Cis-epigenetic control of developmental timing

Nicholas Pease

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Hao Yuan Kueh, Chair

Christine Disteche

Steven Henikoff

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

Cis-epigenetic control of developmental timing

Nicholas Pease

Chair of the Supervisory Committee:
Hao Yuan Kueh
Bioengineering

Development of a single multipotent progenitor into distinct cell-types of the correct number and proportion requires careful coordination of gene expression patterns over time. Within the context of these complex gene regulatory networks, lineage-specifying genes must be activated at a specific time and in order. Because these lineage-specifying events are unfolding within proliferating progenitors, variations in the relative timing of these activation events can alter the final numbers and relative proportions of each lineage. While extrinsic signals induce and guide these development processes, the temporal schedule of lineage-commitment events is frequently upheld in cell-autonomous manner. The mechanisms underlying cell-autonomous timing control and their role in regulating cell-type proportions remains unclear.

In this dissertation, I first review our current understanding of cell-autonomous control of development timing by gene regulatory networks and provide an overview of early thymocyte development as a model system to study the fundamental principles of gene regulation and developmental cell fate control. In Chapter 2, I provide evidence for a *cis*-epigenetic switching mechanism that enables tunable temporal control of developmental gene activation and lineage-specification. In Chapter 3, I detail our approach to efficiently interrogate a collection of non-coding regulatory elements that separately control the timing, initiation and maintenance of the *cis*-epigenetic switch. Finally, in Chapter 4, I provide evidence that the timing of this *cis*-epigenetic switch is important for regulating the number and types of cells that emerge during thymocyte development.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I thank my mentor, Hao Yuan Kueh, for his guidance, support, patience and friendship. You've changed the way I think about science and the world around me more generally. Your enthusiasm, curiosity, optimism, and creativity will continue to inspire me.

I thank all the past and present members of the Kueh group. Specifically, I would like to thank: Kenny Ng, who patiently taught me all the skills essential to the Kueh Lab and generated the mouse models that made all of this work possible. Sam Nguyen, who was an essential collaborator through this whole process was always willing to patiently and elegantly explain computational modeling in terms I could understand. Allan Wang, who did much of the heavy lifting in molecular cloning and viral production and was always willing to help out anything else. Kathleen Abadie, Matt Wither, Jon Chu, Will White, Marcus Woodworth, Elisa Clarke and Lihua Chen, who all went out of their way to help out with different aspects of this work and constantly provided encouragement and constructive feedback.

I thank my committee members Christine Disteche, Steven Henikoff, Cole Trapnell and Brian Iritani for sharing their time and expertise inside and outside of committee meetings throughout the entire process. You all made this work better.

I thank my past mentors Lisa Privette Vinnedge and Susa Wells, who helped get me to where I am now and continue to provide support and guidance.

I thank my partner, Megan McQuade, without whom this process would have been inconceivably more difficult. Thank you for your patience and support and for always keeping things in perspective.

Funding sources:

University of Washington Institute of Stem Cell and Regenerative Fellowship; and F31 NRSA Fellowship (National Heart, Lung, and Blood Institute: HL151090)

Chapter 1. INTRODUCTION

In this dissertation, I will provide evidence that *cis*-epigenetic switching, involving cooperation among transcription factors, chromatin-associated proteins, and non-coding regulatory elements, can provide a mechanism for cell-autonomous control of developmental timing. Furthermore, I will provide evidence that this *cis*-epigenetic switching mechanism can be calibrated to modulate the cell-type proportions. Below, I discuss the historical evidence for cell-autonomous control of developmental timing, review the mechanisms underlying developmental gene regulation, and provide an overview of early thymocyte development as a model system to study the intersection of the two phenomena.

1.1 CELL-AUTONOMOUS CONTROL OF DEVELOPMENTAL TIMING

Chapter 1.1 is adapted from the following manuscript:

Nguyen, P.H.B, Pease, N.A., Kueh, H.Y. (2021) Scalable control of developmental timetables by epigenetic switching networks. *under revision and resubmission*, JRSI.

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 (De Beer, 1940; Gould, 1977; Huxley et al., 1942). Changes to the timing of individual lineage specification events can give rise to innovations in shape or form (Alberch et al., 1979; Chen and Goldhamer, 2004; Gérard et al., 1997; Gould, 1977; Moreau et al., 2019), whereas changes to the speed at which developmental timetables unfold can generate proportionally scaled changes in organ and organism sizes (Calder, 1996; Matsuda et al., 2019; Okamoto et al., 2016). As development is not coupled to an external clock, the timetables for lineage specification must be encoded by the embryo itself (Ebisuya and Briscoe,

2018; Rayon and Briscoe). 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 partly determined 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. The same developmental timetable unfolds *in vitro* in the absence of any intact tissue organization, suggesting it is largely set in a cell-autonomous manner (Eiraku et al., 2008; Gaspard et al., 2008). Furthermore, though progenitors from different species often share conserved gene regulatory networks and follow the same lineage specification order *in vitro*, they can differ substantially in the speed at which they traverse this schedule (van den Ameele et al., 2014; Barry et al., 2017; Davis-Dusenbery et al., 2014; Rayon et al., 2020), which can even be upheld in chimeric conditions (Espuny-Camacho et al., 2013; Fuhrmann et al., 2020; Otani et al., 2016). While the timing of developmental events can be regulated by cell cycle counting mechanisms (Amodeo 2015, Newport Kirschner 1982, Levine and Elowitz 2014), many of the cell-autonomous developmental processes found in vertebrate are uncoupled from cell cycle and thus must be regulated by cell division-independent timing mechanisms (Burton et al., 1999, 1999; Gao et al., 1997; Heinzel et al., 2017; Okamoto et al., 2016; Otani et al., 2016).

In summary, developmental schedules need to be carefully choreographed to specify the size, form and functions of different tissues. These schedules are often executed by cell-autonomous timing mechanisms which possess two unique features. First, developmental schedules can be adjusted across species (Rayon et al., 2020) or in response to different signaling

inputs (Heinzel et al., 2017; Raff, 2007), thus the molecular mechanisms upholding the schedules need to be tunable such that schedules can be delayed or accelerated without completely disrupting the entire process. Second, these schedules can also function independently of cell cycle rates, thus these cell-autonomous molecular mechanisms must track time and not cell divisions. While transcription/translation dynamics (Matsuda et al., 2019) and protein stability (Rayon et al., 2020) have recently been proposed as mechanisms for cell-autonomous control of developmental timing, I will present evidence in *Chapters 2-4* that *cis*-epigenetic switching of gene expression states can also serve as a compatible mechanism that enables both tunable and division-independent control of developmental timing.

1.2 REGULATION OF GENE EXPRESSION DYNAMICS DURING DEVELOPMENT

Chapter 1.2 includes text and figures adapted from the following manuscripts:

*Nguyen, P.H.B, *Pease, N.A., Kueh, H.Y. (2021) Scalable control of developmental timetables by epigenetic switching networks. In revision, JRSI.

Chu, J.M., Pease, N.A., and Kueh, H.Y. (2021). In search of lost time: Enhancers as modulators of timing in lymphocyte development and differentiation. Immunological Reviews *300*, 134–151.

1.2.1 *Epigenetic switches in the context of developmental gene regulatory networks*

Gene regulatory networks (GRN), involving cascades of transcription factor gene activation events, execute lineage and cell-type specification during development. Although GRNs are initiated and guided by extrinsic signals, they often progress in a cell-autonomous manner to establish cell states which are generally capable of retaining their unique properties through future cell generations in absence of the initiating signals. This phenomenon has formed the basis for

ideas of developmentally regulated ‘epigenetic landscapes’ and subsequent ‘epigenetic memory’ (Goldberg et al., 2007; Steffen and Ringrose, 2014) in which the stable differentiated cell states are established as a result of non-genetic changes that are heritable across cell generations. These non-genetic changes are generally the result of changes in gene expression states which can be propagated across cell division even in the absence of the inducing signals. Mark Ptashne first coined the term ‘genetic switch’ to describe this type of phenomenon in phage lambda in which lysogenic genes and lytic genes switch between binary states of expression that are heritable through cell division (Ptashne, 2004). He and others have adapted the term to ‘epigenetic switch’ to emphasize that the heritable binary switch in gene expression states are executed by non-genetic changes within the cell (Ptashne, 2009). These epigenetic switches have traditionally been studied in bacteria and thus have been assumed to be regulated solely by transcription factors and their cognate DNA binding sites which alone generate feedback loops that stabilize changes in gene expression states in response to extrinsic signals. However, as I will explain in the sections below, this is an incomplete picture of epigenetic switching in eukaryotes and would be insufficient for the explaining the unique properties of developmental gene regulatory networks in multicellular organisms.

While studies of epigenetic switching in bacteria have provided a useful quantitative framework for studying fundamental gene regulation principles in single cell organisms, the classical model of epigenetic switching becomes problematic in the context of GRNs controlling multicellular organism development. In response to extrinsic signals, multipotent progenitors do not immediately develop into fully differentiated cell types but instead undergo many rounds of cell division over many days or weeks before fully developing (Gao et al., 1997; Otani et al., 2016). Thus, a critical feature of the developmental GRNs governing this process is that individual genes

within them must frequently be capable switching to active states after long time delays that are tunable across many cell divisions. In a classical model of epigenetic switching, in which transcription factors rapidly bind to DNA and recruit polymerases within seconds to minutes, the activation timing of individual genes would be regulated solely by transcription factor levels and mRNA synthesis rates (Ackers et al., 1982; Alon, 2007; Bintu et al., 2005; Bolouri and Davidson, 2002). In this classical framework, gene activation dynamics are constrained by mRNA and protein stability and thus generally respond to instructive extrinsic signals over timescales shorter than one cell cycle (Levine and Elowitz, 2014). From computational simulations not presented here, we have also found that these epigenetic switches can only generate delays longer than one cell cycle over a very narrow range of input signals. This means that any small deviation in the extrinsic input signal levels would cause the epigenetic switch to occur within one cell cycle or never occur at all. Therefore, a classical model for epigenetic switching, which is only regulated by transcription factor levels, would be incompatible with developmental GRNs that involve long and tunable time delays spanning many cell generations.

1.2.2 *Cis-epigenetic switches: regulation beyond transcription factors*

In eukaryotes, in addition to being regulated by *trans*-acting transcription factors, epigenetic switches can also be regulated by *cis*-acting factors operating at individual gene loci independently. This is best illustrated by the phenomenon of mitotically stable monoallelic expression, in which only one of two gene copies (i.e. alleles) is heritably expressed across many cell divisions (Gendrel et al., 2016). This has historically been studied in the context of X-chromosome inactivation, genomic imprinting, olfactory receptor or antigen receptor gene activation; however, it has more recently been observed across a diverse set of autosomal genes (Rv et al., 2021). Importantly, the expression state of each allele is clonally inherited and thus

distinguishable from other phenomena such as stochastic transcriptional bursting which occurs on timescales much shorter than cell division (Larsson et al., 2019a, 2019b). Furthermore, despite existing in the same nuclear environment with a constant exposure to the same concentration of transcription factors, the two alleles retain opposite expression states and thus demonstrate that additional *cis*-acting factors beyond *trans*-acting transcription factors must be required for this type of epigenetic switch. Therefore, I will refer to these epigenetic switches as "*cis*-epigenetic switches" because each gene copy is additionally regulated independently *in cis.* This is not meant to suggest that *trans*-acting transcription factors are not also important for the switch but instead is meant to emphasize that *cis*-acting factors provide an additional layer of regulation to the switch.

1.2.3 *Potential mechanisms underlying cis-epigenetic switches*

What are the *cis*-acting factors involved in regulating *cis*-epigenetic switches? During mitotically stable monoallelic expression, information about the activation state of each allele must be propagated after DNA replication and cell division. Studies of genomic imprinting, in which only one parental allele is specifically silenced in each cell, have established that DNA methylation and histone 3 lysine 27 tri-methylation (H3K27me3), are critical for the *cis*-epigenetic regulation of the heritably silent allele (Barlow and Bartolomei, 2014; Chen and Zhang, 2020). DNA methylation and H3K27me3 are uniquely suited for regulating *cis*-epigenetic switches because these covalent modifications can be restored at the same genomic location after DNA replication by methyltransferases (Margueron and Reinberg, 2011; Petryk et al., 2021). In the case of DNA methylation, the DNMT1 methyltransferase recognizes hemi-methylaed dinucleotides after DNA replication and converts them to symmetrically methylated dinucleotides (Petryk et al., 2021). In the case of H3K27me3, the Polycomb repressive complex 2 (PRC2) contains one subunit, Eed,

which binds to existing H3K27me3 and another subunit, Ezh2, which is allosterically activated by H3K27me3-bound Eed and methylates newly incorporated histones after replication (Margueron et al., 2009). Therefore, DNA methylation and H3K27me3 are each capable of propagating information about the inactive state of a specific gene locus across cell division as result of read-write positive feedback loops. Furthermore, antagonistic demethylases can actively remove each of the methylation modifications (Williams et al., 2014; Wu and Zhang, 2017) and thus could provide a mechanism for triggering a *cis*-epigenetic switch from heritably inactive to heritably active states. The work presented in this dissertation focuses solely on investigating the role H3K27me3 in regulating a *cis*-epigenetic switch for reasons explained in Chapter 2; however, we do not rule out the possibility that DNA methylation could also contribute to this *cis*-epigenetic switch or others.

1.2.4 *Regulation of cis-epigenetic switches by transcriptional enhancers*

How can individual *cis*-epigenetic switches be selectively regulated in the context of developmental GRNs? The enzymes that regulate DNA methylation and H3K27me3 lack sequence specificity and thus act broadly across the genome. In contrast, non-coding regulatory elements referred to as 'enhancers' have a well-established role in selectively promoting the expression of their target gene(s) *in cis*. Enhancers often serve as binding sites for groups of cell-type specific transcription factors and thus are critical for establishing cell-type specific expression patterns during development and are thought to have played a major role in the evolution of developmental gene networks (Carroll, 2008; Levine, 2010; Long et al., 2016). While some enhancers regulate the amplitude of gene expression within individual cells, others regulate the probability of a binary switch from inactive to active expression states at individual gene loci (Bender et al., 2012; Guy et al., 1996; Ng et al., 2018; Simeonov et al., 2017; Vijayanand et al.,

2012; Walters et al., 1995; Weintraub, 1988). Early studies of the latter type of enhancer found that they often confer position-independent expression of transgenes and suppress position-effect variegation involving the probabilistic and heritable silencing of transgenes depending on the local chromatin environment at their genomic integration sites (Festenstein et al., 1996; Milot et al., 1996; Walters et al., 1996). This suggests that enhancers can mediate all-or-none switching in expression states by counteracting local and heritably repressive heterochromatin (which generally refers to chromosomal regions that are densely packaged by chromatin-associated proteins and assumed to be transcriptionally inactive for reasons explained in *Chapter 2.2.4*). This probabilistic switching in expression states at individual gene loci after long time delays spanning multiple cell divisions is a general feature of *cis*-epigenetic switches thought to be regulated by all-or-none switches in chromatin states (i.e. loosely packaged and 'accessible' or densely packaged and 'inaccessible') (Berry et al., 2015; Bintu et al., 2016; Dobrinić et al., 2020; Mariani et al., 2010). Although mechanisms remains inconclusive, enhancers may counteract repressive heterochromatin states by recruiting transcription factors capable of binding to densely packaged genomic regions (Iwafuchi et al., 2020; Iwafuchi-Doi and Zaret, 2014) and localizing enzymes which modify covalent histone modifications (e.g. H3K27me3) (Estarás et al., 2013; Park et al., 2014; Seenundun et al., 2010; Vernimmen et al., 2011; Williams et al., 2014) or induce conformational changes by displacing or altering the structure of histone-DNA complexes (i.e. nucleosomes) (Kadoch et al., 2017; Zhou et al., 2016).

Importantly, enhancers that promote a binary switch in activation states do so stochastically at individual gene loci over time. Thus, these types of enhancers can be referred to "timing enhancers" because their presence accelerates the activation time of their target gene in an all-or-none manner (**Figure 1.1**). In contrast to "amplitude enhancers" which regulate gene expression

levels uniformly across a population of cells, timing enhancers can alter the fraction of cells that switch to a discrete expression state. Importantly, timing enhancers, which harbor transcription factor binding sites, are responsive to changes in transcription factor inputs such that increasing/decreasing transcription factor concentration or the number of binding sites within the enhancer results in increases/decreases in the activation time of the target gene (Dufourt et al., 2018; Ng et al., 2018). Thus, timing enhancers integrate *trans*-acting information in the form of transcription factors to finely regulate *cis*-epigenetic switching in a cell-context dependent manner. Furthermore, when implemented at lineage-specifying genes, timing enhancers can alter the pace of a developmental cell fate decisions and the numbers or types of cells specified (Gérard et al., 1997; Simeonov et al., 2017; Zheng et al., 2010), an idea that I will explore in more detail in *Chapter 4*.

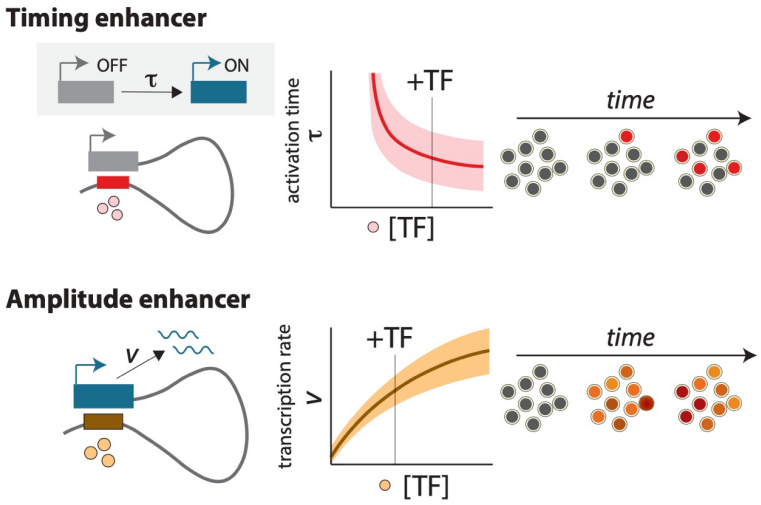


Figure 1.1 Timing enhancer versus amplitude enhancer

(A) A timing enhancer alters the activation time (τ) for a gene locus to switch from an inactive to an active expression state. An increase in transcription factor concentration [TF] shortens activation time as its primary action. A timing enhancer is predicted to produce stable subpopulations of cells with discrete gene expression states. Modulations in timing enhancer activity change the probability that these subpopulations arise over time, without affecting gene expression magnitude. **(B)** An amplitude enhancer alters the transcription rate (v), which measures transcript production as a result of RNA polymerase II loading and elongation. An increase in [TF] increases transcription rate. An amplitude enhancer is predicted to give rise to a single population of cells with graded changes in expression magnitude. Modulations in amplitude enhancer activity change expression magnitude without altering timing.

1.3 EARLY THYMOCYTE DEVELOPMENT AS A MODEL SYSTEM

1.3.1 *Overview of T cell development*

In mice, the thymus exports over a million mature T cells every day (Scollay et al., 1980). This is the product of a continuous yet highly dynamic developmental process ongoing throughout an organism's life. From the ~160 bone marrow derived hematopoietic progenitors seeding the thymus every day, ~100 million immature $CD4^+/CD8^+$ "double-positive" (DP) thymocytes will be generated before the vast majority of which are culled during the process of T cell receptor (TCR) selection (Krueger et al., 2017). This dramatic nearly million-fold expansion in cell numbers is critical for generating a large pool of diverse TCR receptors and results from a preceding slow developmental process that can take several weeks to complete (Krueger et al., 2017). Upon entering the thymus, multipotent progenitors do not immediately commit to the T cell lineage but instead progress through a series of multipotent developmental stages while simultaneously migrating through the thymus and expanding in cell number (Manesso et al., 2012; Petrie and Zúñiga-Pflücker, 2007). Each of these stages is marked by the expression a unique subset of cell surface markers which have enabled for exceptionally fine molecular and functional interrogation of these early developmental transitions and the associated lineage-restriction events (**Figure 1.2**) (Yui et al., 2010). These $CD4^-/CD8^-$ "double-negative" (DN) stages begin with the early T cell lineage progenitors (ETP; $Kit^+/CD44^+/CD25^-$), which retain the most potential for divergence towards myeloid lineages but can also generate innate lymphoid cells (Rothenberg et al., 2016). After spending up to 12 days at the cortico-medullary junction, these cells simultaneously transition to the DN2a stage ($Kit^+/CD44^+/CD25^+$) and migrate to the thymic cortex where they will transition to the DN2b stage ($Kit^{int}/CD44^+/CD25^+$) roughly 2-3 days later (Krueger et al., 2017;

Petrie and Zúñiga-Pflücker, 2007). It's generally been thought that alternative non-T cell lineage potential is first relinquished during this DN2a to DN2b transition (Kueh et al., 2016); however, recent studies suggest that potential for the innate lymphoid cell lineage may exist, albeit with diminished capacity, into the DN3 stage ($Kit^-/CD44^-/CD25^-$) (Qian et al.). DN3 progenitors then migrate to the outer cortex where the *TCRβ* locus will undergo recombination and the choice between TCRαβ and TCRγδ T cell lineages will be made. DN3 progenitors will then migrate to the subcapsular zone where they briefly pass through the DN4 stage ($Kit^-/CD44^-/CD25^-$) before expressing CD8 [immature single-positive (SP) thymocytes] and then CD4 expression [immature double-positive (DP) thymocytes]. DP thymocytes will then migrate back through the cortex, undergo positive selection and silence CD4 or CD8 to become mature SP thymocytes, which will undergo negative selection in the outer medulla prior to exiting the thymus (Petrie and Zúñiga-Pflücker, 2007).

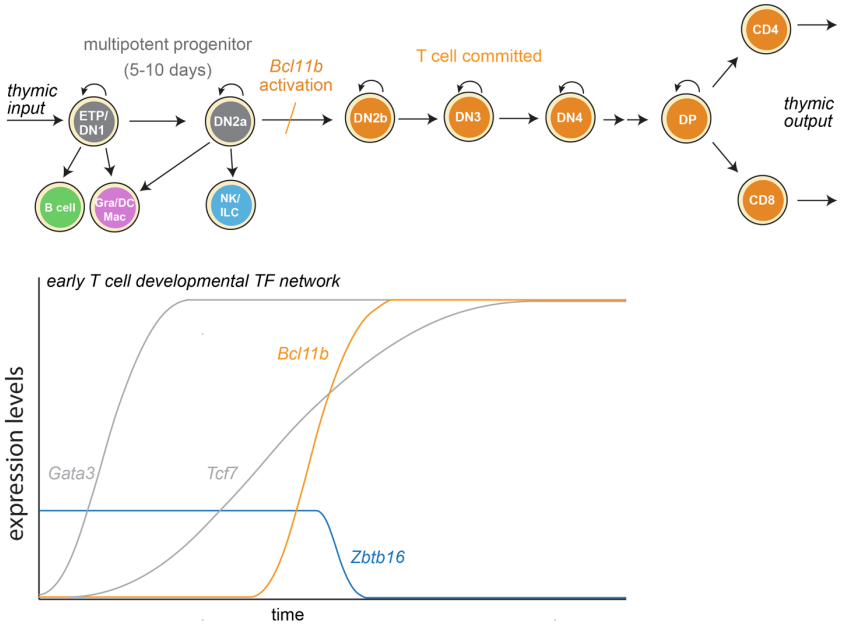


Figure 1.2 Early T cell development and transcription factor gene network
Schematic of T cell development in the thymus (top) with gene expression levels of transcription factors with critical roles in specifying thymic T cell (orange) and ILC (blue) lineages. Expression levels represent approximate percentage of maximum expression during process (note: *Zbtb16* is normally repressed before it can reach its maximum expression).

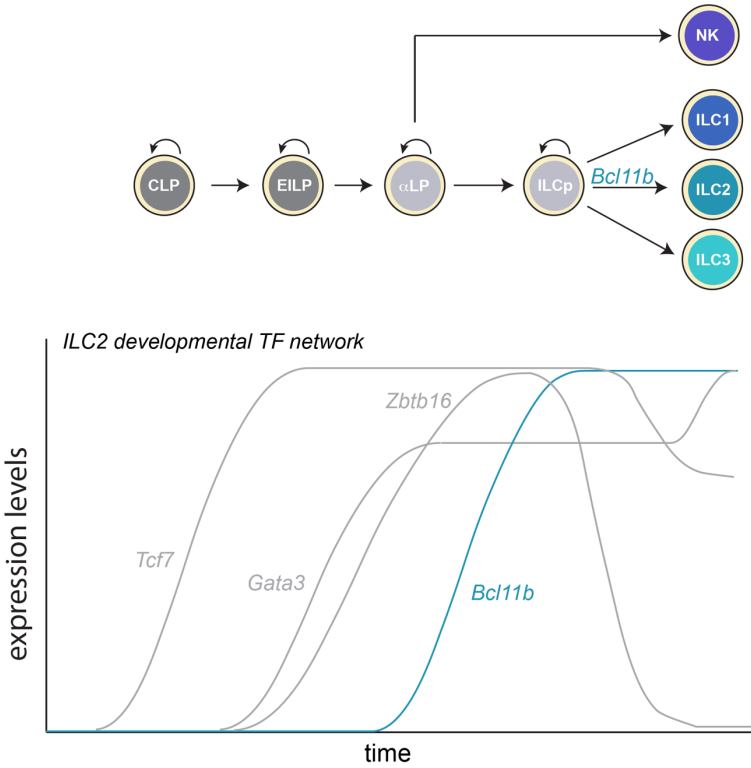


Figure 1.3 ILC2 development and transcription factor gene network
Schematic of ILC2 development from common lymphoid progenitor (CLP) in the bone marrow (*top*) and expression levels of transcription factors with important roles in ILC2 differentiation (note: unlike in T cell development, *Zbtb16* is up-regulated to its maximum before *Bcl11b* is expressed). Adapted from (Huang et al., 2017).

Amazingly, all of the early developmental transitions (ETP-DP) can be studied *in vitro* by co-culture of immature ($Lin^{-}Kit^{+}$) bone marrow progenitors on OP9-DL1 stromal cells. The OP9-DL1 stromal cell line ectopically expresses the Notch ligand, Delta-like 1 (Dll1), which in combination with exogenous IL-7 and Flt3-L, are capable of recapitulating the critical signaling components of the thymus to simultaneously induce self-renewal and development of T cell progenitors from immature bone marrow progenitors (Holmes and Zuniga-Pflucker, 2009). Importantly, the earliest stages (ETP-DN3), during which lineage-restriction events take place, are robustly recapitulated *in vitro* amid a constant signaling environment. Thus, although initiated by extrinsic signals, these early developmental transitions are regulated by cell-autonomous mechanisms which unfold over the course of many days and cell divisions (~10 days and cell divisions on average) (Manesso et al., 2012). This makes early thymocyte development an excellent model system for studying cell-autonomous regulation of developmental timing. Not only can early T cell progenitors be precisely isolated at each stage of development for molecular analysis, but they can also be re-cultured *in vitro* for functional analysis. Thus, unlike traditional developmental models in which this is generally not possible, the developmental potential and dynamics of early thymocyte development can be uniquely interrogated.

1.3.2 *The early T cell developmental gene regulatory network*

Delta-like 1 (Dll1), normally expressed by thymic stromal cells, triggers Notch signaling which is essential for initiating a developmental gene regulatory network that will unfold over the course of weeks (Yui and Rothenberg, 2014). During this time, the genome undergoes dramatic changes in transcription, covalent histone modification distribution, and three-dimensional chromosomal structural and spatial organization (Hu et al., 2018; Zhang et al., 2012). While the Notch intracellular domain (Notch-IC) and its coactivator RBPJ rapidly enter the nucleus and bind to a

suite of target genes, these Notch-target genes are induced at different times during development (Yui and Rothenberg, 2014). Thus, Notch signaling serves as a constant input but requires additional factors which collaborate in driving T cell development and lineage commitment. During the ETP stage, Notch-IC/RBPJ directly bind and activate *Tcf7* (encoding Tcf1) and directly or indirectly activate *Gata3*. During the ETP and DN2a stages, Tcf1 and Gata3 are responsible for establishing a transcriptional landscape suitable for both self-renewal and developmental progression to the T cell lineage (Rothenberg, 2019). There is evidence that Tcf1 and Gata3 both serve as "pioneer" transcription factors which can uniquely bind to densely packaged heterochromatin regions and promote the binding of other transcription factors by inducing DNA bending or recruiting chromatin remodeling enzymes (Fernandez Garcia et al., 2019; Giese et al., 1995; Johnson et al., 2018; Zaret and Carroll, 2011). During the DN2a stage, these factors both directly bind and promote the switch-like activation of *Bcl11b*, which encodes a transcription factor required for T cell lineage commitment and progression beyond the DN2 stage. In contrast to Tcf1 and Gata3, ablation of *Bcl11b* does not affect cell survival or self-renewal during early thymocyte development but instead results in a perpetually self-renewing population of developmentally arrested DN2-like progenitors (Li et al., 2010a). Bcl11b-deficient progenitors retain expression of hematopoietic stem and progenitor genes and disrupting *Bcl11b* expression at later stages can lead to the up-regulation of alternative lineage-specifying genes associated with innate lymphoid cells (ILC) (Li et al., 2010b; Longabaugh et al., 2017). Driven by a shared set of transcription factors, ILC and T cell progenitors share a remarkably similar differentiation into sub-lineages [T cell: CD8 cytotoxic T cells and Th1, Th2, and Th17 helper cells; ILC: natural killer (NK) cells, ILC1, ILC2, and ILC3 helper cells]. Of the ILC sub-lineages, DN2/DN3 progenitors appear to be most primed for the ILC2 lineage based on the co-expression of the

transcription factors Gata3, Tcf1 and Bcl11b which are all critical for both T cell and ILC2 development (**Figure 1.2** and **Figure 1.3**) (Huang et al., 2017; Yu et al., 2015). The ultimate step in the divergence of ILC and T cell lineages appears to be the regulation of and by the E proteins HEB and E2A, which when perturbed, results in a dramatic increase in thymic ILC2s (Qian et al.). Bcl11b serves as the penultimate regulator of lineage-specification by directly repressing Id2 which encodes a negative regulator of E protein activity. In the absence of Bcl11b, Id2 inhibits E protein activity and enables the up-regulation of the pro-ILC transcription factor *Zbtb16* (Hosokawa et al., 2018). Paradoxically, Bcl11b is also required for ILC2 development where its binding preferences are altered due to the co-expression of a unique set of transcription factors and consequently it regulates a set of genes distinct from its target genes in T cell progenitors (Hosokawa et al., 2020). Thus, the relative timing of *Bcl11b* activation is likely important for determining which cells will be destined for T cell maturation at the DN3 stage, which is thought to be the critical checkpoint in regulating T cell output (Ramos et al., 2020).

1.3.3 *Bcl11b as a model locus for studying cis-epigenetic switching*

Bcl11b is initially activated at the DN2a stage but is maintained throughout the lifetime of T cell despite dramatic changes in the signaling environment and thus *Bcl11b* can serve as model epigenetic switch that is also functionally critical for a developmental process. Although many of the required transcription factors are present at high levels during early thymocyte development at the ETP and DN2a stages, the *Bcl11b* epigenetic switch isn't activated until several days and cell divisions later (**Figure 1.2**) (Kueh et al., 2016; Zhou et al., 2019). During the DN2 stage, the *Bcl11b* locus, which resides in a 2 megabase gene desert, undergoes dramatic physical and chemical changes including DNA and H3K27 demethyation, nuclear repositioning to the interior, and increased interactions with distal putative non-coding regulatory elements (Hu et al., 2018;

Isoda et al., 2017; Zhang et al., 2012). Therefore, our group previously hypothesized that one or more of these *cis*-acting factors could be involved in regulating the timing of the *Bcl11b* epigenetic switch. In order to test this, one must be able to detect these *cis*-acting events in isolation from *trans*-acting events occurring in parallel. Inspired by studies of mitotically stable monoallelic expression (see *Chapter 1.2.2*), our group generated a dual-allele reporter mouse in which each of the two *Bcl11b* alleles are non-disruptively tagged with distinguishable fluorescent proteins (Ng et al., 2018). This enabled our group to separately monitor the activation state of each allele during T cell development. Since each allele is exposed to the same nuclear environment, they predicted that if *Bcl11b* activation timing was controlled solely by the concentration of *trans*-acting transcription factors (i.e. the classical model of epigenetic switching), then each *Bcl11b* allele should turn on at roughly the same time within individual progenitors (**Figure 1.4**A, (Chu et al., 2021)). In contrast, if the activation timing was additionally controlled by *cis*-acting factors, then each allele could turn on at different times within individual progenitors. Indeed, they found that individual thymocytes at the DN2 stage frequently expressed *Bcl11b* from a single allele but nearly all thymocytes had expression from both alleles at the DN3 stage and beyond.

These results suggested that *cis*-acting factors were indeed regulating the timing of *Bcl11b* activation. Nonetheless, it remained unclear if this was a consequence of epigenetic regulation heritable across cell generations or whether this was a result of dynamically unstable expression states at the DN2 stage. Therefore, they isolated DN2a thymocytes negative for both *Bcl11b* fluorescent reporters, co-cultured them with OP9-DL1 stromal cells and monitored the two allelic fluorescent reporters in clonal lineages by live cell imaging. They found that developing progenitors frequently expressed one allele multiple days and cell divisions before expressing the other. This demonstrates that the timing of *Bcl11b* is regulated by *cis*-acting factors which function

at each allele independently and across cell division -thus making this a *cis*-epigenetic switch. Furthermore, this dual-allele approach also revealed that the *Bcl11b cis*-epigenetic switch was regulated by a timing enhancer which, when removed from a single allele, delayed *Bcl11b* activation by ~3 days compared to the unperturbed allele in the same cell (**Figure 1.4**B, (Chu et al., 2021))

In contrast to *cis*-epigenetic regulation of genomic imprinting but similar to the *cis*-epigenetic switches regulating X-inactivation, olfactory genes, and antigen receptor genes (Magklara and Lomvardas, 2013), activation of each *Bcl11b* allele occurs stochastically with no parent-of-origin effects. However, unlike regulation of X-inactivation, olfactory genes, and antigen receptor genes, the *Bcl11b cis*-epigenetic switch doesn't involve feedback loops that restrain expression to a single allele in individual cells (i.e. allelic exclusion) and thus both *Bcl11b* alleles will eventually switch to active states when given enough time. This means that instead of regulating gene dosage or antigen specificity, the *Bcl11b cis*-epigenetic switch is instead important for regulating the timing of a lineage-specifying event during development. In the work below, I use the *Bcl11b* and early thymocyte cell development as models to interrogate the molecular mechanisms (Chapter 2 and 3) and physiological implications (Chapter 4) of *cis*-epigenetic switches controlling developmental timing.

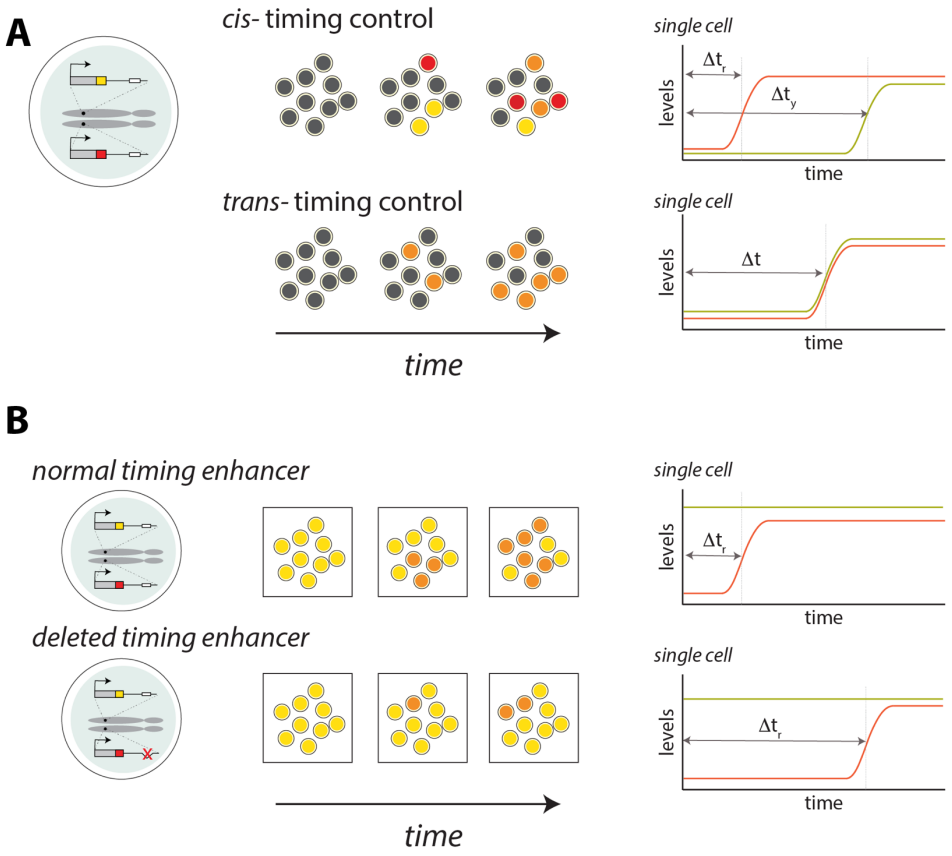


Figure 1.4 Tracking two copies of the same gene in single cells identifies cis-epigenetic switches and timing enhancers
(A) Separately tracking the activation of two gene copies, marked with distinguishable fluorescent reporters, distinguishes cis versus trans control of gene activation timing. When activation is limited by cis events at single loci, the two alleles turn on asynchronously in single cells, with time differences that can span extended timescales. In contrast, when activation is limited by trans events occurring in the nucleus, the two alleles turn on synchronously. (B) Single-allele perturbations of non-coding regulatory elements enable the unperturbed wild-type allele to serve as a same-cell internal control to ensure all trans- factors necessary for gene expression are present. To interrogate the function of a candidate enhancer, single-cells expressing the unperturbed allele (yellow) can be isolated and re- cultured. This allows monitoring of the perturbed allele (red) by live-cell imaging. Single- cell, single-allele tracking enables quantification of the cis-activation timing delay and uncovers the function of the candidate enhancer.

Chapter 2. TUNABLE, DIVISION-INDEPENDENT CONTROL OF GENE ACTIVATION TIMING BY A POLYCOMB SWITCH

Chapter 2 is adapted with minimal modification from the following manuscript:

Pease, N.A., Nguyen, P.H.B., Woodworth, M.A., Ng, K.K.H., Irwin, B., Vaughan, J.C., and Kueh, H.Y. (2021). Tunable, division-independent control of gene activation timing by a polycomb switch. Cell Reports *34*, 108888.

2.1 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 in order to precisely control population sizes of differentiated cell-types. In diverse contexts, cell differentiation delays are generated by division-independent timing mechanisms operating autonomously in single cells (Burton et al., 1999; Gao et al., 1997; Heinzel et al., 2017; Okamoto et al., 2016; Osmond, 1991; Otani et al., 2016) (see also *Chapter 1.1*). 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 tunable, 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., 2016; 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 (Park et al., 2014; Seenundun et al., 2010; Williams et al., 2014). 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 *cis*-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., 2010a) (**Figure 2.1**A). 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) (**Figure 1.2**). 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 *cis*-epigenetic switch acting independently at individual *Bcl11b* loci. This epigenetic switch activates probabilistically, with a multi-day time constant that is controlled by Gata3 and TCF-1 (Kueh et al., 2016), along 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 *cis*-epigenetic switch itself or its basis for generating controllable timing delays in gene activation. We address these outstanding questions here.

2.2 RESULTS

2.2.1 *A timed epigenetic switch delays Bcl11b activation and T cell lineage commitment*

To study the *cis*-acting 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 2.1**B). Using fluorescence-activated cell sorting (FACS), we purified monoallelic *Bcl11b* expressing DN2 progenitors from dual-allelic reporter mice and analyzed 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). Inactive *Bcl11b* alleles turned on after a long time delay such that the fraction of biallelically expressing cells increased progressively over the course of five days (**Figure 2.1**C-D). Activation kinetics were similar for both YFP and RFP alleles 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.

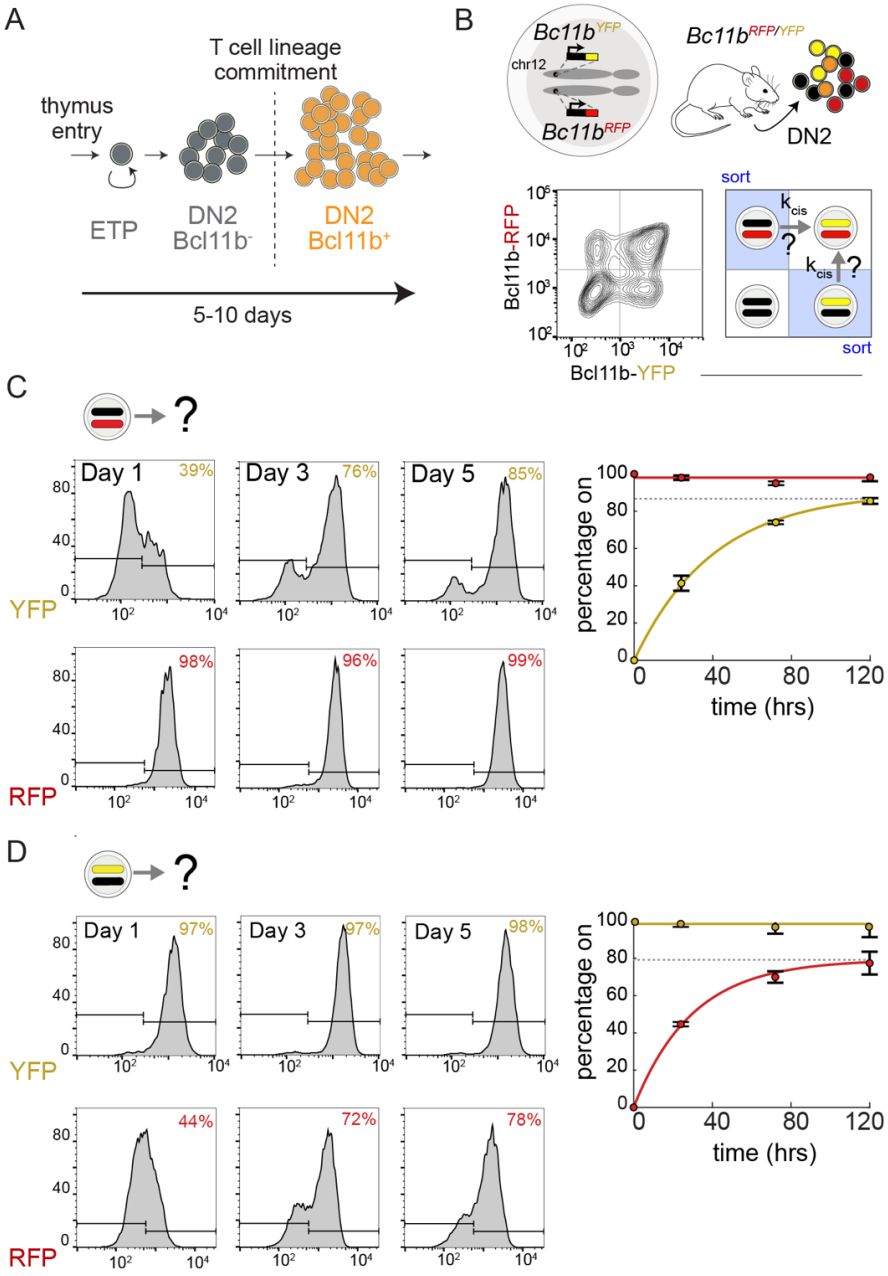


Figure 2.1 A timed *cis*-epigenetic switch delays *Bcl11b* activation and T cell lineage commitment

(A) *Bcl11b*, a transcription factor that drives T cell lineage commitment, turns on with a multi-day time delay. (B) Dual-allelic *Bcl11b* reporter mouse (top), along with flow cytometry plot showing levels of each *Bcl11b* allele in DN2 progenitors (bottom left) and with strategy to purify *Bcl11b* monoallelic expressing progenitors for live-cell analysis of epigenetic switch timing, k_{cis}. (C-D) DN2 monoallelic *Bcl11b* expressing progenitors were purified, cultured on OP9-DL1 feeders with 5 ng/mL IL-7 and Flt3L, and analyzed by flow cytometry. Data represent means and 95% confidence intervals 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 YFP activation and $k = 0.034$ hrs^{-1} +/- 0.009 for RFP activation.

2.2.2 *H3K27me3 levels at the Bcl11b locus tune 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). Therefore, H3K27me3 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 progenitor populations having different numbers of active *Bcl11b* loci. In bone marrow progenitors, where both *Bcl11b* alleles are inactive, there was an abundance of H3K27me3 across the 5' end of *Bcl11b* (**Figure 2.2**A). These broad H3K27me3 peaks were roughly halved in monoallelic *Bcl11b* expressing DN2 progenitors and were almost completely absent in biallelic *Bcl11b* DN2 progenitors. These results show H3K27me3 is lost from the *Bcl11b* locus concurrently with its activation.

H3K27me3 may modulate the timing of *Bcl11b* activation; alternatively, its loss 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) or the H3K27 demethylases Kdm6a/b (GSK-J4) (**Figure 2.2**B), resulted in a ~60% decrease and ~40% increase, respectively, in H3K27me3 abundance at the *Bcl11b* promoter in monoallelic *Bcl11b* expressing DN2 progenitors (**Figure 2.2**C), indicating that they actively modulate H3K27me3 levels at inactive *Bcl11b* loci.

To determine whether H3K27me3 levels regulate *Bcl11b* activation timing, we assayed the expression of inactive *Bcl11b* alleles in DN2 progenitors 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 2.2**D-F). Kdm6a/b demethylase inhibition decreased the activation rate for each silent allele in a graded manner (27-37%). Conversely, Ezh2 inhibition increased the activation rate for each silent allele (41-58%). 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 2.3**A). Furthermore, short hairpin RNA mediated knock-down of another essential PRC2 subunit, *Eed*, also increased the activation rate of silent *Bcl11b* alleles (**Figure 2.3**A,C,D), arguing against non-specific pharmacological effects. Notably, all H3K27me3 perturbations tested altered the fraction of activated cells without altering expression levels in activated cells (**Figure 2.3**A), indicating a specific role of these modifications in tuning gene activation timing. Importantly, perturbing H3K27 methylation rates had no effect on the apoptosis frequency among different populations, indicating that these changes in activation fractions are not due to differential cell death in these populations (**Figure 2.3**E). Taken together, these results show that H3K27me3 levels at the *Bcl11b* locus, set by opposing PRC2 methyltransferase and Kdm6a/b demethylase activities, control the timing of *Bcl11b* activation.

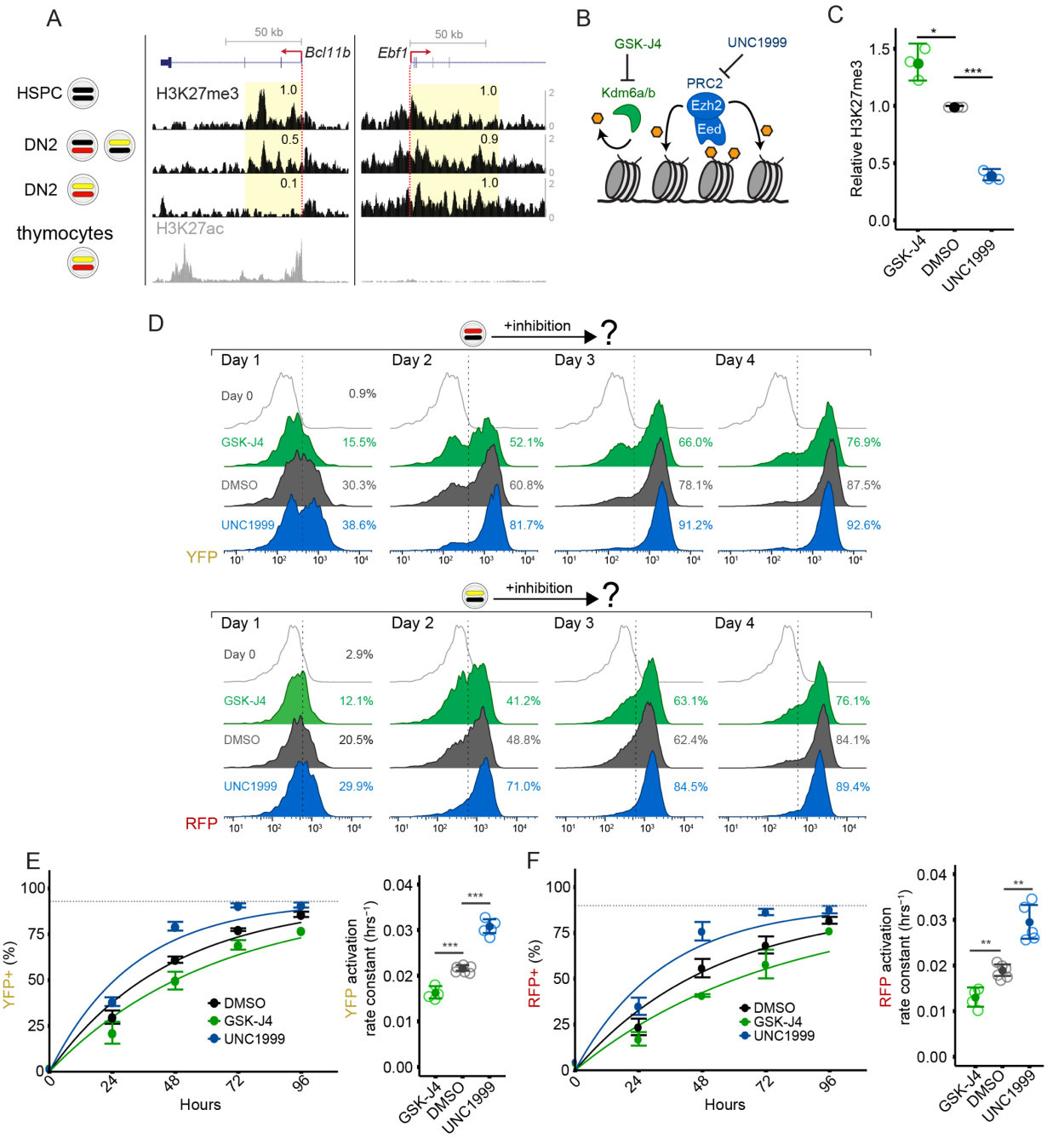


Figure 2.2 H3K27me3 levels, set by PRC2 and Kdm6a/b demethylases, modulate *Bcl11b* activation timing

(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 ENCODE accession number ENCSR000CCH (Davis et al., 2018), demarcate transcribed region. Data are representative of two independent experiments. (B) Schematic depicting inhibition 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 shown for $n = 3$ independent experiments (two-sample t-test, one-tailed: * $p < 0.05$, *** $p < 0.001$). (D) Purified DN2 monoallelic expressing cells were re-cultured with the indicated inhibitors 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 $y = F(1\text{-}e^{-kt})$, where F = maximum percentage of cells positive for assayed allele (represented by the dotted grey lines). (E-F, right) Data represent 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).

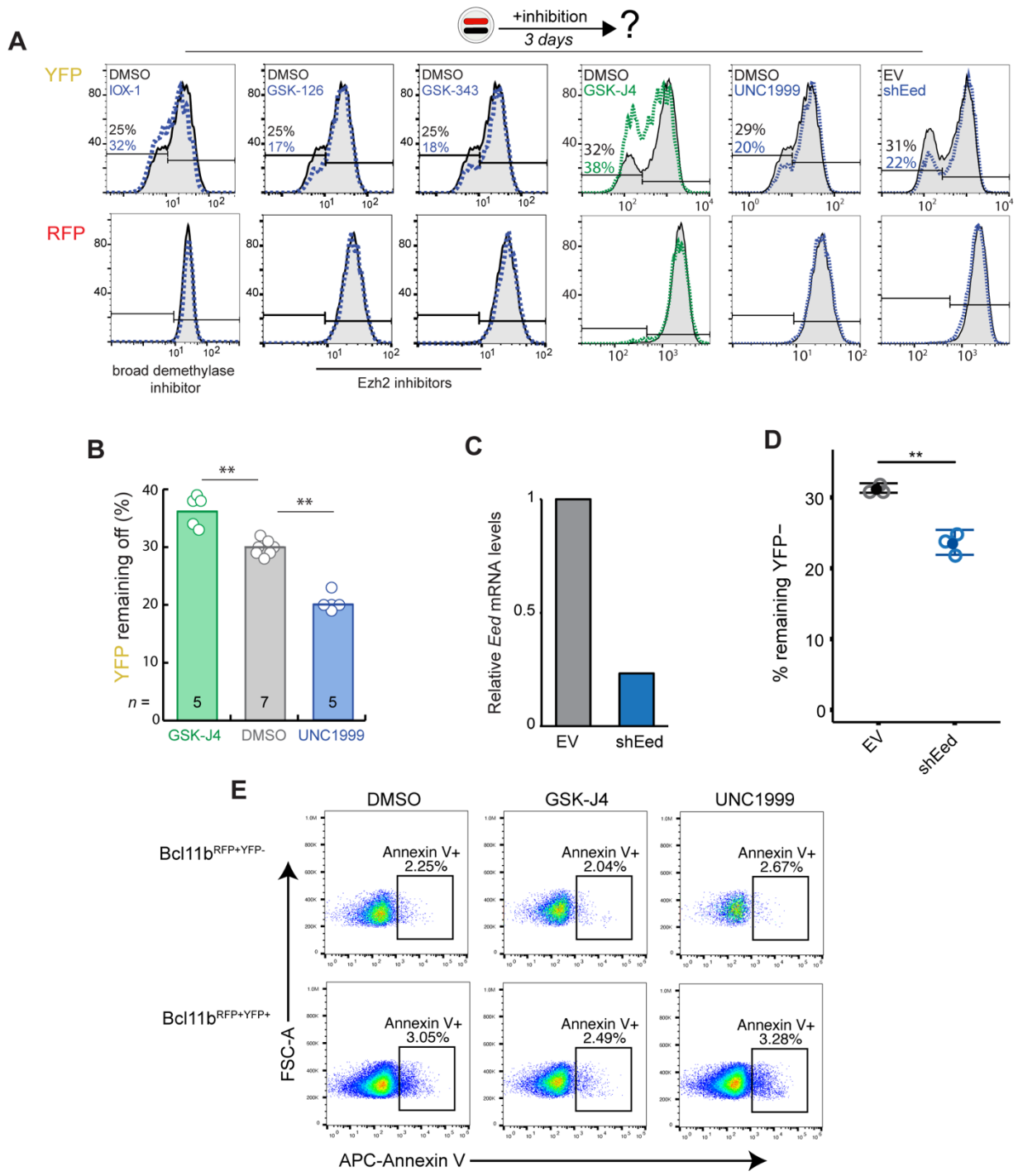


Figure 2.3 Unrelated H3K27me3 perturbations modulate *Bcl11b* activation timing.
(A) DN2 Bcl11b$^{RFP+/YFP-}$ monoallelic expressing cells were sorted, re-cultured on OP9-DL1 cells in the presence of different small molecule inhibitors and analyzed by flow cytometry 3 days later. Structurally unrelated inhibitors show similar effects of decreasing or increasing *Bcl11b* activation probabilities as observed in **Figure 2.2** (left). All histograms show that while the perturbations affect the all-or-none activation probability for the initially inactive alleles (top), the perturbations have no effect on the expression maintenance nor magnitude of the initially active alleles (bottom). (B) Mean percentage of cells remaining Bcl11b YFP^{-} (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 Bcl11b$^{RFP+/YFP-}$ monoallelic cells were transduced with retroviral constructs and precultured for 3 days (two-sample t-test, one-tailed, **p < 0.01, n = 3 independent experiments, error bars = 95% confidence interval). (E) Progenitors from Figure 2D (day 3) were stained with Annexin V to detect apoptotic cells.

2.2.3 *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; Jadhav et al., 2020; Jiang and Berger, 2017). However, the observed regulation of *Bcl11b* activation timing by H3K27 demethylases that counteract PRC2-mediated methylation (**Figure 2.2**) suggests an active mechanism is involved that could potentially operate independently of cell division. To determine whether *Bcl11b* activation timing depends on cell division, we accelerated the cell division rate in DN2 RFP^{+}/YFP^{-} progenitors by transducing them with the proto-oncogene c-Myc and used quantitative live-cell imaging to measure the activation kinetics of the silent YFP allele (**Figure 2.4**A). As expected, c-Myc overexpression resulted in a ~two-fold increase in the cell division rate (**Figure 2.4**B; **Figure 2.5**; **Table 2.1**). However, despite accelerating cell division, c-Myc overexpression did not change *Bcl11b* activation timing, with control and accelerated progenitors activating the silent YFP allele with the same exponential time constant (~136 hrs, **Figure 2.4**C). This lack of change in activation timing following cell division acceleration was also observed by flow cytometry for silent RFP alleles (**Figure 2.4**D). 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.

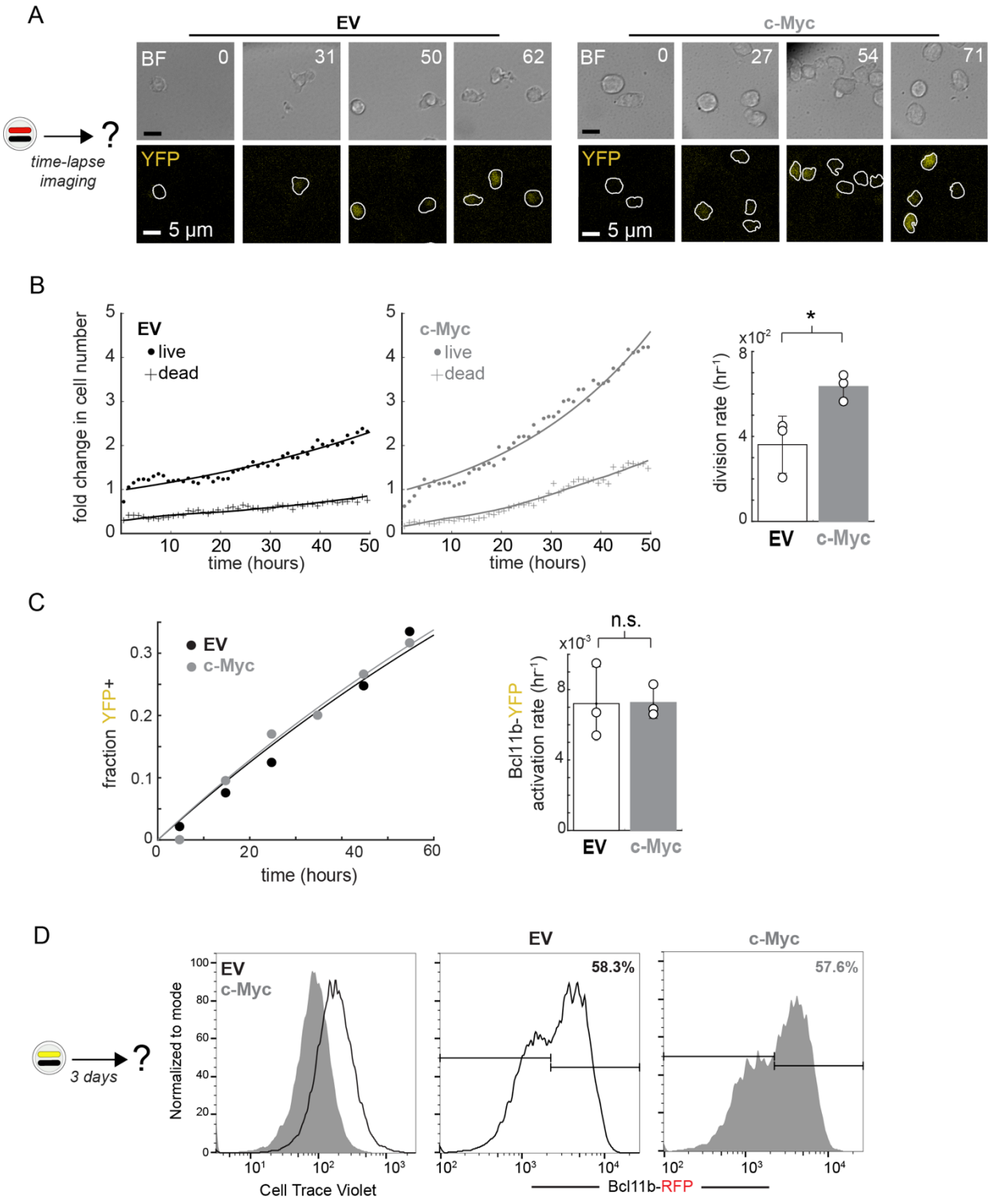


Figure 2.4 *Bcl11b* activation timing is independent of cell division speed
Bcl11b$^{YFP-/RFP+}$ DN2 progenitors transduced with either an empty vector (EV) or c-Myc over-expressing retroviral vector were purified, re-cultured on DL1-coated plates and monitored by timelapse imaging. (A) Timelapse images. White boundaries show automated cell segmentation. Numbers show elapsed time in hours. (B, left) Time evolution of live and dead cell numbers. Data

was fitted to a population dynamics model as shown in Figure S2. (B, right) Data represent mean and standard deviation of cell division rates for $n = 3$ independent experiments (two-sample t-test, one-tailed, * $p < 0.025$). (C, left) Fraction of YFP+ cells over time. (C, right) Data represent mean and standard deviation of *Bcl11b-YFP* activation rates for $n = 3$ independent experiments, (two-sample t-test, one-tailed, n.s. not significant). (D) $Bcl11b^{RFP+/YFP-}$ DN2 progenitors transduced with either EV or c-Myc were re-cultured on OP9-DL1 stromal monolayers for 3 days before analyzing by flow cytometry.

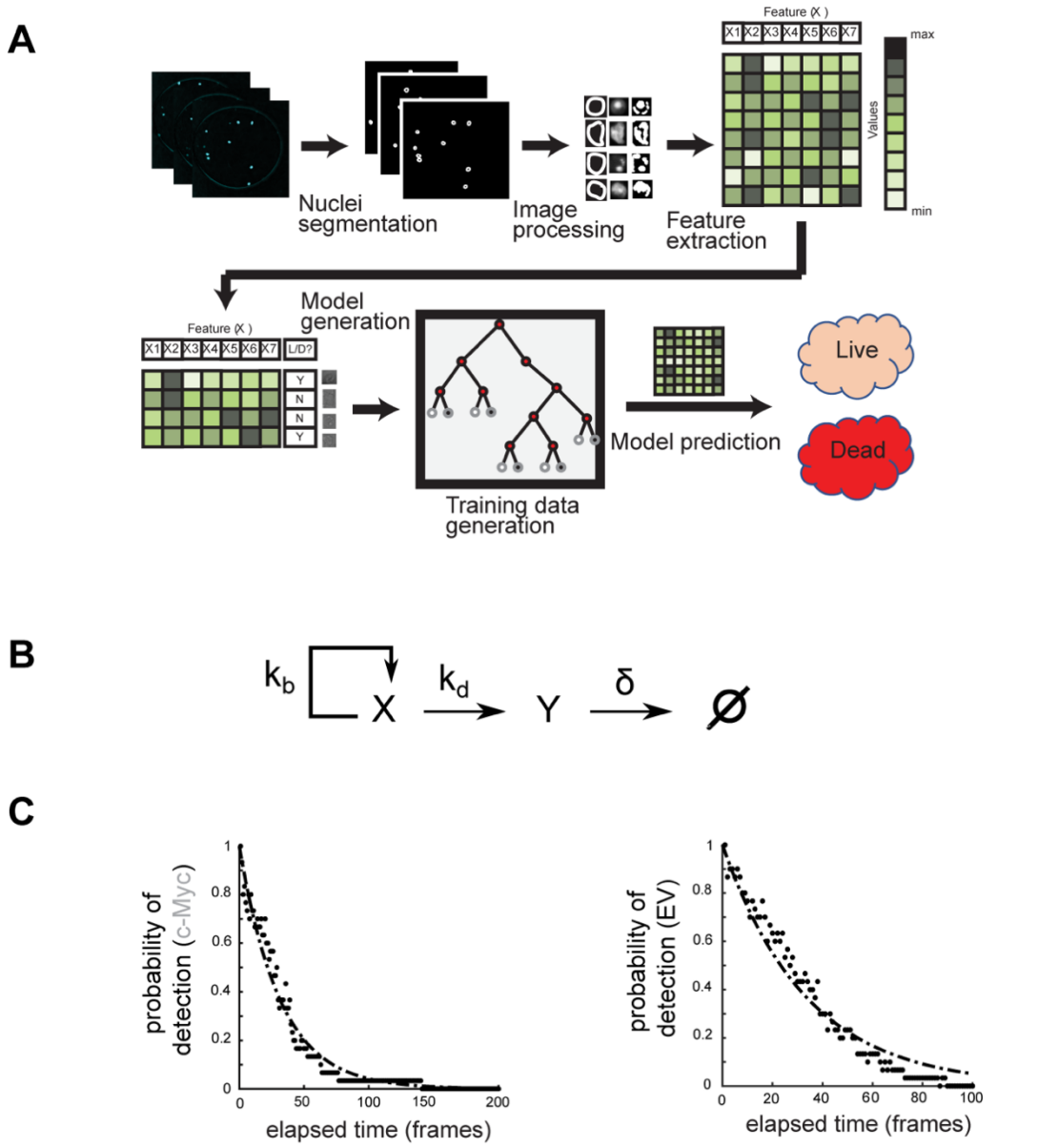


Figure 2.5 Image processing pipeline for identification of live and dead cells in imaging experiments

(A) Cyan fluorescent protein (CFP) signals from transduced cells are used for segmentation of individual cells. Segmented populations were analyzed using image 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 Bcl11b RFP^+/YFP^- cells transduced with c-Myc and empty vector (EV). The model includes a population of live cells (X) with a birth rate k_b and a death rate k_d 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 individual 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 $\delta = 0.030$ per frame for EV population.

	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

Table 2.1 Doubling and death rates by live cell imaging
Tabulated doubling (K) and death (k_d) rates calculated from data fitting of live and dead populations from three independent imaging experiments. Data was fit to population dynamics model described in Statistical and Quantitative Analysis section.

2.2.4 *A methylation-compaction mechanism for tunable, division-independent timing control*

The molecular mechanism underlying the timed *Bcl11b cis*-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

used mathematical modeling to analyze a series of candidate mechanisms. H3K27me3 can bind PRC2 at an allosteric site and activate its methyltransferase activity (Margueron et al., 2009), allowing these modifications to spread to neighboring nucleosomes and be maintained across cell division. Previous work has shown that such positive feedback, mediated by a read-write mechanism for histone methylation, can generate bistable switching that occurs over timescales spanning many cell generations (Dodd et al., 2007; Zhang et al., 2014). Therefore, we first analyzed a simple model of this positive feedback mechanism. In this methylation read-write (M) model, the *Bcl11b* locus comprises a linear array of *N* nucleosomes, each of which can be methylated or demethylated (**Figure 2.6**A). Each nucleosome is methylated at a rate that increases with the number of nearby methylated nucleosomes and is demethylated at a first-order rate. During DNA replication, nucleosomes randomly segregate to the two daughter strands, resulting in half of the methylated nucleosomes being replaced by demethylated ones (Coleman and Struhl, 2017).

Our simulations revealed that single nucleosome arrays could switch from a repressed H3K27 methylated state to a demethylated state with stochastic time delays spanning multiple cell divisions (**Figure 2.6**A, center), consistent with previous work (Dodd et al., 2007; Zhang et al., 2014). However, in our simulations, activation timing was extremely sensitive to H3K27 methylation levels in the silent state, with minor changes in methylation levels (~10%) causing dramatic changes activation timing (~300 fold) (**Figure 2.6**B). This extreme sensitivity far exceeded the sensitivity coefficient observed experimentally (**Figure 2.2**C-F, **Figure 2.6**B) across a wide range of methyltransferase reach lengths (**Figure 2.7**A-B) and was also seen in other studies (Dodd et al., 2007; Zhang et al., 2014), indicating that it represents a general feature of positive-feedback mediated switching and not a specific aspect of our model. By analyzing this system

using a transition state theory framework (**Figure 2.7**C-E, see also Mathematical Appendix), 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 2.2**C-F).

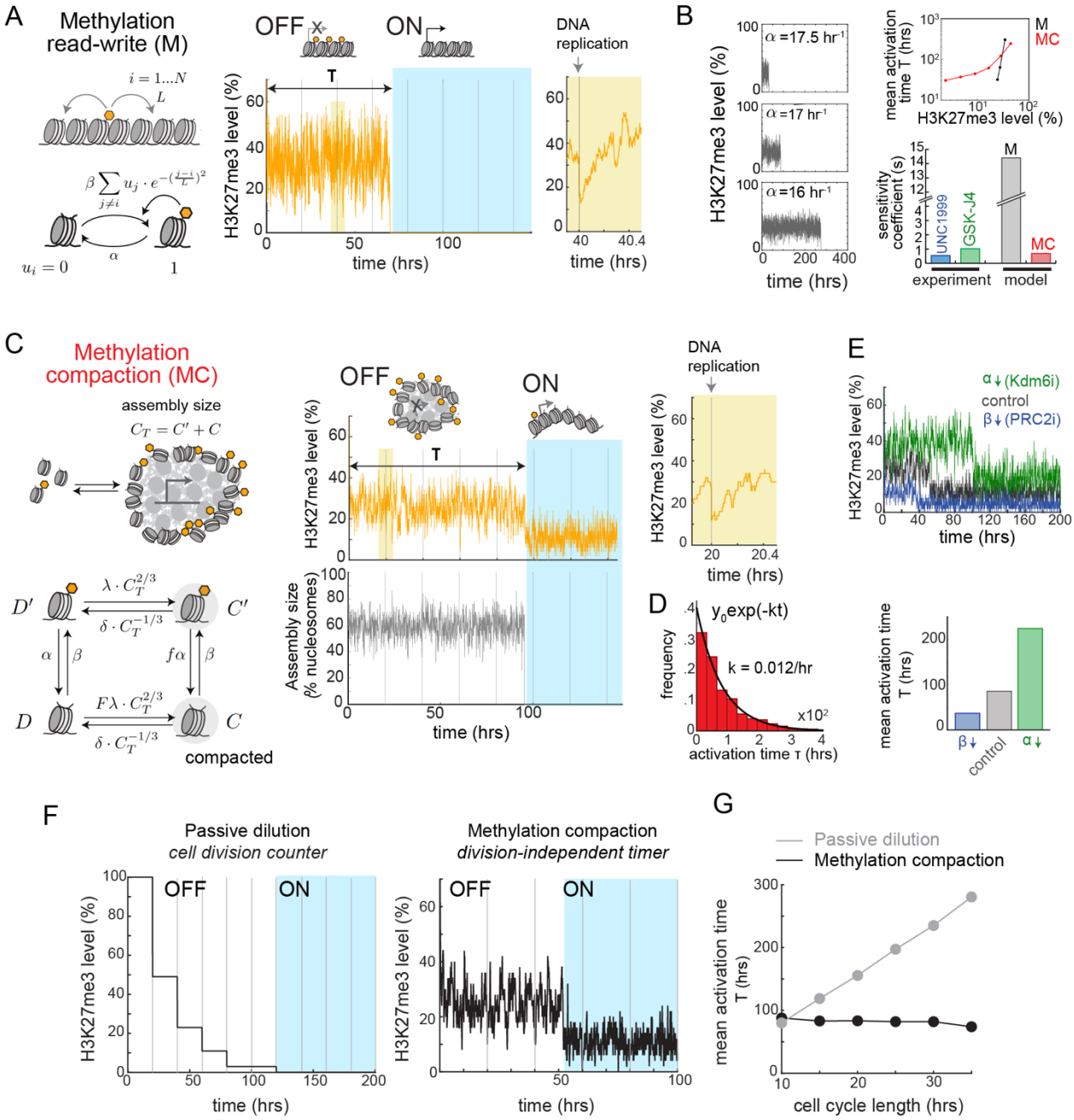


Figure 2.6 A methylation-compaction switching mechanism generates tunable, division-independent delays in gene activation

(A) Methylation read-write (M) model (left), along with representative simulation (right). (B) Representative simulations of M model with different demethylation rates α (left). Mean activation times against H3K27me3 levels (top right) and sensitivity coefficients for this relationship (bottom right). (C) Methylation compaction (MC) model (left), along with representative simulation (right). (D) Histogram shows distribution of activation times, along with exponential fit. (G) Representative simulations of MC model, simulating PRC2 or Kdm6a/b inhibition (top), along with mean activation times (bottom). (F) Simulations of passive dilution and MC models. Vertical lines indicate DNA replication events. (G) Mean activation times as a function of cell cycle length.

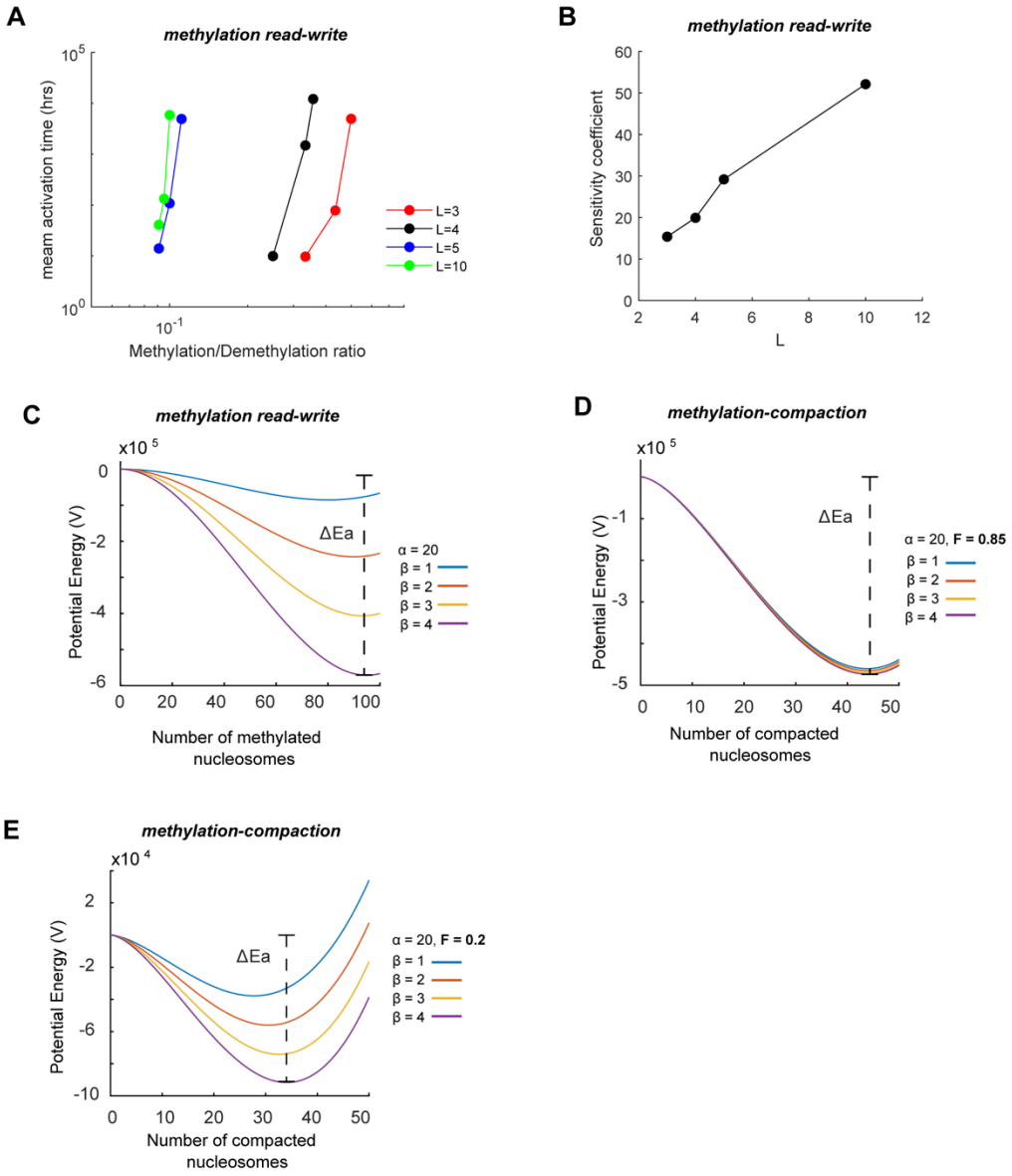


Figure 2.7 Activation energy is robust to changes in methylation rate when interaction affinities between methylated and demethylated nucleosomes are similar

(A) Mean activation time as a function of the methylation/demethylation ratio derived from methylation read-write model simulations while decreasing the number of nucleosomes within 'reach' of a the PRC2 complex, L. (B) Sensitivity coefficient (ΔlogY/ΔlogX) as a function of L. (C) Potential energy landscapes of methylation read-write model. (D-E) Potential energy landscapes of the methylation-compaction model. Parameter F dictates how sensitive 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 parameter as indicated by the curve's color see Mathematical Appendix).

H3K27me3 modifications repress gene expression by promoting the formation of compacted chromatin assemblies inaccessible to the transcription machinery. Compaction may occur through H3K27me3-dependent recruitment of Polycomb repressive complex (PRC1), which can self-associate through intrinsically disordered domains on its Cbx2 subunit to promote chromatin condensation (Plys et al., 2019; Tatavosian et al., 2018); alternatively, compaction may involve direct interactions between nucleosomes that are modulated by the H3K27me3 state (Gibson et al., 2019; Sanulli et al., 2019). In both cases, weak multivalent interactions between nucleosomes drive chromatin condensation through phase separation.

In light of these recent findings, we developed a second model, where H3K27me3 does not directly repress transcription *per se*, but instead enhances the strength of nucleosomal interactions to keep the locus compacted and restrain its activation (**Figure 2.6**C; see Mathematical Appendix). In this methylation-compaction (MC) model, the *Bcl11b* locus consists of an ensemble of nucleosomes that can exist in methylated or demethylated states and can be included in or excluded from a compacted nucleosome assembly. Methylation and demethylation of nucleosomes occur with first order rates, whereas association and dissociation of nucleosomes with the compacted chromatin assembly occur at rates proportional to its surface area, which we assume scales with the two-thirds' power of the number of nucleosomes in the assembly, C_T. This rate dependency assumes that nucleosomes enter (or exit) the compacted assembly only by formation (or breakage) of weak multivalent interactions with nucleosomes on the assembly surface. Methylation and compaction are coupled such that the methylation state of a nucleosome affects its rate of association with a compacted assembly, and vice versa. Below a threshold number of nucleosomes, the chromatin assembly is unstable and dissolves, leading to locus decompaction and consequent

gene expression. This stability threshold reflects a minimum nucleus size needed to maintain a phase-separated condensate.

Simulations of the methylation-compaction model revealed 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 2.6**C, 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 2.6**D). However, in contrast with the methylation read-write model but in concordance with our experimental results (**Figure 2.2**), changing H3K27me3 levels by varying methylation or demethylation rates changed gene activation timing in a much more graded manner (**Figure 2.6**B, E). This tunability was robust over different parameter ranges (**Figure 2.6**B, top right), different degrees of cooperativity for H3K27 methylation (**Figure 2.8**) and different degrees of assembly disruption after DNA replication (**Figure 2.9**A-E). A transition state theory analysis (see Mathematical Appendix and **Figure 2.7**C-E) showed that in order for gene activation timing to be finely tunable, nucleosomes must be able to associate with each other even without H3K27me3 modifications, such that methylated and demethylated 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 (Francis et al., 2004; Gibson et al., 2019; Larson et al., 2017; Sanulli et al., 2019; Strom et al., 2017).

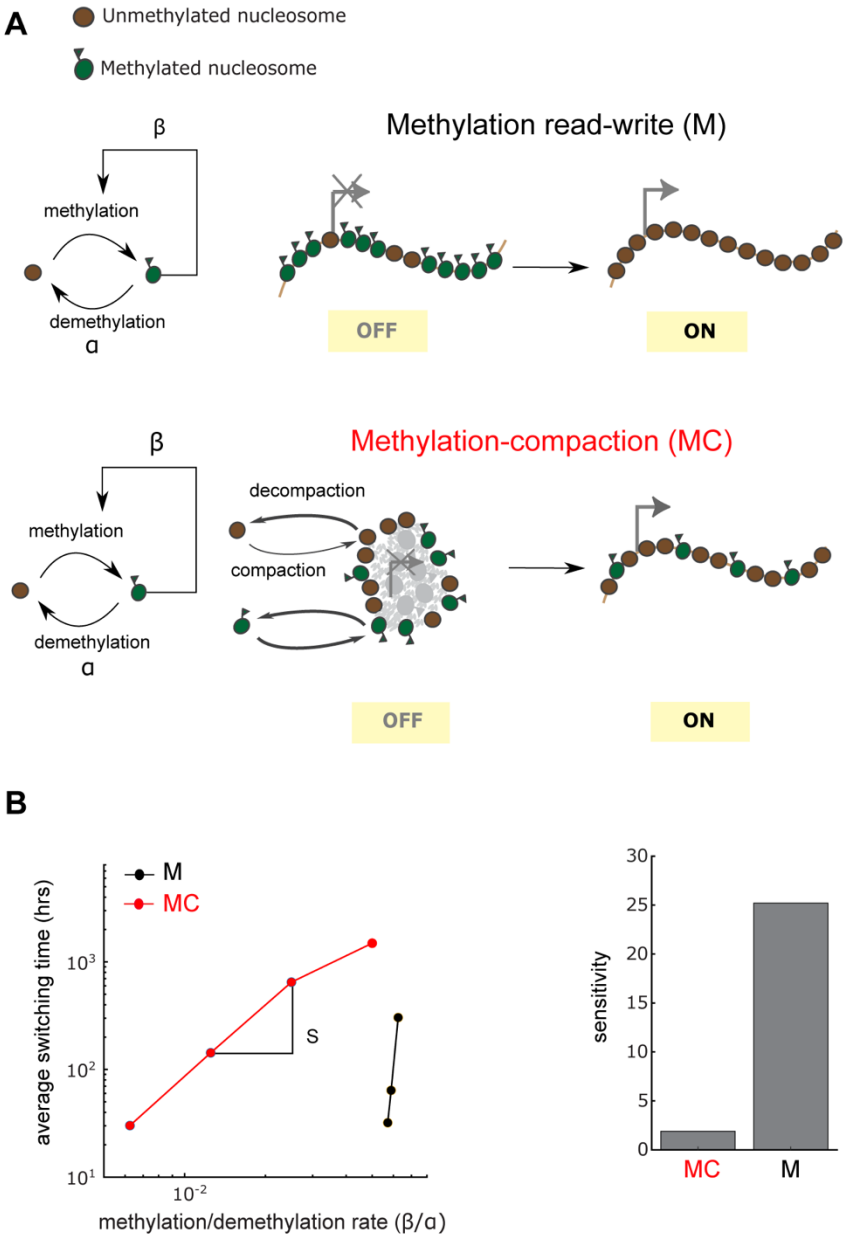


Figure 2.8 Coopertivity in methylation-compaction model increases switching time tunability compared to methylation read-write model

(A, top) Methylation read-write 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 times as a function of methylation (β) and demethylation rate constant (α) ratio for the methylation read-write (black) and methylation-compaction (red) models. Sensitivity coefficient (ΔlogY/ΔlogX) for each plot was calculated by taking the slope of the linear fit y = ax+b for the methylation model data set and the last 5 data points for the compaction model.

Because cell division rates did not affect timing delays in *Bcl11b* activation (**Figure 2.4**), we tested whether the methylation-compaction mechanism also generates cell division-independent activation delays. Indeed, in contrast to a passive dilution model for H3K27me3 loss (**Figure 2.6**F), the methylation-compaction model generated activation delays that were constant over a range of cell cycle speeds (**Figure 2.6**G, **Figure 2.9**A-E). We note that this cell division independence was lost when histone methylation and demethylation rates were reduced to be slower than the cell division rates (**Figure 2.9**F), implying a need for rapid histone methylation dynamics for upholding cell cycle independence. Indeed, histone methylation and nucleosome compaction dynamics occur with reported timescales of minutes and seconds, respectively (Kristensen et al., 2011; Larson et al., 2017; Sneeringer et al., 2010). This explains why the epigenetic state recovers rapidly after DNA replication as observed in our simulations (**Figure 2.6**A-B, right).

These modeling results suggest that timing control by H3K27me3 has the following characteristics: First, H3K27me3 loss does not directly result in gene activation, but instead modulates a separate process that acts as the gatekeeper for gene transcription. Given recent insights into how nucleosomes can interact to form phase-separated structures (Gibson et al., 2019), and how these interactions can be modulated by histone tail modifications and/or binding proteins (Larson et al., 2017; Plys et al., 2019; Strom et al., 2017; Tatavosian et al., 2018), we propose that H3K27me3 loss weakens nucleosomal interactions at the *Bcl11b* locus, promoting gene locus decompaction and gene activation. Second, the compaction process itself must be partially independent of H3K27me3 such that the compacted nucleosome assembly can be maintained by other chromatin-associated proteins (Francis, 2004; Larson et al., 2017).

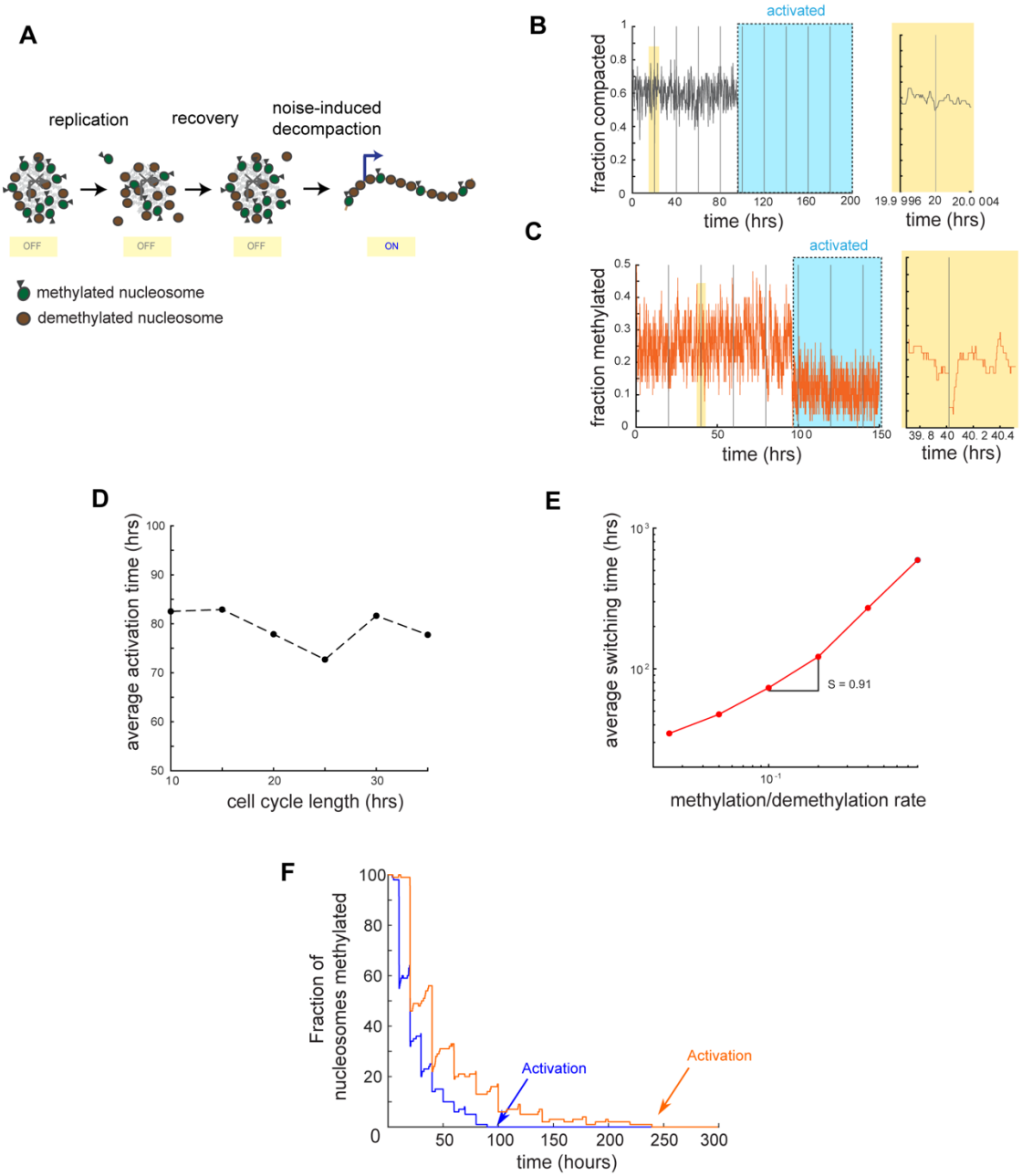


Figure 2.9 Perturbations to the compacted state by DNA replication does not affect tunability or division-independence in the methylation compaction model

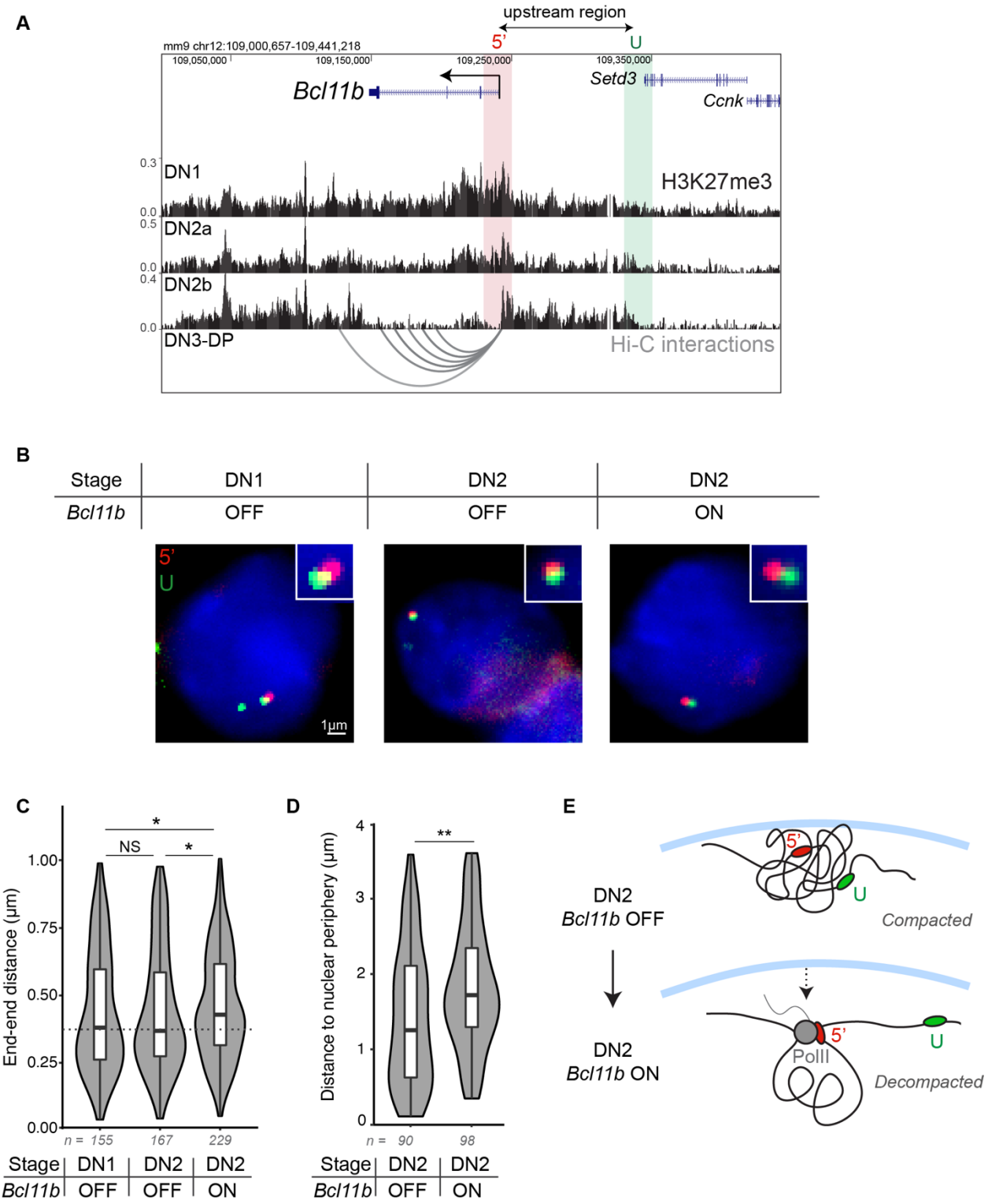
(A) Modified methylation compaction model where every cell division leads to 50% reduction in methylation state and 10% reduction in compaction state. (B-C) Compaction and methylation state as a function of time. Zoomed in first replication event. (D) Average switching time of the system as a function of cell cycle length. (E) Average switching times as a function of methylation and demethylation rate ratio. Tunability coefficient S ($\Delta \log Y / \Delta \log X$) for each plot was calculated by taking the slope of the linear fit $y = ax+b$ for the methylation model data set. (F) Fractions of methylated histones are shown for the methylation read-write model with 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 for more details).

2.2.5 *The Bcl11b locus switches to an extended conformation with activation*

The methylation-compaction mechanism above assumes that the *Bcl11b* locus is compacted prior to activation but switches to an extended, decompacted state during activation. To validate this assumption, 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). To estimate the degree of compaction independently from RNAPII-mediated DNA looping at the *Bcl11b* locus (Hu et al., 2018; Zheng et al., 2019), we designed a pair of FISH probe sets that flank a 100kb region upstream from the promoter. This upstream region begins at the *Bcl11b* promoter and resides at the edge of the putative heterochromatin compaction domain that encompasses *Bcl11b* (**Figure 2.10**A). We performed DNA-FISH in early T cell progenitors (DN1) and DN2 progenitors before or after *Bcl11b* activation (RFP^-/YFP^- vs. RFP^+/YFP^+), and measured end-to-end distances using three-dimensional imaging.

The *Bcl11b* locus showed similar end-to-end distances in RFP^-/YFP^- DN1 and DN2 progenitors; however, in RFP^+/YFP^+ DN2 progenitors, where *Bcl11b* first turns on, this end-to-end distance increased significantly, consistent with this locus maintaining a condensed state through early T cell development but switching abruptly to an extended state during *Bcl11b* activation (**Figure 2.10**B-C). Repressed heterochromatin 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). Therefore, we examined whether *Bcl11b* moves from the nuclear periphery to the interior as it decompacts and turns on, as suggested from previous studies (Isoda et al., 2017). Indeed, the distance between the *Bcl11b* promoter and the nuclear periphery was higher in *Bc111b* expressing

DN2 progenitors compared to *Bcl11b* non-expressing progenitors at the same DN2 stage, consistent with a transition from the nuclear periphery to the interior upon gene activation (**Figure 2.10**D). This finding, in conjunction with the observed increase in the end-end distance of the *Bcl11b* locus, suggests that *Bcl11b* transitions from a condensed heterochromatin-associated state to an accessible euchromatin-associated state during transcriptional activation (**Figure 2.10**E), consistent with the methylation-compaction model.



see figure legend on next page

Figure 2.10 The *Bcl11b* locus switches to an extended conformation with activation
(A) UCSC Genome Browser view of H3K27me3 ChIP-seq results in T cell progenitors (Zhang et al., 2012) and Hi-C representation maps of interactions between the *Bcl11b* TSS and other DNaseI hypersensitivity sites in Bcl11b^{+} DN3-DP T cell progenitors (Hu et al., 2018). (B) Representative images for each condition. Centroids for each pair of foci shown exist in the same *z*-plane and thus provide visual representation of the Euclidean distance. (C) Violin plots show the results from 3D Euclidean distance measurements between each probe pair (Mann Whitney U test significance: *$p < 0.05$, *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 upstream end-end probes above. (D) Violin plots show the results from 2D measurements between the *Bcl11b* 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.

2.2.6 *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 interact independently of H3K27me3. H3K9me2/3 is also associated with repressive heterochromatin domains and can serve as a binding site for HP1α, which can 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 2.11**A); therefore, we tested whether H3K9 methylation also regulates *Bcl11b* activation timing. To do so, we sorted DN2 progenitors with one active *Bcl11b* allele, re-cultured them with OP9-DL1 stromal cells in the presence of inhibitors targeting H3K9me3-modifying enzymes, and quantified the *Bcl11b* activation state after three days. We found that inhibition of Lsd1, a H3K9 demethylase, decreased the fraction of biallelic *Bcl11b* expressing cells, whereas inhibition of G9a, an H3K9 methyltransferase, increased the fraction of biallelic *Bcl11b* expressing cells (**Figure 2.11**B-C). The dose-dependent, yet moderate degree to which H3K9 methylation perturbations altered *Bcl11b* activation rate was similar to that for H3K27 methylation

perturbations as seen in **Figure 2.2**. Thus, consistent with the methylation-compaction model, H3K27me3-independent mechanisms also modulate *Bcl11b* activation timing, possibly by working together to modulate nucleosomal interactions at the gene locus.

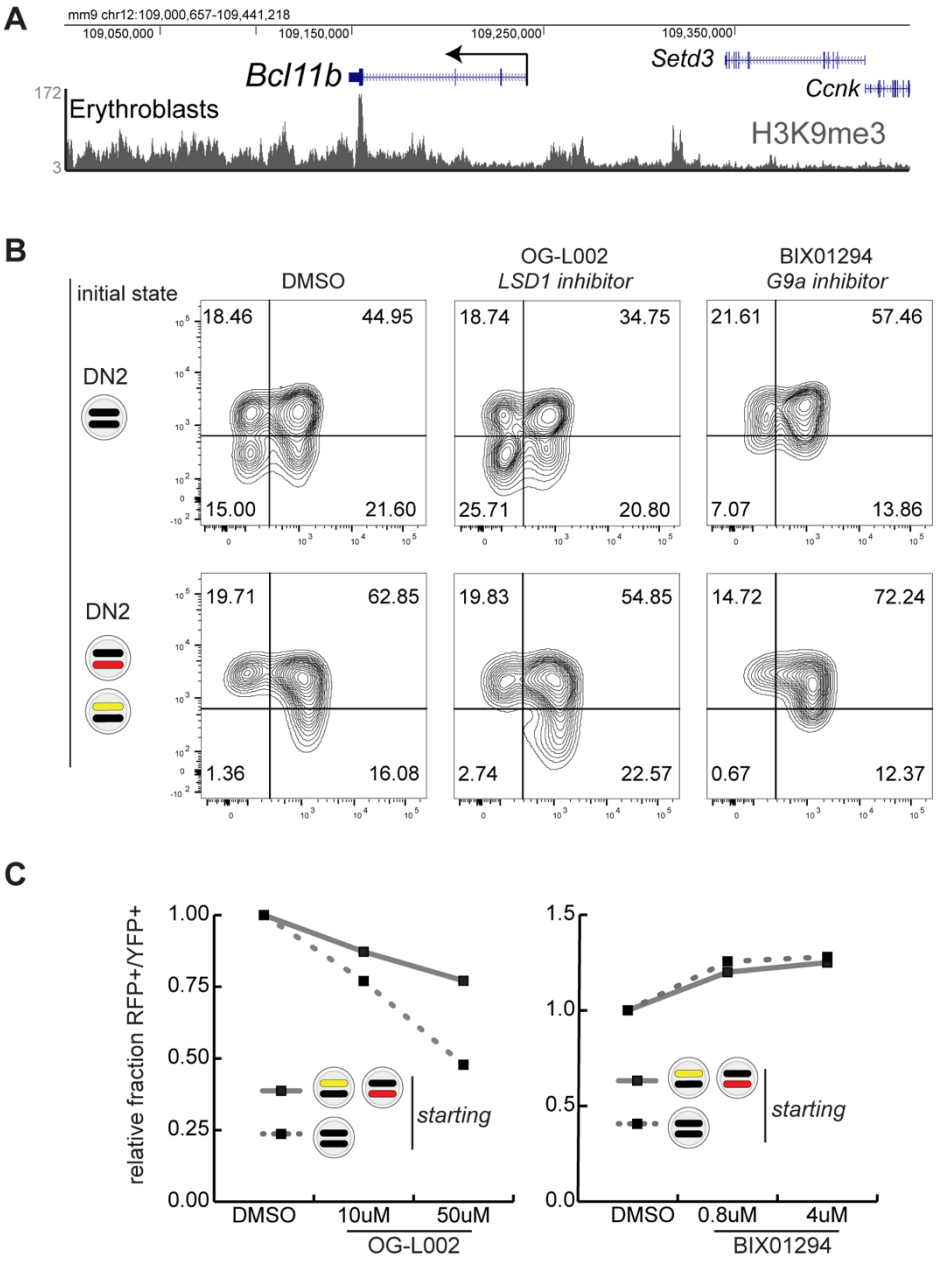


Figure 2.11 H3K9-modifying enzymes regulate Bcl11b activation probability
(A) H3K9me3 ChIP-sequencing data from mouse erythroblasts (Davis et al., 2018; ENCODE accession ENCSR000DHN) visualized in the UCSC Genome Browser. (B) DN2 progenitors, either Bcl11b$^{RFP-/YFP-}$ (top) or Bcl11b$^{RFP+/YFP-}$ and Bcl11b$^{RFP-/YFP+}$ (bottom), were purified and re-cultured on OP9-DL1 cells for 3 days in the presence of OG-L002 (LSD1 inhibitor) or BIX01294 (G9a inhibitor) before analyzing reporter expression by flow cytometry. (C) Relative fraction of progenitors RFP^{+}/YFP^{+} after 3 days normalized to the respective DMSO controls.

2.2.7 *Transcription factors can tune gene activation time delays*

Bcl11b activation timing is tunable not only by chromatin-modifying enzymes, as shown above (**Figure 2.2**), but also by two transcription factors, Gata3 and TCF-1, through a distal enhancer that physically interacts with the *Bcl11b* promoter before activation (Hu et al., 2018; Isoda et al., 2017; Kueh et al., 2016; Ng et al., 2018). These findings are consistent with broader literature showing that transcription factors and their binding sequences can modulate target gene activation probabilities (Dufourt et al., 2018; Walters et al., 1996; 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 nucleosomes and block their association with other nucleosomes in the compacted assembly (**Figure 2.12**A). Indeed, both Gata3 and TCF-1 have been identified as pioneer factors, which possess affinity to nucleosomes in addition to specific DNA sequences (Fernandez Garcia et al., 2019; Johnson et al., 2018; Meers et al., 2019; Zaret and Carroll, 2011). 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).

Our simulations revealed that the rate constant for gene activation varied with both transcription factor concentration and the number of transcription factor binding sites (**Figure 2.12**B). Increasing the number of bound nucleosomes N_B increased the activation rate in a synergistic manner, with a progressive increase in the maximal activation rate with addition of each binding site (**Figure 2.12**B, top right). Interestingly, increasing the number of binding sites also increased the transcription factor concentration at which half-maximal activation occurred

(**Figure 2.12**B, bottom right). These findings suggest that transcription factors can finely control the timing of gene activation by blocking internucleosomal interactions and chromatin compaction.

Besides blocking nucleosomal interactions, transcription factors may 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 second mechanism, where transcription factors induce the demethylation of N_R nucleosomes around their binding vicinities (**Figure 2.12**C). Our simulations revealed that transcription factors could modulate activation timing, as above, but only if they could induce demethylation of a large number of nucleosomes in their binding vicinity (**Figure 2.12**D). When a smaller number of nucleosomes were impacted, activation rates shifted only slightly when transcription factor levels increased (**Figure 2.12**D). Consistent with this idea, transcription factors bound to a single binding site can often induce chromatin changes over a significant area spanning many nucleosomes (Hass et al., 2015; Heinz et al., 2010).

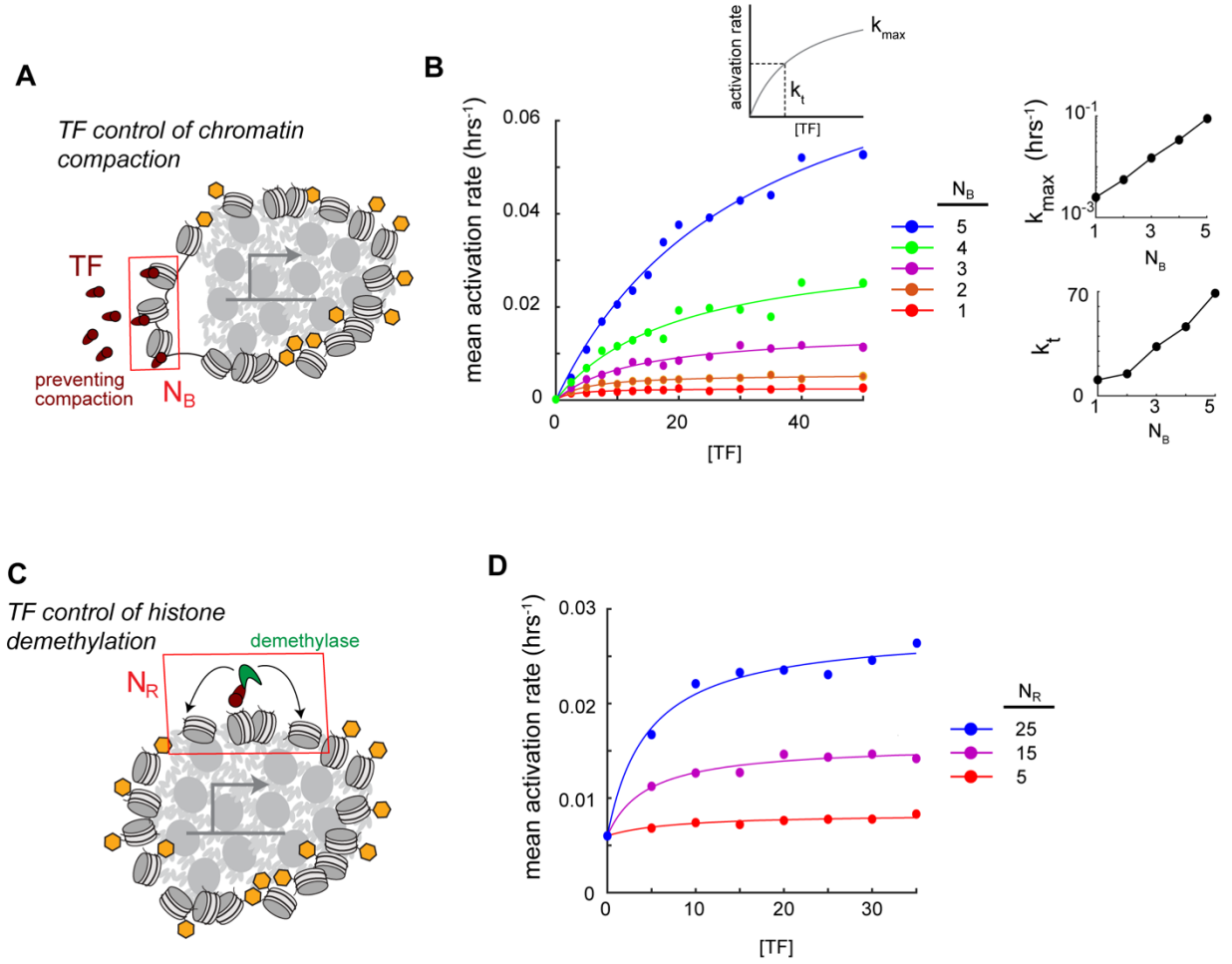


Figure 2.12 Transcription factors can control gene activation timing by modulating nucleosome methylation or compaction

(A) In the MC model, pioneer TFs can bind to nucleosomes, preventing them from entering the compacted assembly. N_B indicates the number of TF binding sites, assumed to occur on different nucleosomes. (B) Mean simulated gene activation rate against TF concentration for different values of N_B (left). Maximal activation rate (k_{max}) and half-maximal concen- tration (k_t) as a function of the number of nucleosomes bound (N_B) (right). (C) TFs can also induce demethylation of a number of nucle- osomes, N_R, within its vicinity of binding. (D) Mean gene activation rate as a function of TF concentration for different values of N_R.

2.3 DISCUSSION

During development, progenitor cells differentiate with time delays spanning multiple days and cell generations. These delays are often independent of cell division and are tunable by regulatory inputs. Here, we elucidated the mechanism of a time-delayed epigenetic switch controlling the activation of the T cell specifying gene *Bcl11b* and developed a mathematical model to explain its emergent properties. We show that H3K27me3 levels at the *Bcl11b* locus, set by opposing methyltransferase and demethylase activities, modulate the *Bcl11b* activation time delay by controlling its switch from a compacted, silent state to an extended, actively-expressing state. Activation delays generated by this methylation-compaction mechanism robustly span multiple cell generations, can be finely adjusted by both histone-modifying enzymes and transcription factors, and are set independently from cell division.

In contrast to previous epigenetic switching models, which only consider the dynamics of histone modification (Dodd et al., 2007; Zhang et al., 2014), the methylation-compaction model we propose couples histone methylation to nucleosomal interactions. Consistent with this idea, both H3K27me3 and H3K9me2/3 can promote 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., 2019; Sanulli et al., 2019; Tatavosian et al., 2018; Wang et al., 2019). Importantly, in order 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. The concept that modification states of proteins modulate their interaction affinities is well established in the study of cytoskeletal polymers (Howard, Jonathon, 2001; Mitchison, 1992), but could provide a fresh perspective on the

relationship between chromatin modifications and chromatin structure. Further testing of the methylation-compaction model will require direct interrogation of chromatin states at individual gene loci in single cells, 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; Woodworth et al., 2020; Xu et al., 2018).

The methylation-compaction switching mechanism could underlie diverse cell-autonomous timers that have been observed to work independently of cell division (Burton et al., 1999; Gao et al., 1997; Heinzel et al., 2017; Okamoto et al., 2016; Osmond, 1991; Otani et al., 2016). Measuring elapsed time independently of cell division could enable unique 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 a separate study (Nguyen et al., *in revision*). Our simulations revealed that such division-independent timing control requires active turnover of H3K27me3 and nucleosome compaction dynamics to be rapid compared to the cell cycle length (Mathematical Appendix, **Figure 2.9**F). 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 2.2**), their fast catalysis rates (Kristensen et al., 2011; Sneeringer et al., 2010), along with the observation that 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), suggest that H3K27me3 indeed turns over at a much faster timescale than that of cell division.

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. Thus, it could serve as a modular building block for gene regulatory networks that enable robust, yet adjustable control of developmental timing.

2.4 METHODS

Animal Models

C57BL/6 *Bcl11b*$^{RFP/YFP}$ mice were generated as described before (Ng et al., 2018). Briefly, *Bcl11b*$^{YFP/YFP}$ mice were generated by inserting an IRES-H2B-mCitrine-neo cassette into the 3' UTR of *Bcl11b* and *Bcl11b*$^{RFP/RFP}$ mice were generated by inserting an IRES-H2B-mCherrry-neo cassette into the same location. Dual allelic *Bcl11b*$^{RFP/YFP}$ mice with identical *Bcl11b* alleles except for fluorescent protein reporters were generated by breeding *Bcl11b*$^{RFP/RFP}$ mice to *Bcl11b*$^{YFP/YFP}$ mice. Bone marrow derived from F1 *Bcl11b*$^{RFP/YFP}$ mice at 2-4 months of age were used for all *in vitro* T cell development assays. Sex was determined not to be influential for these studies, thus male and female bone marrow were combined and analyzed together. All animals were bred and maintained at the University of Washington. All animal protocols were reviewed and approved by the Institute Animal Care and Use Committee at the University of Washington (Protocol No: 4397-01).

Cell Line Culture

Primary cells isolated from bone marrow were cultured on a OP9-DL1 monolayer stromal cells (Holmes and Zuniga-Pflucker, 2009) at 37˚C in 5% CO_2 conditions with standard culture medium [80% aMEM (Gibco), 20% Fetal Bovine Serum (Corning), Pen-Strep-Glutamine (Gibco)] supplemented with appropriate cytokines (described in Method Details). Phoenix-Eco cells were

cultured at 37˚C in 5% CO_2 with standard culture medium [90% DMEM (Gibco), 10% Fetal Bovine Serum (Corning), Pen-Strep-Glutamine (Gibco)] All cell lines were tested and found to be negative for mycoplasma contamination.

METHOD DETAILS

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 *Bcl11b*$^{RFP/YFP}$ mice. CD117 MicroBeads (Miltenyi Biotec) were used to enrich HPSCs which were frozen in 90% FBS and 10% DMSO at 10^6 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 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-enriched bone marrow progenitors were cultured on OP9-DL1 stromal cell monolayers as described before using standard culture medium [80% αMEM (Gibco), 20% Fetal Bovine Serum (Corning), Pen-Strep-Glutamine (Gibco)], grown at 37ºC 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 to 8 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 (FACS) 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 (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 pMSCV-mTagBFP2-shEed retroviral construct was generated by PCR cloning as previously described (Fellmann et al., 2013) using the mir-30-shEed PCR template sequence and pMSCV-mir-30 backbone described in the Key Resources Table.

Retroviral particles were generated using the Phoenix-Eco packaging cell line. Viral supernatants were collected at 2 and 3 days after transfection and immediately frozen at -80^{o}C. 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 32^{o}C. 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 10^6 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 described (Skene et al., 2018) with the following modifications: 1-2.5x10^5 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 4^{o}C with 5ug of anti-H3K27me3 (Active Motif) 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 $CaCl_2$ to induce MNase cleavage activity for 30 minutes at 0^{o}C. The reaction was stopped with 2xSTOP buffer (200 mM NaCl, 20 mM EDTA, 4 mM EGTA, 50 μg/mL RNase A,

50 µg/mL glycogen, and 2pg/mL of yeast spike-in DNA). 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 (Quinlan and Hall, 2010). 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 mouse *Bcl11b* promoter (see Key Resources Table for sequences). 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 derived from the *in vitro* differentiation protocol above were harvested and infected with either a MSCV empty vector or c-Myc 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 a 24-well glass bottom plate (MatTeck). To prepare the stromal-free differentiation system, which facilitates cell identification during imaging, 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 4^{o}C. This step prevents subsequent binding of retronectin to the side of the micromesh walls. Blocked micromeshes 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 (Corning), 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. 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.

Oligopaint DNA-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-CD44 APC (Clone IM7, eBioscience), and biotinylated antibodies against a panel of lineage markers (CD19, CD11b, CD11c, NK.1.1, Ter119, CD3ε, Gr-1 and B220 (BioLegend)). DN1 progenitors ($CD25^{-}/CD44^{+}/Bcl11b^{RFP-/YFP-}$), 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 Oligopaint DNA-FISH protocol (Beliveau et al., 2017). Permeabilized cells were

incubated in 0.1N HCl for 5 minutes, followed by RNaseA (100ug/ml) for 1 hour at 37ºC. Cells were then washed with 2x SSCT (2x saline sodium citrate + 0.1% Tween-20) and incubated with 2x SSCT + 50% formamide for 20 minutes at 60ºC. A hybridization mixture was prepared containing 50% formamide, 2x SSCT, 3mM 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 78ºC before incubating overnight in a humidifier chamber at 37ºC. Approximately 18 hours later, cells were washed with pre-heated 2x SSCT for 5 minutes at 60ºC. 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), catalase, 1mM methyl viologen hydrate, 1mM ascorbic acid).

FISH imaging and analysis

Cells were imaged with an inverted widefield fluorescence microscope (Leica DMi8) using a 100X oil objective, using an sCMOS camera (Photometrics Prime 95B) and a motorized stage (ASI MS-2000). Z-sections were collected at a step size of 100nm. Chromatic aberrations were corrected for using Fiji (Schindelin et al., 2012) and 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 focus 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.

QUANTIFICATION AND STATISTICAL ANALYSIS

The following statistical tests were used in this study: two-sample, one-tailed *t*-test (**Figure 2.2**E-F; **Figure 2.4**B-C, **Figure 2.3**D); and the Mann Whitney U-test (**Figure 2.10**C-D). Details for statistical tests performed are described in the indicated figure legends. All statistical tests were performed using MATLAB or R.

Modeling simulations

All models were simulated using the Gillespie algorithm provided in the Tellurium package in Python 2.7 (Choi et al., 2018). Plotting of simulation results was done in MATLAB. A detailed description of the models can be found in the mathematical appendix (see Methods S1: Mathematical Appendix, related to STAR Methods).

Image analysis of time-lapse movies

Image segmentation

Cell segmentation was performed in MATLAB (Mathworks, Natick, MA) using custom scripts described previously (Kueh et al., 2016; Ng et al., 2018). The segmentation algorithm was performed on CFP fluorescent signals as all transduced cells carried an H2B-mCerulean CFP reporter cassette. Briefly, images underwent (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

In movies of cMyc or empty vector (EV) transduced cells, live and dead cells possessed distinct morphological features as observed in the CFP fluorescence channel. Live cell nuclei had a round, smooth oval shape while dead cell nuclei tended to be more granular, with small but very bright puncta. To provide unbiased, automated recognition of live and dead cells based on these features, we applied a Laplacian mask filter to each segmented cell to delineate the 'smoothness' of its signal, then applied a threshold-cutoff to identify regions with high CFP signal. The resultant list of object features was recorded for each cell object: 1) nuclei area, 2) perimeter, 3) fluorescent intensity, 4) puncta number, 5) mean puncta area, 6) mean puncta perimeter, and 7) area for puncta above the cut-off threshold. Approximately one hundred individual cell images (10% of each data set) were then manually annotated as 'live' or 'dead'. Annotations were then linked to the above feature matrix, and a decision tree supervised machine learning algorithm was then used to generate a model based on the annotated live/dead classification and matrix features of the training images (Figure S2A-C). 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. This approach was utilized to provide an objective, automated method to distinguish between live and dead populations.

Bcl11b activation rate fitting

The following procedure was used to quantify *Bcl11b* activation rate from timelapse movies: first, the YFP and RFP signal intensity of segmented cells were calculated. Next, each cell object was classified as 'live' or 'dead', using classification prediction by trained model described in the previous section. Cells classified as 'live' were selected, and their YFP RFP fluorescence 2D histograms were then fitted to a two-component mixed 2D Gaussian model to obtain the fraction

of YFP-OFF and YFP-ON cells in the population at a given time. To calculate background-corrected fluorescent values of the Bcl11b YFP and Bcl11b RFP signals, we calculated the pixel intensity of an annulus surrounding the segmented cell and subtracted this value from the raw signal intensity in the cell interior. This approach eliminates autofluorescence from the bottom of the glass plate as well as at the edge of the PDMS microwell.

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 subsequent calculations. 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:

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

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). N_i 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 an early time point ($0 < t < 20$) to $f_1(r, y)$ to obtain the means, standard deviations, and correlation coefficients ($u_{r,1}, \sigma_{r,1}, u_{y,1}, \sigma_{y,1}, p_1$)) of the Bcl11b RFP mono-allelic population. At this early time point, cells remained inactive for the 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 $f_i^{obs}(t)$ is given by:

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

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:

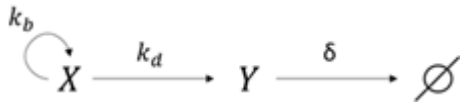
$$F_{bi}(t) = 1 - e^{-\lambda t} \quad (5)$$

Where λ is the rate for activation of the initially silent *Bcl11b YFP* allele. 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 the MATLAB fit function and a 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 3B) and fit these data to population dynamics models to obtain division and death rates:

The model includes a population of live cells (*X*) with a division rate k_b- and a death rate k_d to generate the observable dead cell population (*Y*). This population in turn has a clearance rate δ representing the process by which CFP level degrades and dead cells become undetectable.



We then obtained the division rate k_b- and the death rate k_d through a two-step curve fitting process. First the difference between these two rates $K = k_b - k_d$was obtained by fitting the number of live cells over time: :

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

where X_o is the initial number of live cells at the start of imaging. The death rate was then obtained by fitting the number of observable dead cells over time, which, given the above transition scheme, is given by the following: :

$$Y(T) = Y_o P(T) + \sum_{t=1}^{T} X(t) \cdot k_d \cdot P(T - t) \qquad (7)$$

Here $Y_o$is initial observed number of dead cells at the start of imaging and $P(\tau)$gives the probability of the dead cell remaining observable in the CFP fluorescent channel a period of time τ after its first appearance. In this model, whenever a cell starts to die, its probability of being detected decreases as per function $P(\tau)$,and the number of dead cells at a given time T is the sum of all the still-detectable dead cells generated since the start of imaging up until T. This decrease in detection probability arises because progressive dimming of CFP fluorescence, together with morphological changes following death results in the failure of the cells to be segmented.

We determined $P(\tau)$ empirically for EV and cMyc population separately by manually following 30 different dead cells and recording the time period in which it was detected and undetected until complete disappearance. Then we calculated what fraction of dead cells remained detectable after a given time had elapsed. An exponential function decay function was used to fit this 'fraction detected' curve and to estimate value for clearance rate (Figure S2B-C):

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

Here, $P(\tau)$ is the probability of a given dead cell to be detected under the CFP fluorescent channel after a period of time since its initial death. δ is the clearance rate of this process.

To fit imaging data to equation (6). We classified segmented cell objects as live or dead using a 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 a MATLAB fit function and 95% confidence interval for the fit was recorded.

To fit imaging data to equation (7), we tested a series of candidate $k_{d,i}$values; for each $k_{d,i}$, a predicted $Y_{p,i}(t)$curve was generated based on equation (7) where $T = t_1, t_2, t_3, \ldots$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 value $k_{d,i}$, whose $sse_i$is the smallest.

In order to calculate the confidence bound of the fit, we first performed a nonlinear regression by calculating the residuals of the model's predicted values $Y_p(t_i)$:

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

We then calculated to the Jacobian of the model function to estimate the covariance at each time point, 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.

A summary of results from data fitting are tabulated in **Table 2.1**.

2.5 Mathematical Appendix

2.5.1 *Introduction*

To understand the timed epigenetic switch controlling the *Bcl11b* activation, we used mathematical modeling to analyze a series of candidate biophysical mechanisms. This mathematical modeling analysis seeks to uncover 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 models. In the methylation read-write model (M), 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 methylation compaction (MC) model, individual nucleosomes are also methylated and demethylated; in addition, these nucleosomes also interact to form a compacted assembly with rates dependent on their H3K27me3 state. Unlike the methylation read-write model, 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 methylation-compaction mechanism, but not the methylation read-write mechanism, explains the emergent behaviors of the timed epigenetic switch controlling *Bcl11b* activation, and thus represents our favored model.

2.5.2 *Model I: The Methylation Read-Write 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 should also 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. The positive feedback generated by this methylation read-write mechanism provides a basis for bi-stability in this model. Here, demethylation is taken to occur at a first order rate. We assume there is no spontaneous methylation in the absence of existing methylated nucleosomes; thus, once all nucleosomes are demethylated, the system irreversibly enters an activated state.

Methylation. We explicitly model mark binding and methyltransferase activities of the PRC2 complex, as well as the methylation state of each individual nucleosome. Take the gene locus to a linear array of *N* nucleosomes. Let $i = 1..N$ denote the index for the *i*th nucleosome, and let p_i be its H3K27 methylation state. $p_i = 0$ 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 $i \in \{1,..,N\}$:

$$u' : (p_i = 0) \rightarrow (p_i = 1)$$
$$u : (p_i = 1) \rightarrow (p_i = 0)$$

With:

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

The parameter *L* can be interpreted as the 'reach' of the PRC2 complex to neighboring nucleosomes. A large value of *L* indicates a long length scale for nucleosome interactions. 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 one-half probability of inheriting a nucleosome that is methylated. Experimental evidence suggests that approximately half 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 read-write mechanism can generate a time-delayed, stochastic switches from an inactive H3K27me3-high state to an active state without H3K27me3 (**Figure 2.6**A); however, switching times are hypersensitive to mild changes to methylation and de-methylation rates (**Figure 2.6**B), and therefore inconsistent with the graded changes in switching times observed upon inhibition of PRC2 methyltransferase or Kdm6a/b demethylases (**Figure 2.2**). To understand the origins of this hypersensitivity, we re-formulate this model using a chemical kinetics framework amenable to analysis using transition state theory. To do so, we first consider 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, we can completely describe the state of the system by a single variable, 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:

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

where

$$k_1 = \beta N'(N_T - N') \quad (3)$$
$$k_2 = \alpha N' \quad (4)$$

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

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

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:

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

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

$$v(x) = k_1(x) - k_2(x) \quad (7)$$
$$D(x) = k_1(x) + k_2(x) \quad (8)$$

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*. 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:

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

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

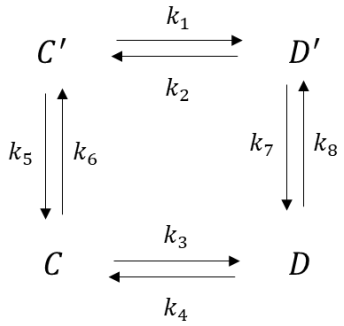
$$V(x) = \frac{\beta}{3}x^3 + \frac{\alpha - \beta N_T}{2}x^2 + W \qquad (10)$$

where W is an arbitrary number. A plot of $V(x)$ is demonstrated in **Figure 2.7**C. 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 (**Figure 2.7**C), 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.

2.5.3 *Model II: The Methylation Compaction Mechanism (MC)*

Because the methylation read-write mechanism above does not account for the tunable characteristics of the *Bcl11b* activation timing switch, we considered a second model, where histone methylation facilitates interactions between nucleosomes to enable the stable maintenance of a repressed, compacted chromatin state at the *Bcl11b* locus. There are multiple mechanisms by which H3K27me3 could facilitate interactions between nucleosomes: H3K27me3 could recruit Polycomb repressive complex 1 (PRC1), which could oligomerize through contacts on its Bmi1 or Phc subunits (Eskeland et al., 2010; Gray et al., 2016; Isono et al., 2013; Kahn et al., 2016), or undergo weak, multivalent interactions on its Cbx2 subunit that result in liquid-liquid phase separation (Howard, Jonathon, 2001; Larson et al., 2017). Alternatively, H3K27me3 could modulate affinities of weak multivalent interactions between nucleosomes (Gibson et al., 2019), and thereby modulate their ability to phase separate.

The methylation compaction model consists of two main modules: (1) a H3K27 methylation and demethylation mechanism, and (2) a dynamic chromatin decompaction mechanism linked to H3K27me3 modification state that ultimately underlies gene switching. In our description of compaction dynamics, we do not explicitly model the spatial extent of the compacted nucleosomal 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:



Here, C', C, D', and D denote the number of nucleosomes in these states, respectively. k_1 to k_8 denote the transition rates between them, which will be defined below. The gene is taken to be activated when all nucleosomes exist in a de-compacted 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:

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

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 of the PRC2 complex and Kdm6a/b demethylases, 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 read-write effect would not significantly alter the conclusion of the model. Kdm6a/b 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:

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

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 $C_T > C_N$

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

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

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

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

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 phase-separated assemblies held together by weak-multivalent, 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 read-write 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 one half probability 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 2.6**E), as experimentally observed (**Figure 2.2**). This ability to tune activation rates

by changing H3K27me3 densities distinguishes this methylation compaction model from the methylation read-write 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 give rise to 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.

To gain insights into the origins of tunability for the methylation compaction model, we adopt an approach, where we reduce this problem to using the Fokker-Planck approach, as utilized to analyze the methylation read-write mechanism (**Figure 2.7**D,E). 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 C_T, 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 that the system is at 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:

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

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

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

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

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

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

A plot of $V(x)$ is demonstrated in **Figure 2.7**D-E. 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 2.7E-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).

2.5.4 *Model II.1: The Methylation Compaction Mechanism, with Compaction Disrupted by Division (**Figure 2.9**)*

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 $S = [S_1, S_2, S_3, \dots, S_n]$ 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 $S_i(\pi_1, \pi_2, \pi_3, \dots, \pi_n)$ be the i^{th} element of 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:

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

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

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

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

From stochastic simulations (**Figure 2.9**-E), 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.

2.5.5 *Model II.2: The Methylation Compaction Mechanism with Cooperative Methylation (**Figure 2.8**)*

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:

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

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

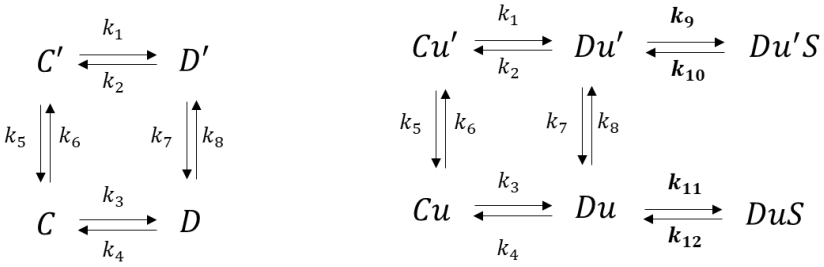
$$k_8{}^* = [\mu + \beta(C' + D')]D \qquad (41)$$

From stochastic simulations (**Figure 2.8**), 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.

2.5.6 *Model II.3: Transcription factor tuning in the methylation-compaction model (**Figure 2.12**)*

Model II.3.1. Transcription factors prevent chromatin compaction

Here, 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 associate with other nucleosomes in the compacted assembly:



The 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:

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

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

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

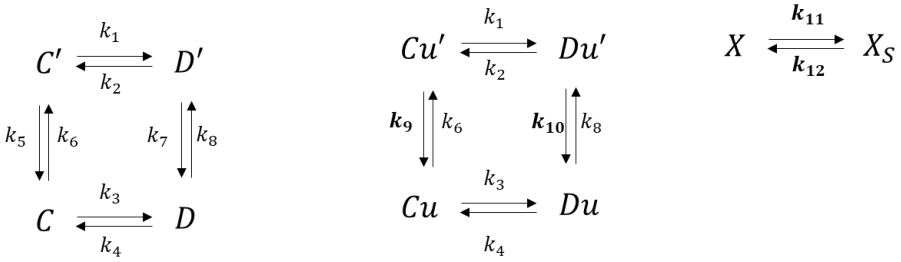
$$k_{12} = K_{OFF} \cdot DuS \qquad (45)$$

Here [*TF*] is the transcription factor concentration. Let C_T be the total number of compacted nucleosomes and let N_B be number of nucleosomes that carry transcription factor binding sites. We run our simulation with the assumption that the total number of nucleosomes is constant, and N_B can be varied.

From simulations (**Figure 2.12**), we find 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 induce histone demethylation

We next consider an alternative mechanism where a transcription factor is recruited to the nucleosome assembly at a single site. Once bound, the transcription factor acts as a demethylase to removes methyl marks on the nucleosomes in its vicinity:



Here *Cu, Cu', Du,* and *Du'* are nucleosomes that are in the recruited demethylase's reach. The unrecruited demethylase is denoted as X, and X_S is the recruited transcription factor–bound demethylase. Since the simulation involves one nucleosomal array, the number of demethylases is limited to 1, such that $X + X_S = 1$. Conversion rate probability between X and X_S 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, α_{TF} is the transcription factor recruited demethylation rate constant, and f is the fraction reduction of the rate when the nucleosome is decompacted. In our simulations, C_T is the total number of compacted nucleosomes, as before, and N_R is the maximum number of nucleosomes

that can be affected by the recruited demethylase. We run our simulation varying $N_{R,}$ while keeping the total number of nucleosomes is constant. From simulations, we find that, unlike transcription factors that impact chromatin compaction, transcription factors that recruit demethylase according to this mechanism requires a large effective range to appreciably tune switching rate (**Figure 2.12**D).

Parameter List

a. Methylation Read-Write Model (Figure 2.6A-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. Compaction Methylation Model (Figure 2.6C-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. Cell division dependence of Pure Dilution vs Compaction Methylation Models (Figure 2.6F-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. Effects of Transcription Factors on Methylation Compaction Model's Activation Timing (Figure 2.12)

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 nudeosome is in demethylated stated	0.85	
N	number of simulated nucleosomes	50	
C_N	nudeation 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 nudeosomes 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 nudeosome is in demethylated stated	0.85	
N	Total number of simulated nudeosomes	50	
C_N	nudeation 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_B	total number of nudeosomes 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. Methylation model with different cooperative reach parameter L (Figure 2.8)

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. Potential energy landscapes analysis for pure methylation model and methylation compaction model (Figure 2.7C - 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. Compaction with Cooperative Methylation Model (Figure 2.8)

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. Compaction Methylation Model with Compaction State Disruption by Cell Division (Figure 2.9A-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. Pure Dilution Model with Minimal Methylation and Demethylation Rates (Figure 2.9F)

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	

2.6 ACKNOWLEDGMENTS

We thank members of the Kueh group, Steve Henikoff, James Briscoe, Ellen Rothenberg, Michael Elowitz and Joe Levine for discussions and feedback on this work. We also thank Steve Henikoff for the generous gift of pA-MNase for CUT&RUN experiments. This work was funded by a NIH Pathway to Independence Award 5R00HL119638 (to H.Y.K.), NIH/NHLBI grants R01HL146478 (to H.Y.K), National Science Foundation URoL EF-2021552 (to H.Y.K.), and a Tietze Foundation Stem Cell Scientist Award (to H.Y.K.), University of Washington Institute for Stem Cell and Regenerative Medicine Scholarships (to N.A.P.), and NIH F31 Fellowships F31 HL151090 (to N.A.P.) and F31 HL142132 (to M.A.W.).

2.7 RESOURCE AVAILABILITY

The pMSCV-mTagBFP2-shEed plasmid is available from the Lead Contact upon request. H3K27me3 CUT&RUN data have been deposited to the NCBI Gene Expression Omnibus under accession number GSE134749. Scripts used for imaging analysis and simulations of mathematical models are available from GitHub (https://github.com/KuehLabUW/Pease_et_al.2021). All other data supporting the findings of this paper will be available from the Lead Contact.

Chapter 3. A TRIPARTITE ENHANCER CLUSTER REGULATES THE ACTIVATION, TIMING, AND MAINTENANCE OF *BCL11B* EXPRESSION

3.1 INTRODUCTION

During development, lineage-specific genes are regulated *in cis* by non-coding regulatory elements referred to as enhancers. Enhancers act as binding sites for collections of lineage-specific transcription factors (TFs) and are crucial for specifying the conditions under and the extent to which a gene is expressed during a developmental transition (Long et al., 2016). Genes are typically regulated by multiple enhancers, each of which contain multiple transcription factor binding sites (TFBSs) (Carey, 1998; Chaudhri et al., 2020; Huang et al., 2016). This layered multiplicity of control elements is likely important for proper dynamic control of gene expression during development; however, it remains unclear how multiple enhancers and multiple TFBSs within them work together to establish precise spatiotemporal expression of their target genes (Panigrahi and O'Malley, 2021).

The timing, amplitude and maintenance of gene expression need to be tightly controlled. Different *cis*-regulatory elements may be required for these different dynamic features of gene expression. However, the contribution of enhancers to these distinct aspects of gene expression during development is often uninterpretable or simply ignored. One problem is that enhancers are frequently studied by measuring the expression of their target gene in bulk at a single time point, making it impossible to distinguish between different modes of enhancer-mediated gene regulation. For example, removal of an enhancer that reduces the expression levels of its target gene in bulk is often interpreted as regulation of expression amplitude at the single cell level

(Hay et al., 2016; Shin et al., 2016). However, as discussed in *Chapter 1.2.4*, enhancers can also regulate the binary expression state of a gene and thus this observation could also be explained by a reduced fraction of cells (and/or alleles within cells) expressing the gene in an all-or-none manner (Simeonov et al., 2017; Walters et al., 1996; Weintraub, 1988). By measuring gene expression at the single cell level one can distinguish enhancers that regulate the probability of a binary switch in expression states (i.e. "timing enhancers") from enhancers that regulate the amplitude of the final active expression state (i.e. "amplitude enhancers") (Bender et al., 2012; Chu et al., 2021). Furthermore, enhancers can regulate the initiation and/or maintenance of the binary switch in expression states and thus must be interrogated before and after the onset of their target gene's expression (Zheng et al., 2010; Zou et al., 2001).

Additionally, multiple enhancers or TFBSs within them may act redundantly, additively, or synergistically to control one or more of these features at a single gene (Long et al., 2016). However, while there is evidence that multiple enhancers can function cooperatively or independently in regards to expression levels in bulk (Hay et al., 2016; Man et al., 2019; Shin et al., 2016), it remains difficult to thoroughly investigate how multiple enhancers may work together to regulate the different aspects of gene expression during development as described above. This is largely due to the limitations of the current approaches. Knockout mouse strains can be used to investigate enhancers and TFBSs in their normal developmental and genomic context but this is too laborious for scaling up to systematically interrogate both enhancers and the TFBS within them (Bender et al., 2012). Transgenic reporter assays, in which putative enhancers are linked to a gene of interest and delivered to the cell as a transgene, have traditionally been used to identify functional enhancers in a high-throughput manner (Inoue and Ahituv, 2015; Muerdter et al., 2015). However, these assays don't interrogate enhancers in their endogenous context which is likely

critical to their functionality particularly when investigating the potential interplay between spatially segregated enhancers. CRISPR-based approaches have recently been adopted to overcome these limitations by targeting putative enhancers with catalytically inactive Cas9-fusion proteins that recruit chromatin remodeling enzymes to activate or repress regions of interest (Gasperini et al., 2020). However, these approaches are difficult to interpret because the targeted enhancers frequently interact with their cognate promoters and thus the effects observed could be the result of incidentally repressing or activating the promoter directly and independently of an enhancer's normal function. In contrast, using catalytically active Cas9 in conjunction with pairs of guide RNAs (pgRNAs) can be used to introduce "programmed deletions" of enhancers which serves as a more direct and interpretable loss-of-function assay (Diao et al., 2017; Gasperini et al., 2020). However, large programmed deletions (>1kb) can be rare and frequently occur only on single alleles which makes it difficult to detect changes in gene expression activation or maintenance (Canver et al., 2014). Furthermore, the process of non-homologous end joining (NHEJ) following Cas9-induced double-stranded breaks can be unpredictable and frequently generates "unprogrammed deletions" that extend far beyond the intended target (Kosicki et al., 2018). Traditionally, multiplexed pgRNA deletion screens are readout by quantifying the distribution of pgRNA constructs across populations with different expression states to indirectly estimate the effects of the intended programmed deletions. However, the heterogeneity in unprogrammed deletions induced by individual pgRNAs makes it difficult to accurately interpret these results by indirect detection of the pgRNA constructs by themselves.

To overcome these limitations and efficiently investigate the functions of multiple enhancers and TFBSs and their potential relationships during a developmental process, we adapted the CRISPR/Cas9 pgRNA deletion approach by 1) combining it with a dual-allele reporter mouse

strain to monitor and detect single allele perturbations before and after their initial activation during development and 2) directly quantifying the programmed or unprogrammed deletions separately by qPCR and long-read sequencing. We have applied this approach to carefully dissect a 40kb enhancer cluster regulating the T cell commitment gene, *Bcl11b*. As described in more detail in *Chapter 1.3.3*, the onset and maintenance of *Bcl11b* expression is critical for specifying and maintaining T cell identity and thus must be tightly regulated during T cell development. Activation of *Bcl11b* at the DN2 progenitor stage coincides with the activation of a putative enhancer cluster located at the locus of a long non-coding RNA (lncRNA) referred to as *ThymoD* (**Figure 3.1**A). Despite being located in a 2Mb gene desert ~850kb downstream of *Bcl11b*, *ThymoD* frequently interacts with the *Bcl11b* promoter and harbors the hallmarks of an active enhancer cluster which include non-coding RNA transcription, DNase I hypersensitivity (DHS) and enrichment of H3K27ac and H3K4me1 modifications (**Figure 3.1**A). Furthermore, the orthogonal semi-conserved region in humans is translocated in over 20% of pediatric T cell leukemia patients where it drives ectopic expression of newly juxtaposed oncogenes (Nagel et al., 2007; Su et al., 2006). While there is evidence that *ThymoD* enhancer cluster is important for activation and potentially maintenance (Isoda et al., 2017; Li et al., 2013; Ng et al., 2018), it remains unclear if and how the constituent enhancers and the TFBSs within them cooperate to regulate multiple aspects of *Bcl11b* expression.

By introducing genomic deletions in developing T cell progenitors either before or after the initiation of *Bcl11b* expression, we uncovered two functionally distinct regions within the enhancer cluster, one required for *Bcl11b* activation but dispensable for maintenance, and the other required for maintenance but dispensable for activation. Within the initiation-promoting region, we identified one enhancer that was essential for *Bcl11b* activation and an additional enhancer that

was not required but simply accelerated activation timing. Finally, by pooling pgRNA constructs and performing multiplexed long-read sequencing to dissect smaller genomic regions within this "timing enhancer", we further identified multiple functional TFBS regions with quantitatively distinct contributions to *Bcl11b* activation as well as intervening regions that were functional but removed from putative TFBSs. Together, this work establishes a general method for efficiently characterizing the functional components of large non-coding regulatory regions and provides evidence for a model of gene regulation by division of labor within a single enhancer cluster.

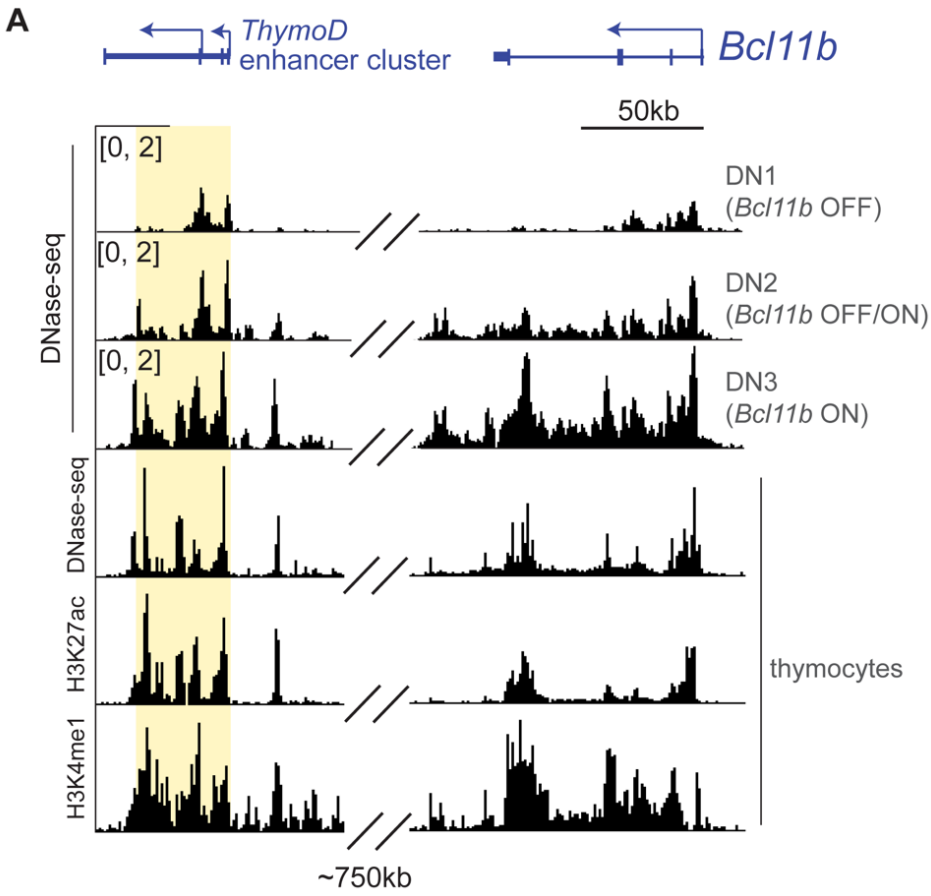


Figure 3.1 Expression of T cell commitment factor *Bcl11b* coincides with the activation of a distal enhancer cluster

(A) DNase-sequencing for early T cell progenitors (Hu et al., 2018) and whole thymus (ENCODE: ENCSR000COB). H3K27ac (ENCODE: ENCSR000CCH) and H3K4me1 (ENCODE: ENCSR000CCI) ChIP-sequencing for whole thymus visualized with the UCSC Genome Browser (Kent et al., 2013).

3.2 RESULTS

3.2.1 *ThymoD enhancer cluster possesses hierarchical constituent enhancers that regulate the onset of Bcl11b expression*

To detect rare monoallelic deletions occurring on single gene copies we generated a mouse strain, $H11^{Cas9}$; $Bcl11b^{RFP/YFP}$, which constitutively expresses Cas9 and has each endogenous copy of *Bcl11b* non-disruptively tagged with distinguishable fluorescent reporters -a mCherry red fluorescent protein (RFP) and a mCitrine yellow fluorescent protein (YFP) (Ng et al., 2018) (**Figure 3.2**A). In order to efficiently and systematically interrogate the functional constituents of the *ThymoD* enhancer cluster in the context of normal T cell development, we isolated immature bone marrow progenitors (Lin^-/Kit^+) from $H11^{Cas9}/Bcl11b^{RFP/YFP}$ mice, infected them with unique pgRNA retroviral constructs and co-cultured them on OP-DL1 stromal cells to induce T cell development. After 9-10 days of co-culture, when most of the progenitors have progressed to the Bcl11b-expressing DN2/DN3 stages, we sorted T cell progenitors with monoallelic- and biallelic-expression of *Bcl11b* into separate tubes. Then, to directly quantify the relative frequencies of each deletion, we performed quantitative real-time PCR (qPCR) with primers flanking the targeted region and calculated an activation score (A) defined as the ratio of deletion frequencies (normalized to an untargeted region of the locus) in biallelic-expressing cells to that of monoallelic-expressing cells ($A = F^{bi}/F^{mono}$). Thus, deletions with no effect on Bcl11b activation would have $A = 1$, while deletions that reduce activation would have $A < 1$.

We initially tested three pgRNA constructs to introduce deletions spanning the entire *ThymoD* enhancer cluster (Δ1-3), the distal half of the enhancer cluster (Δ1-2) and the proximal half of the enhancer cluster (Δ2-3). As a negative control, we used a pgRNA targeting a random inaccessible region of the *Bcl11b* intergenic gene desert (ΔRandom); as a positive control, we used a pgRNA

targeting the fluorescent protein reporters in the *Bcl11b* 3'UTR (ΔFP) (**Figure 3.2**B). We found that while removal of the distal half of the cluster had no effect on *Bcl11b* activation (A = 0.9), removal of either the entire enhancer cluster or the proximal half resulted in an activation score of zero, suggesting that one or more enhancers in the proximal half are required for *Bcl11b* activation. To further delineate the functional regions of the proximal, initiation-promoting region of the cluster, we tested two additional pgRNA constructs targeting either the DNase I hypersensitivity sites at the first *ThymoD* transcription start site (ΔTSS-1) or the second *ThymoD* transcription start site (ΔTSS-2). We found that while TSS-2 was strictly required for *Bcl11b* activation (A = 0.0), TSS-1 was not required but increases the probability of activation when present (A = 0.16) (**Figure 3.2**C). This corroborates our previous findings that when TSS-1 is removed, *Bcl11b* is still activated but with a 2-3 day time delay in the onset of expression compared to the wild-type allele (Ng et al., 2018). Together, these results provide evidence for a potential hierarchical structure to the initiation-promoting region of the *ThymoD* enhancer cluster in which a "timing enhancer" (TSS-1) increases the rate of *Bcl11b* activation but only in the presence of a "core enhancer" (TSS-2) which is strictly required for onset of *Bcl11b* expression.

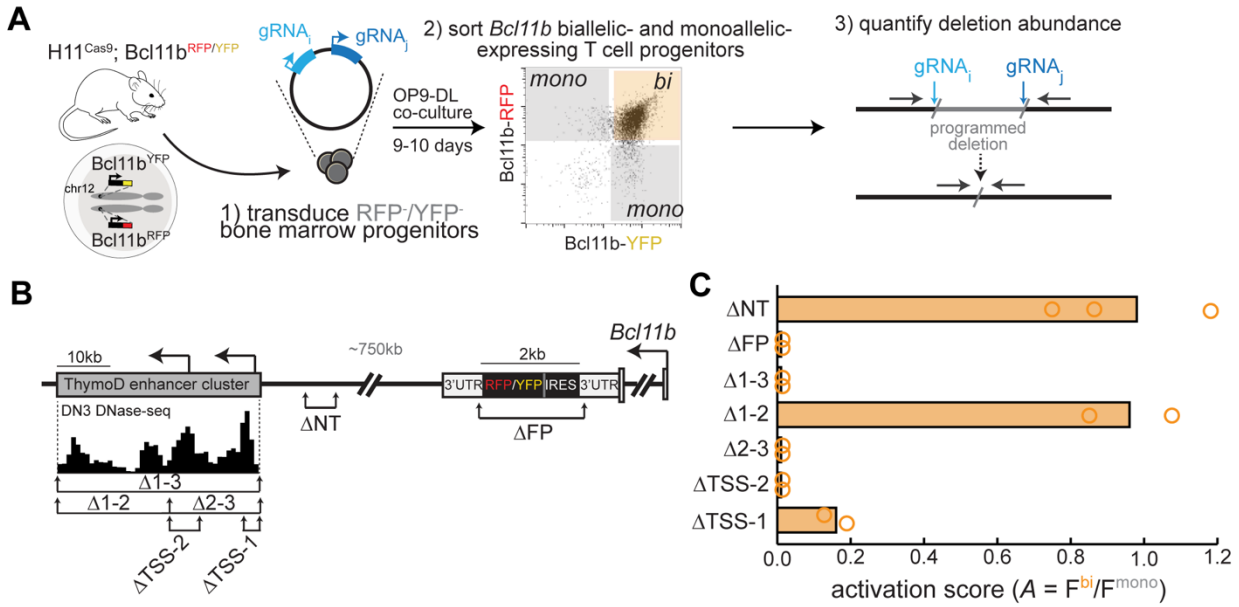


Figure 3.2 *ThymoD* enhancer cluster harbors hierarchical constituent enhancers that regulate the onset of *Bcl11b* expression
(A) Schematic for dual-allele pgRNA deletion assay. (B) Schematic showing the various paired gRNA targets used overlaid with UCSC Genome Browser (Kent et al., 2013) view of DN3 thymocyte DNase-seq results (Hu et al., 2018). (C) Bar plots displaying activation scores for each programmed deletion (n = 2-3 independent experiments).

3.2.2 *CTCF binding site is not required for core enhancer to initiate Bcl11b expression*

In addition to harboring a second transcription start site for *ThymoD*, the core enhancer also possesses a binding motif and ChIP-sequencing peak for the transcription factor CTCF which is thought to play a critical role in establishing topologically-associated domains (TADs) and DNA looping between distal regulatory sites (Isoda et al., 2017). Therefore, we investigated whether the CTCF binding site within the core enhancer was essential for its function in initiating *Bcl11b* expression or whether the putative TFBS cluster downstream in exon 3 was sufficient. To distinguish between these possibilities, we designed two additional pgRNA constructs: one to delete the accessible and transcribed region of exon 3 (Δexon3) downstream of the CTCF binding site and one to delete the CTCF binding site alone (ΔCTCF). (**Figure 3.3**A). We found that

removal of exon 3 recapitulated the deletion of the entire core enhancer (ΔTSS-2) (**Figure 3.3**B) but removal of the CTCF binding site alone (ΔCTCF) was roughly equally distributed across the Bcl11b monoallelic and biallelic expressing populations, demonstrating that the CTCF binding site is not required for the core enhancer functionality.

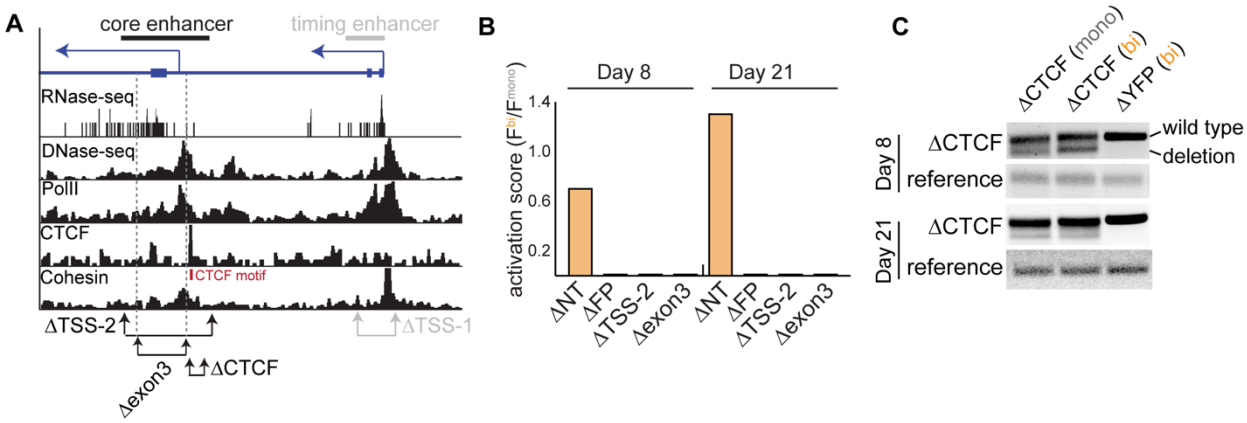


Figure 3.3 CTCF binding site is not required for the core enhancer to initiate *Bcl11b* expression
(A) Schematic showing the 5' region of *ThymoD*. RNA-sequencing, DNase-sequencing, and ChIP-sequencing results from DN2 T cell progenitors (Isoda et al., 2017; Hu et al., 2018) were visualized using the WashU EpiGenome Browser (Daofeng et al., 2019). (B) Activation scores for programmed deletions 8 or 21 days after bone marrow progenitor co-culture on OP9-DL1 cells following retroviral infection. (C) PCR analysis of ΔCTCF deletion distribution across each population.

3.2.3 *The distal region of the ThymoD enhancer cluster is required for robust maintenance of Bcl11b expression*

Unlike the proximal, initiation-promoting half of the *ThymoD* enhancer cluster, which is accessible during the DN1 progenitor stage before *Bcl11b* is expressed, the distal half of the cluster does not become fully accessible until after *Bcl11b* is activated at the DN2-DN3 stage (**Figure 3.1**A). Therefore, despite not being involved in initiating *Bcl11b* expression, the distal half of the

enhancer cluster might be involved in maintaining *Bcl11b* expression after activation which is critical for maintaining T cell identity throughout the entire lifetime of a T cell. To test this, we isolated T cell progenitors already expressing both alleles of *Bcl11b* (RFP^+/YFP^+), transduced them with unique pgRNA retroviral constructs, and co-cultured them on OP9-DL1 stromal cells for 8 days before measuring *Bcl11b* reporter expression by flow cytometry (**Figure 3.4**A). As expected, 100% of the cells transduced with the ΔRandom construct retained biallelic expression of *Bcl11b*, consistent with the irreversible nature of *Bcl11b* activation (Kueh et al. 2016). When transduced with the ΔFP construct, a small percentage of cells (9.8%) lost expression from one *Bcl11b* allele, consistent with the expected frequency of these large programmed deletions (Canver et al., 2014) (Figure 4B). Transduction with the ΔTSS-2 construct targeting the core enhancer had little to no effect on the maintenance of *Bcl11b* expression, indicating that it is not necessary for maintaining *Bcl11b* expression. In contrast, transduction with the Δ1-2 construct targeting the distal half of the enhancer cluster resulted in a substantial percentage of cells (10.5%) losing expression from one *Bcl11b* allele, suggesting a role for this region in maintaining the expression of *Bcl11b* activation. At this time, without qPCR results to quantify the deletion frequency in each sample and population, we currently lack information about the precise quantitative contribution of each region to maintaining *Bcl11b* expression. Nonetheless, these results suggest that while the core enhancer is relatively dispensable for maintenance of *Bcl11b* expression after activation, the distal, late-opening region of the enhancer cluster is required for robust maintenance of the *Bcl11b* active state.

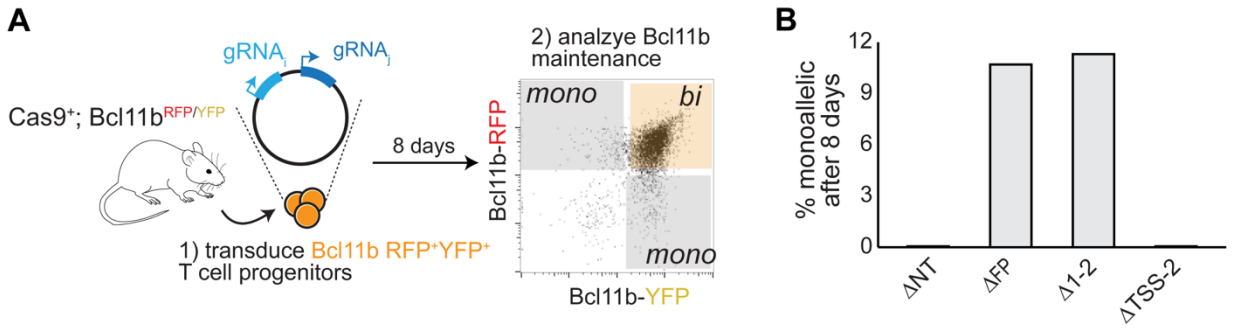


Figure 3.4 The distal half of the *ThymoD* enhancer cluster is required for robust maintenance of *Bcl11b* expression
(A) *Bcl11b* biallelic-expressing T cell progenitors were purified and separately infected with pgRNA retroviral constructs. After 8 days of co-culture with OP9-DL1 cells, infected progenitors were analyzed by flow cytometry. (B) Percentage of T cell progenitors with Bcl11b monoallelic expression 8 days after transduction.

3.2.4 *Long-read sequencing provides a direct readout of multiplexed pgRNA deletions*

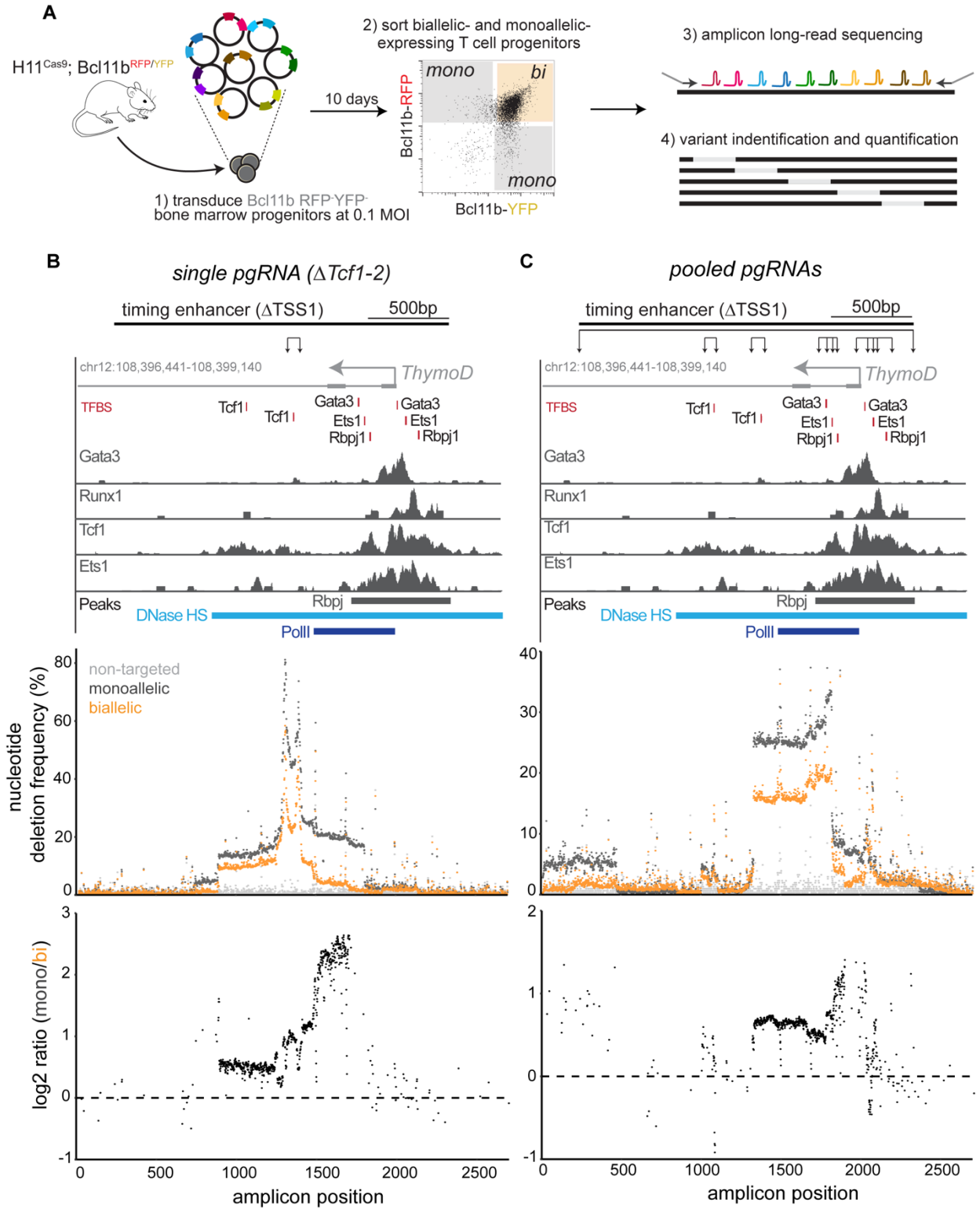
While the qPCR approach to quantifying deletions used above is useful for iteratively dissecting large non-coding regulatory regions, it is limited in its throughput and resolution. Each deletion has to be assayed with individually designed primers and deletions less than 1kb can be difficult to distinguish from the wild-type alleles by qPCR. Therefore, we sought to develop a multiplexed pgRNA approach that would enable us to screen multiple regions with finer resolution in parallel. Multiplexed Cas9-based screening approaches have generally relied on estimating the effect of a deletion indirectly by sequencing and quantifying the relative abundance of each gRNA or pgRNA construct (Diao et al., 2017). However, repair of CRISPR/Cas9-induced double stranded breaks can be unpredictable and potentially result in a heterogenous pool of variants, making it difficult to accurately interpret the actual cause of the pgRNA construct distribution (Kosicki et al., 2018). Thus, to overcome these limitations, we instead used long-read sequencing to directly detect and quantify the resulting pgRNA CRISPR/Cas9 edits. Critically, long-read sequencing enables measurement of combinatorial contributions of distal sequences that would otherwise be

impossible by next generation sequencing. As a pilot experiment, we transduced *H11*^{Cas9}*; Bcl11b*$^{RFP/YFP}$ immature bone marrow progenitors with either a single pgRNA construct targeting a single putative TFBS (Tcf1-2) or a pool of pgRNA constructs targeting 8 different putative TFBSs spread across the *Bcl11b* timing enhancer (ΔTSS-1) (**Figure 3.5**). As before, we then sorted *Bcl11b* biallelic- and monoallelic-expressing T cell progenitors after 9-10 days of co-culture with OP9-DL1 stromal cells. After extracting genomic DNA from each population, we used primers specific to the ~2.7kb timing enhancer to tag each genomic copy with unique molecular identifiers (UMI) and then PCR amplified the targeted region before performing Nanopore long-read sequencing. Sequencing reads were then processed with an Oxford Nanopore pipeline to cluster UMIs and generate high-fidelity consensus sequences which were then aligned to the amplicon sequence to identify variant alleles with deletions (**Figure 3.5**A).

As expected, samples from the single pgRNA (ΔTcf1-2) transduction showed high nucleotide deletion frequency (70-80%) at the two intended cut sites (-3bp from each PAM site) with lower nucleotide deletion frequencies in between the pair of cut sites, reflecting the reduced rates of excising the intervening sequence (**Figure 3.5**B, middle). Surprisingly, the nucleotide deletion frequency remained high (10-25%) well above the background frequency for positions in the amplicon up to several hundred nucleotides from the intended cut sites, indicating that the pgRNAs can frequently generate large deletions that extend far beyond the intended programmed deletion. Indeed, we found that of the reads with a deletion spanning at least one of the intended cut sites, 27% of them contained programmed deletions that extended >30bp from the intended excision region -highlighting the importance of directly sequencing the resulting variants (**Figure 3.6**A-B). Similar to the single pgRNA samples, the pooled pgRNA sample also showed nucleotide deletion frequencies above the background frequency at all of the intended gRNA cut sites (**Figure**

3.5C, middle); however, there was a high degree of variability in deletion frequencies among gRNA sites which likely reflects sequence specific heterogeneity gRNA binding and/or Cas9 cutting efficiencies.

Interestingly, when comparing the relative deletion frequencies between samples with monoallelic and biallelic *Bcl11b* expression, we found that while the nucleotide deletions at the Tcf1-2 motif were moderately enriched the monoallelic sample (~2 fold), deletions hundreds of nucleotides upstream of the motif were far more dramatically enriched (~5-fold) (**Figure 3.5**B, bottom). Similarly, we found that many of the nucleotide deletions most enriched in the *Bcl11b* monoallelic expressing population of the pooled pgRNA experiment were located far outside of the intended programmed deletions. Furthermore, the magnitude of nucleotide deletion enrichment near the Tcf1-2 motif was attenuated in the pooled pgRNA experiment when compared to the same nucleotide deletions in the single pgRNA experiment. This was likely because nucleotides were frequently deleted as part of a longer contiguous deletion and thus the relative enrichment of each nucleotide deletion is inherently influenced by all the other nucleotides also removed in a contiguous deletion. Together, these results suggest that multiple sub-regions contribute to the function of the *Bcl11b* timing enhancer, but the precise coordinates of these regions remain unclear.



see figure legend on next page

Figure 3.5 Long-read sequencing provides a direct readout of multiplexed pgRNA induced variants

(A) Schematic of multiplexed pgRNA CRISPR/Cas9 assay followed by long-read sequencing and variant identification and quantification. (B-C) Results from a pilot experiment targeting the *Bcl11b* timing enhancer with a single pgRNA (B) or a pool of pgRNAs (C). JASPAR TFBS motifs (Fornes et al., 2019) along with ChIP-sequencing results (provided by the Ellen Rothenberg group) for transcription factors with known roles in *Bcl11b* expression and T cell development visualized in the UCSC Genome Browser (Kent et al., 2013) (top). Nucleotide deletion frequency for each sample (middle) and the relative enrichment of each deletion in the *Bcl11b* monoallelic-expressing samples (bottom).

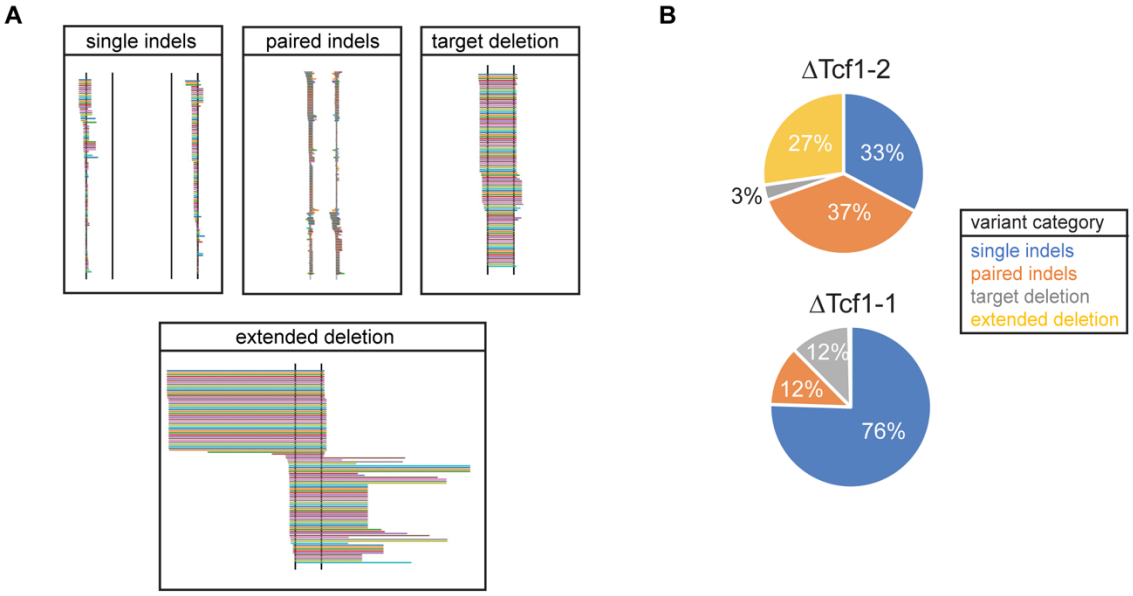


Figure 3.6 Individual pgRNAs generate heterogeneous pool of variants

(A) Representative deletions for each variant category from the single pgRNA experiment targeting the Tcf1-2 motif. Each row represents a single read and each bar marks a contiguous sequence missing from that read. Vertical lines denote gRNA cut sites (-3bp from PAM sequence for the gRNAs targeting Tcf1-2). (B) Quantification of variant heterogeneity for the Tcf1-2 motif targeted in the single pgRNA experiment and the Tcf1-1 motif targeted in the pooled pgRNAs experiment. Percentage of reads in each category among those with at least one deletion spanning at least one cut site are presented.

3.2.5 *Multiple regions of the ThymoD timing enhancer provide quantitatively distinct contributions to Bcl11b activation*

In order to take advantage of the heterogeneous variants generated by pgRNAs and more precisely define the functional regions of TFBSs within the *Bcl11b* timing enhancer, we developed a pipeline to quantify the relative enrichment of each unique contiguous deletion. For this analysis, we removed reads with multiple deletions to avoid cases of potentially confounding deletions, but we are actively developing a pipeline to interpret the combinatorial effects of non-contiguous deletions in the future. We then enumerated all contiguous deletions and binned the counts into 30 x 30 (deletion start by deletion end) nucleotide bins. Although the resulting enrichment matrices were sparse for this pilot experiment, we were still able to more precisely quantify the contribution of individual sub-elements (**Figure 3.7**A-B). In the single pgRNA experiment, we were able to confirm that intended programmed deletions of the Tcf1-2 motif were indeed enriched in the *Bcl11b* monoallelic-expressing population (**Figure 3.7**A). This result suggests that this Tcf1 motif is important to function of the timing enhancer and is consistent with previous results from a transgene reporter assay (Li et al., 2013),. Additionally, we found that deletions spanning both the Tcf1-2 motif and the first cluster of TFBSs within the first *ThymoD* intron resulted in a far greater enrichment score than deletions of the Tcf1-2 motif alone. While we would need greater coverage of the intervening sequencing to make any definitive conclusions, this suggests that TFBSs in the first intron may function additively or synergistically with the Tcf1-2 motif. As expected, the pooled pgRNA experiment provided finer quantitative information about the contributions of different sub-regions (**Figure 3.7**B). First, deletions of the Tcf1-1 were mildly enriched the *Bcl11b* monoallelic expressing samples. Second, while deletions beginning at the Tcf1-2 motif and ending in the first *ThymoD* intron were moderately enriched, deletions of the first intron alone were not enriched, suggesting that either the Tcf1-2 motif and the first intron TFBS cluster function

synergistically or that the functional region only extends into the second exon. Third, large deletions of the first exon are highly enriched to far greater magnitude than smaller deletions on either side of the exon, suggesting that the first exon region is critical for the function of the timing enhancer. While a more comprehensive variant library which includes removal of the entire timing enhancer will be required to make more precise conclusions in the future, the limited library generated in this pilot experiment was capable of refining the functional regions. Furthermore, these results demonstrate that this approach can be used to assess, in a combinatorial manner, the contribution of individual sub-elements within an enhancer region to gene activation.

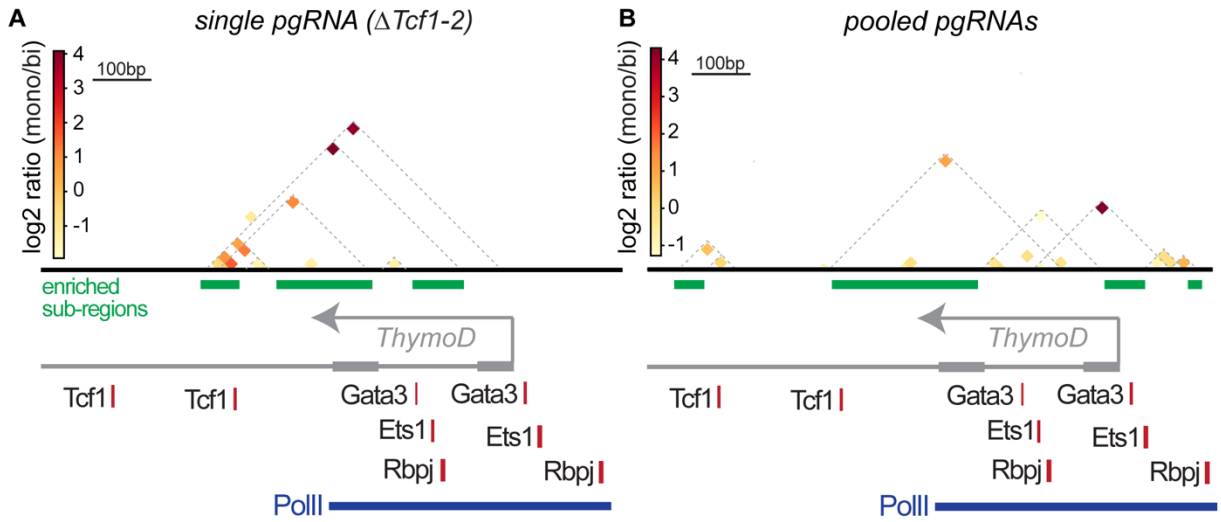


Figure 3.7 Multiple regions of the *ThymoD* timing enhancer provide quantitatively distinct contributions to *Bcl11b* activation

Heatmap representation of the relative enrichment of different variants in *Bcl11b* monoallelic-expressing progenitors from the single pgRNA experiment (A) and the pooled pgRNAs experiment (B). Deletions less than 10bp and reads within multiple deletions were removed from the analysis. Each square represents a 30bp x 30bp bin of deletion sequences that begin and end within those coordinates. Bins with less than 30 reads (single pgRNA samples) or less than 20 reads (pooled pgRNA samples) were excluded.

3.3 DISCUSSION

Most lineage-specifying genes are assumed to be regulated by multiple transcriptional enhancers which each contain multiple TFBS (Chaudhri et al., 2020; Huang et al., 2016). The contribution of each enhancer and its TFBS and the relationships among them has been difficult to study in their natural genomic and developmental context. Here, we overcome previous limitations by 1) using a dual-allele reporter mouse strain to monitor and detect rare perturbations occurring on single alleles and 2) directly quantifying variants from single pgRNAs by qPCR or variants from pools of multiplexed pgRNAs by long-read sequencing.

Gene regulation by multiple enhancers has often been studied in the context of linear clusters of putative enhancers frequently referred to as 'locus-control regions', 'super-enhancers', or 'stretch enhancers', defined based on a set of qualifying features (Li et al., 2002; Parker et al., 2013; Pott and Lieb, 2015). These classifications all imply that the enhancer clusters possess emergent properties as a result of synergy between constituent enhancers which distinguishes them from a simple collection of traditional enhancers that function independently in an additive or redundant manner (Bahr et al., 2018; Bender et al., 2012; Hay et al., 2016; Osterwalder et al., 2018). The debate about whether constituent enhancers within these clusters actually behave cooperatively or independently is often limited to evidence based on their regulation of expression levels frequently only measured in bulk well after the relevant developmental process has taken place (Choi et al., 2021; Dukler et al., 2017; Hay et al., 2016; Moorthy et al., 2017; Shin et al., 2016). Our iterative and combinatorial approach to interrogate the *ThymoD* enhancer cluster in single cells before and after the onset of *Bcl11b* activation provides a new perspective on the potential functions of enhancer clusters. Specifically, these results suggest that enhancer clusters can serve as all-in-one modular units with different constituent enhancers regulating different

aspects of gene regulation beyond expression amplitude. The hierarchical structure inherently built into this division of labor model uniquely enables gene expression timing and maintenance to be modulated independently without completely abrogating the *cis*-epigenetic switch. A similar idea has been proposed for a set of super-enhancers which contain 'hub' and 'non-hub' enhancers (Huang et al., 2016, 2018). In this model, hub enhancers are thought to mediate the most critical components of maintaining gene expression and exhibit a dominant effect over non-hub enhancers. However, functional evidence for this is limited and only involves measurements of target gene expression in bulk. Although hub-enhancers are enriched for CTCF binding sites, these sites are not always critical to the function of the hub-enhancer, which is consistent with the results of our CTCF binding site deletion within the *ThymoD* core enhancer (**Figure 3.3**C).

This study does not investigate the possibility that different elements regulate expression amplitude or maintenance at later stages of development. Different transcription factors are required to maintain expression levels at different stages of development and thus it is reasonable to hypothesize that the some enhancers bound by these transcription factors fulfill their responsibilities at these later stages (Kueh et al., 2016). *Foxp3*, for example, is regulated by at least three 'conserved non-coding DNA sequences' which initiate or maintain all-or-none expression each at different stages of regulatory T cell differentiation (Zheng et al., 2010).

Results from our multiplexed interrogation of the *ThymoD* timing enhancer at the sub-kilobase scale suggest that the first two exons are important for increasing the probability of initiating the *Bcl11b cis*-epigenetic switch. Although not addressed in this study, previous work suggests that it is unlikely that the *ThymoD* transcripts themselves play an important role in regulating *Bcl11b* expression *in trans* (Isoda et al., 2017). Thus, it's more likely that these exonic regions of *ThymoD* harbor critical TFBS or their presence is critical for the process of active

transcription and/or recruitment of splicing factors which have been proposed to play an important role in lncRNA-mediated gene regulation (Engreitz et al., 2016). Our results also suggest that deletion of the individual putative TFBSs alone have a mild effect or no effect at all on *Bcl11b* activation. This might be because the TFBSs within the timing enhancer behave cooperatively as suggested by two recent studies using single molecule footprinting to measure transcription factor binding at multiple sequences simultaneously (Rao et al 2021, Sonmezer et al 2021). Another possibility, which has been observed recently and doesn't exclude the possibility of cooperativity (Lee et al., 2021), is that non-canonical binding sites within the timing enhancer may function redundantly or non-redundantly to compensate for the loss of the canonical TFBSs we targeted. Therefore, our current approach to interrogating the timing enhancer can be improved in the future by using a more comprehensive and unbiased library of pgRNAs not limited to the predicted canonical TFBSs.

Our long-read sequencing results revealed a large degree of heterogeneity in the size and type of deletions resulting from individual pgRNAs (**Figure 3.6**). This heterogeneity also appeared to be highly variable across different pgRNAs, likely reflecting the sequence-dependent nature of NHEJ outcomes (Shen et al., 2018). This variant heterogeneity within pgRNAs and across pgRNAs raises major concerns about approaches that indirectly estimate the effects of a region by only sequencing the gRNA constructs themselves. Many of the resulting deletions we detected extended far beyond the intended programmed deletion or didn't span it at all, which makes it difficult to accurately interpret results from these experiments without directly sequencing the variants. While this heterogeneity is problematic for traditional pgRNA quantification approaches, it dramatically increases the complexity of the variant libraries assayed which can be harnessed to

derive more quantitative information with finer resolution of the targeted regions if directly sequenced by long-read sequencing.

3.4 METHODS

Animal models

C57BL/6 *Bcl11b*$^{RFP/RFP}$ *and Bcl11b*$^{YFP/YFP}$ mice were generated as described in *Chapter 2.4 Methods* and as before (Ng et al., 2018). *Bcl11b*$^{YFP/YFP}$ mice were crossed with *H11*^{Cas9} mice (The Jackson Laboratory; Stock No: 027650) to generate *H11*^{Cas9}; *Bcl11b*$^{YFP/YFP}$ which were then cross to *Bcl11b*$^{RFP/RFP}$ to generate *H11*^{Cas9}; *Bcl11b*$^{YFP/RFP}$ mice. Bone marrow derived from F1 *H11*^{Cas9}; *Bcl11b*$^{YFP/RFP}$ mice at 2-4 months of age were used for all *in vitro* T cell development assays. Sex was determined not to be influential for these studies, thus male and female bone marrow were combined and analyzed together. All animals were bred and maintained at the University of Washington. All animal protocols were reviewed and approved by the Institute Animal Care and Use Committee at the University of Washington (Protocol No: 4397-01).

Cell line culture

See Chapter 2.4 Methods

Cell purification

Bone marrow cells were harvested from femurs and tibias of 2 to 4 month-old *H11*^{Cas9}; *Bcl11b*$^{YFP/RFP}$ mice. CD117 MicroBeads (Miltenyi Biotec) and LS magnetic columns (Miltenyi Biotec) were used to enrich HPSCs using which were frozen in 90% FBS and 10% DMSO at 10^6 cells/mL. To isolate Bcl11b$^{RFP+/YFP+}$ progenitors, Lin$^-$/Kit$^-$/CD44$^-$/CD25$^+$/RFP$^+$/YFP$^+$ DN3 progenitors were purified from CD4/CD8-depleted thymoyctes. CD4/CD8-depletion was performed by staining dissociated thymocytes with biotinylated CD4 and CD8 antibodies

(Biolegend) and Streptavidin MicroBeads (Miltenyi Biotec) before performing negative depletion using LS magnetic columns (Miltenyi Biotec).

In vitro differentiation of T cell progenitors

See Chapter 2.4 Methods

Flow cytometry and cell sorting

See Chapter 2.4 Methods

Retroviral construct and transduction

Paired gRNA retroviral constructs were generated as described in (Vidigal and Ventura, 2015) with the following modifications. Custom primers with overhangs specific to the two gRNA sequences of interest and BbsI cut sites were designed to amplify the gRNA scaffold and mU6 portion of the pDonor_mU6 (a gift from A. Vidigal; Addgene plasmid #69350). PCR products from this reaction were digested with BbsI and ligated to a pBanshee-gRNA-BFP construct. The pBanshee-gRNA-BFP was generated by sub-cloning the mU6 and gRNA cassette from PX459 (a gift from F. Zhang; Addgene plasmid #62988) and the mTagBFP2 from pBAD-mTagBFP2 (a gift from V. Verkhusha, Addgene plasmid #34632).

Retroviral particles were generated as in *Chapter 2.4 Methods.* Retroviral transductions were performed as in *Chapter 2.4 Methods* with the following modification. 36-48 hours before retroviral transduction, CD117-enriched bone marrow progenitors were stimulated with 50ng/mL of Flt3-L and murine-SCF and 10ng/mL of IL-7 (Peprotech) in standard OP9 culture medium.

Quantitative real-time PCR

Genomic DNA was isolated using the Extracta DNA Prep reagents (Quantabio). PowerUp SYBR Green Master Mix (ThermoFisher Scientific) and CFX96 Real-Time PCR Detection System (Bio-

Rad) were used for quantitative PCR. The relative enrichment of each programmed deletion was quantified using the ΔΔCq method.

Library preparation for Nanopore long-read sequencing

Nanopore long-read sequencing libraries were prepared as described by the SQK-LSK109 sequencing library preparation kit and described in (Karst et al., 2021). First, UMI tagging was performed by using 2X Platinum SuperFi II Master Mix (Thermo Fisher Scientific) and custom primers targeting the flanking regions of the *ThymoD* timing enhancer with a random UMI (NNNYRNNNYRNNNYRNNN) and a synthetic priming site for the sample multiplexing primers. The UMI tagging reaction was then used for PCR amplification using standard multiplexing primers (BC07-BC011). PCR clean-up was performed with AxyPrep Magnetic Beads (Axygen) after 10 cycles and then again after another 25 cycles. End-prep and adapter ligation was then performed using NEBNext FFPE DNA Repair Mix, NEBNext Ultra II End repair/dA-tailing Module, and NEBNext Quick Ligation Module (New England Biolabs).

Long-read sequencing analysis

The pipeline-umi-amplicon pipeline (https://github.com/nanoporetech/pipeline-umi-amplicon) was used to generate consensus sequences using unique molecular identifiers (Li, 2018; Li et al., 2009; Rognes et al., 2016) with the following options: umi_errors = 3, min_reads_per_cluster = 1, max_reads_per_cluster = 60, min_overlap = 0.2, balance_strands = False, medaka_model = "r941_min_high_g360". The resulting consensus sequences were then demultiplexed using Minibar (Krehenwinkel et al., 2019) with the following settings: -p .8 -T -F. Demultiplexed consensus sequences were then aligned to the reference sequence for variant detection using CRISPResso2 (Clement et al., 2019) with no gRNA input specified. All deletions detected by

CRISPResso2 were then run through a custom pipeline to calculate the enrichment score for each variant (log2 ratio of monoallelic variant frequency to biallelic variant frequency).

3.5 ACKNOWLEDGMENTS

This work was done in collaboration with Allan Wang, Kathleen Abadie, Nick Cardozo, and Jeff Nivala. Allan performed the majority of the molecular cloning and retroviral infections. Kathleen performed the library preparation and the downstream quantitative analysis of the long-read sequencing data. Nick Cardozo performed the long-read Nanopore sequencing with guidance from Jeff Nivala. We also thank Jeff, Nick and the Kueh group for their discussions and feedback.

Chapter 4. *BCL11B* ACTIVATION TIMING REGULATES THE BALANCE OF T CELL AND INNATE LYMPHOID CELL DEVELOPMENT IN THE THYMUS

4.1 INTRODUCTION

Generating and maintaining specialized cell-types at the correct numbers and relative proportions is critical for multicellular organismal development and disease, yet we lack a complete understanding of how these cell population sizes are precisely regulated. During development, stem and progenitor cells progress through a series of lineage-restriction events prior to lineage commitment and terminal differentiation. The timing at which progenitors progress through these intermediate states in response to signals is tightly defined and is likely important for determining the number and proportion of different cell-types that are generated (Ebisuya and Briscoe, 2018). While exogenous signals are required to initiate and spatially constrain developmental processes, the speed at which they unfold is often regulated by timing mechanisms operating autonomously in single cells, as described in *Chapter 1.1*. Nonetheless, it remains unclear if and how these cell-autonomous timing mechanisms are involved in specifying cell-type numbers or whether cell population sizes are ultimately strictly confined by cell non-autonomous mechanisms.

The functional consequences of timing control on development can be studied by altering *cis*-regulatory elements that control activation kinetics of lineage-specifying genes (Chen and Goldhamer, 2004; Gérard et al., 1997). It is generally difficult to experimentally perturb the timing of a single lineage commitment step in isolation. Knocking out a lineage-specifying gene completely abrogates development and modulating exogenous signals or *trans*-acting factors generally results in systemic pleiotropic effects that are difficult to disentangle. In contrast, mutating *cis*-regulatory elements specific to a single lineage commitment gene can selectively

modulate the timing of that lineage commitment event (see *Chapter 1.2.4*) . For example, changes to *cis*-acting enhancer elements often specifically alter the all-or-none activation probability of a single gene without affecting the final expression amplitude once the gene is active. At the population level, these types of enhancers regulate the timing at which their target gene is expressed and thus can be referred to as 'timing enhancers' (**Figure 1.1**, **Figure 1.4**) (Chu et al., 2021). Timing enhancers at lineage-specifying genes have been observed across a diverse set developmental and differentiation systems ranging from skeletal and muscle development to immune cell development and differentiation (Chen and Goldhamer, 2004; Gérard et al., 1997; Mao et al., 2017; Simeonov et al., 2017; Vijayanand et al., 2012). Mutations at these timing enhancers frequently result in moderate changes in lineage-specification timing and thus could provide a unique tool to investigate the role of cell-autonomous timing mechanisms in regulating cell-type numbers.

Here, we studied the regulation of thymocyte development and cell numbers by a timing enhancer at the T cell commitment gene *Bcl11b*. Upon entering the thymus, multipotent hematopoietic progenitors referred to as CD4-/CD8- "double-negative" (DN) progenitors progress through discrete stages of development (DN1-DN4), during which they proliferate extensively such that a single thymic seeding progenitor will generate more than a million CD4+/CD8+ "double-positive" (DP) cells (**Figure 4.1**A). Only ~10% of DP cells will survive the subsequent selection process which makes the preceding expansion during the DN progenitor stages critical for generating optimal T cell numbers (Krueger et al., 2017). However, upon entering the thymus, DN thymocytes retain potential for alternative non-T cell lineages for 5-10 days before expression of the Bcl11b transcription factor is thought to silence alternative lineage genes and commit progenitors to the T cell lineage during the DN2 stage (Kueh et al., 2016). As described in more

detail in *Chapter 1.3.3*, potential for the innate lymphoid cell (ILC) lineage is the final alternative fate to be relinquished some time during the DN2 to DN3 transition (Rothenberg, 2019). Despite being ultimately expressed in both T cell and ILC lineages, the timing of *Bcl11b* activation relative to the timing of other lineage-specifying transcription factors likely determines lineage fate choice, raising the possibility that it may regulate the numbers of each cell-type generated during thymopoiesis.

To investigate this possibility, we generated a *Bcl11b* timing enhancer mutant (ΔTE) mouse in which *Bcl11b* activation was delayed by ~3 days but maintenance of *Bcl11b* expression and subsequent T cell function remained normal once *Bcl11b* was activated. Delayed T cell lineage commitment in these mice reduced thymic T cell output but enhanced the generation of thymic innate lymphoid cells (ILCs). From single cell RNA-sequencing experiments, we found that delayed *Bcl11b* activation led to the emergence of an early pro-ILC transcriptional program in multipotent progenitors which was marked by heightened expression of the pro-ILC transcription factor PLZF. Despite the delay in *Bcl11b* expression, ILC-primed progenitors were still capable of later activating *Bcl11b* to permit the generation of Bcl11b-dependent ILC2 precursors. However, we found that ILC differentiation after divergence from the T cell lineage was disrupted in progenitors that had previously experienced a delay in *Bcl11b* activation. These results highlight the importance of precise temporal regulation of cell-autonomous lineage-specific gene programs in controlling the number and proportion of differentiated cell-types.

4.2 RESULTS

4.2.1 *Deletion of timing enhancer selectively delays the onset of Bcl11b expression*

We have previously found that removal of the *Bcl11b* timing enhancer (ΔTSS-1 from **Figure 3.2**) from a single *Bcl11b* allele delays the activation of that allele but has no effect on the final

expression levels once activated *in vivo* (**Figure 1.4**B)(Ng et al., 2018). Therefore, to investigate the functional consequences of delayed *Bcl11b* activation, we generated a mouse (ΔTE) in which the Bcl11b timing enhancer was removed from each copy of *Bcl11b* which was also non-disruptively tagged with a mCitrine yellow fluorescent protein (YFP) to track *Bcl11b* expression. To examine the effects of delayed *Bcl11b* expression on the kinetics of T cell development, we purified DN1 progenitors derived from bone marrow of mice with homozygous wild-type (WT) or deleted timing enhancers (ΔTE) and co-cultured them on OP9-DL1 stromal cells to induce and monitor T cell development *in vitro*. In agreement with our previous results studying single copy mutants, ΔTE DN1 progenitors activated *Bcl11b* ~3 days later than WT DN1 progenitors but were able to maintain relatively normal *Bcl11b* expression after activation (**Figure 4.1**B-C). We found that while this delay had no effect on the DN1 to DN2 transition timing, it did cause a delay in the DN2 to DN3 transition and ultimately led to the generation of fewer fully differentiated $CD4^+/CD8^+$ double-positive (DP) cells (13% in ΔTE and 21% in WT) and increased the proportion of progenitors persisting at DN2 stage after 18 days in culture (30% in ΔTE vs. 10% in WT) (**Figure 4.1**C). Together, these results demonstrate that removal of the *Bcl11b* timing enhancer selectively delays *Bcl11b* activation and T cell lineage commitment at the DN2 stage without altering *Bcl11b* expression levels or the downstream stages T cell development.

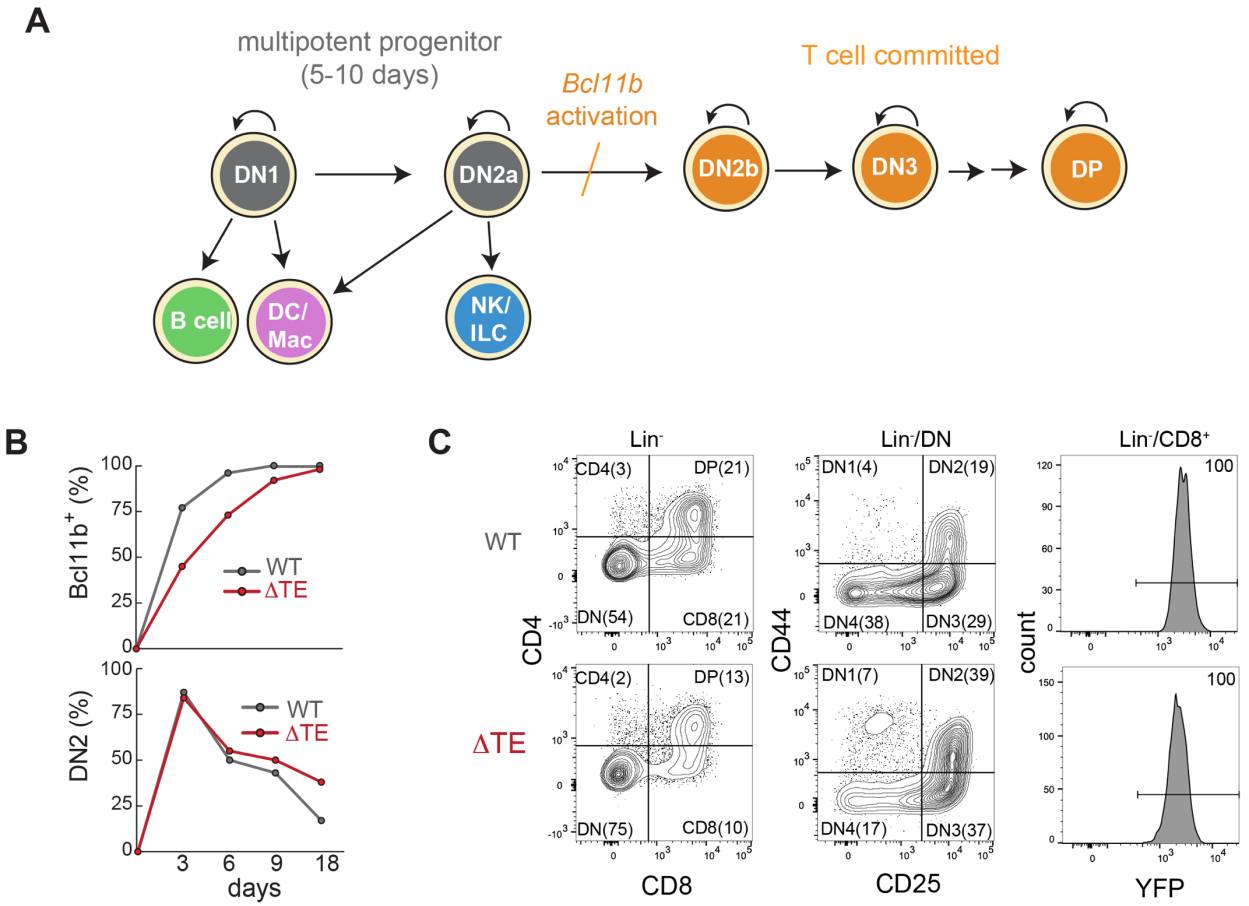


Figure 4.1 Deletion of timing enhancer selectively delays *Bcl11b* expression and T cell commitment
(A) Schematic of early T cell development (B) Bcl11b^{-}/CD44^{+}/CD25^{-} DN1 progenitors from *Bcl11b*$^{WT/WT}$ or *Bcl11b*$^{\Delta TE/\Delta TE}$ were purified and co-cultured on OP9-DL1 cells before analyzing by flow cytometry. (C) Flow cytometry scatterplots showing T cell developmental markers after 18 days of co-culture.

4.2.2 *Delayed Bcl11b activation delays T cell development and decreases T cell output in vivo*

To investigate the importance of tightly regulating lineage commitment timing during development, we compared T cell output and peripheral T cell numbers in WT and ΔTE mice by profiling T cells in their thymi and spleens. We found that delayed *Bcl11b* activation in the ΔTE mice significantly increased the percentage of pre-committed DN2a thymocytes (~1.5-fold) yet

did not disrupt T cell development and maturation after lineage commitment (**Figure 4.2**A-B), consistent with a selective delay in T cell lineage commitment at the DN2 stage upon TE deletion. This moderate delay in T cell lineage commitment resulted in significant decreases in the total number of thymocytes (20%) and T cells in the spleen (27%) (**Figure 4.2**C-D). However, the number of B cells in the spleen remained constant, suggesting that these changes in T cell numbers are not due to systemic changes in immune cell development in ΔTE mice. These results demonstrate that the timing of T cell lineage commitment, regulated by *Bcl11b* activation, plays an important role in regulating T cell output from the thymus.

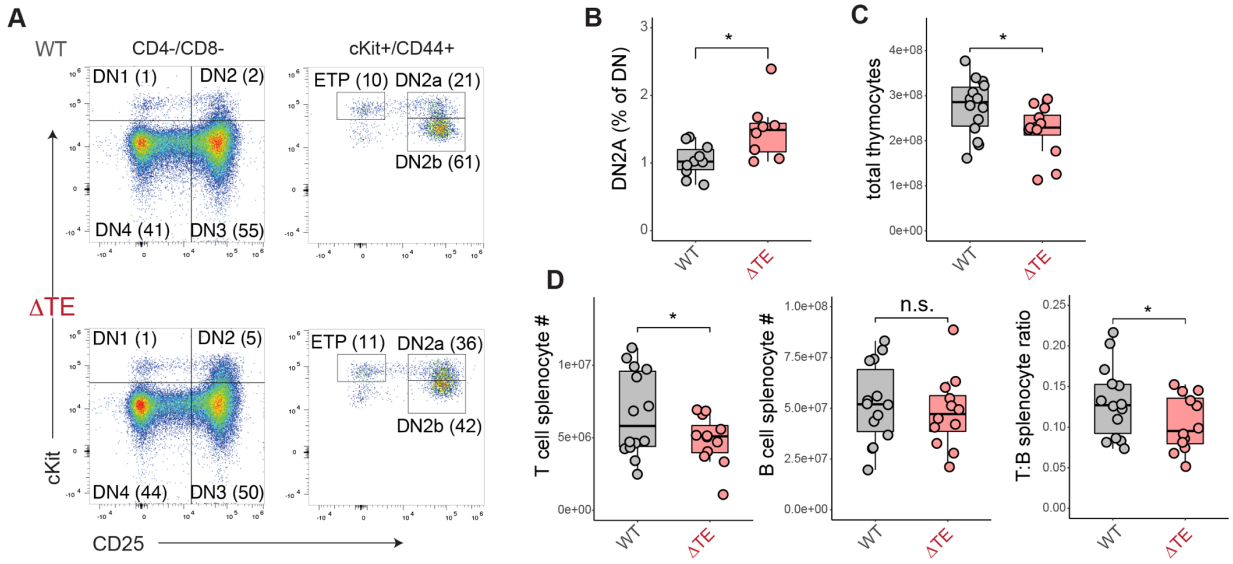


Figure 4.2 Delayed *Bcl11b* activation delays T cell development and decreases T cell output *in vivo*

(A) Representative flow cytometry scatterplots showing DN thymocytes negative for mature lineage markers (CD4/CD8/CD19/CD11b/CD11c/NK1.1/Ter119/Gr-1). (B-D) Boxplots of cell numbers in WT and ΔTE mice (n = 10-11, two-sample t-test, one-sided).

4.2.3 *Delayed Bcl11b activation increases divergence towards pro-ILC transcriptional program*

Activation of *Bcl11b* is generally thought to be one of final steps in relinquishing multipotency during T cell development due to its well characterized role as transcription factor that silences stem cell and non-T cell lineage specifying genes (Hosokawa et al., 2018). Therefore, we hypothesized that delayed *Bcl11b* activation and the consequential extension of multipotency at the DN2 stage increases progenitor cell divergence towards alternative non-T cell lineages. To test this hypothesis and simultaneously examine all possible alternative fate outcomes, we used fluorescence-activated cell sorting (FACS) to purify bone marrow derived DN1, DN2 Bcl11b^{-}, and DN2 Bcl11b^{+} progenitors from WT and ΔTE mice and re-cultured them on OP9-DL1 stromal cells for four days before harvesting them for single-cell RNA-sequencing (scRNA-seq). If our hypothesis were true, we would expect the ΔTE DN2 Bcl11b^{-} progenitors to generate more non-T cell lineages but the ΔTE DN1 and DN2 Bcl11b^{+} progenitors to progress normally to the DN2 and DN3 stages, respectively. By projecting the single-cell gene expression profiles onto a two-dimensional space using the UMAP algorithm (McInnes et al., 2018) and then performing clustering, we identified six clusters (**Figure 4.3**A-B) corresponding to myeloid, DN1, DN2a, DN2a/DN2b-hybrid, DN3, and innate lymphoid cell (ILC) progenitors. As predicted, DN1 progenitors from WT and ΔTE mice generated cells in each cluster with roughly equivalent frequencies, reflecting an early multipotent state which would not yet be perturbed by a forthcoming delay in *Bcl11b* activation. In contrast, ΔTE derived DN2 Bcl11b^{-} progenitors showed a substantial increase in frequency of the ILC-like cluster but not the myeloid-like cluster. This suggests that delayed *Bcl11b* activation at the DN2 stage specifically increases the potential for an ILC transcriptional program to emerge and induce divergence away from the T cell lineage. Surprisingly, we found that both WT and ΔTE derived DN2 Bcl11b^{+} progenitors

generated ILC-like cells even more frequently than DN2 Bcl11b^{-} progenitors. Once again, ΔTE derived DN2a Bcl11b^{+} progenitors generated over twice as many ILC-like progenitors compared to WT derived DN2a Bcl11b^{+} progenitors. Together, these results suggest that DN2 progenitors retain ILC developmental potential and that delayed *Bcl11b* activation and T cell lineage commitment promotes divergence towards a pro-ILC transcriptional program. Strikingly, ILC potential appears to be retained in DN2 progenitors that have initiated *Bcl11b* expression, indicating that a branch point for T and ILC lineages may occur downstream of *Bcl11b* expression.

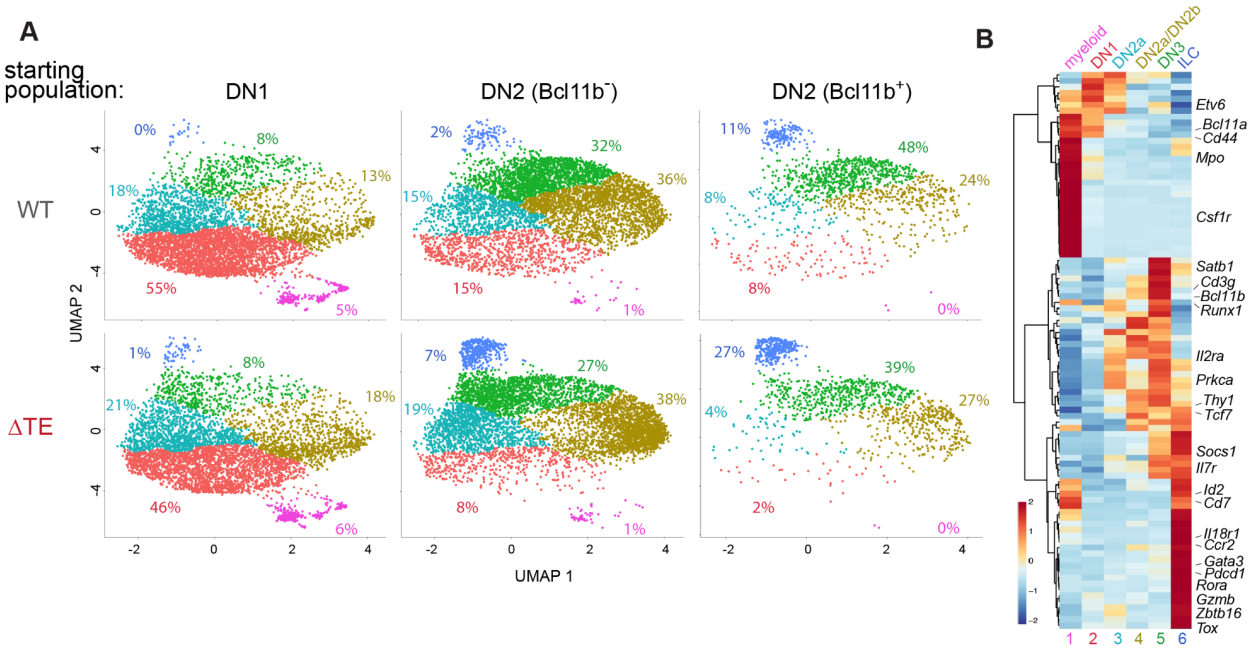
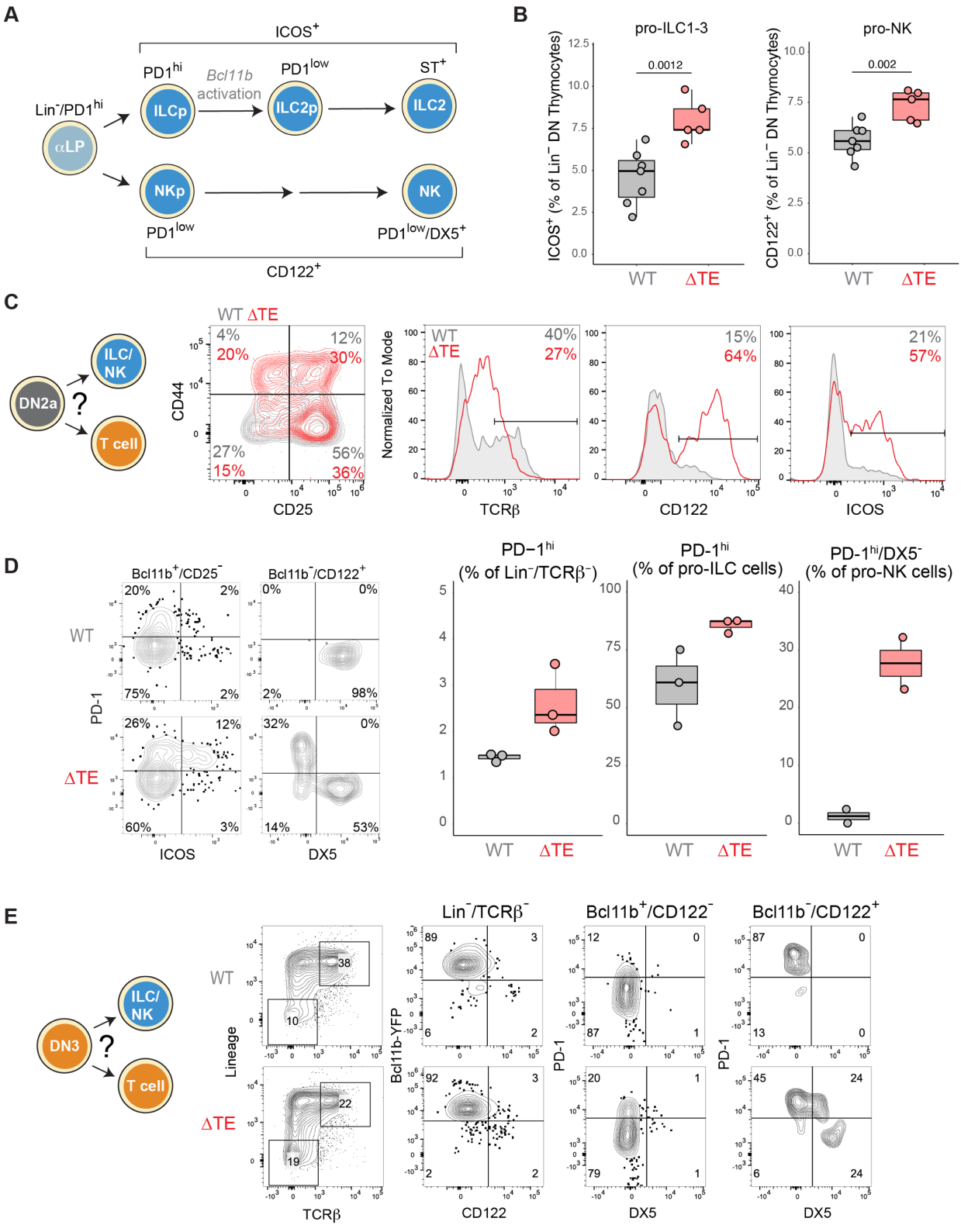


Figure 4.3 Delayed *Bcl11b* activation increases divergence towards pro-ILC transcriptional program

Bone marrow progenitors from WT or ΔTE littermates were co-cultured on OP9-DL1 stromal cells with 10ng/mL of IL-7 and Flt3-L for 7 days to initiate T cell development. After 7 days of co-culture, DN1 (CD44^{+}/CD25^{-}/Bcl11b^{-}), DN2 (CD44^{+}/CD25^{+}/Bcl11b^{-}), and DN2 (CD44^{+}/CD25^{+}/Bcl11b^{+}) progenitors were purified and co-cultured on OP9-DL1 stromal cells for 4 days before harvesting for sci-RNA-seq. (A) UMAP plots with cells color coded by cluster identification. (B) Differential gene expression analysis of immune-related genes across identified clusters.

4.2.4 *Delayed T cell lineage commitment primes DN2/DN3 thymocytes for divergence to ILC lineages but disrupts downstream ILC maturation*

Delayed *Bcl11b* activation may also increase DN2 progenitor divergence towards ILC lineages in the thymus and thus contribute to the diminished T cell output observed in ΔTE mice (**Figure 4.4**B). To investigate this possibility, we measured the expression of ICOS and CD122, which are cell surface receptors that broadly mark cells committed to the ILC1-3 and NK cell lineages, respectively (Huang et al., 2017; Kernfeld et al., 2018; Yu et al., 2016) (**Figure 4.4**A). We found that the relative proportions of $Lin^-/ICOS^+$ pro-ILC1-3 cells and $Lin^-/CD122^+$ pro-NK cells were both significantly increased (~30%) in ΔTE mice, suggesting that delayed *Bcl11b* activation indeed increases the ratio of ILC to T cell progenitor development *in vivo* (**Figure 4.4**B). Nonetheless, ILCs predominately develop in the bone marrow and migrate to peripheral tissues, thus it was possible that the increased pro-ILC numbers were due to pleiotropic effects not related to delayed T cell lineage commitment in the thymus. To directly assay the lineage potential of DN2 thymocytes and determine whether delayed *Bcl11b* activation poises these progenitors for divergence to the ILC lineage, we compared the lineage fate potential of WT and ΔTE derived DN2a ($Bcl11b^-$) thymocytes by sorting and co-culturing them with OP9-DL1 stromal cells in the presence of both pro-ILC and pro-T cell cytokines. After ten days of co-culture, ΔTE derived DN2a progenitors generated fewer $TCR\beta^{hi}$ cells but more $CD44^+/CD25^-$ DN1-like progenitors (20% vs 4%) within the $TCRB^-$ population, indicative of an increase in non-T cell lineage divergence with delayed *Bcl11b* activation (**Figure 4.4**C). Furthermore, in agreement with our *in vivo* data, ΔTE derived DN2a progenitors developed elevated expression of the pro-ILC1-3 and pro-NK cell markers ICOS and CD122, suggesting *Bcl11b*-delayed DN2a thymocytes were indeed more poised for ILC development.



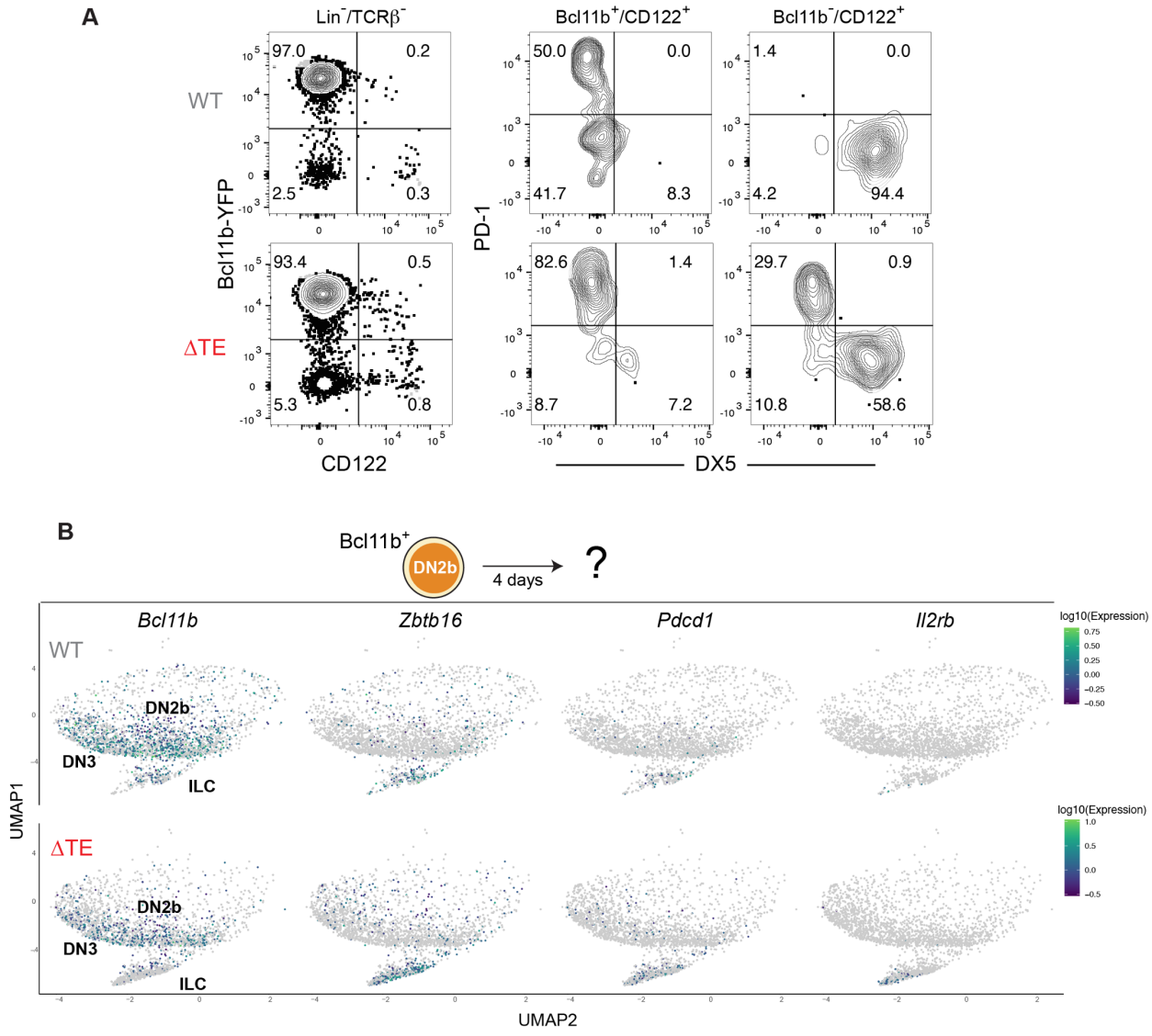
see figure legend on next page

Figure 4.4 Delayed T cell lineage commitment primes DN2/DN3 thymocytes for divergence to ILC lineages but disrupts downstream ILC maturation
(A) Schematic depicting expression of putative ILC lineage markers (Huang et al., 2017). (B) Boxplots of pro-ILC1-3 and pro-NK cell analysis within the DN thymocytes negative for lineage markers (CD4/CD8/CD19/TCRγδ/Ter119/Gr-1). (*n* = 6-7 mice, two-sample t-test, one-tailed). (C) Primary DN2a thymocytes ($Kit^+/CD44^+/CD25^+/Bcl11b^-/CD122^-/ICOS^-/DX5^-$) were FACS purified and re-cultured on OP9-DL1 stromal cells with 5ng/mL Flt3-L, IL-7, IL-33, IL-15, and IL-2 and analyzed 10 days later by flow cytometry. (D) Flow cytometry scatter plots and accompanying boxplots quantifying PD-1^{hi} populations (*n* = 2-3 littermates) (E) Primary DN3 thymocytes ($Kit^-/CD44^-/CD25^+/Bcl11b^+/CD122^-/ICOS^-/DX5^-$) were FACS purified and re-cultured on OP-DL1 stromal cells with 1ng/mL Flt3-L, IL-7, IL-15, and IL-2 and analyzed 6 days later by flow cytometry. Lineages markers include CD4, CD8, CD19, TCRγδ, Ter119, and Gr-1.

The results above provide strong evidence that delayed T cell lineage commitment increases pro-ILC lineage potential in DN2a thymocytes which possibly increases ILC generation at the expense of T cell generation. ILC2 and NK cells have been found to be the dominant ILC subtypes in neonates (Ferreira et al., 2021; Jones et al., 2018); thus, we measured the expression of ST2 and DX5 which are cell surface receptors that mark fully mature ILC2 and NK cells, respectively. Surprisingly, delayed *Bcl11b* activation in ΔTE mice had relatively no effect on the percentage of ST2 or DX5-positive progenitors in the DN thymocyte compartment (data not shown). Therefore, to determine if delayed *Bcl11b* activation affects ILC maturation after commitment, we measured PD-1 expression, which is highly expressed in ILC-restricted progenitors but subsequently downregulated as these progenitors become terminally differentiated in both ILC1-3 and NK lineages (**Figure 4.4**A). As expected, ΔTE thymi had an increased percentage of PD-1^{hi} progenitors, consistent with increased divergence into an ILC-restricted precursor state when T cell commitment is delayed (**Figure 4.4**D). We found that ΔTE thymocytes were capable of expressing ICOS and Bcl11b, which restricts ILC precursors (ILCp) to the ILC2 lineage; however, a much higher percentage of these ILCp cells retained high PD-1 expression in ΔTE thymi, suggesting that the progression to the ILC2p stage was impaired by a previous delay

in *Bcl11b* activation. Additionally, within the NK thymocyte lineage, we found that ΔTE thymocytes were capable of generating fully mature $DX5^+$ NK cells; however, 32% of the $Bcl11b^-$ /$CD122^+$ NK precursors (NKp), which are normally exclusively $PD\text{-}1^-/DX5^+$, were $PD\text{-}1^{hi}/DX5^-$ in ΔTE mice, suggesting that NK cell maturation was also impaired in progenitors with delayed Bcl11b activation. While it was less surprising that ILC2 differentiation, which depends on Bcl11b expression, might be perturbed when *Bcl11b* activation is delayed, it was more surprising that this delay also perturbed NK cell differentiation which is thought to occur independently of Bcl11b expression. Nonetheless, we noticed that a small fraction of $Bcl11b^+$ cells expressed low levels of CD122 which increases as *Bcl11b* expression decreases (**Figure 4.5**A). In agreement with these results, our scRNA-seq experiment revealed that $Bcl11b^+$ DN2 progenitors can generate NKp-like cells that express *Pdcd1* (PD-1) and/or *I2rb* (CD122) (**Figure 4.5**B). Therefore, we hypothesized that thymic NK cells are at least partially derived from precursors that transiently express *Bcl11b* and that delays in this transient expression perturb downstream NK cell maturation. To test this hypothesis, we purified $CD122^-/DX5^-/CD44^-/CD25^+/Bcl11b^+$ DN3 progenitors from WT and ΔTE mice and co-cultured them with OP9-DL1 cells in the presence of both pro-ILC and pro-T cell cytokines. After 6 days of co-culture, the ΔTE derived progenitors generated a smaller proportion of DP T cells ($Lin^+/TCRB^+$) but a larger portion of $PD\text{-}1^{hi}$ cells, suggesting that despite specifically expressing markers of T cell commitment, DN3 thymocytes that previously experienced a delay in *Bcl11b* activation potentially have impaired T cell maturation and/or are more poised for divergence to non-T cell lineages (**Figure 4.4**E). Furthermore, a small fraction of both WT and ΔTE derived DN3 thymocytes became $Bcl11b^-/CD122^+$, suggesting that *Bcl11b* can indeed be transiently expressed prior to CD122 expression. Unlike $Bcl11b^+/CD122^-$ cells which are predominantly $PD\text{-}1^{low}/DX5^-$, $Bcl11b^-/CD122^+$ cells are predominantly $PD\text{-}1^{hi}$ and/or $DX5^{hi}$ in the

case of progenitors derived from ΔTE thymocytes, suggesting that Bcl11b$^+$ DN3 progenitors possess a latent potential for NK cell divergence which may be enhanced by delayed *Bcl11b* activation. Together, these results demonstrate that the timing of *Bcl11b* activation is not only critical for regulating thymic ILC and T cell lineage bifurcation but also important for downstream differentiation of these lineages after Bcl11b is expressed, suggesting that the timing of lineage-specifying gene activation relative to other gene activities is tightly regulated to enable proper development.



see figure legend on next page

Figure 4.5 Rare Bcl11b$^+$ progenitors appear to silence *Bcl11b* while up-regulating NK-lineage markers
(A) Representative flow cytometry scatterplots from two littermates showing the existence of Bcl11b+/CD122+ progenitors (from data used in **Figure 4.4**D). Lineages markers include CD4, CD8, CD19, TCRγδ , Ter119, and Gr-1. (B) UMAP plots displaying normalized gene expression levels for individual cells in low dimensional space. Bone marrow derived DN2 (CD44$^+$/CD25$^+$) Bcl11b$^+$ progenitors were re-cultured on OP9-DL1 cells for 4 days before harvesting for sci-RNA-seq (from data presented **Figure 4.3**).

4.2.5 *PLZF expression is elevated in DN2 progenitors with delayed Bcl11b activation*

To better understand the mechanisms involved in promoting pro-ILC divergence when *Bcl11b* activation is delayed, we used our sc-RNA-seq data to identify genes differentially expressed between WT and ΔTE progenitors within the DN2a and DN2a/b hybrid clusters where *Bcl11b* expression is normally initiated. As expected, pro-T cell genes (*Bcl11b*, *Cd3g*, and *Runx1*) were among the genes most significantly downregulated in ΔTE progenitors, confirming that T cell commitment was indeed delayed in this experiment (**Figure 4.6Error! Reference source not found.**A). In contrast, several key pro-ILC transcription factor genes (*Gata3*, *Runx3*, and *Zbtb16*) were among those most significantly upregulated in ΔTE progenitors, suggesting that delayed *Bcl11b* activation enhances transcriptional priming towards ILC lineages during the DN2 stage. Furthermore, several pro-hematopoietic stem cell maintenance genes (*Mpo*, *Fli1*, and *Meis1*) were also among the genes most significantly downregulated in ΔTE progenitors, raising the possibility that the elevated expression of the pro-ILC transcription factors is also prematurely silencing stem cell legacy genes. *Zbtb16*, which encodes the pro-ILC transcription factor PLZF, is directly bound and repressed by Bcl11b during T cell commitment (Hosokawa et al., 2018, 2020), making it a likely candidate responsible for promoting ILC lineage divergence when Bcl11b expression is delayed (see also **Figure 1.2** and **Figure 1.3**). In agreement with those reports, we found that *Bcl11b* and *Zbtb16* expression were anti-correlated in low-dimensional space (**Figure 4.5**B) with

ΔTE progenitors showing an increased frequency of expression across the *Bcl11b* negative portions of the DN2-associated clusters as well elevated expression at the branching point between DN3 and ILC-associated clusters (**Figure 4.6**B).

To investigate whether PLZF may be promoting divergence toward ILC lineages *in vivo*, we measured its expression levels in thymi of WT and ΔTE littermates. In agreement with our *in vitro* scRNA-seq results, PLZF was transiently upregulated at the DN2a stage before being downregulated as Bcl11b expression increased at the DN3 stage (**Figure 4.6**C). Furthermore, as identified by the *in vitro* scRNA-seq experiments, this transient expression of PLZF was elevated (~50%) at the DN2a stage and persisted longer through the DN2b stage in ΔTE mice, resulting in the emergence of a small population of $PLZF^{hi}$/$Bcl11b^{-}$ DN3 progenitors which are not observed in WT mice. Together, these results suggest that when *Bcl11b* activation is delayed, the increased divergence towards the ILC lineage is likely driven by the up-regulation of *Zbtb16* which establishes a pro-ILC transcriptional program before the onset *Bcl11b* expression.

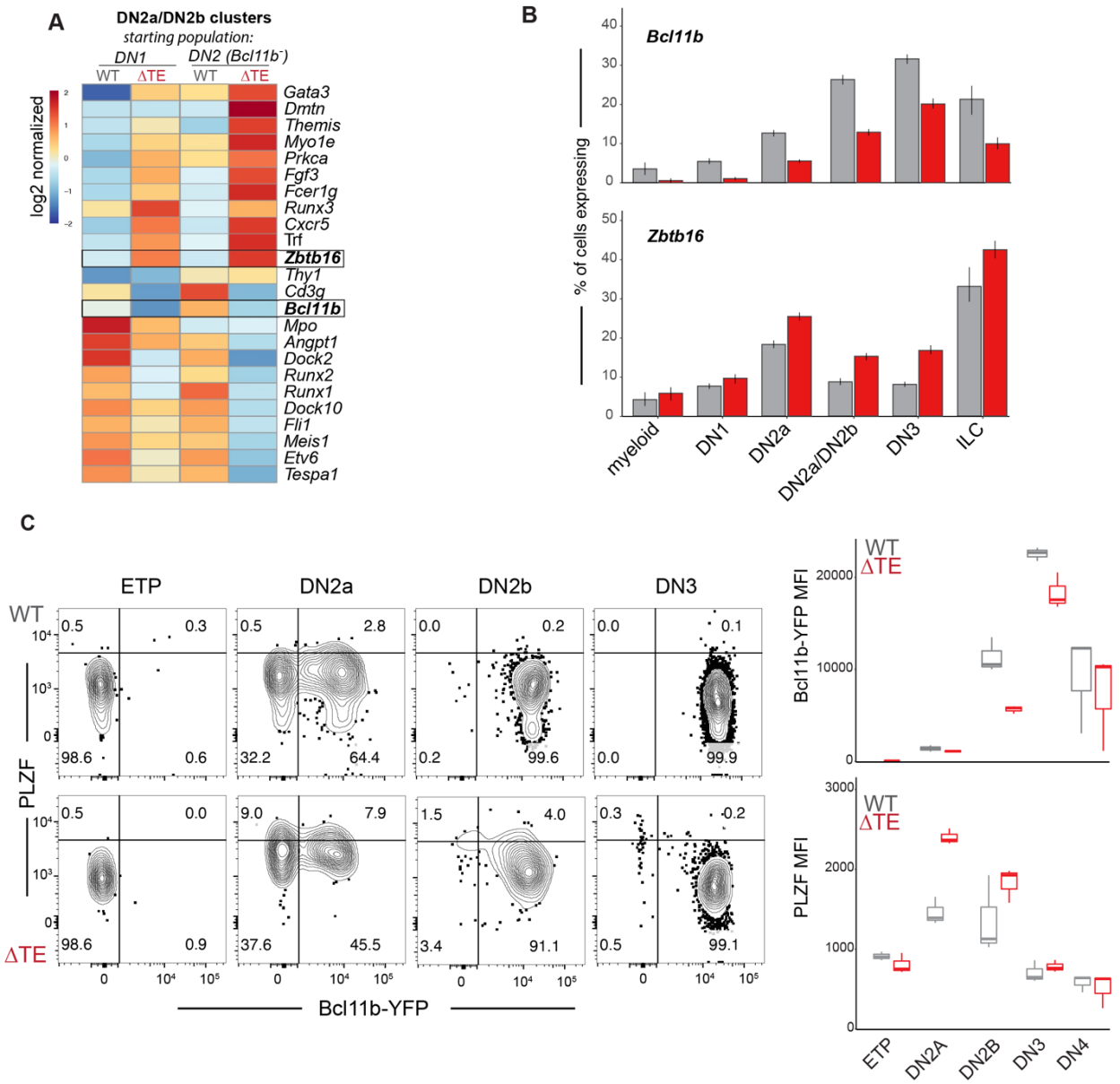


Figure 4.6 Delayed *Bcl11b* activation increases PLZF expression in DN2 progenitors
(A) Top differentially expressed immune-related genes within DN2a/DN2b cells (clusters 3 and 4 in **Figure 4.3**) across samples. (B) Percentage of cells expressing *Bcl11b* and *Zbtb16* (minimum normalized value of 0.1) across clusters (from data presented in **Figure 4.3**) (C) Bcl11b-YFP reporter and PLZF protein expression analyzed by flow cytometry with contour plots (left) and median fluorescence intensity (right) for each stage of early T cell development. Cells are gated on lineage-negative cells (CD4, CD8, CD19, TCRγδ , Ter119, and Gr-1).

4.3 DISCUSSION

The mechanisms that precisely regulate cell-type numbers and proportions during development have remained unclear. Here we provide evidence that a cell-autonomous timing mechanism in the form of a *cis*-epigenetic switch, can be modulated to alter the relative proportions of differentiated immune cell-types. These results also provide new insights into regulatory principles underlying lineage-specification during thymoctye development. It has traditionally been thought that T cell lineage commitment occurs concurrently with the onset *Bcl11b* expression during the DN2b stage (Rothenberg et al., 2016). Here, we present three pieces of evidence that suggest ILC lineage potential is retained after *Bcl11b* expression and into the DN3 stage (**Figure 4.7**). First, our scRNA-seq results showed a prominent cluster with a highly distinct pro-ILC transcriptional program that was most closely related to the Bcl11b^{+} DN3-like cluster (**Figure 4.3**A-B). Second, many of the cells in the pro-ILC cluster were derived from purified and re-cultured DN2 Bcl11b^{+} progenitors (**Figure 4.3**A). Third, a small subset of CD122^{+} NKp cells also expressed Bcl11b which decreases as CD122 and DX5 expression increases *in vivo* (**Figure 4.4**E, **Figure 4.5**A-B). Lastly, purified CD122^{-}/Bcl11b^{+} DN3 progenitors were capable of silencing *Bcl11b* and up-regulating the pro-ILC makers PD-1, CD122, DX5 and ICOS, suggesting these cells retain potential for both NK and ILC2 lineages. Thus, we propose that rare latent ILC-potent progenitors persist into the DN3 stage (**Figure 4.7**), possibly only arising from DN2 progenitors that experienced a non-stereotyped order of transcription factor expression as explored in more detail below.

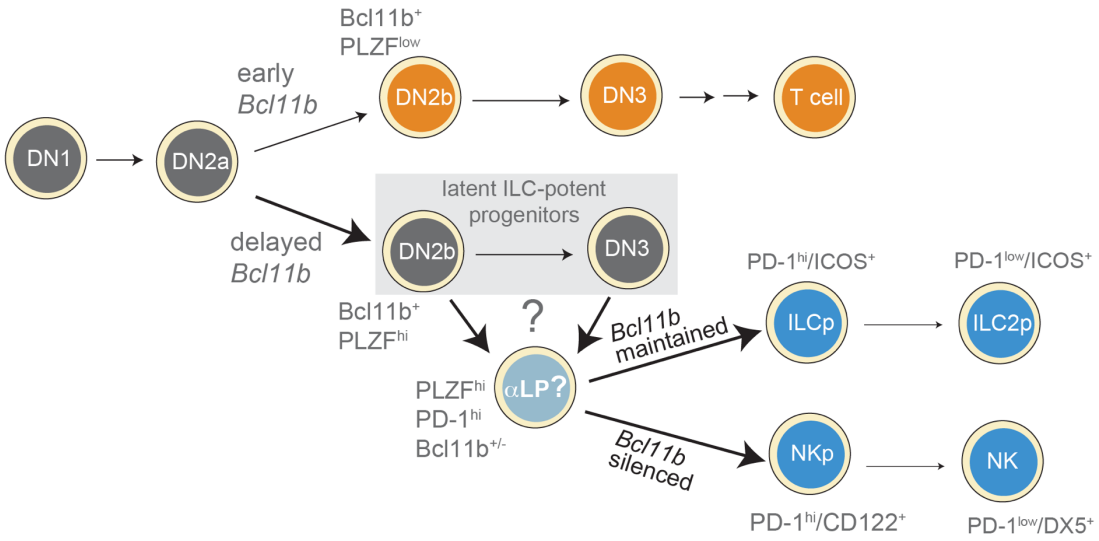


Figure 4.7 Hypothetical model for enhanced ILC lineage commitment from T cell progenitors with delayed *Bcl11b* activation
In this model, DN2 progenitors which up-regulate PLZF before Bcl11b can enter a "latent ILC-potent" progenitor state that appears indistinguishable from T cell committed DN2b and DN3 progenitors. These ILC-potent DN2b and DN3 progenitors may pass through an α lymphoid progenitor (αLP)-like state before diverging into ILC2- or NK-lineage precursors. *Bcl11b* is either maintained or up-regulated to drive ILC2-commitment and maturation from ILCp (ILC precursors) or silenced during NK-commitment and maturation from NK precursors (NKp). When Bcl11b activation is delayed, more progenitors enter the pool of "latent ILC-potent" progenitors but appear to be more likely to stall during ILCp and NKp differentiation. Arrow line width represents increases/decreases in that step in ΔTE thymocytes.

Given that the ILC2 lineage requires the expression of transcription factors that are normally expressed during DN2/DN3 stages (Huang et al., 2017; Kueh et al., 2016), it is perhaps unsurprising that T cell progenitors at this stage can diverge into the ILC2 lineage. However, our results shed light on the cell-autonomous mechanisms that regulate the balance of thymic ILC2 and T cell lineage specification. Specifically, the divergence from the T cell lineage to an alternative ILC lineage appears to be driven by a rare change in relative timing of gene activation events in progenitors at the DN2a stage, even in wild-type mice. Bcl11b is known to negatively regulate *Zbtb16*, thus it's possible that *Bcl11b* and *Zbtb16* or one of its target genes behave as competing timed switches during the DN2a stage. It's likely that in order to generate a latent ILC-

potent DN2/DN3 progenitor, the *Zbtb16*-mediated switch must occur before the *Bcl11b* switch, thus resulting in a shift in transcription activity of Bcl11b, once expressed, from an pro-T cell program to a pro-ILC program (Hosokawa et al., 2020). In an unperturbed system, it appears the *Bcl11b* switch occurs far more quickly on average than that of the *Zbtb16*-mediated switch and thus it is rare that the *Zbtb16*-mediated switch occurs first. However, removal of the *Bcl11b* timing enhancer appears to increase the likelihood of this non-stereotyped order by reducing the rate of the *Bcl11b* switch and would explain why ΔTE mice have more thymic pro-ILC than WT mice. This may also explain why others have found that, under wild type conditions, a small fraction of peripheral ILC2s appear to be derived from thymic DN3 progenitors but that fraction dramatically increases when the *Zbtb16* repressors, E2A/HEB E proteins, are disrupted (Qian et al.). Interestingly, *Zbtb16* appears to also be regulated by its own timing enhancer during natural killer T cell development which supports the hypothesis that *Bcl11b* and *Zbtb16* may behave as competing timed switches during T cell and ILC lineage bifurcation (Mao et al., 2017).

This model for competing timed switches regulating a latent ILC-potent DN2/DN3 programs can be tested in the future by purifying and re-culturing bone marrow derived Bcl11b^{-} DN2 progenitors at early and late time points during T cell development. If the model were true, we would expect that the Bcl11b^{-} DN2 progenitors at the late time point to generate more ILC lineage committed progenitors that those at the early time point as a result of being enriched for progenitors that had initiated the pro-ILC transcriptional program before *Bcl11b* was activated. Additionally, it may be that elevated PLZF expression during the DN2 stage induces this latent ILC-potent DN2/DN3 progenitor state. Similarly, PLZF expression levels appear to play a critical role in inducing invariant natural killer T cell (iNKT) differentiation from immature T cells later in development (Park et al., 2019).

Perhaps the most surprising result was that timely but transient *Bcl11b* expression appears to be important for the maturation of NKp cells even after *Bcl11b* is silenced and NK-lineage commitment has occurred (**Figure 4.4**D). Our thymocyte re-culturing experiments demonstrated that Bcl11b^{+}/Lin^{-} DN3 progenitors can silence *Bcl11b* and generate CD122^{+}/DX5^{+} mature NK cells *in vitro* even in the constant presence of Notch signaling (**Figure 4.4**E). Furthermore, Bcl11b^{+} DN3 progenitors from ΔTE mice appear to have greater NK cell lineage potential and differentiate more quickly into mature DX5^{+}/PD-1^{-} NK cells, suggesting that they possess normal, if not enhanced, NK cell maturation (**Figure 4.4**E). Additionally, we found that delaying Bcl11b activation *in vivo* generates a rare population of Bcl11b^{-}/PLZF^{hi} DN3 progenitors which may be poised to differentiate into NK cells without ever transiently expressing *Bcl11b* (**Figure 4.6**C), which we have previously shown is more difficult to activate when delayed to DN3 stage (Ng et al., 2018). *Zbtb16* is generally quickly silenced during NK cell development (Constantinides et al., 2015) and thus it's possible that transient expression of Bcl11b, a negative regulator of *Zbtb16*, serves to facilitate this silencing during NK cell maturation. Without this rapid silencing, PLZF may generate a more stalled progenitor state marked by elevated PD-1 expression (**Figure 4.4**D). Thus, not only is *Bcl11b* activation timing critical regulating the initial bifurcation of thymic ILC and T cells, but it is also important for downstream differentiation of ILC sub-lineages even when usually only transiently expressed. Together, these results emphasize the importance of lineage-specific gene activation timing in regulating parallel gene circuits which need to unfold in a particular order relative to one another. Here, we have shown that changes in the timing and order of lineage-specifying gene activation can finely calibrate cell-type numbers and functionality and may play critical a role in generating phenotypic variation.

4.4 METHODS

Animal models

C57BL/6 *Bcl11b*$^{YFP/YFP}$ mice (i.e. "WT" mice) were generated as described in *Chapter 2.4 Methods. Bcl11b*$^{YFP\Delta Enh/YFP\Delta Enh}$ mice (i.e. "ΔTE") were generated as previously describe (Ng et al., 2018). *Bcl11b*$^{YFP/YFP}$ were crossed to *Bcl11b*$^{YFP\Delta Enh/YFP\Delta Enh}$ and the resulting *Bcl11b*$^{YFP/YFP\Delta Enh}$ heterozygotes were used as breeding pairs to generate littermate match controls for primary cell analysis. All animals were bred and maintained at the University of Washington. All animal protocols were reviewed and approved by the Institute Animal Care and Use Committee at the University of Washington (Protocol No: 4397-01).

Cell line culture

See Chapter 2.4 Methods

Cell purification

Bone marrow progenitors used for *in vitro* T cell development assays were purified described in *Chapter 3.4 Methods*. Thymi and spleens from 3-week old mice were mechanically dissociated before re-suspending in Fc blocking solution with 2.4G2 hybridoma supernatant. Early stage thymocytes (ETP-DN4) and ILC populations were depleted of CD4 and CD8 thymocytes before analysis or sorting. Thymocyte suspensions were labeled with biotinylated CD4 and CD8 antibodies incubated with MACS Streptavidin Microbeads (Miltenyi, Biotec) in HBH buffer (HBSS (Gibco), 0.5% BSA (Sigma-Aldrich), 10 mM HEPES, (Gibco)), pre-filtered through cell separation magnet (BD Biosciences), and passed through a magnetic column (Miltenyi Biotec).

***In vitro* differentiation of T cell progenitors**

See Chapter 2.4 Methods and figure legends for experiment specific cytokine conditions used.

Flow cytometry and cell sorting

See Chapter 2.4 Methods and figure legends for experiment specific cell surface marker used.

Single cell RNA-sequencing

Bone marrow progenitors attached to OP9-DL1 stromal cells were wash three times with PBS before scraping in HBH buffer and passing through a 70uM mesh filter to minimize OP9-DL1 stromal cell contamination. Cells were then washed with PBS and fixed in 4% paraformaldehyde for 15 minutes on ice. After fixation, cells were pelleted at 500xg for 3 minutes at 4°C before resuspending in 1mL of PBSR [PBS, pH 7.4, 1% BSA, 1% SUPERase-In RNase Inhibitor (ThermoFisher Scientific), and 1% 10mM DTT] and pelleting again at 500xg for 5 minutes at 4°C. The Brotman Baty Institute (BBI) then performed the single cell library construction, sequencing, and data processing. Only cells with less than 30% mitochondrial UMI and between $10^{2.5}$ to $10^{3.5}$ total UMIs were included in the downstream analysis.

Uniform Manifold Approximation and Projection (UMAP) (McInnes et al., 2018) and Louvain clustering (Blondel et al., 2008) were used in the form of the reduce_dimension and cluster_cells functions in Monocle3 package (v0.2.1). The following reduce_dimension parameters were used: umap.min_dist = 0.2, umap.n_neighbors = 25, umap.fast_sgd = TRUE. The cluster_cells parameter *k* was set to 30. For differential gene expression analyses, we first created a subset of the data with only genes from the immune system process gene ontology list (GO:0002376) before using the fit_models function in Monocle3. Differentially expressed genes with a q_value < 0.01 were then visualized using the pheatmap function from the pheatmap package. The plot_percent_cells_positive function from Monocle 3 was used to generate bar plots of the percentage of cells positive for *Bcl11b* or *Zbtb16* with min_expr parameter set to 0.1.

4.5 ACKNOWLEDGMENTS

This work was done in collaboration with Kenny Ng, who generated all of the transgenic mouse strains, and Lihua Chen, who helped process primary tissues for flow cytometry analysis. This work was also done in collaboration with the Brotman Baty Institute for Precision Medicine at University of Washington School of Medicine (BBI). The BBI performed the single cell library construction, sequencing, and data processing. We thank the Kueh group for constant discussion and feedback.

Chapter 5. CONCLUDING REMARKS

Here, I have detailed new approaches to interrogating the molecular mechanisms that regulate the timing of gene activation and lineage-specification during development. With these approaches, we have provided new insights into some of the regulatory components involved in controlling the temporal gene expression patterns and their role in regulating the number and types of cells that emerge during a developmental process. At the core of this cell-autonomous control of developmental timing is a *cis*-epigenetic switching mechanism that possesses two unique features that are compatible developmental gene regulatory networks. First, this mechanism is finely tunable by regulatory inputs over long timescales spanning many cell generations and thus long delays in lineage-specification can be precisely calibrated. Second, this mechanism operates independently of cell division and thus temporal schedules of developmental events can be upheld amid variable cell cycle rates. While I have focused on the implications *cis*-epigenetic switches in the context of developmental GRNs, we believe that they likely also have important roles more broadly in other contexts of effector functions in differentiated cells. For example, many cytokine genes are also regulated by *cis*-epigenetic switches which could be important for precise temporal control of immune effector cell functions and the balance of pathogen clearance and self-tolerance (Chu et al., 2021; Oyler-Yaniv et al., 2020).

While reflecting on the work presented in this dissertation, we have been reminded of John Gerhart and Marc Kirschner's theory of 'facilitated variation' in which a set of conserved core components and processes facilitate viable phenotypic variation on which the process of selection can act (Gerhart and Kirschner, 2007). A critical feature of this theory is that the 'core processes' have unique properties that allow for genetic changes to result in regulatory changes that are both novel and viable. The regulatory changes involve the reuse of the same core components but with

different behaviors. We propose that *cis*-epigenetic switches offer a mechanism by which the behavior (i.e. timing) of a core process (i.e. gene expression) can be altered to generate viable phenotypic variation (i.e. changes in the numbers and types of cells within a tissue). Importantly, we show that the *cis*-epigenetic switching mechanism is controlled by multiple regulatory inputs: chromatin modifying enzymes, transcription factors and *cis*-regulatory elements. Mutations in chromatin modifying enzymes could allow entire networks of *cis*-epigenetic switches to be simultaneously temporally scaled such that the duration of each developmental stage remains proportional, which has been observed for changes in PRC2 activity in the context of cerebral cortex development (Pereira et al., 2010). Alternatively, as we show in *Chapter 4*, mutations in *cis*-regulatory elements can allow for adjustments in the timing of individual *cis*-epigenetic switches that can change timing of one developmental event relative to another. Therefore, through numerous mechanisms, *cis*-epigenetic switching can enable phenotypic variation to emerge. Gaining a more complete understanding of the molecular mechanisms that carefully calibrate these *cis*-epigenetic switches and how they interact with cell non-autonomous mechanisms could provide a new perspective on the sources of disease-causing variation and offer new tools for synthetic biology and bioengineering.

Chapter 6. REFERENCES

Ackers, G.K., Johnson, A.D., and Shea, M.A. (1982). Quantitative model for gene regulation by lambda phage repressor. Proc. Natl. Acad. Sci. U. S. A. *79*, 1129–1133.

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.

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

Alon, U. (2007). Network motifs: theory and experimental approaches. Nat. Rev. Genet. *8*, 450–461.

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

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.

Bahr, C., von Paleske, L., Uslu, V.V., Remeseiro, S., Takayama, N., Ng, S.W., Murison, A., Langenfeld, K., Petretich, M., Scognamiglio, R., et al. (2018). A Myc enhancer cluster regulates normal and leukaemic haematopoietic stem cell hierarchies. Nature *553*, 515–520.

Barlow, D.P., and Bartolomei, M.S. (2014). Genomic Imprinting in Mammals. Cold Spring Harb. Perspect. Biol. *6*, a018382–a018382.

Barry, C., Schmitz, M.T., Jiang, P., Schwartz, M.P., Duffin, B.M., Swanson, S., Bacher, R., Bolin, J.M., Elwell, A.L., McIntosh, B.E., et al. (2017). Species-specific developmental timing is maintained by pluripotent stem cells ex utero. Dev. Biol. *423*, 101–110.

Bender, M.A., Ragoczy, T., Lee, J., Byron, R., Telling, A., Dean, A., and Groudine, M. (2012). The hypersensitive sites of the murine -globin locus control region act independently to affect nuclear localization and transcriptional elongation. Blood *119*, 3820–3827.

Berry, S., Hartley, M., Olsson, T.S.G., Dean, C., and Howard, M. (2015). Local chromatin environment of a Polycomb target gene instructs its own epigenetic inheritance. ELife *4*.

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., Buchler, N.E., Garcia, H.G., Gerland, U., Hwa, T., Kondev, J., and Phillips, R. (2005). Transcriptional regulation by the numbers: models. Curr. Opin. Genet. Dev. *15*, 116–124.

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.

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. J. Stat. Mech. Theory Exp. *2008*, P10008.

Bolouri, H., and Davidson, E.H. (2002). Modeling transcriptional regulatory networks. BioEssays *24*, 1118–1129.

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.

Calder, W.A. (1996). Size, Function, and Life History (Courier Corporation).

Canver, M.C., Bauer, D.E., Dass, A., Yien, Y.Y., Chung, J., Masuda, T., Maeda, T., Paw, B.H., and Orkin, S.H. (2014). Characterization of Genomic Deletion Efficiency Mediated by Clustered Regularly Interspaced Palindromic Repeats (CRISPR)/Cas9 Nuclease System in Mammalian Cells. J. Biol. Chem. *289*, 21312–21324.

Carey, M. (1998). The Enhanceosome and Transcriptional Synergy. Cell *92*, 5–8.

Carroll, S.B. (2008). Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution. Cell *134*, 25–36.

Chaudhri, V.K., Dienger-Stambaugh, K., Wu, Z., Shrestha, M., and Singh, H. (2020). Charting the cis-regulome of activated B cells by coupling structural and functional genomics. Nat. Immunol. *21*, 210–220.

Chen, J.C.J., and Goldhamer, D.J. (2004). The core enhancer is essential for proper timing of MyoD activation in limb buds and branchial arches. Dev. Biol. *265*, 502–512.

Chen, Z., and Zhang, Y. (2020). Maternal H3K27me3-dependent autosomal and X chromosome imprinting. Nat. Rev. Genet. *21*, 555–571.

Choi, J., Lysakovskaia, K., Stik, G., Demel, C., Söding, J., Tian, T.V., Graf, T., and Cramer, P. (2021). Evidence for additive and synergistic action of mammalian enhancers during cell fate determination. ELife *10*, e65381.

Chu, J.M., Pease, N.A., and Kueh, H.Y. (2021). In search of lost time: Enhancers as modulators of timing in lymphocyte development and differentiation. Immunol. Rev. *300*, 134–151.

Clement, K., Rees, H., Canver, M.C., Gehrke, J.M., Farouni, R., Hsu, J.Y., Cole, M.A., Liu, D.R., Joung, J.K., Bauer, D.E., et al. (2019). CRISPResso2 provides accurate and rapid genome editing sequence analysis. Nat. Biotechnol. *37*, 224–226.

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*.

Constantinides, M.G., Gudjonson, H., McDonald, B.D., Ishizuka, I.E., Verhoef, P.A., Dinner, A.R., and Bendelac, A. (2015). PLZF expression maps the early stages of ILC1 lineage development. Proc. Natl. Acad. Sci. *112*, 5123–5128.

Davis-Dusenbery, B.N., Williams, L.A., Klim, J.R., and Eggan, K. (2014). How to make spinal motor neurons. Development *141*, 491–501.

De Beer, G. (1940). Embryos and ancestors (Clarendon Press).

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.

Diao, Y., Fang, R., Li, B., Meng, Z., Yu, J., Qiu, Y., Lin, K.C., Huang, H., Liu, T., Marina, R.J., et al. (2017). A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. Nat. Methods *14*, 629–635.

Dobrinić, P., Szczurek, A.T., and Klose, R.J. (2020). PRC1 drives Polycomb-mediated gene repression by controlling transcription initiation and burst frequency. BioRxiv 2020.10.09.333294.

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*.

Dukler, N., Gulko, B., Huang, Y.-F., and Siepel, A. (2017). Is a super-enhancer greater than the sum of its parts? Nat. Genet. *49*, 2–3.

Ebisuya, M., and 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., and 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.

Engreitz, J.M., Haines, J.E., Perez, E.M., Munson, G., Chen, J., Kane, M., McDonel, P.E., Guttman, M., and Lander, E.S. (2016). Local regulation of gene expression by lncRNA promoters, transcription and splicing. Nature *539*, 452–455.

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.

Espuny-Camacho, I., Michelsen, K.A., Gall, D., Linaro, D., Hasche, A., Bonnefont, J., Bali, C., Orduz, D., Bilheu, A., Herpoel, A., et al. (2013). Pyramidal neurons derived from human pluripotent stem cells integrate efficiently into mouse brain circuits in vivo. Neuron *77*, 440–456.

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.

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.

Ferreira, A.C.F., Szeto, A.C.H., Heycock, M.W.D., Clark, P.A., Walker, J.A., Crisp, A., Barlow, J.L., Kitching, S., Lim, A., Gogoi, M., et al. (2021). RORα is a critical checkpoint for T cell and ILC2 commitment in the embryonic thymus. Nat. Immunol. *22*, 166–178.

Festenstein, R., Tolaini, M., Corbella, P., Mamalaki, C., Parrington, J., Fox, M., Miliou, A., Jones, M., and Kioussis, D. (1996). Locus Control Region Function and Heterochromatin-Induced Position Effect Variegation. Sci. New Ser. *271*, 1123–1125.

Francis, N.J. (2004). Chromatin Compaction by a Polycomb Group Protein Complex. Science *306*, 1574–1577.

Fuhrmann, J.F., Buono, L., Adelmann, L., Martinez-Morales, J.R., and Centanin, L. (2020). Genetic developmental timing revealed by inter-species transplantations in fish. Development *147*.

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.

Gaspard, N., Bouschet, T., Hourez, R., Dimidschstein, J., Naeije, G., van den Ameele, J., Espuny-Camacho, I., Herpoel, A., Passante, L., Schiffmann, S.N., et al. (2008). An intrinsic mechanism of corticogenesis from embryonic stem cells. Nature *455*, 351–357.

Gasperini, M., Tome, J.M., and Shendure, J. (2020). Towards a comprehensive catalogue of validated and target-linked human enhancers. Nat. Rev. Genet. *21*, 292–310.

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.

Gendrel, A.-V., Marion-Poll, L., Katoh, K., and Heard, E. (2016). Random monoallelic expression of genes on autosomes: Parallels with X-chromosome inactivation. Semin. Cell Dev. Biol. *56*, 100–110.

Gérard, M., Zákány, J., and Duboule, D. (1997). Interspecies Exchange of a Hoxd Enhancer in Vivo Induces Premature Transcription and Anterior Shift of the Sacrum. Dev. Biol. *190*, 32–40.

Gerhart, J., and Kirschner, M. (2007). The theory of facilitated variation. Proc. Natl. Acad. Sci. *104*, 8582–8589.

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.

Giese, K., Kingsley, C., Kirshner, J.R., and Grosschedl, R. (1995). Assembly and function of a TCR alpha enhancer complex is dependent on LEF-1-induced DNA bending and multiple protein-protein interactions. Genes Dev. *9*, 995–1008.

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.

Goldberg, A.D., Allis, C.D., and Bernstein, E. (2007). Epigenetics: A Landscape Takes Shape. Cell *128*, 635–638.

Gould, S.J. (1977). Ontogeny and Phylogeny (Cambridge, Mass: Harvard University Press).

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.

Guy, L.G., Kothary, R., DeRepentigny, Y., Delvoye, N., Ellis, J., and Wall, L. (1996). The beta-globin locus control region enhances transcription of but does not confer position-independent expression onto the lacZ gene in transgenic mice. EMBO J. *15*, 3713–3721.

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.

Hay, D., Hughes, J.R., Babbs, C., Davies, J.O.J., Graham, B.J., Hanssen, L.L.P., Kassouf, M.T., Oudelaar, A.M., Sharpe, J.A., Suciu, M.C., et al. (2016). Genetic dissection of the α-globin super-enhancer in vivo. Nat. Genet. *48*, 895–903.

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.

Hosokawa, H., Romero-Wolf, M., Yang, Q., Motomura, Y., Levanon, D., Groner, Y., Moro, K., Tanaka, T., and Rothenberg, E.V. (2020). Cell type–specific actions of Bcl11b in early T-lineage and group 2 innate lymphoid cells. J. Exp. Med. *217*, e20190972.

Howard, Jonathon (2001). Mechanics of Motor Proteins and the Cytoskeleton (Sinauer Associates).

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.

Huang, J., Liu, X., Li, D., Shao, Z., Cao, H., Zhang, Y., Trompouki, E., Bowman, T.V., Zon, L.I., Yuan, G.-C., et al. (2016). Dynamic Control of Enhancer Repertoires Drives Lineage and Stage-Specific Transcription during Hematopoiesis. Dev. Cell *36*, 9–23.

Huang, J., Li, K., Cai, W., Liu, X., Zhang, Y., Orkin, S.H., Xu, J., and Yuan, G.-C. (2018). Dissecting super-enhancer hierarchy based on chromatin interactions. Nat. Commun. *9*, 943.

Huang, Q., Seillet, C., and Belz, G.T. (2017). Shaping Innate Lymphoid Cell Diversity. Front. Immunol. *8*.

Huxley, J.S., Pigliucci, M., and Müller, G.B. (1942). Evolution: The Modern Synthesis (London: Allen and Unwin).

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.

Inoue, F., and Ahituv, N. (2015). Decoding enhancers using massively parallel reporter assays. Genomics *106*, 159–164.

Isoda, T., Moore, A.J., He, Z., Chandra, V., Aida, M., Denholtz, M., Piet van Hamburg, J., Fisch, K.M., Chang, A.N., Fahl, S.P., et al. (2017). Non-coding Transcription Instructs Chromatin Folding and Compartmentalization to Dictate Enhancer-Promoter Communication and T Cell Fate. Cell *171*, 103-119.e18.

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.

Iwafuchi, M., Cuesta, I., Donahue, G., Takenaka, N., Osipovich, A.B., Magnuson, M.A., Roder, H., Seeholzer, S.H., Santisteban, P., and Zaret, K.S. (2020). Gene network transitions in embryos depend upon interactions between a pioneer transcription factor and core histones. Nat. Genet. *52*, 418–427.

Iwafuchi-Doi, M., and Zaret, K.S. (2014). Pioneer transcription factors in cell reprogramming. Genes Dev. *28*, 2679–2692.

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

Jadhav, U., Nalapareddy, K., Saxena, M., O'Neill, N.K., Pinello, L., Yuan, G.-C., Orkin, S.H., and Shivdasani, R.A. (2016). Acquired Tissue-Specific Promoter Bivalency Is a Basis for PRC2 Necessity in Adult Cells. Cell *165*, 1389–1400.

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 *78*, 141-151.e5.

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

Johnson, J.L., Georgakilas, G., Petrovic, J., Kurachi, M., Cai, S., Harly, C., Pear, W.S., Bhandoola, A., Wherry, E.J., and Vahedi, G. (2018). Lineage-Determining Transcription Factor TCF-1 Initiates the Epigenetic Identity of T Cells. Immunity *48*, 243-257.e10.

Jones, R., Cosway, E.J., Willis, C., White, A.J., Jenkinson, W.E., Fehling, H.J., Anderson, G., and Withers, D.R. (2018). Dynamic changes in intrathymic ILC populations during murine neonatal development. Eur. J. Immunol. *48*, 1481–1491.

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. gkw701.

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.

Karst, S.M., Ziels, R.M., Kirkegaard, R.H., Sørensen, E.A., McDonald, D., Zhu, Q., Knight, R., and Albertsen, M. (2021). High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing. Nat. Methods *18*, 165–169.

Kernfeld, E.M., Genga, R.M.J., Neherin, K., Magaletta, M.E., Xu, P., and Maehr, R. (2018). A Single-Cell Transcriptomic Atlas of Thymus Organogenesis Resolves Cell Types and Developmental Maturation. Immunity *48*, 1258-1270.e6.

Kosicki, M., Tomberg, K., and Bradley, A. (2018). Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. Nat. Biotechnol. *36*, 765–771.

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

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.

Krehenwinkel, H., Pomerantz, A., Henderson, J.B., Kennedy, S.R., Lim, J.Y., Swamy, V., Shoobridge, J.D., Graham, N., Patel, N.H., Gillespie, R.G., et al. (2019). Nanopore sequencing of long ribosomal DNA amplicons enables portable and simple biodiversity assessments with high phylogenetic resolution across broad taxonomic scale. GigaScience *8*.

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.

Krueger, A., Ziętara, N., and Łyszkiewicz, M. (2017). T Cell Development by the Numbers. Trends Immunol. *38*, 128–139.

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 action 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.

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.

Larsson, A.J.M., Johnsson, P., Hagemann-Jensen, M., Hartmanis, L., Faridani, O.R., Reinius, B., Segerstolpe, Å., Rivera, C.M., Ren, B., and Sandberg, R. (2019a). Genomic encoding of transcriptional burst kinetics. Nature.

Larsson, A.J.M., Reinius, B., Jacob, T., Dalessandri, T., Hendriks, G.-J., Kasper, M., and Sandberg, R. (2019b). Transcriptional bursts explain autosomal random monoallelic expression and affect allelic imbalance (Genomics).

Lee, H.K., Willi, M., Kuhns, T., Liu, C., and Hennighausen, L. (2021). Redundant and non-redundant cytokine-activated enhancers control Csn1s2b expression in the lactating mouse mammary gland. Nat. Commun. *12*, 2239.

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.

Levine, M. (2010). Transcriptional Enhancers in Animal Development and Evolution. Curr. Biol. *20*, R754–R763.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics *34*, 3094–3100.

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, L., Leid, M., and Rothenberg, E.V. (2010a). An early T cell lineage commitment checkpoint dependent on the transcription factor Bcl11b. Science *329*, 89–93.

Li, L., Zhang, J.A., Dose, M., Kueh, H.Y., Mosadeghi, R., Gounari, F., and Rothenberg, E.V. (2013). A far downstream enhancer for murine Bcl11b controls its T-cell specific expression. Blood *122*, 902–911.

Li, P., Burke, S., Wang, J., Chen, X., Ortiz, M., Lee, S.-C., Lu, D., Campos, L., Goulding, D., Ng, B.L., et al. (2010b). Reprogramming of T Cells to Natural Killer-Like Cells upon Bcl11b Deletion. Science *329*, 85–89.

Li, Q., Peterson, K.R., Fang, X., and Stamatoyannopoulos, G. (2002). Locus control regions. Blood *100*, 3077–3086.

Long, H.K., Prescott, S.L., and Wysocka, J. (2016). Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. Cell *167*, 1170–1187.

Longabaugh, W.J.R., Zeng, W., Zhang, J.A., Hosokawa, H., Jansen, C.S., Li, L., Romero-Wolf, M., Liu, P., Kueh, H.Y., Mortazavi, A., et al. (2017). Bcl11b and combinatorial resolution of cell fate in the T-cell gene regulatory network. Proc. Natl. Acad. Sci. *114*, 5800–5807.

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.

Magklara, A., and Lomvardas, S. (2013). Stochastic gene expression in mammals: lessons from Olfaction. Trends Cell Biol. *23*, 449–456.

Man, J.C.K., Mohan, R.A., Boogaard, M. van den, Hilvering, C.R.E., Jenkins, C., Wakker, V., Bianchi, V., Laat, W. de, Barnett, P., Boukens, B.J., et al. (2019). An enhancer cluster controls gene activity and topology of the SCN5A-SCN10A locus in vivo. Nat. Commun. *10*, 4943.

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.

Mao, A.-P., Ishizuka, I.E., Kasal, D.N., Mandal, M., and Bendelac, A. (2017). A shared Runx1-bound Zbtb16 enhancer directs innate and innate-like lymphoid lineage development. Nat. Commun. *8*, 863.

Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. Nature *469*, 343–349.

Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury Iii, 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.

Mariani, L., Schulz, E.G., Lexberg, M.H., Helmstetter, C., Radbruch, A., Löhning, M., and Höfer, T. (2010). Short-term memory in gene induction reveals the regulatory principle behind stochastic IL-4 expression. Mol. Syst. Biol. *6*.

Matsuda, M., Hayashi, H., Garcia-Ojalvo, J., Yoshioka-Kobayashi, K., Kageyama, R., Yamanaka, Y., Ikeya, M., Toguchida, J., Alev, C., and Ebisuya, M. (2019). Species-specific oscillation periods of human and mouse segmentation clocks are due to cell autonomous differences in biochemical reaction parameters (Developmental Biology).

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.

McInnes, L., Healy, J., Saul, N., and Großberger, L. (2018). UMAP: Uniform Manifold Approximation and Projection. J. Open Source Softw. *3*, 861.

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.

Milot, E., Strouboulis, J., Trimborn, T., Wijgerde, M., de Boer, E., Langeveld, A., Tan-Un, K., Vergeer, W., Yannoutsos, N., Grosveld, F., et al. (1996). Heterochromatin Effects on the Frequency and Duration of LCR-Mediated Gene Transcription. Cell *87*, 105–114.

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

Moorthy, S.D., Davidson, S., Shchuka, V.M., Singh, G., Malek-Gilani, N., Langroudi, L., Martchenko, A., So, V., Macpherson, N.N., and Mitchell, J.A. (2017). Enhancers and super-enhancers have an equivalent regulatory role in embryonic stem cells through regulation of single or multiple genes. Genome Res. *27*, 246–258.

Moreau, C., Caldarelli, P., Rocancourt, D., Roussel, J., Denans, N., Pourquie, O., and Gros, J. (2019). Timed Collinear Activation of Hox Genes during Gastrulation Controls the Avian Forelimb Position. Curr. Biol. *29*, 35-50.e4.

Muerdter, F., Boryń, Ł.M., and Arnold, C.D. (2015). STARR-seq — Principles and applications. Genomics *106*, 145–150.

Nagel, S., Scherr, M., Kel, A., Hornischer, K., Crawford, G.E., Kaufmann, M., Meyer, C., Drexler, H.G., and MacLeod, R.A.F. (2007). Activation of TLX3 and NKX2-5 in t(5;14)(q35;q32) T-Cell Acute Lymphoblastic Leukemia by Remote 3'-BCL11B Enhancers and Coregulation by PU.1 and HMGA1. Cancer Res. *67*, 1461–1471.

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.

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.

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.

Osterwalder, M., Barozzi, I., Tissières, V., Fukuda-Yuzawa, Y., Mannion, B.J., Afzal, S.Y., Lee, E.A., Zhu, Y., Plajzer-Frick, I., Pickle, C.S., et al. (2018). Enhancer redundancy provides phenotypic robustness in mammalian development. Nature *554*, 239–243.

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.

Oyler-Yaniv, J., Oyler-Yaniv, A., Maltz, E., and Wollman, R. (2020). TNF controls a speed-accuracy tradeoff in the apoptotic decision to restrict viral spread (Systems Biology).

Panigrahi, A., and O'Malley, B.W. (2021). Mechanisms of enhancer action: the known and the unknown. Genome Biol. *22*, 108.

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.

Park, J.-Y., DiPalma, D.T., Kwon, J., Fink, J., and Park, J.-H. (2019). Quantitative Difference in PLZF Protein Expression Determines iNKT Lineage Fate and Controls Innate CD8 T Cell Generation. Cell Rep. *27*, 2548-2557.e4.

Parker, S.C.J., Stitzel, M.L., Taylor, D.L., Orozco, J.M., Erdos, M.R., Akiyama, J.A., van Bueren, K.L., Chines, P.S., Narisu, N., NISC Comparative Sequencing Program, et al. (2013). Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. Proc. Natl. Acad. Sci. *110*, 17921–17926.

Pereira, J.D., Sansom, S.N., Smith, J., Dobenecker, M.-W., Tarakhovsky, A., and Livesey, F.J. (2010). Ezh2, the histone methyltransferase of PRC2, regulates the balance between self-renewal and differentiation in the cerebral cortex. Proc. Natl. Acad. Sci. *107*, 15957–15962.

Petrie, H.T., and Zúñiga-Pflücker, J.C. (2007). Zoned Out: Functional Mapping of Stromal Signaling Microenvironments in the Thymus. Annu. Rev. Immunol. *25*, 649–679.

Petryk, N., Bultmann, S., Bartke, T., and Defossez, P.-A. (2021). Staying true to yourself: mechanisms of DNA methylation maintenance in mammals. Nucleic Acids Res. *49*, 3020–3032.

Plys, A.J., Davis, C.P., Kim, J., Rizki, G., Keenen, M.M., Marr, S.K., and Kingston, R.E. (2019). Phase separation of Polycomb-repressive complex 1 is governed by a charged disordered region of CBX2. Genes Dev. *33*, 799–813.

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.

Pott, S., and Lieb, J.D. (2015). What are super-enhancers? Nat. Genet. *47*, 8–12.

Ptashne, M. (2004). A Genetic Switch: Phage Lambda Revisited (CSHL Press).

Ptashne, M. (2009). Binding reactions: epigenetic switches, signal transduction and cancer. Curr. Biol. *19*, R234–R241.

Qian, L., Bajana, S., Georgescu, C., Peng, V., Wang, H.-C., Adrianto, I., Colonna, M., Alberola-Ila, J., Wren, J.D., and Sun, X.-H. Suppression of ILC2 differentiation from committed T cell precursors by E protein transcription factors. 16.

Raff, M. (2007). Intracellular Developmental Timers. Cold Spring Harb. Symp. Quant. Biol. *72*, 431–435.

Ramos, C.V., Ballesteros-Arias, L., Silva, J.G., Paiva, R.A., Nogueira, M.F., Carneiro, J., Gjini, E., and Martins, V.C. (2020). Cell Competition, the Kinetics of Thymopoiesis, and Thymus Cellularity Are Regulated by Double-Negative 2 to 3 Early Thymocytes. Cell Rep. *32*, 107910.

Rayon, T., and Briscoe, J. Cross-species comparisons and in vitro models to study tempo in development and homeostasis. Interface Focus *11*, 20200069.

Rayon, T., Stamataki, D., Perez-Carrasco, R., Garcia-Perez, L., Barrington, C., Melchionda, M., Exelby, K., Lazaro, J., Tybulewicz, V.L.J., Fisher, E.M.C., et al. (2020). Species-specific pace of development is associated with differences in protein stability. Science *369*.

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.

Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. PeerJ *4*, e2584.

Rothenberg, E.V. (2019). Programming for T-lymphocyte fates: modularity and mechanisms. Genes Dev. *33*, 1117–1135.

Rothenberg, E.V., Kueh, H.Y., Yui, M.A., and Zhang, J.A. (2016). Hematopoiesis and T-cell specification as a model developmental system. Immunol. Rev. *271*, 72–97.

Rv, P., Sundaresh, A., Karunyaa, M., Arun, A., and Gayen, S. (2021). Autosomal Clonal Monoallelic Expression: Natural or Artifactual? Trends Genet. *37*, 206–211.

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.

Scollay, R.G., Butcher, E.C., and Weissman, I.L. (1980). Thymus cell migration: Quantitative aspects of cellular traffic from the thymus to the periphery in mice. Eur. J. Immunol. *10*, 210–218.

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.

Shen, M.W., Arbab, M., Hsu, J.Y., Worstell, D., Culbertson, S.J., Krabbe, O., Cassa, C.A., Liu, D.R., Gifford, D.K., and Sherwood, R.I. (2018). Predictable and precise template-free CRISPR editing of pathogenic variants. Nature *563*, 646–651.

Shin, H.Y., Willi, M., Yoo, K.H., Zeng, X., Wang, C., Metser, G., and Hennighausen, L. (2016). Hierarchy within the mammary STAT5-driven Wap super-enhancer. Nat. Genet. *48*, 904–911.

Simeonov, D.R., Gowen, B.G., Boontanrart, M., Roth, T.L., Gagnon, J.D., Mumbach, M.R., Satpathy, A.T., Lee, Y., Bray, N.L., Chan, A.Y., et al. (2017). Discovery of stimulation-responsive immune enhancers with CRISPR activation. Nature *549*, 111–115.

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. *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.

Steffen, P.A., and Ringrose, L. (2014). What are memories made of? How Polycomb and Trithorax proteins mediate epigenetic memory. Nat. Rev. Mol. Cell Biol. *15*, 340–356.

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.

Su, X.-Y., Della-Valle, V., Andre-Schmutz, I., Lemercier, C., Radford-Weiss, I., Ballerini, P., Lessard, M., Lafage-Pochitaloff, M., Mugneret, F., Berger, R., et al. (2006). HOX11L2/TLX3 is transcriptionally activated through T-cell regulatory elements downstream of BCL11B as a result of the t(5;14)(q35;q32). Blood *108*, 4198–4201.

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.

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.

Vernimmen, D., Lynch, M.D., De Gobbi, M., Garrick, D., Sharpe, J.A., Sloane-Stanley, J.A., Smith, A.J.H., and Higgs, D.R. (2011). Polycomb eviction as a new distant enhancer function. Genes Dev. *25*, 1583–1588.

Vidigal, J.A., and Ventura, A. (2015). Rapid and efficient one-step generation of paired gRNA CRISPR-Cas9 libraries. Nat. Commun. *6*.

Vijayanand, P., Seumois, G., Simpson, L.J., Abdul-Wajid, S., Baumjohann, D., Panduro, M., Huang, X., Interlandi, J., Djuretic, I.M., Brown, D.R., et al. (2012). Interleukin-4 Production by Follicular Helper T Cells Requires the Conserved Il4 Enhancer Hypersensitivity Site V. Immunity *36*, 175–187.

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.

Walters, M.C., Magis, W., Fiering, S., Eidemiller, J., Scalzo, D., Groudine, M., and Martin, D.I. (1996). Transcriptional enhancers act in cis to suppress position-effect variegation. Genes Dev. *10*, 185–195.

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.

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.

Wu, X., and Zhang, Y. (2017). TET-mediated active DNA demethylation: mechanism, function and beyond. Nat. Rev. Genet. *18*, 517–534.

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.

Yu, Y., Wang, C., Clare, S., Wang, J., Lee, S.-C., Brandt, C., Burke, S., Lu, L., He, D., Jenkins, N.A., et al. (2015). The transcription factor Bcl11b is specifically expressed in group 2 innate lymphoid cells and is essential for their development. J. Exp. Med. *212*, 865–874.

Yu, Y., Tsang, J.C.H., Wang, C., Clare, S., Wang, J., Chen, X., Brandt, C., Kane, L., Campos, L.S., Lu, L., et al. (2016). Single-cell RNA-seq identifies a PD-1hi ILC progenitor and defines its development pathway. Nature *539*, 102–106.

Yui, M.A., and Rothenberg, E.V. (2014). Developmental gene networks: a triathlon on the course to T cell identity. Nat. Rev. Immunol. *14*, 529–545.

Yui, M.A., Feng, N., and Rothenberg, E.V. (2010). Fine-Scale Staging of T Cell Lineage Commitment in Adult Mouse Thymus. J. Immunol. *185*, 284–293.

Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes Dev. *25*, 2227–2241.

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.

Zheng, Y., Josefowicz, S., Chaudhry, A., Peng, X.P., Forbush, K., and Rudensky, A.Y. (2010). Role of conserved non-coding DNA elements in the Foxp3 gene in regulatory T-cell fate. Nature *463*, 808–812.

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.

Zou, Y.-R., Sunshine, M.-J., Taniuchi, I., Hatam, F., Killeen, N., and Littman, D.R. (2001). Epigenetic silencing of CD4 in T cells committed to the cytotoxic lineage. Nat. Genet. *29*, 332–336.