycle speed by dramatically reducing histone methylation and demethylation rates such that they are slower than the cell division rates (Figure S1.5F).

Together, these modeling results clarify how a H3K27me3-dependent epigenetic switch can give rise to timing delays in gene activation spanning multiple cell divisions, that are both tunable by chromatin regulators and are cell division independent. First, H3K27me3 loss does not directly result in gene activation, but instead modulates a separate cooperative process that acts as the gatekeeper for gene transcription. Given recent insights into how nucleosomes can self-associate to form phase-separated structures (Gibson et al., 2019), and how this self-association can be modulated by histone tail modifications and their associated binding proteins (Larson et al., 2017; Strom et al., 2017), we propose that H3K27me3 loss weakens inter-nucleosomal contacts at the Bcl11b locus, leading to gene locus decompaction and gene activation. Second, the compaction process itself must be partially independent from H3K27me3 such that the compacted nucleosome assembly is maintained and modulated by other chromatin-associated proteins (Francis et al., 2004; Larson et al., 2017; Strome et al., 2014).

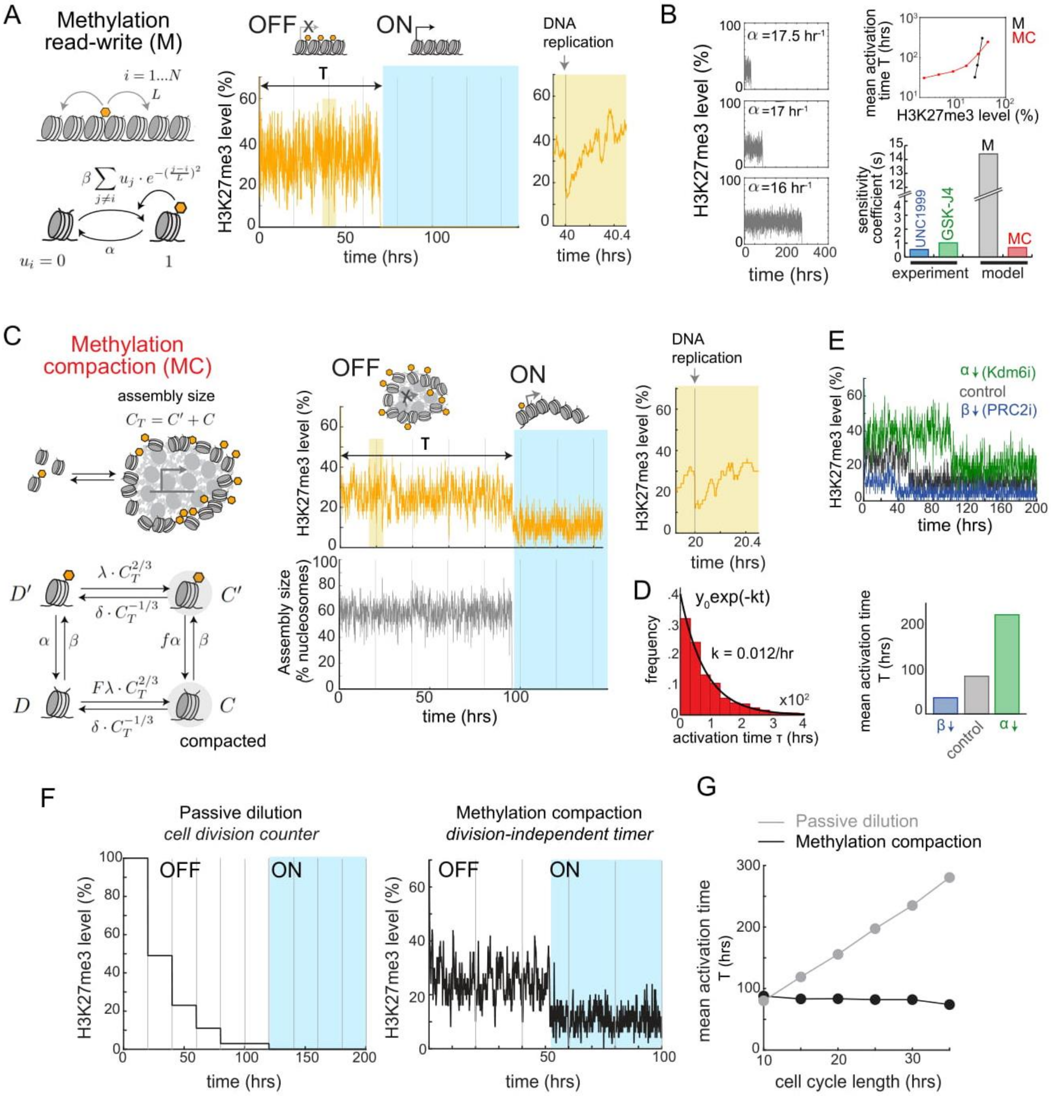
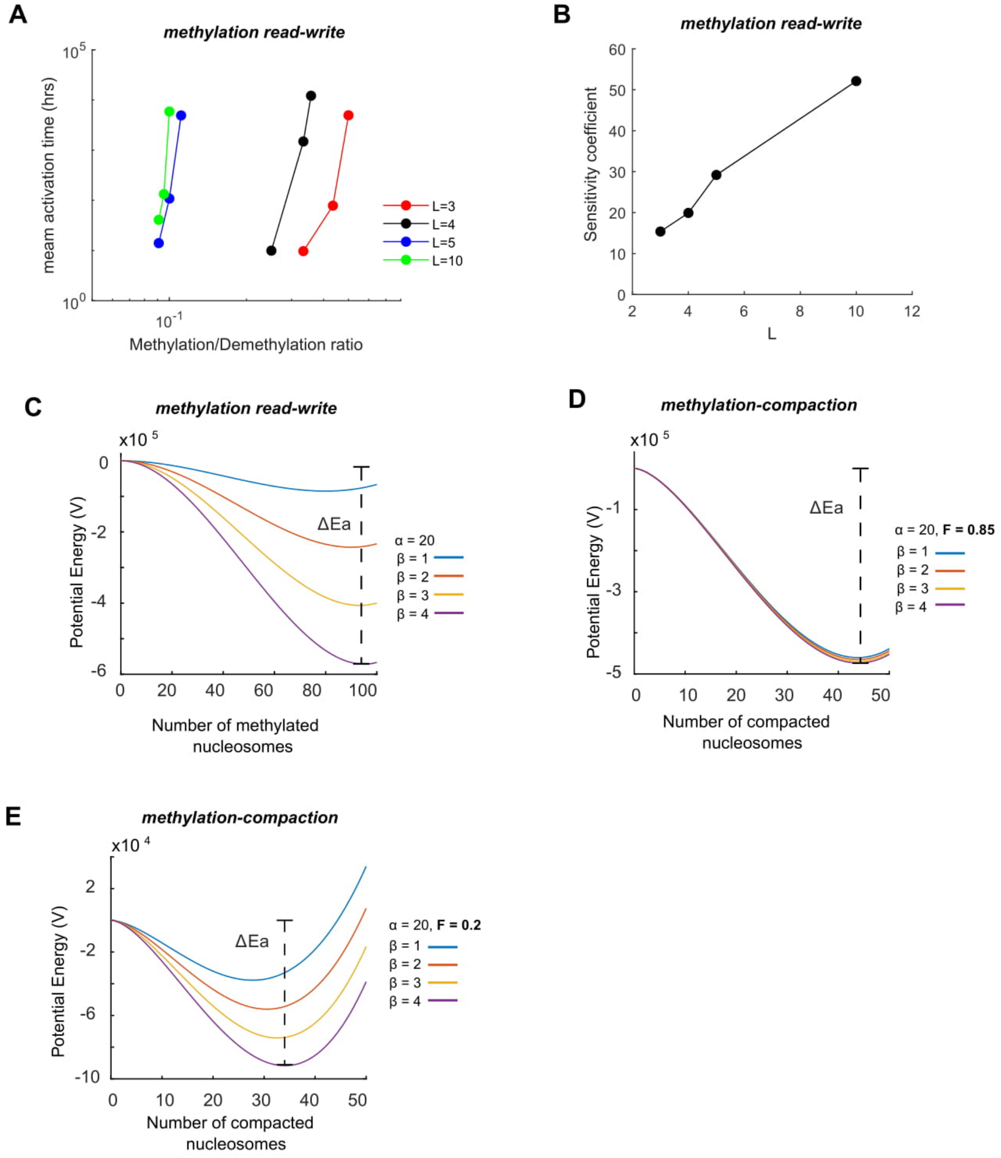


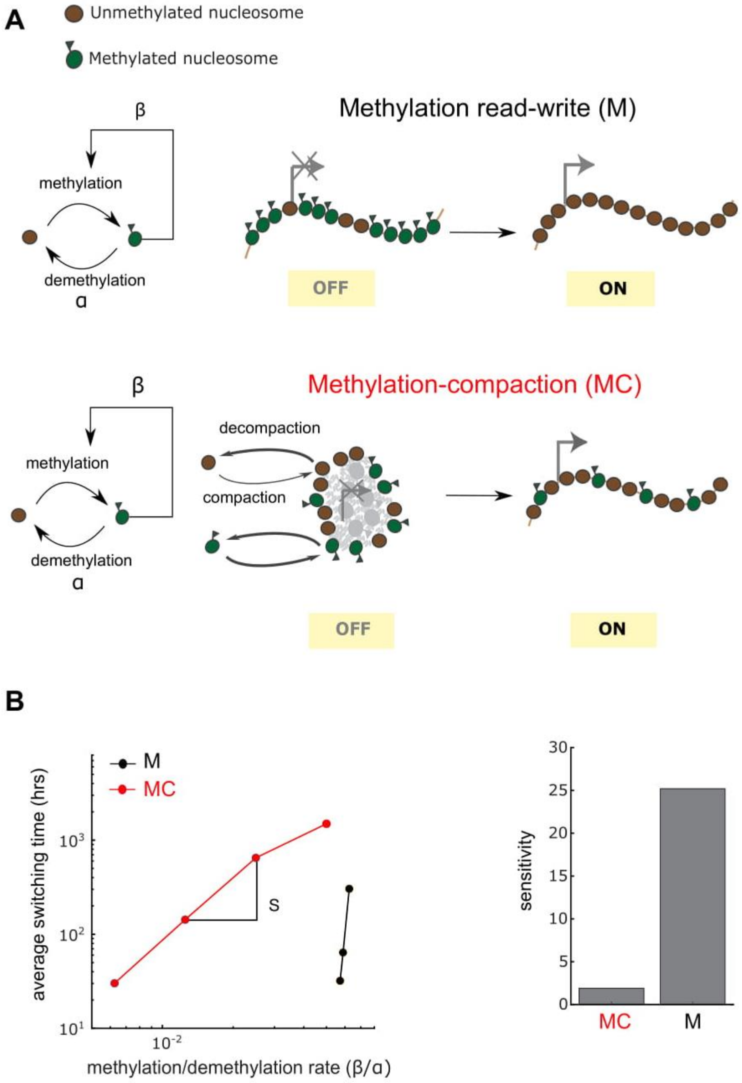
Figure 1.4. A methylation-compaction mechanism generates tunable, division-independent time delays in epigenetic switching. Two candidate mechanisms for the *Bcl11b* epigenetic switch were analyzed using mathematical modeling, a methyl-ation-only mechanism (A), and a coupled methylation-compaction mechanism (C-G). (A) Methylation only model, along with fraction H3K27me3 nucleosomes over time from a stochastic simulation, inset shows H3K27me3 loss and recovery after DNA replication. (B) Sensitivity analysis of the methylation-only model (gray) and methylation-compaction model (red). Simulated H3K27me3 time traces for indicated parameters in the methylation model (left). Mean simulated activation time as a function of H3K27me3 levels (top right) and calculated sensitivity coefficients for this relationship compared to that of experimental results (bottom right). (C) The methylation-compaction model, along with time traces of the fraction of H3K27me3-marked nucleo-somes (top middle), and

the fraction of nucleosomes in a compacted assembly (bottom middle). Inset (top right) shows H3K27me3 recovery after cell division. (D) Histogram shows distribution of activation times, along with fit to the exponential function $y = e^{-kt}$ with $k = 0.012\ hr^{-1}$. (G) H3K27me3 time traces (top) and bar charts (bottom) show impact of simulated PRC2 or Kdm6a/b inhibition on initial H3K27me3 levels and mean activation times. Results represent average of 200 simulations. (F) Representative simulation of methylation dynamics for cell division dependent (passive dilution) and independent (methylation compaction) models of gene activation. (G) Mean simulated activation time as a function of cell cycle length.



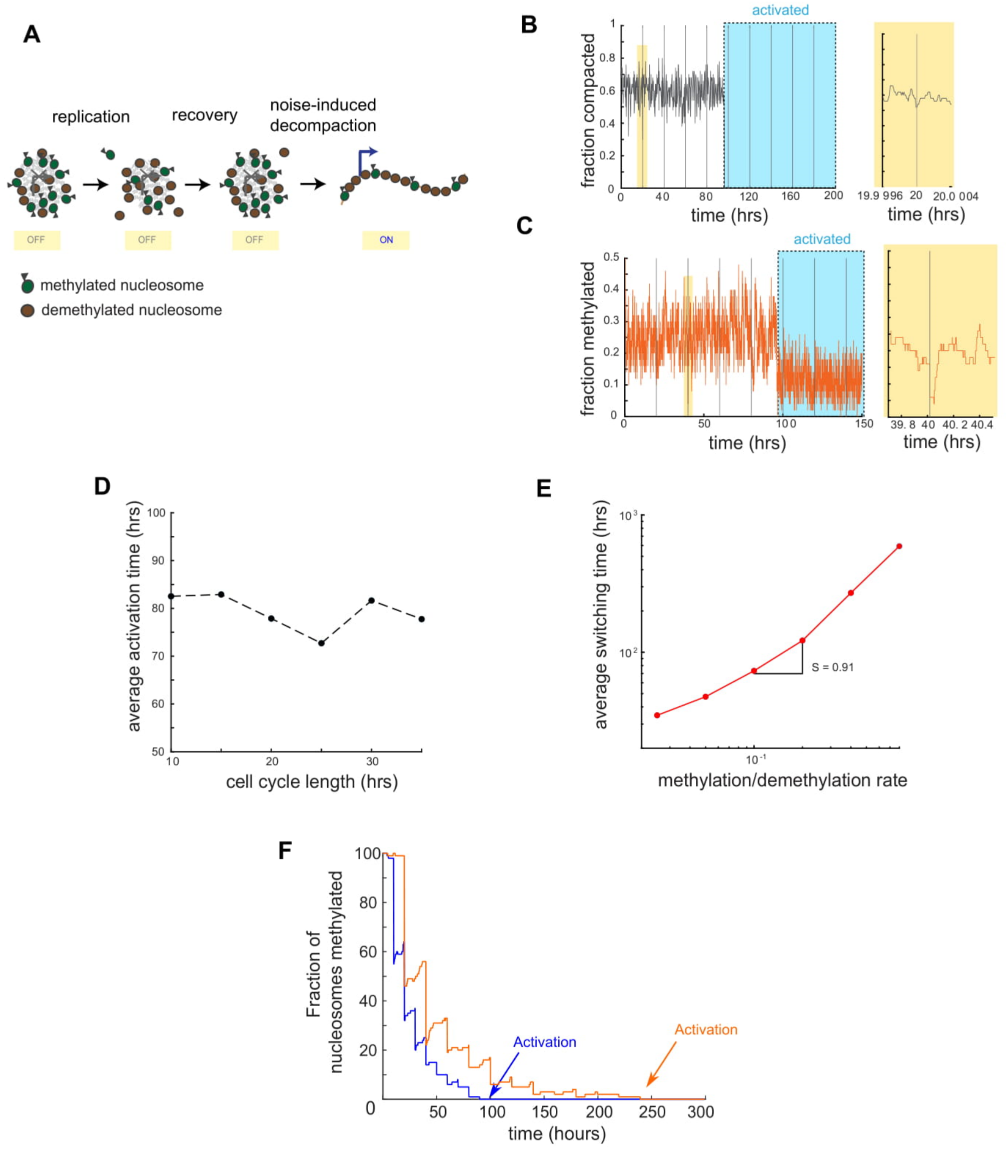
Supplementary Figure S1.3. Analytical formulations of methylation-only and

methylation compaction models (A) Mean activation time as a function of the methylation/demethylation ratio derived from methylation-only model simulations while decreasing the number of nulceosomes within ‘reach’ of a PRC2 Complex L. (B) Sensitivity coefficient (ΔlogX/ ΔlogY) as a function of L. (C) Potential energy landscapes of methylation model. (D-E) Potential energy landscapes of the methylation-compaction model. Parameter F dictates how sensitve nucleosome compaction affinity is to demethylation (i.e. when F is high, the compaction affinity is only moderately affected by demethylation; see Mathematical Appendix for more details). The activation energy barrier (E_a) is defined as the potential energy (V) height between the local maximum and local minimum of the potential energy landscape. Each potential curve was plotted with demethylation parameter set to 20 hrs^{-1} and methylation rate parame-ter as indicated by the curve’s color see Mathematical Appendix).



Supplementary Figure S1.4. Methylation compaction model with cooperative

methylation rate (A, *top*) Methylation model enables gene activation via complete eviction of methylation marks. (A, *bottom*) Methylation compaction model with cooperative methylation rate. A nucleosome's methylation rate increases with the number of methylated nucleosomes in the system. (B) Average switching time as function of methylation (β) and demethylation rate (α) ratio for methylation (black) and methylation compaction models (red). Sensitivity coefficient ΔlogX/ΔlogY) for each plot was calculated by taking the slope of the linear fit y = ax + b for the methylation dataset and the last 5 data points for the methylation compaction model.



Supplementary Figure S1.5. Methylation compaction model with assembly disruption at cell division points. (A) Modified methylation compaction model where every cell division leads to 50% reduction in methylation state and 10% reduction in compaction state. (B-C) Model’s compaction and methylation states as a function of time. Zoomed in first replication event. (D) Average switching time of the system as a function of cell cycle length. (E) Average switching time as a function of methylation and demethylation rate ratio. Tunability coefficient S (ΔlogX/ΔlogY) for each plot was calculated by taking the slope of the linear fit y = ax + b for the methylation dataset. (F) Fractions of methylation histones are shown for the methylation only model with the cell division lengths set to be 10 hrs (blue) and 20 hrs (orange). Methylation and demethylation rates were set to 0.001 per hour (see Mathematical Analysis).

1.6 THE BCL11B LOCUS SWITCHES TO AN EXTENDED CONFORMATION WITH ACTIVATION

Here, we test a key prediction of the methylation-compaction mechanism, described above, that the Bcl11b locus is compacted prior to activation, but switches to an extended, decompacted state during activation. To test this prediction, we measured the end-to-end distances between genomic regions at the Bcl11b locus using DNA fluorescence in-situ hybridization (FISH), an established approach to estimate the degree of chromatin compaction at a gene locus (Eskeland et al., 2010; Giorgetti et al., 2015). For example, a study of the repressed Hoxd gene cluster, a locus that is highly compacted in stem cells, revealed that the distance between two ends of the gene cluster increased upon transcriptional activation (Eskeland et al., 2010). We initially utilized a pair of FISH probe sets flanking the gene body (i.e. at the 5’ and 3’ ends of the gene body separated by 100kb) (Figure 1.5A). However, active promoters can interact with downstream regions of the gene body through RNA polymerase II (RNAPII) mediated contacts (Lee et al., 2015; Zheng et al., 2019), and such interactions are indeed observed in chromosome conformation capture data for the Bcl11b locus in its active state (Hu et al., 2018) (Figure S1.6). Therefore, to estimate the degree of compaction independently from RNAPII-mediated DNA looping, we also utilized an additional pair of FISH probe sets flanking a 100kb region upstream from the promoter. This upstream region resides at the edge of the putative compaction domain that encompasses Bcl11b, and is enriched for H3K27me3 and H3K9me3 modifications (Figure S1.6). We performed DNA-FISH in early T cell progenitors (DN1), as well as in DN2 progenitors before or after Bcl11b activation (RFP-/YFP- vs. RFP+/YFP+), and measured end-to-end distances for both regions using three-dimensional imaging.

From these experiments, we found that Bcl11b upstream region showed similar end-to-end distances in DN1 and Bcl11b RFP-/YFP- DN2 progenitors, consistent with the locus adopting a similar degree of compaction in both Bcl11b non-expressing populations. However, in Bcl11b RFP+/YFP+ progenitors, where Bcl11b first turns on, this end-to-end distance increased significantly (Mann Whitney U test significance: DN1 (RFP-/YFP-) vs. DN2 (RFP+/YFP+), $p = 0.043$; DN2 (RFP-/YFP-) vs. DN2 (RFP+/YFP+), $p = 0.035$), consistent with the Bcl11b locus switching to an extended conformation. In contrast, the end-to-end distance of Bcl11b gene body was greatest in DN1 progenitors, and decreased progressively both before and after transcriptional activation at the DN2 stage (Figure 1.5B, left). This observation is consistent with the formation of RNAPII-mediated DNA loops within the body that may obscure potential effects of chromatin decompaction (Hu et al., 2018). However, the results from the Bcl11b upstream region measurements are consistent with a key prediction of the methylation-compaction model, namely, that the locus switches from compacted state to an extended, decompacted state concurrently with transcriptional activation.

Repressed heterochromatic regions of the genome frequently reside at the nuclear periphery, where interactions between nuclear lamina proteins and nucleosomes are thought to facilitate chromatin compaction (van Steensel and Belmont, 2017; Ulianov et al., 2019). In contrast transcriptionally active regions tend to reside in the nuclear interior. Therefore, we used our DNA-FISH data to examine whether Bcl11b moves from the nuclear periphery to the interior as it decompacts and turns on, as suggested from previous studies (Isoda et al. 2018). Indeed, the distance between the Bcl11b promoter and the nuclear periphery was higher in DN2 progenitors expressing Bcl11b compared to those inactive for Bcl11b expression, consistent with a transition from the nuclear periphery to the interior upon gene activation (Figure 1.5D). This finding, in conjunction with the observed increase in the end-end distance of the Bcl11b upstream region, suggests that the Bcl11b locus transitions from a compacted to decompacted state during transcriptional activation (Figure 1.5E), consistent with a separate cooperative process working together with H3K27me3 modifications to regulate gene activation timing.

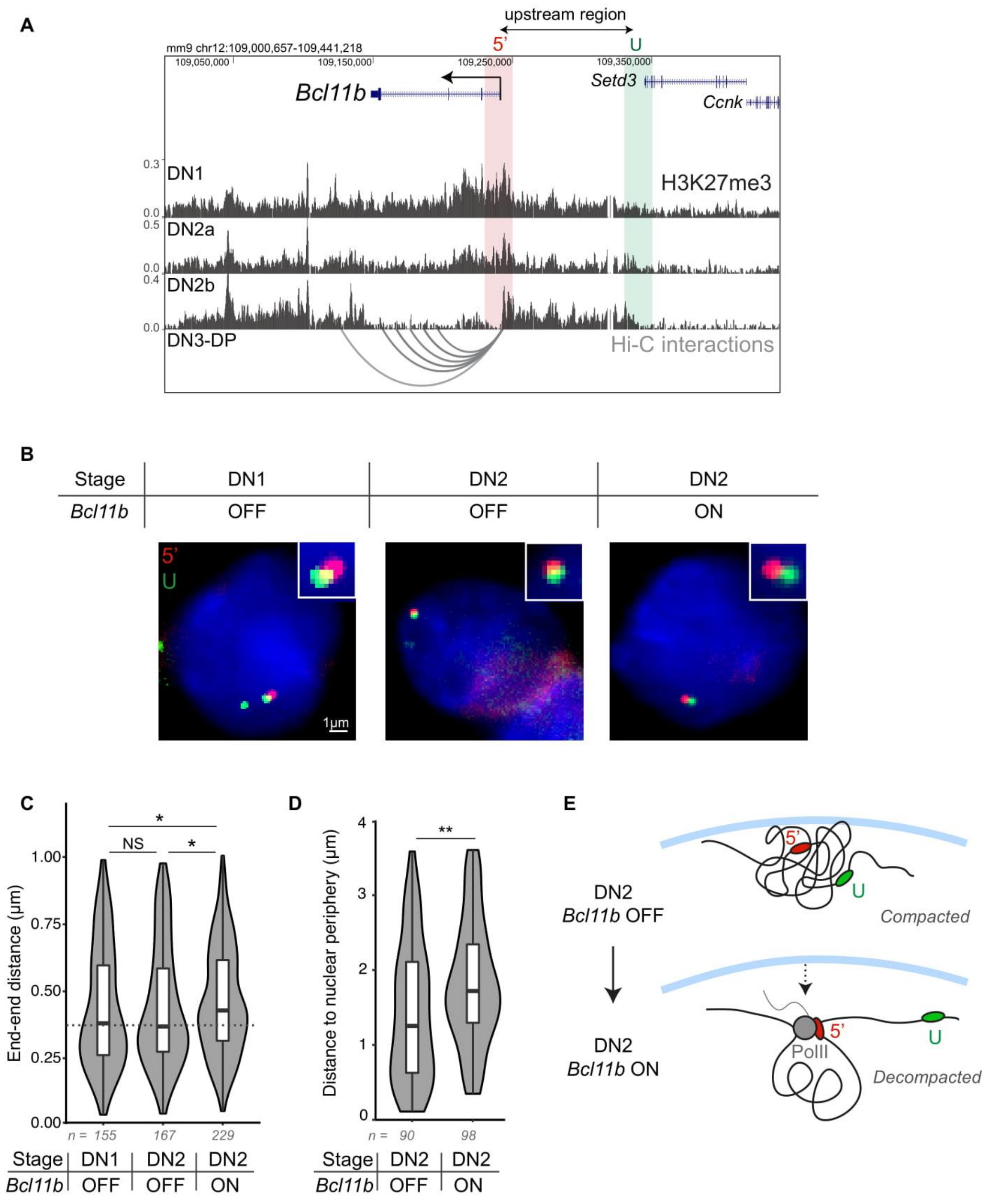
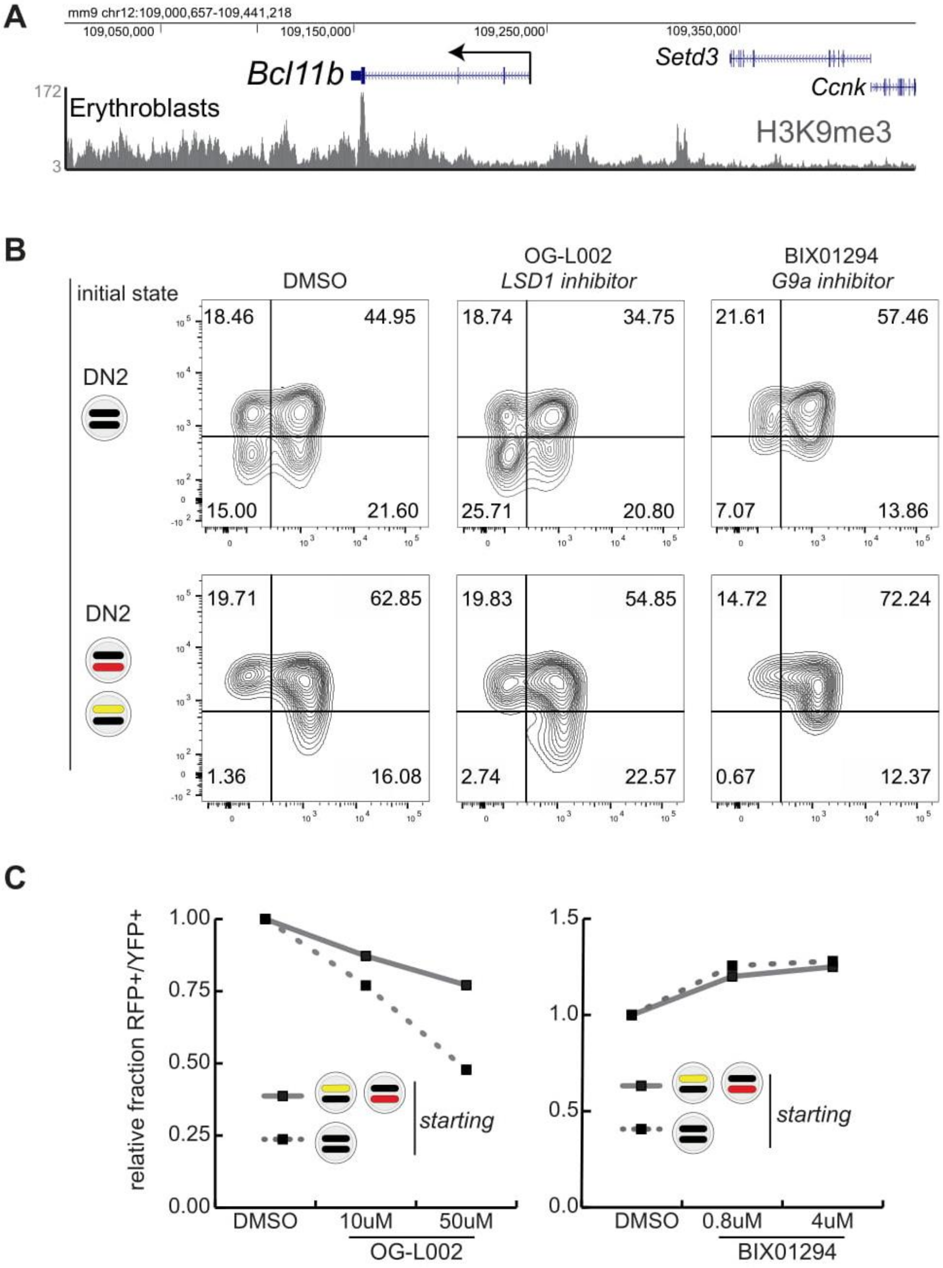


Figure 1.5. The *Bcl11b* locus switches to an extended conformation upon transcriptional activation. (A) Oligopaint probes span-ning 20kb were designed to separately label each end of the *Bcl11b* gene body (*3'* and *5'*) or the upstream region (*5'* and *U)* . (B) Violin plots show the results from 3D Euclidean distance measurements between each probe pair (Mann Whitney *U* test significance: *p < 0.05 and ** p < 0.01, n = number of foci pairs). T cell progenitors were sorted based on cell surface markers and *Bcl11b* reporter expression (either Bcl11b OFF (RFP-/YFP-) or Bcl11b ON (RFP+/YFP+)) before performing DNA-FISH with probes above. (C) Repre-sentative

images for each condition. Centroids for each foci pair shown exist in the same *z*-plane and thus provide visual representation of the Euclidean distance (pixel size = 110nm). (D) Violin plots show the results from 2D measurments between the promoter and the nuclear periphery (Mann Whitney *U* test significance: ** $p < 0.01$, n = number of foci). (E) Schematic depicting decompaction model. In the OFF state, *Bcl11b* exists in a compacted conformation residing at the nuclear lamina. In the ON state, the locus moves away from the nuclear periphery and becomes decompacted resulting in increased distance between the promoter, *5'*, and the upstream, *U*, regions. As a result of PolII-mediated loop extrusion during transcriptional elongation (Zheng et al. 2019), the end-end distance at the gene body decreases



Supplementary Figure S1.6. Bcl11b activation timing in the presence of H3K9me3 small molecule inhibitors. (A) UCSC Genome Browser view of ChIP-seq results for H3K9me3 (Mouse ENCODE, Ross Hardison), H3K27me3 (Zhang et al., 2012) and Hi-C representation maps interactions between the *Bcl11b* TSS and other DNaseI hypersensitivity sites (DHS) in Bcl11b+ DN3-DP T-cell progenitors (Hu et al., 2018). (B) DN2 progenitors, either Bcl11bRFP-/YFP- (*top*) or Bcl11bRFP-/YFP+ and Bcl11bRFP+/YFP- (*bottom*), were purified and re-cultured on OP9-DL1 cells for 3 days in the presence of OG-L002 or BIX01294 before analyzing reporter expression. (C)

Relative fraction of progenitors RFP+/YFP+ after 3 days normalized to the respective DMSO controls.

1.7 H3K27ME3-INDEPENDENT REGULATION OF BCL11B EPIGENETIC SWITCH TIMING

The methylation compaction model predicts that, in order for activation time delays to be tunable, nucleosomes must retain an ability to self-associate, even without H3K27me3 modification. Methylation of H3K9 (H3K9me2/3) is also associated with repressive heterochromatin domains and can serve as a binding site for HP1α, which could facilitate nucleosome adhesion in the absence of H3K27me3 (Poleshko et al., 2013; Sanulli et al., 2019; Wang et al., 2019). H3K9me3 is enriched at the Bcl11b locus in non-T cell lineages (Figure S1.6); therefore, we tested whether H3K9 methylation also regulates Bcl11b activation timing, and could potentially provide additional nucleosomal interactions necessary for tunable timing control, as predicted by the model. To do so, we sorted DN2 progenitors with one active Bcl11b allele (Bcl11b RFP+YFP- and Bcl11b RFP-YFP+), as above in Figure 1.2, recultured them with inhibitors targeting H3K9me3-modifying enzymes, and assayed for activation of the Bcl11b locus after three days. We targeted Lsd1, an H3K9 demethylase, and G9a, an H3K9 methyltransferase. We found that Lsd1 inhibition dampened Bcl11b activation, decreasing the fraction of Bcl11b RFP+/YFP+ cells, whereas G9a inhibition enhanced Bcl11b activation, increasing the fraction of Bcl11b RFP+/YFP+ cells (Figure S1.6). The degree to which H3K9me inhibition affected Bcl11b activation timing was similar to that for H3K27me inhibition (Figure 1.2), suggesting that these two repressive modifications may contribute similarly to the nucleosomal self-interactions posited in the methylation compaction model. Thus, as predicted by the methylation-compaction model, other histone modifications work together with H3K27me3 to modulate the timed epigenetic switch at the Bcl11b locus.

1.8 TRANSCRIPTION FACTORS SET TIME DELAYS IN EPIGENETIC SWITCHING

The activation time constant of the Bcl11b epigenetic switch is tunably controlled not only by chromatin modifying enzymes, but by two transcription factors, Gata3 and TCF-1, that act via a distal enhancer (Kueh et al., 2016; Ng et al., 2018). These findings are consistent with a broader literature showing that transcription factors and the number of their binding sequences can modulate target gene activation probabilities (Dufourt et al., 2018; Walters et al., 1995; Weintraub, 1988), though it remains unknown how they achieve such tunable timing control. Here, we tested whether the methylation-compaction mechanism could integrate information about transcription factor levels to control activation timing. We first considered a scheme, where transcription factors bind to free nucleosomes, and block their association with other nucleosomes in the compacted assembly (Figure 1.6A). Indeed, a class of transcription factors, termed pioneer factors, can possess affinity to nucleosomes in addition to specific DNA sequences (Fernandez Garcia et al., 2019; Meers et al., 2019). Such disruption of compaction could also occur via the activation of gene or non-coding RNA transcription (Rinn et al., 2007; Tu et al., 2017), or by recruitment of factors that disrupt interactions between nucleosomes (Kraushaar et al., 2013; Talbert and Henikoff, 2017; Zhou et al., 2016). In our simulations, we adjusted the number of transcription factor binding sites, each of which occupies a different nucleosome; for each binding site number, we also varied transcription factor concentrations in the nucleus.

From simulation results, we found that the rate constant for epigenetic switching, which varies inversely with the activation time delay, varies in a graded manner with both transcription factor concentration and the number of transcription factor binding sites (Figure 1.6B). For a given number of binding sites NB, the activation rate increased with transcription factor concentration until it reached a maximal activation rate k_{max}. Increasing the number of binding sites increased the activation rate in a synergistic manner, with about a 2.5 fold increase in the maximal activation rate with addition of each binding site (Figure 1.6B, top right). Interestingly, increasing the number of binding sites also increased the transcription factor concentration at which half-maximal activation occurred (Figure 1.6B, bottom right). These findings show that transcription factors, when binding to multiple nucleosomes to prevent their self-association, can tunably control the timing of gene activation as controlled by a methylation-compaction mechanism.

Besides blocking nucleosomal interactions, transcription factors could also promote activation by inducing histone demethylation. Histone demethylation could occur by direct recruitment of Kdm6a/b demethylases (Estarás et al., 2013; Seenundun et al., 2010; Williams et al., 2014), or through PRC2 eviction by chromatin-remodeling enzymes (Kadoch et al., 2017). Therefore, we considered a mechanism, where transcription factors induce the demethylation of a fixed number of nucleosomes NR around their binding vicinity. In simulations, we varied both this number of nucleosomes reached and transcription factor levels, and measured resultant effects on epigenetic switching rate.

From simulations, we found that transcription factors can control activation rates, but only if they induced demethylation of a relatively large number of nucleosomes (Figure 1.6C-D). When only five nucleosomes were accessible to a bound transcription factor, activation rates shifted only slightly when transcription factor levels increased (Figure 1.6D); in contrast, when twenty-five nucleosomes were accessible, activation rates now increased by over five-fold (Figure 1.6D). Taken together, these results show that transcription factors can alter chromatin modifications or compaction states to control switching times in the methylation-compaction model.

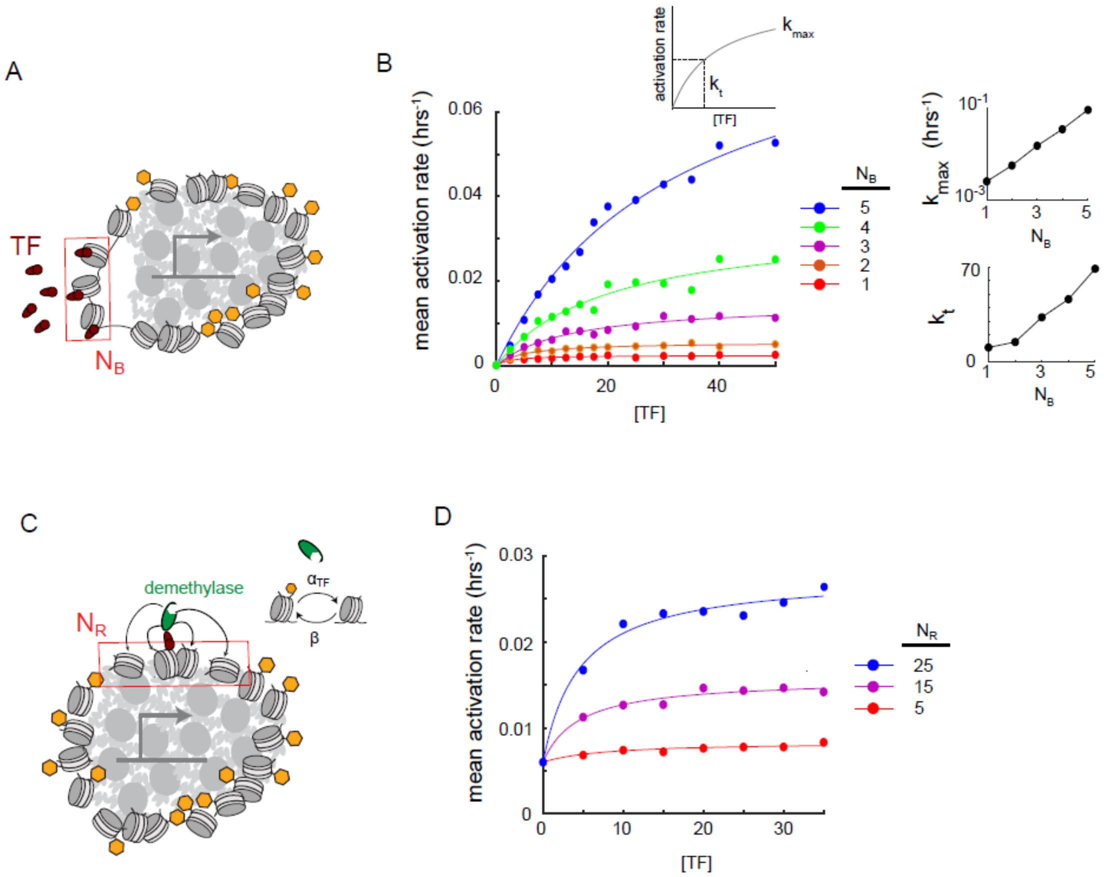


Figure 1.6. Transcription factors can tunably modulate epigenetic switching time delays set by a methylation compaction mechanism . (A) Model in which transcription factors can prevent a number of nucleosomes (NB) from entering the compacted assembly. (B) Mean simulated gene activation rate as a function transcription factor concentration (left). Maximal activation rate (k_{max}) and half-maximal concentration (k_t) as a function of the number of nucleosomes bound (NB) (right). (C) Model in which transcription factors can recruit demethylases to a number of nucleosomes (NR). (D) Mean simulated gene activation rate as a function of transcription factor concentration. mean activation rate (hrs^{-1}) TF demethylase NB k_t [TF] activation rate k_{max}

1.9 METHODS

1.9.1. EXPERIMENTAL TECHNIQUES

Cell purification

To isolate hematopoietic stem and progenitor cells (HSPCs) for in vitro differentiation or CUT&RUN experiments, bone marrow cells were harvested from femurs and tibias of 2 to 4 month-old Bcl11bRFP/YFP mice. CD117 MicroBeads (Miltenyi Biotec) were used to enrich for HPSCs which were frozen in 90% FBS and 10% DMSO at 106 cells/mL. For CUT&RUN experiments, HPSCs were further purified by staining with anti-CD117 APC-eFluor780 (ThermoFisher Scientific) and with biotinylated antibodies against a panel of bone marrow lineage markers (CD19, CD11b, CD11c, NK.1.1, Ter119, CD3ε, Gr-1 and B220 (BioLegend)). Cells were then washed with HBH (Hank Balanced Salt Solution (HBSS) with 0.1% bovine serum albumin (BSA) and 10mM HEPES) and stained with streptavidin-PerCP/Cy5.5 (BioLegend).

In vitro Differentiation of T cell Progenitors

To generate double-negative (DN) T cells in vitro, thawed CD117+ cells were cultured on OP9-DL1 stromal cell monolayers as described before using standard culture medium [80% αMEM (Gibco), 20% Fetal Bovine Serum (Sigma-Aldrich), Pen-Strep-Glutamine (Gibco)], grown at 37^{o}C in 5% CO2 conditions]. All in vitro T cell generation cultures were supplemented with 5ng/mL Flt3-L and 5 ng/mL IL-7 (Peprotech), and were sorted after 6 or 7 total days of culture before transducing with retroviral vectors or treating with small molecule inhibitors. DN2 cells were re-cultured in the same conditions following all cell sorting experiments.

Flow Cytometry and Cell Sorting

Fluorescence activated cell sorting was used to isolate DN2 cells of interest with the following protocol. Bone marrow derived cell cultures were scraped and incubated in 2.4G2 Fc blocking solution and stained with anti-CD25 APC-eFluor 780 (Clone PC61.5, eBioscience) and with biotinylated antibodies against a panel of lineage markers (CD19, CD11b, CD11c, NK.1.1, Ter119, CD3ε, Gr-1 and B220 (BioLegend)). Stained cells were washed with HBH (Hank Balanced Salt Solution (HBSS) with 0.1% bovine serum albumin (BSA) and 10mM HEPES and stained with streptavidin-PerCP/Cy5.5 (BioLegend). Stained cells were washed, resuspended in HBH, and filtered through a 40-um nylon mesh for sorting with a BD FACS Aria III (BD Biosciences) with assistance from the University of Washington Pathology Flow Cytometry Core Facility. A benchtop MacsQuant VYB flow cytometer (Miltenyi Biotec) and a benchtop Attune Nxt Flow Cytometer (ThermoFisher Scientific) were used to analyze time course and perturbation experiments and acquired data were analyzed with FlowJo software (Tree Star).

Retroviral construct and transduction

Overexpression of c-Myc was achieved using cMyc H2B-mCerulean MSCV retroviral vector which was described previously(Kueh et al., 2016). Retroviral mir30-based constructs (a gift from J. Zuber) were used as a backbone for delivering short hairpin RNA (shRNA)(Fellmann et al., 2013). pBAD-mTagBFP2 (a gift from V. Verkhusha, Addgene plasmid #34632) was used to substitute mTagBFP2 for the existing GFP using PCR cloning with the restriction enzymes NcoI and SalI. The RetroE-shEed retroviral construct was generated by PCR cloning as previously described(Fellmann et al., 2013) with the following PCR template:

TGCTGTTGACAGTGAGCG**AAGGCATTATAAGAATAATTAA**TAGTGAAGCCACAGATGTA**TTAATTATTCTTATAATGCCTC**TGCCTACTGCCTCGGA.

Retroviral particles were generated using the Phoenix-Eco packaging cell line as previously described(Kueh et al., 2016). Viral supernatants were collected at 2 and 3 days after transfection and immediately frozen at -80oC. To infect bone marrow derived T cell progenitors, 33 μg/mL retronectin (Clontech) and 2.67 μg/mL of DL1-extracellular domain fused to human IgG1 Fc protein (a gift from I. Bernstein) were added in a volume of 250 μL per well in 24-well tissue culture plates (Costar, Corning) and incubated overnight. Viral supernatants were added the next day into coated wells and centrifuged at 2000 rcf for 2 hours at 32oC. Bone marrow derived derived T cell progenitors used for viral transduction were cultured for 6-7 days according to conditions described above, disaggregated, filtered through a 40-μm nylon mesh, and 106 cells were transferred onto each retronectin/DL1-coated virus-bound well supplemented with 5 ng/mL SCF (Peprotech), 5 ng/mL Flt3-L, and 5 ng/mL IL-7.

CUT&RUN H3K27me3 profiling

CUT&RUN experiments were carried out as previously described28 with the following modifications: 1-2.5x105 cells were isolated by FACS as described in sections above, bound to Concanavalin A coated magnetic beads (Bangs Laboratories), and permeabilized with 0.025% (wt/vol) digitonin. Permeabilized cells were incubated overnight at 4oC with 5ug of anti-H3K27me3 (Active Motif 39156) and then washed before incubating with protein A-MNase fusion protein (a gift from S. Henikoff) for 15 minutes at room temperature. After washing, cells were incubated in CaCl2 to induce MNase cleavage activity for 30 minutes at 0oC. The reaction was stopped with 2XSTOP buffer (200 mM NaCl, 20 mM EDTA, 4 mM EGTA, 50 mg/mL RNase A and 40 mg/mL glycogen) with 2pg of yeast spike-in DNA per sample. Histone-DNA complexes were isolated from insoluble nuclear chromatin by centrifugation and DNA was extracted with a NucleoSpin PCR Clean-up kit (Macherey-Nagel). For CUT&RUN quantitative PCR, human Kasumi-1 cell line (ATCC CRL-2724™) were added before binding the cells to Concanavalin A beads for internal standard instead of yeast spike-in DNA.

CUT&RUN Library Preparation and Sequencing

Library preparation from CUT&RUN products was completed with KAPA Hyper Prep Kit (KAPA Biosystems) following standard protocol with PCR amplification settings adjusted so that annealing and extension steps are combined into one step at 60oC for 10s. Library products were size selected to be within 200 - 300 bp range using AMPure beads (Agencourt). Libraries were sequenced using an Illumina MiSeq system with paired-end 25 bp sequencing read length and TruSeq primer standard for approximately 5 millions reads per sample.

CUT&RUN sequencing analysis

Paired-end sequencing reads were aligned separately to mouse (NCBI37/mm9) and yeast

(SacCer_Apr2011/sacCer3) genomes using Bowtie2 (Langmead and Salzberg, 2012) with the following setting: --local --very-sensitive-local --no-unal --no-mixed --no-discordant -I 10 -X 700 as suggested for mapping CUT&RUN sequencing data (Skene et al., 2018). The alignment setting was designed to specifically search with high stringency for only appropriately paired reads with the proper orientation. The resulting alignments were converted to BAM files with SAMtools (Li et al., 2009) and then converted to BED files with BEDTools. Reads were sorted and filtered to remove random chromosomes. BEDTools genomecov was used to generate histograms for the mapped reads using a scaling factor that is the product of the number of spiked-in yeast reads and the number of input cells. The resulting bedGraph files were visualized using the UCSC Genome Browser (Davis et al., 2018; Kent et al.).

CUT&RUN qPCR

Extracted DNA from CUT&RUN samples was size selected with Ampure XP magnetic beads (Beckman Coulter) to remove fragments >800bp. Primers were designed to detect the the mouse Bcl11b promoter (F - TCCACCTACCAGACCCCGAA, R - CTTCTTCAAAGTGCTTGGCCTC) and the human PAX5 promoter (F - CCAGGATGTGCTGCTGTCCCAG, R - CTCCCTGGTGCTGTGCACTGA). PowerUp SYBR Green Master Mix (ThermoFisher Scientific) and CFX96 Real-Time PCR Detection System (Bio-Rad) were used for quantitative PCR. Since Kasumi-1 cells were used as internal standard, relative enrichment of H3K27me3 at Bcl11b was quantified by the ΔΔCq method using the human PAX5 promoter for normalization to account for differences in efficiency and sample loss during processing.

Cell Preparation for Time-lapse Imaging

T cell progenitors underwent in vitro differentiation protocol as described above. Cells were then harvested and infected with either a MSCV empty vector or cMyc overexpression vector harboring an IRES-H2B-mCerulean reporter cassette. 16-24 hours later CFP-positive cells were purified by FACS and seeded onto PDMS micromesh (250 µm hole diameter, Microsurface) mounted on top of 24-well glass bottom plate (MatTeck). To prepare the stromal-free differentiation system, the top face of PDMS micromesh was first blocked by incubating in solution of 130 µg/ml BSA while mounted on top of a 24-well plate overnight at 4oC. This step prevents subsequent binding of retronectin to the side of the meshm, allowing the cells to climb out of the microwells. Blocked micromesh was then transferred to a clean 24-well glass bottom plate. The well and mesh constructs were incubated in a solution of 10 µg/ml retronection and 3 µg/ml DL-1 overnight at 4oC. The well was then washed with PBS, and culture media [80% αMEM (Gibco), 20% Fetal Bovine Serum (Sigma-Aldrich), Pen-Strep-Glutamine (Gibco), 5 ng/ml IL-7 (Clonetech), 5 ng/ml Flt-3 (Clonetech), 50 ng/ml mSCF (Clonetech), 50 µM beta-mercaptoethanol (Sigma) grown at 37oC in 5% CO2 conditions] was added, and sorted cells were introduced at a concentration of 5-10 cells per microwell.

OligoDNA-PAINT FISH

The OligoMiner pipeline was used to design oligopaint libraries (Beliveau et al., 2018). 35-52bp probes were designed to target 20kb regions at a density of approximately 12-14 probes per kilobase. Bone marrow progenitors were grown on OP9-DL1 stromal cells for 8 days under normal growth conditions. Cells were filtered through a 70uM filter and incubated with 2.4G2 blocking buffer before staining with anti-CD25 APC-eFluor 780 (Clone PC61.5, eBioscience), anti-CD117 APC (Clone 2B8, ThermoFisher), and biotinylated antibodies against a panel of lineage markers (CD19, CD11b, CD11c, NK.1.1, Ter119, CD3ε, Gr-1 and B220 (BioLegend)). DN1 progenitors (CD25+/CD117hi/CD44-/Bcl11b RFP-YFP-), DN2a progenitors (CD25+/CD44+/Bcl11b RFP-YFP-) and DN2b progenitors (CD25+/CD44+/Bcl11b RFP+YFP+) were purified by FACS and centrifuged on top of poly-L-lysine coated 18-well chambered glass coverslips (Ibidi). Cells were then fixed with 4% paraformaldehyde for 10 minutes and permeabilized for 10 minutes in 0.1% TritonX-100 before performing the OligoDNA-PAINT protocol (Beliveau et al., 2017). Permeabilized cells were incubated in 0.1N HCl for 5 minutes, followed by RNaseA (100ug/m) for 1 hour at 37C. Cells were then washed with 2X SSCT and incubated with 2XSCCT/50% formamide for 20 minutes at 60C. A hybridization mixture was prepared containing 50% formamide, 2XSCC, 0.1% Tween-20, 300mM sodium azide, 10% dextran sulfate, 100nM of adapter oligos, 100nM of fluorescently labeled reporter oligos, and 500nM of probes. The hybridization mixture was added and incubated for 3 minutes at 78C before incubating overnight in a humidifier chamber at 37C. Approximately 18 hours later, cells were washed with pre-heated 2XSCCT for 5 minutes at 60C. This step was repeated four times before performing the final wash at room temperature. Cells were then stained with 10ug/mL Hoechst 33342 (ThermoFisher) for 15 minutes before washing with PBS and imaging in with a photoprotective buffer (10% glucose, 200mM Tris, glucose oxidase (GLOX), tris(2-carboxyethyl) phosphine (TCEP), 1mM methyl viologen hyrdrate, 1mM ascorbic acid).

FISH imaging and analysis

Cells were imaged with a Leica DMi8 using a 100X oil objective, a pixel size of 110nm x 110nm, and a z stack step size of 100nm. Chromatic aberrations were corrected for using BUnwarpJ (Arganda-Carreras et al., 2006) as described previously (Giorgetti et al., 2015). After nuclei segmentation, the z slice with the maximum intensity for each foci was chosen as the z coordinate. Each selected z slice was then fit to a two-dimensional Gaussian to determine the xy coordinates for the centroid of the foci. Euclidean distances between each pair of foci and between each labeled promoter focus and the nearest nuclear edge were calculated.

1.9.2. QUANTIFICATION AND STATISTICAL ANALYSIS

Modeling simulations

All models were simulated using Gillespie algorithm provided in the Tellurium package in Python 2.7(Choi et al., 2018). Plotting of simulation results were done in MATLAB. A detailed description of the models can be found in the mathematical appendix.

Image analysis of time-lapse movies

Image segmentation

Segmentation of progenitor cells were performed in MATLAB (Mathworks, Natick, MA) using custom scripts previously (Kueh et al., 2016; Ng et al., 2018). The segmentation algorithm was performed on CFP fluorescent signal as all transduced cells carried a H2B-CFP reporter cassette. Briefly, images undergo (1) correction by subtraction of uneven background signal stemming from the bottom of the glass plate or the side of the PDMS microwells (2) gaussian blur followed by pixel value saturation to fix uneven signal intensity within the nucleus of the cell and (3) Laplacian edge detection algorithm to identify the nucleus boundary. Non-cell objects were excluded via size and shape limit exclusions, and segmentation parameters were chosen such that the number of non-cell objects are <1% of the total segmented cells.

Identification of live and dead cell population

While imaging cMyc or empty vector (EV) transduced cells, we noticed that live and dead cells possess different CFP nuclear signals. Particularly, live cell nuclei have CFP fluorescence constituting a round, smooth oval shape while dead cell nuclei CFP tend to be more granular, containing distinctively small but very bright puncta. To provide unbiased recognition of live and dead cells, individual segmented cell’s CFP image patch underwent Laplacian mask filer to delineate the ‘smoothness’ of the signal and then threshold-cutoff to identify regions with high CFP signal. Resulting features such as the object's areas, perimeter, log(CFP), and puncta numbers are recorded for each cell object. Approximately 100 individual cell image patches (10% of each data set) were then manually annotated as ‘live’ or ‘dead’ by trained scientists. The results were then linked to the above feature matrix. Decision tree supervised machine learning algorithm implemented by fitctree function in MATLAB was then used to generate a model based on the annotated live/dead classification and matrix features of the training images (Figure S3). Finally, built-in MATLAB model evaluation functions resubLoss and crossval were used to validate that mis-assignment error is below 15% for all data sets. Such an approach offers an objective, automated method to distinguish between live and dead populations. All scripts for this procedure are available upon request.

Bcl11b activation rate fitting

To measure Bcl11b activation rate, experimentally, Bcl11b RFP+ cells were cultured on stromal cell-free, DL1-coated system(Varnum-Finney et al., 2003), and activation of Bcl11b YFP allele was monitored in time-lapsed imaging. This stromal-free system enables a greater fold enhancement of cell division rate by cMyc transduction and better resolution for imaging as well as recapitulating Bcl11b activation and T cell lineage commitment, but supports a lower baseline rate of proliferation in unmodified cells compared to the OP9-DL1 system.

For quantitative measurement of this activation rate, first, the YFP and RFP signal intensity of segmented cells were calculated. Then each cell object underwent live/dead classification prediction by trained model as described in the previous section. Only cells that are classified as ‘live’ were selected, and their YFP RFP fluorescence 2D histogram is fitted to a two-component mixed 2D Gaussian model to obtain the fraction of YFP OFF and ON cells in the population at a given time. All the following procedures were implemented in MATLAB. Specifically:

To calculate fluorescent values of the Bcl11b YFP and Bcl11b RFP signals, pixel intensity of an annulus surrounding the segmented cell were calculated and subtracted from the raw signal intensity of the cell interior. This is to eliminate autofluorescence from the bottom of the glass plate as well as the PDMS microwells' edge.

To obtain the time evolution of Bcl11b biallelic population fractions from initial Bcl11b YFP-RFP+ population, cells were first filtered based on their 'live/dead' category, and only 'live' cells were included in further calculation. We used a modified version of least-squares fit of a two-component mixed 2D Gaussian function described by (Ng et al., 2018) to fit the 2D histogram of Bcl11b YFP and Bcl11b RFP fluorescence levels. Specifically, let y and r be the intensity of Bcl11b YFP and Bcl11b RFP fluorescence, respectively, the overall fit, F(r,y), is given by:

$$F(r,y) = \sum_{i=1}^{2} f_i(r,y) \qquad (1)$$

Each 2D gaussian f is given by:

$$f_i(r,y) = \frac{N_i}{2\pi\sigma_{r,i}\sigma_{y,i}\sqrt{1-\rho_i^2}} \cdot \exp\left(-\frac{1}{2(1-\rho_i^2 r^2)}\left[\frac{(r-\mu_{r,i})^2}{\sigma_{r,i}^2} + \frac{(y-\mu_{y,i})^2}{\sigma_{r,i}^2} + \frac{2\rho(r-\mu_{r,i})(y-\mu_{r,i})}{\sigma_{r,i}\sigma_{y,i}}\right]\right) \qquad (2)$$

Here, i=1,2 correspond to the red mono-allelic and biallelic populations, since all starting cells are red mono-allelic, we excluded the other two populations (non-expressing and yellow mono-allelic). Ni is the volume under the gaussian curve when integrated over r and y and is the approximation for the number of cells in each population in Bcl11b RFP mono-allelic and biallelic states.

To fit our data to F(r,y), we followed a two-step process described previously (Ng et al., 2018): (1) We fitted Bcl11b YFP/RFP 2D histogram at early time point ($0 < t < 20$) to $f_1(r,y)$ to obtain the means, standard deviations, and correlation coefficients ($\mu_{r,1}$, $\sigma_{r,1}$, $\mu_{y,1}$, $\sigma_{y,1}$, ρ_1) of the Bcl11b RFP mono-allelic population. At this time point, all cells have inactive Bcl11b YFP allele. (2) Next, we fitted the 2D histograms of Bcl11b YFP/RFP levels at successive time bins of 20 hours, fixing the parameter of the first Gaussian $f_1(r,y)$, and enabling the parameters for the second Gaussian $f_2(r,y)$, to vary within bounds observed in the fluorescent distributions of Bcl11b biallelic populations. After fitting, the fraction of biallelic cell at a given time window centered on time t is given by:

$$f_2^{obs}(t) = \frac{N_2(t)}{N_1(t) + N_2(t)} \qquad (3)$$

The confident bounds for $f_2^{obs}(t)$ is given by:

$$\delta f_i^{obs}(t) = f_i^{obs} \sqrt{\left(\frac{\delta N_i}{N_i}\right)^2 + \frac{\sum_{i=1}^{2} \delta N_i^{\,2}}{\left(\sum_{j=1}^{2} N_j\right)^2}} \qquad (4)$$

Afterward, the resulting fraction of biallelic cells as a function of time window centered at time t from the mixed Gaussian fit was then fitted to the probability density function of a first order process:

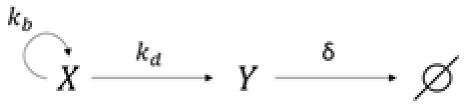
$$F_{bi}(t) = 1 - e^{-\lambda t} \qquad (5)$$

Where is the activation rate of Bcl11b from RFP-monoallelic to biallelic state. We chose this function for activation rate fitting since our histone dynamics simulations suggested that Bcl11b activation can be estimated as a first order stochastic process (see Figure 4D). For this function, fitting was done using MATLAB fit function, and 95% confidence interval for the fit was recorded.

Population dynamics model and fitting

We built a mathematical model to describe the population dynamics of progenitor cells transfected with an empty vector (EV) and c-Myc. From initial inspection of time-lapse movies (Figure 3), progenitors transduced with c-Myc appear to expand more quickly than control progenitors, as expected. Faster expansion of c-Myc-transduced cells could be due to faster cell cycling or slower cell death. To disentangle these two effects, we quantified numbers of both live and dead cells over time (Figure S4) and fit these data to population dynamics models to obtain division and death rates:

In general, the model includes a population of live cell (X) with a birthrate k_b and a death rate k_d to generate the dead cell population (Y). This population in turn has a clearance rate δ denoting the process in which CFP level degrades and the dead cells become undetected.



Live population at a given time T can be described as a simple first order process:

$$X(T) = X_o e^{(k_b - k_d)T} = X_o e^{KT} \qquad (6)$$

Where $K = k_b - k_d$ and X_o is the initial number of live cells at the start of imaging. On the other hand, we adopt a stepwise approach to model the dead cell population:

$$Y(T) = Y_o P(T) + \sum_{t=0}^{T} X(t) \cdot k_d \cdot P(T-t) \qquad (7)$$

Here Y_o is initial number of live cells at the start of imaging and $P(\tau)$ is an exponential decay function describing the fraction of CFP-positive dead population remaining after a period time from its first appearance. Since dead cell's fluorescent signal is dim, segmentation of these cells tends to 'fickle' before completely disappear. In this model, whenever a cell starts to die, its probability of being detected decreases as per function P(T), and the number of dead cells at a given time T is the sum of all the still-detected dead cells generated since the start of imaging up until T. This discrete approach allows us to fit cell death data to a relatively simpler function compared to a more sophisticated two-component system of ODE model without sacrificing the complexity of dead cell detection phenomenon.

We determined $P(\tau)$ empirically for EV and cMyc population separately by manually following 30 different dead cells and record the time period in which it was detected and undetected until complete disappearance. Then we plotted how many dead cells out of 30 were detected after a given time has elapsed. An exponential function decay function was used to fit this 'fraction detected' curve and to estimate value for clearance rate (Figure S4B):

$$P(\tau) = e^{-\delta\tau} \qquad (8)$$

Here, P(τ) is the probability of a given dead cell to be detected under the CFP fluorescent channel after period of time since its initial death.   is the clearance rate of this process. To fit imaging data to equation (6). We classified segmented cell object as live or dead using trained machine learning model as described in ‘Image analysis of time-lapse movies’ section. Number of live cells as a function of time was fitted to equation (6) using MATLAB fit function to estimate , and 95% confidence interval for the fit was recorded.

To fit imaging data to equation (7), we performed fine scanning of candidate $k_{d,i}$ values from the set K_D={0,0.0005,0.001,0.0015,...,0.0495,0.05}. For each $k_{d,i}$, a predicted $Y_{p,i}$(T) curve was generated based on equation (7) where T = t_1,t_2,t_3,... with t_i being the time point at which experimental measurement took place. $Y_{p,i}$(T) is then compared to the experimentally observed dead cell number Y_{exp}(T) using sum square error method:

$$sse_i = \sum_{t=t_1,t_2,t_3,\ldots} \left[Y_{exp}(t) - Y_{p,i}(t)\right]^2 \qquad (9)$$

The best fit k_d value is chosen to be the $k_{d,i}$ value whose sse_i is the smallest.

In order to calculate the confidence bound of the fit, we utilized nonlinear regression method by first calculating the residue of the model’s predicted values $Y_p(t_i)$:

$$dY_{t_i} = Y_{exp}(t_i) - Y_p(t_i) \qquad (10)$$

Then we calculate to the Jacobian of the model function to estimate the covariance at each time point and is given by:

$$J_{t_i} = \frac{\partial Y_p(t_i)}{\partial k_d} \qquad (11)$$

These inputs were used to estimate 95% confidence interval using MATLAB ‘Nonlinear regression parameter confidence intervals’ function nlparci.

Summary of results from fitting of data to activation rate and population models are tabulated in Table 1.

1.9.3. MATHEMATICAL APPENDIX

Modeling of the polycomb-dependent epigenetic timer

Introduction

To understand the epigenetic timer controlling the *Bcl11b* activation, we used mathematical modeling to analyze a series of candidate biophysical mechanisms. In this modeling, we seek to explain the essential emergent properties of the switch, namely (1) its irreversible, all-or-none nature; (2) its long, stochastic time delay; (3) the heritability of its inactive and active states over DNA replication; and (4) its tunability with respect to changes in H3K27me3 levels and modifying-enzyme activity.

We consider two main candidate mechanisms. In the methylation only mechanism (Model I), individual nucleosomes within in a one-dimensional lattice can be methylated or unmethylated. Gene expression is assumed to occur when the total fraction of methylated nucleosomes in this lattice falls below a threshold value. In the coupled methylation compaction mechanism (Model II), individual nucleosomes are also methylated and demethylated; in addition, these nucleosomes also interact to form a compacted assembly with rates dependent on the H3K27me3 marking. Unlike Model I, gene expression does not depend directly on H3K27me3 levels, but on the compaction state of the nucleosome assembly, which in turn depends on methylation states of individual nucleosomes . Both models explicitly model DNA replication as a process involving random segregation of modified nucleosomes into daughter strands. From our analysis, we find that the coupled methylation-compaction mechanism alone explains the emergent behaviors of the *Bcl11b* epigenetic timer we observe experimentally, and thus represents our favored model for polycomb based timing control.

Model I: The Methylation Only Mechanism (M)

Here, we adopt a standard framework for histone modification dynamics previously shown to generate multi-stability (Angel et al., 2011; Dodd et al., 2007). In this model, individual nucleosomes reside in a one-dimensional lattice, and exist in two states, a methylated state, corresponding to an H3K27 tri-methylated state, and demethylated state. We do not describe multiple demethylated states in our model (i.e. mono-methylation, di-methylation, and an un-methylated state), though our analysis, together with previous work (Dodd et al., 2007), indicates that our main conclusions hold in more complex models with additional states. As with previous models, the methylation rate of a given nucleosome depends on the number and distance of methylated nucleosomes in its vicinity, reflecting observations that PRC2 can bind and be activated by H3K27me3-marked nucleosomes to write H3K27me3 on neighboring nucleosomes. Demethylation is taken to occur at a first order rate. In this model, we assume there is no spontaneous methylation in the absence of existing methylated nucleosomes; thus, once all nucleosomes are demethylated, no more remethylation is possible and the system enters an irreversibly activated state.

Methylation. We explicitly model mark binding and cooperative activities of the PRC2 complex, as well as the methylation state of each individual nucleosome. Let p_i be the methylation state of the ith nucleosome. $p_i = 0$ denotes the de-methylated state while $p_i = 1$ denotes the methylated state. Let u' and u denote the transitions between the methylation state and demethylation state, respectively. The model is set up as follows:

For i ∈ $\{1, .. , 100\}$:

$$u' : p_i[0] \rightarrow p_i[1]$$

$$u : p_i[1] \rightarrow p_i[0]$$

With:

$$Pr(u') = \beta_{ON} \cdot (1 - p_i) \cdot \sum_{j \neq i} p_j \cdot e^{\left(\frac{j-i}{L}\right)^2} \qquad (1)$$

$$Pr(u) = \alpha_{OFF} \cdot p_i \qquad (2)$$

The parameter L can be interpreted as the ‘reach’ of the PRC2 complex to neighboring nucleosomes. Large value of L indicates a long length scale for interaction with nearby nucleosomes. This effect is set to have a gaussian shape so that nucleosome closest to the anchored PRC2 complex has the highest methylation rate. Similar distributions of activity have been reported for artificially tethered enzymes (Hass et al., 2015), as well as for histone modifications around transcription factor binding sites (Heinz et al., 2010). Moreover, we assume periodic boundary conditions for the one-dimensional lattice, though similar results were observed with other non-repeating boundary conditions (not shown).

Cell division. To model the transmission of histone marks across cell divisions, we assume that methylated nucleosomes segregate randomly to the two daughter DNA strands upon replication; thus, each nucleosome position has a probability $p = 0.5$ of inheriting a nucleosome that is methylated. Experimental evidence suggests that approximately 50% of total global H3K27me3 partitioning of parental marks to the subsequent generations (Alabert et al., 2015).

From stochastic simulations of this model, we find that this methylation only mechanism can generate a slow, heritable, and stochastic gene switch (see Results and Figure 1.4); however, switching times are hypersensitive to mild changes to methylation and de-methylation rates, and therefore inconsistent with experimental results. To understand the origins of this hypersensitivity, we re-formulate this model using a chemical kinetics framework amenable to analysis using transition state theory. By considering the limit where $L \rightarrow \infty$, such that each H3K27me3-bound PRC2 methylates all other un-methylated nucleosomes with the same reaction rate. In this limit, given N as number of unmethylated nucleosome, we can completely describe the state of the system by the number of methylated nucleosomes N’. As the rates of adding or subtracting one methylated nucleosome from the system would reduce to become a function of N’, independent of spatial arrangement. Consequently:

$$N \underset{k_2}{\overset{k_1}{\rightleftarrows}} N'$$

where

$$k_1 = \beta N'(N_T - N') \qquad (3)$$

$$k_2 = \alpha N' \qquad (4)$$

and N_T is the total number of nucleosomes. The master equation describing the time evolution of this system is given by:

$$\frac{\partial p_n}{\partial t} = -[k_1(n) + k_2(n)] \cdot p_n + k_1(n-1)p_{n-1} + k_2(n+1)p_{n+1} \qquad (5)$$

where p_n is the probability of having N’ methylated nucleosomes. When the total number of nucleosomes is large, we can approximate the number of methylated nucleosomes to be a continuous variable x. In this limit, we can rewrite the master equation as Fokker-Planck equation:

$$\frac{\partial p(x,t)}{\partial t} = \frac{\partial}{\partial x}[v(x)p(x)] + \frac{1}{2} \cdot \frac{\partial^2}{\partial x^2}[D(x)p(x)] \qquad (6)$$

where, we have ignored third and higher order terms, and where:

$$v(x) = k_1(x) - k_2(x) \qquad (7)$$

$$D(x) = k_1(x) + k_2(x) \qquad (8)$$

Given the velocity and diffusion constants for this system as a function of methylated nucleosome number, the switching of the system is essentially given by the first-passage time of the system to reach the absorbing state x = 0. However, a closed-form solution of this first-passage time distribution for the given rate functions is hard to obtain; Nevertheless, we note that our system operates in the regime where the timescales of individual methylation and demethylation reactions are much shorter than switching times for this system. In this regime, switching times are well described by the Kramer's theory for escape of a Brownian particle over a potential well (Kramers, 1940), and would thus approximately scale exponentially with the height of a potential energy barrier. We can obtain the functional form of this potential barrier by relating it to the velocity function:

$$-\frac{dV}{dx} = v(x) \qquad (9)$$

From equations (3), (4), and (7), we can then integrate the system to explicitly derive the potential function:

$$V(x) = \frac{\beta}{3}x^3 + \frac{\alpha - \beta N_T}{2}x^2 + W \qquad (10)$$

Where W is an arbitrary number. A plot of $V(x)$ is demonstrated in Figure S1.3C. The energy landscape possesses a local minimum at a nonzero value of x, indicating the metastable state. The landscape has a local maximum near $x = 0$. State switching occurs when the system reaches the absorbing state x = 0. Thus, we define E_a as the height in $V(x)$ between the local maxima and the metastable state minima. When we plotted this potential energy for different values of β (Fig. S1.3C), we found that moderate changes in β led to significant changes in potential well height. As switching rates scale roughly exponentially with well height, we would expect this system would show extreme sensitivity in switching times with respect to changes in methylation rate changes.

Model II: The Methylation Compaction Mechanism (MC)

As the methylation only mechanism above does not fully explain the tunable characteristics of the Bcl11b activation timing switch, we considered a second model, where histone methylation is coupled to chromatin compaction. Indeed, H3K27me3 has been suggested to meditate a condensed, polymerase-inaccessible chromatin conformation. For instance, the polycomb repressive complex 1 (PRC1), which binds H3K27me3 and is important for compaction and gene silencing, oligomerizes through contacts on its Bmi1 or Phc subunits (Eskeland et al., 2010; Gray et al., 2016; Isono et al., 2013; Kahn et al., 2016), and may also undergo weak, multivalent interactions on Cbx2 that result in liquid-liquid phase separation and gene silencing (Howard, 2001; Larson et al., 2017). This model consists of two main modules: (1) a H3K27 methylation and demethylation mechanism, and (2) a dynamic chromatin decompaction mechanism linked to H3K27me3 modification stat that ultimately underlies gene switching. In our description of compaction dynamics, we do not explicitly model the spatial extent of the compacted assembly; instead, we adopt a mean-field approach that is established in models of cytoskeletal polymer dynamics (Erickson and Pantaloni, 1981; Jackson and Berkowitz, 1980). With this approach, the numbers of un-methylated and methylated nucleosomes within a compacted assembly are given by C and C’ respectively, along with those outside the assembly are given by D and D’ respectively. As a result, the dynamical system is described by four states: 1) Compacted-Methylated 2) Compacted-Demethylated 3) Decompacted-Methylated and 4) Compacted-Demethylated:

$$\begin{tikzcd}
C' \arrow[r, shift left, "k_1"] \arrow[d, shift right, "k_5"'] & D' \arrow[l, shift left, "k_2"] \arrow[d, shift right, "k_7"'] \\
C \arrow[r, shift left, "k_3"] \arrow[u, shift right, "k_6"'] & D \arrow[l, shift left, "k_4"] \arrow[u, shift right, "k_8"']
\end{tikzcd}$$

Here, C’, C, D’, and D denote the number of nucleosomes in these states, respectively. $K_{1\text{-}8}$ denote the transition rates between them, which will be defined below. Gene activation is defined to be the system state where all nucleosomes exist in a decompacted state. The two mechanisms are intertwined so that methylation states affect compaction rates and vice versa. Detailed descriptions of the rates are given below:

Methylation. In this model, un-methylated nucleosomes convert into a methylated state with a first-order rate constant β. We assume this rate constant is the same regardless of whether nucleosomes are inside or outside the compacted assembly. Methylated nucleosomes convert into a de-methylated state with a rate constant of α if the nucleosome is outside the assembly (D’), or a lower rate constant of $f\alpha$, ($f < 1$) if the nucleosome is inside the assembly. This lower rate constant assumes that the demethylation reaction is less efficient on compacted nucleosomes, possibly due to competition for demethylase binding by compaction proteins, or due to the exclusion of demethylases through steric occlusion or phase separation. The rates describing these reactions on the four nucleosomal species are given by:

$$C' \underset{k_6}{\overset{k_5}{\rightleftarrows}} C \qquad\qquad D' \underset{k_8}{\overset{k_7}{\rightleftarrows}} D$$

Where:

$$k_5 = f\alpha\, C' \qquad (11)$$
$$k_6 = \beta\, C \qquad (12)$$

$$k_7 = \alpha\, D' \qquad (13)$$
$$k_8 = \beta\, D' \qquad (14)$$

H3K27 methylation and demethylation rates are based on the catalytic activity of the EZH2 subunit the PRC2 complex and JMDJ3/UTX proteins, respectively. Specifically, these rate constants were chosen to represent the conversion between H3K27me2 and H3K27me3. For simplicity, we do not model the H3K27me-binding dependent H3K27 methylation activity previously described (Margueron et al., 2009), though we show below that explicit modeling of this cooperative effect would not significantly alter the conclusion of the model. JMJD3 and UTX demethylate H3K27me3 (Agger et al., 2007), and to our knowledge no cooperative activity of these complexes have not been reported.

Compaction. We adopt a mean-field description of the compacted nucleosomal assembly, following kinetic models of multi-stranded cytoskeletal polymer assembly (Howard, 2001). This description assumes that the nucleosome assembly is a roughly spherical structure held together by weak, multivalent interactions between individual nucleosomes, and can add or lose individual nucleosomes at its surface. Both methylated and demethylated nucleosomes can incorporate into the assembly; thus the assembly has a total size of:

$$C_T = C + C' \qquad (15)$$

where C and C' represent the number of methylated and demethylated nucleosomes in the assembly, respectively. Unlike other polymer models (MacPherson et al., 2018; Nuebler et al., 2018) , we do not explicitly model physical connections between nucleosomes due to DNA; such connections would be expected to result in a spatial dependence of reaction rates within this chromatin domain; however, as the entire domain (100 nucleosomes) has a length scale greater than the persistence length of chromatin (~15-20 nucleosomes, from (Arbona et al., 2017), and would thus enable free interactions between non-neighboring nucleosomes, we would expect the essential properties of our minimal model in a more realistic physical model that incorporates nucleosome connectedness.

The addition and removal of methylated and demethylated nucleosomes from the assembly is described by the following rate equations:

$$C' \underset{k_2}{\overset{k_1}{\rightleftarrows}} D' \qquad\qquad C \underset{k_4}{\overset{k_3}{\rightleftarrows}} D$$

Where:

$$k_1 = \frac{\delta}{C_T^{\frac{1}{3}}} C' \qquad (16)$$

$$k_2 = \lambda\, C_T^{\frac{2}{3}} D' \qquad (17)$$

$$k_3 = \frac{\delta}{C_T^{\frac{1}{3}}} C \qquad (18)$$

$$k_4 = F\, \lambda\, C_T^{\frac{2}{3}} D \qquad (19)$$

if $C_T > C_N$

$$k_1 = \frac{\delta}{C_T^{\frac{1}{3}}} C'' \qquad (20)$$

$$k_2 = 0 \qquad (21)$$

$$k_3 = \frac{\delta}{C_T^{\frac{1}{3}}}\, C \qquad (22)$$

$$k_4 = 0 \qquad (23)$$

$$\text{if } C_T < C_N$$

Here, methylated nucleosomes incorporate into the compacted assembly with a rate constant λ; however, importantly, demethylated nucleosomes can also incorporate into the assembly with a reduced rate constant $F\lambda$ (where $F < 1$). The effect of methylation state on compaction rate is experimentally observed in instances such as recruitment of PRC1 complex by H3K27me3 marks (Kahn et al., 2016). The complex's subunits such as Ring1B and Phc-1 have been shown to be important in chromatin compaction and gene silencing (Eskeland et al., 2010; Francis et al., 2004; Isono et al., 2013). However, as PRC1 recruitment is not the only compaction mechanism in vivo, and because PRC1 can bind to nucleosomes independently of H3K27me3 (Francis et al., 2004), this model treats methylation as only a part, but not solely responsible for chromatin condensation. In choosing rate constants; we assume that compaction and decompaction is faster than histone methylation and demethylation rates, though timescales for both processes are assumed to be much faster than that for cell division. Fast compaction kinetics relative of modification is supported by in vitro studies of H3K27me3 methylation and demethylation kinetics, as well as in vitro DNA compaction by HP1α and chromatin condensation experiments (Kristensen et al., 2011; Ladoux et al., 2000; Larson et al., 2017; Sneeringer et al., 2010).

The reaction rates for nucleosome incorporation (loss) scales with assembly size as $\sim C_T^{\frac{2}{3}}$ ($\sim C_T^{-\frac{1}{3}}$), as these reactions only take place on the surface of the assembly. Assuming a compacted nucleosome complex is spherical, the compaction rate would thus be proportional to the surface area. Likewise, the decompaction rate is also proportional to the surface area but reversely proportional to the total number compacted nucleosome in the complex.

In this description, there is a critical threshold number of compacted nucleosomes, C_N, below which the complex is thermodynamically unstable. The existence of a minimal nucleus size is a fundamental property of multivalent polymers, whereby addition of a new subunit to an already formed complex is thermodynamically more favorable than formation of the initial nucleus itself (Erickson and Pantaloni, 1981; Jackson and Berkowitz, 1980). Below this critical threshold C_N, the compacted assembly disintegrates, and gene turns on.

Cell division. Heritability of histone marks and chromatin states are crucial in maintaining gene expression states across cellular generations. As with the methylation only model above, we assume that methylated nucleosomes partition randomly between two daughter strands upon replication; the total number of nucleated nucleosomes is then obtained by sampling from binomial distribution with p = 0.5 and N equal to the total number of nucleosomes at the point of DNA replication. Furthermore, we assume that compacted nucleosomes persist within a compacted assembly reside upon passage of DNA polymerase. This model feature assumes that new nucleosomes rapidly incorporate into a compacted assembly after passage of DNA polymerase; however, in the subsequent version of this model below, we will relax this assumption to allow for disruption of compaction state by DNA polymerase passage (see below).

From Monte-Carlo simulations, we found that this dynamic methylation compaction model can recapitulate all the essential emergent properties of the Bcl11b activation switch. Specifically, this model shows the following dynamic properties:

1) Irreversible all-or-none switching to an H3K27me3-low, de-compacted state. From simulations, we found that the system adopts a stable compacted assembly of nucleosomes with higher H3K27me3 marking density, but switches abruptly to a de-compacted state with lower H3K72me3 levels. As there is no re-nucleation of the compacted assembly after its elimination, this de-compacted state represents an absorbing, permanently active expressing state. The abrupt decrease in the H3K27me3 levels arises because compacted nucleosomes demethylate at a lower rate; thus, upon total decompaction, the percent of methylated nucleosomes lowers to a new steady state level.

2) Noise induced gene activation. Transition to the completely decompacted state, or gene activated state, occurs via stochastic deviation of the system from its compaction meta-stable state. Activation is triggered when the system reaches below the threshold number of compacted nucleosomes.

3) Tunable activation rates. The model is able to generate a gene switch with slow, tunable activation rate. Delay in activation is in order of days and can be finely adjusted by modifying methylation and demethylation rates, and/or changing H3K27me3 levels at the gene locus (Figure 1.4E), as experimentally observed (Figure 1.2). This ability to tune activation rates by changing H3K27me3 densities uniquely distinguishes this methylation compaction model from the methylation only model above, and thus represents a more plausible model for describing the activation mechanism of this switch. Why is this model uniquely tunable? In this model, locus de-compaction and gene activation are determined by a dynamic balance between rates of nucleosome entry or exit from a compacted assembly. The system still be sensitive to changes in these rates; however, as demethylated nucleosomes can still enter and exit a compacted assembly at a reduced rate, changes in the fraction of demethylated nucleosomes would cause a fine change in these entry or exit rates, and thus represent a plausible tuning parameter for controlling activation timing.

4) Division-independent timing control. When the cell cycle length is changed in this model, activation kinetics remain largely unaffected, implying that the methylation-compaction mechanism functions as a cell division-independent delay timer. These conclusions hold, as long as the dynamic methylation and compaction mechanisms operate on timescales much faster than the cell cycle length.

Why is the methylation compaction tunable? To answer this question, we adopt an approach, where we reduce this problem to using the Fokker-Planck approach, as utilized to analyze the Methylation only mechanism (Figure S1.3C). The full system with both methylation and compaction reactions would correspond to diffusive motion of a particle in a three-dimensional state space describing both chemical and physical states of nucleosomes. However, to simplify this problem to gain intuition, we will first take the methylation and demethylation reactions to be fast compared to the compaction and de-compaction reactions, such that the system can be described a single parameter NC, corresponding to the total number of compacted nucleosomes. At any given time, the number of methylated and demethylated nucleosomes in the compacted state is at quasi-steady state, with values:

$$C' = \frac{\beta}{\beta + f\alpha} \cdot C_T \qquad (24)$$

and

$$C = \frac{f\alpha}{\beta + f\alpha} \cdot C_T \qquad (25)$$

Similarly, assuming quasi steady-state, the number of methylated and demethylated nucleosomes in the uncompacted state is given by:

$$D' = \frac{\beta}{\beta + \alpha} \cdot D_T \qquad (26)$$

and

$$D = \frac{\alpha}{\beta + \alpha} \cdot D_T \qquad (27)$$

Let $N_T = C_T + D_T$. With this approximation, the averaged rate of adding or removing a nucleosome from the compacted assembly is then given by:

$$k_{add} = k_2 + k_4 = \lambda C_T^{\frac{2}{3}} \frac{\beta}{\beta + \alpha} \cdot D_T + F \cdot \lambda C_T^{\frac{2}{3}} \frac{\alpha}{\beta + \alpha} \cdot D_T$$

$$= \left(\frac{\beta}{\alpha} + F\right) \lambda C_T^{\frac{2}{3}} \cdot \frac{N_T - C_T}{1 + \frac{\beta}{\alpha}} \qquad (28)$$

$$k_{remov} = k_1 + k_3 = \frac{\delta}{C_T^{\frac{1}{3}}} C' + \frac{\delta}{C_T^{\frac{1}{3}}} C$$

$$= \delta C_T^{\frac{2}{3}} \qquad (29)$$

Let the total number of compacted nucleosomes C_T be x. By writing down the master equation for this system, and by further applying the Fokker-Planck approximation, as performed in (5) and (6) we then have:

$$\frac{\partial p(x,t)}{\partial t} = \frac{\partial}{\partial x}[v(x)p(x)] + \frac{1}{2} \cdot \frac{\partial^2}{\partial x^2}[D(x)p(x)] \qquad (30)$$

where:

$$v(x) = \delta x^{\frac{2}{3}} - \left(\frac{\beta}{\alpha} + F\right) \lambda x^{\frac{2}{3}} \cdot \frac{N_T - x}{1 + \frac{\beta}{\alpha}} \qquad (31)$$

$$D(x) = \delta x^{\frac{2}{3}} + \left(\frac{\beta}{\alpha} + F\right) \lambda x^{\frac{2}{3}} \cdot \frac{N_T - x}{1 + \frac{\beta}{\alpha}} \qquad (33)$$

As before, we define a potential energy for this system:

$$-\frac{dV}{dx} = v(x) \qquad (34)$$

The analytical solution for the potential energy $V(x)$ for the methylation compaction model is:

$$V(x) = \frac{3}{5}\left[\delta - \frac{N}{1+\frac{\beta}{\alpha}}\left(\frac{\beta}{\alpha}+F\right)\lambda\right]x^{\frac{5}{3}} + \frac{3}{8}\cdot\frac{1}{1+\frac{\beta}{\alpha}}\left(\frac{\beta}{\alpha}+F\right)\lambda x^{\frac{8}{3}} + W \qquad (35)$$

A plot of $V(x)$ is demonstrated in Figure S1.3B-C. We found that increasing methylation rate results in a much more attenuated increase in activation energy E_a with the methylation compaction model. This confirms that the improved switching rate tunability in the MC model stems from the decreased sensitivity to changes in activation barrier height by methylation rate. This result intuitive explains why this system shows significantly more graded changes in switching times when methylation rates are changed.

This tunability of switching times with respect to histone methylation depends on the relative association strengths of demethylated and methylated nucleosomes for each other in forming a compacted assembly. In our initial simulations, demethylated nucleosomes show only a moderate decrease in affinity for other compacted nucleosomes relative to methylated nucleosome (F = 0.85). However, when the binding strength of a demethylated nucleosome is much weaker than that of a methylated nucleosome (F = 0.2), we find changes in potential well heights become more significant, indicating that the system loses its tunability with respect to methylation changes (see Figure S1.6E-D). This prediction, that methylated and demethylated nucleosomes have comparable strengths of association for a compacted assembly agrees well with evidence that unmethylated nucleosomes can nonetheless aggregate through a variety of H3K27me-independent mechanisms (Larson et al., 2017; Strom et al., 2017).

Model II.1: The Methylation Compaction Mechanism, with Compaction Disrupted by Division (Figure S1.5)

This version of the model includes modified cellular division process in which upon replication, 50% of methylated nucleosomes become demethylated and 10% of compacted nucleosomes become uncompacted. This exit of nucleosomes from a compacted assembly due to DNA replication reflects the possibility that as the DNA replication machinery enters the compacted nucleosomal structure, it creates decompaction ‘defects’ in the condensed locus because nucleosomes near the replication forks are replaced. However, we reason that such defect would have a small effect to the overall stability of the structure because, at any given time, the site of replication would only take up a small region of the entire compacted domain.

In order to simulate both changes in methylation and compaction at the point of DNA replication, we must describe probabilistically how each of the four nucleosomal species are affected: 1) The Compacted-Methylated species (C’); 2) the Compacted-Demethylated species (C); 3) the Decompacted-Methylated species (D’); and 4) the Compacted-Demethylated species (D). Since methylation state is reduced by 50%, approximately half of Decompacted-Methylated species is transferred to Decompacted-Demethylated pool. Similarly, on average, 10% of the Compacted-Demethylated species are transferred to Decompacted-Demethylated pool due to DNA replication. Compacted-Methylated species have 50% chance to demethylate and 10% chance to decompact. Assuming these are two independent processes, this species has 5% chance to convert into Decompacted-Demethylated or Decompacted-Methylated and 45% chance to become Compacted-Demethylated. These observations are implemented as follows:

Let vector $\boldsymbol{S} = [S_1, S_2, S_3, \dots, S_n]$ be the result from sampling a multinomial distribution with probabilities $\pi_1, \pi_2, \pi_3, \dots, \pi_n$, where $\pi_1 + \pi_2 + \pi_3 + \cdots + \pi_n = 1$. Let $S_i(N|\pi_1, \pi_2, \pi_3, \dots, \pi_n)$ be the i^{th} element of $\boldsymbol{S}$ and N be the sample size. Let c', c, d', d be the number of compacted-methylated, compacted-unmethylated, decompacted-methylated, and decompacted-unmethylated nucleosomes, respectively immediately preceding the cellular division event. Partitioning of each species occurs as follows:

$$C' = S_4(c'|0.45,0.05,0.05,0.45) \tag{36}$$

$$D' = S_1(d'|0.5,0.5) + S_2(c'|0.45,0.05,0.05,0.45) \tag{37}$$

$$C = S_2(c|0.1,0.9) + S_1(c'|0.45,0.05,0.05,0.45) \tag{38}$$

$$D = d + S_2(d'|0.5,0.5) + S_1(c|0.1,0.9) + S_3(c'|0.45,0.05,0.05,0.45) \tag{39}$$

From stochastic simulations (Figure S1.5), we find that this modified methylation compaction model shows similar dynamic characteristics compared to the original methylation compaction model (Model II): it shows stochastic, all-or-none switching between inactive and active states; has an activation delay that can be tuned by changing H3K27me levels and enzyme activity; and shows division-independence in its activation time delay. Thus, we conclude that the essential features of this model hold, even upon mild disruption of the inactive, compacted assembly by passage of DNA polymerase.

Model II.2: The Methylation Compaction Mechanism with Cooperative Methylation (Figure S1.4)

PRC2 is known to be allosterically activated by H3K27me3 binding via its EED subunit (Margueron et al., 2009). Here, we consider this cooperative property of PRC2 by specifying that methylation rate increases with the total number of methylated nucleosomes in the model system. This assumption is likely valid when the number of nucleosomes in the condensed structure is small, and all the nucleosomes are more or less in close proximity with each other. To simulate this, we modified the methylation rates K_6 and K_8 so that their magnitude has a spontaneous term μ and the cooperative term β that is proportional to the total number of methylated species in the simulation:

$$C' \underset{k_6^*}{\overset{k_5}{\rightleftarrows}} C \qquad\qquad D' \underset{k_8^*}{\overset{k_7}{\rightleftarrows}} D$$

$$k_6{}^* = [\mu + \beta(C' + D')]C \tag{40}$$

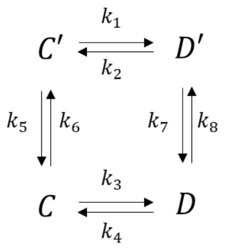
$$k_8{}^* = [\mu + \beta(C' + D')]D \tag{41}$$

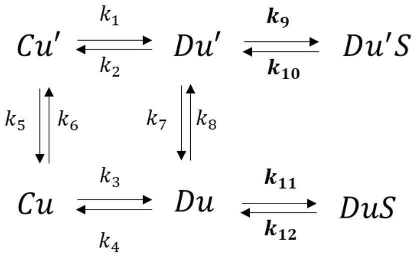
From stochastic simulations (Figure S1.4), we find that this system is also capable of generating long, stochastic delays in all-or-none switching in locus compaction state, and that switching times can be finely tuned by changing H3K27me3 levels, as with our simpler methylation compaction model (Model II). We conclude that incorporation of a cooperative H3K27me3 methylation rate in our methylation compaction model does not alter its main conclusions.

Model II.3: Transcription action in activation rate tuning in compaction methylation mechanism (Figure 1.6)

Model II.3.1. Transcription factor affects compaction mechanism

To simulate the potential effect of transcription factors on tuning the activation timing in the methylation compaction model. We first consider a scenario where the transcription factors prevent a small number of nucleosomes from compaction. This can be accomplished by designating a small portion of nucleosomes to carry binding sites for the transcription factors. Upon binding to these factors, these nucleosomes can no longer form compaction:





'Unique' nucleosomes carrying transcription factor binding sites are labeled Cu, Cu', Du, and Du'. Reaction rates $k_{1\text{-}8}$ are the same as the standard model. When these nucleosomes are decompacted, they can bind to the transcription factors with the following rates:

$$k_9 = K_{ON} \cdot TF \cdot Du' \qquad (42)$$

$$k_{10} = K_{OFF} \cdot Du'S \qquad (43)$$

$$k_{11} = K_{ON} \cdot TF \cdot Du \qquad (44)$$

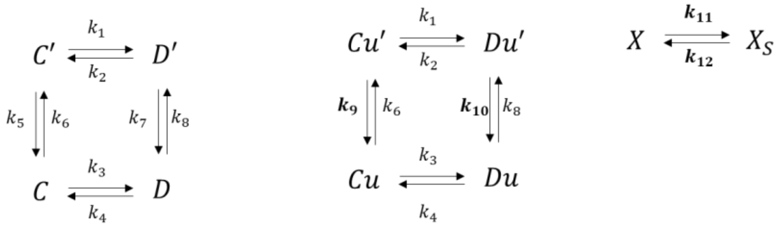
$$k_{12} = K_{OFF} \cdot DuS \qquad (45)$$

Here TF is the concentration of transcription factors. Let C_T be the total number of compacted nucleosomes and let N_T be number of nucleosomes that carries the transcription factor binding sites. We run our simulation with the assumption that the total number of nucleosomes is constant, and N_T can be varied.

Our simulations (Figure 1.6A) suggest that activation timing can be tuned with the transcription factor copy number. Additionally, activation rate increases synergistically with number of binding sites. Therefore, timing modulation can be achieved via transcription factor acting as nucleosome sequester.

Model II.3.2. Transcription factor affects methylation mechanism

We next consider an alternative mechanism where a single transcription is recruited to the nucleosome assembly. There, it acts as a demethylase and removes methyl marks on the neighboring nucleosomes within a specific range:



Here Cu, Cu', Du, and Du' are nucleosomes that are in the recruited demethylase's reach. The unrecruited demethylase is denoted as X, and XS is the recruit, transcription factor (TF) – bound demethylase. Since the simulation involves one compaction assembly, the number of demethylase is limited to 1, or $X + X_S = 1$. Conversion rate probability between X and XS and the demethylation rate of the nucleosomes inside the reach of the methylase are described below:

$$k_9 = \alpha_{TF} \cdot X \cdot f \cdot Cu' \qquad (46)$$

$$k_{10} = \alpha_{TF} \cdot X_S \cdot Du' \qquad (47)$$

$$k_{11} = K_{ON} \cdot TF \cdot X \qquad (48)$$

$$k_{12} = K_{OFF} \cdot X_S \qquad (49)$$

Here, $\alpha_{TF}$is the transcription factor recruited demethylation rate constant, and f is the fraction reduction of the rate when the nucleosome is decompacted. Unlike transcription factors that affect compaction, transcription factor that recruits demethylase according to this mechanism require a large effective range (approximately 50% of the overall number of nucleosomes) to appreciably tune switching rate (Figure 1.6D).

Parameter List

a. Supplementary Table S1.2. Model Parameters: Pure Methylation Model (Figure 1.4A,B)

Pure Methylation Model			
Parameters	**Description**	**Value**	**Comments/references**
β_{ON}	cooperative methylation rate constant	1 hrs^{-1}	Sneeringer et al., 2010
α_{OFF}	demethylation rate constant	16.5 hrs^{-1}	[16 , 17 , 17.5] for parameter scan
			Kristensen et al., 2011
L	reach of anchored methylating enzyme to neighboring nucleosome	15	Hass et al., 2015
N	number of simulated nucleosomes	100	
cell division length	cell division length	20 hrs	

b. Supplementary Table S1.3. Model Parameters: Compaction Methylation Model (Figure 1.4C,D)

Compaction Methylation Model			
Parameters	**Description**	**Value**	**Comments/references**
β	methylation rate constant	1 hrs^{-1}	[0.2 , 0.4 , 0.8 , 1.6 , 3.2 , 6.4] for parameter scan
			Sneeringer et al., 2010
α	demethylation rate constant	8 hrs^{-1}	Kristensen et al., 2011
f	fraction of de-methylation rate when nucleosome is in compacted state	0	
λ	compaction rate constant	310 hrs^{-1}	Larson et al., 2017; Ladoux et al., 2000
δ	decompaction rate constant	5300 hrs^{-1}	Larson et al., 2017; Ladoux et al., 2000
F	fraction of compaction rate when nucleosome is in demethylated stated	0.85	
N	Number of simulated nucleosomes	50	
C_N	nucleation threshold for compacted nucleosomal complex	5	
cell division length	cell division length	20 hrs	

c. Supplementary Table S1.4. Model Parameters: Cell division dependence of Pure Dilution vs Compaction Methylation Models (Figure 1.4 F-G)

Pure Dilution Model			
Parameters	**Description**	**Value**	**Comments/references**
β_{ON}	cooperative methylation rate constant	0	
α_{OFF}	demethylation rate constant	0	
L	reach of anchored methylating enzyme to neighboring nucleosome	0	
N	number of simulated nucleosomes	100	
cell division length	cell division length	20 hrs	[10 , 15 , 20 , 25 , 30 , 35] for cycle scan
Compaction Methylation Model			
Parameters	**Description**	**Value**	**Comments/references**
β	methylation rate constant	1 hrs^{-1}	Sneeringer et al., 2010
α	demethylation rate constant	8 hrs^{-1}	Kristensen et al., 2011
f	fraction of de-methylation rate when nucleosome is in compacted state	0	
λ	compaction rate constant	310 hrs^{-1}	
δ	decompaction rate constant	5300 hrs^{-1}	
F	fraction of compaction rate when nucleosome is in demethylated stated	0.85	
N	Number of simulated nucleosomes	50	
C_N	nucleation threshold for compacted nucleosomal complex	5	
cell division length	cell division length	20 hrs	[10 , 15 , 20 , 25 , 30 , 35] for cycle scan

d. Supplementary Table S1.5. Model Parameters: Effects of Transcription Factors on Methylation Compaction Model's Activation Timing (Figure 1.6)

Transcription Factor Affecting Compaction Model			
Parameters	**Description**	**Value**	**Comments/references**
β	methylation rate constant	6.4 hrs^{-1}	Sneeringer et al., 2010
α	demethylation rate constant	8 hrs^{-1}	Kristensen et al., 2011
f	fraction of de-methylation rate when nucleosome is in compacted state	0	
λ	compaction rate constant	315 hrs^{-1}	Larson et al., 2017; Ladoux et al., 2000
δ	decompaction rate constant	5300 hrs^{-1}	Larson et al., 2017; Ladoux et al., 2000
F	fraction of compaction rate when nucleosome is in demethylated stated	0.85	
N	number of simulated nucleosomes	50	
C_N	nucleation threshold for compacted nucleosomal complex	5	
K_{ON}	transcription factor association constant	1 hrs-1	
K_{OFF}	transcription factor dissociation constant	5 hrs-1	
TF	transcription factor concentration	0-100	
N_B	total number of nucleosomes carrying transcription factor binding sites	1-5	
cell division length	cell division length	20 hrs	
Compaction Methylation Model With Permanently demethylated Nucleosomes			
Parameters	**Description**	**Value**	**Comments/references**
β	methylation rate constant	6.4 hrs^{-1}	Sneeringer et al., 2010
α	demethylation rate constant	8 hrs^{-1}	Kristensen et al., 2011
f	fraction of de-methylation rate when nucleosome is in compacted state	0	
λ	compaction rate constant	295 hrs^{-1}	Larson et al., 2017; Ladoux et al., 2000
δ	decompaction rate constant	5300 hrs^{-1}	Larson et al., 2017; Ladoux et al., 2000
F	fraction of compaction rate when nucleosome is in demethylated stated	0.85	
N	Total number of simulated nucleosomes	50	
C_N	nucleation threshold for compacted nucleosomal complex	5	
K_{ON}	transcription factor association constant	1 hrs-1	
K_{OFF}	transcription factor dissociation constant	5 hrs-1	
TF	transcription factor concentration	0-35	
N_R	total number of nucleosomes carrying transcription factor binding sites	5-25	
α_{TF}	recruited demethylase rate constant	1600 hrs-1	
cell division length	cell division length	20 hrs	

e. Supplementary Table S1.6. Model Parameters: Methylation model with different cooperative reach parameter L (Supplementary Figure 1.3A-B)

Pure Methylation Model			
Parameters	**Description**	**units**	**values**
β_{ON}	cooperative methylation rate constant	hrs^{-1}	1
α_{OFF}	demethylation rate constant	hrs^{-1}	[2, 2.3, 3] for L = 3
			[2.8, 3, 4] for L = 4
			[9, 10, 11] for L = 5
			[10, 10.5, 11] for L = 10
L	reach of anchored methylating enzyme to neighboring nucleosome	nucleosomes	[3, 4, 5 ,10]
N	number of simulated nucleosomes	nucleosomes	100
cell division length	cell division length	hrs	20

f. Supplementary Table S1.7. Model Parameters: Potential energy landscapes analysis for pure methylation model and methylation compaction model (Supplementary Figure 1.3C - E)

Pure Dilution Model		
Parameters	**Description**	**Value**
β	cooperative methylation rate constant	1-4 hrs^{-1}
α	demethylation rate constant	20 hrs^{-1}
N	number of simulated nucleosomes	100
W	arbitrary constant	100
Compaction Methylation Model		
Parameters	**Description**	**Value**
β	methylation rate constant	1-4 hrs^{-1}
α	demethylation rate constant	20 hrs^{-1}
λ	compaction rate constant	31 hrs^{-1}
δ	decompaction rate constant	530 hrs^{-1}
F	fraction of compaction rate when nucleosome is in demethylated stated	0.85
N	Number of simulated nucleosomes	50

g. Supplementary Table S1.8. Model Parameters: Compaction with Cooperative Methylation Model (Supplementary Figure 1.4)

Compaction Cooperative Methylation Model			
Parameters	**Description**	**Value**	**Comments/references**
β	cooperative methylation rate constant	0.02 hrs^{-1}	
μ	spontaneous rate constant	[0.05 , 0.1 , 0.2 , 0.4] hrs^{-1}	
α	demethylation rate constant	8 hrs^{-1}	Kristensen et al., 2011
f	fraction of de-methylation rate when nucleosome is in compacted state	0	
λ	compaction rate constant	310 hrs^{-1}	Larson et al., 2017; Ladoux et al., 2000
δ	decompaction rate constant	5300 hrs^{-1}	Larson et al., 2017; Ladoux et al., 2000
F	fraction of compaction rate when nucleosome is in demethylated stated	0.85	
N	number of simulated nucleosomes	50	
C_N	nucleation threshold for compacted nucleosomal complex	5	
cell division length	cell division length	20 hrs	
Pure Methylation Model			
Parameters	**Description**	**Value**	**Comments**
β_{ON}	cooperative methylation rate constant	1 hrs^{-1}	Sneeringer et al., 2010
α_{OFF}	demethylation rate constant	[16 , 17 , 17.5] hrs^{-1}	for parameter scan, Kristensen et al., 2011
L	reach of anchored methylating enzyme to neighboring nucleosome	15	Hass et al., 2015
N	number of simulated nucleosomes	100	
cell division length	cell division length	20 hrs	

h. Supplementary Table S1.9. Model Parameters: Compaction Methylation Model with Compaction State Disruption by Cell Division (Supplementary Figure 1.5A-E)

Compaction Methylation Model with Compaction State Affected by Cell Division			
Parameters	**Description**	**Value**	**Comments/references**
β	methylation rate constant	1 hrs^{-1}	[0.2 , 0.4 , 0.8 , 1.6 , 3.2 , 6.4] for parameter scan in Supplemental Figure 2E
			Sneeringer et al., 2010
α	demethylation rate constant	8 hrs^{-1}	Kristensen et al., 2011
f	fraction of de-methylation rate when nucleosome is in compacted state	0	
λ	compaction rate constant	310 hrs^{-1}	Larson et al., 2017; Ladoux et al., 2000
δ	decompaction rate constant	5300 hrs^{-1}	Larson et al., 2017; Ladoux et al., 2000
F	fraction of compaction rate when nucleosome is in demethylated stated	0.85	
N	Number of simulated nucleosomes	50	
C_N	nucleation threshold for compacted nucleosomal complex	5	
cell division length	cell division length	20 hrs	[10 , 20 , 25 , 30 , 35] for parameter scan in Supplemental Figure 2D

I. Supplementary Table S1.10. Model Parameters: Pure Dilution Model with Minimal Methylation and Demethylation Rates (Supplementary Figure 1.5F)

Dilution Model With Minimal Enzymatic Activities			
Parameters	**Description**	**Value**	**Comments/references**
β_{ON}	cooperative methylation rate constant	0.001 hrs-1	
α_{OFF}	demethylation rate constant	0.001 hrs-1	
L	reach of anchored methylating enzyme to neighboring nucleosome	15	Hass et al., 2015
N	number of simulated nucleosomes	100	
cell division length	cell division length	10 and 20 hrs	

1.10 DISCUSSION

Here, we studied a timed epigenetic switch controlling the activation of Bcl11b, an essential regulator of T cell lineage commitment. In response to regulatory inputs, this epigenetic switch activates with a stochastic time delay that spans multiple cell divisions, that is tunably controlled independently of cell cycle length by chromatin regulators and transcription factors. Through integration of experiments and modeling, we identify a mechanism for timed epigenetic switching, where removal of repressive histone H3K27me3 modifications weakens inter-nucleosomal contacts at the gene locus, leading to chromatin decompaction and gene activation. We find that this methylation-compaction mechanism generates long time delays that are both tunable and division-independent, as observed experimentally. Together, our results establish a robust, tunable timing mechanism that may be broadly deployed in multicellular organism development and function.

Our experiments implicate histone methyltransferases and demethylases as opposing, tunable controllers of gene activation and developmental timing. Histone demethylases are frequently found to be critical for activating cell type specific repressed genes that are selectively targeted for activation during development (Li et al., 2014; Park et al., 2014), thus the prevailing mode for regulating development genes in their normal context may indeed involve such active removal mechanisms. Furthermore, these chromatin regulators exhibit weak DNA-sequence specificity, which allows them to work concurrently at multiple regulatory gene loci during development and differentiation. In an accompanying paper, we explore the potential functions of chromatin regulators as master regulators of gene regulatory network dynamics and developmental speed (Nguyen et al., in submission).

In contrast to previous models of epigenetic switching (Dodd et al., 2007; Zhang et al., 2014), which consider only histone modification dynamics, the methylation-compaction model couples histone modifications to nucleosomal self-association and chromatin compaction. Such coupling is necessary to ensure that activation time delays can be robustly and tunably controlled by chromatin-modifying enzymes or transcription factors (Figure 4B, 6A-D). Emerging evidence indicates that repressive histone modifications – H3K27me3 and H3K9me2/3 – each function by promoting nucleosomal self-association though histone-tail interactions (Gibson et al., 2019) or by recruiting protein complexes that self-associate to form phase-separated condensates (Plys et al., 2018; Sanulli et al., 2019; Tatavosian et al., 2018; Wang et al., 2019) -physical processes that can generate the cooperativity needed for all-or-none transitions observed in the methylation-compaction model (Figure 4C). Importantly, for the activation times to be tunable, nucleosomes must retain some self-interaction affinity without H3K27me3, such that methylation promotes but is not strictly necessary for nucleosomal association. Consistently, we found that H3K9me2/3 modifications (Larson et al., 2017; Strom et al., 2017), which recruit HP-1 to promote chromatin phase separation, also tunes Bcl11b activation timing, and may thus mediate nucleosomal interactions in the absence of H3K27me3 (Larson et al., 2017; Strom et al., 2017). This concept – that modification states of proteins modulate their interaction affinities – is well established in the study of cytoskeletal polymer dynamics (Howard, 2001; Mitchison, 1992; Phillips et al., 2012), but could provide a fresh perspective on understanding the relationship between chromatin modification states and structure. In principle, other weak, multivalent interactions that are tunable by histone modifications – such as the formation of transcriptional condensates at enhancers (Shrinivas et al., 2018) – could also account for the unique features of the timed epigenetic switch at Bcl11b. However, we favor a mechanism where H3K27 and H3K9 methylation adjusts nucleosomal interactions to control chromatin compaction, because there is direct biochemical evidence linking these processes, and because these repressive modifications can be propagated across cell division, explaining how loci could reside in an repressed state for multiple cell generations prior to activation. Further testing of the methylation-compaction model will require measurements of interaction affinities of nucleosomes in different chemical modification states, as well as identification of the molecular players that mediate these interactions. It will also require direct interrogation of the relationship between these modification states and higher order chromatin structure at individual gene loci, work that will be aided by new methods to simultaneously visualize histone modification states and chromatin folding at single gene loci in single cells (Kundu et al., 2017; Xu et al., 2018).

The methylation-compaction switching mechanism could account for diverse cell-autonomous timers that function independently from cell division (Burton et al., 1999; Gao et al., 1997; Heinzel et al., 2017; Li et al., 2019; Okamoto et al., 2016; Osmond, 1991; Otani et al., 2016). While there are classic examples of embryonic processes with cell cycle-linked time delays, through inhibitor dilution-based mechanisms (Amodeo et al., 2015; Newport and Kirschner, 1982), many developmental processes instead appear to measure and control elapsed time independently from cell division. Division-independent timing control could enable multiple functions including operation in non-dividing cells and constancy amid changes to cell proliferation, which could allow for tunable population size control, an idea we explore in an accompanying manuscript (Nguyen et al., in submission). From model simulations, we find that such division-independent timing control requires H3K27me3 turnover and nucleosome compaction dynamics to be rapid compared to the cell cycle length (Mathematical Appendix, Figure S9). We currently lack methods to measure H3K27me3 turnover kinetics at specific genomic loci in vivo; however, the active roles of PRC2 and Kdm6a/b demethylases in modulating H3K27me3 levels at the Bcl11b locus (Figure 2), along with the fast catalysis rates of these enzymes measured in vitro (Kristensen et al., 2011; Sneeringer et al., 2010), suggest that H3K27me3 indeed turns over at a much faster timescale than that of cell division. Furthermore, nucleosomes within polycomb domains are replaced with kinetics much faster than that of cell division (1.5 hrs vs. 20 hrs) (Deal et al., 2010), consistent with the idea that H3K27 methylation turns over many times during a cell cycle lifetime.

Cell type specification during multicellular development is controlled by gene regulatory networks whose dynamics unfold over time-scales spanning many cell generations. The division-independent timed epigenetic switch we describe here is uniquely tunable at multiple levels of gene regulation including histone modifications, transcription factors and non-coding cis-regulatory elements. Therefore, it could serve as a modular building block within gene regulatory networks that could be flexibly adjusted throughout evolution. In a separate study, we explore this hypothesis by constructing model gene regulatory networks composed of timed epigenetic switches (Nguyen et al., in submission). We find evidence that adjusting epigenetic switches can modulate various features of multicellular tissues such as allometric scaling and the numbers and relative proportions of each cell type. Thus, we propose that epigenetic timing control could provide an underlying molecular mechanism by which the developmental timing is flexibly altered to give rise to changes in organism size and morphology.

1.11 REFERENCE

Agger, K., Cloos, P.A.C., Christensen, J., Pasini, D., Rose, S., Rappsilber, J., Issaeva, I., Canaani, E., Salcini, A.E., and Helin, K. (2007). UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. Nature 449, 731–734.

Alabert, C., Barth, T.K., Reverón-Gómez, N., Sidoli, S., Schmidt, A., Jensen, O.N., Imhof, A., and Groth, A. (2015). Two distinct modes for propagation of histone PTMs across the cell cycle. Genes Dev. 29, 585–590.

Amodeo, A.A., Jukam, D., Straight, A.F., and Skotheim, J.M. (2015). Histone titration against the genome sets the DNA-to-cytoplasm threshold for the Xenopus midblastula transition. Proc. Natl. Acad. Sci. 112, E1086–E1095.

Angel, A., Song, J., Dean, C., and Howard, M. (2011). A Polycomb-based switch underlying quantitative epigenetic memory. Nature 476, 105–108.

Arbona, J.-M., Herbert, S., Fabre, E., and Zimmer, C. (2017). Inferring the physical properties of yeast chromatin through Bayesian analysis of whole nucleus simulations. Genome Biol. 18, 81.

Arganda-Carreras, I., Sorzano, C.O.S., Marabini, R., Carazo, J.M., Ortiz-de-Solorzano, C., and Kybic, J. (2006). Consistent and Elastic Registration of Histological Sections Using Vector-Spline Regularization. In Computer Vision Approaches to Medical Image Analysis, R.R. Beichel, and M. Sonka, eds. (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 85–95.

Beliveau, B.J., Boettiger, A.N., Nir, G., Bintu, B., Yin, P., Zhuang, X., and Wu, C. -ting (2017). In Situ Super-Resolution Imaging of Genomic DNA with OligoSTORM and OligoDNA-PAINT. In Super-Resolution Microscopy, H. Erfle, ed. (New York, NY: Springer New York), pp. 231–252.

Beliveau, B.J., Kishi, J.Y., Nir, G., Sasaki, H.M., Saka, S.K., Nguyen, S.C., Wu, C., and Yin, P. (2018). OligoMiner provides a rapid, flexible environment for the design of genome-scale oligonucleotide in situ hybridization probes. Proc. Natl. Acad. Sci. 115, E2183–E2192.

Berry, S., Dean, C., and Howard, M. (2017). Slow Chromatin Dynamics Allow Polycomb Target Genes to Filter Fluctuations in Transcription Factor Activity. Cell Syst. 4, 445-457.e8.

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

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature 441, 349–353.

Burton, P.B.J., Raff, M.C., Kerr, P., Yacoub, M.H., and Barton, P.J.R. (1999). An Intrinsic Timer That Controls Cell-Cycle Withdrawal in Cultured Cardiac Myocytes. Dev. Biol. 216, 659–670.

Choi, K., Medley, J.K., König, M., Stocking, K., Smith, L., Gu, S., and Sauro, H.M. (2018). Tellurium: An extensible python-based modeling environment for systems and synthetic biology. Biosystems 171, 74–79.

Coleman, R.T., and Struhl, G. (2017). Causal role for inheritance of H3K27me3 in maintaining the OFF state of a Drosophila HOX gene. Science 356, eaai8236.

Davis, C.A., Hitz, B.C., Sloan, C.A., Chan, E.T., Davidson, J.M., Gabdank, I., Hilton, J.A., Jain, K., Baymuradov, U.K., Narayanan, A.K., et al. (2018). The Encyclopedia of DNA elements (ENCODE): data portal update. Nucleic Acids Res. 46, D794–D801.
Deal, R.B., Henikoff, J.G., and Henikoff, S. (2010). Genome-Wide Kinetics of Nucleosome Turnover Determined by Metabolic Labeling of Histones. Science 328, 1161–1164.

Dodd, I.B., Micheelsen, M.A., Sneppen, K., and Thon, G. (2007). Theoretical analysis of epigenetic cell memory by nucleosome modification. Cell 129, 813–822.

Dufourt, J., Trullo, A., Hunter, J., Fernandez, C., Lazaro, J., Dejean, M., Morales, L., Nait-Amer, S., Schulz, K.N., Harrison, M.M., et al. (2018). Temporal control of gene expression by the pioneer factor Zelda through transient interactions in hubs. Nat. Commun. 9.

Erickson, H.P., and Pantaloni, D. (1981). The role of subunit entropy in cooperative assembly. Nucleation of microtubules and other two-dimensional polymers. Biophys. J. 34, 293–309.

Eskeland, R., Leeb, M., Grimes, G.R., Kress, C., Boyle, S., Sproul, D., Gilbert, N., Fan, Y., Skoultchi, A.I., Wutz, A., et al. (2010). Ring1B Compacts Chromatin Structure and Represses Gene Expression Independent of Histone Ubiquitination. Mol. Cell 38, 452–464.

Estarás, C., Fueyo, R., Akizu, N., Beltrán, S., and Martínez-Balbás, M.A. (2013). RNA polymerase II progression through H3K27me3-enriched gene bodies requires JMJD3 histone demethylase. Mol. Biol. Cell 24, 351–360.

Fellmann, C., Hoffmann, T., Sridhar, V., Hopfgartner, B., Muhar, M., Roth, M., Lai, D.Y., Barbosa, I.A.M., Kwon, J.S., Guan, Y., et al. (2013). An Optimized microRNA Backbone for Effective Single-Copy RNAi. Cell Rep. 5, 1704–1713.

Fernandez Garcia, M., Moore, C.D., Schulz, K.N., Alberto, O., Donague, G., Harrison, M.M., Zhu, H., and Zaret, K.S. (2019). Structural Features of Transcription Factors Associating with Nucleosome Binding. Mol. Cell 75, 921-932.e6.

Francis, N.J., Kingston, R.E., and Woodcock, C.L. (2004). Chromatin compaction by a polycomb group protein complex. Science 306, 1574–1577.

Gao, F.-B., Durand, B., and Raff, M. (1997). Oligodendrocyte precursor cells count time but not cell divisions before differentiation. Curr. Biol. 7, 152–155.

García-Ojeda, M.E., Klein Wolterink, R.G.J., Lemaître, F., Richard-Le Goff, O., Hasan, M., Hendriks, R.W., Cumano, A., and Di Santo, J.P. (2013). GATA-3 promotes T-cell specification by repressing B-cell potential in pro-T cells in mice. Blood 121, 1749–1759.

Gaydos, L.J., Wang, W., and Strome, S. (2014). H3K27me and PRC2 transmit a memory of repression across generations and during development. Science 345, 1515–1518.

Germar, K., Dose, M., Konstantinou, T., Zhang, J., Wang, H., Lobry, C., Arnett, K.L., Blacklow, S.C., Aifantis, I., Aster, J.C., et al. (2011). T-cell factor 1 is a gatekeeper for T-cell specification in response to Notch signaling. Proc. Natl. Acad. Sci. U. S. A. 108, 20060–20065.

Gibson, B.A., Doolittle, L.K., Schneider, M.W.G., Jensen, L.E., Gamarra, N., Henry, L., Gerlich, D.W., Redding, S., and Rosen, M.K. (2019). Organization of Chromatin by Intrinsic and Regulated Phase Separation. Cell 179, 470-484.e21.

Giorgetti, L., Piolot, T., and Heard, E. (2015). High-Resolution 3D DNA FISH Using Plasmid Probes and Computational Correction of Optical Aberrations to Study Chromatin Structure at the Sub-megabase Scale. In Nuclear Bodies and Noncoding RNAs, S. Nakagawa, and T. Hirose, eds. (New York, NY: Springer New York), pp. 37–53.

Gray, F., Cho, H.J., Shukla, S., He, S., Harris, A., Boytsov, B., Jaremko, Ł., Jaremko, M., Demeler, B., Lawlor, E.R., et al. (2016). BMI1 regulates PRC1 architecture and activity through homo- and hetero-oligomerization. Nat. Commun. 7, 13343.

Hass, M.R., Liow, H.-H., Chen, X., Sharma, A., Inoue, Y.U., Inoue, T., Reeb, A., Martens, A., Fulbright, M., Raju, S., et al. (2015). SpDamID: Marking DNA Bound by Protein Complexes Identifies Notch-Dimer Responsive Enhancers. Mol. Cell 59, 685–697.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–589.

Heinzel, S., Binh Giang, T., Kan, A., Marchingo, J.M., Lye, B.K., Corcoran, L.M., and Hodgkin, P.D. (2017). A Myc-dependent division timer complements a cell-death timer to regulate T cell and B cell responses. Nat. Immunol. 18, 96–103.

Holmes, R., and Zuniga-Pflucker, J.C. (2009). The OP9-DL1 System: Generation of T-Lymphocytes from Embryonic or Hematopoietic Stem Cells In Vitro. Cold Spring Harb. Protoc. 2009, pdb.prot5156-pdb.prot5156.

Hosogane, M., Funayama, R., Shirota, M., and Nakayama, K. (2016). Lack of Transcription Triggers H3K27me3 Accumulation in the Gene Body. Cell Rep. 16, 696–706.

Hosokawa, H., Romero-Wolf, M., Yui, M.A., Ungerbäck, J., Quiloan, M.L.G., Matsumoto, M., Nakayama, K.I., Tanaka, T., and Rothenberg, E.V. (2018). Bcl11b sets pro-T cell fate by site-specific cofactor recruitment and by repressing Id2 and Zbtb16. Nat. Immunol. 19, 1427–1440.

Howard, J. (2001). Mechanics of Motor Proteins and the Cytoskeleton (Sunderland, Mass: Sinauer Associates is an imprint of Oxford University Press).

Hu, G., Cui, K., Fang, D., Hirose, S., Wang, X., Wangsa, D., Jin, W., Ried, T., Liu, P., Zhu, J., et al. (2018). Transformation of Accessible Chromatin and 3D Nucleome Underlies Lineage Commitment of Early T Cells. Immunity 48, 227-242.e8.

Howard, J. (2001). Mechanics of Motor Proteins and the Cytoskeleton (Sunderland, Mass: Sinauer Associates is an imprint of Oxford University Press).

Ikawa, T., Hirose, S., Masuda, K., Kakugawa, K., Satoh, R., Shibano-Satoh, A., Kominami, R., Katsura, Y., and Kawamoto, H. (2010). An Essential Developmental Checkpoint for Production of the T Cell Lineage. Science 329, 93–96.

Isono, K., Endo, T.A., Ku, M., Yamada, D., Suzuki, R., Sharif, J., Ishikura, T., Toyoda, T., Bernstein, B.E., and Koseki, H. (2013). SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing. Dev. Cell 26, 565–577.

Jackson, M.B., and Berkowitz, S.A. (1980). Nucleation and the kinetics of microtubule assembly. Proc. Natl. Acad. Sci. 77, 7302–7305.

Jadhav, U., Manieri, E., Nalapareddy, K., Madha, S., Chakrabarti, S., Wucherpfennig, K., Barefoot, M., and Shivdasani, R.A. (2020). Replicational Dilution of H3K27me3 in Mammalian Cells and the Role of Poised Promoters. Mol. Cell S1097276520300381.

Jiang, D., and Berger, F. (2017). DNA replication–coupled histone modification maintains Polycomb gene silencing in plants. Science 357, 1146–1149.

Kadoch, C., Williams, R.T., Calarco, J.P., Miller, E.L., Weber, C.M., Braun, S.M.G., Pulice, J.L., Chory, E.J., and Crabtree, G.R. (2017). Dynamics of BAF–Polycomb complex opposition on heterochromatin in normal and oncogenic states. Nat. Genet. 49, 213–222.

Kahn, T.G., Dorafshan, E., Schultheis, D., Zare, A., Stenberg, P., Reim, I., Pirrotta, V., and Schwartz, Y.B. (2016). Interdependence of PRC1 and PRC2 for recruitment to Polycomb Response Elements. Nucleic Acids Res. 44, 10132–10149.

Kaikkonen, M.U., Spann, N.J., Heinz, S., Romanoski, C.E., Allison, K.A., Stender, J.D., Chun, H.B., Tough, D.F., Prinjha, R.K., Benner, C., et al. (2013). Remodeling of the Enhancer Landscape during Macrophage Activation Is Coupled to Enhancer Transcription. Mol. Cell 51, 310–325.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. The Human Genome Browser at UCSC. 12.
Kraushaar, D.C., Jin, W., Maunakea, A., Abraham, B., Ha, M., and Zhao, K. (2013). Genome-wide incorporation dynamics reveal distinct categories of turnover for the histone variant H3.3. Genome Biol. 14, R121.

Kramers, H.A. (1940). Brownian motion in a field of force and the diffusion model of chemical reactions. Physica 7, 284–304.

Kristensen, J.B.L., Nielsen, A.L., Jørgensen, L., Kristensen, L.H., Helgstrand, C., Juknaite, L., Kristensen, J.L., Kastrup, J.S., Clausen, R.P., Olsen, L., et al. (2011). Enzyme kinetic studies of histone demethylases KDM4C and KDM6A: Towards understanding selectivity of inhibitors targeting oncogenic histone demethylases. FEBS Lett. 585, 1951–1956.

Kueh, H.Y., Yui, M.A., Ng, K.K.H., Pease, S.S., Zhang, J.A., Damle, S.S., Freedman, G., Siu, S., Bernstein, I.D., Elowitz, M.B., et al. (2016). Asynchronous combinatorial activation of four regulatory factors activates Bcl11b for T cell commitment. Nat. Immunol. 17, 956–965.

Kundu, S., Ji, F., Sunwoo, H., Jain, G., Lee, J.T., Sadreyev, R.I., Dekker, J., and Kingston, R.E. (2017). Polycomb Repressive Complex 1 Generates Discrete Compacted Domains that Change during Differentiation. Mol. Cell 65, 432-446.e5.

Ladoux, B., Quivy, J.-P., Doyle, P., Roure, O. du, Almouzni, G., and Viovy, J.-L. (2000). Fast kinetics of chromatin assembly revealed by single-molecule videomicroscopy and scanning force microscopy. Proc. Natl. Acad. Sci. 97, 14251–14256.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359.

Larson, A.G., Elnatan, D., Keenen, M.M., Trnka, M.J., Johnston, J.B., Burlingame, A.L., Agard, D.A., Redding, S., and Narlikar, G.J. (2017). Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. Nature 547, 236–240.

Lee, K., Hsiung, C.C.-S., Huang, P., Raj, A., and Blobel, G.A. (2015). Dynamic enhancer–gene body contacts during transcription elongation. Genes Dev. 29, 1992–1997.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K., et al. (2006). Control of Developmental Regulators by Polycomb in Human Embryonic Stem Cells. Cell 125, 301–313.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079.

Li, H., Natarajan, A., Ezike, J., Barrasa, M.I., Le, Y., Feder, Z.A., Yang, H., Ma, C., Markoulaki, S., and Lodish, H.F. (2019). Rate of Progression through a Continuum of Transit-Amplifying Progenitor Cell States Regulates Blood Cell Production. Dev. Cell.
Li, L., Leid, M., and Rothenberg, E.V. (2010). An early T cell lineage commitment checkpoint dependent on the transcription factor Bcl11b. Science 329, 89–93.

Li, Q., Wang, H.Y., Chepelev, I., Zhu, Q., Wei, G., Zhao, K., and Wang, R.-F. (2014). Stage-Dependent and Locus-Specific Role of Histone Demethylase Jumonji D3 (JMJD3) in the Embryonic Stages of Lung Development. PLoS Genet. 10, e1004524.

MacPherson, Q., Beltran, B., and Spakowitz, A.J. (2018). Bottom–up modeling of chromatin segregation due to epigenetic modifications. Proc. Natl. Acad. Sci. 115, 12739–12744.

Manesso, E., Chickarmane, V., Kueh, H.Y., Rothenberg, E.V., and Peterson, C. (2012). Computational modelling of T-cell formation kinetics: output regulated by initial proliferation-linked deferral of developmental competence. J. R. Soc. Interface 10, 20120774–20120774.

Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury, W.J., Voigt, P., Martin, S.R., Taylor, W.R., De Marco, V., et al. (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. Nature 461, 762–767.

Mayran, A., Khetchoumian, K., Hariri, F., Pastinen, T., Gauthier, Y., Balsalobre, A., and Drouin, J. (2018). Pioneer factor Pax7 deploys a stable enhancer repertoire for specification of cell fate. Nat. Genet. 50, 259–269.

Meers, M.P., Janssens, D.H., and Henikoff, S. (2019). Pioneer Factor-Nucleosome Binding Events during Differentiation Are Motif Encoded. Mol. Cell 75, 562-575.e5.

Mitchison, T.J. (1992). Compare and contrast actin filaments and microtubules. Mol. Biol. Cell 3, 1309–1315.

Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N., and Mirny, L.A. (2018). Chromatin organization by an interplay of loop extrusion and compartmental segregation. Proc. Natl. Acad. Sci. 115, E6697–E6706.

Newport, J., and Kirschner, M. (1982). A major developmental transition in early xenopus embryos: I. characterization and timing of cellular changes at the midblastula stage. Cell 30, 675–686.

Ng, K.K., Yui, M.A., Mehta, A., Siu, S., Irwin, B., Pease, S., Hirose, S., Elowitz, M.B., Rothenberg, E.V., and Kueh, H.Y. (2018). A stochastic epigenetic switch controls the dynamics of T-cell lineage commitment. ELife 7.

Okamoto, M., Miyata, T., Konno, D., Ueda, H.R., Kasukawa, T., Hashimoto, M., Matsuzaki, F., and Kawaguchi, A. (2016). Cell-cycle-independent transitions in temporal identity of mammalian neural progenitor cells. Nat. Commun. 7, 11349.

Osmond, D.G. (1991). Proliferation kinetics and the lifespan of B cells in central and peripheral lymphoid organs. Curr. Opin. Immunol. 3, 179–185.

Otani, T., Marchetto, M.C., Gage, F.H., Simons, B.D., and Livesey, F.J. (2016). 2D and 3D Stem Cell Models of Primate Cortical Development Identify Species-Specific Differences in Progenitor Behavior Contributing to Brain Size. Cell Stem Cell 18, 467–480.

Park, D.H., Hong, S.J., Salinas, R.D., Liu, S.J., Sun, S.W., Sgualdino, J., Testa, G., Matzuk, M.M., Iwamori, N., and Lim, D.A. (2014). Activation of Neuronal Gene Expression by the JMJD3 Demethylase Is Required for Postnatal and Adult Brain Neurogenesis. Cell Rep. 8, 1290–1299.

Phillips, R. (2015). Napoleon Is in Equilibrium. Annu. Rev. Condens. Matter Phys. 6, 85–111.

Phillips, R., Kondev, J., Theriot, J., and Garcia, H. (2012). Physical Biology of the Cell (London : New York, NY: Garland Science).

Plys, A.J., Davis, C.P., Kim, J., Rizki, G., Keenen, M.M., Marr, S.K., and Kingston, R.E. (2018). Phase separation and nucleosome compaction are governed by the same domain of Polycomb Repressive Complex 1. BioRxiv 467316.

Poleshko, A., Mansfield, K.M., Burlingame, C.C., Andrake, M.D., Shah, N.R., and Katz, R.A. (2013). The Human Protein PRR14 Tethers Heterochromatin to the Nuclear Lamina during Interphase and Mitotic Exit. Cell Rep. 5, 292–301.

Porritt, H.E., Gordon, K., and Petrie, H.T. (2003). Kinetics of Steady-state Differentiation and Mapping of Intrathymic-signaling Environments by Stem Cell Transplantation in Nonirradiated Mice. J. Exp. Med. 198, 957–962.
Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., et al. (2007). Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs. Cell 129, 1311–1323.

Rossi, A.M., Fernandes, V.M., and Desplan, C. (2017). Timing temporal transitions during brain development. Curr. Opin. Neurobiol. 42, 84–92.

Sanulli, S., Trnka, M.J., Dharmarajan, V., Tibble, R.W., Pascal, B.D., Burlingame, A.L., Griffin, P.R., Gross, J.D., and Narlikar, G.J. (2019). HP1 reshapes nucleosome core to promote phase separation of heterochromatin. Nature 575, 390–394.

Schmitt, T.M., and Zúñiga-Pflücker, J.C. (2002). Induction of T Cell Development from Hematopoietic Progenitor Cells by Delta-like-1 In Vitro. Immunity 17, 749–756.

Seenundun, S., Rampalli, S., Liu, Q.-C., Aziz, A., Palii, C., Hong, S., Blais, A., Brand, M., Ge, K., and Dilworth, F.J. (2010). UTX mediates demethylation of H3K27me3 at muscle-specific genes during myogenesis. EMBO J. 29, 1401–1411.

Skene, P.J., Henikoff, J.G., and Henikoff, S. (2018). Targeted in situ genome-wide profiling with high efficiency for low cell numbers. Nat. Protoc. 13, 1006–1019.

Sneeringer, C.J., Scott, M.P., Kuntz, K.W., Knutson, S.K., Pollock, R.M., Richon, V.M., and Copeland, R.A. (2010). Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas. Proc. Natl. Acad. Sci. U. S. A. 107, 20980–20985.

van Steensel, B., and Belmont, A.S. (2017). Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. Cell 169, 780–791.

Strom, A.R., Emelyanov, A.V., Mir, M., Fyodorov, D.V., Darzacq, X., and Karpen, G.H. (2017). Phase separation drives heterochromatin domain formation. Nature 547, 241–245.

Strome, S., Kelly, W.G., Ercan, S., and Lieb, J.D. (2014). Regulation of the X chromosomes in Caenorhabditis elegans. Cold Spring Harb. Perspect. Biol. 6.
Sun, B., Looi, L.-S., Guo, S., He, Z., Gan, E.-S., Huang, J., Xu, Y., Wee, W.-Y., and Ito, T. (2014). Timing Mechanism Dependent on Cell Division Is Invoked by Polycomb Eviction in Plant Stem Cells. Science 343, 1248559–1248559.
Talbert, P.B., and Henikoff, S. (2017). Histone variants on the move: substrates for chromatin dynamics. Nat. Rev. Mol. Cell Biol. 18, 115–126.

Tatavosian, R., Kent, S., Brown, K., Yao, T., Duc, H.N., Huynh, T.N., Zhen, C.Y., Ma, B., Wang, H., and Ren, X. (2018). Nuclear condensates of the Polycomb protein chromobox 2 (CBX2) assemble through phase separation. J. Biol. Chem. jbc.RA118.006620.

Temple, S., and Raff, M.C. (1986). Clonal analysis of oligodendrocyte development in culture: Evidence for a developmental clock that counts cell divisions. Cell 44, 773–779.

Tu, S., Yuan, G.-C., and Shao, Z. (2017). The PRC2-binding long non-coding RNAs in human and mouse genomes are associated with predictive sequence features. Sci. Rep. 7.
Ulianov, S.V., Doronin, S.A., Khrameeva, E.E., Kos, P.I., Luzhin, A.V., Starikov, S.S., Galitsyna, A.A., Nenasheva, V.V., Ilyin, A.A., Flyamer, I.M., et al. (2019). Nuclear lamina integrity is required for proper spatial organization of chromatin in Drosophila. Nat. Commun. 10, 1–11.

Varnum-Finney, B. Notch signaling requires immobilized Delta-1. 6.
Varnum-Finney, B., Brashem-Stein, C., and Bernstein, I.D. (2003). Combined effects of Notch signaling and cytokines induce a multiple log increase in precursors with lymphoid and myeloid reconstituting ability. Blood 101, 1784–1789.

Walters, M.C., Fiering, S., Eidemiller, J., Magis, W., Groudine, M., and Martin, D.I. (1995). Enhancers increase the probability but not the level of gene expression. Proc. Natl. Acad. Sci. 92, 7125–7129.

Wang, L., Gao, Y., Zheng, X., Liu, C., Dong, S., Li, R., Zhang, G., Wei, Y., Qu, H., Li, Y., et al. (2019). Histone Modifications Regulate Chromatin Compartmentalization by Contributing to a Phase Separation Mechanism. Mol. Cell 76, 646-659.e6.

Weintraub, H. (1988). Formation of stable transcription complexes as assayed by analysis of individual templates. Proc. Natl. Acad. Sci. 85, 5819–5823.

Williams, K., Christensen, J., Rappsilber, J., Nielsen, A.L., Johansen, J.V., and Helin, K. (2014). The Histone Lysine Demethylase JMJD3/KDM6B Is Recruited to p53 Bound Promoters and Enhancer Elements in a p53 Dependent Manner. PLoS ONE 9, e96545.
Xu, J., Ma, H., Jin, J., Uttam, S., Fu, R., Huang, Y., and Liu, Y. (2018). Super-Resolution Imaging of Higher-Order Chromatin Structures at Different Epigenomic States in Single Mammalian Cells. Cell Rep. 24, 873–882.

Zhang, H., Tian, X.-J., Mukhopadhyay, A., Kim, K.S., and Xing, J. (2014). Statistical Mechanics Model for the Dynamics of Collective Epigenetic Histone Modification. Phys. Rev. Lett. 112, 068101.

Zhang, J.A., Mortazavi, A., Williams, B.A., Wold, B.J., and Rothenberg, E.V. (2012). Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity. Cell 149, 467–482.

Zheng, M., Tian, S.Z., Capurso, D., Kim, M., Maurya, R., Lee, B., Piecuch, E., Gong, L., Zhu, J.J., Li, Z., et al. (2019). Multiplex chromatin interactions with single-molecule precision. Nature.
Zhou, C.Y., Johnson, S.L., Gamarra, N.I., and Narlikar, G.J. (2016). Mechanisms of ATP-Dependent Chromatin Remodeling Motors. Annu. Rev. Biophys. 45, 153–181.

Zhou, W., Yui, M.A., Williams, B.A., Yun, J., Wold, B.J., Cai, L., and Rothenberg, E.V. (2019). Single-Cell Analysis Reveals Regulatory Gene Expression Dynamics Leading to Lineage Commitment in Early T Cell Development. Cell Syst. 9, 321-337.e9.

CHAPTER II. SCALABLE CONTROL OF DEVELOPMENTAL TIMETABLES BY EPIGENETIC SWITCHING NETWORKS

2.1 INTRODUCTION

How are the sizes and proportions of organs and body plans specified so precisely during multicellular development, yet so readily adjustable by evolution? For over a century, it has been recognized that the timetables of developmental events in the embryo are a central determinant of size and form (Haeckel, 1866; De Beer, 1940; Huxley et al., 1942; Gould, 1977). Because progenitors grow and divide during development, the timing at which they specify different cell types will determine the final numbers and fractions of their differentiated progeny; consequently, changes in developmental timetables can engender variation in size or form. Changes to the timing of individual lineage specification events can give rise to innovations in shape or form (Alberch et al., 1979; Huxley et al., 1942; Gould, 1977), whereas changes to the speed at which developmental timetables unfold can generate proportionally scaled changes in organ and organism sizes (Bonner, 1965; Calder, 1984).

As development is not coupled to an external clock, the timetables for developmental lineage specification must be encoded by the embryo itself (Ebisuya and Briscoe, 2018). Most generally, these temporal schedules are the culmination of complex processes unfolding in space and time in the embryo; however, across a growing number of systems, it is now clear that these schedules appear to be determined, in large part, by timekeeping mechanisms operating autonomously in individual progenitor cells (Burton et al., 1999; Gao et al., 1997; Heinzel et al., 2017; Otani et al., 2016; Rosello-Diez et al., 2014; Saiz-Lopez et al., 2015). For example, during cerebral cortex development, progenitors generate different layers of cortical neurons in a defined order, giving rise to inner layer neurons before outer layer neurons. This same developmental timetable unfolds in vitro (Eiraku et al., 2008; Gaspard et al., 2008), even in the absence of any intact tissue organization, suggesting it is largely set in a cell-autonomous manner. Furthermore, though progenitors from different species follow the same lineage specification schedule in vitro, they differ substantially in the speed at which they take to traverse this schedule, ranging from one week in mice to four months in humans (van den Ameele et al., 2014; Barry et al., 2017a; Espuny-Camacho et al., 2013; Otani et al., 2016). These results show that timing mechanisms, operating autonomously in progenitor cells, can set defined timetables for developmental lineage specification, and that these timetables can be scalably expanded or contracted over time across different species.

Gene regulatory networks control the dynamics of development, and ultimately dictate the timing with which cell lineage specification occurs in response to signals (Alon, 2007; Briscoe and Small, 2015; Davidson, 2010; Georgescu et al., 2008). However, it is unclear how gene regulatory networks can set adjustable temporal schedules over developmental timescales spanning many cell generations. One major issue is that genes in regulatory networks are widely assumed to be regulated according to a quantitative framework derived largely from bacterial studies, whereby transcription factor tor levels alone dictate transcription rates (Alon, 2007; Ackers et al., 1982; Bintu et al., 2005; Bolouri and Davidson, 2002). In this classical framework, genes respond to transcriptional inputs over timescales shorter than one cell cycle (Levine and Elowitz, 2014), because their dynamics are constrained by mRNA and protein stability and, by extension, cell cycle speed. Models for temporal regulation of developmental gene networks in animals have historically only incorporated these short, single generation transcription factor-based time delays (Ben-Tabou de Leon and Davidson, 2009) however, development and cellular differentiation frequently unfolds over many cell generations in animals. Another major issue is how temporal schedules set by gene regulatory networks can be readily altered by evolutionary change. Mutations in cis-regulatory elements drive morphological change in evolution (Carroll, 2008); however, under a classical framework, it is not clear how these mutations can change timing of gene regulatory events, when their main effect is to alter transcription factor occupancy and transcription rates. It is also unknown how these temporal schedules can be scalably expanded or compressed in time as is the case in cerebral cortex development between different species such that the relative duration of developmental events remains the same, but the overall speed with which these events unfold can differ.

In eukaryotes, transcription factors not only regulate RNA polymerase recruitment and transcription, as is the case in bacteria, but also control the switching of gene loci between silent and active chromatin states, a process that can involve long rate-limiting steps spanning multiple cell generations. In stem and progenitor cells, lineage-specifying genes are often packaged in a closed chromatin environment enforced by repressive histone H3K27me3 modifications (Boyer et al.; Lee et al., 2006; Nicetto et al., 2019). While transcription factors can bind rapidly to these genes, within seconds, they remove repressive histone modifications over considerably longer timescales (Mayran et al., 2018; Bintu et al., 2016). In our recent studies on Bcl11b, a transcription factor that turns on to drive T cell lineage commitment (Ng et al., 2018; Rothenberg, 2019), we found that this gene locus switches stochastically from a silent to active state with a remarkably long time constant spanning multiple days and cell generations (Ng et al., 2018). The magnitude of this activation time constant was controlled not only by upstream transcription factors binding to a distal enhancer (Kueh et al., 2016; Ng et al., 2018), but also by H3K27me3 levels at the Bcl11b locus (Nguyen et al., 2019), which were actively set by methyltransferase and demethylase enzymes. As histone H3K27me3 and chromatin conformation changes occur frequently on lineage-specifying regulated genes during development, and because transcription factors and their bound cis-regulatory elements have been found to play broader roles in timing control, it is likely that such timed epigenetic switches may also play control gene activation or silencing kinetics at other regulatory gene loci. As such, they could profoundly shape the dynamic properties of developmental gene regulatory networks, and allow them to implement temporal control behaviors over long timescales.

In this study, we analyze the dynamic properties of gene regulatory networks in which timed epigenetic switches regulate the expression of constituent genes. Through mathematical modeling, we show how such epigenetic switching networks, in contrast to classical networks, can robustly maintain timetables for cell specification that can unfold over extended developmental timescales. Next, with a series of evolutionary case studies, we then consider how mutations in epigenetic switching networks can tunably alter these developmental timetables, to give rise to species variation in organism size and proportion. Based on our findings, we propose that epigenetic switching networks underlie the ability of embryos to encode autonomous timetables for development, and the remarkable capacity for evolution to adjust these timetables to facilitate variation in size and form.

2.2 MODELING FRAMEWORK FOR EPIGENETIC SWITCHING NETWORKS

We develop a quantitative framework for modeling gene regulatory networks in which timed epigenetic switches regulate the expression states of constituent genes. W explicitly account for both the expression levels of regulatory genes and the epigenetic states of gene loci. In this framework, we do not explicitly model chromatin states at individual loci, as in the methylation compaction model in our accompanying study (Pease et al., in submission); instead, we use a phenomenological model that captures the essential emergent properties of timed epigenetic switching in the methylation compaction model, including stochastic, first-order switching between heritable states; and tunable, division-independent rate constants for switching. In this framework, we also model the growth and division of individual cells, together with its interplay with gene regulatory network activity.

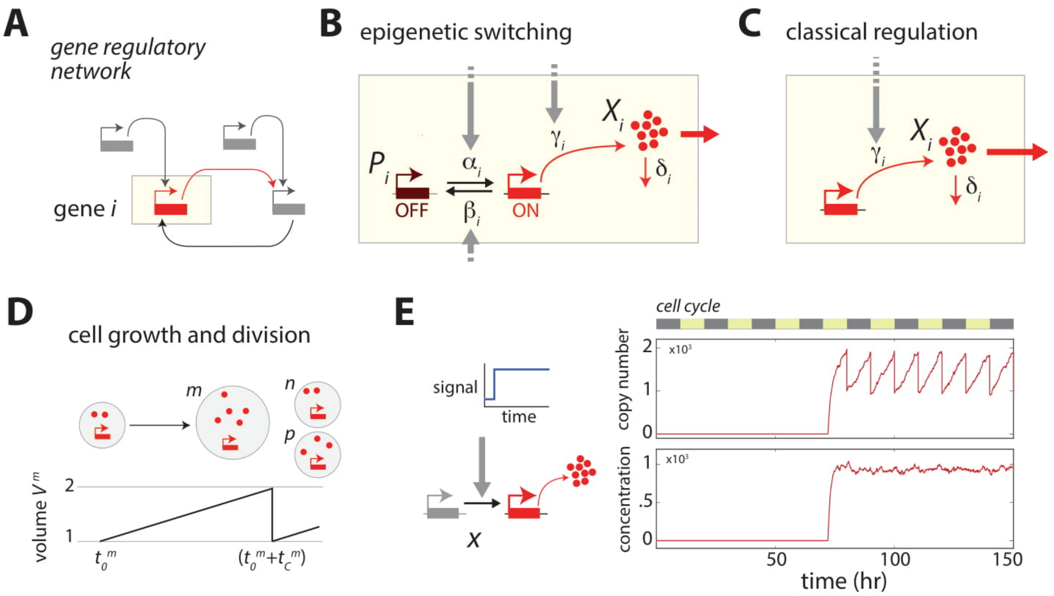


Figure 2.1: Mathematical framework for modeling epigenetic switching networks. (A) Gene regulatory network, with single gene (i) highlighted. (B) Gene promoter undergoing epigenetic switching. (C) Classically-regulated gene promoter. X_i gives the expression level of gene i; P_i gives the epigenetic state of gene promoter i. Also shown are rate constants for promoter activation αi, promoter silencing β_i, protein synthesis γ_i, and protein degradation δ_i. (D) Cell growth and division results in random partitioning of proteins, and inheritance of promoter states. (E) An epigenetic switch for gene X (left) is activated by an upstream signal S at t = 70 hr. Plots show time evolution of copy number (X, top right) and concentration ($x = X/V$, bottom right), obtained using stochastic simulations. Zebra stripe indicates the timing of cell division events.

Consider a network of N regulatory genes (Figure 2.1A). Let P_i be the number of active loci of gene i in the cell. Let X_i be the copy number of its gene product. Let $V\ (t)$ be the volume of the cell, such that $x_i\ =\ X_i = V$ is the concentration of protein i (Figure 2.1B). The time evolution of these variables is given by the following rules:

$$P_i \xrightarrow{\alpha_i(x_1,\ldots,x_N)\cdot V\cdot(P_T-P_i)} P_i+1 \tag{1}$$

$$P_i \xrightarrow{\beta_i(x_1,\ldots,x_N)\cdot V\cdot P_i} P_i-1 \tag{2}$$

$$X_i \xrightarrow{\gamma_i(x_1,\ldots,x_N)\cdot V\cdot P_i} X_i+1 \tag{3}$$

$$X_i \xrightarrow{\delta_i x_i\cdot V} X_i-1 \tag{4}$$

Here, gene i switches from an inactive (active) to an active (inactive) epigenetic state with first-order kinetics, with a stochastic time constant of α_i (β_i). Both locus activation and silencing rates depend on $x_i \ldots x_N$, reflecting the key finding in our studies of the epigenetic switch controlling Bcl11b activation (Pease et al., in submission), that time constants for epigenetic switching depends on levels of regulatory inputs, which may either involve transcription factors or chromatin-modifying enzymes (Pease et al., in submission). Once in an active state, gene i is synthesized at a rate γ_i, and its encoded gene product $X_i$is then actively degraded at rate δ_i. P_T is the total number of gene copies of each gene; in all simulations here, P_T is taken to equal unity, under the assumption that the copy number effects do not alter the emergent properties of the regulatory network considered here. Rates of promoter switching and protein degradation increase linearly with cell volume, reflecting an increase molecular content with cell size; in accordance with previous observations Padovan-Merhar et al. (2015), the rate of transcription also increases with cell volume, a mechanism that keeps steady-state protein concentration constant amid cell size variations (Figure 2.1E). When the promoters of all genes are always in an active state, i.e. when $P_i = 1$, where $i = 1 \ldots N$, this network reduces to one where individual genes are governed exclusively through classical gene regulation functions (Figure 2.1C).

To model this gene regulatory network within cell lineages, we account for individual cells within lineage trees. Let M be the total number of cells in a given lineage, and let the m^{th} cell in this lineage have promoter states P_i^m and protein levels X_i^m . Let this cell have volume V_m, birth time t_0^m, and life time t_c^m. Finally, let the cell cycle state of the cell be captured by a binary variable c_m, that has a value of one when the cell is actively cycling, and zero when the cell has exited the cell cycle.

If cell m in the lineage is actively cycling c_m= 1, its volume increases linearly during its lifetime to reflect cell growth (Figure 2.1D):

$$V^m(t) = V_0 \cdot \left[1 + (t - t_0^m)/t_c^m\right] \qquad (5)$$

Where:

$$t_0^m < t < (t_0^m + t_c). \qquad (6)$$

At the end of its lifetime, cell m divides to generate two daughter cells, n and p. Its regulatory proteins partition randomly into the two daughters at the time of cell division, such that their numbers follow a binomial distribution:

$$\begin{aligned} X_i^n &= \mathrm{bin}(X_i^m, 0.5) \qquad (7) \\ X_i^p &= X_i^m - X_i^n. \end{aligned}$$

On the other hand, gene promoters maintain their epigenetic states across cell division, such that:

$$\begin{aligned} P_i^n &= P_i^m \qquad (8) \\ P_i^p &= P_i^m. \qquad (9) \end{aligned}$$

Cells that are not cycling ($c_m = 0$) are taken to be irreversibly differentiated, and are not further analyzed by modeling, though they are counted to obtain differentiated cell numbers. Following previous findings (Spencer et al., 2013), the decision to exit the cell cycle is made at the cells' birth, and will depend on both regulatory network state and cell division mechanism, as we discuss below.

To obtain regulatory network dynamics in a cell lineage, together with the structure of the lineage tree itself, we simulate the stochastic chemical reactions describing the network (Eq. 1-4) in single cells using the Gillespie algorithm (Gillespie, 1977), while setting lifetimes, growth rates and initial conditions of simulated cells according to the rules above (Eqs. 5-9). For each lineage, we begin by simulating a single progenitor cell. At the end of its lifetime, this cell divides; daughter cells inherit their mother's regulatory network state (Eqs. 8-9), and also set their cycling states based of the states of their regulatory networks. Each daughter cell is then simulated, and this process continues iteratively until all cells withdraw the cell cycle, or until a simulation time limit is reached. This algorithm generates a list of cells, a lineage tree showing the relationships between these cells, and Monte Carlo simulation results for each cell.

2.3 EPIGENETIC SWITCHES GENERATE TUNABLE TIMING DELAYS

To understand temporal control by gene regulatory networks with timed epigenetic switches, we first elucidate the temporal properties of individual genes regulated by epigenetic switches, and compare them to those of classically regulated genes, whose transcription rates depend solely on levels of transcription factor inputs. For both types of genes, we measure the relationship between the magnitude of an input signal and the time delay to an output response, to determine the conditions under which they can robustly respond with delays that are tunably controlled by regulatory inputs.

Classically regulated genes typically respond to regulatory inputs within one cell generation, as their expression dynamics are shaped by the mRNA and protein half-life, which are in turn determined by cell cycle duration (Alon, 2007; Levine and Elowitz, 2014). Epigenetic switching genes, on the other hand, could in principle respond and be controlled over a much wider range of timescales, as their switching dynamics are uncoupled from cell division and are instead determined by heritable chromatin modification and conformation states at the gene locus (Pease et al., in submission). To test this idea, we analyzed a series of simple gene regulatory networks, where a regulatory gene X receives an input from an upstream signal S and in turn activates a downstream response gene Y [Figure 2.2A, (I)-(V)] (Mathematical Appendix, Part I). X

is either regulated by a timed epigenetic switch, whose activation rate varies with input signal levels S [Figure 2.2A, (I)-(II)], or regulated classically, with transcription rates controlled by signal S [Figure 2.2A, (I)-(II)]. For each network, we performed simulations to measure its response delay, defined as the time between signal onset and expression of the output gene above a threshold. We considered both signal-induced promoters, that promote response gene expression [Figure 2.2A, (I),(III),(IV)]; and signal repressed promoters, that inhibit response gene expression [Figure 2.2A, (II),(V)]. In all cases, we both performed stochastic simulations and calculated analytical expressions for response times (Mathematical Appendix, Part II), to determine the generality of our findings.

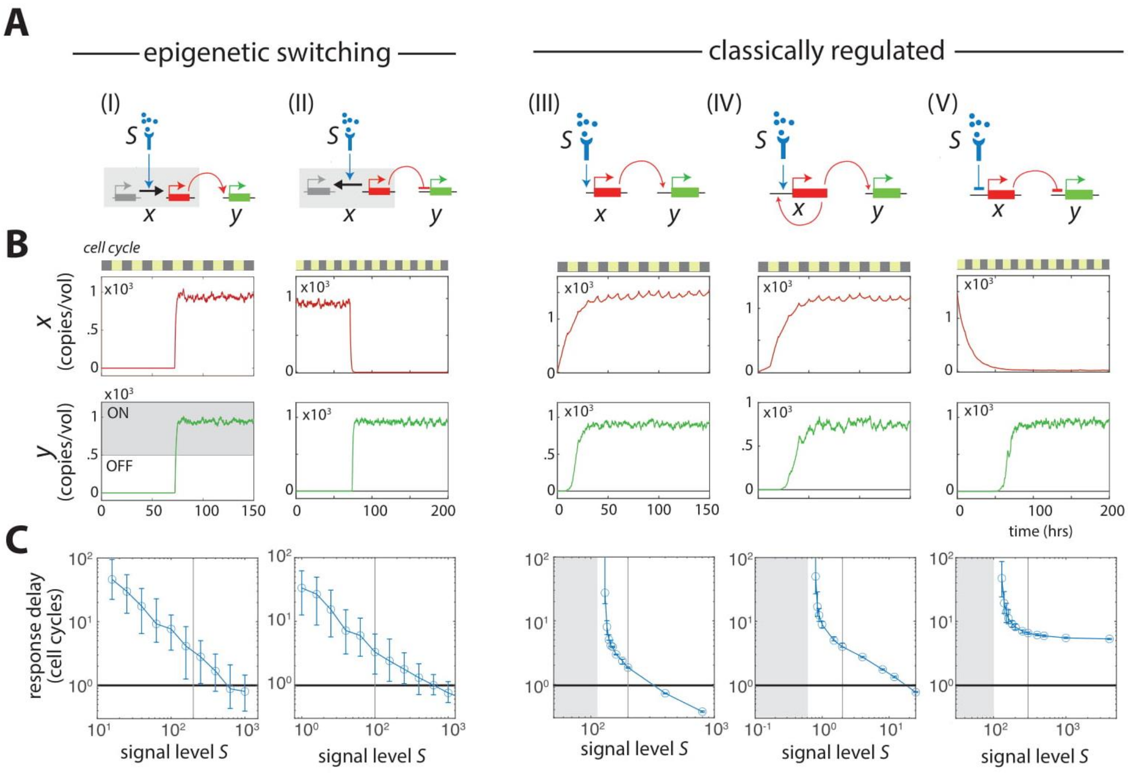


Figure 2.2: Epigenetic switches generate long, tunable delays in gene activation. Using stochastic simulations, we determined response times of different gene promoters (I-V), as indicated. Response times are defined as the delay between onset of an input signal S and activation of a response gene Y. For each promoter type, we show circuit diagrams (A), simulation time traces (B) and response delays, plotted against signal levels (C). Circles and error bars represent the median and inter-quartile range of the response time distribution, obtained over N = 100 simulations; the gray shaded area indicates the range of signals over which the circuit fails to respond.

From both simulations and analytical work, we found that epigenetic switches generated extended response delays that are tunable over a wide range of input signal levels. When exposed to strong signals, both signal-induced or signal-repressed promoters switched rapidly, resulting in rapid response gene induction; however, when signal levels decreased, gene switching rates decreased substantially, leading to longer delays spanning multiple cell generations in the response gene induction [Figure 2.2B,C, (I),(II)]. Response delays were variable from cell-to-cell, reflecting the inherently probabilistic nature of switching. However, in both cases, response time distributions were well defined, with average values that could be controllably varied over a wide range of timescales by varying signal magnitude (Mathematical Appendix, Part II).

Classically-regulated genes, on the other hand, were more limited in their ability to generate long, tunable response delays. As these gene promoters express immediately upon signal exposure, the response delays they generate would depend on the time at which their expressed gene products reach a threshold for downstream gene induction. For signal-induced promoters, this threshold is either crossed within one cell generation, or, when signals are weak, not crossed at all [Figure 2.2C,(III)], resulting in a sharp divergence in response times below a certain threshold level of signal. This `now-or-never' behavior arises because cell cycle duration imposes an upper limit on the half-life of the expressed protein, which determines its rise kinetics. Positive feedback on the gene prolonged response times, as previously observed [Figure 2.2C,(IV)] (Levine and Elowitz, 2014); however, delays exceeding one cell generation still occurred only over a narrow signal range, and there remained a large range of signals over which the system did not respond. To determine the generality of the result, we solved analytically for the system response times, both with and without positive feedback. In both cases, we found that defined response delays exceeding more than one cell generation could be upheld only over a very narrow range of system parameters; slight changes to the parameters resulted in either a failure to respond, or a substantial shortening of the response delay.

For classical signal-repressed promoters, responses could be delayed over many cell generations, if the threshold of X below which gene Y turns on is much lower than initial levels of X [Figure 2.2B,C,(V)]. Furthermore, this response delay is much less variable from cell-to-cell when compared to those generated by epigenetically-switching gene promoters. However, simulations and analytical work revealed that this long response delay cannot be tuned by levels of S; it remains largely constant over a range of signals, and diverges below a critical threshold, resulting in a failure to respond [Figure 2.2C,(V)]. Signal-dependent tuning of response delays could occur when signal levels were close to this critical threshold (Mathematical Appendix, Part II); however, as with signal-induced genes, small changes in signal levels in this regime would lead to a failure to respond, or substantially shorten response delays. Taken together, these results indicate that classically-regulated genes, in contrast to epigenetic switching genes, are fundamentally limited in their ability to generate tunable response delays that span multiple cell generations.

2.4 NETWORKS OF EPIGENETIC SWITCHES UPHOLD TEMPORAL SCHEDULES FOR GENE EXPRESSION

As single epigenetic switches generate long, tunable delays in gene activation that span multiple cell generations; networks of interacting epigenetic switches could potentially exhibit more elaborate temporal behavior over long timescales, and plausibly maintain timetables for gene expression and cell lineage specification during development.

A regulatory gene network's behavior depends on its topology, or the way in which its nodes are connected. Thus, we here consider specific network motifs - or recurrent connectivity patterns (Milo et al., 2002) - to see whether they may support the generation of particular types of temporal schedules. We specifically analyzed two network motifs frequently found developmental gene networks: negative feedback loops (Li et al., 2013; Kohwi and Doe, 2013) and cross-inhibitory loops (Alon, 2007; Singh et al., 2014). We simulated both these network motifs using timed epigenetic switches as nodes, and determined their resultant temporal responses.

Negative feedback loops are common in signaling and gene regulatory networks. A widely-established function of negative feedback loops is to enable adaptation in signaling responses (Ma et al., 2009; Ferrell, 2016). However, negative feedback loops may not all generate adaptation, but may instead play roles in setting temporal schedules for gene expression. For instance, in the gene networks controlling nervous system development in Drosophila (Li et al., 2013; Kohwi and Doe, 2013), negative feedback loops involving lineage-specifying transcription factors are thought to restrict the expression of transcription factors to sequential time windows. Similar types of negative feedback may also control lineage specification during nervous system development in vertebrates (Rossi et al., 2017).

Using mathematical modeling, we ask whether negative feedback loops in regulatory networks of epigenetic switches can generate sequential temporal windows of regulatory gene expression. We consider a simple network, where a signal S activates the promoter of regulatory gene X_1 (Figure 2.3A, top left) (Mathematical Appendix, Part II). X_1 then activates the promoter of regulatory gene X_2, which feeds back to inactivate the promoter of X_1. From Monte Carlo simulations, we found that this network generated sequential temporal windows for X_1 and X_2 expression (Figure 2.3A; bottom, right). These temporal windows each spanned multiple cell generations, and were largely mutually exclusive, a consequence of rapid silencing of the X_1 promoter by X_2. These simulation results show that negative feedback between two epigenetic switches can generate sequential temporal windows of gene expression.

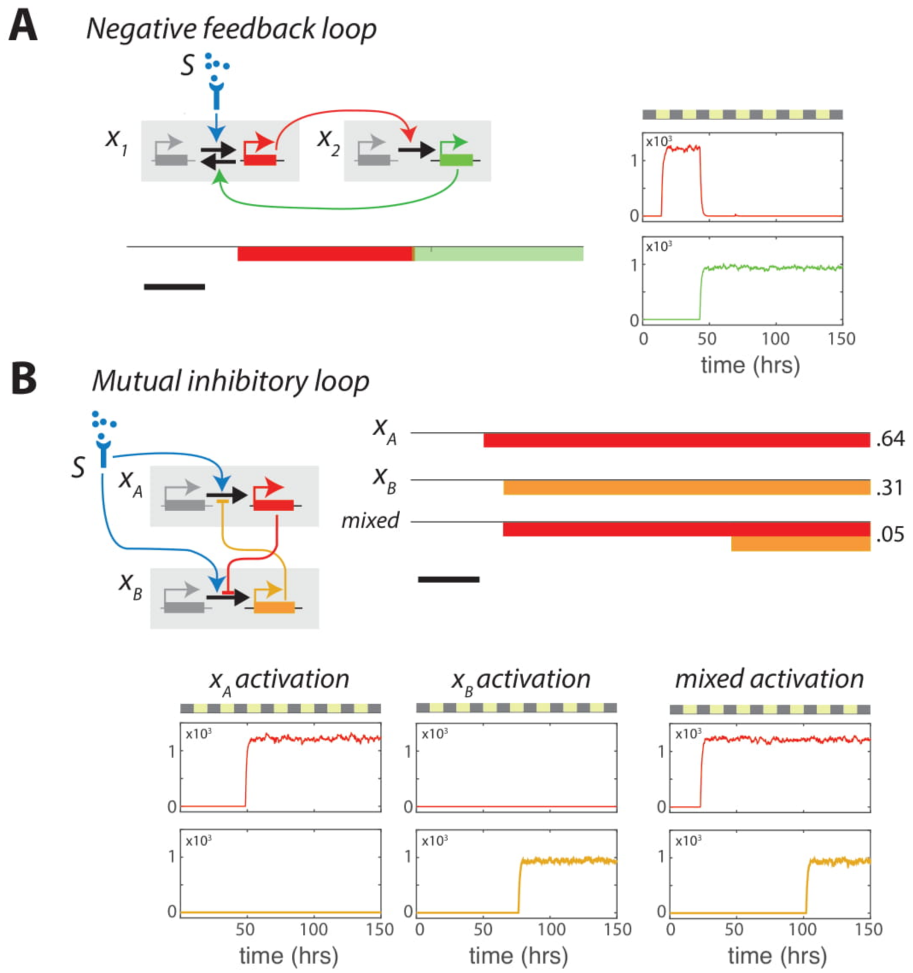


Figure 2.3: Epigenetic switching networks uphold robust temporal schedules for gene activation and lineage specification. (A) Negative feedback loops generate sequential temporal windows for gene activation. Shown are a network diagram (top left), an example simulation trace (right), and a timeline showing the averaged time window for expression of x_1 and x_2 (bottom left). (B) Mutual inhibitory loops allow concurrent gene expression time windows in different cell populations. Shown are a network diagram (top left), together with example simulations (bottom) and averaged gene activation timelines for cells activating x_A only, x_B only, or both genes. The fraction of cells within these categories are shown to the right of the timelines (top right).

Mutual inhibitory loops are common in developmental gene networks (Singh et al., 2014; Huang et al., 2007; Johnston and Desplan, 2010), and are frequently associated with the stabilization of multiple cell types. However, for cross-inhibitory loops between classically-regulated genes, which cell type is produced, and in what numbers, typically depend strongly on what external signaling inputs are present, and the direction in which they bias switching of the cross-inhibitory loop. In some developmental contexts, spatial heterogeneity in signaling inputs indeed drives heterogeneous fate outcomes (Gregor et al., 2005); however, in other systems, such a spinal motor column formation (Johnston and Desplan, 2010), multiple cell types can emerge in parallel over time from a single progenitor population, even in the apparent absence of heterogeneous external cues.

Due to the stochastic nature of epigenetic switching, mutual inhibitory loops involving epigenetic switches are expected to show substantial intrinsic variability in switching behavior; as such, they could enable progenitors to initiate multiple lineage specification schedules that unfold in parallel in their descendants, even in response to homogeneous signals. To test this idea, we modeled a simple mutual inhibitory network consisting of two regulatory genes, x_A and x_A, both of which are regulated by timed epigenetic switches. Both epigenetic switches activate in response to a common signal S, and both are inhibited in their activation by each other's expression (Figure 2.3B, top left) (Mathematical Appendix, Part III).

From simulations, we found that this mutual inhibitory loop of epigenetic switches could give rise to three parallel temporal schedules for gene expression upon exposure to the same inducing signals. Two of these schedules involved the activation of either x_A or x_B alone, whereas the third schedule involved the concurrent activation of x_A and x_B in the same cell (Figure 2.3B, top right). Such stable co-expression of the two regulatory genes results from incomplete inhibition of each other's activation, and may indeed occur in a number of systems as a way to diversify cell type identity and functionality (Zhu, 2018). We further note that these distinct temporal schedules emerged in defined fractions of cells, which were determined by the kinetic parameters of epigenetic switching as triggered by different regulatory inputs (not shown). Together, these results show that mutual inhibitory loops of epigenetic switches allow multiple timetables to unfold in parallel from a commo progenitor population in response to uniform external signals.

2.5 NETWORKS OF EPIGENETIC SWITCHES GENERATE MULTIPLE CELL TYPES IN DEFINED NUMBERS

While epigenetic switching networks are able generate robust temporal schedules for regulatory gene expression and lineage specification, these schedules would vary substantially from cell-to-cell due to the stochastic nature of epigenetic switching. (Figure 2.2C, I-II). Amid this variability, it is unclear how progenitors could nonetheless expand and give rise to precisely defined numbers and proportions of different cell types, as is the case during development.

Whether or not precise output cell numbers could arise from epigenetic switch networks would likely depend on the exact coupling schemes through which gene regulatory networks control cell division and differentiation. Here, we analyze different coupling schemes by which gene regulatory networks control division and differentiation, to determine whether there are particular schemes that support robust population size control. We first considered a simple scheme, where gene activation directly induces differentiation (Mathematical Appendix, Part IV). In this scheme, progenitors divide and self-renew indefinitely, until a signal S activates a gene Y that induces cell cycle exit and terminal differentiation (Figure 2.4A, I-II). Consistent with this scheme, a number of regulatory genes can directly induce cell cycle exit and differentiation (Kueh et al., 2013; Crescenzi et al., 1990; Sorrentino et al., 1990). We then coupled this scheme to an epigenetic switching network, where activation of Y is controlled by an upstream epigenetic event on an upstream regulator X (Figure 2.4A, II). For comparison, we also considered a classical regulatory gene network, where activation of Y occurs due to dilution of an upstream inhibitor X that is in turn repressed by signals (Figure 2.4A, I).

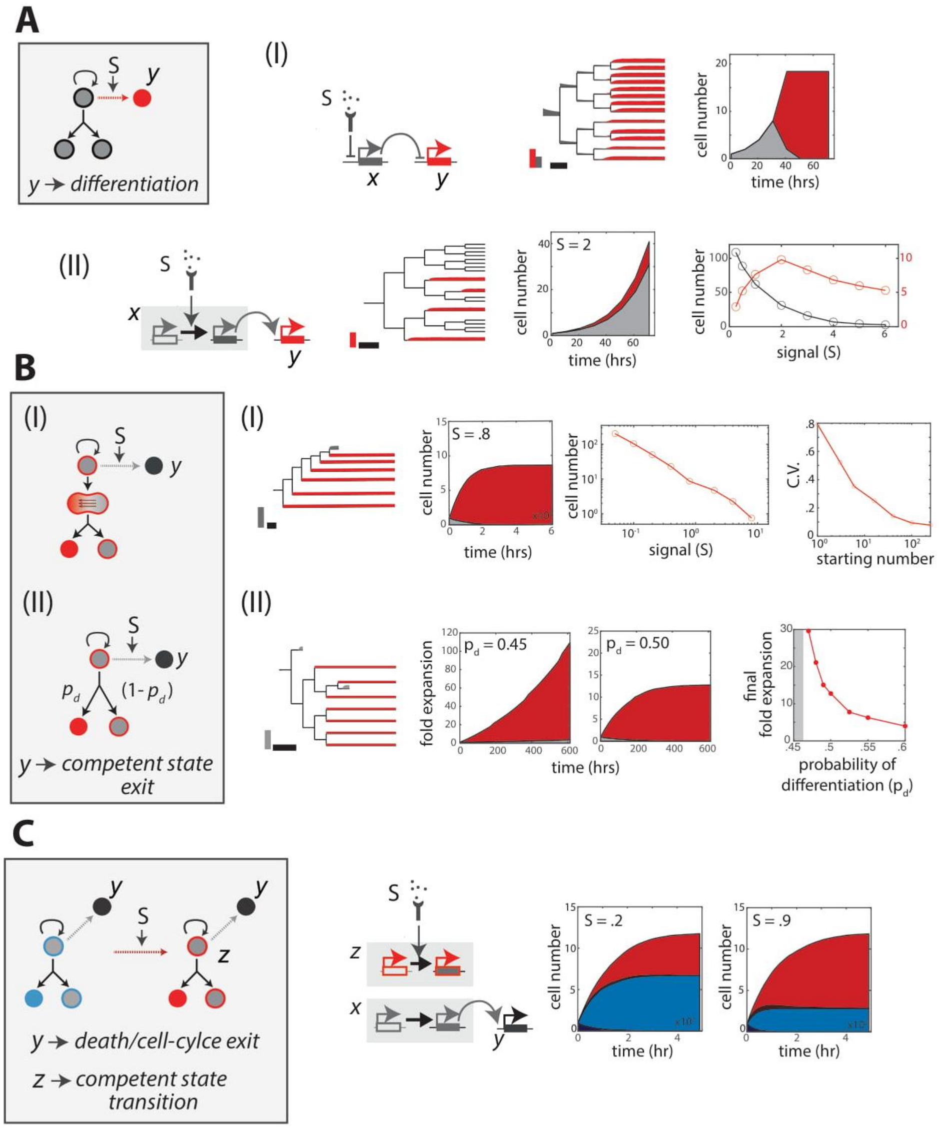


Figure 2.4: Epigenetic switching networks enable the generation of multiple cell populations in defined sizes and fractions. Two cell division coupling schemes are considered: one where gene activation induces differentiation and cell cycle exit (A), and another where gene activation controls the duration of a competent state, where non-dividing progeny are generated by asymmetric division (B.I, C) or stochastic mechanisms (B.II). In the first scheme, gene activation is induced by repression of an upstream inhibitor (A.I), or by epigenetic switching of an upstream activator (A.II). In the second scheme, epigenetic switching of an activator either induces a gene that promotes exit from a competent state (B), or promotes entry into second competent state (C). Each panel includes representative lineage trees superposed with colored areas representing gene product concentrations, area plots showing averaged cell numbers over time, as well as additional plots as indicated. Vertical bars denote concentration (gray = 10^3 copies/cell

(A.I), 4 x 10^3 copies/cell (B.I); red = 5 x 10^3 copies/cell (A.I-II)), and horizontal scale bars denote time (10 hrs).

When gene activation was controlled through classical trans- mechanisms (Figure 2.2, III), progenitors could expand over a fixed number of cell generations to generate larger progeny numbers (Figure 2.4A, I). However, when gene activation was gated by an upstream epigenetic switching event (Figure 2.4A, II), progenitors were not able to generate fixed progeny numbers. When input signals were high ($S > 4$), gene activation and differentiation occurred quickly, within a single cell generation; in this regime, progenitors could give rise to defined progeny numbers, but only with a minimal degree of cell expansion. On the other hand, input signals decreased, gene activation timing became protracted over many cell generations; however, progenitor numbers increased in an unbounded manner, reflecting the inability of rare probabilistic events to curtail exponential growth.

In some developmental systems (Kohwi and Doe, 2013), regulatory genes do not directly induce differentiation and cell cycle arrest, but instead control the entry or exit of progenitors from a competent state, where they are able to differentiate. In this competent state, cells could differentiate by asymmetric cell division, where they give rise to a progenitor and a terminally differentiated cell. Such asymmetry could occur through uneven partitioning of fate determinants, as observed in a number of systems (Knoblich, 2010a; Venkei and Yamashita, 2018). Alternatively, it could occur through a separate stochastic mechanism, where differentiated progeny are generated with a fixed probability during each cell division (Simons and Clevers, 2011a).

Here, using simulations, we asked whether epigenetic switching networks, when coupled to schemes involving competent state entry or exit, could generate defined numbers of output cells (Figure 2.4B). We first considered a simple scheme, where progenitors initially adopt a competent state, but eventually exit this state upon activation of a regulatory gene Y . Y is activated by X, whose promoter undergoes an epigenetic switching event controlled by signal S. We considered both differentiation strategies discussed above, where competent progenitors differentiate either by deterministic asymmetric cell division, or through stochastic mechanisms (Figure 2.4B, I-II). From simulations, we found that coupling epigenetic switching to competent state exit allowed for generation of differentiated progeny from rare starting progenitors. For both asymmetric and stochastic strategies for differentiation control, individual starting progenitors gave rise to heterogeneous numbers of offspring, as expected; however, larger cohorts of starting progenitors showed precise final fold increases in cell number, by averaging across multiple cell lineages (Figure 2.4B, I). In this scheme, population sizes increased with decreasing signal strength, reflecting the lengthened period of time a progenitor cell spends in a competent state. For stochastic control strategies, defined output population sizes were maintained when differentiation probabilities were close to one half per cell generation (Figure 2.4B, II). Notably, when competent state differentiation probabilities decreased beyond a threshold level, the number of output cells diverged in an unbounded manner (Figure 2.4B, II), thus reflecting the importance for these probabilities to be specified within a narrow range. Such tight control of differentiation probabilities is important for adult stem cell homeostasis (Simons and

Clevers, 2011a), and could also occur during embryonic development as well.

When developmental timetables set by epigenetic switching networks involve activation of multiple regulatory genes, they could give rise to multiple competent states that specify different cell types. Such a coupling may enable progenitors to not only control the numbers of different cell types, but also their relative proportions. To explore this possibility, we analyzed a simple epigenetic switching network that mediates the switching of progenitors from one competent state to another. In this scheme, progenitors exist in a first competent state (Figure 2.4C, blue), but enter a second competent state upon activation of Z by an upstream signal S (Figure 2.4C, red). The expression of Z is controlled by an epigenetic switch, whose activation probability is tunably controlled by S. In both competent states, a second gene Y activated by an upstream regulator X works independently to drive cell cycle exit. The activation of X is controlled by an epigenetic switch on its locus. In both competent states, differentiated cells are generated through asymmetric cell division; however, similar conclusions are expected when differentiation from the competent state is controlled through stochastic mechanisms.

From simulations, we found that this epigenetic switching network, when used to control transitions between two competent states, generates two cell types with controllable numbers and proportions. As expected, progenitors switched from one competent state (blue) to the next (red), giving rise to the two progeny types in sequence (Figure 2.4C). Increasing signal strength S, which accelerates switching to the second competent state, proportionally increased the numbers of the second cell type (red). Notably, changing signal levels S did not affect the total number of output cells, which would instead be modulated by the activation of gene Y and, consequently, by the activation timing of the epigenetic switch controlling the expression of X. Thus, regulatory inputs into epigenetic switching networks can be tuned to control the relative proportions of multiple cell types that emerge without altering their total numbers.

2.6 FLEXIBLE ADJUSTMENT OF DEVELOPMENTAL TIMETABLES DURING EVOLUTION

During evolution, developmental timetables vary across individuals and species, leading to variation in organ and organism sizes and proportions (Alberch et al., 1979; Keyte and Smith, 2014). While the temporal schedules set by epigenetic switching networks are robust over developmental timescales, they can also be flexibly modified by altering the strength of network components (Figure 2.4B,C); thus it is reasonable to propose that evolution may act upon epigenetic switching networks to vary developmental timetables and, consequently, organism size and form.

Here, by analyzing two paradigmatic evolutionary cases, we present evidence that mutations in epigenetic switching networks, occurring during evolution, could indeed alter developmental timetables to generate changes in sizes and proportions across species. In the first case, on body axis patterning by the Hox gene network, we show how non-coding mutations affecting epigenetic switching of individual genes can alter developmental events in timetables, and lead to shifts in cell type proportions. In the second case, on cerebral cortex development, we consider how changes to chromatin modifying enzymes can uniformly expand or contract developmental timetables, leading to scaled changes in total organism size. In both cases, we analyze simple mathematical models of epigenetic switching networks, coupled with appropriate cell proliferation control schemes, to illustrate the effects of evolutionary change on network activity and population sizes and proportions.

2.7 NON-CODING MUTATIONS AFFECTING EPIGENETIC SWITCHING CAN ALTER BODY PLAN PROPORTIONS

Patterning of vertebrate body axis requires the Hox genes, a family of transcription factors that form a regulatory network to specify segment identity. Proper temporal regulation of Hox genes is crucial for proper control of body axis proportions. Hox genes follow a defined timetable for activation in the embryo, where those specifying the anterior turn on before those specifying the posterior (Figure 2.5A) (Mallo and Alonso, 2013; Keyte and Smith, 2014; Zakany et al., 1997; Wakamatsu and Suzuki, 2019; Juan and Ruddle, 2003). Changes in the relative timing of Hox gene activation, caused by mutations in non-coding regulatory sequences, can shift anterior/posterior proportions: targeted deletion of an enhancer for Hoxd11 - a posterior-specifying Hox gene - delayed its activation, and shifted the sacrum towards the posterior (Zakany et al., 1997). Conversely, substitution of this enhancer for its more active fish counterpart accelerated gene activation, resulting in an anterior shift in sacrum positioning (Gerard et al., 1997).

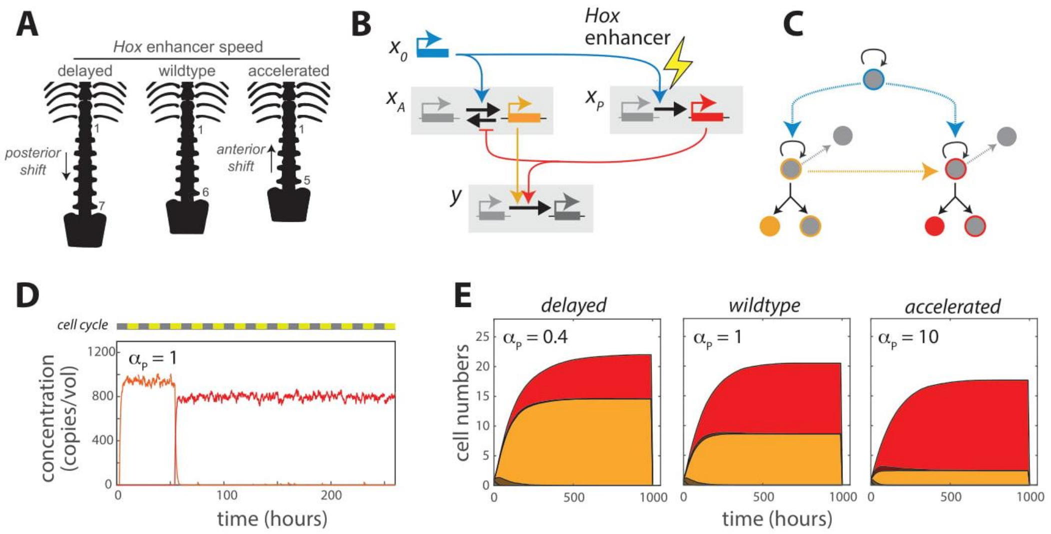


Figure 2.5: Non-coding mutations in epigenetic switching networks can alter developmental timing and cell type proportions. (A) Anterior-posterior positioning by Hox gene activation (G´erard et al., 1997). (B) Minimal epigenetic switching Hox gene network, including an anterior Hox gene (x_A), a posterior Hox gene (x_P), an upstream regulator (x_0), and a cell-cycle exit regulator (y). (C) Asymmetric division scheme for cell differentiation. Cell states are colored according expression of network regulators, as shown in (B). (D) Representative time traces of anterior and posterior Hox gene levels (x_A and x_P) in simulated single progenitor lineages. (E) Averaged numbers of anterior and posterior-specified cells over time, for three different relative rates of x_P activation (f). Increasing this rate accelerates entry into a posterior Hox gene expressing state, resulting in a higher proportion of posterior-specified cells.

Evidences from a wide range of studies indicate that that Hox genes, similarly to Bcl11b, are regulated by timed epigenetic switches that activate with long, tunable delays. Similarly to Bcl11b, Hox gene activation requires active histone H3K27me3 removal controlled by PRC2 and demethylases (Soshnikova and Duboule, 2009; Eskeland et al., 2010), and also involves decompaction of individual gene loci (Fabre et al., 2015; Eskeland et al., 2010). Moreover, like Bcl11b, Hox genes show variable regulation not only between cells, but also between individual loci within individual cells (Fabre et al., 2018, 2015), consistent with stochastic epigenetic control occurring at single loci. Thus, it is likely that the Hox gene network is, at its core, a network of epigenetic switches, where the activation timing of individual nodes can be tuned by regulatory inputs acting on enhancer elements.

Here, by analyzing a minimal model of an epigenetic switching network of Hox genes, we ask whether non-coding mutations previously found alter Hox gene expression (Zakany et al., 1997; Gerard et al., 1997) can indeed alter gene activation timing and anterior/posterior proportions, as experimentally observed (Figure 2.5B) (Zakany et al., 1997; Juan and Ruddle, 2003) (Mathematical Appendix, Part V). For simplicity, we aggregate anterior and posterior-specifying Hox genes into two groups, x_A and x_P . Both anterior and posterior gene groups are activated by a common upstream input x_0, whose function is to trigger an epigenetic switch leading to gene activation. Anterior (posterior) Hox genes have a faster (slower) time constant for activation, reflecting the order of activation encoded in the embryo. Once activated, posterior Hox genes silence anterior Hox gene expression, reflecting a phenomenon known as posterior dominance (Mallo and Alonso, 2013). Gene silencing is taken to be fast, such that anterior and posterior Hox gene expression is mutually exclusive. When expressed, the anterior (posterior) Hox gene group cause progenitors to reside in a competent state, where they give rise to anterior (posterior) cell types. Here, we assume these progenies are generated through asymmetric cell division (Figure 2.5C), as described (Figure 2.4), though we expect similar conclusions when they are instead generated through stochastic mechanisms.

Simulations show that this minimal Hox gene network recapitulated key features of Hox gene regulation and anterior and posterior specification. As expected, Hox genes turned on in an ordered temporal sequence with the anterior Hox gene group turning on first, and subsequent activation of the posterior Hox gene resulting in rapid anterior gene silencing (Figure 2.5D). Gene activation led to concomitant generation of anterior and posterior cells, with both populations expanding to reach a defined final size. To model the effects of disrupting the Hoxd11 enhancer, we changed the activation rate constant for the posterior Hox gene group α_P , either decreasing it to mimic deletion, or increasing it to mimic a more active variant. Consistent with experimental findings, lowering (raising) this rate constant diminished (enlarged) the posterior cell population relative to the anterior cell population, by shortening (lengthening) the time window over which this cell population was specified (Figure 2.5E). These results show how non-coding mutations affecting epigenetic switching rates in gene regulatory networks can readily change developmental timetables and cell type proportions in an organism.

2.8 CHROMATIN-MODIFYING ENZYMES CAN SCALABLY EXPAND OR CONTRACT DEVELOPMENTAL SCHEDULES

Evolutionary mutations lead not only to changes in the timing of events within developmental timetables, but also to the speed at which these timetables unfold (Bonner, 1965; Calder, 1984). During cerebral cortex development, neural progenitors follow a partly cell-autonomous temporal schedule for generating different cortical layer neurons that spans a week in mice (Eiraku et al., 2008; Gaspard et al., 2008), but expands to two months in macaques and four months in humans (van den Ameele et al., 2014; Espuny-Camacho et al., 2013; Otani et al., 2016; Barry et al., 2017b) (Figure 2.6A). These scalable changes in the timetable for cerebral cortex specification are likely to underlie the dramatic differences in brain sizes - and organ sizes in general - across different species.

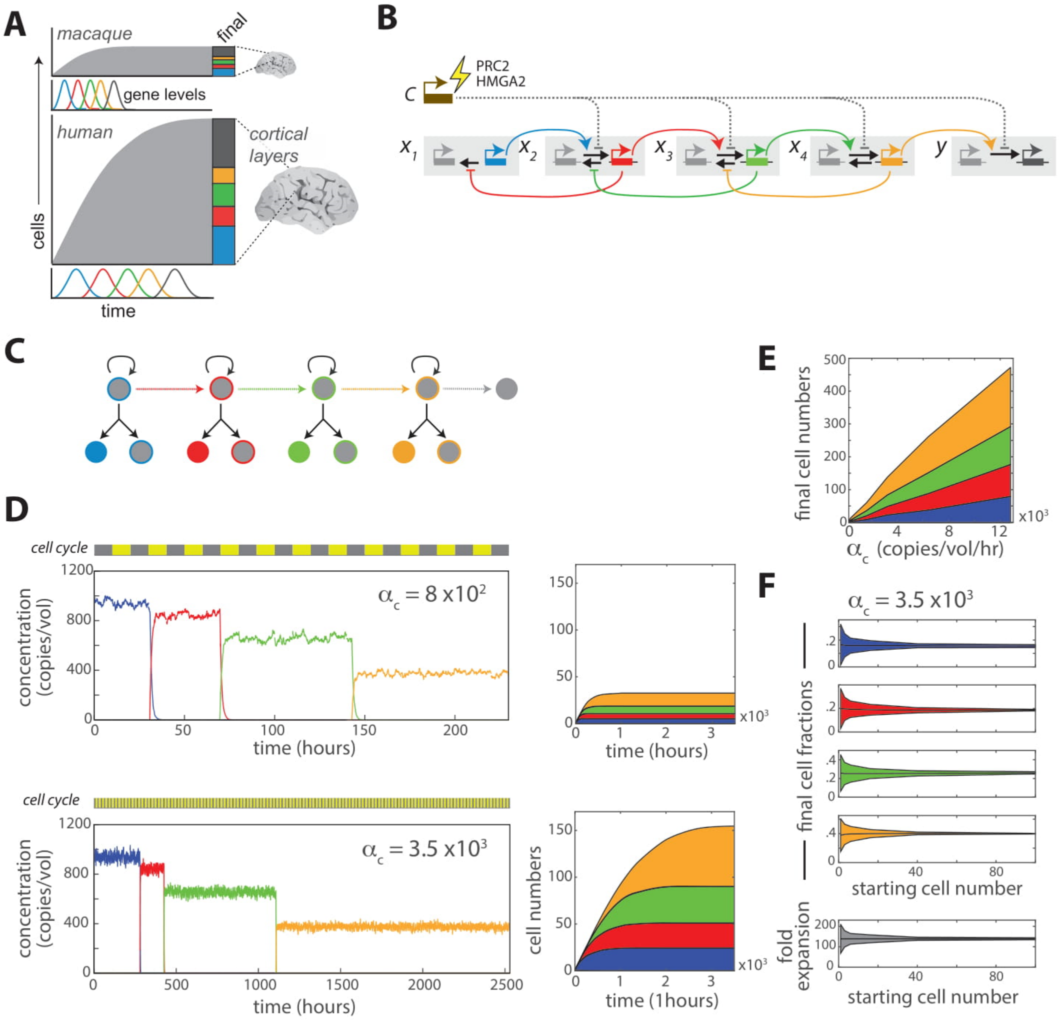


Figure 2.6: Chromatin regulators can scalably expand or contract developmental timetables to modify total organ size. (A) Cerebral cortical layers in different species (Otani et al., 2016). (B) Gene regulatory network controlling the specification of different cerebral cortical layers. (C) Asymmetric division scheme for neuronal differentiation,

with cell states colored according to expression of network regulators in (B). (D) Representative time traces (left) and averaged cell number area plots (right) for single lineage simulations, performed with higher (top) or lower (bottom) rates of chromatin regulator transcription α_c. (E) Output numbers of different cortical layer neuron types as a function of α_c. (F) Output fraction of neuron types (top), along with total fold expansion in cell number (bottom), plotted as a function of starting cell number. Line and shaded area represent mean and standard deviation. These results show that progenitors can generate multiple cortical layer neurons with precisely defined fractions, and that final cell numbers can be scalably adjusted by varying the chromatin regulator levels.

Chromatin modifying enzymes, with their broad specificity and their ability to tunably modulate epigenetic switching dynamics at multiple gene loci, could impact the speed at which temporal schedules set by epigenetic switching networks unfold. In a separate study (Pease et al., in submission), we found that the polycomb repressive complex 2 (PRC2) and Kdm6a/b, the writer and erasers of the H3K27 trimethylation mark respectively, tunably altered activation time delay for Bcl11b. Because PRC2 and Kdm6a/b broadly target lineage-specifying gene loci in stem cells (Boyer et al.; Lee et al., 2006), they could concurrently control switching times at multiple nodes in a regulatory gene network; consequently, their activity could control the overall speed at which temporal schedules, set by regulatory networks, unfold over time. Consistent with this idea, disrupting the PRC2 complex in cerebral cortex development does not simply change the timing at which a single neuronal layer type is specified, but alters the specified cation dynamics of multiple neuron types (Pereira et al., 2010), causing the entire temporal schedule for their generation to be completed in a shorter total duration.

Here, we test the idea that chromatin-modifying enzymes can act on epigenetic switching networks to control the speed of development. To do so, we developed a regulatory gene network model for cerebral cortex development. This network has a core architecture that is best elucidated from work in Drosophila, but may also underlie neural development in vertebrates (Rossi et al., 2017; Kohwi and Doe, 2013). This architecture consists of a cascade of transcription factors that activate each other's expression in succession (Figure 2.6B; Mathematical Appendix, Part VI). When activated, the regulators induce a competent states where a distinct cortical neuron type is produced through asymmetric progenitor division, or, in the case of the last regulator in the sequence, induce cell cycle exit, as observed (Kohwi and Doe, 2013) (Figure 2.6C). These regulators also repress their upstream activators in a negative feedback loop, to ensure their mutually exclusive expression (Figure 2.6B). Finally PRC2 provides an additional input into all other regulatory genes in the sequence, working to dampen their rates of promoter activation.

From simulations, we found that regulatory genes in this cascade turned on in a defined temporal order, consistent with experimental findings. Activation of a gene led to rapid silencing of its preceding regulator, as expected (Figure 2.6D, top), allowing these regulators to be expressed in sequential waves in single cell lineages. Activation times of regulators varied from cell to cell, leading to heterogeneity in the final numbers and fractions of different cortical layer types generated from a single progenitor (Figure 2.6F). However, as starting progenitor numbers increased, variability in cell type fractions decreased, showing that timing control that is stochastic at the single cell level can lead to precise control of cell type proportions at the population level.

When PRC2 levels were increased, progenitors activated regulatory genes in the same order, but proceeded through this temporal schedule for activation at a slower speed

(Figure 2.6D, bottom). This scaled expansion of the timetable for cortical layer neuron specification, combined with an unchanged rate of cell division, led to larger final numbers of different cortical layer neuron types (Figure 2.6D, E). Importantly, while total cell numbers increased, the proportions of neurons in different cortical layers remained the same. From an analytical solution of the system (Mathematical Appendix, Part VI), we find that such scaled changes in timing and population sizes occur over a range of parameters, as long as PRC2 similarly alters activation timing all regulatory gene loci when perturbed. These results show that chromatin modifying enzymes, acting on epigenetic switching networks, can scalably adjust the speed at which developmental schedules unfold and generate proportional changes in organ and organism sizes.

2.9 INDEPENDENT GROWTH AND TEMPORAL CONTROL

Evolutionary changes in organ and organism sizes often arise not from changes to developmental timetables, but from changes in cell proliferation amid an invariant developmental schedule. For example, while different dog breeds have birth weights that can vary by over five fold, they follow a similar embryonic timetable with a gestational period of two months (Mugnier et al., 2019; Okkens et al.). Relatively few genetic variants can explain the majority of the variation in breed sizes. Remarkably, most of these variants affect genes involved in the growth hormone/insulin-like growth factor (GH/IGF) axis (Rimbault et al., 2013). This signaling pathway controls the rate of cell growth and division and acts as a master controller of embryonic growth (Baker et al., 1993; Verhaeghe et al., 1993). Mutations altering GH/IGF signaling activity change the overall rate of proliferation in the embryo without affecting the temporal schedules for organ formation, resulting in proportionately scaled body sizes at birth (Baker et al., 1993; Gagliardi et al., 2005; Garfunkel et al., 2015). Similarly, decreasing c-Myc expression in mice reduces embryonic proliferation rates without affecting developmental schedules, resulting in neonates with fewer cells that retain their basic anatomical proportions (Trumpp et al., 2001; Hofmann et al., 2015).

It has been unclear how changes to cell proliferation control can occur without concomitant changes in developmental timetables. Classical gene regulatory networks that are capable of generating long delays (Figure 2.2III-V) tend to have protein and mRNA species that are stable over the cell cycle, such that their levels depend on their rate of dilution with cell division. In contrast, the timed epigenetic switches we have identified operate independently from cell division (Pease et al., in submission); thus, epigenetic switching networks could potentially set developmental schedules invariant to cell proliferation changes (Figure 2.7A). To examine this idea, we developed a toy regulatory network for body plan formation, and analyzed this network to define the conditions under which epigenetic switching networks could generate invariant developmental schedules that allowed for tunable control of cell population sizes and fractions through control of cell division rate.

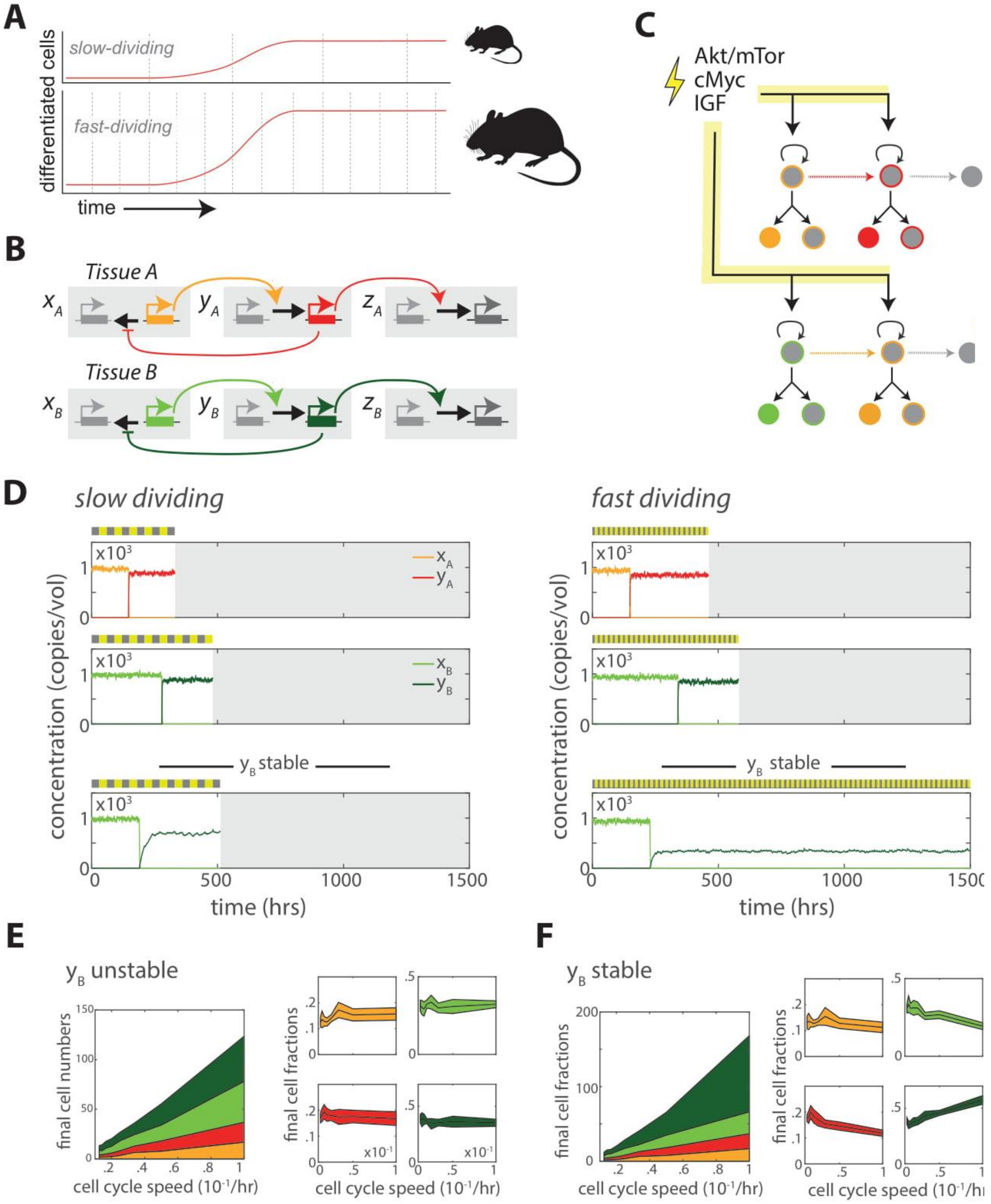


Figure 2.7: Master growth regulators, acting on division-independent developmental schedules, can proportionally vary population sizes. (A) Perturbations to cell cycle regulators result in coordinated scaling of independently developing tissues. (B) Two gene regulatory networks for the two tissue types, each with an upstream regulator (X_A, X_B), a downstream regulator (Y_A, Y_B), and a cell cycle exit regulator (Z_A,Z_B). (C)

Asymmetric division scheme for cell type specification, with cell states colored according to their corresponding regulators. (D) Representative time traces showing concentrations of regulatory genes, in single cell lineages with slower (left) or faster (right) cell division rates. In both instances, we considered two parameter sets: one where YB is unstable (middle), and another where it is stable (bottom). (E-F) Averaged cell number plots (left) and final cell fractions (right), plotted as a function of cell cycle speed, obtained under regimes where Y_B is either unstable (E) or stable (F). These results show that changes in cell division rates can lead to proportional changes in cell population sizes, but only when regulators have short half-lives and, consequently, concentrations that do not vary with cell cycle speed.

To capture the autonomous development of multiple parts of a body plan, we incorporate into our model two epigenetic switching networks, each operating in parallel in different progenitor populations to generate a total of four differentiated cell populations (Figure 2.7B, C) (Mathematical Appendix, Part VII). While these progenitors develop independently from each other, they share a cell cycle control mechanism that can be adjusted by evolutionary mutations. These mutations could occur in the GH/IGF signaling pathway, as in the case for dogs, or in other cell division speed regulators, such as c-Myc or Akt/mTOR signaling (Chen et al., 2001; Peng et al., 2003). We then simulate the time evolution of each regulatory network independently, to determine the effects of changing cell division rates on sizes and proportions of differentiated cell populations.

From stochastic simulations, we found that progenitors could maintain a constant developmental schedule over a range of cell cycle speeds (Figure 2.7D, top and middle; E). When the speed of cell cycle progression increased, the population sizes of all differentiated cell types increased in a proportional manner, such that the fractions of different cell types remained largely constant.

Developmental schedules maintained constancy only when proteins encoded by regulatory network genes were all unstable, such that their half-lives were shorter than the cell cycle duration. When a regulatory protein in this network was set to be stable, such that its levels now varied with changing cell cycle length, there was a change in the activation timing of its target gene, which in turn led to disproportionate increase in the size of its corresponding cell population (Figure 2.7D, bottom, F). These results indicate that rapid protein turnover in epigenetic switching networks, relative to cell division, is a requirement for maintaining developmental timetables that are independent of cell division speed.

2.10 METHODS AND MATHEMATICAL APPENDIX

Here, we describe the gene regulatory network models analyzed in this study. Some models follow regulatory dynamics of single cells within individual lineage branches (Sections I-III), whereas others extend these models to include lineage information (Sections IV-VII). Simulation parameters are provided in Table 2.1. All models were simulated using Tellurium, a python environment for biochemical network modeling (Choi et al., 2018). Simulations were run using the Gillespie algorithm (Gillespie, 1977).

For simulations of cell lineages, individual cells within lineages were simulated separately, and the ending conditions of each cell were used to compute initial conditions for its two daughters. Data from simulations were then exported and subsequently analyzed and/or plotted using MATLAB (Mathworks, Natick, MA).

COMPARISON OF CLASSICAL AND EPIGENETIC SWITCHING GENES

Here, we will describe models of networks with either classical or epigenetic switching genes, to determine whether epigenetic switching genes can generate more reliable response delays that are tunable over timescales spanning multiple cell divisions. In each model, an upstream signal S activates (represses) a regulatory gene X with a promoter P_X, which in turn activates (represses) a downstream response gene Y (Figure 2.2). For a promoter activated by epigenetic switching, the reactions describing this network circuit is given by:

$$\begin{array}{lcll}
P_x & \xrightarrow{\alpha_X \cdot S \cdot (1-P_x)} & P_x + 1 & \\
X & \xrightarrow{\gamma_X \cdot V \cdot P_x} & X + 1 & \\
X & \xrightarrow{\delta_X \cdot X} & X - 1 & \\
Y & \xrightarrow{\gamma_Y V \cdot x^n/(x^n+K^n)} & Y + 1 & \qquad\qquad (10) \\
Y & \xrightarrow{\delta_Y \cdot Y} & Y - 1 & \qquad\qquad (1)
\end{array}$$

where X is the rate constant for X promoter activation, X is a rate constant of X transcription from an active promoter, X is the first-rate constant for X degradation, Y

is the maximal rate constant of Y transcription, $x \ = \ X = V$ is the concentration of x in the cell, K is the constant for half-maximal transcription of Y by X, and Y is the rate of Y degradation. Here, we assume that S controls of X promoter activation with a first-order rate, whereas X regulates transcription of Y following a Michaelis-Menten relationship. The rate constants for simulations shown in (Figure 2.2) are given in Table S2.1.

For a promoter repressed by epigenetic switching, the underlying reactions are given by:

$$P_x \xrightarrow{\beta_X \cdot S \cdot P_x} P_x - 1$$

$$X \xrightarrow{\gamma_X \cdot V \cdot P_x} X + 1$$

$$X \xrightarrow{\delta_X \cdot X} X - 1$$

$$Y \xrightarrow{\gamma_Y V \cdot K^n/(x^n + K^n)} Y + 1 \qquad (11)$$

$$Y \xrightarrow{\delta_Y \cdot Y} Y - 1 \qquad (2)$$

Here, S turns off the promoter of X with rate constant α_X, and X in turn inhibits Y transcription with a hyperbolic function.

For a classically-regulated promoter turned on by a signal S, the underlying reactions are given by:

$$\begin{array}{ccc} X & \xrightarrow{\gamma_X V\cdot S} & X+1 \\ X & \xrightarrow{\delta_X\cdot X} & X-1 \\ Y & \xrightarrow{\gamma_Y V\cdot x^n/(x^n+K^n)} & Y+1 \\ Y & \xrightarrow{\delta_Y\cdot Y} & Y-1 \end{array} \qquad (12)$$

Here, S directly up-regulates transcription of X with first-order rate constant γ_X; X then up-regulates transcription of Y following a Hill function.

For a classically-regulated promoter activated by both a signal S and by itself, through a positive feedback loop, the underlying chemical reactions are:

$$\begin{array}{ccc} X & \xrightarrow{\gamma_0 V\cdot S+\gamma_1 V\cdot x^m/(x^m+K^m)} & X+1 \\ X & \xrightarrow{\delta_X\cdot X} & X-1 \\ Y & \xrightarrow{\gamma_Y V\cdot x^n/(x^n+L^n)} & Y+1 \\ Y & \xrightarrow{\delta_Y\cdot Y} & Y-1 \end{array} \qquad (13)$$

Here, S activates X with a rate constant γ_0, and X also cooperatively activates itself, following a Hill function with maximal rate constant γ_1, Hill coefficient m and half-maximal constant J.

For a classically-regulated promoter repressed by both a signal S, the underlying chemical reactions are:

$$
\begin{array}{lcl}
X & \xrightarrow{\gamma_X V\cdot K/(K+S)} & X+1 \\
X & \xrightarrow{\delta_X \cdot X} & X-1 \\
Y & \xrightarrow{\gamma_Y V\cdot L^n/(x^n+L^n)} & Y+1 \\
Y & \xrightarrow{\delta_Y \cdot Y} & Y-1
\end{array}
\qquad (14)
$$

Here, signal S inhibits transcription of X following a hyperbolic function with half maximal constant K.

In all cases, we simulated these chemical reaction scheme in a single cell with cell cycle duration t_c^m. As described above in the modeling, cell volumes change V with cell cycle; at the end of each cell cycle, proteins are then partitioned randomly, whereas promoter activation states are inherited from the parental cell. At time $t = 0$, signal S is provided to the system; the response time is then defined as the time when levels of the response gene Y increase above a certain threshold Y_T.

Classically-regulated genes do not robustly generate long, tunable response delays.

Our simulation results suggest that the classically-regulated genes are fundamentally limited in their ability to generate robust, yet tunable delays that exceed one cell generation. To demonstrate the generality of the result, we consider the deterministic equivalent of the model for a signal-activated gene that is classically regulated, as follows:

$$\frac{dx}{dt} = \gamma_x S - \delta_x x \qquad (15)$$

where α is the synthesis rate of the gene, and δ_x is the rate of protein removal, by either degradation or cell cycle mediated dilution. Here, we assume that the protein shows negligible degradation over the course of a cell cycle, such that $\delta_x \approx 1/\tau_c$, where τ_c is the duration of the cell cycle.

We now consider a situation, where the gene is silent, but is subject to signals at $t = 0$ that activate its expression. The expression of x increases progressively, until it reaches a threshold x_t, upon which it triggers a downstream response. By analytically solving for this differential equation, we can obtain the response delay D_t, defined as the time when x reaches the response threshold x_t, as follows:

$$D_t = \frac{1}{\delta_c} \cdot \ln\left[\frac{\gamma_x S}{\gamma_x S - \delta_c x_t}\right]$$

Based on the equation, we now calculate parameter conditions for which D_t must activate after a given number of cell cycles N, (i.e. when $D\_t > N/\delta_t$). By substituting this parameter condition into the above equation, we get the following constraints on the model parameters:

$$\left[1 - \frac{1}{e^N}\right] < \frac{\gamma_x S}{\delta_c \cdot x_t} < 1 \qquad (16)$$

Based on this equation, it is clear that, when $N \ll 1$, the range of parameters over which this constraint is satisfied becomes very narrow. For instance, when $N = 5$, then:

$$0.99 < \frac{\gamma_x S}{\delta_c \cdot x_t} < 1 \qquad (17)$$

Thus, even a very minor change to any of the parameters in the model will lead to a failure in satisfying this condition.

We can perform a similar analysis for a signal-induced gene with positive feedback, as follows:

$$\frac{dx}{dt} = \Gamma(S) - \delta_x X \qquad (18)$$

Where:

$$\Gamma(S) = \gamma_0 S + \gamma_1 \cdot \frac{x^m}{x^m + K^m}$$

Here, the rate constants are defined equivalent to those in the stochastic model above. To derive a general expression for the response delay as function of parameters, we first take the limit where the positive feedback is highly ultra-sensitive, such that $m \to \infty$. In this regime, the rate of gene synthesis can be approximated by the following piecewise linear function:

$$\Gamma(S) \approx \gamma_0 S \qquad \text{if} \qquad x < K \tag{19}$$

$$\approx \gamma_0 S + \gamma_1 \qquad \text{if} \qquad x > K \tag{20}$$

In this limit, we can solve for the delay D_t for gene levels to reach a threshold x_t, as follows:

$$D_t = \frac{1}{\delta_C} \cdot \left[\ln\left(\frac{\gamma_0 S}{\gamma_0 S - \delta_x K}\right) + \ln\left(\frac{\gamma_0 S + \gamma_1 - K}{\gamma_0 S + \gamma_1 - x_t}\right) \right] \tag{21}$$

This expression reveals that long response delays can arise under the following regimes: when either the first or second terms on the right hand side are large.

The parameter conditions satisfying the first regime is given by:

$$1 < \frac{\gamma_0 S}{\delta_x \cdot K} < \frac{e^N}{e^N - 1} \tag{22}$$

The parameter conditions satisfying the second regime is given by:

$$1 < \frac{\gamma_1 + \gamma_0 S}{\delta_x \cdot x_t} < \frac{e^N}{e^N - 1} \tag{23}$$

In the first regime, long response delays arise because of long times required to activate the positive feedback loop; in the second regime, delays arise because of long times required to reach a threshold for downstream gene expression. In either case, when delay exceeds one cell generation, the right hand side of the condition converges rapidly to unity, providing only a narrow range of parameter over which a long delay can be generated. For instance, when $N = 5$ cell generations, the right hand side of both expression is only slightly higher than unity (1.0067), allowing for only 0.7% variation in the combination of parameters above.

We next analyze a classically-regulated gene that is repressed by a signal. The ordinary differential equation describing this system is given by:

$$\frac{dx}{dt} = \frac{\gamma_x K}{S + K} - \delta_x x \qquad (24)$$

We now calculate time required for x to drop to a threshold value, x_t, below which a response occurs. Assume that the system starts its initial steady state value in the absence signal, such that $x(0) = \gamma_x/\delta_x$. At time $t = 0$, the signal S is introduced.

Let as assume, as is the case for the simulations that the response threshold x_t is much lower than the initial levels $x(0)$, such that there is a multi-generational response delay, even in the presence of strong signals. In this case, we can solve analytically for the time at which x drops to its response threshold value:

$$D_t = \frac{1}{\delta_C} \cdot \ln\left(\frac{x_0}{x_t - x_0 \cdot K/S}\right) \qquad (25)$$

where

$$x_0 = \gamma_x/\delta_x \qquad (26)$$

is the initial level of x prior to introduction of signal S. When the signal is strong, such that $x_t \gg (x_0 \cdot K/S)$, a response delay of N cell generations or greater can only occur when:

$$\frac{x_t}{x_0} < \frac{1}{e^N} \qquad (27)$$

In this regime, where signals are strong, the response delay varies with the response threshold x_t and the maximal steady state level x_0, but is largely invariant to changes in the level of signal S. When signals are weaker, there is a regime over which response delays can vary with signal levels. In this regime, response delays occur when:

$$0 < \left(\frac{x_t}{x_0} - \frac{K}{S}\right) < \frac{1}{e^N} \qquad (28)$$

However, in this regime, when N exceeds one cell generation, this condition is satisfied only for a very narrow range of S. When S falls below the value $S = x_t/(x_0 K)$, the system fails to respond. Thus, while classical signal-repressed genes can mount delayed responses, these responses are tunable over only a narrow range of signal levels S.

Epigenetic switching genes generate signal-tunable response delays

For epigenetic switching genes, response delays are dominated by the waiting time for the gene promoter to switch to an active or silent state in response to signals. As switching is governed by first-order kinetics, occurring with a constant probability per unit time, the waiting time distribution for switching is exponential, as follows:

$$p(\tau) = k \exp\left(-k \cdot \tau\right) \qquad (29)$$

where k is the first-order rate constant for switching. For a gene that is activated in response to a signal, this first-order rate constant can be tunably controlled by signal strength $k = \alpha_X S$, as we characterize in our accompanying study (Pease et al., in prep); consequently, we can integrate over this waiting time distribution to obtain the mean response delay:

$$\begin{aligned} \langle D_t \rangle &= \int \tau \cdot p(\tau) d\tau && (30) \\ &= \int_0^\infty \tau \cdot \left[(\alpha_X S) e^{-\alpha_X S \tau}\right] d\tau && (31) \\ &= \frac{1}{\alpha_X \cdot S} && (32) \end{aligned}$$

We note that, in contrast to classically-regulated genes, the response time of the system does not diverge when signals fall below a threshold value; as long as there is a non-zero level of signal, there will be a non-zero probability of switching. Thus, for epigenetic switching events that give rise to activation, response times will be generally be tunably over a wide range of signal magnitudes.

We can perform the same analysis for a gene that is epigenetically silenced by a signal with first-order kinetics. While our accompanying paper did not explicitly model signal-induced switching to a repressed state, other studies has found that signal-induced epigenetic switching similarly occurs stochastically, with first-order kinetics that are tunable by signaling inputs (Bintu et al.). In this case, the mean response delay is given by:

$$\langle D_t \rangle = \frac{1}{\beta_X \cdot S} \tag{33}$$

A NEGATIVE FEEDBACK LOOP OF EPIGENETIC SWITCHES

We consider a simple network, where a signal S activates the promoter of regulatory gene X. X then activates the promoter of a second regulatory gene Y , which feeds back to inactivate the promoter of X. The chemical reactions underlying this network

Model		Description	Value
Fig. 2.2(I)	α_X	epigenetic activation rate constant of X promoter	10^{-4}/hr
	γ_X	synthesis rate of X	10^3 copies/hr
	δ_X	degradation rate of X	1/hr
	γ_Y	maximal synthesis rate of Y	10^3 copies/hr
	n	Hill coefficient for Y synthesis	10
	K	level of x for half-maximal Y synthesis	200 copies
	δ_Y	degradation rate of Y	1/hr
Fig. 2.2(II)	β_X	epigenetic silencing rate constant of X promoter	2×10^{-3}/hr
	γ_X	synthesis rate of X	10^3 copies/hr
	δ_X	degradation rate of X	1/hr
	γ_Y	maximal synthesis rate of Y	10^3 copies/hr
	n	Hill coefficient for Y synthesis	10
	K	level of x for half-maximal inhibition of Y synthesis	40 copies
	δ_Y	degradation rate of Y	1/hr

Table 2.1: Model Parameters (I)

are as follows:

$$
\begin{array}{lcl}
P_x & \xrightarrow{\alpha_X \cdot S \cdot (1-P_x)} & P_x + 1 \\
P_x & \xrightarrow[12]{\beta_X \cdot P_x \cdot y/(y+K)} & P_x - 1 \\
X & \xrightarrow{\gamma_X \cdot V \cdot P_x} & X + 1 \\
X & \xrightarrow{\delta_X \cdot X}| & X - 1 \\
P_y & \xrightarrow{\alpha_Y S \cdot (1-P_y) \cdot x/(x+L)} & P_y + 1
\end{array}
$$

$$Y \xrightarrow{\gamma_Y V \cdot P_y} Y+1$$

$$Y \xrightarrow{\delta_Y \cdot Y} Y-1$$

(34)

Here, assume that S activates the promoter of X with first-order kinetics at a rate constant X; X activates the promoter of Y following Michalis-Menten kinetics, with half-maximal constant L, and maximal rate constant Y ; Y then switches off P_X with hyperbolic rates, with half-maximal constant K and maximal rate X.

This network was simulated in single cells with cell cycle duration t_c^m. As before, cell volumes change V with cell cycle. Finally, to obtain timetables for the expression of X and Y, simulations were repeated, and activation times for these two genes were recorded and averaged.

A MUTUAL INHIBITORY LOOP OF EPIGENETIC SWITCHES

We describe a simple network, where a signal S activates the promoters of two regulatory genes X and Y . Once activated, these genes inhibit the signal-dependent activation of the other. This network is described by this reaction scheme:

$$\begin{array}{lcl}
P_x & \xrightarrow{\alpha_X S\cdot K/(y+K)\cdot(1-P_x)} & P_x+1 \\
X & \xrightarrow{\gamma_X\cdot V\cdot P_x} & X+1 \\
X & \xrightarrow{\delta_X\cdot X} & X-1 \\
P_y & \xrightarrow{\alpha_Y S\cdot(1-P_y)\cdot L/(x+L)} & P_y+1 \\
Y & \xrightarrow{\gamma_Y V\cdot P_y} & Y+1 \\
Y & \xrightarrow{\delta_Y\cdot Y} & Y-1
\end{array} \qquad (35)$$

Here, we assume that P_X and P_Y both activate in response to S with first-order kinetics. Their gene products X and Y then inhibit each other following a hyperbolic function, with half-maximal constant K, for Y , and L for X. As before, single cells were simulated with cell cycle duration t_c^m multiple simulations were run to obtain averaged time windows for expression of X and Y .

COUPLING OF EPIGENETIC SWITCHING NETWORKS TO CELL PROLIFERATION CONTROL

For these and all subsequent models, we will simulate gene regulatory networks across on entire cell lineages. We do so in order to understand how cell numbers evolve over time, and how this time evolution may depend on the strategy for coupling cell differentiation with gene regulatory network activity.

We considered two coupling strategies, one where gene activation directly induces differentiation and cell cycle exit, and another where gene activation controls the entry or exit of a cell from a competent state where it can differentiate. For the first strategy, expression of a regulatory gene Y above a certain threshold induces cell differentiation arrest. In this scheme, progenitors initially divide with cell cycle time t_c^m, generating two daughters that can both divide further. However, when a progenitor expresses the regulatory gene Y above a certain threshold Y_T anytime during its initial cell cycle lifetime, it will no longer divide; instead, it will exit from the cell cycle to generate a differentiated, growth arrested cell. We coupled this cell division control scheme to two regulatory networks, the classical signal-induced repression model (Model I a, Eqs. 5), and the epigenetic promoter activation model (Model I a, 1). In both cases, we simulated single cells recursively using this coupling scheme, until all the cells stopped dividing, or until a time limit t_{max} was reached.

For the second strategy, expression of a regulatory gene Y controls the entry or exit of progenitors from a differentiation competent state. In this competent state, progenitors divide with cell cycle time t_c^m, as before; however, with each cell division, they can either divide asymmetrically, giving rise to one progenitor and one differentiated, growth-arrested cell; alternatively, with each cell division, each daughter can differentiate with a fixed probability p.

For the first model (Figure 2.4B), cells initially reside in this competent state, and expression of Y above a threshold concentration y_T leads to death of the cell at its point of division (Figure 2.3B.I). Here, the signal-dependent epigenetic activation of X results in the expression of Y, as given by Eqs. 1).

For the second model (Figure 2.4C), cells initially reside in a competent state where they give rise to cell state 0. The signal S then leads to the epigenetic activation of the promoter of Z_S, whose expression beyond a concentration threshold z_T switches the cell to a second competent state 1, where they generate a second cell type through asymmetric division. Independently, epigenetic activation X then causes the cell to turn on the growth arrest Y regulator. When Y reaches a concentration threshold y_T , the progenitor, irrespective of its state, undergoes cell death at the end of its lifetime. The following reaction scheme describes this simple network:

$$\begin{array}{ccc}
P_x & \xrightarrow{\alpha_X\cdot(1-P_x)} & P_x+1 \\
X & \xrightarrow{\gamma_X V\cdot P_x} & X+1 \\
X & \xrightarrow{\delta_X\cdot X} & X-1 \\
Y & \xrightarrow{\gamma_Y V\cdot x^n/(x^n+K^n)} & Y+1 \\
Y & \xrightarrow{\delta_Y\cdot Y} & Y-1 \\
P_z & \xrightarrow{\alpha_Z S\cdot(1-P_z)} & P_z+1 \\
Z & \xrightarrow{\gamma_Z V\cdot P_Z} & Z+1 \\
Z & \xrightarrow{\delta_Z\cdot Z} & Z-1
\end{array} \qquad (36)$$

Here, we assume that signal S activates P_X with a first-order kinetics; also, we assume that X activates Y in a cooperative manner, with half-maximal concentration K, Hill coefficient n, and maximal rate constant Y . As with the above models, single cells are simulated recursively, until there are no more proliferating cells, or until a time limit t_{max} is reached.

Competent state control enables generation of defined population sizes

Here, we derive an analytical expression for the number of differentiated progeny generated in the competent state control schemes described above. We consider the model (Figure 2.4B), where cells initially exist in a differentiation competent state, but then exit this state upon activation of an epigenetic switch. Let τ be the random variable representing the time to epigenetic switching, and let the distribution of this waiting time be $p(\tau)$. Cells in this competent state have a cell cycle duration of τ_c; at each division, we then assume here that they divide asymmetrically, generating one progenitor and one differentiated cell. Thus, for a cell that undergoes epigenetic switching at time τ , the number of differentiated cells generated are then given by:

$$n(\tau) = \lfloor \tau / \tau_c \rfloor \qquad (37)$$

We can now integrate over the waiting distribution to get the average number of progeny generated by a single cell:

$$\langle n\rangle \approx \int(\tau/\tau_c)\cdot p(\tau)d\tau \tag{38}$$

$$\approx \int_0^{\infty}(\tau/\tau_c)\cdot ke^{-k\tau}d\tau \tag{39}$$

$$\approx \frac{1}{\tau_c\cdot k}. \tag{40}$$

Here, $k = \alpha\cdot S$ is the rate of signal-dependent epigenetic switching. This expression shows that population sizes can be increased either by shortening the cell cycle duration, τ_c, or by decreasing the first-order rate of epigenetic switching, k.

MINIMAL HOX GENE NETWORK

Here, we describe a minimal model of the Hox gene network. In this model, we do not try to comprehensively capture the regulatory dynamics of all Hox genes, but instead seek to capture the essential architecture and dynamic features of this network. In this model, we partition anterior and posterior-specifying Hox genes into two groups, X_A and X_P . The posterior Hox gene X_P, once activated, can promote the silencing of the anterior Hox gene promoter P_A. Both Hox genes, once activated, can serve to activate a cell cycle terminating gene Y. The following equations describe the chemical reactions that underlie this network:

$$\begin{array}{ccc}
X_0 & \xrightarrow{\gamma_0 V} & X_0 + 1 \\
X_0 & \xrightarrow{\delta_0 X_0} & X_0 - 1 \\
P_A & \xrightarrow{\alpha_A X_0 \cdot (1 - P_x)} & P_A + 1 \\
P_A & \xrightarrow{\beta_A X_P \cdot (1 - P_A)} & P_A - 1 \\
X_A & \xrightarrow{\gamma_A V \cdot P_A} & X_A + 1 \\
X_A & \xrightarrow{\delta_A \cdot X} & X_A - 1 \\
P_P & \xrightarrow{f \cdot \alpha_P X_0 \cdot (1 - P_P)} & P_P + 1 \\
X_P & \xrightarrow{\gamma_P V \cdot P_P} & X_P + 1 \\
X_P & \xrightarrow{\delta_P \cdot X_P} & X_P - 1 \\
P_Y & \xrightarrow{\alpha_Y \cdot (X_A + X_P) \cdot (1 - P_Y)} & P_Y + 1 \\
Y & \xrightarrow{\gamma_Y V \cdot P_Y} & Y + 1 \\
Y & \xrightarrow{\delta_Y \cdot Y} & Y - 1
\end{array} \tag{41}$$

Here, we have assumed that all activating interactions, as well as the silencing of X_A by X_P follows first-order kinetics. We further define f to be a dimensionless parameter that scales the rate of P_P switching, that would depend on the strength of the enhancer controlling its activation. In this model, progenitors initially divide symmetrically, but switch to an asymmetrically dividing competent state that either specifies anterior cells, when X_A is on and exceeds a threshold concentration α_T , or posterior cells, when X_P is on p_T . Note that we have chosen parameters such that cells do not stably express both regulatory genes; however, in the event that there is transient expression of both genes, we take the cell to specify a posterior cell, are reflection of the posterior dominance in Hox gene expression. Finally, once Y exceeds a threshold concentration y_t, the cell ceases to divide, and dies at the end of its cell cycle lifetime.

EPIGENETIC SWITCHING NETWORK FOR CEREBRAL CORTEX DEVELOPMENT

Here, we describe a gene regulatory network model for cerebral cortex development. This network consists of a sequence of the regulatory genes $X_1 - X_4$, each with promoters $P_1 - P_4$. Each gene induces activates the promoter of the next gene in succession, and silences the promoter of the preceding gene. The last gene product, X_4, activates gene Y that drives cell cycle exit; additionally, a repressive chromatin regulator C is constitutively expressed and dampens the epigenetic activation of all regulators. The chemical reactions describing this network is given by:

$$
\begin{array}{lcl}
X_1 & \xrightarrow{\gamma_1 V\cdot P_1} & X_1+1 \\
X_1 & \xrightarrow{\delta_1\cdot X_1} & X_1-1 \\
P_2 & \xrightarrow{\alpha_2 X_1\cdot(1-P_2)/(C/V)} & P_2+1 \\
P_2 & \xrightarrow{\beta_2 X_3\cdot P_2} & P_2-1 \\
X_2 & \xrightarrow{\gamma_2 V\cdot P_2} & X_2+1 \\
X_2 & \xrightarrow{\delta_2\cdot X_2} & X_2-1 \\
P_3 & \xrightarrow{\alpha_3 X_2\cdot(1-P_3)/(C/V)} & P_3+1 \\
P_3 & \xrightarrow{\beta_3 X_4\cdot P_3} & P_3-1 \\
X_3 & \xrightarrow{\gamma_3 V\cdot P_3} & X_3+1 \\
X_3 & \xrightarrow{\delta_3\cdot X_3} & X_3-1 \\
P_4 & \xrightarrow{\alpha_4 X_3\cdot(1-P_4)/(C/V)} & P_4+1 \\
X_4 & \xrightarrow{\gamma_4 V\cdot P_4} & X_4+1 \\
X_4 & \xrightarrow{\delta_4\cdot X_4} & X_4-1 \\
P_Y & \xrightarrow{\alpha_Y X_4\cdot(1-P_Y)/(C/V)} & P_Y+1 \\
Y & \xrightarrow{\gamma_Y V\cdot P_Y} & Y+1 \\
Y & \xrightarrow{\delta_Y\cdot X_Y} & Y-1 \\
C & \xrightarrow{\gamma_C V 22} & C+1 \\
C & \xrightarrow{\delta_C C} & C-1
\end{array}
\tag{42}
$$

For simplicity, we have assumed that all regulatory interactions obey First-order kinetics, being far from saturation. Furthermore, we assume expression of the regulators $X_1 - X_4$ above concentration thresholds $x_{1T} - X_{4T}$ switch the cells into a competent state, where they produce neuronal cell types 1-4, following the asymmetric cell division scheme described above. Furthermore, the expression of Y above a threshold y_T then causes cell death at the end of the progenitor lifetime.

In simulations we varied the transcription rate of C, γ_C, to model the effects of the mutations that may change chromatin regulator expression or activity. As above, we obtained lineages by simulating single cells recursively, until there are no more proliferating cells, or until the time limit t_{max} is reached.

Let k_i be the effective activation time constant for the gene X_i. Assuming that the preceding activator X_{i-1} rapidly reaches steady state upon switching to an active epigenetic state, this time constant for activation is approximately:

$$k_2 = \frac{\alpha_2 \gamma_1}{\delta_1 c} \quad (43)$$

$$k_3 = \frac{\alpha_3 \gamma_1}{\delta_2 c} \quad (44)$$

$$k_4 = \frac{\alpha_4 \gamma_1}{\delta_3 c} \quad (45)$$

$$k_y = \frac{\alpha_y \gamma_1}{\delta_4 c} \quad (46)$$

Where $c = C/V$ is the steady-state concentration of chromatin regulator in the cell.

Furthermore, assuming that activation of the gene X_i causes rapid silencing of gene X_{i-1} through negative feedback, we get that the number of progeny of type i, N_i generated is approximated by:

$$N_1 = \frac{\tau_c \delta_1 c}{\alpha_2 \gamma_1} \qquad (47)$$

$$N_2 = \frac{\tau_c \delta_2 c}{\alpha_3 \gamma_2} \qquad (48)$$

$$N_3 = \frac{\tau_c \delta_3 c}{\alpha_4 \gamma_3} \qquad (49)$$

$$N_4 = \frac{\tau_c \delta_4 c}{\alpha_y \gamma_4} \qquad (50)$$

These expressions show that the population sizes of all differentiated cell type can be adjusted by changing chromatin regulator concentrations c. On the other hand, the fractions of different cell types are given by:

$$f_1 = \frac{(\delta_1/\alpha_2\gamma_1)}{(\delta_1/\alpha_2\gamma_1) + (\delta_2/\alpha_3\gamma_2) + (\delta_3/\alpha_4\gamma_3) + (\delta_4/\alpha_y\gamma_4)} \qquad (51)$$

$$f_2 = \frac{(\delta_2/\alpha_3\gamma_2)}{(\delta_1/\alpha_2\gamma_1) + (\delta_2/\alpha_3\gamma_2) + (\delta_3/\alpha_4\gamma_3) + (\delta_4/\alpha_y\gamma_4)} \qquad (52)$$

$$f_3 = \frac{(\delta_3/\alpha_4\gamma_3)}{(\delta_1/\alpha_2\gamma_1) + (\delta_2/\alpha_3\gamma_2) + (\delta_3/\alpha_4\gamma_3) + (\delta_4/\alpha_y\gamma_4)} \qquad (53)$$

$$f_4 = \frac{(\delta_4/\alpha_y\gamma_4)}{(\delta_1/\alpha_2\gamma_1) + (\delta_2/\alpha_3\gamma_2) + (\delta_3/\alpha_4\gamma_3) + (\delta_4/\alpha_y\gamma_4)} \qquad (54)$$

In contrast to total population sizes, population fractions are independent of chromatin regulator levels c, demonstrating that that population sizes can be scalably adjusted while maintain relative proportions. Interesting, these fractions are also independent of cell cycle time τ_c, suggesting that population sizes can be scalably adjusted by changing cell cycle duration. This is a point we will explore further below.

INDEPENDENT EPIGENETIC SWITCHING NETWORKS FOR TISSUE FORMATION

As a toy model for the independent formation of two tissue during embryogenesis, we consider two that regulatory gene networks evolving in time in two progenitor populations. These two populations of cells do not interact, but share a common cell cycle control machinery that dictates their rate of division.

Each network, a initial regulator (X_1 or X_3) activates the promoter of a second regulatory gene (X_2 or X_4), which represses the first gene's promoter through negative feedback, and activates the promoter a cell cycle exit regulator (Y or Z).

The reaction scheme for the first regulatory network is given by:

$$
\begin{array}{ccc}
P_1 & \xrightarrow{\beta_1 X_2 \cdot P_1} & P_1 - 1 \\
X_1 & \xrightarrow{\gamma_1 V \cdot P_1} & X_1 + 1 \\
X_1 & \xrightarrow{\delta_1 \cdot X_1} & X_1 - 1 \\
P_2 & \xrightarrow{\alpha_2 X_1 \cdot (1 - P_2)} & P_2 + 1 \\
X_2 & \xrightarrow{\gamma_2 V \cdot P_2} & X_2 + 1 \\
X_2 & \xrightarrow{\delta_2 \cdot X_2} & X_2 - 1 \\
P_Y & \xrightarrow{\alpha_Y X_2 \cdot (1 - P_Y)} & P_Y + 1 \\
Y & \xrightarrow{\gamma_Y V \cdot P_Y} & Y + 1 \\
Y & \xrightarrow{\delta_Y \cdot Y} & Y - 1
\end{array}
\tag{55}
$$

whereas the chemical reactions for the second regulatory network is:

$$
\begin{array}{ccc}
P_3 & \xrightarrow{\beta_3 X_4 \cdot P_3} & P_3 - 1 \\
X_3 & \xrightarrow{\gamma_3 V \cdot P_3} & X_3 + 1 \\
X_3 & \xrightarrow{\delta_3 \cdot X_3} & X_3 - 1 \\
P_4 & \xrightarrow{\alpha_4 X_1 \cdot (1-P_4)} & P_4 + 1 \\
X_4 & \xrightarrow{\gamma_4 V \cdot P_4} & X_4 + 1 \\
X_4 & \xrightarrow{\delta_4 \cdot X_4} & X_4 - 1 \\
P_Z & \xrightarrow{\alpha_Z X_4 \cdot (1-P_Z)} & P_Z + 1 \\
Z & \xrightarrow{\gamma_Z V \cdot P_Z} & Z + 1 \\
Z & \xrightarrow{\delta_Z \cdot Z} & Z - 1
\end{array}
\tag{56}
$$

In both networks and cell populations, progenitors exist in a competent state where they generate cell-cycle arrested progeny through asymmetric cell division; the type of progeny generated is determined by the expression of the regulatory gene above a certain threshold (x_{1T} , x_{2T} , x_{3T} and x_{4T}). In all competent states, the cell cycle lifetime is given by t_{mc}. We note that this lifetime is determined by activity of cell growth pathways such as the IGF pathway. To model the effects mutations altering the activity of these pathways, we will concurrently vary t_{mc} in both gene regulatory networks in our simulations.

As above, we perform simulations recursively in single cells to obtain complete lineages. The two regulatory gene networks are simulated independently, and the time traces and cell numbers obtained are then combined for analysis.

Model		Description	Value
Fig.2.2(III)	γ_X	synthesis rate of X	0.5/hr
	δ_X	degradation rate of X	0/hr
	γ_Y	maximal synthesis rate of Y	10^3 copies/hr
	n	Hill coefficient for Y synthesis	10
	K	level of x for half-maximal Y synthesis	10^3 copies
	δ_Y	degradation rate of Y	1/hr
Fig. 2.2(IV)	γ_0	signal-dependent synthesis rate of X	5 copies/hr
	γ_1	positive feedback-dependent synthesis rate of X	70 copies/hr
	m	Hill coefficient for positive feedback of X	15
	K	level of x for half-maximal synthesis of X	100 copies
	δ_X	degradation rate of X	0/hr
	γ_Y	maximal synthesis rate of Y	10^3 copies/hr
	n	Hill coefficient for Y synthesis	10
	L	level of x for half-maximal Y synthesis	10^3 copies
	δ_Y	degradation rate of Y	1/hr
Fig. 2.2(V)	γ_X	maximal synthesis rate of X	100 copies/hr
	K	level of S for half-maximal inhibition of synthesis of X	5 copies
	δ_X	degradation rate of X	0/hr
	γ_Y	maximal synthesis rate of Y	10^3 copies/hr
	n	Hill coefficient for Y synthesis	10
	L	level of x for half-maximal Y synthesis	40 copies
	δ_Y	degradation rate of Y	1/hr

Table 2.2: Model Parameters (I) cont.

Model		Description	Value
Fig. 2.3.A	α_X	epigenetic activation rate constant of X promoter	0.02/hr
	β_X	epigenetic silencing rtae constant of X	20/copies/hr
	K	level of y for half-maximal X promoter silencing	30 copies
	γ_X	synthesis rate of X	1300 copies/hr
	δ_X	degradation rate of X	1/hr
	L	level of x for half-maximal promoter Y activation	30 copies
	γ_Y	synthesis rate of Y	10^3/hr
	δ_Y	degradation rate of Y	1/hr
Fig. 2.3B	α_X	maximal epigenetic activation rate constant of X promoter	0.02/hr
	K	level of y for half-maximal inhibition of epigenetic activation of X	20 copies
	γ_X	synthesis rate of Y	300 copies/hr
	δ_X	degradation rate of X	1/hr
	L	level of x for half-maximal inhibition of the promoter Y activation	20
	γ_Y	synthesis rate of Y	10^3 copies/hr
	δ_Y	degradation rate of Y	1/hr

Table 2.3: Model Parameters (II-III)

Model		Description	Value
Fig.2.4A(II) Fig.2.4B(I)	α_X	epigenetic activation rate constant of X promoter	0.01/hr
	γ_X	synthesis rate of X	10^3 copies/hr
	δ_X	degradation rate of X	1/hr
	γ_Y	maximal synthesis rate of Y	10^3 copies/hr
	n	Hill coefficient for synthesis of Y	10
	K	level of x for half-maximal epigenetic activation of Y	200 copies
	δ_Y	degradation rate of Y	1/hr
Fig.2.4B(II)	α_X	epigenetic activation rate constant of X promoter	0.008/hr
	γ_X	synthesis rate of X	10^3 copies/hr
	δ_X	degradation rate of X	1/hr
	γ_Y	maximal synthesis rate of Y	10^3 copies/hr
	n	Hill coefficient for synthesis of Y	10
	K	level of x for half-maximal epigenetic activation of Y	200 copies
	δ_Y	degradation rate of Y	1/hr
	α_Z	epigenetic activation rate constant of Z promoter	0.03/hr
	γ_Z	synthesis rate of Z	500 copies/hr
	δ_Z	degradation rate of Z	1/hr

Table 2.4: Model Parameters (IV)

Model		Description	Value
Fig. 2.5	γ_0	synthesis rate of X_0	10^3 copies/hr
	δ_0	degradation rate of X_0	1/hr
	α_A	epigenetic activation of promoter P_A	9×10^{-5}/hr
	β_A	epigenetic silencing of promoter P_A	0.1/hr
	γ_A	maximal synthesis rate of X_A	10^3 copies/hr
	δ_A	degradation rate of X_A	1/hr
	α_P	epigenetic activation rate of promoter P_P	1.2×10^{-4}/hr
	γ_P	synthesis rate of X_P	850 copies/hr
	δ_P	degradation rate of X_P	1/hr
	α_Y	epigenetic activation of promoter P_Y	$6 \times 10^{-6}/copy/hr$
	γ_Y	synthesis rate of Y	10^3 copies/hr
	δ_Y	degradation rate of Y	1/hr

Table 2.5: Model Parameters (V)

Model		Description	Value
Fig. 2.6	β_1	epigenetic silencing rate of promoter P_1	100/copy/hr
	γ_1	synthesis rate of X_1	10^3 copies/hr
	δ_1	degradation rate of X_1	1/hr
	α_2	epigenetic activation rate of promoter P_2	0.01/hr
	β_2	epigenetic silencing rate of promoter P_2	100/copy/hr
	γ_2	synthesis rate of X_2	900 copies/hr
	δ_2	degradation rate of X_2	1/hr
	α_3	epigenetic activation rate of promoter P_3	0.01/hr
	β_3	epigenetic silencing rate of promoter P_3	100/copy/hr
	γ_3	synthesis rate of X_3	700 copies/hr
	δ_3	degradation rate of X_3	1/hr
	α_4	epigenetic activation rate of promoter P_4	0.01/hr
	γ_4	synthesis rate of X_4	10^3 copies/hr
	δ_4	degradation rate of X_4	1/hr
	α_Y	epigenetic activation rate of promoter P_Y	0.01/hr
	γ_Y	synthesis rate of Y	10^3 copies/hr
	δ_Y	degradation rate of Y	1/hr
	γ_C	synthesis rate of C	10^3 copies/hr
	δ_C	degradation rate of C	1/hr

Table 2.6: Model Parameters (VI)

Model		Description	Value
VII. 11	β_1	epigenetic silencing rate of promoter P_1	100/copy/hr
	γ_1	synthesis rate of X_1	10^3 copies/hr
	δ_1	degradation rate of X_1	1/hr
	α_2	epigenetic activation rate of promoter P_2	4×10^{-6} copies/hr
	β_2	epigenetic silencing rate of promoter P_2	100/copies/hr
	γ_2	synthesis rate of X_2	900 copies/hr
	δ_2	degradation rate of X_2	1/hr
	α_Y	epigenetic activation rate of promoter P_Y	$4 \times 10^{-6}/copy/hr$
	γ_Y	synthesis rate of Y	10^3 copies/hr
	δ_Y	degradation rate of Y	1/hr
	β_3	epigenetic silencing rate of promoter P_3	100/copy/hr
	γ_3	synthesis rate of X_3	10^3 copies/hr
	δ_3	degradation rate of X_3	1/hr
	α_4	epigenetic activation rate of promoter P_4	2×10^{-6}/copy/hr
	γ_4	synthesis rate of X_4	10^3 copies/hr
	δ_4	degradation rate of X_4	1/hr
	α_Z	epigenetic activation rate of promoter P_Z	2×10^{-6}/copy/hr
	γ_Z	synthesis rate of Z 35	10^3 copies/hr
	δ_Z	degradation rate of Z	1/hr

Table 2.7: Model Parameters (VII)

2.11 DISCUSSION

In this study, we sought to understand how progenitors can maintain autonomous timetables for cell lineage specification that robustly unfold over long developmental timescales, but are yet flexibly adjustable by evolution. We found that epigenetic switching networks, in contrast to classical networks governed by the bacterial quantitative paradigm, indeed maintains robust timetables for gene activation and lineage specification. When used to control the transitions between different competent cell states for differentiation, epigenetic switching networks could generate multiple output cell types in defined numbers and proportions. Through analyzing a series of case studies, we further show how evolution can act on epigenetic switching networks to alter developmental timetables and, consequently, organism size and proportion. Our results establish a quantitative paradigm for understanding how embryos can set and flexibly adjust developmental timetables for controlling organism size and proportions.

Epigenetic switching networks, in contrast to classical gene regulatory networks, can uniquely generate long response delays that are tunable by signaling inputs. While classical gene regulatory networks can also generate multi-generational response delays, these delays could only be tuned by signals over a narrow range. Furthermore, this tunable signal range was close to the lower signal threshold below which progenitors fail to respond, making them susceptible to stalling in a proliferative state. Thus, classically regulatory networks, while ideal for functional processes that require rapid responses, may be less suitable for developmental and physiological processes requiring protracted temporal control over many cell generations.

Epigenetic switching networks can generate multiple cell types in defined sizes and proportions, provided that they control the transitions of progenitors between competent states that generate differentiated progeny through either asymmetric cell division (where fate determinants are unevenly partitioned during cell division) or through stochastic differentiation control (Figure 2.4). In either case, cell populations grow linearly with time, allowing their sizes to be capped by stochastic timing delays that arise through epigenetic switching. There is evidence for both asymmetric and stochastic mechanisms in specific vertebrate developmental systems (Knoblich, 2010b; Fish et al., 2008; Simons and Clevers, 2011b; Klein and Simons, 2011); however, it will be useful to analyze developmental processes at the single lineage level in order to ascertain whether these competent states occur generally in the embryo. Emerging techniques for lineage reconstruction at the level of whole organisms may allow us to perform this analysis in a systematic manner (McKenna et al., 2016).

Our studies in cerebral cortex development implicate epigenetic-modifying enzymes as master regulators of developmental speed (Figure 2.6). Because of their broad sequence specificity, chromatin regulators likely bind and act on multiple genes within epigenetic switching networks (Boyer et al., 2006; Lee et al., 2006); consequently, changes in their levels or activity would alter network dynamics at a global level, and could potentially change the speed at which their temporal schedules unfold. Consistent with this idea, disruptions in polycomb complex activity accelerates differentiation across multiple cell lineages in different contexts (Xiao et al., 2018; Ezhkova et al., 2009; Mirzamohammadi et al., 2016; Endoh et al., 2017; Fujimura et al., 2018; Zhang et al., 2015). Deletion of the PRC2 methyltransferase subunit Ezh2 accelerates the temporal schedule for cerebral cortex development, leading to reduced cortical tissue size while preserving the temporal order of neuronal subtype differentiation (Pereira et al., 2010). Evolutionary changes in polycomb activity could stem from non-coding mutations a affecting expression levels of polycomb components; however, because the PRC1 and PRC2 families have both undergone expansion during vertebrate evolution (Sowpati et al., 2015), it is also possible that such speed changes may also involve mutations that alter protein function. We point out that while H3K27me3 modifications modulate activation timing in our system, our modeling and experimental results indicate that they must collaborate with other chromatin modifications to set activation timing (Pease et al., in submission). Thus, multiple chromatin systems may ultimately work together to dictate the pace of development. The repressive mark H3K9me3 also regulates lineage specification (Nicetto et al., 2019), and may co-exist with H3K27me3 to establish repression at developmental genes (Boros et al., 2014; de la Cruz et al., 2007; Yamamoto et al., 2004). Another intriguing chromatin-regulating factor is HMGA2, a DNA-binding protein that controls body size and developmental timing in vertebrates (Chung et al., 2018; Lamichhaney et al., 2016). HMGA proteins also have broad genome-specificity (Ozturk et al., 2014) and could operate on many loci simultaneously to control developmental speed.

Temporal control mechanisms, as implemented by epigenetic switching networks, must ultimately work together with spatial patterning mechanisms to ensure proper development. It is still largely unclear how spatial and temporal control mechanisms work together to determine tissue and organ sizes; however, based on existing work, we suggest two possible strategies: Firstly, temporal control mechanisms could set the size of domains on which spatial mechanisms act. For instance, some morphogen gradients can scale with the dimensions of a domain (Ben-Zvi et al., 2011; Inomata, 2017; Rogers and Schier, 2011), though these gradients do not control the domain size itself (Averbukh et al., 2014; Fried and Iber, 2014). Temporal mechanisms, with their ability to precisely specify output cell numbers, could work in conjunction with spatially scalable patterning mechanisms to ensure proportionality in both size and form. Secondly, temporal control mechanisms could generate defined cell populations that then self organize in space. For example, during spinal cord development, an epigenetic switching network of mutually-inhibitory genes is thought to drive motor neuron specification in the spinal motor column (Johnston and Desplan, 2010). While these different motor neuron types are initially randomly distributed in space, they express unique cadherin combinations (Price et al., 2002) that allow them to cluster together and generate defined structures. In general, future studies in developmental biology will benefit from closer consideration of how spatial and temporal control mechanisms work together to specify growth and form during multicellular development.

2.12 REFERENCE

Ackers, G.K., Johnson, A.D., Shea, M.A., 1982. Quantitative model for gene regulation by lambda phage repressor. Proceedings of the National Academy of Sciences of the United States of America 79, 1129-1133.

Alberch, P., Gould, S.J., Oster, G.F., Wake, D.B., 1979. Size and Shape in Ontogeny and Phylogeny. Paleobiology 5, 296-317.

Alon, U., 2007. Network motifs: theory and experimental approaches. Nature Reviews Genetics 8, 450-461.

van den Ameele, J., Tiberi, L., Vanderhaeghen, P., Espuny-Camacho, I., 2014. Thinking out of the dish: what to learn about cortical development using pluripotent stem cells. Trends in Neurosciences 37, 334-342.

Baker, J., Liu, J.P., Robertson, E.J., Efstratiadis, A., 1993. Role of insulin-like growth factors in embryonic and postnatal growth. Cell 75, 73-82.

Barry, C., Schmitz, M.T., Jiang, P., Schwartz, M.P., Dun, B.M., Swanson,S., Bacher, R., Bolin, J.M., Elwell, A.L., McIntosh, B.E., Stewart, R., Thomson, J.A., 2017a. Species-specific developmental timing is maintained by pluripotent stem cells ex utero. Developmental Biology 423, 101-110.

Barry, C., Schmitz, M.T., Jiang, P., Schwartz, M.P., Dun, B.M., Swanson, S., Bacher, R., Bolin, J.M., Elwell, A.L., McIntosh, B.E., Stewart, R., Thomson, J.A., 2017b. Species-specific developmental timing is maintained by pluripotent stem cells ex utero. Developmental Biology 423, 101-110.

Bintu, L., Buchler, N.E., Garcia, H.G., Gerland, U., Hwa, T., Kondev, J., Phillips, R., 2005. Transcriptional regulation by the numbers: models. Current Opinion in Genetics & Development 15, 116-124.

Bintu, L., Yong, J., Antebi, Y.E., McCue, K., Kazuki, Y., Uno, N., Oshimura, M., Elowitz, M.B., 2016. Dynamics of epigenetic regulation at the single-cell level 351, 720-724. Publisher: American Association for the Advancement of Science Section: Report.

Bolouri, H., Davidson, E.H., 2002. Modeling transcriptional regulatory networks. BioEssays 24, 1118-1129. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/bies.10189.

Bonner, J.T., 1965. Size and Cycle: An Essay on the Structure of Biology. Princeton University Press, Place of publication not identified.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., Bell, G.W., Otte, A.P., Vidal, M., Giford, D.K., Young, R.A., Jaenisch, R., . Polycomb complexes repress developmental regulators in murine embryonic stem cells 441, 349-353.

Briscoe, J., Small, S., 2015. Morphogen rules: design principles of gradient-mediated embryo patterning. Development 142, 3996-4009.

Burton, P., Ra, M., Kerr, P., Yacoub, M., Barton, P., 1999. An Intrinsic Timer That Controls Cell-Cycle Withdrawal in Cultured Cardiac Myocytes. Developmental Biology 216, 659-670.

Calder, W.A., 1984. Size, Function, and Life History. Courier Corporation. Carroll, S.B., 2008. Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution. Cell 134, 25-36.

Chen, W.S., Xu, P.Z., Gottlob, K., Chen, M.L., Sokol, K., Shiyanova, T., Roninson, I., Weng, W., Suzuki, R., Tobe, K., Kadowaki, T., Hay, N., 2001. Growth retardation and increased apoptosis in mice with homozygous disruption of the akt1 gene. Genes & Development 15, 2203-2208.

Choi, K., Medley, J.K., Konig, M., Stocking, K., Smith, L., Gu, S., Sauro, H.M., 2018. Tellurium: An Extensible Python-based Modeling Environment for Systems and Synthetic Biology. Bio Systems 171, 74-79.
Gillespie, D.T., 1977. Exact stochastic simulation of coupled chemical reactions. The Journal of Physical Chemistry 81, 2340-2361.

Crescenzi, M., Fleming, T.P., Lassar, A.B., Weintraub, H., Aaronson, S.A., 1990. MyoD induces growth arrest independent of differentiation in normal and transformed cells. Proceedings of the National Academy of Sciences 87, 8442-8446. Publisher: National Academy of Sciences Section: Research Article.

Davidson, E.H., 2010. Emerging properties of animal gene regulatory networks. Nature 468, 911-920.

De Beer, G., 1940. Embryos and ancestors. Clarendon Press. Ebisuya, M., Briscoe, J., 2018. What does time mean in development? Development 145, dev164368.

Eiraku, M., Watanabe, K., Matsuo-Takasaki, M., Kawada, M., Yonemura, S., Matsumura, M., Wataya, T., Nishiyama, A., Muguruma, K., Sasai, Y., 2008. Self-organized formation of polarized cortical tissues from ESCs and its active manipulation by extrinsic signals. Cell Stem Cell 3, 519-532.

Endoh, M., Endo, T.A., Shinga, J., Hayashi, K., Farcas, A., Ma, K.W., Ito, S., Sharif, J., Endoh, T., Onaga, N., Nakayama, M., Ishikura, T., Masui, O., Kessler, B.M., Suda, T., Ohara, O., Okuda, A., Klose, R., Koseki, H., 2017. PCGF6-PRC1 suppresses premature differentiation of mouse embryonic stem cells by regulating germ cell-related genes. eLife 6,
e21064. Publisher: eLife Sciences Publications, Ltd.

Eskeland, R., Leeb, M., Grimes, G.R., Kress, C., Boyle, S., Sproul, D., Gilbert, N., Fan, Y., Skoultchi, A.I., Wutz, A., Bickmore, W.A., 2010. Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. Molecular Cell 38, 452-464.

Espuny-Camacho, I., Michelsen, K.A., Gall, D., Linaro, D., Hasche, A., Bonnefont, J., Bali, C., Orduz, D., Bilheu, A., Herpoel, A., Lambert, N., Gaspard, N., Peron, S., Schi mann, S.N., Giugliano, M., Gaillard, A., Vanderhaeghen, P., 2013. Pyramidal neurons derived from human pluripotent stem cells integrate efficiently into mouse brain circuits in vivo. Neuron 77, 440-456.

Ezhkova, E., Pasolli, H.A., Parker, J.S., Stokes, N., Su, I.h., Hannon, G., Tarakhovsky, A., Fuchs, E., 2009. Ezh2 Orchestrates Gene Expression for the Stepwise Differentiation of Tissue-Specific Stem Cells. Cell 136, 1122-1135.

Fabre, P.J., Benke, A., Manley, S., Duboule, D., 2015. Visualizing the HoxD Gene Cluster at the Nanoscale Level. Cold Spring Harbor Symposia on Quantitative Biology 80, 9-16.

Fabre, P.J., Leleu, M., Mascrez, B., Lo Giudice, Q., Cobb, J., Duboule, D., 2018. Heterogeneous combinatorial expression of Hoxd genes in single cells during limb development. BMC biology 16, 101.

Ferrell, J.E., 2016. Perfect and Near-Perfect Adaptation in Cell Signaling. Cell Systems 2, 62-67.

Fish, J.L., Dehay, C., Kennedy, H., Huttner, W.B., 2008. Making bigger brains-the evolution of neural-progenitor-cell division. Journal of Cell Science 21, 2783-2793. Publisher: The Company of Biologists Ltd Section: Commentary.

Fujimura, N., Kuzelova, A., Ebert, A., Strnad, H., Lachova, J., Machon, O., Busslinger, M., Kozmik, Z., 2018. Polycomb repression complex 2 is required for the maintenance of retinal progenitor cells and balanced retinal differentiation. Developmental Biology 433, 47-60.

Gagliardi, A.D., Kuo, E.Y.W., Raulic, S., Wagner, G.F., DiMattia, G.E., 2005. Human stanniocalcin-2 exhibits potent growth-suppressive properties in transgenic mice independently of growth hormone and IGFs. American Journal of Physiology-Endocrinology and Metabolism 288, E92-E105.

Gao, F.B., Durand, B., Ra, M., 1997. Oligodendrocyte precursor cells count time but not cell divisions before differentiation. Current Biology 7, 152-155.

Garfunkel, B.P., Arad, S., Le, P.T., Bustin, M., Rosen, C.J., Gabet, Y., Orly, J., 2015. Proportionate Dwarfism in Mice Lacking Heterochromatin Protein 1 Binding Protein 3 (HP1BP3) Is Associated With Alterations in the Endocrine IGF-1 Pathway. Endocrinology 156, 4558-4570. Publisher: Oxford Academic.

Gaspard, N., Bouschet, T., Hourez, R., Dimidschstein, J., Naeije, G., van den Ameele, J., Espuny-Camacho, I., Herpoel, A., Passante, L., Schimann, S.N., Gaillard, A., Vanderhaeghen, P., 2008. An intrinsic mechanism of corticogenesis from embryonic stem cells. Nature 455, 351-357.

Georgescu, C., Longabaugh, W.J.R., Scripture-Adams, D.D., David-Fung, E.S., Yui, M.A., Zarnegar, M.A., Bolouri, H., Rothenberg, E.V., 2008. A gene regulatory network armature for T lymphocyte specification. Proceedings of the National Academy of Sciences of the United States of America 105, 20100-20105.

Gillespie, D.T., 1977. Exact stochastic simulation of coupled chemical reactions. The Journal of Physical Chemistry 81, 2340-2361.

Gould, S.J., 1977. Ontogeny and Phylogeny. Harvard University Press, Cambridge, Mass. reprint edition.

Gregor, T., Bialek, W., van Steveninck, R.R.d.R., Tank, D.W., Wieschaus, E.F., 2005. Diffusion and scaling during early embryonic pattern formation. Proceedings of the National Academy of Sciences 102, 18403-18407.

Gerard, M., Zakany, J., Duboule, D., 1997. Interspecies exchange of a Hoxd enhancer in vivo induces premature transcription and anterior shift of the sacrum. Developmental Biology 190, 32-40.

Haeckel, E., 1866. Generelle Morphologie der Organismen. Georg Reimer, Berlin. Heinzel, S., Binh Giang, T., Kan, A., Marchingo, J.M., Lye, B.K., Corcoran, L.M., Hodgkin, P.D., 2017. A Myc-dependent division timer complements a cell-death timer to regulate T cell and B cell responses. Nature Immunology 18, 96-103.

Hofmann, J., Zhao, X., De Cecco, M., Peterson, A., Pagliaroli, L., Manivannan, J., Hubbard, G., Ikeno, Y., Zhang, Y., Feng, B., Li, X., Serre, T., Qi, W., Van Remmen, H., Miller, R., Bath, K., de Cabo, R., Xu, H., Neretti, N., Sedivy, J., 2015. Reduced Expression of MYC Increases Longevity and Enhances Healthspan. Cell 160, 477-488.

Huang, S., Guo, Y.P., May, G., Enver, T., 2007. Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. Developmental Biology 305, 695-713.

Huxley, J.S., Pigliucci, M., Muller, G.B., 1942. Evolution: The Modern Synthesis. Allen and Unwin, London. Johnston, R.J., Desplan, C., 2010. Stochastic Mechanisms of Cell Fate Specification that Yield Random or Robust Outcomes. Annual Review of Cell and Developmental Biology 26, 689-719.

Juan, A.H., Ruddle, F.H., 2003. Enhancer timing of Hox gene expression: deletion of the endogenous Hoxc8 early enhancer. Development (Cambridge, England) 130, 4823-4834.

Keyte, A.L., Smith, K.K., 2014. Heterochrony and developmental timing mechanisms: changing ontogenies in evolution. Seminars in Cell & Developmental Biology 34, 99-107.

Klein, A.M., Simons, B.D., 2011. Universal patterns of stem cell fate in cycling adult tissues. Development (Cambridge, England) 138, 3103-3111.

Knoblich, J.A., 2010a. Asymmetric cell division: recent developments and their implications for tumour biology. Nature Reviews Molecular Cell Biology 11, 849-860. Number: 12 Publisher: Nature Publishing Group.

Knoblich, J.A., 2010b. Asymmetric cell division: recent developments and their implications for tumour biology. Nature Reviews. Molecular Cell Biology 11, 849-860.

Kohwi, M., Doe, C.Q., 2013. Temporal Fate Specification and Neural Progenitor Competence During Development. Nature reviews. Neuroscience 14, 823-838.

Kueh, H.Y., Champhekar, A., Champhekhar, A., Nutt, S.L., Elowitz, M.B., Rothenberg, E.V., 2013. Positive feedback between PU.1 and the cell cycle controls myeloid differentiation. Science (New York, N.Y.) 341, 670-673.

Kueh, H.Y., Yui, M.A., Ng, K.K.H., Pease, S.S., Zhang, J.A., Damle, S.S., Freedman, G., Siu, S., Bernstein, I.D., Elowitz, M.B., Rothenberg, E.V., 2016. Asynchronous combinatorial action of four regulatory factors activates Bcl11b for T cell commitment. Nature Immunology 17, 956-965.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K.i., Koseki, H., Fuchikami, T., Abe, K., Murray, H.L., Zucker, J.P., Yuan, B., Bell, G.W., Herbolsheimer, E., Hannett, N.M., Sun, K., Odom, D.T., Otte, A.P., Volkert, T.L., Bartel, D.P., Melton, D.A., Giford, D.K., Jaenisch, R., Young, R.A., 2006. Control of Developmental Regulators by Polycomb in Human Embryonic Stem Cells. Cell 125, 300-313.

Ben-Tabou de Leon, S., Davidson, E.H., 2009. Modeling the dynamics of transcriptional gene regulatory networks for animal development. Developmental Biology 325, 317-328.

Levine, J.H., Elowitz, M.B., 2014. Polyphasic feedback enables tunable cellular timers. Current biology: CB 24, R994-995.

Li, X., Chen, Z., Desplan, C., 2013. Temporal patterning of neural progenitors in Drosophila. Current Topics in Developmental Biology 105, 69-96.

Ma, W., Trusina, A., El-Samad, H., Lim, W.A., Tang, C., 2009. Defining network topologies that can achieve biochemical adaptation. Cell 138, 760-773.

Mallo, M., Alonso, C.R., 2013. The regulation of Hox gene expression during animal development. Development 140, 3951-3963.

McKenna, A., Findlay, G., Gagnon, J., Horwitz, M., Schier, A., Shendure, J., 2016. Whole-organism Lineage Tracing by Combinatorial and Cumulative Genome Editing.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U., 2002. Network Motifs: Simple Building Blocks of Complex Networks. Science 298, 824-827.

Mirzamohammadi, F., Papaioannou, G., Inloes, J.B., Rankin, E.B., Xie, H., Schipani, E., Orkin, S.H., Kobayashi, T., 2016. Polycomb repressive complex 2 regulates skeletal growth by suppressing Wnt and TGF- signalling. Nature Communications 7, 12047. Number: 1 Publisher: Nature Publishing Group.

Mugnier, A., Mila, H., Guiraud, F., Brevaux, J., Lecarpentier, M., Martinez, C., Mariani, C., Adib-Lesaux, A., Chastant-Maillard, S., Saegerman, C., Grellet, A., 2019. Birth weight as a risk factor for neonatal mortality: Breed-specific approach to identify at-risk puppies. Preventive Veterinary Medicine 171, 104746.

Ng, K.K., Yui, M.A., Mehta, A., Siu, S., Irwin, B., Pease, S., Hirose, S., Elowitz, M.B., Rothenberg, E.V., Kueh, H.Y., 2018. A stochastic epigenetic switch controls the dynamics of T-cell lineage commitment.

Nicetto, D., Donahue, G., Jain, T., Peng, T., Sidoli, S., Sheng, L., Montavon, T., Becker, J.S., Grindheim, J.M., Blahnik, K., Garcia, B.A., Tan, K., Bonasio, R., Jenuwein, T., Zaret, K.S., 2019. H3K9me3-heterochromatin loss at protein-coding genes enables developmental lineage specification, 5.

Okkens, A., Teunissen, J., Van Osch, W., Van Den Brom, W., Dieleman, S., Kooistra, H., Influence of litter size and breed on the duration of gestation in dogs , 57:193-197.

Otani, T., Marchetto, M.C., Gage, F.H., Simons, B.D., Livesey, F.J., 2016. 2D and 3D Stem Cell Models of Primate Cortical Development Identify Species-Specific Differences in Progenitor Behavior Contributing to Brain Size. Cell Stem Cell 18, 467-480.

Padovan-Merhar, O., Nair, G.P., Biaesch, A.G., Mayer, A., Scarfone, S., Foley, S.W., Wu, A.R., Churchman, L.S., Singh, A., Raj, A., 2015. Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms. Molecular Cell 58, 339-352.

Park, I.k., Qian, D., Kiel, M., Becker, M.W., Pihalja, M., Weissman, I.L., Morrison, S.J., Clarke, M.F., 2003. Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells. Nature 423, 302-305.

Pease, N.A., Nguyen, P., Ng, K.K., Woodworth, M., Irwin, B., Vaughan, J., Kueh, H.Y., in submission. Tunable, division-independent control of developmental timing by an epigenetic switch.

Peng, X.D., Xu, P.Z., Chen, M.L., Hahn-Windgassen, A., Skeen, J., Jacobs, J., Sundararajan, D., Chen, W.S., Crawford, S.E., Coleman, K.G., Hay, N., 2003. Dwarfism, impaired skin development, skeletal muscle atrophy, delayed bone development, and impeded adipogenesis in mice lacking Akt1and Akt2. Genes & Development 17, 1352-1365.

Pereira, J.D., Sansom, S.N., Smith, J., Dobenecker, M.W., Tarakhovsky, A., Livesey, F.J., 2010. Ezh2, the histone methyltransferase of PRC2, regulates the balance between self-renewal and differentiation in the cerebral cortex. Proceedings of the National Academy of Sciences of the United States of America 107, 15957-15962.

Price, S.R., Garcia, N.V.D.M., Ranscht, B., Jessell, T.M., 2002. Regulation of Motor Neuron Pool Sorting by Differential Expression of Type II Cadherins. Cell 109, 205-216. Publisher: Elsevier.

Rimbault, M., Beale, H.C., Schoenebeck, J.J., Hoopes, B.C., Allen, J.J., Kilroy-Glynn, P., Wayne, R.K., Sutter, N.B., Ostrander, E.A., 2013. Derived variants at six genes explain nearly half of size reduction in dog breeds. Genome Research 23, 1985-1995.

Rosello-Diez, A., Arques, C.G., Delgado, I., Giovinazzo, G., Torres, M., 2014.Diffusible signals and epigenetic timing cooperate in late proximo-distal limb patterning. Development 141, 1534-1543.

Rossi, A.M., Fernandes, V.M., Desplan, C., 2017. Timing temporal transitions during brain development. Current Opinion in Neurobiology 42, 84-92.

Rothenberg, E.V., 2019. Programming for T-lymphocyte fates: modularity and mechanisms. Genes & Development 33, 1117-1135.

Saiz-Lopez, P., Chinnaiya, K., Campa, V.M., Delgado, I., Ros, M.A., Towers, M., 2015. An intrinsic timer species distal structures of the vertebrate limb. Nature Communications

6.

Simons, B.D., Clevers, H., 2011a. Strategies for Homeostatic Stem Cell Self-Renewal in Adult Tissues. Cell 145, 851-862. Publisher: Elsevier.

Simons, B.D., Clevers, H., 2011b. Strategies for homeostatic stem cell self-renewal in adult tissues. Cell 145, 851-862.

Singh, H., Khan, A.A., Dinner, A.R., 2014. Gene regulatory networks in the immune system. Trends in Immunology 35, 211-218.

Sorrentino, V., Pepperkok, R., Davis, R.L., Ansorge, W., Philipson, L., 1990. Cell proliferation inhibited by MyoD1 independently of myogenic differentiation. Nature 345, 813-815.

Soshnikova, N., Duboule, D., 2009. Epigenetic Temporal Control of Mouse Hox Genes in Vivo. Science 324, 1320-1323.

Sowpati, D.T., Ramamoorthy, S., Mishra, R.K., 2015. Expansion of the polycomb system and evolution of complexity. Mechanisms of Development 138, 97-112.

Spencer, S.L., Cappell, S.D., Tsai, F.C., Overton, K.W., Wang, C.L., Meyer, T., 2013. The proliferation-quiescence decision is controlled by a bifurcation in CDK2 activity at mitotic exit. Cell 155, 369-383.

Trumpp, A., Refaeli, Y., Oskarsson, T., Gasser, S., Murphy, M., Martin, G.R., Bishop, J.M., 2001. c-Myc regulates mammalian body size by controlling cell number but not cell size. Nature 414, 768.

Venkei, Z.G., Yamashita, Y.M., 2018. Emerging mechanisms of asymmetric stem cell division. Journal of Cell Biology 217, 3785-3795. Publisher: The Rockefeller University Press.

Verhaeghe, J., Van Bree, R., Van Herck, E., Laureys, J., Bouillon, R., Van Assche, F.A., 1993. C-peptide, insulin-like growth factors I and II, and insulin-like growth factor binding protein-1 in umbilical cord serum: Correlations with birth weight. American Journal of Obstetrics and Gynecology 169, 89-97.

Wakamatsu, Y., Suzuki, K., 2019. Sequence alteration in the enhancer contributes to the heterochronic Sox9 expression in marsupial cranial neural crest. Developmental Biology 456, 31-39.

Xiao, S., Zhang, W., Manley, N.R., 2018. Thymic B cell development is controlled by the B potential of progenitors via both hematopoietic intrinsic and thymic microenvironment-intrinsic regulatory mechanisms. PLOS ONE 13, e0193189.

Zhang, J., Taylor, R.J., La Torre, A., Wilken, M.S., Cox, K.E., Reh, T.A., Vetter, M.L., 2015. Ezh2 maintains retinal progenitor proliferation, transcriptional integrity, and the timing of late differentiation. Developmental Biology 403, 128-138.

Zhu, J., 2018. T Helper Cell Differentiation, Heterogeneity, and Plasticity. Cold Spring Harbor perspectives in biology 10.

Zakany, J., Gerard, M., Favier, B., Duboule, D., 1997. Deletion of a HoxD enhancer induce transcriptional heterochrony leading to transposition of the sacrum. The EMBO journal 16, 4393-4402.

CHAPTER III. AN UNSUPERVISED, DEEP-LEARNING WORKFLOW FOR EXPLORATION OF CELLULAR PHENOTYPIC STATES AND THEIR DYNAMICS FROM LIGHT MICROSCOPY MOVIES

3.1 INTRODUCTION

Cells maintain and switch between distinct phenotypic states in a dynamic manner. The facile identification of these states and understanding the basis for and dynamics by which they interconvert is a central challenge in biology. Modern single-cell analysis methods, such as single cell RNA sequencing and multiparameter flow cytometry or mass cytometry (Baumgarth and Roederer, 2000; Chapman, 2000; Gong et al., 2014; Jaitin et al., 2014; Qiu et al., 2017), are widely used to define cell states in heterogeneous populations; while powerful, these methods provide incomplete readouts of cell phenotypes, and typically do not report on stability or transition dynamics. Transmitted light microscopy images directly reveal cell morphology and have historically formed the basis for identifying cell types and cell states in diverse fields, ranging from cell biology to neuroscience (Cajal, 1905; Flemming, 1882). These images can then be acquired at successive timelapse intervals and over long times, with minimal phototoxicity and without prior labeling or genetic manipulation, and resultant live cell movies can reveal additional information about the dynamics of these cell phenotypic states.

Traditionally, cell phenotypes have mostly been identified in transmitted light images by human visual inspection and interpretation. The advent of modern machine learning, however, is enabling high-throughput automated analysis of cell morphology, and could open doors for new deep learning approaches for the systematic, unbiased extraction of cell morphological states and their transition dynamics from these imaging data sets. However, several barriers remain to the development of such methods. First, current deep learning pipelines for cell image analysis rely heavily on predetermined knowledge to generate classification training datasets, or on large sets of heuristic formulations to capture the diverse cellular shapes (Bhaskar et al., 2019; Buggenthin et al., 2017; Carpenter et al., 2006; Eulenberg et al., 2017). When examining novel biological processes with minimal to no preconceived information, it can be difficult for investigators to determine what the important labels are without manual intervention and feature selection. Second, current machine learning pipelines generate features that are often not readily interpretable. A variety of unsupervised methods can generate reduced dimensionality representations from complex data, including principal component analysis (PCA), adversarial autoencoders (Makhzani et al., 2015), generative adversarial network (Salimans et al., 2016; Schlegl et al., 2017), and self-supervised deep learning approaches (Lu et al., 2019; Yao et al., 2019). However, these methods are limited in their ability to generate interpretable morphological features that allow scientists to investigate and understand the machine-identified cell states. Finally, current movie analysis methods cannot infer state transition dynamics from live cell movies in an automated, systematic manner (Held et al., 2010). Cell state transitions are typically observed from trajectories of single cells; however, despite recent advances (Moen et al., 2019), current tracking algorithms still typically require considerable parameter adjustment and manual error correction for generation of cell trajectories (Ng et al., 2018).

Here, we present an end-to-end deep learning method for elucidating cell phenotypic states and their dynamics from brightfield movies of living cells. This method, termed UPSIDE (for Unsupervised Phenotypic State IDEntification), is designed to facilitate unsupervised discovery of cellular phenotypic states, elucidation of morphological features that define these states, and inference of state transition dynamics. UPSIDE segments cells directly from brightfield images, then utilizes the variational autoencoder architecture (VAE) (Kingma and Welling, 2013) to learn intuitive latent features that can be clustered to reveal distinct morphological states, and also decoded to extract human-interpretable meaning. In order to demonstrate the use and versatility of UPSIDE, we first identified distinct hematopoietic cell phenotypes in a mixed dataset. We then analyzed live imaging movies of leukemic cells from an acute myeloid leukemia (AML) patient to identify morphologically-distinct cell states associated with stemness, and determined the rates of transition to and from these states. These results demonstrate the utility of UPSIDE as a tool for unbiased exploration of cellular states and their dynamics from large, time-resolved imaging datasets.

3.2 DESCRIPTION OF THE UPSIDE PLATFORM

UPSIDE is designed to be a versatile machine-learning pipeline for unsupervised exploration cell morphological states in transmitted light images, and subsequent elucidation of their transition dynamics from movies (Figure 3.1A, see Methods section for detailed description of the pipeline). In this pipeline, cells are first segmented using a neural network that converts brightfield images of unlabeled cells into synthetic fluorescent images of cytoplasm for segmentation (Ounkomol et al., 2018). This neural network is trained using a set of images of cells stained for their cytoplasm (Supplementary Figure S3.1). This approach allows the network to autonomously tailor its parameters, and to accommodate a wide range of different cell types in order to optimize performance without human input. Dead cells and other debris are eliminated from identified cell sub-images through a convolutional classifier that is trained on manually annotated brightfield cell crops labeled as live or dead (Supplementary Figure S3.2).

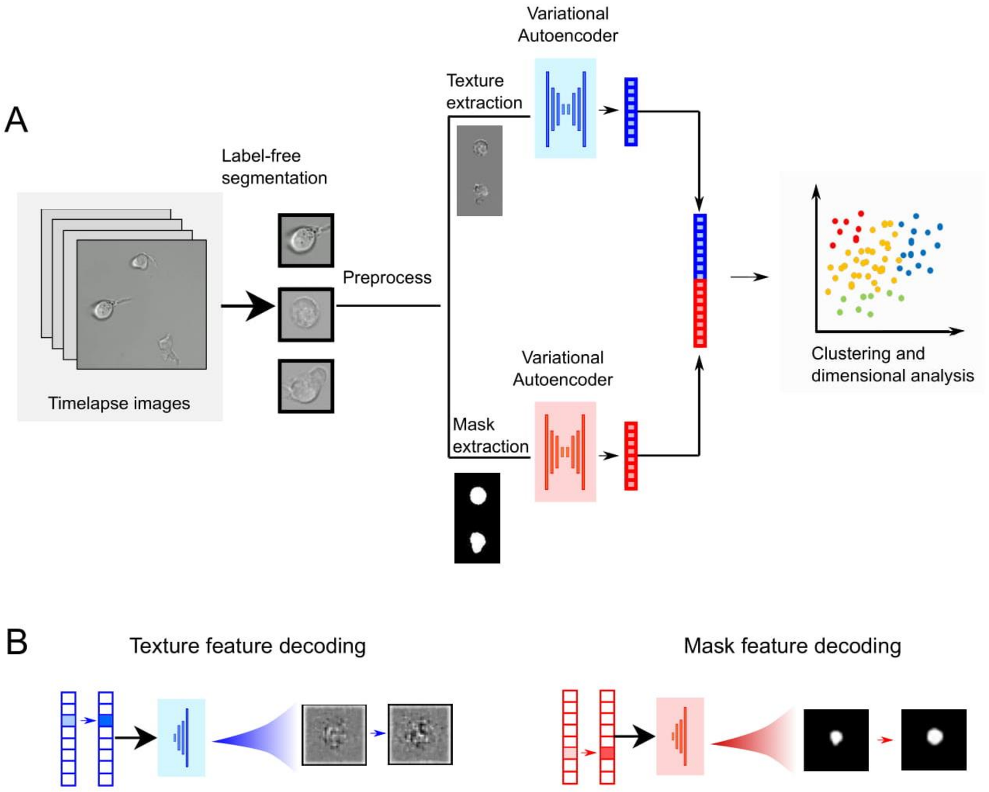
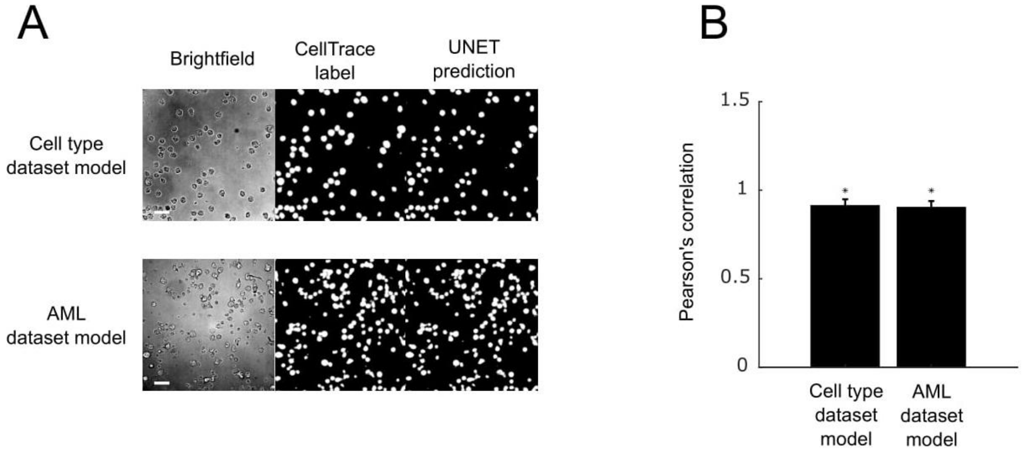
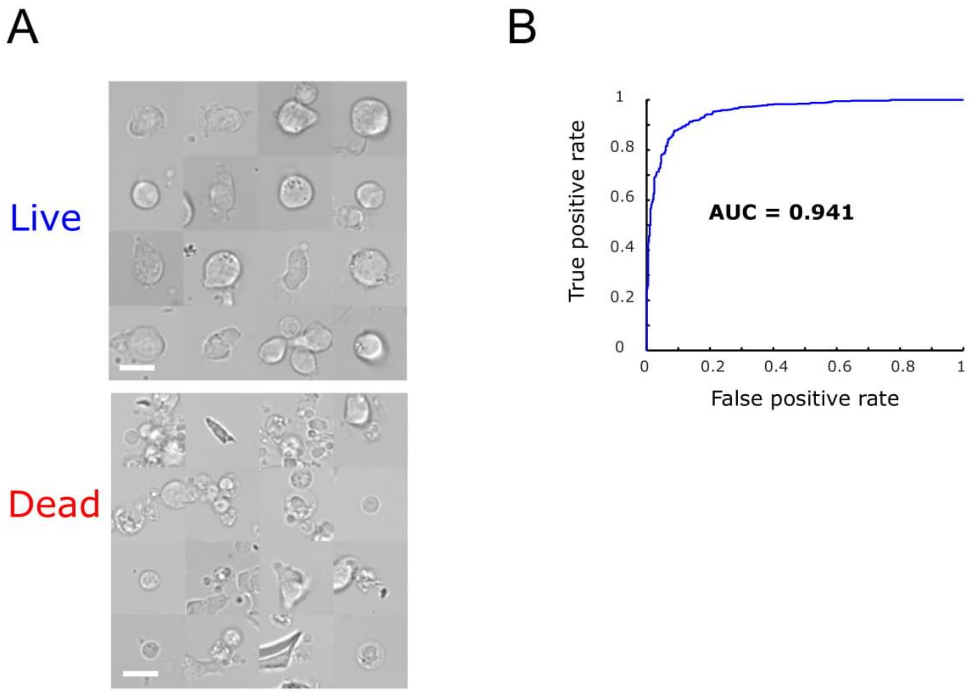


Figure 3.1. Description of the UPSIDE workflow. (A) Single cells are segmented

directly from brightfield images and deep learning UNET architecture to predict synthetic fluorescent images (Ounkomol et al., 2018) . Segmented cells are then pre-processed to generate separate mask and texture images, which are then used to concurrently train two variational autoencoders (VAEs). The shape and texture encodings learnt by these two VAEs are then concatenated and used for downstream data analysis. (B) Encoded latent vectors are then decoded into a shape and texture image to aid the interpretation of the encoded features.



Supplementary Figure S3.1. Robust label-free prediction of cell area using the UNET architecture. (A) Sample label-free results from models trained for the cell type dataset and the Acute Myeloid Leukemia dataset. Scale bar represents 20 **μ**m. Brightfield images and respective ground truth fluorescent images are used to compare and calculate Pearson's correlation values (B) with respect to their predicted synthetic images. * represents the theoretical upper limit of the model's performance for each dataset. Such a model would perfectly predict the fluorescent level of each cell but not be able to predict fluorescent noises that arise from the instrumentation (Ounkomol et al., 2018).



Supplementary Figure S3.2. A live dead cell classifier is trained to remove dead cells from the dataset. Brightfield crops of selected cells are labeled as either 'Live' or 'Dead' using a convolutional classifier that is trained to recognize dead cells using a manually labeled dataset. (A) Representative cell crops classified as 'Live' or 'Dead'. Scale bar represents 10 **μ**m. (B) Receiver operating curve (ROC) measuring the prediction performance of the trained classifier. AUC: Area under the ROC Curve.

UPSIDE utilizes a variational autoencoder (VAE) architecture to learn morphological features of segmented cells. Preprocessed masks of the cell and the cellular texture inside the mask boundary are then used to train two concurrent VAEs, one that encodes the cell shape, and another that encodes cell texture, through a binary mask. The mask VAE learns features related to overall cell shape and size, while the texture VAE learns pixel value variation within the mask itself while accounting for size and shape to some degree. Latent space encodings representing the learned mask and texture features are then multiplied with varying contributing weights, then concatenated for subsequent clustering and dimensionality reduction. Specifically, encodings are clustered into groups using the Louvain method (Blondel et al., 2008), represented on a 2D plane using the uniform manifold and projection algorithm (UMAP) (McInnes et al., 2018) Additionally, mask and texture vectors are subject to decoding, through variation of magnitudes of specific features or groups of features, followed by generation of synthetic images in observable image space (Figure 3.1B). This approach allows latent features to be visually displayed for human inspection and interpretation.

3.3 UPSIDE DISTINGUISHES BETWEEN MORPHOLOGICALLY DISTINCT CELL TYPES IN A HETEROGENEOUS BLOOD CELL POPULATION

To test UPSIDE's ability to learn cell type-defining morphological features, we first determined whether this pipeline could distinguish between different cell types based on their morphologies in a mixed cell dataset. We chose four blood cell types that, despite having distinct size, shape and textural features, were similar in their gross morphologies (Figure 3.2A, Supplementary Figure S3.3A): a mouse T cell leukemia line (Scid.ADH2), a mouse macrophage cell line (Raw246.7), a human acute myeloid leukemia cell line (Kasumi-1), and primary patient-derived leukemia stem cells (CD34+CD38- AML LSCs). Brightfield images from each cell population were captured; cells were segmented using the neural network described above; and image crops of segmented cells were mixed together and encoded into the latent space with UPSIDE's VAE (Supplementary Figure S3.4A). To quantify UPSIDE's performance, we devised a cell type homogeneity score, which reflects how closely cells of the same type cluster together in their latent space (see Methods section). We ran the VAE for this dataset, optimizing combination weights between the learned mask and texture encodings to achieve the maximum mean homogeneity score across the four cell types (Supplementary Figure S3.5A-B and Methods section). To compare the performance of the VAE to other deep learning methods, we repeated this analysis with several alternative architectures such as a vanilla autoencoder (AE) (Ballard, 1987), the adversarial autoencoder with latent dimension encoding trained to fit a normal distribution or mixed gaussian distribution (Makhzani et al., 2015) (1xAAE and 4xAAE, respectively), and the ClusterGAN architecture (Mukherjee et al., 2019) (ClusGAN) (see Methods section).

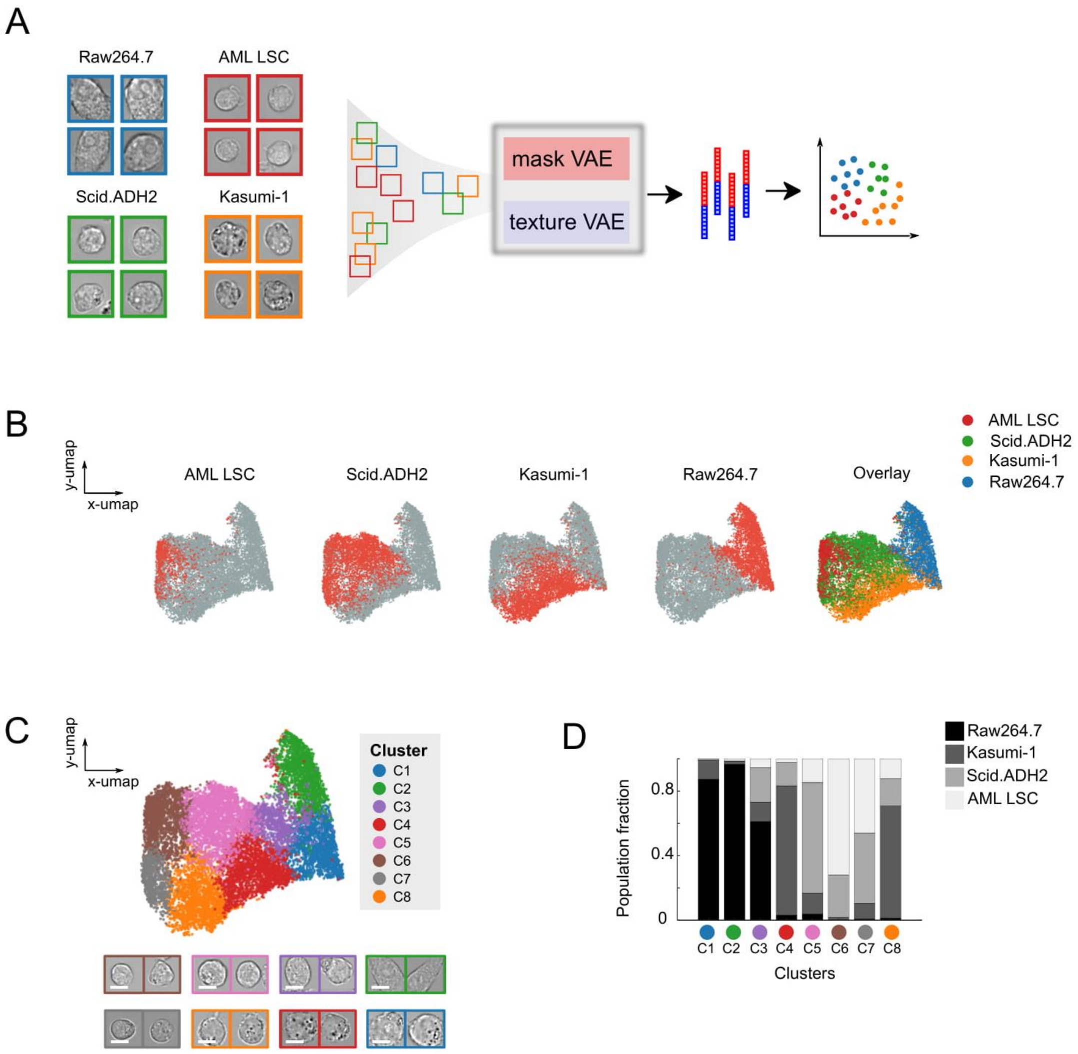
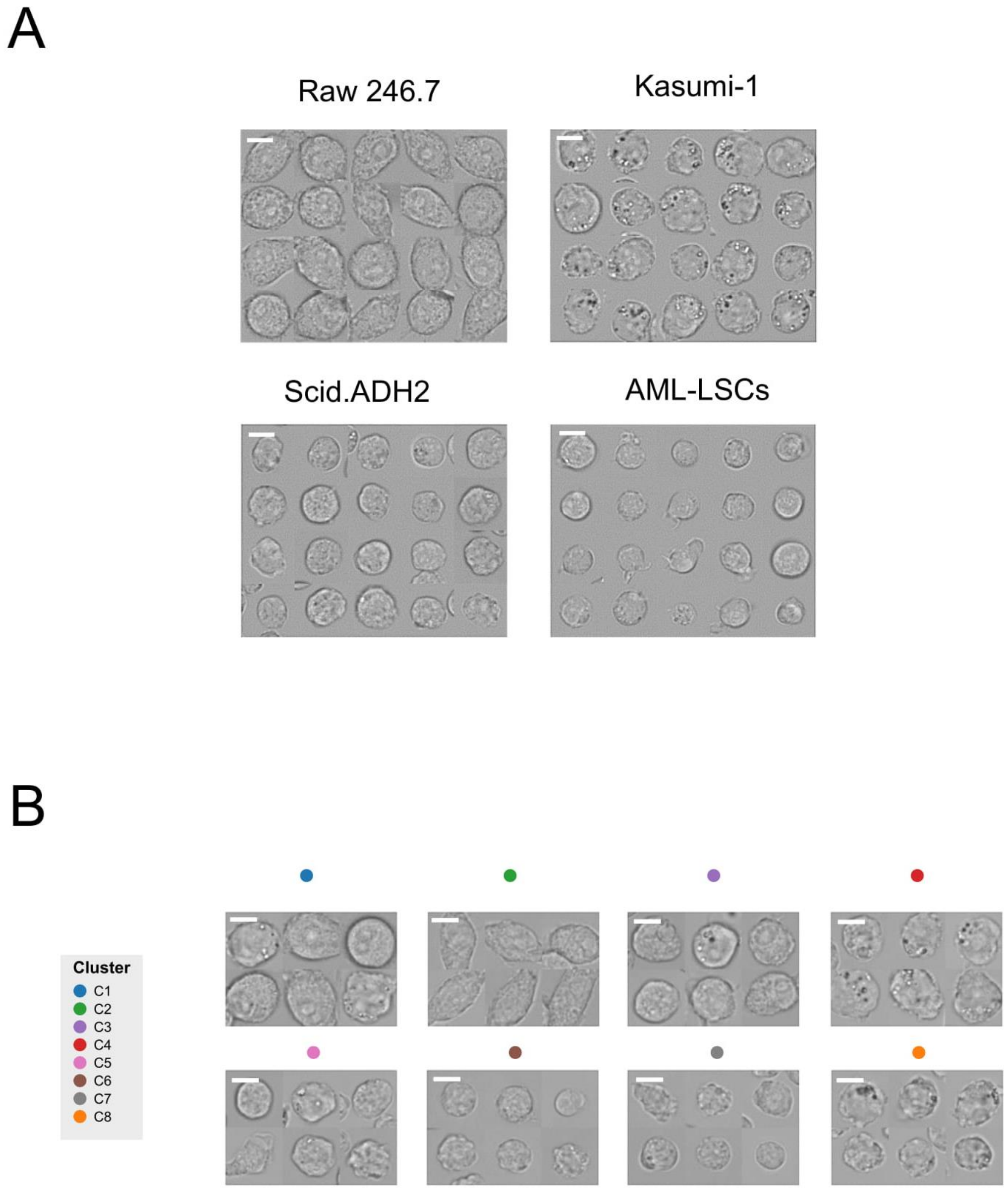
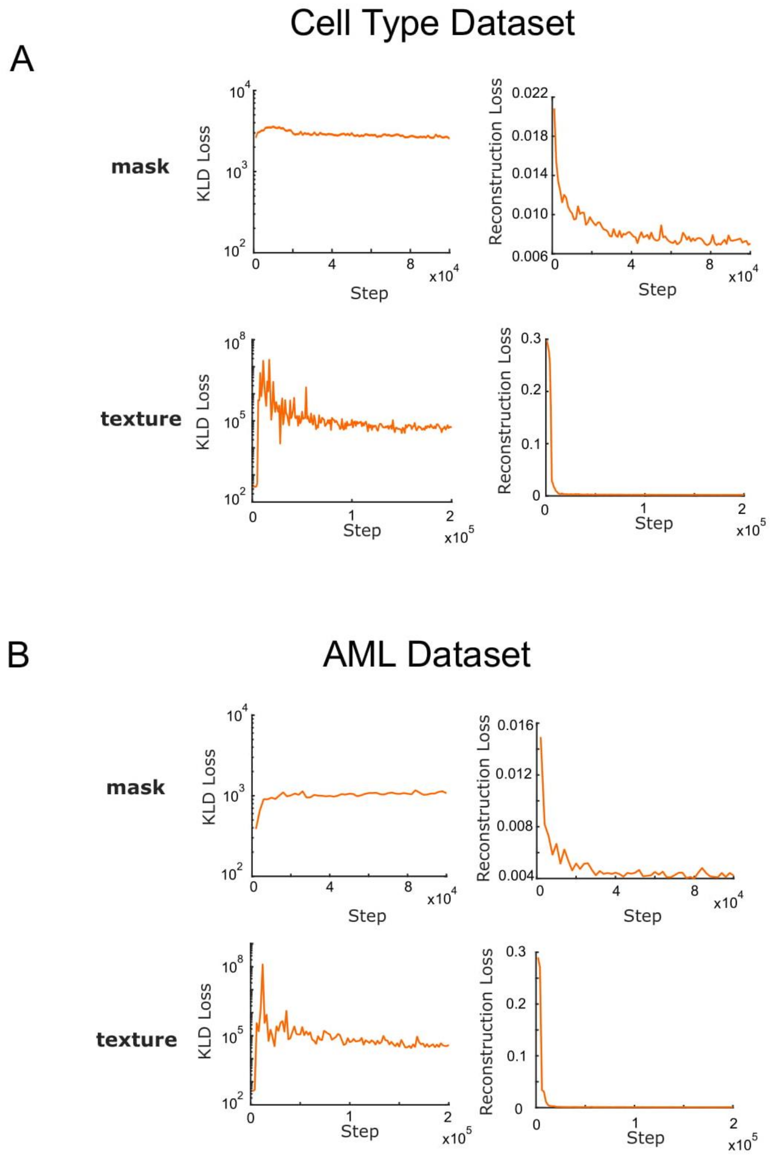


Figure 3.2. UPSIDE distinguishes morphologically-distinct blood cell types in a heterogeneous population. (A) Images of four different blood cell types were mixed together and passed through the UPSIDE workflow. Resultant shape and texture images were used to train concurrent VAEs. Output latent encodings were weighted relative to each other, concatenated, then projected onto a 2D plane using UMAP. (B) Dot plots show distribution of each cell type projected on 2D UMAP space made by UPSIDE. (C) 2D UMAP projection of the VAE-generated encodings that have been grouped into different morphological clusters using Louvain clustering algorithm. Representative

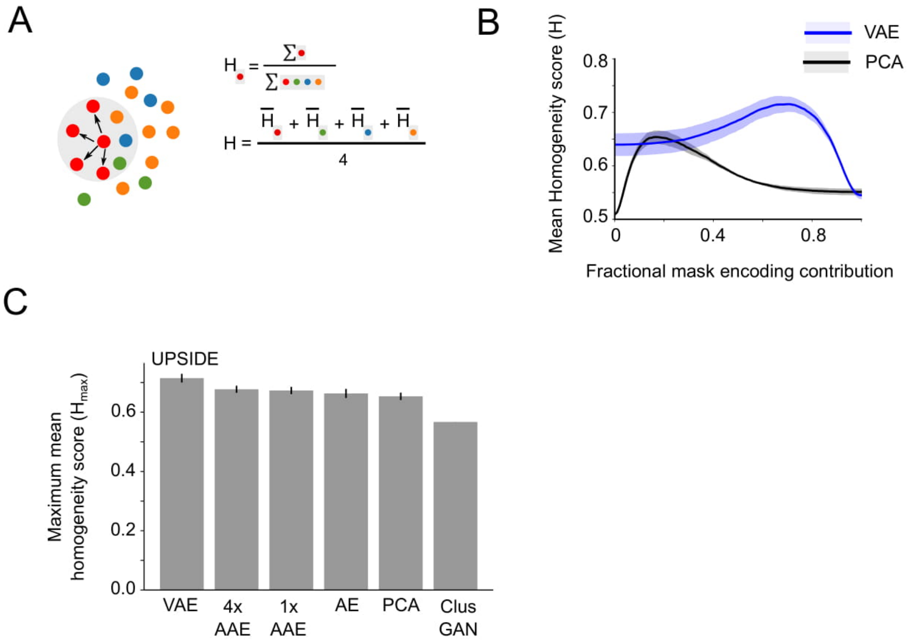
brightfield cell crop images from the different clusters were listed. Scale bar represents 5 μm. (D) Cell type fractional composition within each cluster. A fixed number of cells from each cell type were sampled, and the cluster-wised cell type composition was calculated from this pooled population.



Supplementary Figure S3.3. Cell images from blood cell types analyzed by UPSIDE. (A) Representative images from four analyzed blood cell types. (B) Representative images from eight different morphological clusters identified by Louvain clustering of the UPSIDE-generated latent vectors from each cell type. Scale bar represents 5 **μ**m.



Supplementary Figure S3.4. Concurrent training of shape and texture variational autoencoders for the cell type dataset. Reconstruction and Kubeck-Leibler Divergence (KLD) losses of the models for the Cell Types Dataset (A) and the Acute Myeloid Leukemia Differentiation Dataset (B).



Supplementary Figure S3.5. Cell type homogeneity scores were obtained using different data encoding methods. (A) Homogeneity score is calculated to measure how well different cell types are separated in latent space (B) Mean nearest neighbor score (H) across 4 cell types obtained with different relative mask weight contribution for encodings generated by either VAE vs PCA method. (C) Maximum nearest neighbor scores (H_{max}) between VAE, PCA, and other alternative deep learning architectures. Hmax is defined as the highest Mean nearest neighbor score across all weight combinations of mask and texture contributions. VAE: Variational Autoencoder, 4x AAE: Adversarial Autoencoder with latent dimension trained to fit a 4 mixed gaussian distribution, 1x AAE: Adversarial Autoencoder with latent dimension trained to fit a normal distribution, Clus GAN: Cluster Generative Adversarial Autoencoder with the one hot encoding component module removed, PCA: Principal Component Analysis.

Our VAE outperformed these other approaches, generating approximately 6% higher homogeneity scores compared to the adversarial autoencoders, 9% higher compared to PCA, and 26% higher than the ClusterGAN architecture (Supplemental Figure S3.5C). Adversarial autoencoders performed better than the vanilla encoder, though worse than the VAE, possibly because it is difficult to train the discriminator to perfectly fit the latent encoding to the desired distribution. Surprisingly, ClusGAN architecture performed the worst, even though it generated quite realistic-looking generated cell texture and mask images (data not shown). This weaker performance may stem from an inability to consistently generate direct, regularized encoded representations. These comparisons suggest that the VAE architecture is particularly well suited for learning morphological features for cell type discrimination.

To further visualize and analyze the representation of cells in latent space, we projected the encodings from the VAE into two dimensions using the UMAP algorithm (McInnes et al., 2018) (Figure 3.2B). From the UMAP projection, we found that the cell types largely segregated into distinct regions in this two-dimensional space (Figure 3.2B). Raw264.7 macrophages occupied a region that was largely distinct from regions occupied by other three cell types, reflecting their markedly different cell size and shape distribution. The three other cell types occupied partially overlapping regions, likely reflecting greater similarities in morphology among these cells (Supplemental Figure S3.3A). Interestingly, primary human AML stem cells (identified by their CD34+CD38- surface marker phenotype) overlapped parts of the Scid.ADH2 region, suggesting some of Scid.ADH2 cells look quite similar to their AML counterparts. Despite these overlaps, there are substantial areas in the two-dimensional space occupied by these regions containing only one cell type, indicating the presence of morphological features that distinguish each of these three cell types from another and allow them to be identified in mixed populations.

To better understand the morphological features that drive cell type discrimination in this learned latent space, we clustered cell representations in the latent space using the Louvain method, then visualized cells and the morphological attributes that defined each cluster. Eight clusters were identified, with each enriched for different cell types (Figure 3.2C-D, Supplemental Figure S3.3B). Clusters C1-3 were highly enriched for Raw264.7 macrophages that are larger phagocytic cells than their progenitor cells. Clusters C4 and C8 were highly enriched for Kasumi-1 cells that contained circular profile cells with dark granules, a unique distinguishing observable feature of these cells. Cluster C5 was enriched for Scid.ADH2 cells, which were also circular, but lacked granules. Clusters C6 and C7 were enriched for both LSCs and Scid.ADH2 cells, both of which were small and lacked granules. Cells in Cluster C7 had darker interiors and less well-defined cell boundaries when compared with Cluster C6 cells, which indicate they are flatter and may be more substrate adherent. The morphological differences within these clusters indicate the existence of distinct morphological sub-states within individual cell types.

To understand the morphological features that separate cells into distinct groups in latent space, we performed hierarchical clustering on the averaged latent space representation for cells from different clusters (Figure 3.3A). From this analysis, we found that each morphological cluster of cells was associated with a specific set of latent features, with magnitudes that are higher than population average. To decode these latent features, we transformed them back into synthetic images in visual space (Figure 3.3B-C, top). First, we generated a mean mask or texture vector by averaging over all cells in the dataset. From these mean vectors, we then selectively increased the magnitudes of the feature (or groups of features) of interest to generate a new vector. Using the VAE decoder module, we then transformed the feature-dominated vector and the mean vector into synthetic images for interpretation.

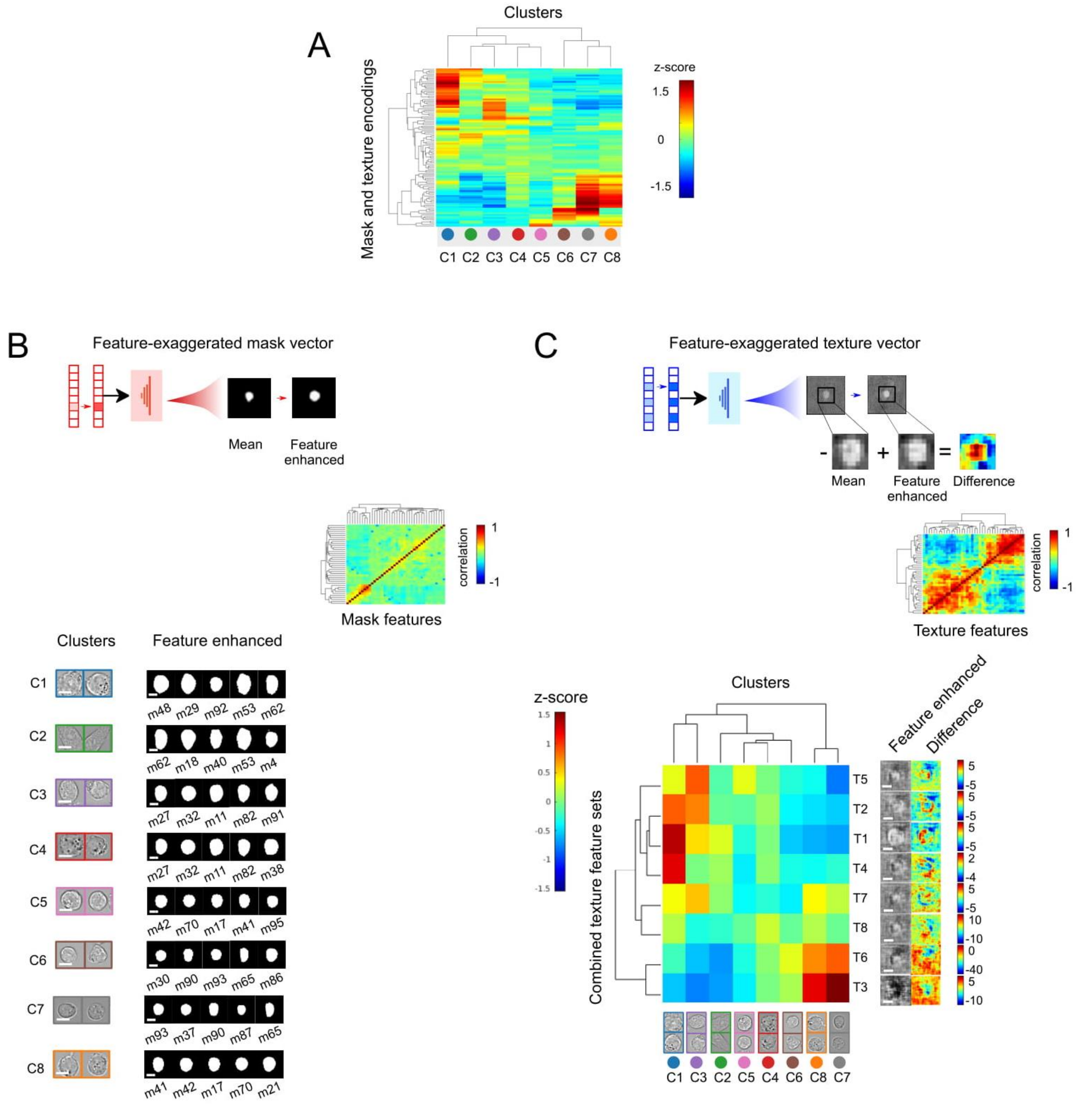


Figure 3.3. Cell-type specific morphological features can be interpreted by decoding the latent space cell representation. (A) Clustergram of average z-scores for latent shape and texture features for different cell clusters (see Methods section for how z-scores values were calculated). (B) Five mask features with highest z-scores for each morphological cluster are decoded and visualized. Inset: Clustergram shows matrix of correlation coefficients for forty mask features having the highest standard variation in

the dataset. Scale bar represents 5 μm. (C) Individual texture features were clustered into eight groups (T1-T8) according to their correlation with each other (inset and clustergram). Each group was decoded into brightfield difference images for interpretation (see Methods). Scale bar represents 5 μm.

We first examined the synthetic decoded images from the five most enriched mask features for each morphology-defined cluster (Figure 3.3B, bottom). Clusters C1-4 contained large cells with large, round profiles. As expected, cluster C2 contained elongated cells with numerous elongated mask features. Clusters C5-8, in contrast, contained smaller cells enriched in features representing small, round profile celler shapes. These mask features are in general agreement with sizes and shapes for cells found within individual clusters (Figure 3.2C).

Taken together, these features reveal how UPSIDE separates cells into distinct morphological clusters by their size, shape and distinct textural features. This ability can be seen readily in Clusters C3 and C4, where cells of similar size and profile can be discriminated based on their cell edge texture features. Cells with similar textural features can also be discriminated using other features, e.g., Custer C7 and C8 are both enriched with dark cell interior textures but differ in size with Cluster C7 cells larger on average than those in Cluster C8. These results demonstrate that UPSIDE can generate meaningful learned morphological features in an unsupervised manner, and these features can be effectively decoded into images to aid interpretability. This ability allows UPSIDE to extract valuable morphological properties by simply observing cells over time without prior manipulation or human annotations.

3.4 UPSIDE UNCOVERS MORPHOLOGICALLY DISTINCT CELL STATES IN CULTURED LEUKEMIC STEM CELLS

Leukemia stem cells (LSCs) play critical roles in acute myeloid leukemia (AML) disease propagation and drug resistance (Bonnet and Dick, 1997a; Ishikawa et al., 2007). LSC and other AML cell subpopulations are typically identified and characterized by a combination of cell staining for granule content and cell surface markers as well as by their gene expression signatures (Ho et al., 2020; Kaufmann et al., 2019). All of these classification approaches can be further extended by transmitted light imaging and analyses to provide complementary information about leukemic cell types and states that is not readily obtainable through more conventional classification approaches. In particular, live cell movies that resolve phenotypic states over time and in response to pharmacological treatment could provide unique insights into cellular heterogeneity and responses that could better inform therapeutic decision-making.

Towards this end, we employed UPSIDE to profile human LSCs cultured under cytokine conditions promoting expansion and differentiation, and filmed using brightfield imaging (Figure 3.4A, left). We directly isolated leukemic stem cells from an adult AML patient using as markers of stemness expression of the cell surface marker CD34 together with an absence of the differentiation marker CD38 (Bonnet and Dick, 1997a; Plesa et al., 2013; Roshal et al., 2013) (i.e., the CD34+CD38- cell fraction). To profile the self-renewal and differentiation dynamics of these sorted cells, we then cultured LSCs with either IL6 and TPO, to induce differentiation, or in the presence of Aryl hydrocarbon receptor inhibitors (AhRi) UM729 and SR1 to maintain stemness and suppress differentiation (Boitano et al., 2010; Fares et al., 2014; Pabst et al., 2014). We then filmed these cells in the brightfield channel for ~4 days at high temporal resolution (3-minute intervals, Figure 3.4A). To determine the association between observed cell morphological states, stemness and differentiation, we also added fluorescently-labeled anti-CD34 and anti-CD38 antibodies in situ, and took fluorescent images each hour to follow expression of these markers in imaged cells (Figure 3.4A, top right). Such in situ antibody labeling allows real-time visualization of cell surface marker protein expression with minimal effects on cell viability (Coutu and Schroeder, 2013). UPSIDE is well-suited to facilitate these types of time course analyses and image-based profiling: the use of brightfield imaging obviates the need for genetic engineering of fluorescent reporters to allow a wider range of analyses to be performed on primary patient-derived cell samples. Furthermore, a reliance on brightfield imaging minimizes cellular phototoxicity, thus enabling long-term cell observation at high temporal resolution.

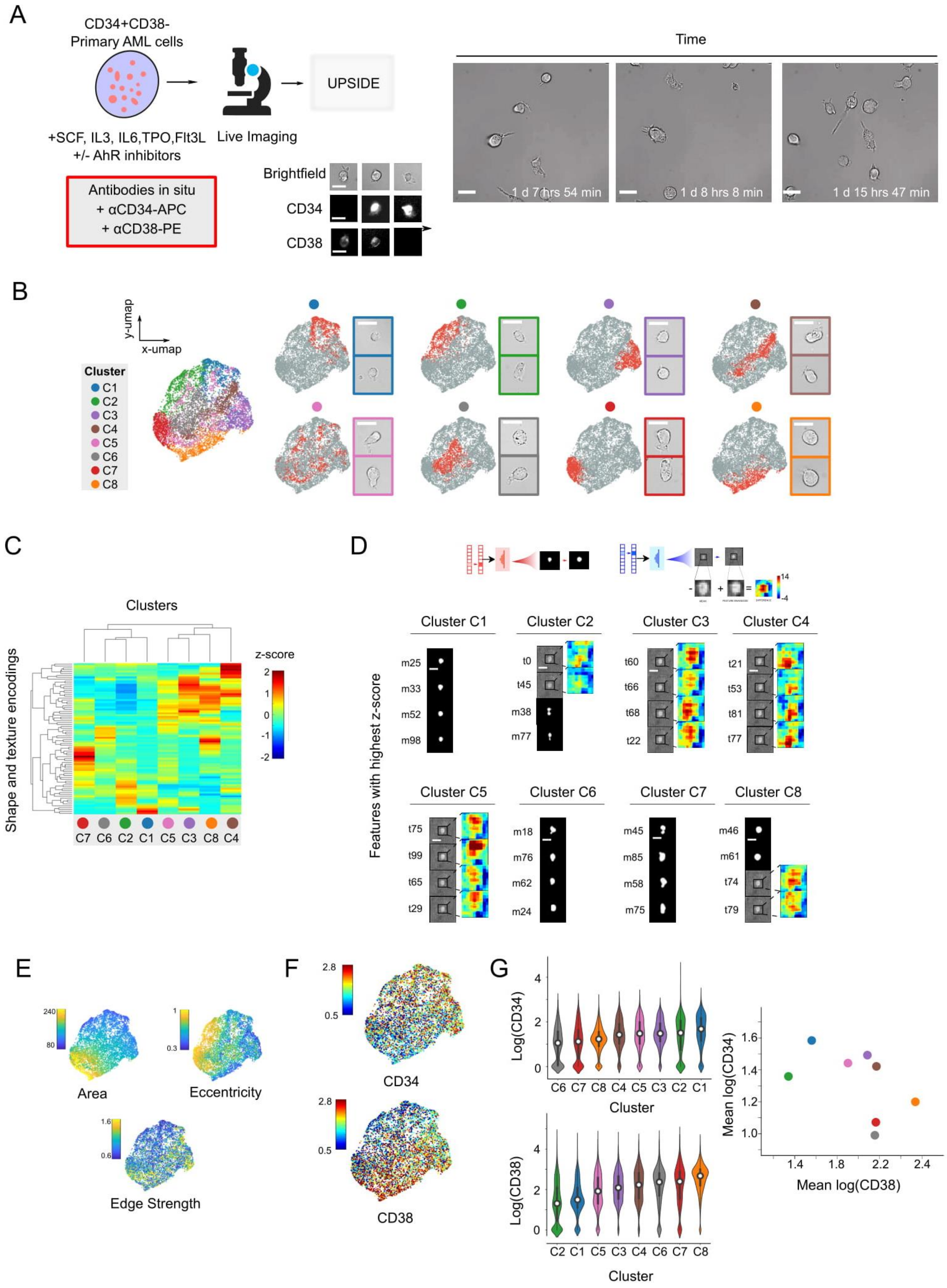
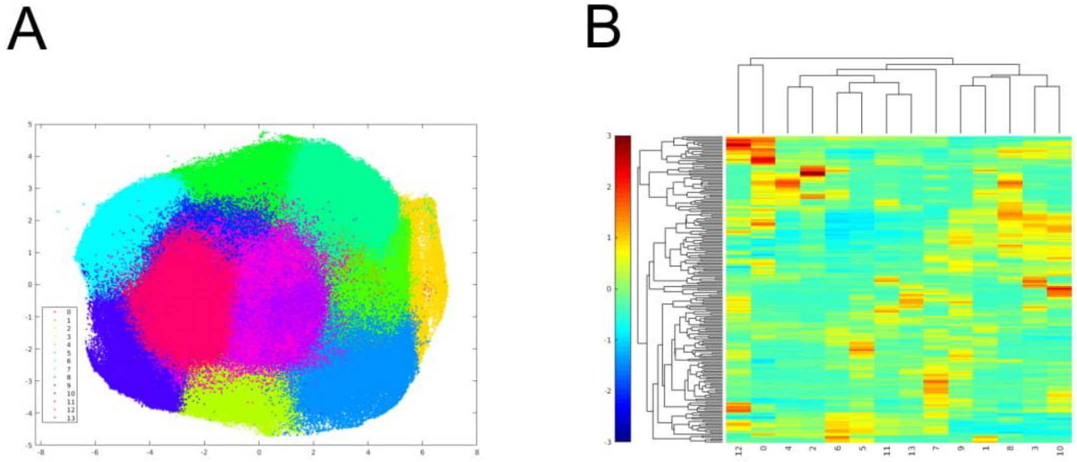


Figure 3.4. UPSIDE identifies stem cell-associated morphological states from patient-derived AML leukemic cells. (A) LSCs (CD34+CD38-) from an acute myeloid leukemia patient were cultured in cytokines with or without AhR inhibitors (UM729 and

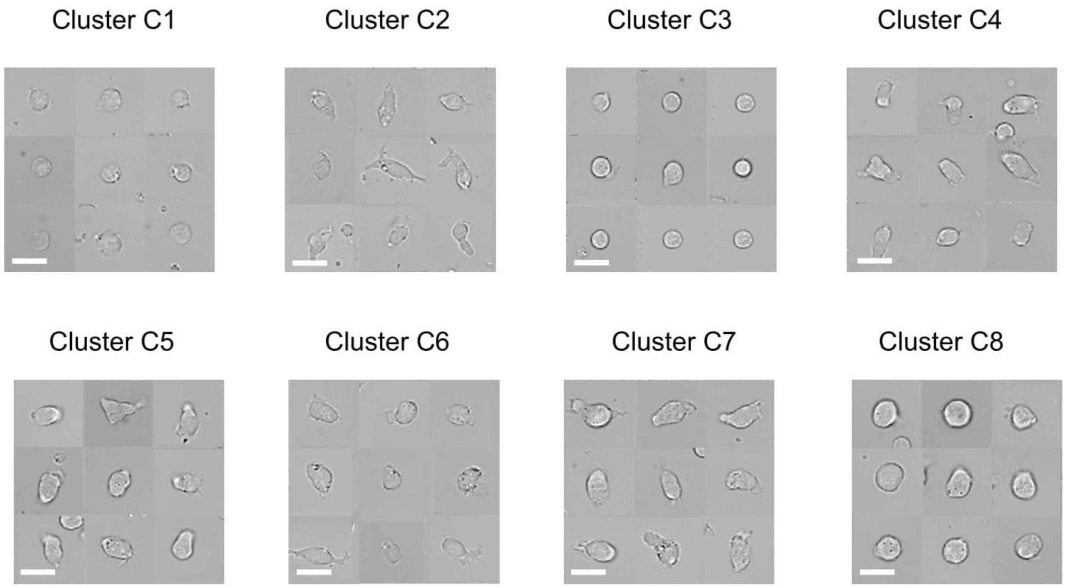
StemRegnin1) filmed for ~5 days (left). Brightfield images were captured once every 3-5 minutes. αCD34-APC and αCD38-PE antibodies were added in situ, and fluorescent images were captured once every hour (top right). Still frames show representative time lapse images of AML cells (bottom right). Scale bar represents 10 μm. (B) UMAP 2D projection of the UPSIDE generated latent space cell representations. Individual morphological clusters were identified using the Louvain Clustering algorithm, then grouped manually based on their proximity to each other in the latent space (See Supplementary Figure 6B). Representative cell images from each cluster were also shown. Scale bar represents 10 μm. (C) Clustergram shows Z-scores of the latent mask and texture encodings for each morphological cluster. (D) Decoded images of the four most enriched features for each morphological state. Texture features were visualized using difference maps that were zoomed in around the decoded cells. (E) Area, eccentricity, and edge strength for each cell were calculated and mapped to the UMAP latent space representation. (F) CD34 and CD38 levels were mapped onto the UMAP. (G) Violin plots show distributions of CD34 and CD38 expression levels in different morphological clusters (left). Scatterplot shows mean CD34 levels against CD38 levels for each morphological cluster (right).



Supplementary Figure S3.6. Clustering of AML cell morphologies in the latent space with the Louvain method. (A) 2D UMAP projection of learned mask and texture encodings from combined AML datasets. Each cell was colored based on the raw Louvain clustering result over all datasets. (B) Clustergram of the z-score from morphological groups defined by Louvain methods. Groups with closely related z-score patterns were combined into larger morphological clusters.

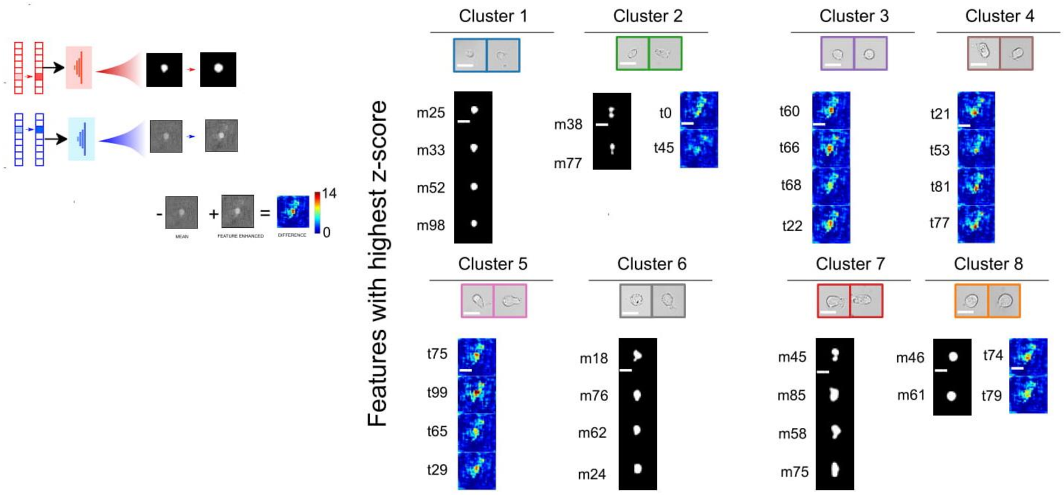
From the resultant timelapse images, we observed considerable heterogeneity in the morphologies of observed cells, with these cells differing in their sizes, shapes and textures, as observed by their observed contrast from transmitted light images (Figure 3.4A, right). To better understand the morphological states of these cells, we fed these images into the UPSIDE pipeline. Cells across all time points were segmented, then encoded using separate shape and texture VAEs (Supplementary Figure S3.4B). We then organized representations of cells in latent space into morphological groups using the Louvain clustering method (Supplemental Figure S3.6). Based on proximity in the latent space, we further combined some of these groups into larger clusters. To visualize obtained cell clusters in two-dimensional space, we then projected these cells' latent representations onto a two-dimensional plane using the uniform manifold approximation and projection algorithm (UMAP). This projection revealed the locations of the discrete clusters, along with their overlap regions. In this two-dimensional visualization, some cell clusters showed considerable boundary overlap with others, likely reflecting the continuous nature of the latent features encoded by the VAE.

In order to gain insight into the features that drive the separation of the cell encodings into distinct clusters, we performed hierarchical clustering on averaged cell encodings from each group (Figure 3.4C), then decoded the specific mask or textural features with the highest z-scores in each group to generate feature-dominated synthetic images as described above (Figure 3.3B-C). These synthetic images highlight significant morphological features that display coherence across cells within a cluster but differ between cells in different clusters (Figure 3.4D, Supplementary Figure S3.7-S3.8). Examples of coherent morphological features include size, with some having smaller cells (Clusters C1, C2) and others having larger cells (Clusters C6, C7, C8); cellular elongation or eccentricity, with some clusters displaying rounder cell profiles (Clusters C1, C8) and others have more elongated cells (Clusters C2 and C7); and the presence of finer morphological features, such as cytoplasmic protrusions from the cell body (Clusters C6 and C7). Another important coherent morphologic feature was cell edge and boundary morphology. Cluster C8 cells had pronounced cell edges all around the boundary, while other cells had dimmer, asymmetric cell edges (Clusters C4, C5), as visualized with synthetic difference images of cell texture. These features indicate cell flattening or protrusion out from the imaging plane in the Z-direction and may reflect differences in cell adhesion to substrate (Supplementary Figure S3.7). These decoding results indicate that clusters identified by UPSIDE contain leukemic cells with distinct morphological states.



`

Supplementary Figure S3.7. Representative brightfield images of identified cells from each grouped morphological cluster. Scale bar represents 10 µm.



Supplementary Figure S3.8. Four most enriched mask and texture features for each morphological cluster are decoded into the image space. Decoded texture images are accompanied by unzoomed pixel difference maps. Scale bar represents 10 **µ**m.

To verify that these differences in decoded features indeed reflect systematic morphological differences between cells in different clusters, we calculated cell area, eccentricity, and edge strength – defined by the maximum value of the cell's gradient image – and then plotted these quantities onto the 2D projection of the latent space (Figure 3.4E). Indeed, regions occupied by the different cell clusters had area, eccentricity, and edge strength values consistent with what was generally observed in the decoded cell images: Clusters C1-3 resided in the region with small cell areas, whereas Clusters C6, C7 and C8 resided in the region with larger cell areas. Elongated cells in Clusters C2 and C7 resided in regions with high eccentricity, whereas cells with darker cell edges in Clusters C3 and C4 resided in regions with high edge strength. Together, this analysis shows that UPSIDE can elucidate defining characteristic shape and textural features of cells in different morphological states.

3.5 DISTINCT MORPHOLOGICAL STATES ARE ASSOCIATED WITH DIFFERENT DEGREES OF STEMNESS

Cells in the different morphological states identified above may exhibit different degrees of AML cell stemness or differentiation. To test this idea, we quantified CD34 and CD38 expression levels for each cell, and mapped them onto the 2D projection of AML's learned latent dimensions (Figure 3.4F-G). We then calculated average CD34 and CD38 levels for all cells in a cluster, and plotted these values against each other. These plots show that the different morphological cell clusters differ in their average CD34 and CD38 levels, indicating different degrees of cellular maturation. Cells in clusters C1 and C2 had high CD34 but low CD38 expression. Cells in these clusters differed in their degree of elongation, but were mostly smaller and flat, indicating that the most stem-like cells that are characterized by their small size and their flatness, possibly because of strong substrate-attachment through the expression of cell adhesion proteins(Healy et al., 1995). Cells in Clusters C3, C4 and C5, which were larger and appear less flattened, had high expression of both CD34 and CD38, indicating that they may reside in an intermediate maturation state. Finally, cells in Clusters C6, C7 and C8, which were uniformly larger but had varying degrees of flatness, had low CD34 but high CD38 expression. Thus more mature cells may be generally larger, despite continuing to display variable shapes and degrees of substrate adherence. These results highlight the ability of

UPSIDE to discover subtle yet biologically relevant cell morphological states in heterogeneous populations.

3.6 POPULATION DYNAMICS OF MORPHOLOGICAL STATES

To gain insight into the population dynamics of cells in different morphological states, we examined how the numbers of cells in different clusters evolved over time, both with and without suppressing differentiation with AhR inhibitors (AhRi; Figure 3.5A, Supplementary Figure S3.9A) . In the absence of AhRi, cells with stem-like morphological states (Clusters C1 and C2) were progressively depleted, while those with differentiation-associated morphological states (Cluster C8) expanded, consistent with the maturation of LSCs into more differentiated cells over time. In the presence of AhRi, there was a less pronounced increase in differentiated morphological states with a concomitant decrease in stem-like states. This reflects the known effects of AhRi in maintaining stem cell self-renewal. Interestingly, the fractions of cells in intermediate morphological states (Clusters C3-7) remained largely unchanged, regardless of the presence of AhRi, suggesting that AhR inhibition may affect intermediate state transitions without driving specific outcomes. At the same time, CD34 levels decreased whereas CD38 levels increased over time, with both these changes becoming less pronounced with the addition of AhRi (Figure 3.5B, Supplementary Figure S3.9B). Together, these results provide insights into the kinetics of LSC self-renewal and differentiation, and how these kinetics are affected by pharmacological compounds that modulate self-renewal.

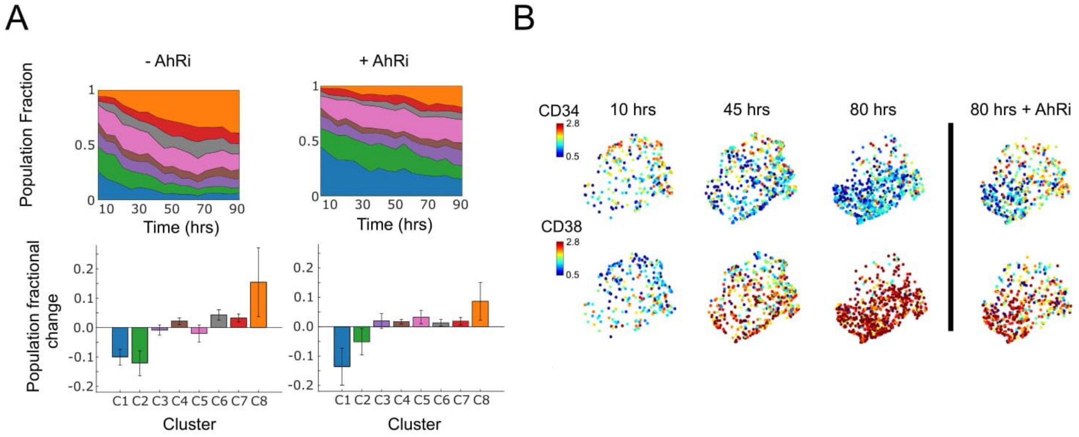
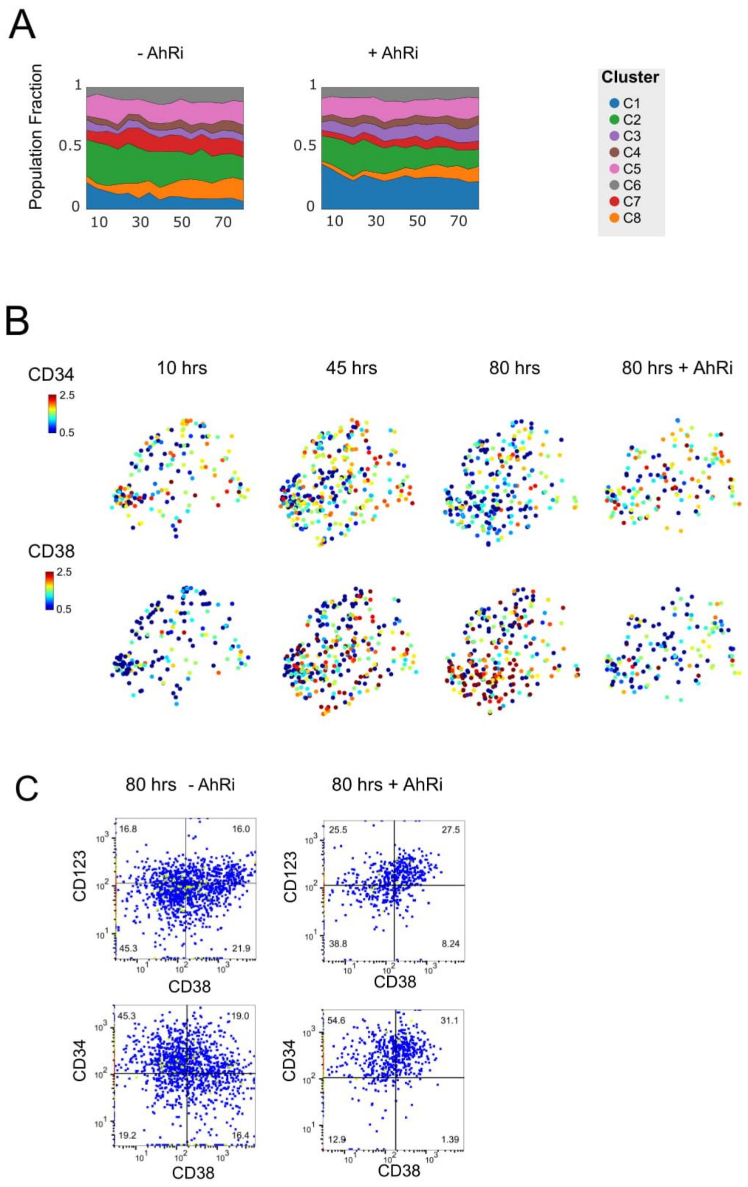


Figure 3.5. Population dynamics of identified morphological states. (A) Population fraction dynamics over time for each morphological cluster with (right) or without (left) AhR inhibitors (top). Population fraction difference for each cluster between the first 20 hrs and last 10-20 hrs of culture (bottom). (B) UMAP showing CD34 and CD38 expression levels at different time points, in the presence or absence of AhRi.



Supplementary Figure S3.9. Experimental replication for CD34 and CD38 evolution along with morphological change during LSC differentiation. (A) Dynamics of cells' CD34 and CD38 expression and their morphological properties during differentiation process in the presence or absence of AhRi. (B) Dynamics of cells' CD34 and CD38 expression and their morphological properties during differentiation process in the presence or absence of AhRi. (C) Flow cytometry analysis of CD34, CD38, and CD123 expressions of patient LSCs after parallel 80 hrs of non-microscopy culture with or without AhR inhibitors alongside the imaging experiment.

Parallel experiments were used to further explore morphological changes that reflect LSCs maturation. We cultured LSCs (CD34+CD38-) in parallel with live imaging experiments, and analyzed them after three days for expression of CD34, CD38, and CD123, another common LSC marker(Mony et al., 2008) (Supplementary Figure S3.9 C). Compared to untreated samples, cells treated with AhRi show higher expression of CD34 and CD123. On the other hand, CD38 expression magnitude was higher in the untreated sample, indicating greater differentiation in this population. Of note, a population of cells expressed both CD38 and CD34; this result indicates that expression of these markers is not mutually exclusive. Parallel live imaging experiments of AhRi-treated cells showed slower expansion of large round cell morphology clusters compared to the untreated counterpart (Figure 3.5A, Supplementary Figure S3.9A). Together, these results demonstrate that distinct cell morphological states identified using UPSIDE indeed reflect leukemic cells in different states of maturation.

3.7 INFERENCE OF MORPHOLOGICAL STATE TRANSITION RATES BY CELL LINKAGE ANALYSIS

The high temporal resolution of the brightfield movie analyzed above enables tracking of individual cells from frame-to-frame and, in conjunction with UPSIDE, inference of the rates at which leukemic cells transition between different morphological states. Here, we develop an analysis routine to automatically infer transition rates from brightfield movies. In particular, this UPSIDE-enabled analysis obviates the need for the generating individual cell tracks, which are typically error prone and requires considerable manual intervention. We paired cells from adjacent frames together, based on their close proximity. We then identified the morphological states of linked cell pairs using the

VAE above (see Methods section), and calculated state transition probabilities based on the frequencies of linked cell pairs with specific initial and final morphological states (Figure 3.6A). By repeating this calculation over all possible pairs of morphological states, we obtained a matrix, describing the transition probabilities between different morphological states (Figure 3.6B, left).

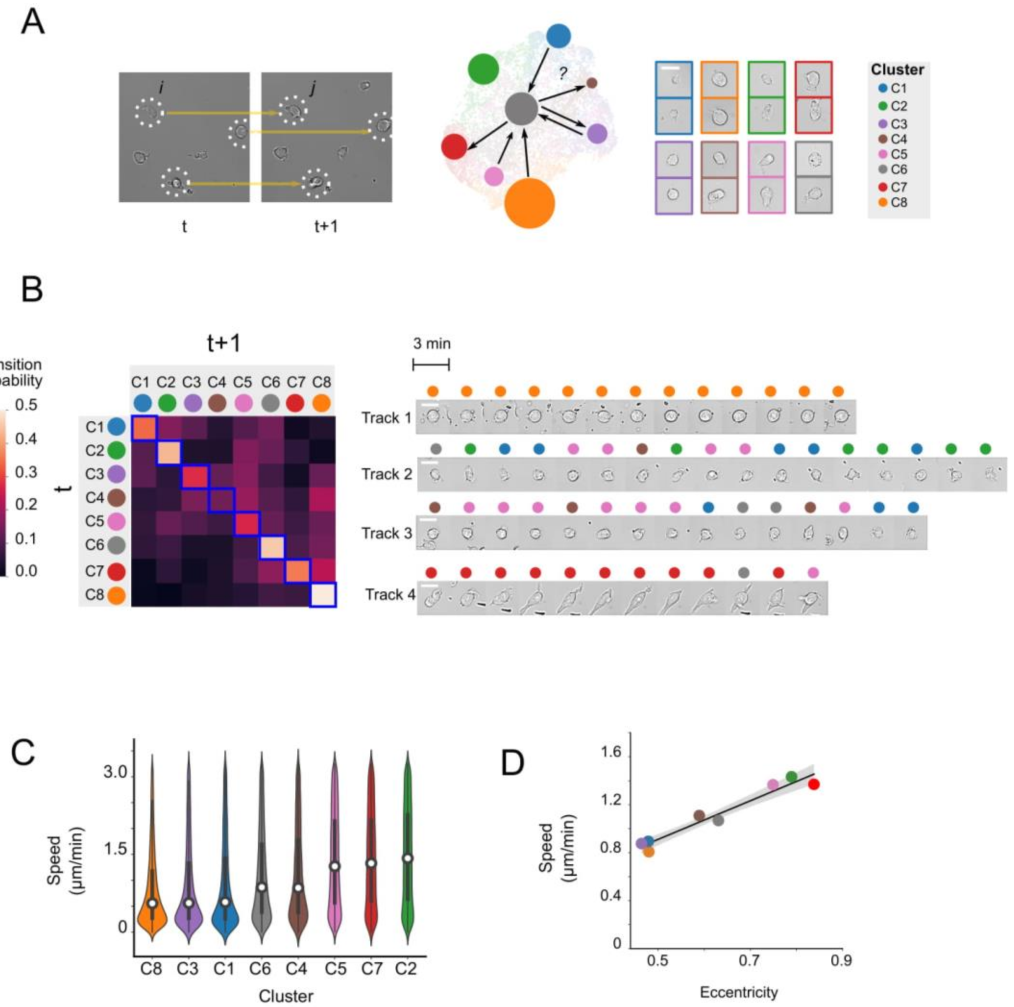


Figure 3.6. Calculation of morphological state transition probabilities by cell linkage analysis. (A) Cell pairs found in proximity across on successive time points were linked

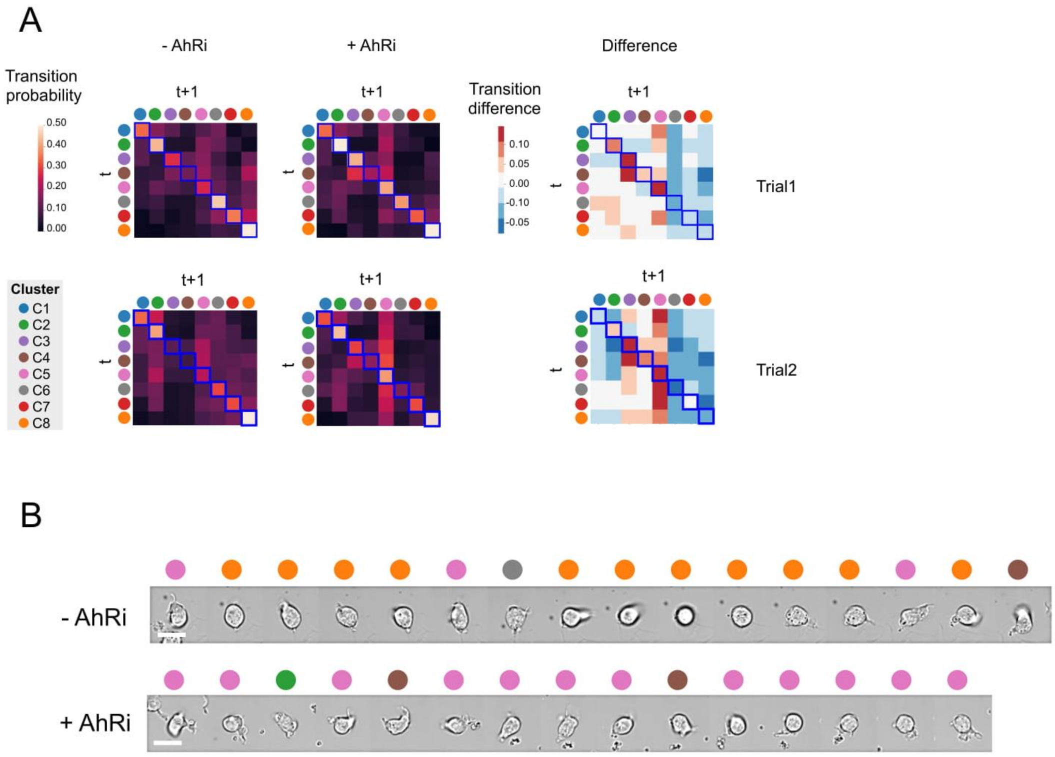
(left). Cell linkages, along with assigned morphological states of linked cells, were used to calculate transition probabilities between all states. (B) Heatmap shows transition probability matrix between all morphological clusters (left); image montage shows representative cell tracks identified from the culture (right). Scale bar represents 10 **μ**m. (C) Distribution of cellular velocity for linked cells for each morphological cluster. (D) Plot shows mean cell velocity against mean cell eccentricity for each morphological cluster.

This analysis showed that different morphological states showed considerably different stability and transition preferences (Figure 3.6B, left). Some states represented by Clusters C2, C6, C7 and C8 had high stability, as indicated by a high probability of remaining in the same state from one frame to the next. These represent a variety of morphologies, from small, elongated cells to larger, rounder cells. Other states – such as clusters C3, C4 and C5 – showed lower probabilities of remaining in the same state, and higher probability of transitioning into a different state by the next frame. To verify these inferred transition rates, we manually generated and inspected a number of different single tracks for cells in different morphological states (Figure 3.6B, right). Indeed, cells starting in states C7 and C8 (tracks 1 and 4) tended to remain in the same state, whereas cells starting in states C1, C3 and C5 were highly dynamic, switching from one state to another rapidly in successive frames (tracks 2 and 3). Notably, transitions occurred preferentially between particular groups of states. For instance, track 2 showed frequent transitions between states C1 and C2, whereas track 3 showed frequent transitions between states C4, C5 and C6, consistent with the elevated transition probabilities between these states as observed in the transition matrix (Figure 3.6B, left).

The morphological state transitions observed above occurred over tens of minutes (Figure

3.4B), a timescale much shorter than that for cell differentiation, which occurs progressively over the course of movie observation (Figure 3.5A). As such, these transitions are unlikely to directly report on cell differentiation events, but are instead more likely to reflect more rapid cell phenotypic state changes. Cells tend to polarize as they crawl on a substrate; thus, it is possible that some of these observed transitions could reflect transitions from a stationary to a motile state. To test this hypothesis, we derived the instantaneous velocities of cells in different morphological states, by calculating the displacement between successive frames for each state (Figure 3.6C). From this analysis, we found that cells with elongated morphologies, such as those in states C2, C4, C5, and C7, showed a higher movement velocity compared to other states. Consistently, there was a strong correlation between instantaneous velocity and cell eccentricity, averaged over all cells in individual clusters (Figure 3.6D). Thus, it is likely the case that the observed morphological transitions involve the rapid switching between stationary and mobile states (Figure 6B, left, tracks 2-3). Interestingly, both stem-like (CD34-high) and more differentiated (CD34-low) cells contained both stationary and moving cells populations. This indicates that cell mobility, or lack thereof, is not necessarily strongly tied to the differentiation process.

The switches between cells in different motility states; however, these transitions are likely to be influenced by cell differentiation transitions as well. To test whether this was the case, we repeated this state transition analysis for cells treated with AhR differentiation inhibitors (Supplementary Figure S3.10A). In the presence of AhRi, there was a decrease in the transition probabilities into the more differentiated states (C6, C7 and C8) Together with an increase in the transition probabilities into and amongst intermediate cell states (C3, C4 and C5). This suggests that AhRi blocks the conversion from a stem-like morphological state to a differentiated state by stalling cells in an intermediate state. Single cell track analyses showed that cells often transitioned rapidly from an intermediate state (C5) to more mature state (C8) without AhRi, but stalled at an intermediate state (C5) when AhRi is present (Supplementary Figure S3.10B). Taken together, these observations show that changes in cell differentiation dynamics can influence the probabilities at which cells transition between different morphological states.



Supplementary Figure S3.10. Transition dynamics between morphological clusters for AML in differentiation and stemness-preserving condition. (A) transition probability matrix between identified morphological states with and without AhRi in two replicates (B) Representative cell tracks in cultures with and without AhRi. Scale bar represents 10 **μ**m.

3.8 METHODS

3.8.1. EXPERIMENTAL TECHNIQUES

Culture conditions for hematopoietic cell lines. Kasumi-1 (gift from Pamela Becker), Scid.ADH2, and RAW246.7 cell lines were cultured Eagle's minimal essential medium (DMEM), phenol-red free containing 10% Fetal Bovine Serum (FBS), Penicillin-Streptomycin-Glutamine (Gibco 10378016) for 2 days before imaging. For 5 cell type imaging experiments, each cell line was imaged in separate individual well in the same 96-well glass-bottomed plate. AML211 $CD34^+CD38^-$ subpopulation was cultured in 'Differentiation Media Condition' for 2 days before imaging at the same time with the cell lines.

Culture conditions for AML211 subpopulations. Primary acute myeloid leukemia samples (AML211) were provided by Pamela Becker. Frozen cells were thawed and cultured in 'Long Term Bone Marrow Media' (Iscove's Modified Dulbecco's Medium (IMDM) with glutamine, 15% FBS, 1% Penicillin-Streptomycin (Gibco 15140148), 15% Horse Serum, 50 μM beta-mercaptoethanol (Sigma M6250), 0.043% Monothiolglycerol (Sigma-Aldrich, M1753), and 10 ng/ml Recombinant Human SCF (Peprotech 300-07)) for 2 days. Cultured cells underwent Ficoll separation to remove dead cells and stained with CD34 (ThermoFisher 17-0349-42), CD38 (ThermoFisher 12-0388-41), and CD45 (VWR 10758-692) for flow cytometry analysis and sorting. Sorted $CD45^+CD34^+CD38^-$ subpopulation from AML211 samples were cultured in 'Differentiation Media Condition' (adapted from (Klco et al., 2013)) which comprises of Eagle's minimal essential medium (DMEM), phenol-red free containing 10% FBS, Penicillin-Streptomycin-Glutamine (Gibco 10378016), 100 ng/ml Recombinant Murine SCF (Prepotech 250-03), 50 μM beta-mercaptoethanol (Sigma M6250), 10 ng/ml Recombinant Human IL-3 (Prepotech 200-03), 20 ng/ml Recombinant Human IL-6 (Prepotech 200-06), 10 ng/ml Recombinant Human TPO (Prepotech 300-18), 10 ng/ml Recombinant Human Flt3-Ligand (Prepotech 300-19), or 'Maintenance Media Condition' (adapted from (Pabst et al., 2014)) which comprises of minimal essential medium (DMEM), phenol-red free containing 10% FBS, Penicillin-Streptomycin-Glutamine (Gibco 10378016), 100 ng/ml Recombinant Murine SCF (Prepotech 250-03), 50 μM beta-mercaptoethanol (Sigma M6250), 20 ng/ml Recombinant Human IL-3 (Prepotech 200-03),), 50 ng/ml Recombinant Human Flt3-Ligand (Prepotech 300-19), 1 μM UM729 (STEMCELL Technologies 72332), and 500 nM StemRegenin-1 (STEMCELL Technologies 72342). Cells were cultured on treated polystyrene (Corning) of glass-bottomed (Mattek) 96-well culture plate coated overnight with 33.33 μg/ml Retronectin (Takara T202).

For imaging differentiation assay, CD34 Human Monoclonal Antibody (4H11), APC (eBioscience 17-0349-42) and CD38 Human Monoclonal Antibody (HB7), PE (eBioscience 12-0388-41) was spiked into the culture media. Cells were imaged every 3-5 min with brightfield and 60 min with fluorescent light for 4 days.

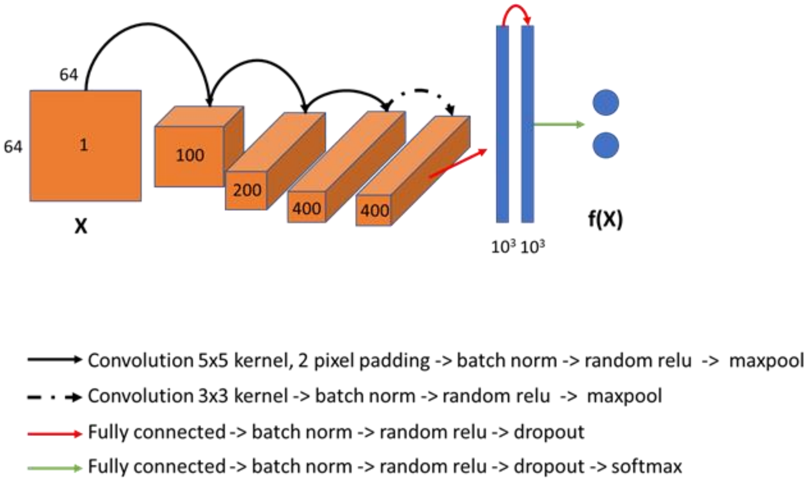
Image Acquisition. Timelapse imaging was performed on Inverted Microscope Platform Leica DMi8 (Leica Microsystem). All image acquisitions were performed using 40X air objective. Fluorescent images were captured using Laser Diode Illuminator, LDI (89 North).

3.8.2. IMAGE ANALYSIS

CellMorph computational pipeline is designed to analyze the morphological diversity of cells from timelapse brightfield images. The method consists of four main modules: 1) Label-free prediction, 2) Image segmentation, 3) Live cell classification, and 4) Unsupervised feature learning. The following section describes each of the modules in further details.

Label-free Imaging and Image Segmentation. CellMorph utilizes label-free imaging method to identify cells from brightfield (BF) images. Here we adapted a U-net-based deep learning technique described by (Ounkomol et al., 2018) to predict fluorescent pictures of the cells' cytoplasm and nuclei from the captured BF images. To complete this task, a small sample of cells to be analyzed was stained with Hoetchs 3342 (ThermoFisher H3570) to label the nuclei, or CellTrace Violet Cell Proliferation dye (ThermoFisher C34557) to label the cytoplasm. Training data was obtained by capturing approximately 300 – 400 BF images and their corresponding nuclei or cytoplasm fluorescent images. This data was then used to train a model that predicts nuclei and another that predicts cellular cytoplasm. These two models were subsequently used to predict fluorescent images for the main time-lapse brightfield image stacks. Object segmentation were performed on predicted nuclei images using ictrack software described in (Ng et al., 2018).

Live Cell Classification. Identified cell crops were then fed through a classifier that separate a living cell for further analysis from dead cells, and other debris that were picked up from the segmentation. We built a convolutional neural network for this classification task (Figure S11).



Supplementary Figure S3.11. Architecture of convolutional classifier neural network for live cell classification

To obtain training data for this network, ~10000 brightfield cell crops were manually annotated as ‘Live’ or ‘Dead.’ The network were trained for approximately 10,000 steps, and cross entropy loss were calculated a Adam optimizer (Kingma and Ba, 2017) were used for weights and biases learning:

$$CELoss = \sum_{i=1}^{J} \left(f(X) \log(y) + \left(1 - f(X)\right) \log(1 - y) \right) \qquad (1)$$

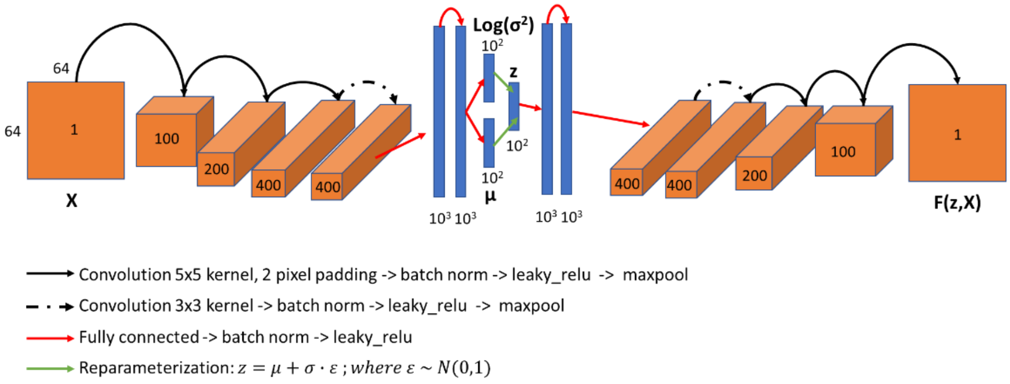
Where *f(X)* is the predicted class of a given cell crop *X* and *y* is its correct label. The remainder of the identified cell crops were then fed to the trained classification model. Crops classified as ‘Dead’ were discarded, and ‘Live’ crops were used for further analysis.

Unsupervised Feature Learning. Morphological feature learning in CellMorph relies on the variational autoencoder architecture (VAE) (Kingma and Welling, 2013) to perform feature extraction. Two information pieces were used to train the VAE: 1) The overall shape of the cell and 2) The cellular texture inside the boundary mask. Predicted CellTrace violet signal of the cell was used to generate the cell shape crop. The following image preprocessing steps were performed to minimize trivial variations between cell crops:

- Object re-centering
- Object rotational orientation to 90°. All cell crop images are then rescaled accordingly to eliminate image's dimension inflation due to rotation.
- Object's vertical and horizontal pixel density reorientation to top and right, respectively

To obtain texture representation, brightfield pixel values distribution inside the cell's mask was scale adjusted to zero mean and unit variation. They are then scaled linearly to be between 0 and 1 to facilitate learning with VAE. All pixel values outside the boundary was set to 0.5.

Preprocessed image crops for shape and texture were used to train two separate VAEs. The overall architecture is as described below:



Supplementary Figure S3.12. Architecture of convolutional variational autoencoder for cell shape and texture learning.

The loss function for the VAE is a weighted combination between reconstruction loss and Kullback-Leibler Divergence loss:

$$L = A[\gamma \cdot MSE + (1 - \gamma) \cdot KLD] \tag{2}$$

Where A is a constant, and γ is between 0 and 1. Additionally,

$$MSE = ||X - F(z, X)||^2 \tag{3}$$

$$KLD = -\frac{1}{2}\sum_{i=1}^{J}(1 + log(\sigma^2) - \mu^2 - \sigma^2) \tag{4}$$

VAE for cell shape feature extraction was first trained for ~100000 steps while VAE for texture feature extraction was first trained for ~200000 steps. Trained weights and biases for the cell shape extraction were then used to encode all cell crops obtained from the movie into 100-element vectors. These vectors were projected onto 2D plane using UMAP (McInnes et al., 2018). Cell crops with defective shapes are gated out using cytometry2 function in ictrack. The remaining crops were then used to train VAE for cell shapes and texture separately. Afterward, cell crops were encoded into 100-element shape vectors and 100-element texture vectors. Each cell crop’s latent vector is represented by a weighted concatenation between the shape and the texture contributions:

$$z = \ concatenate(w \cdot z_{shape}, (1 - w) \cdot z_{texture}) \tag{5}$$

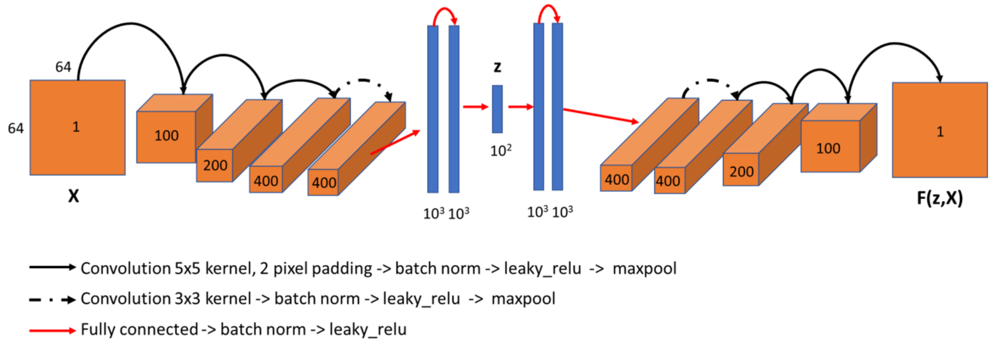
Encoded latent dimension of cell crops are then clustered using Louvain clustering algorithm.

To generate synthetic images, encoded cell barcodes and arithmetic variations were treated as z and fed directly into the decoder.

Comparable Deep Learning Architectures. In addition to utilizing the Variational Auto Encoder architecture to learn the latent dimensions in our imaging datasets, we tested a few other deep learning architectures to compare their performances with our current approach:

Vanilla Auto Encoder (AE) (Ballard, 1987)

In this architecture, each processed shape or texture is fed through a series of convolutional layers and fully connected neural network layers to generate a latent vector with a dimension of 100. The organization of the neural network layers are as follows:



Supplementary Figure S3.13. Architecture of convolutional Vanilla Auto Encoder (AE) for cell shape and texture learning.

The loss function for the AE is:

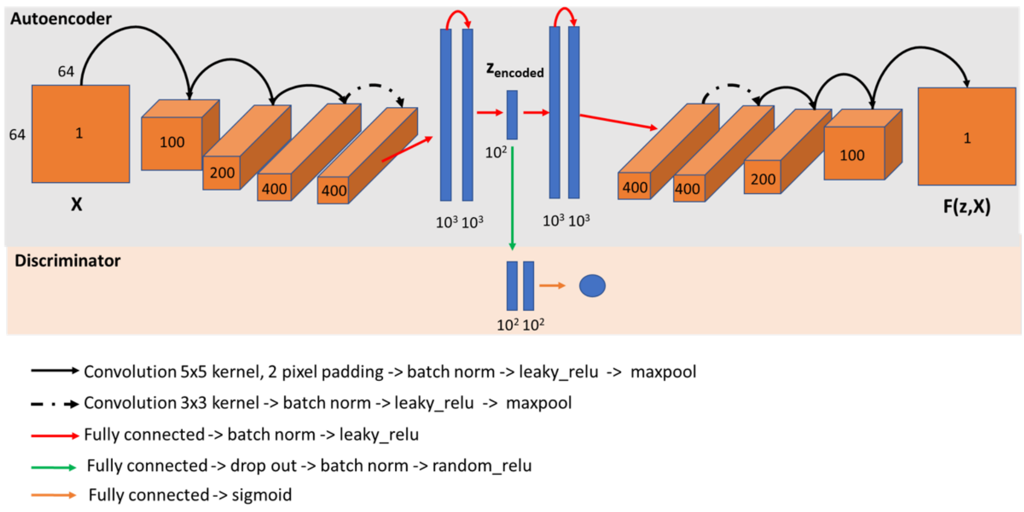
$$L = ||X - F(z, X)||^2 \qquad (6)$$

AE for cell shape feature extraction was first trained for ~100000 steps while AE for texture feature extraction was first trained for ~200000 steps. Each cell crop's latent vector is represented by a weighted concatenation between the shape and the texture contributions:

$$z = \ concatenate(w \cdot z_{shape}, (1 - w) \cdot z_{texture}) \qquad (7)$$

Adversarial Auto Encoder (AAE) (Makhzani et al., 2015)

In this architecture, each processed shape or texture is fed through a series of convolutional layers and fully connected neural network layers to generate a latent vector with a dimension of 100. The latent dimension was then regularized using a discriminator that forces the dimension space into a unit gaussian distribution (1x AAE) or four mixed gaussian distribution (4x AAE). The organization of the neural network layers are as follows:



Supplementary Figure S3.14. Architecture of convolutional Adversarial Auto Encoder (AAE) for cell shape and texture learning.

The loss functions for the VAE are:

$$L_{autoencoder} = ||X - F(z, X)||^2 \qquad (8)$$

$$L_{discriminator} = -\frac{1}{N}\sum_{i=1}^{N}(\log(z_{real}) + \log(1 - z_{encoded})) \qquad (9)$$

Where z_{real} is a 100 element vector sampled from a normal gaussian distribution (1X AAE) or a mixed 4-gaussian distribution with each gaussian's mean to be -1,

-0.5, 0.5, and 0.5 and standard deviation to be 1 (4X AAE).

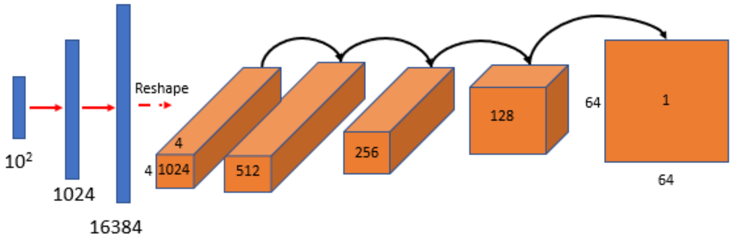
AAE for cell shape feature extraction was first trained for ~100000 steps while AAE for texture feature extraction was first trained for ~200000 steps. Each cell crop’s latent vector is represented by a weighted concatenation between the shape and the texture contributions: (10)

$$z = concatenate(w \cdot z_{shape}, (1 - w) \cdot z_{texture})$$

ClusterGAN (Mukherjee et al., 2019)

This architecture carries an encoder that converts a generated image into a latent dimension which is then forced to match the same starting latent code that was originally used to make the image. This is a semi-supervised architecture where a specific number of classes needs to be predetermined beforehand. To convert this into an unsupervised method, we removed the class module, enabling the GAN to draw data from a normal distribution, without the one-hot class vector input. The neural network organizations for the generator, encoder, and the discriminator are follows:

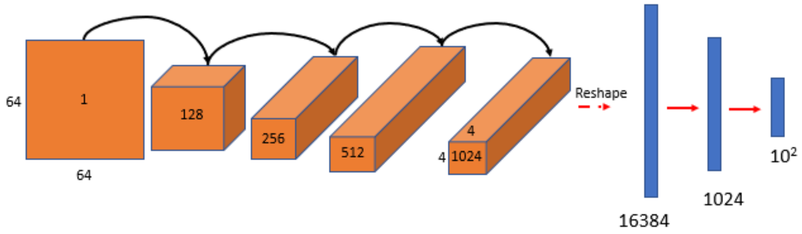
Generator



→ Fully connected -> batch norm -> leaky_relu

→ Transpose convolution 4x4 kernel, 1 pixel padding -> batch norm -> leaky_relu

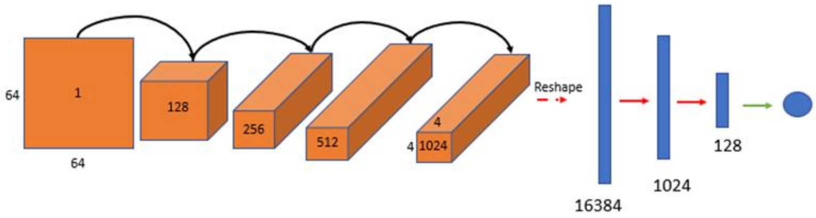
Encoder



→ convolution 4x4 kernel, 1 pixel padding -> batch norm -> leaky_relu

→ Fully connected -> batch norm -> leaky_relu

Discriminator



→ Fully connected -> batch norm -> leaky_relu

→ Convolution 4x4 kernel, 1 pixel padding -> batch norm -> leaky_relu

→ Fully connected -> sigmoid

Supplementary Figure S3.15. Architectures of the generator, encoder, and discriminator module of clusterGAN for cell shape and texture learning.

Loss functions used for training were described previously (Mukherjee et al., 2019). We input the cell crops into the encoder module of ClusterGAN to generate the latent dimensions for the comparative analysis with other architectures.

Cell shape feature extraction was first trained for ~100000 steps while the texture feature extraction was first trained for ~200000 steps. Each cell crop's latent vector is represented by a weighted concatenation between the shape and the texture contributions:

$$z = concatenate(w \cdot z_{shape}, (1-w) \cdot z_{texture}) \qquad (11)$$

3.8.3. ALGORITHMS AND QUANTITATIVE ANALYSIS

Neighbor similarity scoring. The metric is formulated to estimate the degree of homogeneity of the grouping of each cell types in encoding space of the four cell types data. Specifically, the neighbor similarity score H_C for a given cell type C is define as followed:

$$H_C = E_C\left[\frac{\sum_{i=1}^{N} n_i}{N}\right] \qquad (12)$$

Where $E(\cdot)$ represents the expectation value, or the mean over all cells within a cell type C, and N specifies a predetermined number of nearest neighbor cells to a given cell. Furthermore, for each neighboring cell i, $n_i = 1$ if the identity of i is C, and $n_i = 0$ otherwise.

Latent dimension z-score calculation. The z-score $Z_{f,c}$ of a particular feature f of cluster C is defined as the fold difference in standard deviation between the mean of the value of that feature in cluster C compared to that of the complete dataset:

$$Z_{f,C} = \frac{\mu_{f,C} - \mu_f}{\sigma_f} \qquad (13)$$

Here, $\mu_{f,C}$ is the mean value of feature f over all cells in cluster C, and μ_f, σ_f are the mean and standard deviation of feature f over the dataset.

Pairwise cell tracking algorithm. The pairwise cell tracking algorithm was built to ensure the validity of a given paired cell linkage from one frame to another. To achieve this goal, we established stringent requirements for a given cell pair to be considered ‘valid’. Specifically, the linking algorithm concerning all cells in frame t is as followed:

```
|| For all cells $i, j \in N_t$:
||     If $d(i,j) < D_t$:
||         For all cells $k \in N_{t+1}$:
||             If $d(i,k) < D_{t+1}$ and $d(i,k) = min\,(V_i^{t+1})$:
```

|| Link cell i and k together

Here, $N_t$represents a set of all detected cells in frame t; $d(a, b)$ denotes the Euclidean distance between cells a and b, and V_a^trepresents a set of Euclidean distances between cell a in frame $t-1$ to all cells in frame t

Transition probability between cell clusters. In order to estimate the transitional dynamics between identified morphological clusters through time, we determine the transition probability between a given cell X at time t to belong to cluster i to another cluster j in a set of clusters C at time $t + 1$ as followed:

$$p(X_{t+1} = j \mid X_t = i) = \frac{f_{ij}}{\sum_C f_{ik}} \tag{14}$$

Where f_{ij} is the number of transitions from i to j.

3.9 DISCUSSION

Here, we presented UPSIDE, a deep learning platform for unsupervised exploration of dynamic cell morphological states in transmitted light movies. Using UPSIDE, we identified cell morphological features and states in a heterogeneous collection of blood cell types. We found that the UPSIDE VAE learning architecture outperforms other comparable methods in recognizing unique morphological features within each cell type.

We further demonstrated the utility of our method by uncovering morphological states in primary human AML patient-derived leukemic cells displaying different degrees of stemness, differentiation, and cellular mobility. Finally, UPSIDE addresses the issue of latent feature interpretability, one of the most challenging aspects of analyzing deep convolutional networks, to allow more intuitive insight into learned latent morphological features.

UPSIDE provides a versatile new platform to enable the use of timelapse live imaging to discover phenotypic states and analyze their dynamics in heterogeneous cell populations. This method does not require genetic manipulation for labeling, identifying, or tracking cells, and thus enables observation and analysis of primary cells such as patient samples. The lack of reliance on fluorescence imaging reduces phototoxicity burden, which may have the added advantage of allowing imaging at higher time resolution over long times, thus allowing dynamics analyses over a wide range of timescales. Finally, UPSIDE is an unsupervised method which tailors analyses and outputs to differences within each individual dataset. Thus no initial knowledge of the system, feature selection or assumptions are necessary, as the VAE architecture is capable of self-learning the distinguishing features.

The ability of UPSIDE to resolve the dynamics of state transitions opens doors for systematic analysis of cell response dynamics in primary cells and other genetically intractable cell types. In particular, one attractive application for this pipeline is for automated, high throughput analysis of leukemic cell responses to chemotherapy. AML drug resistance poses a significant clinical challenge as the majority of patients eventually develop relapse disease. Growing evidence suggests leukemic stem cell (LSC) populations in AML harbor drug-tolerant sub-populations that survive drug treatment and eventually lead to relapse disease (Bonnet and Dick, 1997b; Roshal et al., 2013). To this end, a large scale screen of hundreds of drug treatment or regimen time courses can be captured via brightfield time lapse imaging (Kubben et al., 2016; Pietarinen et al., 2015). UPSIDE can then be employed as an unbiased method to survey and specify important morphological features associating with therapeutic response, persistence or resistance as a function of cell types, cell states and treatment. We anticipate UPSIDE will provide a valuable tool to enable the broader application of powerful deep learning to a wide range of biological questions.

3.10 REFERENCE

Ballard, D. (1987). Modular Learning in Neural Networks. In AAAI, p.

Baumgarth, N., and Roederer, M. (2000). A practical approach to multicolor flow cytometry for immunophenotyping. J. Immunol. Methods 243, 77–97.

Bhaskar, D., Lee, D., Knútsdóttir, H., Tan, C., Zhang, M., Dean, P., Roskelley, C., and Edelstein-Keshet, L. (2019). A methodology for morphological feature extraction and unsupervised cell classification. BioRxiv 623793.

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks.

Boitano, A.E., Wang, J., Romeo, R., Bouchez, L.C., Parker, A.E., Sutton, S.E., Walker, J.R., Flaveny, C.A., Perdew, G.H., Denison, M.S., et al. (2010). Aryl Hydrocarbon Receptor Antagonists Promote the Expansion of Human Hematopoietic Stem Cells. Science 329, 1345–1348.

Bonnet, D., and Dick, J.E. (1997a). Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nat. Med. 3, 730–737.

Bonnet, D., and Dick, J.E. (1997b). Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nat. Med. 3, 730–737.

Buggenthin, F., Buettner, F., Hoppe, P.S., Endele, M., Kroiss, M., Strasser, M., Schwarzfischer, M., Loeffler, D., Kokkaliaris, K.D., Hilsenbeck, O., et al. (2017). Prospective identification of hematopoietic lineage choice by deep learning. Nat. Methods 14, 403–406.

Cajal, S.R. y (1905). Manual de anatomía patológica general: y fundamentos de bacteriología (N. Moya).
Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J., et al. (2006). CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol. 7, R100.
Chapman, G.V. (2000). Instrumentation for flow cytometry. J. Immunol. Methods 243, 3–12.

Coutu, D.L., and Schroeder, T. (2013). Probing cellular processes by long-term live imaging--historic problems and current solutions. J. Cell Sci. 126, 3805–3815.

Eulenberg, P., Köhler, N., Blasi, T., Filby, A., Carpenter, A.E., Rees, P., Theis, F.J., and Wolf, F.A. (2017). Reconstructing cell cycle and disease progression using deep learning. Nat. Commun. 8, 463.

Fares, I., Chagraoui, J., Gareau, Y., Gingras, S., Ruel, R., Mayotte, N., Csaszar, E., Knapp, D.J.H.F., Miller, P., Ngom, M., et al. (2014). Pyrimidoindole derivatives are agonists of human hematopoietic stem cell self-renewal. Science 345, 1509–1512.

Flemming, W. (1882). Zellsubstanz, Kern und Zelltheilung (Leipzig: F.C.W. Vogel).
Gong, X., Zhao, Y., Cai, S., Fu, S., Yang, C., Zhang, S., and Zhang, X. (2014). Single Cell Analysis with Probe ESI-Mass Spectrometry: Detection of Metabolites at Cellular and Subcellular Levels. Anal. Chem. 86, 3809–3816.

Healy, L., May, G., Gale, K., Grosveld, F., Greaves, M., and Enver, T. (1995). The stem cell antigen CD34 functions as a regulator of hemopoietic cell adhesion. Proc. Natl. Acad. Sci. 92, 12240–12244.

Held, M., Schmitz, M.H.A., Fischer, B., Walter, T., Neumann, B., Olma, M.H., Peter, M., Ellenberg, J., and Gerlich, D.W. (2010). CellCognition: time-resolved phenotype annotation in high-throughput live cell imaging. Nat. Methods 7, 747–754.

Ho, J.M., Dobson, S.M., Voisin, V., McLeod, J., Kennedy, J.A., Mitchell, A., Jin, L., Eppert, K., Bader, G., Minden, M.D., et al. (2020). CD200 expression marks leukemia stem cells in human AML. Blood Adv. 4, 5402–5413.

Ishikawa, F., Yoshida, S., Saito, Y., Hijikata, A., Kitamura, H., Tanaka, S., Nakamura, R., Tanaka, T., Tomiyama, H., Saito, N., et al. (2007). Chemotherapy-resistant human AML stem cells home to and engraft within the bone-marrow endosteal region. Nat. Biotechnol. 25, 1315–1321.

Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Elefant, N., Paul, F., Zaretsky, I., Mildner, A., Cohen, N., Jung, S., Tanay, A., et al. (2014). Massively parallel single cell RNA-Seq for marker-free decomposition of tissues into cell types. Science 343, 776–779.

Kaufmann, K.B., Garcia-Prat, L., Liu, Q., Ng, S.W.K., Takayanagi, S.-I., Mitchell, A., Wienholds, E., van Galen, P., Cumbaa, C.A., Tsay, M.J., et al. (2019). A stemness screen reveals C3orf54/INKA1 as a promoter of human leukemia stem cell latency. Blood 133, 2198–2211.

Kingma, D.P., and Ba, J. (2017). Adam: A Method for Stochastic Optimization. ArXiv14126980 Cs.

Kingma, D.P., and Welling, M. (2013). Auto-Encoding Variational Bayes. ArXiv13126114 Cs Stat.

Kubben, N., Brimacombe, K.R., Donegan, M., Li, Z., and Misteli, T. (2016). A high-content imaging-based screening pipeline for the systematic identification of anti-progeroid compounds. Methods 96, 46–58.

Lu, A.X., Kraus, O.Z., Cooper, S., and Moses, A.M. (2019). Learning unsupervised feature representations for single cell microscopy images with paired cell inpainting. PLOS Comput. Biol. 15, e1007348.

Makhzani, A., Shlens, J., Jaitly, N., Goodfellow, I., and Frey, B. (2015). Adversarial Autoencoders. ArXiv151105644 Cs.

McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. ArXiv180203426 Cs Stat.

Moen, E., Borba, E., Miller, G., Schwartz, M., Bannon, D., Koe, N., Camplisson, I., Kyme, D., Pavelchek, C., Price, T., et al. (2019). Accurate cell tracking and lineage construction in live-cell imaging experiments with deep learning. BioRxiv 803205.

Mony, U., Jawad, M., Seedhouse, C., Russell, N., and Pallis, M. (2008). Resistance to FLT3 inhibition in an in vitro model of primary AML cells with a stem cell phenotype in a defined microenvironment. Leukemia 22, 1395–1401.

Mukherjee, S., Asnani, H., Lin, E., and Kannan, S. (2019). ClusterGAN : Latent Space Clustering in Generative Adversarial Networks. ArXiv180903627 Cs Stat.

Ng, K.K., Yui, M.A., Mehta, A., Siu, S., Irwin, B., Pease, S., Hirose, S., Elowitz, M.B., Rothenberg, E.V., and Kueh, H.Y. (2018). A stochastic epigenetic switch controls the dynamics of T-cell lineage commitment. ELife 7, e37851.

Ounkomol, C., Seshamani, S., Maleckar, M.M., Collman, F., and Johnson, G.R. (2018). Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy. Nat. Methods 15, 917–920.

Pabst, C., Krosl, J., Fares, I., Boucher, G., Ruel, R., Marinier, A., Lemieux, S., Hébert, J., and Sauvageau, G. (2014). Identification of small molecules that support human leukemia stem cell activity ex vivo. Nat. Methods 11, 436–442.

Pietarinen, P.O., Pemovska, T., Kontro, M., Yadav, B., Mpindi, J.P., Andersson, E.I., Majumder, M.M., Kuusanmäki, H., Koskenvesa, P., Kallioniemi, O., et al. (2015). Novel drug candidates for blast phase chronic myeloid leukemia from high-throughput drug sensitivity and resistance testing. Blood Cancer J. 5, e309–e309.

Plesa, A., Chelghoum, Y., Mattei, E., Labussière, H., Elhamri, M., Cannas, G., Morisset, S., Tagoug, I., Michallet, M., Dumontet, C., et al. (2013). Mobilization of CD34+CD38- hematopoietic stem cells after priming in acute myeloid leukemia. World J. Stem Cells 5, 196–204.

Qiu, X., Hill, A., Packer, J., Lin, D., Ma, Y.-A., and Trapnell, C. (2017). Single-cell mRNA quantification and differential analysis with Census. Nat. Methods 14, 309–315.

Roshal, M., Chien, S., Othus, M., Wood, B., Fang, M., Appelbaum, F., Estey, E., Papayannopoulou, T., and Becker, P. (2013). The proportion of CD34 + CD38low or neg myeloblasts, but not side population frequency, predicts initial response to induction therapy in patients with newly diagnosed acute myeloid leukemia. Leukemia 27, 728–731.

Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., and Chen, X. (2016). Improved Techniques for Training GANs. ArXiv160603498 Cs.

Schlegl, T., Seeböck, P., Waldstein, S.M., Schmidt-Erfurth, U., and Langs, G. (2017). Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide

Marker Discovery. ArXiv170305921 Cs.

Yao, K., Rochman, N.D., and Sun, S.X. (2019). Cell Type Classification and Unsupervised Morphological Phenotyping From Low-Resolution Images Using Deep Learning. Sci. Rep. 9, 13467.

CHAPTER IV. PERSPECTIVE AND FUTURE DIRECTION

This thesis has provided insight into how epigenetic mechanisms could control the timing of cell state transitions in mammalian cells, explained how such mechanisms can explain a wide range of population size control schemes in development, and introduced a machine learning tool for identification of cell states and their transitions from brightfield movies. Together, these works contribute to our understanding of cell state transitions well as provide the foundation for future experiments to identify important cell states for biomedical applications.

The epigenetic switching mechanism we elucidate unifies different processes such as histone modification dynamics, chromatin compaction, and gene activation into one comprehensive model that can be the starting point for investigating this mode of regulation at other genomic loci. Our mathematical modeling studies, on the applicability of epigenetic switching mechanisms behind different developmental systems, suggest that networks of epigenetic switches, simultaneously controlled by histone-modifying enzymes and transcription factors, can set developmental timetables that can allow for precise control of body sizes and proportions. Lastly, our computational pipeline, UPSIDE, presents a method to effectively identify cell states and profile their morphologies from brightfield movies, making this type of complex dataset more accessible and analyzable for scientists. Rich brightfield datasets can provide crucial morphological information to complement transcriptomics and protein expression profiles.

Future work will be necessary to further elucidate the mechanism of the timed epigenetic switch, and also determine its generality and functional significance for multicellular development and function. Our model predicts that dynamic histone modification-dependent nucleosome compaction – decompaction is at the heart of tunability in gene activation control; as such follow-up studies should focus on building experimental evidence for such processes operating at genomic locus. Techniques such as Stochastic Optical Reconstruction Microscopy (STORM) (Bates et al., 2013) coupled with DNA fluorescence in-situ hybridization (FISH) and concurrent imaging of histone modifications (Woodworth et al., 2020) can provide information on a methylation and compaction states at a gene locus. Genetic approaches such as recruitment of epigenetic modifiers to the locus, as well as biochemical experiments such reconstituting chromatin components in vitro to monitor compaction would allow for more direct interrogation of these dynamic processes at a genomic locus. Finally, a survey of the existence of such in other genes would further reinforce the generalizability of our model.

Future applications of our machine learning workflow would be in generating comprehensive cell state morphological profiles that complement existing transcriptomic and proteomic information on a given cell population undergoing different biological processes such as differentiation, division, and drug response. In conjunction with visual cell sorting techniques (Hasle et al., 2020), relationships between morphological states and transcriptomic states can be directly investigated. Since morphological information can give clues to a cell's lineage decision and drug resistance (Buggenthin et al., 2017; Domura et al., 2017), UPSIDE can be used to survey different drug response experiments and identify cellular morphologies associated with drug resistance and how they differ from one patient to another. Finally, isolating cells based on their morphologies and investigating their respective transcriptomic and proteomic profiles can lead to discoveries of novel gene and protein targets of chemotherapy.

REFERENCE

Bates, M., Jones, S.A., and Zhuang, X. (2013). Stochastic optical reconstruction microscopy (STORM): a method for superresolution fluorescence imaging. Cold Spring Harb. Protoc.

Buggenthin, F., Buettner, F., Hoppe, P.S., Endele, M., Kroiss, M., Strasser, M., Schwarzfischer, M., Loeffler, D., Kokkaliaris, K.D., Hilsenbeck, O., et al. (2017). Prospective identification of hematopoietic lineage choice by deep learning. Nat. Methods *14*, 403–406.

Domura, R., Sasaki, R., Ishikawa, Y., and Okamoto, M. (2017). Cellular Morphology-Mediated Proliferation and Drug Sensitivity of Breast Cancer Cells. J. Funct. Biomater. *8*.

Hasle, N., Cooke, A., Srivatsan, S., Huang, H., Stephany, J.J., Krieger, Z., Jackson, D., Tang, W., Pendyala, S., Monnat Jr., R.J., et al. (2020). High-throughput, microscope-based sorting to dissect cellular heterogeneity. Mol. Syst. Biol. *16*, e9442.

Woodworth, M.A., Ng, K.K.H., Halpern, A.R., Pease, N.A., Nguyen, P.H.B., Kueh, H.Y., and Vaughan, J.C. (2020). Multiplexed single-cell profiling of chromatin states at genomic loci by expansion microscopy. BioRxiv 2020.11.17.385476.