ABSTRACT

BRADSHAW, JUSTIN MICHAEL. Genetic and Epigenetic Regulation of Transcription during T Lymphocyte Development. (Under the direction of Michael L. Sikes).

The adaptive respond to invading pathogens requires lymphocytes to express a diverse array of antigen receptors. This diversity is generated through V(D)J recombination, a somatic rearrangement of the antigen receptor genes during early lymphopoiesis. Recombination proceeds in a carefully orchestrated sequence of individual rearrangements that are tied to and essential for developmental progression. The patterning of recombination is imposed epigenetically, such that developmentally timed transcription from promoters and enhancers positioned throughout each antigen receptor gene locus modulates recombinational accessibility of the transcribed gene segments. The goal of this doctoral thesis was to elucidate mechanisms that govern activation of antigen receptor germline promoters, and shed light on epigenetic programs that shape the immune repertoire and lineage commitment more broadly.

Activation of the T early alpha (TEA) promoter initiates *Tcra* assembly in $\alpha\beta$ thymocytes, directing accessibility of the proximal Jα segments to recombinase. TEA activity requires the chromosomal organizer CTCF, which facilitates juxtapositioning of TEA with the distal *Tcra* enhancer. However, while TEA is inactive prior to thymopoiesis, it is primed for activation much earlier. To investigate priming, I measured CTCF binding at TEA both before and after thymocyte differentiation. I found that CTCF binding was constitutive, being present in every tissue type or cell line I examined (e.g. placenta, embryonic stem cell (ESC), neuron, and B cell). Using ESC lines cultured under ground state conditions, I found that a roughly 3 kb region that separates the proximal Jα segments from the *Tcrd* locus, and which contains TEA, is targeted for DNA methylation very early in embryogenesis. Only the ~300 bp TEA promoter sequence is protected from this hypermethylation. I used RNAi-mediated knockdown of CTCF expression to demonstrate that in fact, CTCF appears to be the agent that protects TEA against hypermethylation. Moreover, I found

that mono-allelic deletion of TEA resulted in a biallelic increase in DNA methylation both upstream and downstream of the hypermethylated 3 kb region, suggesting a potential inter-allelic association at TEA that modulates the region's epigenetic structure. Together, these findings suggest a novel priming role for CTCF as a protector against TEA DNA methylation. It remains to be determined if CTCF may similarly prime promoters other that TEA.

In a second project, I explored the mechanisms that govern recombinational accessibility. *Tcrb* contains two DJβ cassettes that rearrange simultaneously in response to activation of associated promoters. Although D-to-J joining deletes the primary DJβ2 promoter, it activates a second promoter. To investigate the mechanistic link between recombination and Dβ2 promoter control, I created thymocyte cell lines with distinct DJβ repertoires. Analyses of these cell lines confirmed that dsDNA break-dependent loss of the stress-response factor USF-1 from the Dβ2 repressor site leads to activation of the second promoter, while breaks generated at DJβ1 failed to induce de-repression. Together, these findings suggest that antigen receptor repertoire development is not simply a stochastic accumulation of individual rearrangements, but also involves feedback responses that can alter local recombinational accessibility by redirecting germline transcription.

Genetic and Epigenetic Regulation of Transcription during T Lymphocyte Development

by
Justin Michael Bradshaw

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Microbiology

Raleigh, North Carolina

2013

APPROVED BY:

_____	_____
Dr. Jeffrey Yoder	Dr. Frank Scholle
_____	_____
Dr. Michael L. Sikes Chair of Advisory Committee	Dr. Scott M. Laster

DEDICATION

To my incredible wife Liz, whose motivation and loving assistance made the impossible possible.

BIOGRAPHY

Justin was born and raised in Eastern North Carolina and spent his formative years in Plymouth, a small paper mill town in Washington County. He always had a passion for learning and science, and became interested in immunology as a teenager when he wanted to find a cure for his terrible seasonal allergies. He left his hometown to attend the North Carolina School of Science and Mathematics in his junior and senior years of high school, and upon graduation attended North Carolina State University. He completed dual Bachelor's degrees in Biology in 2005 and Arts Applications: Music in 2006. Along with meeting his wife at NCSU, Justin realized he desired a career in teaching at the tertiary level after working as a professional tutor. He began a Master's in Microbiology in 2007 working with Dr. Sikes and ultimately moved this graduate coursework into a PhD program of Microbiology. Justin lives in Morrisville with his astounding wife Liz and his bunny Hanna.

ACKOWLEDGEMENTS

First and foremost, I would like to thank my advisor Dr. Michael Sikes, without whose funding, guidance and infinite patience this document would not exist. In addition, I would like to thank my current and former committee members (Drs. Scott Laster, Frank Scholle, Gregg Dean and Jeff Yoder) and collaborators Dr. Mike Krangel and Dr. Pierre Bushel as well as members of the Sikes lab past and present for technical assistance and valuable criticism. I appreciate greatly the camaraderie of fellow graduate students in the Department of Microbiology (RIP) as well as the excellent support services and conversations of Cindy Whitehead and Robert Davis. I thank my parents and in-laws for their infinite encouragement, my friends and *a cappella* groups Wolfgang and TTW for letting me blow off some steam, my undergraduate and graduate pupils both in and out of the lab for helping me remember why I started the program in the first place, and my wife and bunny for helping me get through those particularly difficult days.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: Literature Review

CHAPTER 2: CTCF Constancy and Developmental Progression at the TCRα locus

APPENDIX: RAG-DSB Feedback Control of Dβ2 Germline Transcription in a CD4/CD8 Double Negative T cell line acts through USF-1-mediated Promoter Repression

CHAPTER 1

Literature Review

INTRODUCTION

Multicellular organisms require differential collections of proteins in each tissue to correctly fulfill their various function(s). Since all somatic cells in a multicellular organism possess the same genetic code, these organisms possess a variety of factors to regulate the production of individual proteins. Much of this regulation occurs during the steps of gene expression, be it during transcription, splicing, translation or proper translocation of the resulting protein to its appropriate cellular compartment. However, a growing appreciation exists now for the elegant and extensive *cis*-acting mechanisms that regulate the availability of individual DNA sequences to the gene expression machinery. Many, if not most of these non-genetic changes that affect DNA bases directly are preserved during DNA replication and accompanying cell division. Given that such heritable changes alter gene function but not gene sequence, they are collectively referred to as epigenetic, or "above the gene" regulation.

Over the past few decades much has been learned about the various epigenetic marks found in the eukaryotic genome, and how they correlate with transcription and cell identity. Collation of such analyses has led to the positing of an "epigenetic code", though our ability to decipher such a code remains extremely primitive. Nevertheless, changes to a somewhat hierarchical ordering of particular epigenetic marks can clearly be traced through cellular development, correlating with the activation or silencing of individual genes. Progression from this original totipotent progenitor to pluripotent ES cells and then to the more differentiated cell types that make up the body's various tissues is accompanied by specific and numerous changes to the epigenetic code, many of the most substantial of which are noticed at lineage-specifc genes and genes involved in differentiation. This review of the literature regarding epigenetic regulation will (i)

provide a detailed analysis of epigenetic chromatin modifications, (ii) give a summary of our current understanding of cellular differentiation and epigenetic marks identified as relevant in this process, (iii) discuss the role of antigen receptor genes as lineage-specific genes involved in lymphoid development, as well as address the normal function of T cell receptors in T lymphocyte-mediated immune responses, and (iv) outline the central questions addressed by this thesis.

1.1 Overview of Epigenetics

In the simplest of terms, early epigenetics could be considered the intersection of developmental biology and genetics, two fields that took little notice of each other until the 1930s. In 1939, Edinburgh University's Dr. Conrad Waddington coined the term epigenetics from the Greek word epigenesis, a theory that proscribed the early embryo as undifferentiated and requiring a complex genetic program to achieve development of a complete organism [1]. This work was supported in later years by Dr. Ernst Hadorn, who discovered in *Drosophila* the existence of heritable patterns of differentiation among embryonic tissues which could be controlled and manipulated with specific chemical treatments [2].

The mechanisms of this epigenetic control over developmental progression at this point remained unelucidated, however experiments were quickly catching up to theory. The discovery in 1969 of non-canonical histone incorporation and their subsequent connection to chromatin structure and gene expression provided some early evidence of mechanism, though it took many years for this for these histone variants to rise to prominence [3-5]. Meanwhile the Mirsky laboratory identified that specific, reversible changes in histone structure independent of histone-DNA interaction can enhance or inhibit transcription of associated DNA, paving the way for our current understanding of post-translational histone modifications (PTMs) and the histone code [6,

7]. Finally, in the mid-seventies two groups independently suggested an association between epigenetic regulation and the reversible enzymatic methylation of cytosine bases within the DNA, leading to our current understanding of CpG methylation as a primary method of epigenetic control [8, 9]. These three epigenetic mechanisms are by far the most described in the literature and all but define our current understanding of epigenetics, and as such our discussions of epigenetic modifications will be confined to one or more of these mechanisms for the duration of the thesis.

1.2 Epigenetic Regulation

1.2.1 Chromatin configuration

Each human cell contains approximately 2 meters of linear dsDNA that is housed within a nucleus that is roughly 6 to 10 μm in diameter. Consequently, the DNA contained within a eukaryotic cell is heavily compacted through its association with octomeric histone complexes to form chromatin. The core particle of chromatin is the nucleosome. Nucleosomes consist of two copies of each of four histone proteins: H2A, H2B, H3 and H4. Each of these proteins contains a characteristic fold using a three-helix core domain that allows the protein to dimerize with its partner (H2A with H2B, H3 with H4) [10]. In the presence of DNA or high salt conditions, these multiple dimers coalesce into a ball-shaped octomer that is wrapped twice by dsDNA (Figure 1.1). Based on crystallography, each nucleosome contains 147 bp DNA [11], with a stretch of linker DNA that is somewhat variable but is on average around 38 bp in humans [12]. Visualization of nucleosome arrays led to the “beads on a string” description to describe the most-accessible form of chromatin [10, 13]. An additional linker histone, H1, serves to stabilize the DNA between nucleosomes and is also involved in chromatin condensation, with denser H1 concentration (up to one molecule per nucleosome in differentiated cells) correlating with less accessible DNA.

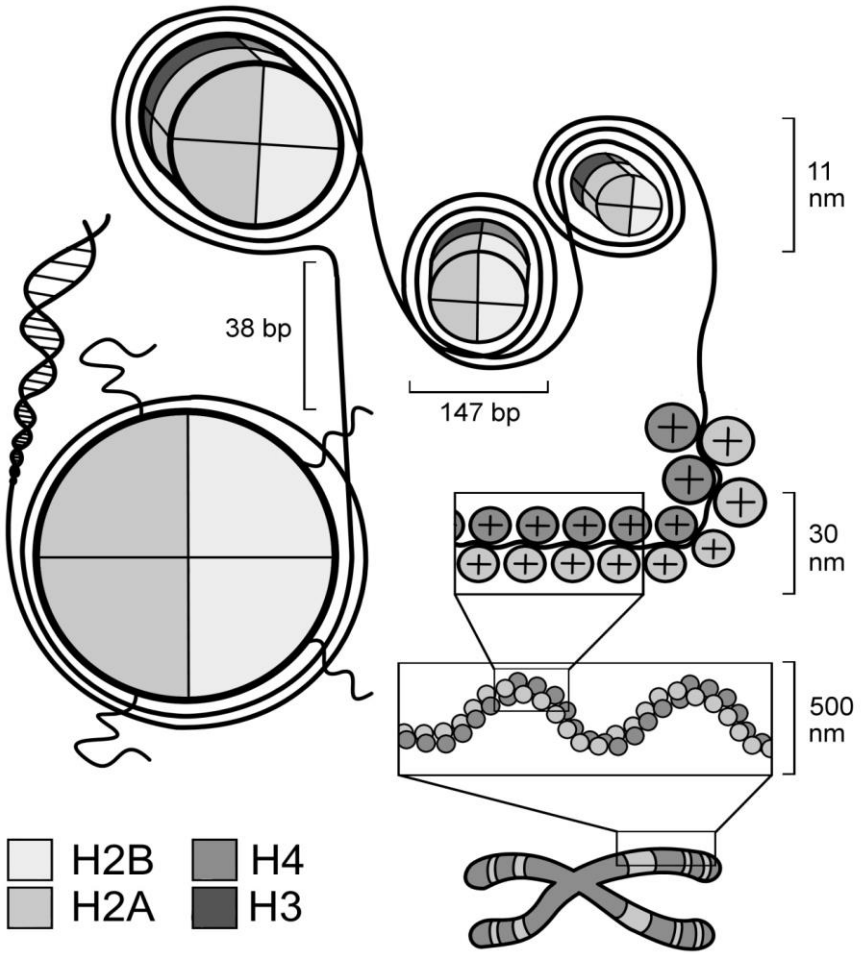


Figure 1.1: Chromatin configuration and DNA packaging. In somatic cells, DNA is compacted by helical winding around octomeric histone complexes to form repeating nucleosomes, where each histone octomer binds 147 bp of DNA. This basal chromatin organization forms the 11 nm "beads on a string" euchromatin fiber. Progressive compaction yields 30 nm solenoidal chromatin fibers or fully condensed 500 nm chromosomes. Each histone possesses an N-terminal tail not involved in chromatin compaction – rather, this histone region can undergo reversible modifications that are central components of epigenetic regulation. Adapted from [14, 15].

The chromatin fiber can subsequently fold into more complex structures, though the mechanisms that drive higher-order chromatin condensation are less well elucidated. However, the differential organization of chromatin into the relatively open "beads on a string" euchromatin structure or the progressively more condensed heterochromatin structures has been strongly implicated in the regulation of transcription, as well as a variety of other biological processes inherent to DNA [16].

Electron microscopy suggests two models for heterochromatic compaction: a solenoid, where intervening linker DNA is bent in a superhelical path with 6 to 8 nucleosomes per turn (Figure 1.1), and, in the absence of sufficient H1 density, a ladder-like configuration of stacked nucleosomes connected by straight linker DNA in a "zig-zag" arrangement. Higher-order folding programs are orchestrated by inter-nucleosomal interactions mediated by reversible covalent modifications of the histone proteins (primarily on histone N-terminal tail regions that actually protrude from the more spherical protein body). These covalent histone modifications are central components of epigenetic regulation, as discussed later.

1.2.2 Histone variants

In humans, many of the histone proteins are encoded by multiple genes, generating an abundant supply of protein for replenishment of the genome. For example, human histone H4 is encoded by 14 genes, though only a single conserved human H4 protein is produced and serves as the docking site for the other histones [17]. The core histone proteins (H2A, H2B, H3, and H4) are primarily assembled with DNA during replication. As DNA strands are forcibly separated, histone cores are displaced from the replication bubble. Likewise, as the replication bubble passes, the newly formed dsDNA is immediately associated with newly synthesized histone cores. During synthesis and non-synthesis phases of the cell cycle, however, non-canonical human histone H2A,

H2B and H3 proteins can supplant the canonical histones and may establish unique patterns of temporal and developmental association; these patterns are an emerging field of epigenetic research [18, 19]. Of the known histone variants, many are associated with sites of DNA damage, and are predicted to play critical roles in damage repair and cell cycle arrest until repairs are complete [20-23]. However, particular variants of histones H2A and H3 have been linked to gene regulation, and will be briefly discussed below.

H3.3, a replication-independent variant of H3, differs from its canonical counterpart by only four amino acids [3]. Incorporation of H3.3 requires the presence of chaperone proteins, either HIRA, which is found in a variety of eukaryotes [24] , or a combination of two other proteins unique to metazoans, death domain associated protein (Daxx) and α-thalassemia X-linked mental retardation protein (ATRX) [25, 26]. These separate modes of H3.3 deposition are not equal across the genome. Rather, HIRA-mediated H3.3 deposition associates with protein-coding regions, while Daxx/ATRX H3.3 deposition is associated with pericentric heterochromatin and intergenic regulatory regions [26, 27]. Evidence in murine ES cells suggests that Daxx/ATRX-mediated H3.3 deposition plays a role in proper chromosome alignment during meiosis in a manner similar to the chromatin-organizing molecules CTCF and cohesin, which will be discussed at length later [28].

Of the core histones, H2A has the largest number of identified variants (19 in humans), differing predominantly in their C-terminal docking domain, which is responsible for entry/exit of DNA from the histone core and for interaction with the H3:H4 tetramer [29-31]. H2A variants also differ in a domain of acidic residues critical for the formation of higher-order chromatin assemblies [29, 32]. As a consequence, H2A variation can have strong impact on chromatin structure/function. Of these variants, H2A.X and H2A.Z were among the first discovered and are perhaps the best

understood [33]. Primarily known for its critical role in DNA damage repair, H2A.X becomes enriched at the site of DNA damage and is required for the proper recruitment of DNA repair machinery, chromatin decondensation at the site of damage, and subsequent repair. However, a more recent study in pre-implantation mouse embryos found that H2A.X rapidly becomes the dominant H2A molecule immediately post-fertilization during the one to four cell stages [34], suggesting that H2A.X may be critical for establishing a vertebrate organism's earliest chromatin configuration. H2A.Z is required for proliferation and differentiation of murine embryos [35]. Though H2A.Z is classically known for flanking nucleosome-free regions and marking transcriptionally active chromatin, its role has become much more nuanced. In the early embryo, H2A.Z is enriched in some regions of constitutive heterochromatin, while adult distribution of H2A.Z is less constrained. H2A.Z can be acetylated, and a number of studies have correlated unacetylated H2A.Z with more stable nucleosome structures and repressive function, while acetylated H2A.Z is more activating and associated with histone variant instability [36-39]. H2A.Z has also been implicated in DNA repair and mitosis [23, 40, 41]. The third most studied variant of H2A, an unusually large structural variant called macroH2A, appears like unacetylated H2A.Z to be associated with transcriptional repression. MacroH2A is best understood as a critical mediator of the process of X inactivation, where it becomes strongly enriched on the mammalian Xi chromosome in a manner that requires ATRX (without Daxx) [19, 42]. Likewise in frogs, macroH2A can act as an epigenetic repressor, preventing X chromosome reprogramming. However, it is also equally well expressed in both males and females in a variety of other vertebrates that do not undergo X inactivation [43, 44].

1.2.3 The Histone Code

Each of the four nucleosomal core histones fold into roughly spherical particles that multimerize to form two co-associating tetramers (2 H2A:H2B dimers + 2 H3:H4 dimers) that are wrapped by DNA. Each of the four core histones also features a rather long and highly flexible N-terminal "tail" that extends from the nucleosomal core (Figure 1.1). The tails of H2B and H3 are closely associated with the DNA that wraps the histone core, running along the minor groove of each strand [45]. On each core histone tail, select amino acid residues can undergo reversible covalent modifications that can alter their charge, configuration and/or DNA binding properties [46]. These modifications do not appear to affect the structure of the nucleosome core particle; however, their functions are numerous. Beyond their role in structural integrity of chromatin fibers [47], they are involved in facilitating interaction of DNA with regulatory proteins and many combinations of these residues correlate directly with important biological processes such as gene regulation [46, 48, 49], DNA repair [22, 50, 51] and replication (both mitotic and meiotic) [46, 52-54]. The pattern of specific histone modifications across a length of chromatin is proposed to encode an epigenetic regulatory map for the contained gene(s), and has been dubbed the "histone code". Though continuing to rapidly evolve, our current understanding of the histone code involves a variety of different modifications (acetylation, methylation, ADP-ribosylation, biotinylation, ubiquitination, phosphorylation and sumoylation) of specific lysine, arginine and serine residues of individual histone tails. For the purposes of this review, I will focus on two modifications in particular – acetylation and methylation of lysine residues along the N-terminal tail of histone H3.

1.2.4 Lysine Acetylation

Modifications of histone N-terminal amino acid tails can modulate chromatin accessibility to DNA-binding proteins. For example, the acetylation of multiple lysine residues on the N-terminal tails of the core histones, and in particular, histones H3 and H4, correlates with increased transcriptional activity of genes associated with such hyperacetylated histones [6, 46, 55, 56]. In general, acetylation neutralizes a lysine's innate positive charge, weakening the histone's interaction with negatively-charged DNA and loosening local chromatin architecture [10]. Beyond this general electrostatic mechanism, acetylation of specific lysine residues is involved in biological processes (including DNA repair/replication and histone deposition) through interactions with proteins that recognize acetylated lysines via bromodomain protein modules [57, 58]. These proteins tend to be associated with transcriptional accessibility, such as activating transcription factors and chromatin-remodeling complexes [57, 59]. Histone lysine acetylation is controlled by two competing enzymatic activities– a variety of histone acetyltransferases (HATs, e.g. Gcn5/PCAF, MOZ, $TAF_{II}250$, CBP/p300, SRC, ATF-2, TFIIIC, and HAT1) [60-62], responsible for acetyl group deposition, and histone deacetylases (HDACs 1 through 11 and SIRT 1-7) [60, 63, 64], responsible for acetyl group removal. Consistent with empirical correlations between histone acetylation and gene activity of the acetylated chromatin, HATs are generally considered activators of transcription. Indeed, most HATs were originally identified as transcriptional coactivators [62, 65]. Likewise, HDACs are generally considered transcriptional repressors [66]. Most HDACs have been linked to the regulation of a surprisingly small subset of genes, and many are relatively restricted in their expression patterns [67, 68]. HDAC1 however has wide ranging effects, and its deletion in targeted mouse mutagenesis experiments is embryonic lethal [69].

With the exception of Lysine 16 on histone H4, histone lysine acetylation events appear to be functionally redundant, such that hyperacetylation of the core histone tails strongly correlates with accessible euchromatin and active transcription [70, 71]. However, recent findings suggest that gene regulation is also influenced by dynamism of the acetylation pattern [72, 73]. In mouse embryo-derived C3H cells, general inhibition of HDAC activity enhanced histone acetylation at *c-fos* and *c-jun*. However, rather than enhancing transcription of the *fos* and *jun* genes, hyperacetylation was instead associated with their transient repression, suggesting that turnover of histone acetylation may be more relevant to gene regulation than steady state histone acetylation levels [73].

At the same time, lysine acetylation is only one of the many covalent modifications that make up the histone code, and the sensitivity of gene transcription to changing acetylation likely reflects the impact of other epigenetic changes in the chromatin. Consistent with this notion, transcriptional response to changes in in histone acetylation, demonstrated experimentally by transient inhibition of HDACs, correlates strongly with the di- and trimethylation of lysine 4 on histone H3 (H3K4me2/H3K4me3) [74, 75]. The challenge in our understanding of histone epigenetics has been in trying to tease apart the contributions of individual marks. Nonetheless, despite a still relatively uninformed view of the overall histone code, histone lysine acetylation remains the clearest understood activating epigenetic mark.

1.2.5 Lysine Methylation

As mentioned above, lysine residues in the N-terminal tails of histones can also be methylated. Like lysine acetylation, histone lysine methylation patterns have been linked to heterochromatin formation, transcription and a variety of other important cellular processes [76].

However, the addition of one or more methyl groups to lysine residues differs significantly from lysine acetylation; rather than near universal association with gene activation, the effects of lysine methylation are strongly linked to the specific residue being modified [66]. In addition, replacement of a variable number of hydrogens on lysine's side-chain amine with methyl groups permits three different states of histone lysine methylation, referred to as mono-, di- and tri-methylation [77]. Different numbers of methyl marks at a specific residue are associated with widely divergent phenomena – for example, monomethylation of lysine 4 on histone 3 (H3K4me1) has been found to characterize enhancer elements, trimethylation of the same residue (H3K4me3) is associated with active transcription at promoters and dimethylation (H3K4me2) correlates with transcriptional accessibility at both promoter and enhancers [78]. In reference to gene activation, modifications such as the aforementioned H3K4me3, H3K79me2 and H3K36me3 are associated with transcription and a localized euchromatic state while modifications correlated with repression and heterochromatin include H3K9me3, H3K27me3 and H4K20me3 [66, 79].

Lysine methylation marks are introduced by a family of histone methyltransferases (HMTs) [80, 81] primarily through catalytic SET domains that use S-adenosyl-L-methionine as a methyl donor (the exception being H3K79me2, which is catalyzed by HMT DOT1L in a method that resembles DNA methyltransferase activity [82]). Whereas activity of individual HAT complexes may target a variety of N-terminal lysines on more than one histone at a given gene, HMT activity appears to be much more restricted to individual lysine residues. For example, there are currently eight known human HMTs specific for H3K9 (e.g. GLP and G9a [81], and a similar number of HMTs specific for H3K4 [81]. Removal of these methyl groups is catalyzed by histone demethylases (HDMs, [81, 83, 84]). Lysine demethylation in particular is initiated by lysine demethylases (KDMs), of which 8 families (KDM1-8)

have currently been identified in humans [85]. Each of these KDMs utilize a mechanism of oxidative demethylation via radical processes, in which the KDM induces oxidization and subsequent excision of the methyl group, releasing formaldehyde as a co-product [85].

1.2.6 DNA methylation

In addition to histone modifications, the DNA of eukaryotes can also be modified by methylation. DNA methylation in eukaryotes is largely restricted to 5-cytosine-guanine-3' dinucleotides, with the 5-carbon of cytosine being targeted for methyl group addition. So-called CpG methylation is common in vertebrate tissues, with up to 85% of all CpG dinucleotides methylated [86, 87]. However, CpGs are not evenly distributed across the vertebrate genome. Rather, most CpGs are found in the promoters and 5' or 3' coding regions of genes, and often in large clusters dubbed "CpG islands". CpG islands have been canonized in the literature as DNA regions of ≥200bp with an observed-to-expected CpG ratio >60% [88]. CpG islands are distinct through possession of a unique epigenetic state; whereas most CpG sites in the genome are methylated, >90% of CpGs within CpG islands are unmethylated [89, 90].

DNA methylation has long been linked to gene repression by altering transcription factor binding and/or chromatin structure. In keeping with a presumed repressive role for DNA methylation, the promoters of constitutively expressed "housekeeping" genes were among the first genes to be associated with unmethylated CpG islands [91, 92]. Shortly after fertilization, gametic DNA methylation patterns are almost entirely erased as totipotency is established, and then a new pattern of methylation is established coincident with embryo implantation [89, 93, 94]. Though as yet poorly supported by experimentation, recent studies have also suggested that specific environmentally-influenced changes in DNA methylation experienced by the embryo's parents may

survive preimplantation erasure to provide a degree of transgenerational epigenetic memory [95, 96]. Such an evolutionary shortcut is provocative, but remains to be confirmed. More accepted is the notion that once established, DNA methylation patterns are largely stable (~90%), and are maintained through every subsequent cell division and to a large extent through tissue differentiation [95, 97].

Changes in the pattern of DNA methylation have been linked to aging and tumor development, in which multiple studies have documented the demethylation of non-CpG island sites and the *de novo* methylation of critical CpG islands [98, 99]. Physiological and disease-related changes in methylation are both mediated by the same molecules. CpG methylation is driven by the Dnmt3 *de novo* methylases, the maintenance methylase Dnmt1 and by specific histone methylases, while demethylation appears to be driven at least in part by Tet-mediated conversion of 5-methylcytosine to an intermediate 5-hydroxymethylcytosine [85, 89, 100, 101]. More recently, studies of embryonic and transformed human and mouse genomes have shown that changes in DNA methylation have a significant impact on the binding pattern of the CCCTC-binding factor (CTCF), a critical sequence-specific regulatory protein that is essential for embryonic development and has been linked to the regulation of a wide array of genes [102-105].

It remains unclear how DNA methylation patterning is initially directed. *Ptcra* is a thymocyte-specific gene, responsible for production of a pre-T cell receptor alpha chain to allow cell-surface TCRβ expression before TCRα recombination. The enhancer of *Ptcra* maintains a window of unmethylated CpGs in ES cells, which appears to persist throughout lineage development regardless of eventual cell fate [106]. In a series of elegant experiments, the Smale laboratory found that stable transfection of insulated and premethylated plasmid constructs carrying the *Ptcra*

enhancer and a Green Fluorescent Protein (GFP) reporter into ES cells resulted in rapid enhancer demethylation, without leading to expression of the accompanying GFP reporter [106]. In contrast, when the same experiment was performed in *Ptcra*-expressing thymocyte cell lines, these differentiated cells proved unable to demethylate and activate the enhancer, though cells transfected with an unmethylated construct readily expressed GFP [106].

The Smale studies clearly suggest that the capacity to remodel DNA methylation patterns is both intrinsic to early embryonic cells and is critical for proper regulation of differentiation-specific genes. What then directs the patterning of DNA methylation in pluripotent embryonic cells? The lack of CpG establishment at tissue-specific genes results at least in part from methylation interference by pre-bound transcription factors. Evidence suggests that this transcription factor binding may be requisite for proper gene transcription upon lineage differentiation [106]. As noted above, the genome-wide pattern of CTCF binding in human and mouse cells is tightly linked to patterns of DNA methylation. Both *in vivo* and *in vitro* studies have suggested that DNA methylation may hinder CTCF binding [102, 104]. Not surprisingly, the vast majority of CTCF binding sites in the mouse ESC genome are free of DNA methylation (Bradshaw, Bushel and Sikes, unpublished observations). However, in as much as DNA methylation may prevent CTCF binding, it remains unknown whether CTCF binding in early embryonic cells may similarly prevent DNA methylation. Indeed, CTCF is abundantly expressed and essential in even the very earliest undifferentiated embryos. Specifically, *Ctcf* nullizygous embryos only developed to the peri-implantation blastocyst stage (E3.5), a point coincident with DNA remethylation [95, 107]. The functional relationship between DNA methylation and CTCF is a central theme of this thesis, and will be discussed in detail in Chapter 2. Beyond CTCF, the wide array of data now suggests that DNA methylation patterns at

tissue specific promoters are established in early embryonic development (at least in part due to the occlusion of DNA methylases by bound regulatory factors), that the DNA methylation pattern is largely stable during development and differentiation, and that it directly shapes gene activation capacity in response to appropriate developmental cues.

1.3 Overview of Stem cell development

1.3.1 The early embryo

After fertilization and the subsequent fusion of sperm and oocyte chromosomes, mammalian preimplantation development starts with a single-cell zygote [108], followed by a series of three cleavage divisions resulting in an eight-celled undifferentiated morula (Figure 1.2). Cells within this morula flatten and polarize, then undergo an additional pair of cell divisions, during which the first distinct populations of embryonic cells are generated; cells that maintain contact with the external environment contribute to the trophectoderm lineage, which forms embryonic support structures such as the placenta, and the inner cells develop into a collective fittingly called the inner cell mass (ICM) [109]. A fluid-filled segmentation cavity formed at the 32-cell stage pushes the ICM to one side of the embryo, forming the blastocyst (E3.5), and cells in direct contact with this cavity form a polarized epithelium and develop into the primitive ectoderm lineage (or hypoblast), which gives rise to the yolk sac that nourishes the developing embryo [108]. However, in contrast to the early morula, lineage differentiation in the ICM occurs independent of location, and segregation consistent with a mature blastocyst is attained through a combination of differential adhesion of cell progenitors and response to directional signals, the details of which remain to be elucidated [110]. Upon completion of this segregation by day E4.5, the mature blastocyst is ready for implantation.

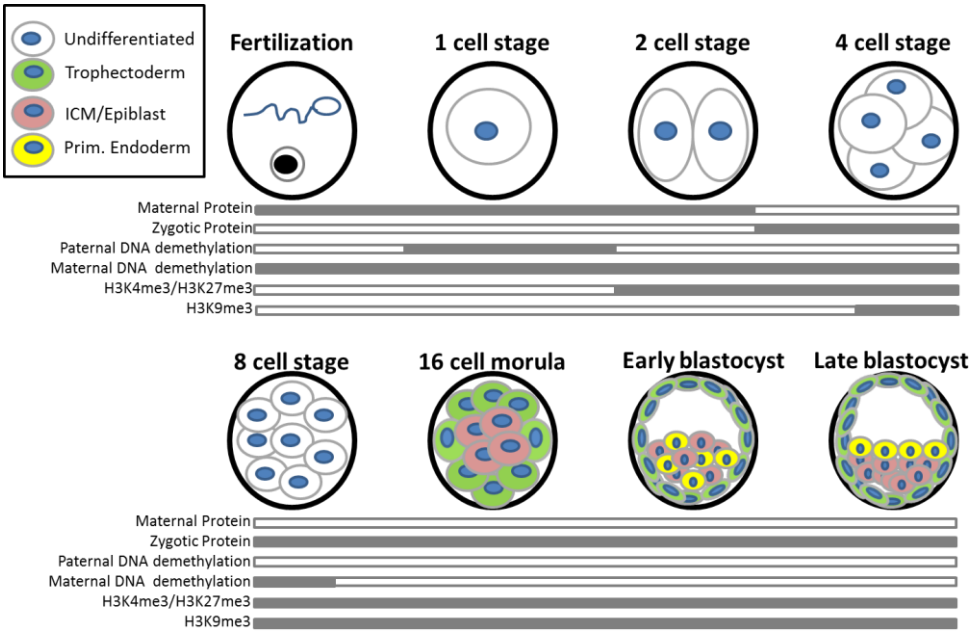


Figure 1.2: Cell fate and epigenetic regulation during early post-fertilization development. The sperm genome is compacted with a mixture of histones and histone-like protamines. After fertilization, the paternal genome is repackaged to remove protamines and allow association with maternal DNA in the zygote. This repackaging is accompanied by a period of active paternal demethylation, followed by passive maternal demethylation over the first few cell divisions, along with establishment of histone modifications involved in gene priming. The morula is the earliest indication of cell fate establishment, in which cells along the outside of the developing embryo become predisposed to the trophectodermal lineage while the ICM does not. Progression to the pre-implantation blastocyst sees further cell fate specification within the ICM, with primitive ectoderm and epiblast cells initially forming in a "salt and pepper" fashion – however, by the late blastocyst these cells have segregated, with cells along the segmentation cavity in the primitive ectoderm and away from this cavity in the epiblast. Adapted from [111].

1.3.2 Progression of the epigenetic map

The development of a new biological organism requires sweeping epigenetic changes to facilitate the combination of two haploid nuclei into a single functional diploid, and subsequent lineage specification comes with specific collections of epigenetic hallmarks. During gametogenesis, sperm lose a majority of their histones and repackage their DNA into a more compact format using proteins called protamines [111, 112]. Maternal gametes maintain histone packaging, however, and within an hour of fertilization maternally generated histones replace protamines on paternal chromatin as well, first with synthesis-independent H3.3 and finally with canonical replication variant H3.1 [113]. This conversion is concurrent with an active demethylation of paternal CpGs and deposition of histone acetylation and methylation marks [111, 114] (Figure 1.2). Maternal DNA, though protected by an unknown mechanism from this active demethylation of the paternal DNA, nonetheless undergoes passive demethylation during successive divisions through exclusion of DNA methylases from the nucleus during replication [115]. Consequently, both maternal and paternal genomes are similarly demethylated by the 8-cell morula stage [116-118]. Since CpG methylation levels at DNA other than the pluripotency genes and imprinting control regions are largely similar in sperm nuclei and ESCs, the purpose of genome-wide active demethylation in sperm, as well as the mechanism and rationale for protection of the maternal genome from this active demethylation remains unknown [111, 116].

Post-translational histone modifications also differ significantly between paternal and maternal genomes prior to the morula stage, primarily due to the ground-up reconstruction of histone modification patterns necessary for the newly histone-rich paternal genome. Though repressive H3K27me3 is established and maintained on both genomes prior to the first cleavage

division, the repressive H3K9me3 histone mark is not established on paternal sequences until after the four-cell stage (Figure 1.2). Regions of the paternal genome destined for H3K9me3 deposition are instead protected by the Polycomb Group PRC1 complex until the 8-cell morula forms, by which point this asymmetry has been rectified [119].

Epigenetic factors play a significant role in progression toward trophectodermal, ectodermal and endodermal cell fates as well. Genes required during embryonic development are often primed for either activation or repression in the earliest stages of embryonic development, to be discussed at length later in the chapter. However, in cells destined for the trophectoderm, these primed genes (denoted by the bivalent or opposing functional marks of H3K4me3 and H3K27me3) are repressed through addition of an additional histone mark, the aforementioned H3K9me3, as well as by DNA hypermethylation [108, 120]. In the ICM, H3K9me3 plays a complementary role in repression of trophectodermal lineage genes, and depletion of the methyltransferase that establishes this mark results in a trophectoderm-like phenotype [121].

1.3.3 Stage-specific transcription factors

Three of the most well-studied transcription factors related to murine embryonic progression are OCT4, NANOG and SOX2, which are necessary in conjunction with Leukemia Inhibitory Factor (LIF) for maintenance of pluripotency in cultured embryonic stem cells of murine origin [109]. OCT4 expression is noticeable as early as the two-cell stage, and is specifically excluded from the trophectodermal lineage through repressive action of the protein CDX2 [109, 122]. CDX2 expression is required for trophectodermal development just as OCT4 is necessary for the ICM [117]. OCT4 promotes expression of fellow transcription factor NANOG by the eight-cell morula stage; NANOG is also required for proper ICM development, but absence of NANOG protein results in

progression down the primitive endoderm pathway [109, 123]. SOX2 can be detected throughout preimplantation development and is required for formation of the epiblast in the ICM. Like OCT4, SOX2 deficiency causes trophectoderm-like morphology and complete absence of SOX2 results in embryonic death soon after implantation [109, 124].

Another transcription factor vital to proper embryonic development is CTCF. CTCF forms homodimers while bound to DNA, leading to chromatin looping between CTCF sites that can facilitate or prevent transcriptional activation [125, 126]. CTCF is highly conserved in all multicellular organisms, plays many vital roles in somatic cells and has well-characterized involvement in gametocyte implementation of imprinting, a process of parent-of-origin specific gene expression that is vital for proper embryonic development [126, 127]. During spermatogenesis, CTCF is specifically downregulated and is absent from mature sperm, instead replaced by a protein with an identical binding site called CTCF-like (CTCFL) or BORIS. However, maternal CTCF is a requisite for proper development, and its absence results in meiotic defects in oocyte maturation and ultimately leads to certain apoptotic death of the developing embryo by the blastocyst stage, as noted earlier [128]. Zygotic CTCF does not begin transcription until the two-cell stage of embryonic development, at which point the more profound effects of CTCF depletion are made manifest. Prior to this, maternal CTCF is sufficient, though the exact role of maternal and zygotic CTCF in embryonic development remains elusive.

CTCF also has a pronounced effect on nucleosome positioning. CTCF binding displaces histones such that CTCF is found at the center of linker DNA, and the CTCF binding site is flanked by at least 6 well-positioned nucleosomes on each side [12]. This is a larger and better-defined window of positioning than seen with any other transcription factor, even larger than transcription start site-

related nucleosome deposition. DNA footprinting evidence suggests that the zinc-fingerless CTCF portion protects DNA from digestion in a characteristic way and CTCF binding increases linker length between neighboring nucleosomes to 118 bp. Nucleosomes flanking CTCF sites share much in common with those flanking promoters, and are often enriched in marks associated with open chromatin and active transcription. These marks, when not associated with full-length transcription itself, can lead to rapid histone turnover, which may help prevent the lateral spread of chromatin states when the DNA remains accessible to modification factors [129]. Indeed, one of the earliest identified functions for CTCF was its ability to insulate β globin genes against encroachment of neighboring repressive heterochromatin [130].

CTCF has long been appreciated for its insulating functions at the *H19* ICR, where its loss leads to increased DNA methylation and biallelic activation of the insulin-like growth factor 2 (*Igf2*) gene [102, 131]. Likewise, CTCF and a number of other factors have been shown to play critical insulating roles at a number of genes in *Drosophila* including the *Hox* gene *Bithorax* [132, 133]. Conversely, a number of studies have shown a role for CTCF in coordinating gene activation in ESCs with Cohesin [134-136] and also with TAF3 [137], and in activating expression of many genes during neuronal and hematopoietic development [138-142]. In mouse, CTCF also appears to play an insulating role in X inactivation [143] and in enhancer blocking during V(D)J recombination of the *Igh* and *Igk* loci [138, 144, 145]. To this point, CTCF remains the only defined mammalian insulator, despite the presence of multiple insulator proteins in *Drosophila.*

1.4 Priming

Multicellular life requires that gene usage be restricted based on tissue type. Consequently, genes can exist in an activated state, if that gene is involved in that specific cell's proper

functionality, or in a silenced state for lineage-inappropriate genes. These alternate states are largely mediated by alterations in a gene's chromatin structure, with open chromatin facilitating transcription of activated genes and condensed chromatin rendering transcription impossible by interfering with binding of transcription factors or RNA Polymerase II through steric hindrance. A third gene state exists, in which genes are not actively being transcribed but can quickly be converted into an activated form. This intermediate or primed state is frequently associated with genes involved in developmental or metabolic processes. Priming of tissue-restricted genes is first imposed in pluripotent stem cells, and then must be maintained during renewal and progressive differentiation until the proper developmental cues trigger gene activation [146]. The establishment and maintenance of priming mechanisms during lineage development is an area of intense interest. Current information on epigenetic marks associated with the phenomenon allows us to sort priming into two categories. One is associated with monovalent histone modifications and narrow stretches of CpG islands that lack DNA methylation. The other is associated with two particular histone N-terminal tail modifications: trimethylation of the 27^{th} lysine of histone 3 (H3K27me3), a mark associated with repression, and the trimethylation of the 4^{th} lysine of histone 3 (H3K4me3), associated with activation [78, 147, 148]. Because of the opposing activities associated with H3K4me3 and H3K27me3, this priming variety is referred to as bivalency. In embryonic stem cells, genes associated with development tend to exhibit bivalent priming [147], while tissue-specific genes lack bivalent priming but many contain histone modifications typically associated with active chromatin as well as windows of unmethylated CpGs [106, 149].

1.4.1 Bivalent Priming

Establishment of H3K27me3, the repressive mark involved in bivalent priming, is contingent upon collections of Polycomb group proteins called polycomb repressive complexes (PRCs) [150]. Two different complexes are involved: PRC2 contains a functional histone methyltransferase (HMT) that establishes H2K27me3, while PRC1 specifically binds to this modification and is capable of inhibiting both transcription and nucleosome remodeling [148, 151-153]. H3K27me3 and PRC2 binding alike are found repressing developmental regulators in ES cells and beyond, until activation through lineage differentiation catalyzes H3K27me3 loss [154]. Murine PRC2 is regulated particularly in ES cells by recently identified PRC2 subunits JARID2, MTF2 and esPRC2p48, which contribute to ES repression of lineage-specific genes. These subunits, unlike other components of the polycomb complexes, are significantly elevated in ES cells [155]. They also are synergistically involved in the potential for pluripotent cell induction (iPS) via Oct4/Sox2/KLF4 transduction, as removal of any of these subunits using RNA interference (RNAi) or through gene knockouts has a profound negative effect on iPS reprogramming efficiency [155]. This repression is mediated through direct effects on establishment and maintenance of H3K27me3, including PRC2 recruitment and protection from histone demethylases. Knockout of PRC2 components in mice results in embryonic lethality *in vivo*. *In vitro*, embryonic stem cells from these knockouts show premature differentiation, widespread de-repression of lineage-specific genes and/or an inability to properly differentiate [155-159]. This indicates that bivalent priming plays an important role in the ES repression of lineage-specific genes.

RNA Polymerase II (Pol II) is still recruited to bivalent chromatin. However it is incapable of productive transcriptional elongation due to PRC1-mediated ubiquitination of histone protein H2A

[160]. Instead, Pol II initiation events at bivalent chromatin result in production of long ncRNAs (up to 8 kb) that can help stabilize PRC2-DNA interaction but avoid establishment of marks associated with transcriptional elongation (H3K79me2) and mRNA splicing (H3K36me3) [161-163]. Genes that are the targets of Polycomb proteins also preferentially avoid the divergent short-range antisense transcription characteristic of active promoters while bivalently marked [162]. Beyond Pol II itself, other elements of transcription machinery may be associated with PRC binding, in particular the histone deacetylase HDAC1 and transcriptional coactivator p300 [160].

Clues to the nature of transcription by Pol II are provided by modifications of specific residues of a long consensus repeat within the polymerase Rpb1 subunit, located at the carboxy-terminal domain and dubbed the "Pol II CTD"[164]. Two of these modifications, phosphorylation of serine 2 (Ser2P) and serine 5 (Ser5P) of this repeat, are relatively well-characterized, and their presence or absence correlate with the transcriptional state of the polymerase [165]. Initial recruitment of the polymerase to a promoter region requires it to be free of CTD phosphorylation [165]. Establishment of Ser5P is associated with transcriptional initiation, which in yeast also correlates with recruitment of the histone methyltransferase (HMT) Set1 to induce H3K4me3 [166, 167]. Following initiation, Ser2P is associated with productive elongation. Establishment of Ser2P by P-TEFb also facilitates polymerase escape from negative regulation by factors NELF and DSIF found at a subset of promoters [167-170]. Both Ser2P and Ser5P have been linked to lineage development through an association with deposition of transcriptional initiation and/or elongation-related histone modifications [171-173]. Bivalent genes are bound by Pol II that possesses a unique CTD architecture, with high levels of Ser5P and an unknown modification or conformational change that blocks access to Ser2 by either anti-Ser2P or anti-Ser2 antibodies [160]. Ser-5 phosphorylated Pol II

is strongly associated with polycomb target genes in ES cells, indicating a role for transcriptional stalling in lineage development [160]. Indeed, stalled transcription often results in so-called "short" RNAs, 50-200 nucleotide ncRNAs which allow for establishment of the activating mark H3K4me3 [166, 174-176]. These short RNAs, as well as longer ncRNAs such as those involved in X chromosome inactivation (e.g. RepA) and the HOXD locus (e.g. HOTAIR) among others, could serve to stabilize interaction between PRC2 and polycomb target genes [174, 177-179]. Expression of short RNAs and mRNA for particular regions are inversely proportional - removal of PRC2-mediated repression during ES cell differentiation leads to a reduction in transcription of localized short RNAs and an increase in the full-length mRNA necessary for gene expression [174].

1.4.2 Alternative Priming Mechanisms

Efforts to identify and characterize enhancers involved in human embryonic development have revealed an additional layer of regulation for proper cell-fate transition. As defined by a number of factors, including evolutionary conservation, DNase I hypersensitivity and binding of enhancer-associated transcriptional coactivators like p300, active enhancers can also be predicted using two specific histone marks, the monomethylation of H3K4 (H3K4me1) and the acetylation of H3K27 (H3K27ac) [180-183]. Using a series of ChIP-seq experiments with antibodies against p300, chromatin regulator BRG1 and various histone modifications the Wysocka laboratory identified two particular chromatin patterns associated with enhancers in human ESCs; one contains localized enrichment of H3K4me1 and H3K27ac and one exchanges acetylation of H3K27 for the mutually exclusive trimethylation of the same residue (H3K27me3), a mark involved in bivalent as discussed earlier in this review [180, 184]. These poised enhancers are associated with cell-type dependent genes and, upon differentiation to that particular cell type, H3K27me3 is replaced by H3K27ac and

the enhancer becomes active. This activation is accompanied by recruitment of Pol II to the enhancer, leading to bidirectional transcription of enhancer RNAs that augment mRNA synthesis at related genes [184, 185]. The nature of the enhancer also reflects the nature of the gene – primed enhancers are associated with primed genes and active enhancers with active genes.

Early studies in yeast found that silent, poised, and active genes could be discriminated by their differential enrichment of H3K4 di- and trimethylation, where poised genes possessed H3K4me2 in the absence of H3K4me3 (H3K4me2+/me3-) while silent and constitutively active genes respectively lacked (H3K4me2-/me3-) or contained (me2+/me3+) both marks [186]. Unlike constitutively active genes, the H3K4 methylation patterns of methionine- or inositol-regulated genes were entirely dependent on culture conditions. In the presence of the appropriate growth supplement, these metabolic genes were active and marked by both di- and trimethylation of H3K4. Absent methionine or inositol, they lost H3K4 trimethylation, but maintained dimethylation [175], suggesting that H3K4me2 reflects an intermediate chromatin state that may allow for future or more rapid conversion to an active conformation.

More recently, differential H3K4 methylation was found in hematopoietic progenitors to specifically mark developmental genes poised for expression [187]. Again, H3K4me2+/me3- genes were transcriptionally inactive, but differential K4 methylation was specifically enriched over the genes' promoter regions, particularly proximal to binding sites for hematopoietic regulatory transcription factors like RUNX1, PU.1, and in correlation with PU.1 occupancy. Importantly, self-renewing hematopoietic stem cells were found to lack this pool of poised genes, suggesting that priming occurs after progenitors lose their totipotency. Finally, lineage commitment correlated with

loss of H3K4 methylation and restriction of the poised gene pool, suggesting that differential H3K4 methylation may be a sensitive marker for mapping lineage decisions during hematopoiesis.

The study of priming is still in its infancy. Bivalently primed genes have been largely identified through genome-wide chromatin immunoprecipitation coupled with massively parallel sequencing (ChIP-seq). However, not all developmental genes are bivalently primed, nor are all bivalently primed genes involved in development. At present, priming is largely defined as a catalog of chromatin phenotypic markings that distinguish a particular stretch of transcriptionally silent DNA from neighboring heterochromatic stretches, and that are suggestive of gene activity. Not surprisingly, different combinations of the modifications discussed above (ie, H3K4me3/H3K27me3 bivalency, RNA Pol II pausing, H3K4me1 enhancer marking, and CpG hypomethylation of promoters) have been identified for various genes. Few unifying themes have emerged as of yet. More significantly, beyond the demonstration that ES-dependent demethylation of the *Ptcra* enhancer is required for its transcriptional activity in thymocytes [106], few studies have been able to show a functional role for priming. The primary focus of this thesis research was to define the epigenetic priming of *Tcra*, and to lay a foundation for functional studies aimed at determining the physiological impact of this priming on *Tcra* activation and thymocyte development.

1.5. Antigen Receptor Genes

1.5.1. Adaptive Immunity

Higher-order vertebrate immunity derives from the actions of separate but complementary innate and adaptive activities. The more primordial innate immune system provides nonspecific host barriers to infection such as the physical and chemical barrier activities associated with epithelial tissue and cellular defenses that respond to a limited collection of chemical targets shared

by different classes of pathogens. In contrast, the adaptive immune system, which is primarily composed of B and T lymphocytes, is capable of very specific, surgical responses to virtually any antigen. Whereas innate responses are extremely rapid but genetically fixed, adaptive responses are slow to activate but rapidly gain in scope and efficacy and produce long-lasting immunological memory.

B cells, so named for the bursa of Fabricius in birds where they were originally discovered, are the primary source of adaptive immune response to extracellular pathogens [188]. Each naive B cell is covered with identical copies of its B cell receptor or BCR, uniquely generated during development in the appropriate lymphoid organ (in the case of most mammals, the bone marrow), and is exposed to environmental stimulus through release and circulation through the blood and lymph. Upon binding of sufficient amounts of its specific antigenic epitope, the B cell ingests this antigen-receptor complex via receptor-mediated endocytosis, degrades it and begins expressing smaller subunits of the antigen in association with Major Histocompatibility Complex (MHC) II, targeting it for recognition by a variety of T cell known as the helper T cell (T_H). If this B cell is subsequently engaged via its MHCII:antigen by antigen-specific T_H cells, it will be stimulated to undergo clonal expansion. Many of these clones will differentiate into plasma cells, a terminal cell form that switches from production of a surface-bound version of the BCR to a secreted form known as antibody. Plasma cells produce and secrete massive quantities of this antibody, which travels throughout the organism and binds to the same antigen recognized by the originally activated B cell, interfering with the antigen's function and targeting it for degradation by other immune components such as macrophages. A small number of these clones instead become memory B cells,

exceedingly long-lived and easily activated cells that provide a faster response upon repeat exposure to the same antigen.

Like B cells, each T cell expresses a heterodimeric antigen receptor on its cell surface. The antigen receptor on most T lymphocytes is encoded by the T cell receptor (TCR) alpha (*Tcra*) and beta (*Tcrb*) genes. The αβ heterodimer assembles on the cell surface with a number of accessory proteins that facilitate both antigen engagement and subsequent intracellular signaling. In a subset of T cells, these same accessory proteins assemble with heterodimers of TCRγ and TCRδ.

All antigen receptor proteins contain folded sequential repeating β motifs that are linked by disulfide bridges. Each α and β TCR chain contains two of these so-called immunoglobulin motifs, a hypervariable N-terminal antigen binding domain and a conserved C-terminal signaling domain. The antigen specificity of each TCR derives from the paired α and β antigen binding domains. Receptor variability is confined to the three hypervariable α helix loops that separate the antiparallel β turns of the antigen binding Ig motifs. Since these α helices establish antigen complementarity, they are called the complementarity determining regions (CDRs 1-3) of the receptor. Together the TCRα and TCRβ CDRs control TCR association with the MHC:antigen complex on the antigen presenting cell (APC). Unlike B cell receptors, which supply their immune functions through mass secretion of antibodies after activation, T cell receptors function exclusively on the T cell surface. Also distinct from B cell receptors, which can recognize a wide variety of antigen structures and compositions, αβ TCRs recognize peptide antigens bound and presented by one of two classes of MHC: those that acquire antigenic peptides from cytosolic (MHC I) or endocytic (MHC II) proteolysis.

When MHC I, expressed by all nucleated cells, presents antigen to its cognate TCR, the MHC:antigen:TCR complex is stabilized by CD8 co-receptors expressed on the T cell surface. T cells

that express CD8 are called CD8+ single positive (SP) T cells. Upon activation, these cells mature into cytotoxic T lymphocytes (CTLs) that seek out and destroy virally-infected cells that present appropriate target antigens on their MHC I. By contrast, only a limited number of cells constitutively express MHC II, which presents extracellular or intravesicular antigens acquired during endocytosis or phagocytosis. These so-called professional antigen presenting cells (APCs) are principally dendritic cells, macrophages and B cells, and their MHC II:antigen interactions with the $\alpha\beta$ TCR are stabilized by CD4 co-receptors expressed principally on naïve helper T cells and regulatory T cells. Upon activation, these CD4 SP cells function to either help drive macrophage or B cell-mediated immunity against extracellular or intravesicular pathogens (hence the term helper T cell) or to limit the scope and duration of cellular immune responses (so-called CD4 regulatory T cells or T_{Reg}s). Together, the CD4 and CD8 SP T cells provide a master regulatory function that impacts all aspect of the adaptive immune response such that genetic defects that exclusively impair the development or function of T cells are sufficient to result in a severe combined immunodeficiency that is fatal without extravagant medical intervention.

1.5.2 Hematopoiesis and T cell development

In the adult bone marrow, a collection of self-renewing, multipotent cells give rise to all of the hematopoietic cell lineages. These hematopoietic stem cells (HSCs) give rise to new lineages through a series of developmental stages in which pluripotency is progressively restricted until a cell ultimately commits to a terminally differentiated lineage [189]. T and B lymphocytes undergo a common developmental pathway to a certain extent: HSCs can differentiate to non-self-renewing multipotent progenitors (MPPs). These MPPs can further differentiate to lymphoid MPPs (LMPPs), which retain the capacity to become lymphoid cells or macrophages, and then lose their

macrophage potential through further differentiation into all lymphoid progenitors (ALPs) [189, 190]. ALPs have the potential to become B-cell biased lymphoid progenitors (BLPs), which serve as the immediate precursors for B cell development. It is thought that early T lineage precursors (ETPs) found in the thymus also derive from ALPs in the bone marrow, though development between the LMPP and ETP stage has yet to be fully elucidated [191-194]. Regardless, ETPs are the only intrathymic precursors shown to significantly contribute to T cell development [191].

ALPs destined to become B cells undergo antigen receptor gene recombination, positive and negative selection in the bone marrow, their tissue of origin. In order to become T cells, however, ALPs must migrate from the bone marrow to the thymus in order to undergo selection and recombination. Once they arrive in the thymus, T cell development occurs in a series of distinct developmental stages that are tied to particular thymic regions, cumulating in an eventual escape to the bloodstream from the thymic medulla if recombination is successful. Extensive research has also found numerous cell surface markers differentially expressed within thymic progenitors that we now use for identification of these developmental stages; these markers include the T cell co-receptors CD4 and CD8, CD3 complex proteins, the adhesion molecule CD44, the alpha chain for the IL-2 receptor (also known as CD25) and c-kit, the receptor for stem cell factor [195, 196]. The most commonly used nomenclature for thymopoiesis refers to the presence or absence of CD4 and CD8; cells expressing both co-receptors are referred to as double positive (DP) [181], cells expressing only one are single positive (SP) and cells expressing neither receptor are called double negative (DN) (triple negative if they also lack expression of CD3, though this discussion will follow the more widely used nomenclature that ignores CD3 expression). DN cells can be further subdivided into four

substages based on the expression of CD44 and CD25 (DN1 or ETP are $CD44^{hi}CD25^{-}$, DN2 are $CD44^{+}CD25^{+}$, DN3 are $CD44^{-}CD25^{+}$, DN4 are $CD44^{-}CD25^{-}$).

In general, mainstream α:β T cell development progresses as such: ETPs enter the thymus from the bloodstream as DN1 cells, and RAG proteins are first detected in late DN1 cells. Transition to DN2 involves the advent of detectable V(D)J recombination of the *Tcrb, Tcrd,* and *Tcrg* genes. Further transition to DN3 is accompanied by completion of β or both γ and δ recombination. If *Tcrd* and *Tcrg* are successfully rearranged prior to successful rearrangement of *Tcrb,* recombination is terminated and cells leave the thymus as mature γδ T cells. If however *Tcrb* is successfully rearranged and expressed on the cell surface as part of the pre-TCR (along with the preTα protein), the cell downregulates CD25 and enters the DN4 stage, where it ceases expression of RAGs, enters into a period of intense proliferation and upregulates both CD4 and CD8 to enter the DP stage. At this point the cell again halts proliferation, re-expresses RAG proteins and re-enters a period of recombination, this time permitting rearrangement of the α locus as well as γ and δ. Because *Tcrd* sits within *Tcra,* initial *Tcra* rearrangements delete the δ locus. As such, the vast majority of DP cells progress to become αβ rather than γδ cells. This distribution is entirely dependent then on the rapid and efficient activation of *Tcra* recombination upon DP transition. The primary focus of this thesis research is aimed at defining the epigenetic priming that facilitates such rapid activation of *Tcra* recombination. Successful α recombination leads to expression of a functional αβ T cell receptor (TCR) and termination of recombination. These αβ pre-T cells undergo positive selection for engagement with the host's MHC molecules concurrent with downregulation of either CD4 (in the case of engagement with MHC I) or CD8 (in the case of engagement with MHC II). Positive selection

is followed by tolerizing negative selection, and cells exit the thymus as mature CD4 SP or CD8 SP T cells.

1.5.3 V(D)J recombination

The power of the adaptive immune system derives from the enormous breadth of B and T lymphocytes' antigen binding repertoire, generated by V(D)J recombination during lymphocyte development. Recombination entails a carefully controlled pattern of DNA cleavages and ligations to create novel exonic segments that, when translated, encode a receptor's antigen binding domain (Figure 1.4). In each two-step rearrangement, a lymphocyte-specific recombinase composed of the Recombination Activating Gene (RAG) 1 and 2 proteins introduce dsDNA breaks (DSBs) at conserved recombinase signal sequences (RSSs) that flank each V (variable), D (diversity) and J (joining) segment. Though rare, RAG targeting of cryptic RSS sequences outside of the antigen receptor genes can result in leukemic translocations. Indeed, the majority of lymphomic translocations are mediated by RAG-1/2 [197, 198].

Importantly, all rearrangements in each of the various antigen receptor genes use a common mechanism initiated by the RAG proteins binding to compatible RSSs and inducing a single-stranded nick in the DNA precisely at the boundary between the RSS and its associated coding segment. RAG-mediated nicking occurs as part of a large synaptic complex formed between the RAG proteins and the two gene segments to be joined, as well as proteins involved in bending the chromatin strand to allow synapsis (Figure 1.4). Once nicking has occurred at both compatible RSSs, the free hydroxyl groups liberated by the single-stranded nicks attack the phosphate backbones of their antiparallel DNA strands in a transesterification reaction that generates hairpinned coding ends and blunt RSS or signal ends [199, 200]. The hairpinned coding ends are reopened through a process

that allows nucleotide addition and/or deletion, and then ubiquitous proteins of the non-homologous DNA end-joining (NHEJ) repair pathway ligate both the signal ends to form an extrachromosomal signal joint and also the modified coding ends to restore chromosome integrity and generate a novel coding joint [201, 202]. After recombination, these signal joint T-cell receptor excision circles (TRECs) remain relatively stable for several days and can be used as a measure of thymic productivity [203].

Once assembled, a V(D)J joint defines the leader and antigen binding exons of the antigen receptor chain. During post-transcriptional processing, these mRNA exons are spliced to downstream constant region exons that encode the extracellular effector domains and transmembrane anchor of the protein. While TCRs exist only in a membrane-bound form, immunoglobulins developmentally switch between surface-bound (ie, the BCR made by B cells) or secreted (antibody made by plasma cells) forms by alternative splicing of the transmembrane anchor exons [204]. Notwithstanding the differences in alternative splicing between TCR and Ig messages, it is remarkable how conserved the process of recombination is between the various antigen receptor genes in B and T cells. The recombinase and end-joining machinery, as well as the RSS DNA targets that flank each V, D, and J gene segment are common to the assembly of all B and T antigen receptor genes.

The TCR genes of mice and humans share many characteristics. The murine *Tcrb* locus consists of 21 functional V segments (along with a number of nonfunctional pseudogenes) and two DJC cassettes, each containing one D segment, 6 functional J segments and a constant region (Figure 1.3). The human *TCRB* locus contains 35 functional Vs followed again by the same two DJC cassettes [205]. The human *TCRG* locus contains twelve V segments, followed by two cassettes that contain a

constant region preceded by 3 Jγ segments and a constant region preceded by 2 Jγ segments, respectively. The mouse *Tcrg,* by contrast, is much less linear, having seven V segments dispersed among each of 4 JγCγ cassettes. The *Tcra* and *Tcrd* genes in both mice and humans are intermingled in a single locus (*Tcra/d*) with a similar genomic location (mouse mm9: chromosome 14, 53040000-54853381); this joint locus consists of a somewhat interchangeable pool of Vα and Vδ segments [206] followed by 2 or 3 Dδ segments, 2 or four Jδ segments, a Cδ constant region, and then an extensive collection of Jα segments. As noted above, this genomic organization necessitates that an attempt at Vα-Jα recombination will eliminate the cell's capacity to become a γδ T cell through elimination of the TCRδ locus. Both the Ig heavy and light chain genes are also organized similarly between mouse and human. Consequently, our understanding of how human antigen receptor genes are assembled comes almost entirely from experimentation on the mouse *Tcr* and Ig genes.

1.5.4 Regulation of V(D)J recombination

As improper recombination leads to symptoms varying from immunodeficiencies to a predisposition to aggressive cancers, V(D)J recombination is a tightly regulated process. Expression of RAGs and recombination are cell lineage specific, restricted to ETPs and BLPs during specific stages of development as discussed above. Within these cell varieties additional restrictions are in place such that only ETPs rearrange T cell receptor loci and only BLPs rearrange B cell receptor loci. A notable exception to this rule is the IgH chain, which undergoes limited D to J joining in T cells, but can only join V to DJ in B cells for reasons as yet unknown [207, 208]. Recombination only occurs in cells paused at G1, and entry into replication leads to ubiquitin-mediated degradation of RAG2 and transcriptional downregulation of both RAG1 and RAG2 [209, 210]. Finally, within each individual cell paired and multi-step recombination events always occur in a stepwise fashion. For example, for

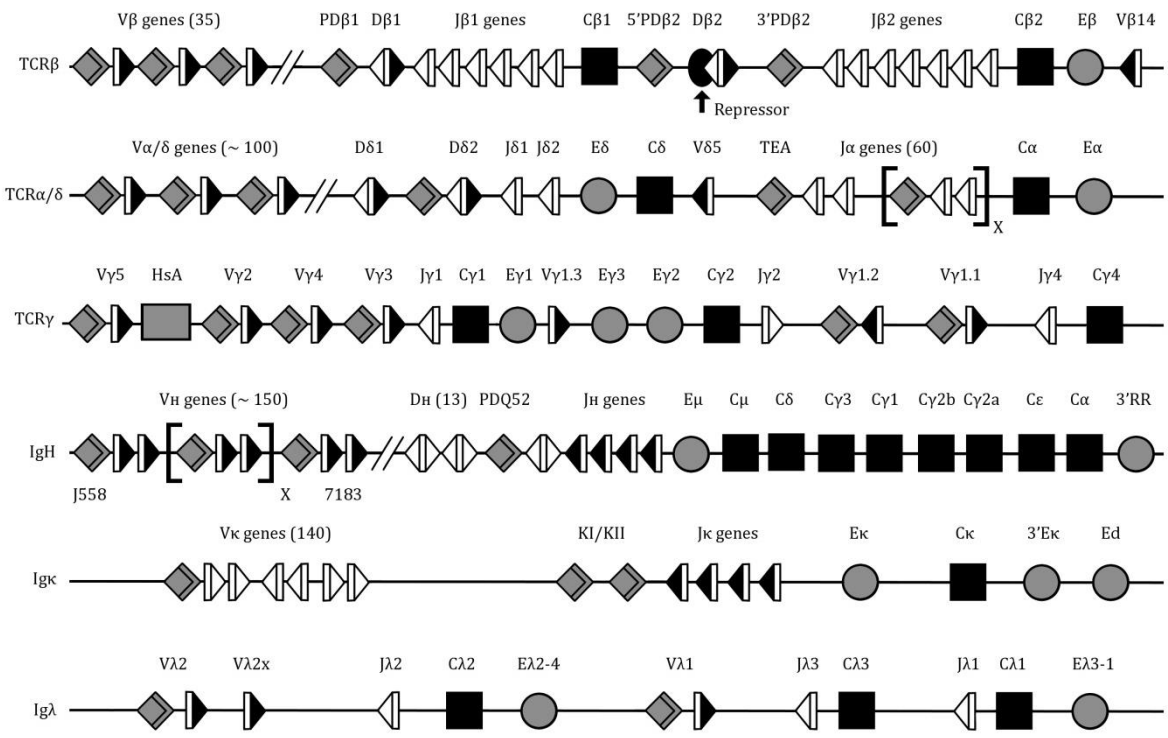


Figure 1.3: Map of murine antigen receptor loci. A representation of gene segments and known regulatory elements identified in each of the murine antigen receptor loci (not to scale). Promoters (arrowed diamonds), enhancers (grey circles), gene segments (small white rectangles), 23-RSSs (black triangles), 12-RSSs (white triangles), and constant regions (black squares) are shown. For the IgH locus the most proximal (7183) and distal (J558) V_H families are shown. Adapted from [211].

a functional αβ T cell receptor, D to J recombination of *Tcrb* necessarily precedes V to DJ recombination of *Tcrb*, which in turn precedes V to J recombination of *Tcra*. Finally, in the case of heavy-chain rearrangements (IgH in B cells, *Tcrb* in T cells), after creation and recognition of a functional joint at one allele V to DJ recombination is turned off for the other allele. This process, termed allelic exclusion, is required for monospecificity during clonal expansion in both B and T cells and remains one of the least-understood mechanisms of recombinational regulation [211, 212]

Targeting of the recombinase complex during rearrangement is mediated by conserved sequences that are present in some number at each V, D and J at every locus involved in rearrangement. These targeting sequences are called recombination signal sequences (RSSs) and come in two forms: one of which contains a palindromic heptamer followed by a poorly conserved 23 bp spacer and AT-rich nonamer (referred to as a 23 RSS), while the other RSS contains the same palindromic heptamer and AT-rich nonamer flanking a spacer of only 12 bases (12 RSS). Evidence suggests that the 12 RSS typically serves as the initial assembly point for the recombinase protein complex, and a compatible 23 RSS is then captured to form a synaptic recombination complex (Figure 1.4) [213]. 12 RSSs are located upstream of Jβ, Jα, Jγ, Dδ and Dβ in T cells and downstream of Igκ Vs, upstream of Igλ Js and flanking D segments in the IgH locus in B cells (Figure 3).

Interestingly, though recombinase can separately bind both 12 and 23 RSSs, recombinations under physiological conditions only occur between 12/23 RSS pairs: this is known as the 12/23 rule and plays a major role in controlling the order of recombination [214]. Adherence to the 12/23 rule *in vitro* was much stricter upon RAG complex assembly initiation at a 12 RSS than at a 23 RSS, providing some early evidence that RAG complexes initially assemble at the 12 RSS and leading to a "12RSS-first" model where the RAG complex first binds to and nicks the 12RSS, then recruits and

nicks the 23 RSS in every instance of recombination [215]. However, recent chromatin immunoprecipitation evidence support an alternate, "nonamer-first" model, one in which RAG complexes always bind to the downstream RSS (J or DJ), regardless of the type of flanking RSS, and nicking occurs first at the 12 RSS due to an enhanced RAG efficiency provided by the decreased distance between nonamer and heptamer inherent in a 12 RSS region [216, 217].

The 12/23 rule is not the only avenue through which RSSs exert control over the rearrangement process. Within *Tcrb* and *Tcrd*, where RSS configuration alone would be insufficient to ensure proper recombination order, additional mechanisms appear to be in place to constrain improper gene assembly, referred to as "beyond" or B12/23 restriction [218]. Wu et al. provided evidence for this additional level of restriction through generation of mice with a 3'Dβ 12 RSS in place of a Vβ 3' 12 RSS. This resulted in a dramatic increase in usage of that particular Vβ segment accompanied by the accumulation of cells containing direct Vβ to Jβ recombination, eschewing Dβ completely [219]. Other studies implicate both the integrity of nonamer and the spacer nucleotides of the 5' Dβ1 RSS in B12/23 restriction [220]. Polymorphism also exists within RSSs of the same type, and this polymorphism has been shown to directly correlate with the representation of the associated gene segments in the antigen receptor repertoire [221]. However, the plasticity of RSS identification by the recombinase complex is not without risk; it can lead to targeting mistakes and improper recombination utilizing so-called "cryptic RSSs", over 10 million of which are estimated to exist in the diploid genome [222, 223]

The majority of proteins in the recombinase complex are ubiquitously expressed components of the NHEJ pathway. In fact, only RAG-1/2 are lymphocyte-specific in their expression. RAG-1 serves as the principal DNA binding component and catalyzes DNA cleavage, while RAG-2

enhances RAG-1-DNA interaction and is required for DNA cleavage [216, 224]. The recombinase complex is proposed to be a tetramer of two RAG-1 proteins and two RAG-2 proteins; during binding, the RAG-1 proteins interact with the conserved nonamer of the RSSs on both strands simultaneously, while the conserved heptamer serves to enhance binding – this brings RAG-2 into association, though it interacts not with the DNA itself but with the local chromatin structure [216]. The enzymatic cleavage function of RAG-1, as well as its points of contact with RAG-2 and DNA, have been mapped to the central and C-terminal areas of the protein [225]. It is also likely that the N-terminus of RAG-1, though disposable to the enzymatic cleavage reaction itself, plays an important role in rearrangement - it possesses E3 ubiquitin ligase activity, which could aid in the disassembly of the protective RAG protein complex from coding joint hairpins to allow ligation of the DNA ends [226]. Transfection of the RAG proteins alone is enough to confer recombinase activity to cells of non-lymphoid lineages providing evidence that other lymphoid-specific genes are not required for enzymatic activity [227]. Though no mandatory role *in vivo* has been proven, *in vitro* evidence suggests that ubiquitous and non-specific DNA architectural high mobility group proteins HMGB1 and HMGB2 could stimulate RAG binding, promote cleavage at 23 RSSs and enhance fidelity to the 12/23 rule [224, 228].

RAG-1 and RAG-2 lie 15 kb apart from each other in an inverted fashion in the mouse genome and are often expressed in a coordinated fashion [229]. *In vivo* tracking of RAG-2 expression using RAG-2-GFP mouse strains reveal RAG expression is first detectable during the late ETP and DN2 stage of T cell development and the comparative A2 stage of B cell development, which correspond with the onset of heavy chain recombination in both cell types [229, 230]. IgH heavy chain or *Tcrb* rearrangement, followed by successful expression of the recombinant protein

on the cell surface in conjunction with cell's respective surrogate light chain (pTα for T cells, VpreB/λ5 for B cells) leads to allelic exclusion, a period of clonal expansion and downregulation of RAG expression [231, 232]. The mechanisms involved in downregulation of RAG expression are unclear, though a decrease in steady-state RAG mRNA levels is detected and protein levels are noticeably reduced [229, 232, 233]. Likewise, after developmental progression and IgL or *Tcra* rearrangement, RAG expression is ultimately terminated in response to antigen receptor cross-linking in T cells and peripheral immature B cells [232, 234, 235]. As receptor cross-linking is required during T cell positive selection, T cells that are mature enough to leave the thymus have already terminated RAG expression. For B cells, however, there appears to be a two-stage system. While the B cell is immature, cross-linking of surface receptors actually induces RAG protein expression; this allows for the rescue of self-reactive B cells through additional recombination in a process known as receptor editing [232, 236, 237]. As the cell matures, the amount of RAG mRNA decreases inversely as surface IgM levels increase and the cell becomes more likely to die by apoptosis upon self-recognition than to undergo receptor editing. Once B cells have undergone negative selection, they respond like DP T cells, eliminating RAG expression upon sufficient receptor cross-linking [232].

Resolution of recombination is achieved in a RAG-independent manner by the repair of DNA double-stranded breaks (DSBs) generated during the process. This repair process introduces three additional methods of receptor diversification: first through the addition of palindromic (P) nucleotides through an inexact opening of coding joint hairpins, secondly through the action of exonucleases that chew back germline nucleotides from these open hairpins in a manner associated with the nucleotide composition of the gene end, and finally through the enzyme terminal

deoxynucleotidal transferase (TdT), which adds random, nontemplated (N) nucleotides to the gene ends before repair completion [201, 238, 239]. Repair of DSBs generated during recombination is accomplished through the NHEJ DNA repair pathway (Figure 1.4) [240]. The NHEJ DNA repair pathway is made up of at least six proteins. Three of these, DNA-binding proteins Ku70 and Ku80 (also called Ku86) as well as a DNA-dependent protein kinase catalytic subunit (DNA-PKcs), make up the DNA-PK holoenzyme, involved in the early recognition of DSBs [201, 241]. The DNA-PKcs subunit of this holoenzyme interacts with a nuclease called Artemis to open RAG-generated hairpins. Upon opening, coding joint ends are ligated together through the actions of DNA ligase IV, stimulated by its complex protein partners XRCC4 and XLF, also called Cernunnos (Figure 1.4). Ligation through the NHEJ pathway is likely forced by the binding of the Ku70/Ku80 heterodimer, as homologous mutation of either Ligase IV or XRCC4 results in embryonic lethality and the inactivation of Ku80 rescues these mutations [242, 243]. The absence of XLF is accompanied by a notable decline in V(D)J recombination in both mice and humans [244-246]. In a more general sense, disruption of proteins involved in the NHEJ pathway result in a variety of karyotypic abnormalities including translocations which can lead to tumors. In fact, mice doubly deficient in any of the six mentioned NHEJ proteins and the tumor suppressor gene p53 have an increased mortality rate and level of tumor incidence, a majority of these tumors being pro-B cell lymphomas [201, 247]. Though there is some evidence that polymorphism of NHEJ factors can affect susceptibility to some cancers, the role NHEJ defects play in human tumors remains largely unknown [201].

1.6 Regulation of Tcra/d

Tcra and *Tcrd* loci co-localize in both the murine and human genomes (Figure 1.3), with the entirety of the δ locus contained between the Vα and Jα segments (*Tcra/d* locus spans over 1.6

million bases on mouse chromosome 14) [248]. This grouping presents a problem for developmental timing, as *Tcrd* undergoes rearrangement in DN cells while *Tcra* rearrangement is restricted to the DP stage of thymic development. Many of the V segments in the combined locus are actually shared between *Tcra* and *Tcrd.* These shared V segments are distributed throughout the 5' portion of the locus, many being proximal to the Dδ and Jδ and intermingled with the V segments that are exclusively used by *Tcrd,* but others being positioned much farther upstream amongst the V segments exclusively used by *Tcra* [249]. Given that a single recombinase complex must distinguish between such intermingled Vα and Vδ RSSs during DN development, *Tcra/d* offers a unique window onto the regulatory programs that govern gene activation and V(D)J recombination.

1.6.1 Accessibility control

Despite the developmental regulation of RAG transcription and translation, additional mechanisms account for stepwise recombination and allelic exclusion. Given that RAGs are required to interact with RSSs in order to perform their function, control of gene segment accessibility to RAG is a key to proper recombinational regulation. Yancopolous and Alt first proposed such an accessibility model in which chromatin configuration controls differential accessibility of gene segments during recombination in 1984 [250]. Indeed, both nucleosome structure and histone tails have been shown to provide a significant impediment to recombinase activity, which can be relieved using nucleosome remodeling complexes and tail modification [251]. Also, mixing RAG proteins with isolated B and T nuclei *in vitro* revealed that pro-B cells could not recombine T cell loci and vice versa, implying a method of locus-specific RAG inhibition inherent to the nucleus [252]. Numerous DNA elements within TCR and BCR loci have been discovered that are implicated in both chromatin

modification and recombinational efficacy as described below [253]. Most importantly, the generation of sterile or non-coding transcripts of unrearranged gene segments is common to all antigen receptor loci. Evidence now strongly suggests that the ability of this germ line transcription to forcibly remodel chromatin is critical for recombinational accessibility [253-255].

1.6.2 Conformational control

Due to the vast size of the *Tcra/d* locus, which contains over 160 V, D or J gene segments, many promoters and two transcriptional enhancers (Eα and Eδ), germ line transcription and recombination require extensive chromatin reorganization to juxtapose promoters with their enhancers and upstream V segments with their more 3' D or J counterparts (Figure 1.3). In 2010, the Krangel laboratory used 3D fluorescence *in situ* hybridization (3D-FISH) to characterize the distance between four different regions of the locus (distal V, central Vs, d-proximal Vs and Eα) in B cells, DN thymocytes and DP thymocytes [248]. They found dramatic locus contraction during the DN stage, consistent with the necessity of δ recombination to target distal Vδ segments. However, upon transition to the DP stage and onset of *Tcra* recombination, the 5' end of the locus underwent a moderate relaxation, rendering the distal Vs less accessible to capture by a recombination complex. This pattern of locus contraction/relaxation correlates well with the observed pattern of Vα rearrangements, where early Vα-to-Jα joints involve proximal Vs and the more distal Vs are reserved for secondary attempts to rescue out-of-frame primary joints [256]. The Krangel group noted in their study that locus contraction is independent of Eα, but that formation of DNA loops that connect Eα to the Jα germ line promoters T early alpha [182] and Jα49 are absolutely dependent on CTCF bound to each of the three elements [257]. Loss of DNA looping at the α/δ locus negatively impacted Jα rearrangement [257].

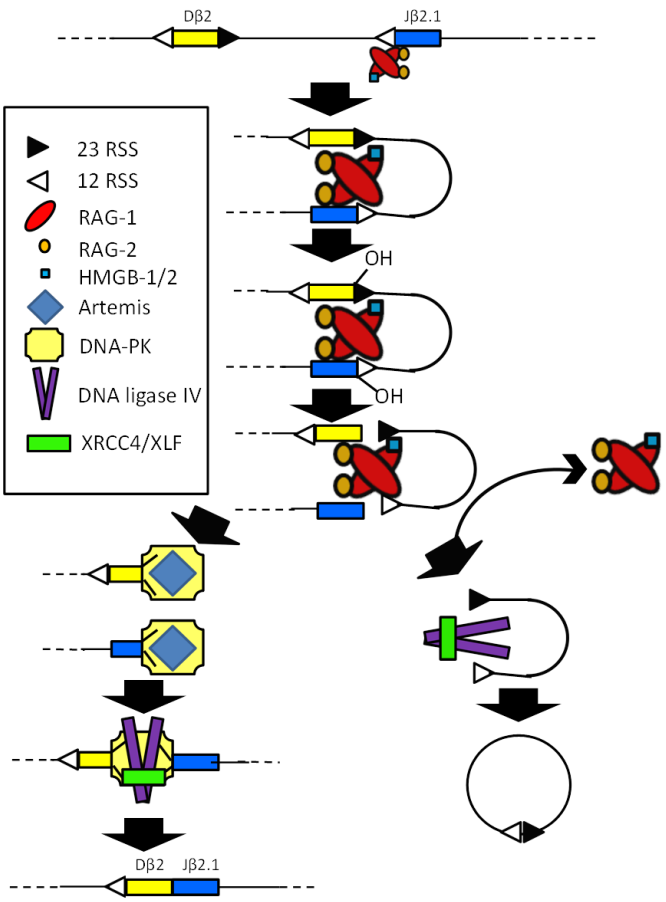


Figure 1.4: Schematic overview of RAG-mediated DNA rearrangement. Initiation of recombination occurs between a 23-RSS and a 12-RSS. Binding of the recombinase complex to a 12 RSS induces DNA looping and recruitment of a compatible 23-RSS, followed by generation of blunt double strand breaks between the gene segment (coding ends) and the RSSs (signal ends). At this point the RAG complex dissociates and action of NHEJ DNA repair proteins resolve the breaks into a genomic coding joint between the two segments and an extragenomic signal joint which contains the RSSs. Adapted from [201, 211, 224].

1.6.3 cis-acting regulatory elements: Enhancers

As noted above for *Tcra,* antigen receptor gene enhancers activate germ line promoters through "DNA looping", in which proteins bound to the enhancers interact directly with proteins bound to promoters to bring these stretches of DNA into close proximity [258]. Also, enhancers can augment chromatin relaxation and remodeling via recruitment of histone-modifying enzymes that increase transcription factor access to promoter regulatory elements [259]. Each antigen receptor gene contains one or more enhancers positioned at the 3' end of the locus (Figure 1.3). The first critical evidence in support of the recombinational accessibility model came from the demonstration that deleting the intronic Ig heavy chain enhancer iEμ significantly attenuated both germ line promoter activity and V(D)J recombination in the heavy chain gene of knockout mice [260] . Subsequent knockout experiments in each of the six remaining antigen receptor loci have confirmed the central role of antigen receptor enhancers in controlling germline promoter activity and recombination [261-268].

In the case of *Tcrb*, the lone enhancer element (Eβ) is located downstream of both DJβCβ cassettes, about 20 kb from Dβ1 and 6 kb from Dβ2 [269]. The Alt laboratory utilized the cre-*loxP* system to analyze the effects of targeted Eβ deletion in chimeric mice. They found specific deletion of Eβ reduced Dβ to Jβ recombination in T cells by 95%, as well as slashing transcription of germline *Tcrb* alleles by over a thousand-fold. Interestingly, replacement of Eβ with iEμ rescued both transcription and recombination in the β locus in T cells, suggesting similar activation potential for these two enhancers [265]. However, though this replacement induces transcription of Cβ in chimeric B cells as well, Dβ to Jβ recombination was not detected in B lineage cells of the mutant mice, indicating that additional cell type-specific mechanisms regulate β recombination *in vivo* [265].

Likewise, replacement of Eα with either iEμ or Eδ in similar experiments failed to rescue *Tcra* recombination [270]. Further studies in mice indicate that a homozygous deletion of Eβ results in a complete αβ-specific block in T cell development, as well as a severe attenuation of germline Vβ and DJβ transcription, cementing the tremendous importance of Eβ to proper *Tcrb* recombination [271].

Activation of Eβ during hematopoietic development follows an interesting pattern. Utilizing a GFP-Eβ transgenic mouse model, along with adoptive transfer to track the eventual fate of cell progenitors, Norris and colleagues found that both HSCs and CLPs had a significant population of GFP+ Eβ+ cells [272]. In the HSC, Eβ activation did not visibly lead to any developmental bias. Surprisingly, in the later CLP stage of hematopoietic development, activation of Eβ was associated with a propensity towards a B cell fate rather than a T cell fate [272]. Since both Eβ+ and Eβ- CLPs were capable of T cell differentiation when injected into the thymus, Eβ activation can be considered part of a broader collection of developmental changes in which fate is determined by the local microenvironment [272].

Differential enhancer activity is also involved in the transcription of recombinase proteins RAG-1 and RAG-2. These proteins are differentially regulated in B and T cells, with one upstream enhancer involved in both B and T cell development and another, more proximal enhancer restricted to the B cell lineage [273, 274]. Using a transgenic system that expresses a reporter protein only after successful recombination of compatible RSSs within the transgene, the Gerstein laboratory found evidence that a functional recombinase complex is expressed not only before and independent of chromatin remodeling at antigen receptor gene loci, but also that this complex is active during the earlier CLP phase of hematopoietic development, which accounts for the accumulation of IgH D to J recombinations in NK and dendritic cells [275, 276].

In *Tcra/d*, the close proximity of Eα and Eδ allows for a kind of cross-activation of the two. In the DN3 stage of T cell development, transcription and recombination of the δ locus are frequent whereas the α locus remains inaccessible [277, 278]. However, upon transition to DP, transcription and recombination of both loci are possible despite a change in protein occupation at Eδ that renders it non-functional [264, 278]. Eα activation, contrastingly, appears not to be associated with a change in protein occupancy; genomic footprinting assays reveal a similar protein-binding pattern at Eα during DN3 and DP development [278]. Rather, epigenetic factors such as nucleosomal distortions and permissive histone modifications at Eα and the promoter TEA likely control their developmental activity [278].

1.6.4 cis-acting regulatory elements: Promoters

Antigen receptor loci contain a large number of promoters - DNA elements that serve as binding sites for factors involved in the initiation of transcription and that exert substantial control over gene accessibility to recombinase. Extensive evidence suggests that antigen receptor promoters function in a highly localized fashion [253, 279]. In each antigen receptor gene locus, promoter elements located upstream of V segments are responsible for driving transcription of novel antigen receptors post-recombination, as well as facilitating access of V segments to recombinase during recombination [211, 250, 280]. Promoters are also heavily involved in accessibility of D and J segments during rearrangement. The work of the Sikes and Oltz laboratories has revealed a total of three promoters in the D region of *Tcrb* (Figure 1.3). One promoter, located immediately upstream of Dβ1, initiates transcription of the Dβ1 coding sequences and is required for DJβ1 rearrangement while having no effect on DJβ2 or later V to DJ rearrangements [281, 282]. The Dβ2 region has both an upstream and a downstream promoter: the germline transcription

required for proper recombination of the Dβ2Jβ2Cβ2 cassette is facilitated by the downstream promoter, while the upstream promoter is held in check by a repressor located in the spacer sequence of the Dβ2 12 RSS [283, 284]. Recombination within this cassette deletes the downstream promoter and leads to elimination of repression at the upstream promoter, likely due to a loss in binding of the transcription factor USF-1 to Dβ2 12 RSS region caused by the activation of DNA repair pathways in response to recombination [284]. This allows the upstream promoter to facilitate transcription of Dβ2Jβ2 rearrangements, making them available for future V to Dβ2Jβ2 recombination [283, 285].

The Krangel laboratory has defined a relatively similar layout of promoters across the *Tcra* locus. In addition to each Vα segment having its own upstream promoter, the Jα region is scattered with promoters that exist in a 5'-to-3' directional hierarchy such that the most 5' [182] is first activated upon DP transition. Its deletion during recombination leads to activation of the more 3' Jα49 promoter, and its deletion in turn leads to activation of yet further 3' promoters [286, 287]. This Jα promoter layout facilitates an "inside-out" method of Vα to Jα recombination that allows for multiple attempts at generating an in-frame *Tcra* rearrangement [288-290]. In *Tcrg*, V segment promoters enforce a developmental bias where fetal recombination favors usage of Vγ3 and adult recombination eschews this segment. This switch is moderated by the promoter region of Vγ3 and Vγ4, which contain sequences repressive to adult recombination [249, 291]. A similar fetal to adult usage switch can also be found in the *Tcrd* locus, where TRDV4 is selectively utilized in fetal thymocytes and suppressed in adult thymocytes. This switch is moderated by Eδ and developmentally regulated histone acetylation that is likely promoter-mediated [286].

Promoter regulation of immunoglobulin assembly parallels that described above for the TCR genes. In the heavy chain B cell locus (IgH), there is one promoter per pair of V segments (Figure 1.3). These promoters all share a highly conserved octameric sequence that is important for germline transcription as well as for facilitating interaction with a downstream enhancer element through a looping mechanism mediated by CTCF and promoter-bound TFIID as well as enhancer-bound OCA-B [211, 292]. Within the D segments of IgH, a germ line promoter (DQ52) associated with the 3'most D segment has been identified. The absence of DQ52 does not dramatically alter either germ line transcription or the recombination efficiency of DJH segments, however, suggesting the presence of additional promoters or unique regulatory elements [293]. Germ line promoters critical for recombination have been identified and characterized in the light-chain *Igk* and *Igl* loci as well [294, 295].

1.6.5 The T Early Alpha Promoter

Germ line transcription across the Jα region of *Tcra* begins at the T early alpha promoter, generating transcription in early DP cells that extends across a 75 kb cassette containing the Jα and Cα gene segments [296]. As noted above, the Jα49 promoter sits about 15kb downstream of TEA, while additional poorly defined promoters are associated with progressively distal J segments [287]. These germ line promoters exist in a functional hierarchy, such that only the most 5' promoter is active, and Jα usage correlates with proximity to the active germ line promoter [287]. As such, prior to rearrangement only TEA shows significant transcriptional activity. Following its deletion by Vα rearrangement to one of the first several Jα segments, PJα49 is activated, and initial secondary rearrangements target Jα segments immediately 3' of PJα49. Deletion of PJα49 then allows activation of promoter activity 5' of Jα37 and recombination of immediately 3' Js, and so on [287].

These usage patterns also coincide with the distribution of H3K4me3 across the Jα cassette [288]. Since germline transcription is conserved in all antigen receptor genes as far back as the cartilaginous fish where antigen receptors appear to have arisen [297-299], it is reasonable to conclude that the programs that govern V(D)J recombination are critically dependent on transcription-induced epigenetic changes in the accessibility of antigen receptor gene regions to recombinase.

1.6.6 The role of TEA priming during ES cell development

As noted above, TEA is activated in a CTCF-dependent manner upon the transition of thymocytes to the DP stage of development, and this activation is essential for normal *Tcra* recombination and development of the αβ T cell repertoire. The role of CTCF in TEA activation is well established now, with separate studies showing that CTCF mediates long-range DNA loops that juxtapose TEA with Eα. Given the precise developmental timing of TEA activation, we were surprised in preliminary studies into the developmental regulation of TEA to find that CTCF binding at TEA and Eα is conserved throughout development and differentiation. This unexpected binding pattern became the focus of my subsequent thesis research. In this dissertation, I present my findings regarding a role for CTCF in embryonic priming of TEA. I show that (i) TEA maintains CTCF binding in ground-state embryonic stem cells equivalent to pre-implantation embryogenesis, (ii) this binding does not involve the formation of DNA loops with Eα that are required for promoter activity in Double Positive T cells, (iii) CTCF binding is required to maintain CpG demethylation across the TEA regulatory region, and (iv) similar CTCF binding patterns at other demethylated tissue-specific promoters in ESCs suggests that CTCF may play a critical role in priming tissue-specific genes across the genome of ESCs by protecting them against repressive DNA methylation. In the appendix to this

thesis, I describe a related project I undertook in collaboration with other members of the Sikes laboratory to define the role of USF-1 in the feedback regulation of Dβ2 transcription in response to RAG-mediated DSBs.

1.7. References

1. CH, W., *Introduction to Modern Genetics*. 1939, London: Allen and Unwin.

2. Hadorn, E., *Developmental Genetics and Lethal Factors*. 1960, London: Methuen.

3. Henikoff, S., E. McKittrick, and K. Ahmad, *Epigenetics, histone H3 variants, and the inheritance of chromatin states.* Cold Spring Harb Symp Quant Biol, 2004. **69**: p. 235-43.

4. Pusarla, R.H. and P. Bhargava, *Histones in functional diversification. Core histone variants.* FEBS J, 2005. **272**(20): p. 5149-68.

5. Holde, K.v., *Chromatin*, ed. A. Wolfe. 1988, New York: Springer-Verlag.

6. Allfrey, V.G., R. Faulkner, and A.E. Mirsky, *Acetylation and Methylation of Histones and Their Possible Role in the Regulation of Rna Synthesis.* Proc Natl Acad Sci U S A, 1964. **51**: p. 786-94.

7. Musselman, C.A., et al., *Perceiving the epigenetic landscape through histone readers.* Nat Struct Mol Biol, 2012. **19**(12): p. 1218-27.

8. Riggs, A.D., *X inactivation, differentiation, and DNA methylation.* Cytogenet Cell Genet, 1975. **14**(1): p. 9-25.

9. Holliday, R. and J.E. Pugh, *DNA modification mechanisms and gene activity during development.* Science, 1975. **187**(4173): p. 226-32.

10. Khorasanizadeh, S., *The nucleosome: from genomic organization to genomic regulation.* Cell, 2004. **116**(2): p. 259-72.

11. Harp, J.M., et al., *Asymmetries in the nucleosome core particle at 2.5 A resolution.* Acta Crystallogr D Biol Crystallogr, 2000. **56**(Pt 12): p. 1513-34.

12. Fu, Y., et al., *The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome.* PLoS Genet, 2008. **4**(7): p. e1000138.

13. Olynik, B.M. and M. Rastegar, *The genetic and epigenetic journey of embryonic stem cells into mature neural cells.* Front Genet, 2012. **3**: p. 81.

14. Bowman, G.D., *Mechanisms of ATP-dependent nucleosome sliding.* Curr Opin Struct Biol, 2010. **20**(1): p. 73-81.

15. Wegel, E. and P. Shaw, *Gene activation and deactivation related changes in the three-dimensional structure of chromatin.* Chromosoma, 2005. **114**(5): p. 331-7.

16. Li, G. and D. Reinberg, *Chromatin higher-order structures and gene regulation.* Curr Opin Genet Dev, 2011. **21**(2): p. 175-86.

17. Marzluff, W.F., et al., *The human and mouse replication-dependent histone genes.* Genomics, 2002. **80**(5): p. 487-98.

18. Boskovic, A., et al., *Analysis of active chromatin modifications in early mammalian embryos reveals uncoupling of H2A.Z acetylation and H3K36 trimethylation from embryonic genome activation.* Epigenetics, 2012. **7**(7): p. 747-57.

19. Banaszynski, L.A., C.D. Allis, and P.W. Lewis, *Histone variants in metazoan development.* Dev Cell, 2010. **19**(5): p. 662-74.

20. Altaf, M., et al., *Connection between histone H2A variants and chromatin remodeling complexes.* Biochem Cell Biol, 2009. **87**(1): p. 35-50.

21. Redon, C., et al., *Histone H2A variants H2AX and H2AZ.* Curr Opin Genet Dev, 2002. **12**(2): p. 162-9.

22. Soria, G., S.E. Polo, and G. Almouzni, *Prime, repair, restore: the active role of chromatin in the DNA damage response.* Mol Cell, 2012. **46**(6): p. 722-34.

23. Xu, Y., et al., *Histone H2A.Z controls a critical chromatin remodeling step required for DNA double-strand break repair.* Mol Cell, 2012. **48**(5): p. 723-33.

24. Du, X.Z., et al., *Identical sequences but different expression patterns of Hira gene in gynogenetic and gonochoristic crucian carps.* Fish Physiol Biochem, 2008. **34**(2): p. 175-84.

25. Drane, P., et al., *The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3.* Genes Dev, 2010. **24**(12): p. 1253-65.

26. Goldberg, A.D., et al., *Distinct factors control histone variant H3.3 localization at specific genomic regions.* Cell, 2010. **140**(5): p. 678-91.

27. Szenker, E., D. Ray-Gallet, and G. Almouzni, *The double face of the histone variant H3.3.* Cell Res, 2011. **21**(3): p. 421-34.

28. Santenard, A., et al., *Heterochromatin formation in the mouse embryo requires critical residues of the histone variant H3.3.* Nat Cell Biol, 2010. **12**(9): p. 853-62.

29. Millar, C.B., *Organizing the genome with H2A histone variants.* Biochem J, 2013. **449**(3): p. 567-79.

30. Smith, S. and B. Stillman, *Stepwise assembly of chromatin during DNA replication in vitro.* EMBO J, 1991. **10**(4): p. 971-80.

31. Burgess, R.J. and Z. Zhang, *Histone chaperones in nucleosome assembly and human disease.* Nat Struct Mol Biol, 2013. **20**(1): p. 14-22.

32. Zhou, J., et al., *The nucleosome surface regulates chromatin compaction and couples it with transcriptional repression.* Nat Struct Mol Biol, 2007. **14**(11): p. 1070-6.

33. West, M.H. and W.M. Bonner, *Histone 2A, a heteromorphous family of eight protein species.* Biochemistry, 1980. **19**(14): p. 3238-45.

34. Ziegler-Birling, C., et al., *Distribution of p53 binding protein 1 (53BP1) and phosphorylated H2A.X during mouse preimplantation development in the absence of DNA damage.* Int J Dev Biol, 2009. **53**(7): p. 1003-11.

35. Faast, R., et al., *Histone variant H2A.Z is required for early mammalian development.* Curr Biol, 2001. **11**(15): p. 1183-7.

36. Bellucci, L., et al., *Activation of p21 by HDAC inhibitors requires acetylation of H2A.Z.* PLoS One, 2013. **8**(1): p. e54102.

37. Kusch, T., et al., *Acetylation by Tip60 is required for selective histone variant exchange at DNA lesions.* Science, 2004. **306**(5704): p. 2084-7.

38. Millar, C.B., et al., *Acetylation of H2AZ Lys 14 is associated with genome-wide gene activity in yeast.* Genes Dev, 2006. **20**(6): p. 711-22.

39. Zlatanova, J. and A. Thakar, *H2A.Z: view from the top.* Structure, 2008. **16**(2): p. 166-79.

40. Nekrasov, M., et al., *Histone H2A.Z inheritance during the cell cycle and its impact on promoter organization and dynamics.* Nat Struct Mol Biol, 2012. **19**(11): p. 1076-83.

41. Kelly, T.K., et al., *H2A.Z maintenance during mitosis reveals nucleosome shifting on mitotically silenced genes.* Mol Cell, 2010. **39**(6): p. 901-11.

42. Chang, C.C., et al., *A maternal store of macroH2A is removed from pronuclei prior to onset of somatic macroH2A expression in preimplantation embryos.* Dev Biol, 2005. **278**(2): p. 367-80.

43. Costanzi, C. and J.R. Pehrson, *Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals.* Nature, 1998. **393**(6685): p. 599-601.

44. Pasque, V., et al., *Epigenetic stability of repressed states involving the histone variant macroH2A revealed by nuclear transfer to Xenopus oocytes.* Nucleus, 2011. **2**(6): p. 533-9.

45. Luger, K., et al., *Crystal structure of the nucleosome core particle at 2.8 A resolution.* Nature, 1997. **389**(6648): p. 251-60.

46. Strahl, B.D. and C.D. Allis, *The language of covalent histone modifications.* Nature, 2000. **403**(6765): p. 41-5.

47. Chambeyron, S. and W.A. Bickmore, *Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription.* Genes Dev, 2004. **18**(10): p. 1119-30.

48. Li, B., M. Carey, and J.L. Workman, *The role of chromatin during transcription.* Cell, 2007. **128**(4): p. 707-19.

49. Richards, E.J. and S.C. Elgin, *Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects.* Cell, 2002. **108**(4): p. 489-500.

50. Ismail, I.H., et al., *BMI1-mediated histone ubiquitylation promotes DNA double-strand break repair.* J Cell Biol, 2010. **191**(1): p. 45-60.

51. O'Hagan, H.M., H.P. Mohammad, and S.B. Baylin, *Double strand breaks can initiate gene silencing and SIRT1-dependent onset of DNA methylation in an exogenous promoter CpG island.* PLoS Genet, 2008. **4**(8): p. e1000155.

52. Ito, T., *Role of histone modification in chromatin dynamics.* J Biochem, 2007. **141**(5): p. 609-14.

53. Govin, J., et al., *Systematic screen reveals new functional dynamics of histones H3 and H4 during gametogenesis.* Genes Dev, 2010. **24**(16): p. 1772-86.

54. Gu, L., Q. Wang, and Q.Y. Sun, *Histone modifications during mammalian oocyte maturation: dynamics, regulation and functions.* Cell Cycle, 2010. **9**(10): p. 1942-50.

55. Escobar, J., et al., *Redox signaling and histone acetylation in acute pancreatitis.* Free Radic Biol Med, 2012. **52**(5): p. 819-37.

56. Jenuwein, T. and C.D. Allis, *Translating the histone code.* Science, 2001. **293**(5532): p. 1074-80.

57. Kurdistani, S.K. and M. Grunstein, *Histone acetylation and deacetylation in yeast.* Nat Rev Mol Cell Biol, 2003. **4**(4): p. 276-84.

58. Dhalluin, C., et al., *Structure and ligand of a histone acetyltransferase bromodomain.* Nature, 1999. **399**(6735): p. 491-6.

59. Winston, F. and C.D. Allis, *The bromodomain: a chromatin-targeting module?* Nat Struct Biol, 1999. **6**(7): p. 601-4.

60. Yang, X.J. and E. Seto, *HATs and HDACs: from structure, function and regulation to novel strategies for therapy and prevention.* Oncogene, 2007. **26**(37): p. 5310-8.

61. Goodman, R.H. and S. Smolik, *CBP/p300 in cell growth, transformation, and development.* Genes Dev, 2000. **14**(13): p. 1553-77.

62. Lee, K.K. and J.L. Workman, *Histone acetyltransferase complexes: one size doesn't fit all.* Nat Rev Mol Cell Biol, 2007. **8**(4): p. 284-95.

63. Mihaylova, M.M. and R.J. Shaw, *Metabolic reprogramming by class I and II histone deacetylases.* Trends Endocrinol Metab, 2013. **24**(1): p. 48-57.

64. Grozinger, C.M. and S.L. Schreiber, *Deacetylase enzymes: biological functions and the use of small-molecule inhibitors.* Chem Biol, 2002. **9**(1): p. 3-16.

65. Brownell, J.E. and C.D. Allis, *Special HATs for special occasions: linking histone acetylation to chromatin assembly and gene activation.* Curr Opin Genet Dev, 1996. **6**(2): p. 176-84.

66. Ehrensberger, A.H. and J.Q. Svejstrup, *Reprogramming chromatin.* Crit Rev Biochem Mol Biol, 2012. **47**(5): p. 464-82.

67. Robyr, D., et al., *Microarray deacetylation maps determine genome-wide functions for yeast histone deacetylases.* Cell, 2002. **109**(4): p. 437-46.

68. Verdin, E., F. Dequiedt, and H.G. Kasler, *Class II histone deacetylases: versatile regulators.* Trends Genet, 2003. **19**(5): p. 286-93.

69. Lagger, G., et al., *Essential function of histone deacetylase 1 in proliferation control and CDK inhibitor repression.* EMBO J, 2002. **21**(11): p. 2672-81.

70. Vaquero, A., R. Sternglanz, and D. Reinberg, *NAD+-dependent deacetylation of H4 lysine 16 by class III HDACs.* Oncogene, 2007. **26**(37): p. 5505-20.

71. Dion, M.F., et al., *Genomic characterization reveals a simple histone H4 acetylation code.* Proc Natl Acad Sci U S A, 2005. **102**(15): p. 5501-6.

72. Clayton, A.L., C.A. Hazzalin, and L.C. Mahadevan, *Enhanced histone acetylation and transcription: a dynamic perspective.* Mol Cell, 2006. **23**(3): p. 289-96.

73. Hazzalin, C.A. and L.C. Mahadevan, *Dynamic acetylation of all lysine 4-methylated histone H3 in the mouse nucleus: analysis at c-fos and c-jun.* PLoS Biol, 2005. **3**(12): p. e393.

74. Wang, Z., et al., *Pro isomerization in MLL1 PHD3-bromo cassette connects H3K4me readout to CyP33 and HDAC-mediated repression.* Cell, 2010. **141**(7): p. 1183-94.

75. Marinova, Z., et al., *Histone deacetylase inhibition alters histone methylation associated with heat shock protein 70 promoter modifications in astrocytes and neurons.* Neuropharmacology, 2011. **60**(7-8): p. 1109-15.

76. Qian, C. and M.M. Zhou, *SET domain protein lysine methyltransferases: Structure, specificity and catalysis.* Cell Mol Life Sci, 2006. **63**(23): p. 2755-63.

77. Rice, J.C., et al., *Histone methyltransferases direct different degrees of methylation to define distinct chromatin domains.* Mol Cell, 2003. **12**(6): p. 1591-8.

78. Heintzman, N.D., et al., *Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome.* Nat Genet, 2007. **39**(3): p. 311-8.

79. Im, H., et al., *Dynamic regulation of histone H3 methylated at lysine 79 within a tissue-specific chromatin domain.* J Biol Chem, 2003. **278**(20): p. 18346-52.

80. Rea, S., et al., *Regulation of chromatin structure by site-specific histone H3 methyltransferases.* Nature, 2000. **406**(6796): p. 593-9.

81. Greer, E.L. and Y. Shi, *Histone methylation: a dynamic mark in health, disease and inheritance.* Nat Rev Genet, 2012. **13**(5): p. 343-57.

82. Jones, B., et al., *The histone H3K79 methyltransferase Dot1L is essential for mammalian development and heterochromatin structure.* PLoS Genet, 2008. **4**(9): p. e1000190.

83. Shi, Y., et al., *Histone demethylation mediated by the nuclear amine oxidase homolog LSD1.* Cell, 2004. **119**(7): p. 941-53.

84. Tsukada, Y., et al., *Histone demethylation by a family of JmjC domain-containing proteins.* Nature, 2006. **439**(7078): p. 811-6.

85. Walport, L.J., R.J. Hopkinson, and C.J. Schofield, *Mechanisms of human histone and nucleic acid demethylases.* Curr Opin Chem Biol, 2012. **16**(5-6): p. 525-34.

86. Razin, A., et al., *Variations in DNA methylation during mouse cell differentiation in vivo and in vitro.* Proc Natl Acad Sci U S A, 1984. **81**(8): p. 2275-9.

87. Gruenbaum, Y., et al., *Methylation of CpG sequences in eukaryotic DNA.* FEBS Lett, 1981. **124**(1): p. 67-71.

88. Bird, A.P., *CpG-rich islands and the function of DNA methylation.* Nature, 1986. **321**(6067): p. 209-13.

89. Hackett, J.A. and M.A. Surani, *DNA methylation dynamics during the mammalian life cycle.* Philos Trans R Soc Lond B Biol Sci, 2013. **368**(1609): p. 20110328.

90. Suzuki, M.M. and A. Bird, *DNA methylation landscapes: provocative insights from epigenomics.* Nat Rev Genet, 2008. **9**(6): p. 465-76.

91. Larsen, F., et al., *CpG islands as gene markers in the human genome.* Genomics, 1992. **13**(4): p. 1095-107.

92. Antequera, F., *Structure, function and evolution of CpG island promoters.* Cell Mol Life Sci, 2003. **60**(8): p. 1647-58.

93. Reik, W., W. Dean, and J. Walter, *Epigenetic reprogramming in mammalian development.* Science, 2001. **293**(5532): p. 1089-93.

94. Smith, Z.D., et al., *A unique regulatory phase of DNA methylation in the early mammalian embryo.* Nature, 2012. **484**(7394): p. 339-44.

95. Reik, W., *Stability and flexibility of epigenetic gene regulation in mammalian development.* Nature, 2007. **447**(7143): p. 425-32.

96. Blewitt, M.E., et al., *Dynamic reprogramming of DNA methylation at an epigenetically sensitive allele in mice.* PLoS Genet, 2006. **2**(4): p. e49.

97. Meissner, A., et al., *Genome-scale DNA methylation maps of pluripotent and differentiated cells.* Nature, 2008. **454**(7205): p. 766-70.

98. Suva, M.L., N. Riggi, and B.E. Bernstein, *Epigenetic reprogramming in cancer.* Science, 2013. **339**(6127): p. 1567-70.

99. Baylin, S.B. and P.A. Jones, *A decade of exploring the cancer epigenome - biological and translational implications.* Nat Rev Cancer, 2011. **11**(10): p. 726-34.

100. Tahiliani, M., et al., *Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1.* Science, 2009. **324**(5929): p. 930-5.

101. Okano, M., et al., *DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development.* Cell, 1999. **99**(3): p. 247-57.

102. Bell, A.C. and G. Felsenfeld, *Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene.* Nature, 2000. **405**(6785): p. 482-5.

103. Kim, T.H., et al., *Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome.* Cell, 2007. **128**(6): p. 1231-45.

104. Wang, H., et al., *Widespread plasticity in CTCF occupancy linked to DNA methylation.* Genome Res, 2012. **22**(9): p. 1680-8.

105. Dixon, J.R., et al., *Topological domains in mammalian genomes identified by analysis of chromatin interactions.* Nature, 2012. **485**(7398): p. 376-80.

106. Xu, J., et al., *Pioneer factor interactions and unmethylated CpG dinucleotides mark silent tissue-specific enhancers in embryonic stem cells.* Proc Natl Acad Sci U S A, 2007. **104**(30): p. 12377-82.

107. Moore, J.M., et al., *Loss of maternal CTCF is associated with peri-implantation lethality of Ctcf null embryos.* PLoS One, 2012. **7**(4): p. e34915.

108. Oron, E. and N. Ivanova, *Cell fate regulation in early mammalian development.* Phys Biol, 2012. **9**(4): p. 045002.

109. Medvedev, S.P., A.I. Shevchenko, and S.M. Zakian, *Molecular basis of Mammalian embryonic stem cell pluripotency and self-renewal.* Acta Naturae, 2010. **2**(3): p. 30-46.

110. Meilhac, S.M., et al., *Active cell movements coupled to positional induction are involved in lineage segregation in the mouse blastocyst.* Dev Biol, 2009. **331**(2): p. 210-21.

111. Albert, M. and A.H. Peters, *Genetic and epigenetic control of early mouse development.* Curr Opin Genet Dev, 2009. **19**(2): p. 113-21.

112. Balhorn, R., *The protamine family of sperm nuclear proteins.* Genome Biol, 2007. **8**(9): p. 227.

113. Santos, F., et al., *Dynamic reprogramming of DNA methylation in the early mouse embryo.* Dev Biol, 2002. **241**(1): p. 172-82.

114. Hemberger, M., W. Dean, and W. Reik, *Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal.* Nat Rev Mol Cell Biol, 2009. **10**(8): p. 526-37.

115. Howell, C.Y., et al., *Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene.* Cell, 2001. **104**(6): p. 829-38.

116. Farthing, C.R., et al., *Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes.* PLoS Genet, 2008. **4**(6): p. e1000116.

117. Palini, S., et al., *Epigenetic regulatory mechanisms during preimplantation embryo development.* Ann N Y Acad Sci, 2011. **1221**: p. 54-60.

118. Saitou, M., S. Kagiwada, and K. Kurimoto, *Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells.* Development, 2012. **139**(1): p. 15-31.

119. Puschendorf, M., et al., *PRC1 and Suv39h specify parental asymmetry at constitutive heterochromatin in early mouse embryos.* Nat Genet, 2008. **40**(4): p. 411-20.

120. Rugg-Gunn, P.J., et al., *Distinct histone modifications in stem cell lines and tissue lineages from the early mouse embryo.* Proc Natl Acad Sci U S A, 2010. **107**(24): p. 10783-90.

121. Yuan, P., et al., *Eset partners with Oct4 to restrict extraembryonic trophoblast lineage potential in embryonic stem cells.* Genes Dev, 2009. **23**(21): p. 2507-20.

122. Hay, D.C., et al., *Oct-4 knockdown induces similar patterns of endoderm and trophoblast differentiation markers in human and mouse embryonic stem cells.* Stem Cells, 2004. **22**(2): p. 225-35.

123. Mitsui, K., et al., *The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells.* Cell, 2003. **113**(5): p. 631-42.

124. Adachi, K., et al., *Role of SOX2 in maintaining pluripotency of human embryonic stem cells.* Genes Cells, 2010. **15**(5): p. 455-70.

125. Pant, V., et al., *Mutation of a single CTCF target site within the H19 imprinting control region leads to loss of Igf2 imprinting and complex patterns of de novo methylation upon maternal inheritance.* Mol Cell Biol, 2004. **24**(8): p. 3497-504.

126. Ohlsson, R., V. Lobanenkov, and E. Klenova, *Does CTCF mediate between nuclear organization and gene expression?* Bioessays, 2010. **32**(1): p. 37-50.

127. Han, L., D.H. Lee, and P.E. Szabo, *CTCF is the master organizer of domain-wide allele-specific chromatin at the H19/Igf2 imprinted region.* Mol Cell Biol, 2008. **28**(3): p. 1124-35.

128. Wan, L.B., et al., *Maternal depletion of CTCF reveals multiple functions during oocyte and preimplantation embryo development.* Development, 2008. **135**(16): p. 2729-38.

129. Mito, Y., J.G. Henikoff, and S. Henikoff, *Histone replacement marks the boundaries of cis-regulatory domains.* Science, 2007. **315**(5817): p. 1408-11.

130. Bell, A.C., A.G. West, and G. Felsenfeld, *The protein CTCF is required for the enhancer blocking activity of vertebrate insulators.* Cell, 1999. **98**(3): p. 387-96.

131. Hark, A.T., et al., *CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus.* Nature, 2000. **405**(6785): p. 486-9.

132. Holohan, E.E., et al., *CTCF genomic binding sites in Drosophila and the organisation of the bithorax complex.* PLoS Genet, 2007. **3**(7): p. e112.

133. Maeda, R.K. and F. Karch, *Making connections: boundaries and insulators in Drosophila.* Curr Opin Genet Dev, 2007. **17**(5): p. 394-9.

134. Kagey, M.H., et al., *Mediator and cohesin connect gene expression and chromatin architecture.* Nature, 2010. **467**(7314): p. 430-5.

135. Wendt, K.S., et al., *Cohesin mediates transcriptional insulation by CCCTC-binding factor.* Nature, 2008. **451**(7180): p. 796-801.

136. Remeseiro, S., et al., *A unique role of cohesin-SA1 in gene regulation and development.* EMBO J, 2012. **31**(9): p. 2090-102.

137. Liu, Z., et al., *Control of embryonic stem cell lineage commitment by core promoter factor, TAF3.* Cell, 2011. **146**(5): p. 720-31.

138. Herold, M., M. Bartkuhn, and R. Renkawitz, *CTCF: insights into insulator function during development.* Development, 2012. **139**(6): p. 1045-57.

139. Kehayova, P., et al., *Regulatory elements required for the activation and repression of the protocadherin-alpha gene cluster.* Proc Natl Acad Sci U S A, 2011. **108**(41): p. 17195-200.

140. Yang, Y., et al., *CTCF is essential for up-regulating expression from the amyloid precursor protein promoter during differentiation of primary hippocampal neurons.* J Neurochem, 1999. **73**(6): p. 2286-98.

141. Majumder, P., et al., *The insulator factor CTCF controls MHC class II gene expression and is required for the formation of long-distance chromatin interactions.* J Exp Med, 2008. **205**(4): p. 785-98.

142. Hadjur, S., et al., *Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus.* Nature, 2009. **460**(7253): p. 410-3.

143. Spencer, R.J., et al., *A boundary element between Tsix and Xist binds the chromatin insulator Ctcf and contributes to initiation of X-chromosome inactivation.* Genetics, 2011. **189**(2): p. 441-54.

144. Ribeiro de Almeida, C., et al., *The DNA-binding protein CTCF limits proximal Vkappa recombination and restricts kappa enhancer interactions to the immunoglobulin kappa light chain locus.* Immunity, 2011. **35**(4): p. 501-13.

145. Guo, C., et al., *CTCF-binding elements mediate control of V(D)J recombination.* Nature, 2011. **477**(7365): p. 424-30.

146. Dillon, N., *Factor mediated gene priming in pluripotent stem cells sets the stage for lineage specification.* Bioessays, 2012. **34**(3): p. 194-204.

147. Bernstein, B.E., et al., *A bivalent chromatin structure marks key developmental genes in embryonic stem cells.* Cell, 2006. **125**(2): p. 315-26.

148. Papp, B. and J. Muller, *Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins.* Genes Dev, 2006. **20**(15): p. 2041-54.

149. Szutorisz, H., et al., *Formation of an active tissue-specific chromatin domain initiated by epigenetic marking at the embryonic stem cell stage.* Mol Cell Biol, 2005. **25**(5): p. 1804-20.

150. Cao, R., et al., *Role of histone H3 lysine 27 methylation in Polycomb-group silencing.* Science, 2002. **298**(5595): p. 1039-43.

151. Wang, L., et al., *Hierarchical recruitment of polycomb group silencing complexes.* Mol Cell, 2004. **14**(5): p. 637-46.

152. Schwartz, Y.B. and V. Pirrotta, *Polycomb complexes and epigenetic states.* Curr Opin Cell Biol, 2008. **20**(3): p. 266-73.

153. Jorgensen, H.F., et al., *Stem cells primed for action: polycomb repressive complexes restrain the expression of lineage-specific regulators in embryonic stem cells.* Cell Cycle, 2006. **5**(13): p. 1411-4.

154. Mikkelsen, T.S., et al., *Genome-wide maps of chromatin state in pluripotent and lineage-committed cells.* Nature, 2007. **448**(7153): p. 553-60.

155. Zhang, Z., et al., *PRC2 complexes with JARID2, MTF2, and esPRC2p48 in ES cells to modulate ES cell pluripotency and somatic cell reprogramming.* Stem Cells, 2011. **29**(2): p. 229-40.

156. Richly, H., L. Aloia, and L. Di Croce, *Roles of the Polycomb group proteins in stem cells and cancer.* Cell Death Dis, 2011. **2**: p. e204.

157. O'Carroll, D., et al., *The polycomb-group gene Ezh2 is required for early mouse development.* Mol Cell Biol, 2001. **21**(13): p. 4330-6.

158. Lee, T.I., et al., *Control of developmental regulators by Polycomb in human embryonic stem cells.* Cell, 2006. **125**(2): p. 301-13.

159. Landeira, D., et al., *Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators.* Nat Cell Biol, 2010. **12**(6): p. 618-24.

160. Stock, J.K., et al., *Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells.* Nat Cell Biol, 2007. **9**(12): p. 1428-35.

161. de Almeida, S.F., et al., *Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36.* Nat Struct Mol Biol, 2011. **18**(9): p. 977-83.

162. Seila, A.C., et al., *Divergent transcription from active promoters.* Science, 2008. **322**(5909): p. 1849-51.

163. Steger, D.J., et al., *DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells.* Mol Cell Biol, 2008. **28**(8): p. 2825-39.

164. Corden, J.L., *Tails of RNA polymerase II.* Trends Biochem Sci, 1990. **15**(10): p. 383-7.

165. Phatnani, H.P. and A.L. Greenleaf, *Phosphorylation and functions of the RNA polymerase II CTD.* Genes Dev, 2006. **20**(21): p. 2922-36.

166. Ng, H.H., et al., *Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity.* Mol Cell, 2003. **11**(3): p. 709-19.

167. Brookes, E. and A. Pombo, *Modifications of RNA polymerase II are pivotal in regulating gene expression states.* EMBO Rep, 2009. **10**(11): p. 1213-9.

168. Kininis, M., et al., *Postrecruitment regulation of RNA polymerase II directs rapid signaling responses at the promoters of estrogen target genes.* Mol Cell Biol, 2009. **29**(5): p. 1123-33.

169. Core, L.J. and J.T. Lis, *Transcription regulation through promoter-proximal pausing of RNA polymerase II.* Science, 2008. **319**(5871): p. 1791-2.

170. Nechaev, S. and K. Adelman, *Promoter-proximal Pol II: when stalling speeds things up.* Cell Cycle, 2008. **7**(11): p. 1539-44.

171. Hanyu-Nakamura, K., et al., *Drosophila Pgc protein inhibits P-TEFb recruitment to chromatin in primordial germ cells.* Nature, 2008. **451**(7179): p. 730-3.

172. Kizer, K.O., et al., *A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation.* Mol Cell Biol, 2005. **25**(8): p. 3305-16.

173. Krogan, N.J., et al., *The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation.* Mol Cell, 2003. **11**(3): p. 721-9.

174. Kanhere, A., et al., *Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2.* Mol Cell, 2010. **38**(5): p. 675-88.

175. Santos-Rosa, H., et al., *Active genes are tri-methylated at K4 of histone H3.* Nature, 2002. **419**(6905): p. 407-11.

176. Levine, M., *Paused RNA polymerase II as a developmental checkpoint.* Cell, 2011. **145**(4): p. 502-11.

177. Bernstein, E., et al., *Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin.* Mol Cell Biol, 2006. **26**(7): p. 2560-9.

178. Rinn, J.L., et al., *Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs.* Cell, 2007. **129**(7): p. 1311-23.

179. Guenther, M.G. and R.A. Young, *Transcription. Repressive transcription.* Science, 2010. **329**(5988): p. 150-1.

180. Hawkins, R.D., et al., *Dynamic chromatin states in human ES cells reveal potential regulatory sequences and genes involved in pluripotency.* Cell Res, 2011. **21**(10): p. 1393-409.

181. Heintzman, N.D., et al., *Histone modifications at human enhancers reflect global cell-type-specific gene expression.* Nature, 2009. **459**(7243): p. 108-12.

182. Creyghton, M.P., et al., *Histone H3K27ac separates active from poised enhancers and predicts developmental state.* Proc Natl Acad Sci U S A, 2010. **107**(50): p. 21931-6.

183. Visel, A., et al., *ChIP-seq accurately predicts tissue-specific activity of enhancers.* Nature, 2009. **457**(7231): p. 854-8.

184. Rada-Iglesias, A., et al., *A unique chromatin signature uncovers early developmental enhancers in humans.* Nature, 2011. **470**(7333): p. 279-83.

185. Kim, T.K., et al., *Widespread transcription at neuronal activity-regulated enhancers.* Nature, 2010. **465**(7295): p. 182-7.

186. Pokholok, D.K., et al., *Genome-wide map of nucleosome acetylation and methylation in yeast.* Cell, 2005. **122**(4): p. 517-27.

187. Orford, K., et al., *Differential H3K4 methylation identifies developmentally poised hematopoietic genes.* Dev Cell, 2008. **14**(5): p. 798-809.

188. Glick, B., *Historical perspective: the bursa of Fabricius and its influence on B-cell development, past and present.* Vet Immunol Immunopathol, 1991. **30**(1): p. 3-12.

189. Murre, C., *Defining the pathways of early adult hematopoiesis.* Cell Stem Cell, 2007. **1**(4): p. 357-8.

190. Pronk, C.J., et al., *Elucidation of the phenotypic, functional, and molecular topography of a myeloerythroid progenitor cell hierarchy.* Cell Stem Cell, 2007. **1**(4): p. 428-42.

191. Benz, C., et al., *The stream of precursors that colonizes the thymus proceeds selectively through the early T lineage precursor stage of T cell development.* J Exp Med, 2008. **205**(5): p. 1187-99.

192. Allman, D., et al., *Thymopoiesis independent of common lymphoid progenitors.* Nat Immunol, 2003. **4**(2): p. 168-74.

193. Gautreau, L., et al., *Gene coexpression analysis in single cells indicates lymphomyeloid copriming in short-term hematopoietic stem cells and multipotent progenitors.* J Immunol, 2010. **184**(9): p. 4907-17.

194. Murre, C., *Developmental trajectories in early hematopoiesis.* Genes Dev, 2009. **23**(20): p. 2366-70.

195. Godfrey, D.I. and A. Zlotnik, *Control points in early T-cell development.* Immunol Today, 1993. **14**(11): p. 547-53.

196. Hayday, A.C. and D.J. Pennington, *Key factors in the organized chaos of early T cell development.* Nat Immunol, 2007. **8**(2): p. 137-44.

197. Nambiar, M., V. Kari, and S.C. Raghavan, *Chromosomal translocations in cancer.* Biochim Biophys Acta, 2008. **1786**(2): p. 139-52.

198. Tsuji, H., et al., *Involvement of illegitimate V(D)J recombination or microhomology-mediated nonhomologous end-joining in the formation of intragenic deletions of the Notch1 gene in mouse thymic lymphomas.* Cancer Res, 2004. **64**(24): p. 8882-90.

199. Roth, D.B., et al., *V(D)J recombination: broken DNA molecules with covalently sealed (hairpin) coding ends in scid mouse thymocytes.* Cell, 1992. **70**(6): p. 983-91.

200. Ma, Y., et al., *Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination.* Cell, 2002. **108**(6): p. 781-94.

201. Rooney, S., J. Chaudhuri, and F.W. Alt, *The role of the non-homologous end-joining pathway in lymphocyte development.* Immunol Rev, 2004. **200**: p. 115-31.

202. Lewis, S.M., *P nucleotide insertions and the resolution of hairpin DNA structures in mammalian cells.* Proc Natl Acad Sci U S A, 1994. **91**(4): p. 1332-6.

203. Wysoczanska, B., *[T-cell receptor rearrangement excision circles (TRECs) as a marker of thymic function].* Postepy Hig Med Dosw (Online), 2008. **62**: p. 708-24.

204. Rogers, J., et al., *Two mRNAs with different 3' ends encode membrane-bound and secreted forms of immunoglobulin mu chain.* Cell, 1980. **20**(2): p. 303-12.

205. Glusman, G., et al., *Comparative genomics of the human and mouse T cell receptor loci.* Immunity, 2001. **15**(3): p. 337-49.

206. Bluestone, J.A., et al., *Structure and specificity of T cell receptor gamma/delta on major histocompatibility complex antigen-specific CD3+, CD4-, CD8- T lymphocytes.* J Exp Med, 1988. **168**(5): p. 1899-916.

207. Kurosawa, Y., et al., *Identification of D segments of immunoglobulin heavy-chain genes and their rearrangement in T lymphocytes.* Nature, 1981. **290**(5807): p. 565-70.

208. Perlot, T. and F.W. Alt, *Cis-regulatory elements and epigenetic changes control genomic rearrangements of the IgH locus.* Adv Immunol, 2008. **99**: p. 1-32.

209. Lin, W.C. and S. Desiderio, *V(D)J recombination and the cell cycle.* Immunol Today, 1995. **16**(6): p. 279-89.

210. Schlissel, M., et al., *Double-strand signal sequence breaks in V(D)J recombination are blunt, 5'-phosphorylated, RAG-dependent, and cell cycle regulated.* Genes Dev, 1993. **7**(12B): p. 2520-32.

211. Cobb, R.M., et al., *Accessibility control of V(D)J recombination.* Adv Immunol, 2006. **91**: p. 45-109.

212. Loffert, D., et al., *Surrogate light chain expression is required to establish immunoglobulin heavy chain allelic exclusion during early B cell development.* Immunity, 1996. **4**(2): p. 133-44.

213. McBlane, J.F., et al., *Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps.* Cell, 1995. **83**(3): p. 387-95.

214. Tonegawa, S., *Somatic generation of antibody diversity.* Nature, 1983. **302**(5909): p. 575-81.

215. Jones, J.M. and M. Gellert, *Ordered assembly of the V(D)J synaptic complex ensures accurate recombination.* EMBO J, 2002. **21**(15): p. 4162-71.

216. Schatz, D.G. and Y. Ji, *Recombination centres and the orchestration of V(D)J recombination.* Nat Rev Immunol, 2011. **11**(4): p. 251-63.

217. Ji, Y., et al., *The in vivo pattern of binding of RAG1 and RAG2 to antigen receptor loci.* Cell, 2010. **141**(3): p. 419-31.

218. Tillman, R.E., et al., *Regulation of T-cell receptor beta-chain gene assembly by recombination signals: the beyond 12/23 restriction.* Immunol Rev, 2004. **200**: p. 36-43.

219. Wu, C., et al., *Restriction of endogenous T cell antigen receptor beta rearrangements to Vbeta14 through selective recombination signal sequence modifications.* Proc Natl Acad Sci U S A, 2007. **104**(10): p. 4002-7.

220. Hughes, M.M., et al., *The B12/23 restriction is critically dependent on recombination signal nonamer and spacer sequences.* J Immunol, 2003. **171**(12): p. 6604-10.

221. Fanning, L., et al., *Mouse RSS spacer sequences affect the rate of V(D)J recombination.* Immunogenetics, 1996. **44**(2): p. 146-50.

222. Cowell, L.G., et al., *Identification and utilization of arbitrary correlations in models of recombination signal sequences.* Genome Biol, 2002. **3**(12): p. RESEARCH0072.

223. Marculescu, R., et al., *Recombinase, chromosomal translocations and lymphoid neoplasia: targeting mistakes and repair failures.* DNA Repair (Amst), 2006. **5**(9-10): p. 1246-58.

224. Swanson, P.C., *The bounty of RAGs: recombination signal complexes and reaction outcomes.* Immunol Rev, 2004. **200**: p. 90-114.

225. Kirch, S.A., P. Sudarsanam, and M.A. Oettinger, *Regions of RAG1 protein critical for V(D)J recombination.* Eur J Immunol, 1996. **26**(4): p. 886-91.

226. Yurchenko, V., Z. Xue, and M. Sadofsky, *The RAG1 N-terminal domain is an E3 ubiquitin ligase.* Genes Dev, 2003. **17**(5): p. 581-5.

227. Oettinger, M.A., et al., *RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination.* Science, 1990. **248**(4962): p. 1517-23.

228. van Gent, D.C., et al., *Stimulation of V(D)J cleavage by high mobility group proteins.* EMBO J, 1997. **16**(10): p. 2665-70.

229. Yu, W., et al., *Coordinate regulation of RAG1 and RAG2 by cell type-specific DNA elements 5' of RAG2.* Science, 1999. **285**(5430): p. 1080-4.

230. Monroe, R.J., et al., *RAG2:GFP knockin mice reveal novel aspects of RAG2 expression in primary and peripheral lymphoid tissues.* Immunity, 1999. **11**(2): p. 201-12.

231. Hesslein, D.G. and D.G. Schatz, *Factors and forces controlling V(D)J recombination.* Adv Immunol, 2001. **78**: p. 169-232.

232. Nagaoka, H., W. Yu, and M.C. Nussenzweig, *Regulation of RAG expression in developing lymphocytes.* Curr Opin Immunol, 2000. **12**(2): p. 187-90.

233. Grawunder, U., et al., *Down-regulation of RAG1 and RAG2 gene expression in preB cells after functional immunoglobulin heavy chain rearrangement.* Immunity, 1995. **3**(5): p. 601-8.

234. Turka, L.A., et al., *Thymocyte expression of RAG-1 and RAG-2: termination by T cell receptor cross-linking.* Science, 1991. **253**(5021): p. 778-81.

235. Hertz, M., et al., *V(D)J recombinase induction in splenic B lymphocytes is inhibited by antigen-receptor signalling.* Nature, 1998. **394**(6690): p. 292-5.

236. Hertz, M. and D. Nemazee, *BCR ligation induces receptor editing in IgM+IgD- bone marrow B cells in vitro.* Immunity, 1997. **6**(4): p. 429-36.

237. Melamed, D., et al., *Developmental regulation of B lymphocyte immune tolerance compartmentalizes clonal selection from receptor selection.* Cell, 1998. **92**(2): p. 173-82.

238. Dudley, D.D., et al., *Mechanism and control of V(D)J recombination versus class switch recombination: similarities and differences.* Adv Immunol, 2005. **86**: p. 43-112.

239. Jackson, K.J., et al., *Exonuclease activity and P nucleotide addition in the generation of the expressed immunoglobulin repertoire.* BMC Immunol, 2004. **5**: p. 19.

240. Rothkamm, K., et al., *Pathways of DNA double-strand break repair during the mammalian cell cycle.* Mol Cell Biol, 2003. **23**(16): p. 5706-15.

241. Smith, G.C. and S.P. Jackson, *The DNA-dependent protein kinase.* Genes Dev, 1999. **13**(8): p. 916-34.

242. Gao, Y., et al., *A critical role for DNA end-joining proteins in both lymphogenesis and neurogenesis.* Cell, 1998. **95**(7): p. 891-902.

243. Karanjawala, Z.E., et al., *The embryonic lethality in DNA ligase IV-deficient mice is rescued by deletion of Ku: implications for unifying the heterogeneous phenotypes of NHEJ mutants.* DNA Repair (Amst), 2002. **1**(12): p. 1017-26.

244. Li, G., et al., *Lymphocyte-specific compensation for XLF/cernunnos end-joining functions in V(D)J recombination.* Mol Cell, 2008. **31**(5): p. 631-40.

245. Lieber, M.R., *The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway.* Annu Rev Biochem, 2004. **79**: p. 181-211.

246. Ahnesorg, P., P. Smith, and S.P. Jackson, *XLF interacts with the XRCC4-DNA ligase IV complex to promote DNA nonhomologous end-joining.* Cell, 2006. **124**(2): p. 301-13.

247. Zhu, C., et al., *Unrepaired DNA breaks in p53-deficient cells lead to oncogenic gene amplification subsequent to translocations.* Cell, 2002. **109**(7): p. 811-21.

248. Shih, H.Y. and M.S. Krangel, *Distinct contracted conformations of the Tcra/Tcrd locus during Tcra and Tcrd recombination.* J Exp Med, 2010. **207**(9): p. 1835-41.

249. Krangel, M.S., *Mechanics of T cell receptor gene rearrangement.* Curr Opin Immunol, 2009. **21**(2): p. 133-9.

250. Yancopoulos, G.D. and F.W. Alt, *Developmentally controlled and tissue-specific expression of unrearranged VH gene segments.* Cell, 1985. **40**(2): p. 271-81.

251. Kwon, J., et al., *Histone acetylation and hSWI/SNF remodeling act in concert to stimulate V(D)J cleavage of nucleosomal DNA.* Mol Cell, 2000. **6**(5): p. 1037-48.

252. Stanhope-Baker, P., et al., *Cell type-specific chromatin structure determines the targeting of V(D)J recombinase activity in vitro.* Cell, 1996. **85**(6): p. 887-97.

253. Osipovich, O. and E.M. Oltz, *Regulation of antigen receptor gene assembly by genetic-epigenetic crosstalk.* Semin Immunol, 2010. **22**(6): p. 313-22.

254. Abarrategui, I. and M.S. Krangel, *Regulation of T cell receptor-alpha gene recombination by transcription.* Nat Immunol, 2006. **7**(10): p. 1109-15.

255. Bolland, D.J., et al., *Antisense intergenic transcription in V(D)J recombination.* Nat Immunol, 2004. **5**(6): p. 630-7.

256. Pasqual, N., et al., *Quantitative and qualitative changes in V-J alpha rearrangements during mouse thymocytes differentiation: implication for a limited T cell receptor alpha chain repertoire.* J Exp Med, 2002. **196**(9): p. 1163-73.

257. Shih, H.Y., et al., *Tcra gene recombination is supported by a Tcra enhancer- and CTCF-dependent chromatin hub.* Proc Natl Acad Sci U S A, 2012. **109**(50): p. E3493-502.

258. Nolis, I.K., et al., *Transcription factors mediate long-range enhancer-promoter interactions.* Proc Natl Acad Sci U S A, 2009. **106**(48): p. 20222-7.

259. Szutorisz, H., N. Dillon, and L. Tora, *The role of enhancers as centres for general transcription factor recruitment.* Trends Biochem Sci, 2005. **30**(11): p. 593-9.

260. Serwe, M. and F. Sablitzky, *V(D)J recombination in B cells is impaired but not blocked by targeted deletion of the immunoglobulin heavy chain intron enhancer.* EMBO J, 1993. **12**(6): p. 2321-7.

261. Oltz, E.M., et al., *A V(D)J recombinase-inducible B-cell line: role of transcriptional enhancer elements in directing V(D)J recombination.* Mol Cell Biol, 1993. **13**(10): p. 6223-30.

262. Takeda, S., et al., *Deletion of the immunoglobulin kappa chain intron enhancer abolishes kappa chain gene rearrangement in cis but not lambda chain gene rearrangement in trans.* EMBO J, 1993. **12**(6): p. 2329-36.

263. Sleckman, B.P., J.R. Gorman, and F.W. Alt, *Accessibility control of antigen-receptor variable-region gene assembly: role of cis-acting elements.* Annu Rev Immunol, 1996. **14**: p. 459-81.

264. Sleckman, B.P., et al., *Function of the TCR alpha enhancer in alphabeta and gammadelta T cells.* Immunity, 1997. **7**(4): p. 505-15.

265. Bories, J.C., et al., *Gene-targeted deletion and replacement mutations of the T-cell receptor beta-chain enhancer: the role of enhancer elements in controlling V(D)J recombination accessibility.* Proc Natl Acad Sci U S A, 1996. **93**(15): p. 7871-6.

266. Redondo, J.M., et al., *A T cell-specific transcriptional enhancer within the human T cell receptor delta locus.* Science, 1990. **247**(4947): p. 1225-9.

267. Conlon, T.M. and K.B. Meyer, *The chicken Ig light chain 3'-enhancer is essential for gene expression and regulates gene conversion via the transcription factor E2A.* Eur J Immunol, 2006. **36**(1): p. 139-48.

268. Xiong, N., C. Kang, and D.H. Raulet, *Redundant and unique roles of two enhancer elements in the TCRgamma locus in gene regulation and gammadelta T cell development.* Immunity, 2002. **16**(3): p. 453-63.

269. Mathieu, N., et al., *Chromatin remodeling by the T cell receptor (TCR)-beta gene enhancer during early T cell development: Implications for the control of TCR-beta locus recombination.* J Exp Med, 2000. **192**(5): p. 625-36.

270. Bassing, C.H., et al., *T cell receptor (TCR) alpha/delta locus enhancer identity and position are critical for the assembly of TCR delta and alpha variable region genes.* Proc Natl Acad Sci U S A, 2003. **100**(5): p. 2598-603.

271. Bouvier, G., et al., *Deletion of the mouse T-cell receptor beta gene enhancer blocks alphabeta T-cell development.* Proc Natl Acad Sci U S A, 1996. **93**(15): p. 7877-81.

272. Norris, H.H., et al., *TCRbeta enhancer activation in early and late lymphoid progenitors.* Cell Immunol, 2007. **247**(2): p. 59-71.

273. Monroe, R.J., et al., *RAG2 is regulated differentially in B and T cells by elements 5' of the promoter.* Proc Natl Acad Sci U S A, 1999. **96**(22): p. 12713-8.

274. Wei, X.C., et al., *Characterization of chromatin structure and enhancer elements for murine recombination activating gene-2.* J Immunol, 2002. **169**(2): p. 873-81.

275. Borghesi, L. and R.M. Gerstein, *Developmental separation of V(D)J recombinase expression and initiation of IgH recombination in B lineage progenitors in vivo.* J Exp Med, 2004. **199**(4): p. 483-9.

276. Borghesi, L., et al., *B lineage-specific regulation of V(D)J recombinase activity is established in common lymphoid progenitors.* J Exp Med, 2004. **199**(4): p. 491-502.

277. Rothenberg, E.V. and T. Taghon, *Molecular genetics of T cell development.* Annu Rev Immunol, 2005. **23**: p. 601-49.

278. Hernandez-Munain, C., B.P. Sleckman, and M.S. Krangel, *A developmental switch from TCR delta enhancer to TCR alpha enhancer function during thymocyte maturation.* Immunity, 1999. **10**(6): p. 723-33.

279. Sikes, M.L., et al., *Regulation of V(D)J recombination: a dominant role for promoter positioning in gene segment accessibility.* Proc Natl Acad Sci U S A, 2002. **99**(19): p. 12309-14.

280. Baker, J.E., D. Cado, and D.H. Raulet, *Developmentally programmed rearrangement of T cell receptor Vgamma genes is controlled by sequences immediately upstream of the Vgamma genes.* Immunity, 1998. **9**(2): p. 159-68.

281. Sikes, M.L., C.C. Suarez, and E.M. Oltz, *Regulation of V(D)J recombination by transcriptional promoters.* Mol Cell Biol, 1999. **19**(4): p. 2773-81.

282. Whitehurst, C.E., S. Chattopadhyay, and J. Chen, *Control of V(D)J recombinational accessibility of the D beta 1 gene segment at the TCR beta locus by a germline promoter.* Immunity, 1999. **10**(3): p. 313-22.

283. McMillan, R.E. and M.L. Sikes, *Differential activation of dual promoters alters Dbeta2 germline transcription during thymocyte development.* J Immunol, 2008. **180**(5): p. 3218-28.

284. Stone, J.L., et al., *DNA double-strand breaks relieve USF-mediated repression of Dbeta2 germline transcription in developing thymocytes.* J Immunol, 2012. **188**(5): p. 2266-75.

285. McMillan, R.E. and M.L. Sikes, *Promoter activity 5' of Dbeta2 is coordinated by E47, Runx1, and GATA-3.* Mol Immunol, 2009. **46**(15): p. 3009-17.

286. Hawwari, A. and M.S. Krangel, *Regulation of TCR delta and alpha repertoires by local and long-distance control of variable gene segment chromatin structure.* J Exp Med, 2005. **202**(4): p. 467-72.

287. Hawwari, A., C. Bock, and M.S. Krangel, *Regulation of T cell receptor alpha gene assembly by a complex hierarchy of germline Jalpha promoters.* Nat Immunol, 2005. **6**(5): p. 481-9.

288. Abarrategui, I. and M.S. Krangel, *Noncoding transcription controls downstream promoters to regulate T-cell receptor alpha recombination.* EMBO J, 2007. **26**(20): p. 4380-90.

289. Hawwari, A. and M.S. Krangel, *Role for rearranged variable gene segments in directing secondary T cell receptor alpha recombination.* Proc Natl Acad Sci U S A, 2007. **104**(3): p. 903-7.

290. Ji, Y., et al., *Promoters, enhancers, and transcription target RAG1 binding during V(D)J recombination.* J Exp Med, 2010. **207**(13): p. 2809-16.

291. Xiong, N., et al., *Gene placement and competition control T cell receptor gamma variable region gene rearrangement.* J Exp Med, 2008. **205**(4): p. 929-38.

292. Ren, X., et al., *Direct interactions of OCA-B and TFII-I regulate immunoglobulin heavy-chain gene transcription by facilitating enhancer-promoter communication.* Mol Cell, 2011. **42**(3): p. 342-55.

293. Afshar, R., et al., *Regulation of IgH gene assembly: role of the intronic enhancer and 5'DQ52 region in targeting DHJH recombination.* J Immunol, 2006. **176**(4): p. 2439-47.

294. Martin, D.J. and B.G. van Ness, *Initiation and processing of two kappa immunoglobulin germ line transcripts in mouse B cells.* Mol Cell Biol, 1990. **10**(5): p. 1950-8.

295. Stiernholm, N.B. and N.L. Berinstein, *A mutated promoter of a human Ig V lambda gene segment is associated with reduced germ-line transcription and a low frequency of rearrangement.* J Immunol, 1995. **154**(4): p. 1748-61.

296. Krangel, M.S., *T cell development: better living through chromatin.* Nat Immunol, 2007. **8**(7): p. 687-94.

297. Miracle, A.L., et al., *Complex expression patterns of lymphocyte-specific genes during the development of cartilaginous fish implicate unique lymphoid tissues in generating an immune repertoire.* Int Immunol, 2001. **13**(4): p. 567-80.

298. Hohman, V.S., et al., *Genomic clone for sandbar shark lambda light chain: generation of diversity in the absence of gene rearrangement.* Proc Natl Acad Sci U S A, 1993. **90**(21): p. 9882-6.

299. Eason, D.D. and G.W. Litman, *Haplotype exclusion: the unique case presented by multiple immunoglobulin gene loci in cartilaginous fish.* Semin Immunol, 2002. **14**(3): p. 145-52; discussion 220.

CHAPTER 2

CTCF Constancy and Developmental Progression at the TCRα locus

ABSTRACT

To maintain pluripotency, cells of the early embryo must remain both undifferentiated and pluripotent. To that end, differentiation- or tissue-specific genes must remain silent but capable of activation. Initial investigations into the so-called priming of tissue-specific genes in early embryonic cells suggest critical roles for pluripotency factors and epigenetic modifications, particularly hypomethylation of DNA within the transcriptional regulatory regions of such genes. Though not a pluripotency factor itself, the multi-faceted transcription factor CTCF is vital for proper differentiation and is instrumental in function of many pluripotency factors. Genome-wide, CTCF binding strongly correlates with hypomethylation of its binding site. However, the functional relationship between CTCF and modulations in DNA methylation during development remains unknown. The T early alpha promoter (TEA) is a tissue-specific transcriptional regulatory element required for expression of the T cell receptor alpha *(Tcra)* gene and proper thymopoiesis. Binding of CTCF at TEA is essential for promoter activity in developing thymocytes, coordinating the formation of a chromatin loop that juxtaposes TEA with the *Tcra* enhancer ~100 kb downstream. To elucidate the embryonic mechanisms that prime TEA, we investigated DNA methylation and CTCF binding in lineage-restricted and pluripotent cells. While TEA activity and TEA:Eα looping appears to be restricted to developing thymocytes, we found that CTCF is constitutively bound to TEA, being present in every cell type assessed including pluripotent embryonic stem cells (ESCs), and that its presence directly correlated with the promoter's hypomethylation despite constitutive methylation of flanking DNA. Disruption of CTCF binding by RNAi-mediated knockdown or mono-allelic deletion of TEA led to hypermethylation of the promoter region on both mutated and wildtype *Tcra* alleles.

Given the generally repressive effect of DNA methylation on transcription, our findings suggest a novel role for CTCF in embryonic gene priming as a constitutive protector against promoter methylation.

2.1 Introduction

Embryonic stem cells (ESCs) are self-renewing and undifferentiated, representing the ground state of development, and at the same time, are pluripotent—competent to give rise to the body's somatic cell lineages. The process of differentiation is one of progressive lineage restriction: as a cell develops along one particular lineage, genes that encode products specific to another lineage are epigenetically silenced [1]. Consequently, to sustain pluripotency the ESC genome must be maintained in a hyperdynamic and relatively open conformation characterized by generally reduced DNA methylation [2], widely disseminated binding of RNA polymerase II (RNAP2) [3, 4], and high levels of acetylation on lysine 9 of histone H3 (H3K9ac) [5-8]. The functional significance of this epigenetic openness is easily seen when cloning by nuclear transfer, where use of nuclei from G_1 ESCs yields a cloning success rate that is 10- to 20-fold higher than that obtained from any type of somatic donor cell yet tested [9]. However, the mechanisms that regulate the ESC epigenetic landscape, and by extension, the patterning and success of differentiation, are only just coming to light.

One of the earliest-acting mechanisms for generation and maintenance of the pluripotent state involves members of the *Polycomb* group (PcG) of transcriptional regulators, which interact to form chromatin-associated Polycomb Repressive Complexes (PRCs). PRCs are strongly affiliated with "bivalent" gene priming, identified by colocalization of histone marks H3K4me3 and H3K27me3 [5]. The major PRCs, known as PRC1 and PRC2 have been extensively characterized for their roles in

repressing transcriptional elongation [10]. Evidence suggests PRC2 is responsible for recruitment of PRC1 and deposition of the repressive histone modification H3K27me3 [11]. Although ESCs that lack PRC2 maintain pluripotency, they begin inappropriate expression of a variety of developmental regulators [10, 12]. PRC2 is made up of PcG proteins EZH2, EED, SUZ12 and RbAp46/48, which are required for its enzymatic activity *in vitro*, as well as additional proteins identified *in vivo* that function to enhance enzymatic activity and facilitate DNA targeting [13].

Beyond PRC-mediated repression, a large number of transcription factors have been implicated in maintaining the ESC properties of pluripotency and self-renewal. In particular, pluripotency is primarily regulated by Oct4, Klf4, Sox2 and Nanog [14, 15]. Indeed, forced expression of these four "pluripotency factors" underlies the establishment of induced pluripotent stem cells (or iPSCs) from terminally differentiated somatic cells [16]. Instrumental in proper functionality of these pluripotency factors is the CCCTC-binding factor (CTCF), a multifaceted scaffolding protein [17]. First identified as a transcriptional regulator of c-*Myc*, CTCF is the only known vertebrate insulator protein, acting to properly delineate boundaries between active and inactive regions of DNA and to isolate transcriptional control elements like promoters from activity at neighboring promoters or enhancers [18]. Widespread incorporation of methods to detect 3-D chromatin conformation *in vivo* have since suggested CTCF may serve as a "master regulator" of chromatin fiber organization and interaction [19]. However, because CTCF complexes with a wide variety of proteins through its relatively unstructured N and C-terminal tails, as well as with variety of DNA sequences through a promiscuous central zinc finger domain, it appears likely that its ever-expanding portfolio of actions reflects the array of complexes in which it is found [20]. Proper CTCF expression is vital for cell growth, survival, proliferation, imprinting, X-chromosome inactivation and

differentiation [21]. CTCF is critical to embryogenesis, and its elimination in mouse models results in peri-implantation embryonic lethality [22, 23]. CTCF is also implicated in bivalent priming, as it interacts with the PRC2 component SUZ12, and drives maternal allele-specific H3K27me3 deposition and transcriptional suppression [24].

DNA methylation patterns are carried by gametes prior to fertilization, and differentially methylated regions (DMRs) in paternal or maternal-origin DNA maintain essential imprinted developmental controls [25]. Following extensive demethylation in the pre-implantation embryo to largely erase the parental epigenetic patterns, a new, embryo-specific methylation pattern is established coincident with implantation [26]. As mentioned above, the embryonic genome is largely hypomethylated in ESCs, while differentiation correlates with progressive hypermethylation [8]. Conversely, successful reprogramming of terminally differentiated somatic cells during iPSC derivation correlates with widespread erasure of DNA methylation [27]. Though by no means uniform, DNA methylation, particularly at the promoter regions of genes, predominantly correlates with transcriptional repression [28]. As such, the dynamism of DNA methylation during development would suggest that deposition of methyl marks directly contributes to the epigenetic lineage restriction seen during differentiation. DNA methylation regulates gene expression through a variety of mechanisms, including interference with transcription factor-DNA interactions both by altering the transcription factor binding site and by recruiting methyl-CpG-binding proteins that block access to the transcription factor binding site. Conversely, DNA methylation has recently been implicated as an inhibitor of both PRC2 binding and subsequent H3K27me3 deposition [29], suggesting that at some gene targets, DNA methylation may be activating rather than repressive. DNA methylation may also be inhibitory to CTCF binding, as most sites occupied by CTCF are

unmethylated in the ESC genome [30, 31] and methylation impairs CTCF binding to reporter targets *in vitro* [31, 32]. This is not universal, however, as CTCF does bind to a low percentage of methylated sites and CTCF remains unbound from other sites despite an absence of methylation [19, 30, 32].

Development of lymphocytes from hematopoietic stem cells involves a series of lineage restrictions and the carefully orchestrated activation of lymphocyte-specific genes. Most dramatic of these are the activation and rearrangement of the antigen receptor genes at discrete stages of B and T cell development which are both dependent on and essential for continued differentiation [33, 34]. In the case of T cells, lineage commitment in DN thymocytes activates the *Tcrb, Tcrg,* and *Tcrd* genes, and progression to the DP stage requires signaling through the Pre-TCR complex that contains a newly expressed TCRβ protein [35, 36]. Only then is *Tcra* activated and rearranged in a manner timed precisely to limit the proportion of γδ T cells and maximize expansion of the αβ repertoire [36]. Like other antigen receptor genes, including the *IgH* gene, activation of *Tcra* in DP cells is dependent upon CTCF [37, 38]. Here the T early alpha promoter (TEA) positioned at the 5' end of the Jα array facilitates non-coding transcription of TEA-proximal Jα segments during T cell development, making them accessible to recombination. More distal Jαs are then targeted in secondary rearrangements as necessary to generate an in-frame joint. Their accessibility to recombinase following deletion of TEA extends primarily from the proximity of the promoter associated with the rearranged Vα segment, though germline promoters scattered through the Jα array show progressive serial activation in response to deletion of the upstream Jα promoters [39]. Proper function of the TEA promoter during T cell development has been well-characterized, and requires the involvement of CTCF and binding partner cohesin to generate an intrachromosomal

loop, bringing the promoter in close contact with a downstream enhancer [40]. Though TEA activity is strictly linked to DP development, initial studies found that even in DN cells the inactive TEA was already widely decorated by transcription factors [37] including CTCF and cohesin, and was free of DNA methylation, as was its enhancer, Eα. How then is TEA activity developmentally regulated?

To address the nature of TEA developmental regulation and the potential role of CTCF in guiding thymopoiesis, we sought to investigate TEA epigenetics in the pluripotent environment of the ESC. By reassessing genome-wide maps of CTCF binding, we show that TEA maintains CTCF occupation in the murine embryonic stem cell stage as well as in all other cell types along both hematopoietic and non-hematopoietic differentiation pathways. This constancy of CTCF binding does not correspond with local stability of either histone modification patterns or DNA intrachromosomal loops. Bisulfite profiling of TEA in ground state embryonic stem cells (established by culturing ESCs in 2i medium, a defined serum-free medium containing inhibitors of the GSK3β and Mek1/2 kinases [41]) reveals that an unmethylated TEA promoter sits within a narrow band of fully methylated DNA, a pattern we see maintained in DN and DP thymocytes. In contrast, more distal sequences including the *TRVD5* and BEAD-1 insulator upstream and Jα61 coding segment downstream are hypomethylated in ground-state ESCs, but become hypermethylated in DP thymocytes. Disruption of CTCF binding by global knockdown of CTCF in ground-state ESCs caused pronounced methylation along a 3.6 kb region that included TEA. Likewise, mono-allelic deletion of TEA from C57Bl6/129SvEv ESCs showed a pronounced increase in DNA methylation both upstream and downstream of TEA across both the mutated and wildtype alleles of ground-state ESCs. Together, these findings imply a role for CTCF in protecting the sequence that separates *Tcrd* from the Jα array, and that contains TEA and the insulator element BEAD-1, against potentially repressive

methylation prior to differentiation. Taken together, our studies of ground-state ESCs reveal a dynamic pattern of DNA methylation in the region of *Tcra/d* containing TEA that requires both CTCF and potential allelic cross-talk, and that developmental expansion of DNA methylation across the region remains specifically occluded from TEA. Given the generally repressive impact of promoter methylation, primordial epigenetic priming likely ensures that TEA remains ready to respond immediately to activating signals upon DP thymocyte transition, a strategy that may well be extensible to the regulation of other differentiation-specific genes across the genome.

2.2 Materials and Methods

2.2.1 Cell culture

All murine cell culture was conducted at 37 °C/5% CO_2 in RPMI 1640 medium supplemented with 10% fetal calf serum, 2 mM L-glutamine, 0.01% penicillin/streptomycin, and 50 µM β-mercaptoethanol. EF-1 embryonic stem cells were feeder-independent and supplemented with 1X ESGRO (Millipore) to prevent differentiation. To establish a ground state of naïve pluripotency, cells were cultured at 37°C in 2i medium (StemDiff NCell N2B27 supplemented with 1X ESGRO, 3µM CHIR99021 and 1 µM PD0325901 (both Selleck Chemicals)) for four-six passages, then returned to standard culture condition for two passages of expansion before analysis.

2.2.2 Bioinformatic analyses

Raw CTCF ChIP-seq, MeDIP and ChIA-PET datasets were obtained from the NCBI Gene Expression Omnibus (GEO). Accession numbers are as follows: ES Cell – GSM747536, DN Thymocytes – GSM1023416, DP Thymocytes – GSM1023418, pro-B – GSM1023421, MEF – GSM722663, Cerebellum – GSM722663, MEF MeDIP – GSM678696-GSM678698, ES cell MeDIP – GSM678684-GSM678686, ESC ChIA-PET – GSM699169-699171. When necessary, file conversion and manipulation were performed using publicly available tools on Galaxy (http://galaxyproject.org/) [42-44]. Visualization was performed using the UCSC Genome Browser and Trackster [45, 46] (http://genome.ucsc.edu/).

2.2.3 Chromatin Preparation

Approximately 10^7 cells were resuspended in 10 mL of RPMI containing 10% FBS (vol/vol) and cross-linked for 5 minutes with shaking through addition of formaldehyde to 1% final concentration. Cross-linking was stopped through addition of glycine to 125 µM and cells were

incubated at RT with shaking for an additional 5 minutes. After centrifugation (4° C, 1100 RPM, 4 minutes) and aspiration, the cell pellet was washed once in 10 mL PBS, and following a repeat centrifugation and aspiration was resuspended in 1 mL Chromatin Lysis Buffer (50 mM Tris-HCl pH 7.5, 1.5 M NaCl, 10 mM EDTA, 5 mM EGTA, 10% Triton X-100, 1% SDS, 1 mM PMSF, 1X Protease Inhibitor Cocktail (PIC, Sigma)) and incubated at 4°C for 30 minutes to lyse cells. The lysate was then subjected to a low-speed spin (4°C 500 RPM, 5 minutes) and after a gentle aspiration was resuspended in 135 µL fresh Chromatin Lysis Buffer. Sonication was conducted using a Covaris S2 machine within the NCSU Genomic Sciences Laboratory (GSL) with continuous degassing, "Frequency Sweeping" power mode. Sonication was performed using a pulsing 200bp shearing preset (SonoLAB S-Series software v2.55) for 8 minutes, optimized to yield chromatin fragments predominantly of 200 to 600 bp length. After sonication was complete, the chromatin was diluted with Chromatin Lysis Buffer to 10^5 cell equivalents per 10 µL, aliquotted and stored at -80^{o}C prior to chromatin immunoprecipitation. Shearing efficiency was characterized using a Bioanalyzer (Agilent).

2.2.4 Chromatin Immunoprecipitation

For one reaction: 5 µl each Protein A- and Protein G-conjugated paramagnetic dynabeads (Dynal) were captured on the side of the tube with a bar magnet, and washed twice in RIPA-DNA (RD) buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1% Triton X-100, 0.1% SDS, 2.5 mg/ml denatured sheared salmon sperm DNA) and were then resuspended in 100 µl of RD buffer supplemented with 1 mM PMSF and 1X PIC. After addition of ~1 µg antibody, the bead mixture was conjugated for 1 hour at 4°C, then washed three times in 100 µl RD buffer to remove unconjugated antibody. The antibody-bead complex was then resuspended in 90 µL RD buffer with 1 mM PMSF and 1X PIC and mixed with 10 µL (10^5 cell equivalents) of sheared chromatin and

incubated at 4°C overnight with rotation. To remove unbound chromatin, the bead complex was washed eight times: twice each with RIPA buffer (10 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1% Triton X-100, 0.1% SDS), High Salt Wash (500 mM NaCl, 20 mM Tris-HCl pH 8.0, 2 mM EDTA, 0.1% SDS, 1% Triton X-100), LiCl Wash (250 mM LiCl, 10 mM Tris-HCl pH 8.0, 1 mM EDTA, 1% NP-40), and TE (10 mM Tris-HCl pH 8.0, 1 mM EDTA). The chromatin-antibody complexes were eluted from the beads by incubation in 100 µL of 0.1 M $NaHCO_3$/1% SDS with vortexing for 15 minutes at room temperature. NaCl was added to the eluate (100 mM final), and cross-links were reversed by 15 minute incubation at 95°C. Proteinase K (1 mg final) was added, and protein digested for 60 minutes at 37°C. Protein digestion was stopped with PMSF (2 mM final), and DNA was purified using the Qiaquick PCR purification system (Qiagen). DNA was resuspended in 100 µl of H_2O and 3 µl of ChIP DNA was used per real-time reaction. Antibodies used: H3K9ac - Millipore (ABE18), H3K4me2 - Abcam (ab32356), H3K4me3 - Active Motif (39139) and Abcam (ab12209), H3K27me3 - Abcam (ab6002), CTCF - Millipore (07-729), IgG - Santa Cruz (sc-2027). QPCR was conducted on a MyIQ thermal cycler (Bio-RAD) using 1X SensiMix *Plus* (BioLine) in triplicate reactions. Primers and annealing temperatures are as indicated (Table S1).

2.2.5 Chromosome Conformation Capture (3C)

3C assays were performed by the Krangel laboratory as described previously [37]. In brief, 10^6 cells were cross-linked in 8 mL of RPMI containing 10% FBS and 2% paraformaldehyde (both v/v) for 10 min at 4°C, and cross-linking was then stopped by addition of glycine to 0.125 M. Following a single PBS wash cells were lysed in 10 mM Tris-HCl pH 8.0, 10 mM NaCl and 0.2% (v/v) NP-40 for 10 min at 4°C. Nuclei were pelleted by centrifugation and resuspended in 1.1X NEB Buffer 2 (New England BioLabs) overnight with 0.3% (w/v) SDS. After a 1 h 37°C incubation, Triton-X was added to

2% final concentration (v/v) and the mixture was returned to 37°C to neutralize SDS. Chromatin was then digested through addition of 200 U of restriction enzyme HindIII (NEB) at 37°C, followed by a second addition of 200 U HindIII for an additional 4 h (Final digestion efficiency ~97%). Digestion was stopped by SDS addition to 0.8% (w/v) and heat inactivation at 68°C for 10 min. Chromatin was purified thorough centrifugation (10°C, 35K RPM, 16 h) through 8 M urea using a Beckman SW40Ti rotor. Purified chromatin was resuspended in buffer TM (30 mM Tris-HCl (pH 7.4) and 10 mM $MgCl_2$) and was dialyzed against buffer TM to remove urea and then diluted through addition of 5 mL of buffer TM supplemented with DTT to 10 mM and ATP to 1 mM. Chromatin was then ligated by addition of 200 U of T4 DNA ligase (NEB) overnight at 16°C, followed by a second addition of 200 U of T4 ligase for 4 h. Ligated DNA was collected after overnight incubation at 65°C with 10 µg/mL Proteinase K and phenol/chloroform extraction followed by ethanol precipitation. Products were quantified by Taqman-based qPCR assays using a LightCycler 480 probe master kit and a LightCycler 480 Real-Time PCR system (Roche). The following PCR program was used: 95°C for 10 min, followed by 48 cycles of 95°C for 10 s and 65°C for 30 s. All reactions were run in duplicate. Primer sequences can be found in Table S1. Standard curves were generated using an unbiased pool of 3C products generated by digestion and relegation of equally mixed bacterial artificial chromosomes (BACs) bMQ-440L6 and bMQ-206H21, which span a significant fraction of the TCR α/δ locus, including the entire Jα region. Samples were normalized by setting the 3C signal between Eα and its 3' neighbor fragment to one.

2.2.6 Bisulfite Mapping

Bisulfite conversion of DNA was performed using the EZ DNA Methylation-Gold Kit (Zymo Research) according to the manufacturer's instructions. Briefly, 600 ng of genomic DNA was mixed

with proprietary conversion reagent, incubated at 98°C for 10 minutes to denature DNA and then reduced to 53°C for four hours to facilitate cytosine conversion. Converted DNA was isolated from the conversion mixture using a Zymo-Spin IC Column purification system, and specific 200-400 bp target regions were amplified via PCR using KAPA HiFi HotStart Uracil+ ReadyMix 2X (KAPA Biosystems) and the indicated primers and melt temps (Table S1), which were designed using the BiSearch primer-design algorithm. Amplification products were purified using the QiaQuick PCR Purification Kit (Qiagen) or gel-purified using the QiaQuick Gel Extraction Kit (Qiagen) based on necessity and according to manufacturer's instructions. Purified fragments were end-phosphorylated using T4 PNK (Fermentas) and blunt-end cloned into the EcoRV site of pBluescript SK (Strategene) using T4 DNA ligase (Fermentas). After ethanol precipitation in the presence of glycogen, ligations were transformed into Thunderbolt GC10 electrocompetent cells and screened by PCR. Individual cloned inserts were sequenced and manually scored for methylation.

2.2.7 Transduction

Lentiviral supernatants containing puromycin resistance and anti-CTCF shRNA were graciously provided by Dr. Zhe Liu, and construction details can be found in [47]. In brief, lentiviral particles were generated through transfection of HEK293T cells with a plasmid containing one of two different anti-CTCF shRNA sequences and puromycin resistance, along with third generation lentiviral packaging vectors, and mature virus was harvested and concentrated using the Fast-Trap Lentivirus Purification and Concentration Kit (Millipore). Using the provided supernatants, transduction of EF-1 ESCs was performed via spinoculation of 10^5 cells with concentrated virus at 1500 x RPM for 5 minutes in ES medium containing ESGRO and 12 μg/mL polybrene (Sigma). 24

hours after infection media was changed, and 36 hours post-infection cells were treated with 1 μg/mL puromycin. Resistant colonies were picked and screened for CTCF knockdown via RT-PCR.

2.2.8 RT-PCR

RNA was extracted using RNAzol RT (MRC) according to the manufacturer's instructions. 300 ng of RNA was reverse transcribed using Moloney murine leukemia virus (M-mulv, New England Biolabs) and oligo d(T) primers. The resultant cDNAs were amplified using intron-spanning primer pairs (Supplemental Table 1) by standard or quantitative PCR as described earlier for ChIP. The relative abundance of cDNAs was quantified following QPCR by ΔΔ threshold cycle (CT) normalization to matched untreated controls and standardized for loading variations by comparison with values obtained for β-actin. Determination of CTCF knockdown levels was performed according to Haimes et al. [48]. In brief, CTCF levels for each sample were normalized to β actin and this value was exponentially transformed to accurately reflect qPCR amplification efficiency. At this point, triplicate samples were averaged and the averages of knockdown clones were normalized to the average of two non-targeting control samples. This value, expressed as a percentage, reflects the effectiveness of CTCF knockdown.

2.3 Results

2.3.1 CTCF binding to the Jα region of *Tcra* is extremely localized and stable throughout development

Previous studies have demonstrated the importance of CTCF binding at the TEA promoter during the double positive (DP) stage of thymocyte development, where it facilitates V(D)J recombination of *Tcra* by tethering TEA to the downstream Eα enhancer in a long-range DNA loop [40]. CTCF is a ubiquitously expressed master regulator of chromatin structure and function from

early embryogenesis on, and has been implicated in the regulation of antigen receptor loci during lymphocyte development [49-51]. Previous studies had found that, while TEA and Eα are inactive prior to DP development, both elements are extensively decorated with transcription factors in earlier DN cells [37, 52]. To determine if the developmental timing of TEA activity might be linked to CTCF binding, we compiled an array of mouse CTCF ChIP-seq datasets from a variety of cell types including DN and DP thymocytes, pro-B cells, cerebellum, ESC, embryonic fibroblasts, and placenta from published sets archived at the Gene Expression Omnibus (GEO) and analyzed each for CTCF binding across a 150 kb stretch of *Tcra*. As expected, CTCF binding was strongly enriched in DP cells at Eα and the antiapoptotic *Defense against death (Dad1)* gene and to a somewhat lesser extent, at TEA (Fig. 1). Weaker CTCF peaks were also observed just downstream of TEA at sequence flanking the PJα49 promoter, which is known to activate following TEA deletion [39]. In earlier DN cells, CTCF binding was detected at TEA, Eα and *Dad1*, but not at Jα49. Despite CTCF, *Tcra* lacks TEA:Eα loops in DN cells, and TEA is transcriptionally inactive [37, 53].

We found that the CTCF peaks at TEA and Eα persisted in all tissues examined, including closely related pro-B cells, but also cerebellar neurons. These peaks were also present in ESCs. In the placenta, an extra-embryonic tissue derived from a different blastocyst compartment than the ESC progenitor, CTCF binding at Eα and Dad1 was retained, though binding at TEA remained undetermined due to lack of sufficient resolution (data not shown). Global genomic analyses have since revealed that around a third of all CTCF binding sites identified in ESCs, many of those in close proximity to the promoter or enhancer elements of tissue specific genes, remain stably occupied by CTCF in various differentiated tissues [54]. Remarkably, the significance of this seemingly constitutive CTCF binding—whether it serves a structural role (e.g. to enhance chromosomal

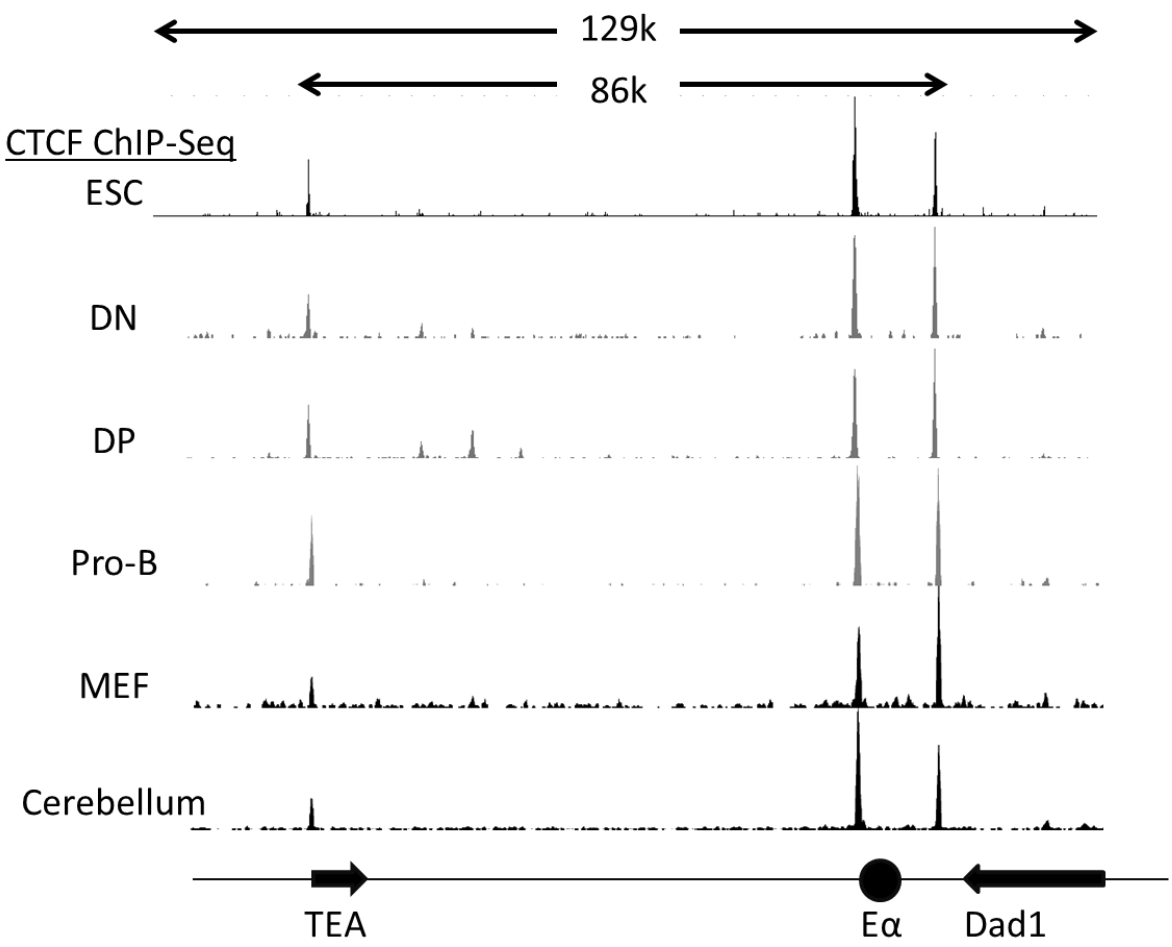


Figure 2.1: CTCF binding across the Jα locus remains constant throughout stem cell development. ChIP-seq datasets compiled from [55] (MEF, Cerebellum), [37] (DN, DP, Pro-B) and [56] (ESC) were analyzed for CTCF enrichment across the murine *Tcrα* Jα region, spanning from upstream of TEA to downstream of the Dad1 gene (reference genome mm9, chr14: 54750000-54879000) All datasets were obtained from the NCBI Gene Expression Omnibus (GEO), and visualization was performed using the UCSC Genome Browser (http://genome.ucsc.edu/).

integrity) or a functional role (e.g. to prime or sustain the priming of genes), or whether it is simply byproduct of ubiquitous CTCF expression—is unknown. However, the rapidly expanding evidence that implicates CTCF in an array of developmental programs and pathologies suggests a critical need to explore the potential significance of CTCF targeting.

2.3.2 CTCF binding at TEA correlates with H3K9 acetylation, but not with Eα looping in ESCs

To determine if CTCF plays a role in establishing an open or primed chromatin conformation at TEA similar to its role at the chicken β-globin gene [18], we first analyzed the distribution of post-translational histone modifications at TEA in the wildtype C57BL6/129SvEv EF-1 ESC line, along with thymocytes freshly isolated from mice deficient for *Rag1* (>95% DN thymocytes), deficient for *Rag1* but expressing a rearranged *Tcrb* transgene (termed Rxβ [57], >98% DP thymocytes) and from Rxβ mice that also carried a homozygous deletion of Eα (>98% DP thymocytes). As shown in Figure 2, significant alterations in histone H3 marks proximal to TEA accompany hematopoietic development. The most defined mark of open chromatin [58], H3K9 acetylation (grey bars), is already established at TEA in ground state ESCs, persists in DN and DP cells, and is not sensitive to the presence of Eα, which is required for H3K9ac of more 3' Jα segments [59]. In contrast, two marks strongly associated with transcriptional activation, H3K4 dimethylation (white bars) and H3K4 trimethylation (black bars), are strongly induced in Eα-containing DP cells (though more modest induction at TEA was retained in the Eα-deficient DP cells). Finally, H3K27 trimethylation (hatched bars), which correlates strongly with transcriptionally silent chromatin, was highest in ESCs, and reduced >8-fold in DN and DP cells that retained Eα. Enhancer deletion in the Eα-Rxβ DP cells led to elevated H3K27me3 levels, suggesting that while Eα is required to fully induce transcription-associated H3K4

methylation at TEA, it also blocks trimethylation of H3K27 at the promoter upon the DP transition. This sensitivity of TEA histone methylation to Eα in DP cells correlates with the developmental stage in which CTCF acts with the protein complex cohesin to induce TEA:Eα chromatin loops [40]. Though no studies have directly assessed TEA histone modifications in DP cells deficient for CTCF, Seitan et al. found a 3-fold reduction in H3K4me3 levels in DP cells deficient for the cohesin protein Rad21. Taken together, these ChIP data suggest that persistent CTCF binding at TEA coincides with stable H3K9 acetylation, but is not sufficient in itself to maintain other histone markers of transcriptional competency, which are instead dependent upon developmental progression and Eα interaction.

As has been mentioned, CTCF:cohesin complexes facilitate an intrachromosomal interaction between TEA and Eα that is required for transcriptional accessibility, and therefore recombinational accessibility, of TEA-proximal Jα segments [37, 60, 61]. To determine if cohesin colocalizes with CTCF at TEA and Eα in ESCs, we compiled ChIP-seq experiments on the genomic distribution of the cohesin subunits Rad21 and Smc1 (data not shown). Indeed, strong binding of the cohesin subunits overlaps CTCF at both elements. However, quite unlike findings in DP cells [37, 40], chromosome conformation capture (3C) demonstrated no significant interaction between TEA and Eα in ESCs (Fig. 3A). These data corroborate more recent findings in DN cells [37], and suggest that while the proteins most associated with intrachromosomal looping are preassembled at TEA and Eα early in pluripotency, other unknown factors must restrict loop formation to DP development. One possibility is that TEA or Eα (or both) are sequestered from one another in earlier staged cells by association with alternative *cis* elements. To test this possibility, we measured genome-wide interactions involving TEA or Eα in recently published ESC datasets of CTCF-mediated chromatin

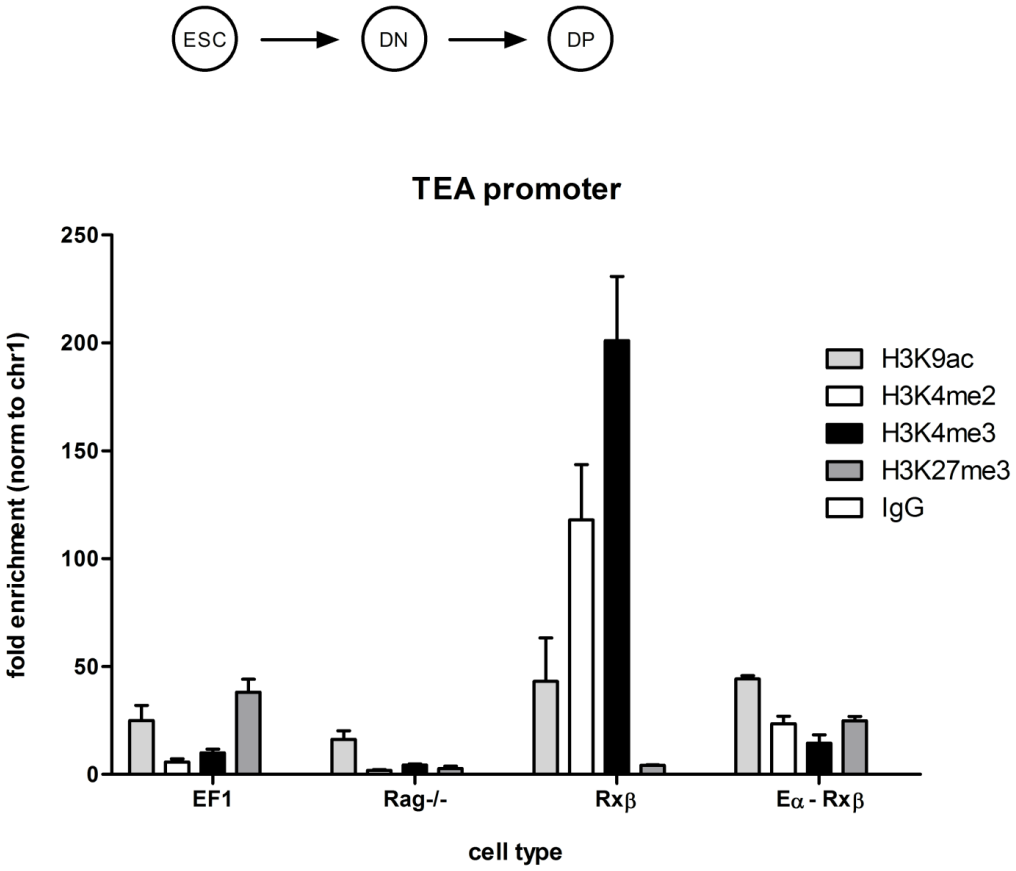


Figure 2.2: CTCF binding at TEA correlates with early establishment of H3K9 acetylation. Chromatin from the B6/129 embryonic stem cell line EF1, Rag1-deficient DN thymocytes (Rag -/-), DP Rag-deficient thymocytes that express a stably transfected β transgene (Rxβ) and Rxβ thymocytes that harbor a homozygous deletion of the downstream α enhancer (Eα- Rxβ) was immunoprecipitated with the indicated antibodies against covalently modified histone H3 or isotype-matched IgG. Resultant DNAs were analyzed by qPCR at the TEA promoter. Enrichment (ΔΔCT) was calculated relative to pre-IP input control levels and was normalized against signals obtained for an intergenic region of chromosome 1. Bars indicate means (± SD, n = 3), and are representative of 2 independent assays.

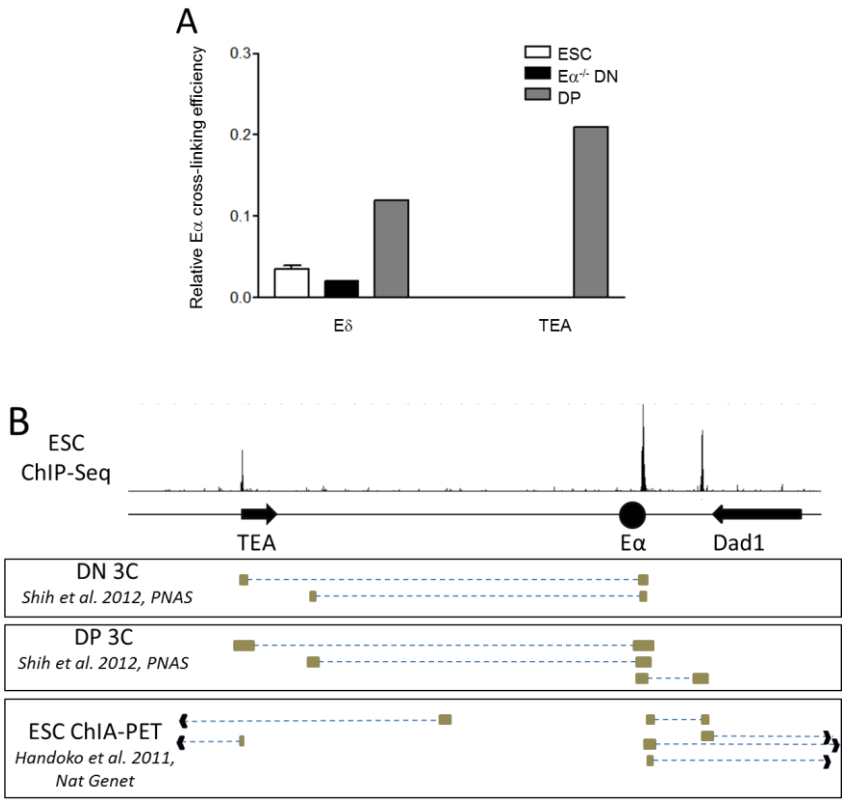


Figure 2.3: TEA:Eα looping is not present in embryonic stem cells. (A) 3C was conducted on chromatin from ESCs, double-negative thymocytes from mice that carry a homozygous deletion of Eα (Eα$^{-/-}$ DN) and Rxβ double-positive (DP) thymocytes. Following cross-linking, DNA isolation, restriction endonuclease digestion and ligation, interactions between HindIII fragments containing Eα and TEA or Eα and Eδ were assessed by real-time PCR. All reactions were run in duplicate, and samples were normalized to Eα internal interaction (set to 1.0). (B) Graphical representation of 3C and ChIA-PET analyses of intrachromosomal interactions involving either TEA or Eα at the indicated developmental stages [37, 62]. Width of boxes indicates frequency of sequencing reads, dashed lines and arrows indicated upstream or downstream intrachromosomal interactions. ESC CTCF ChIP-Seq [56] is shown for reference.

interaction analysis with paired-end tag sequencing (ChIA-PET) [62]. As with 3C our analyses of the ChIA-PET datasets failed to detect any TEA:Eα interactions in ESCs, but did reveal a number of alterative intrachromosomal interactions, including associations between Eα and *Dad1,* and between either Eα or TEA and distal genes (Fig. 3B). ChIA-PET analyses do not yet exist for either DN or DP cells. Consequently, the developmental conservation or potential significance of these interactions remains to be determined.

2.3.3 The TEA promoter remains unmethylated throughout hematopoietic development

Methylation of CpG dinucleotides, while dispensable for division and longevity of embryonic stem cells, is required for their differentiation to early hematopoietic progenitors [63, 64]. CTCF has been linked to the methylation-dependent transcriptional regulation of multiple genes, and its binding is strongly inversely correlated at most sites with DNA methylation [31, 54, 65-67]. Moreover, like the developmental persistence of CTCF, methylation patterns at CTCF binding sites largely remain constant throughout differentiation [54]. The CTCF binding site at TEA (*CAGCGCCCCCTGCTGC*) contains an internal CpG. TEA contains two other CpGs, positioned roughly 120 bp 5' and 140 bp 3' of the CTCF site. The methylation status of TEA has not been previously reported for any stage of murine thymocyte development. Given the prominence of a CpG dinucleotide within the TEA CTCF binding site, we used bisulfite conversion to determine how TEA methylation correlated with CTCF binding during development (Fig. 4). Consistent with CTCF binding, we found that the CpG contained within the CTCF binding site of TEA was unmethylated in every clone analyzed for the ground state EF1 ESC line, *Rag1*-deficient DN thymocytes, and Rxβ DP thymocytes (Fig. 4A). Both the upstream and downstream flanking CpGs were methylated slightly more frequently in ESCs, but both were devoid of methylation in DN and DP thymocytes.

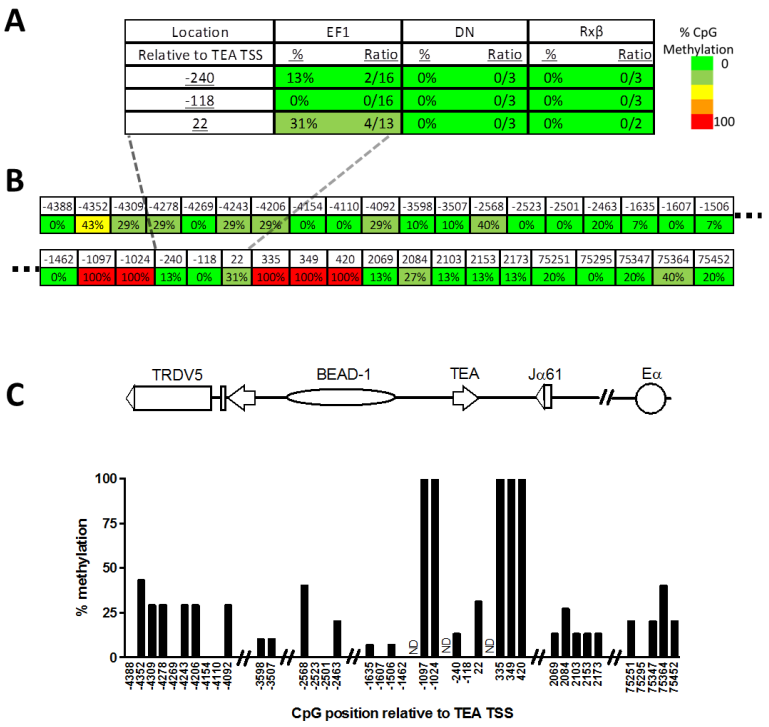


Figure 2.4: The TEA promoter is unmethylated throughout hematopoietic development and is contained within a distinct island of hypermethylation. Methylation of the TEA promoter is shown in the embryonic stem cell line EF1, DN thymocytes and Rxβ thymocytes as determined by bisulfite mapping (A). Bisulfite survey of EF-1 for the region surrounding TEA was conducted using a series of discrete primer sets (B) – the top number represents genomic distance from the TEA transcription start site while the bottom represents the % of DNA methylation as seen over 5-15 clones (for ratios, see Table S2). Methylation levels are represented in a gradation of colors: light green (0-20%), dark green (21-40%), yellow (41-60%), orange (61-80%) and red (81-100%). Methylation results in EF-1 compared to genetic elements Vδ5, BEAD-1, TEA, Jα61 and Eα are shown in a graphical for reference (C). ND – not determined. Breaks in the line indicate regions between primer sets.

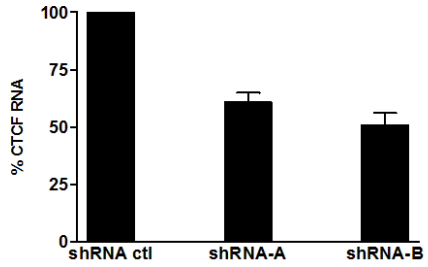
We expanded bisulfite analyses of our ground state EF1 ESC line to encompass most of the CpGs within a 6 kb region from just upstream of the 3' most V segment (Vδ5) to just downstream of the 5' most Jα segment (Jα61). As expected in ground state cells, the majority of CpGs in the vicinity of TEA and within a CpG-rich region proximal to Eα were hypomethylated, with a limited but reproducible increase in methylation across a ~150 bp CpG-rich region of Vδ5 (Fig. 4B-C). Unexpectedly though, narrow bands of fully methylated CpGs were detected less than 1 kb upstream and downstream of TEA. Regions only 350 bp 5' of the upstream methylation peak and 1.5 kb 3' of the downstream peak were again generally hypomethylated, suggesting that the bands of methylated DNA are tightly centered around TEA. MeDIP-seq analyses of murine embryonic stem cells heterozygous for the DNA repair enzyme Thymine DNA glycosylase (TDG) [68] corroborate peaks in methylation immediately upstream and downstream of the TEA CTCF (data not shown). These peaks of hypermethylated DNA do not correlate with any known regulatory element in the region, including the BEAD-1 DNA insulator element [69]. Indeed, CpGs across BEAD-1 were strongly hypomethylated.

2.3.4 CTCF is required to maintain TEA hypomethylation in ESCs

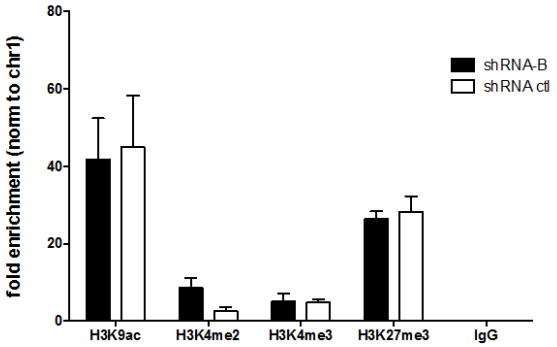
Although CTCF binding is strongly correlated genome-wide with binding site hypomethylation [30], we wished to determine if CTCF's presence at TEA directly contributed to maintaining the promoter's methylation status. To that end, we used two independent lentivirus-mediated shRNAs to stably knock down CTCF expression in EF1 cells (Fig. 5A), and then assessed TEA histone modification and methylation in response to the CTCF knockdown (Fig. 5B-C). Though no significant changes in histone modification patterns were observed, all three CpGs contained within TEA showed a marked increase in methylation in stable lines isolated from each of the two shRNAs.

Figure 2.5: Knockdown of CTCF results in increased DNA methylation at the TEA promoter without effecting local histone modifications. Lentivirus-mediated transduction of CTCF shRNA templates was used to generate EF1 clones with diminished CTCF expression relative to a lentiviral scrambled control shRNA. (A) Knockdown of CTCF was assessed by qRT-PCR and normalized to β actin. Bars represent mean CTCF expression ± SD (n=3 technical replicates) and are representative of 3 independent experiments. Successful clones 1-15 (shRNA-A) and 2-12 (shRNA-B) were isolated and used for all further experiments. (B) Histone modifications at TEA following CTCF knockdown. Chromatin from control and CTCF knockdown clones was immunoprecipitated with antibodies against the indicated histone post-translational modification. Resultant DNAs were analyzed by qPCR at the TEA promoter. Enrichment was calculated relative to pre-IP input control levels and was normalized against signals obtained for an intergenic region of chromosome 1. Bars indicate mean (± SD, n = 3) and are representative of 2 immunoprecipitations. (C) DNA methylation of the TEA promoter following CTCF knockdown. Genomic distance of each CpG dinucleotide from the TEA transcription start site is indicated above the percent DNA methylation in 7-10 independent clones as shown. (D) Tabular and graphical representation of a broader analysis of DNA methylation across the TEA region following CTCF knockdown. Genomic distance of each CpG dinucleotide from the TEA transcription start site is indicated above the percent DNA methylation in 3-9 independent clones as shown (ND – methylation status of intervening CpGs not determined).

A



B



C

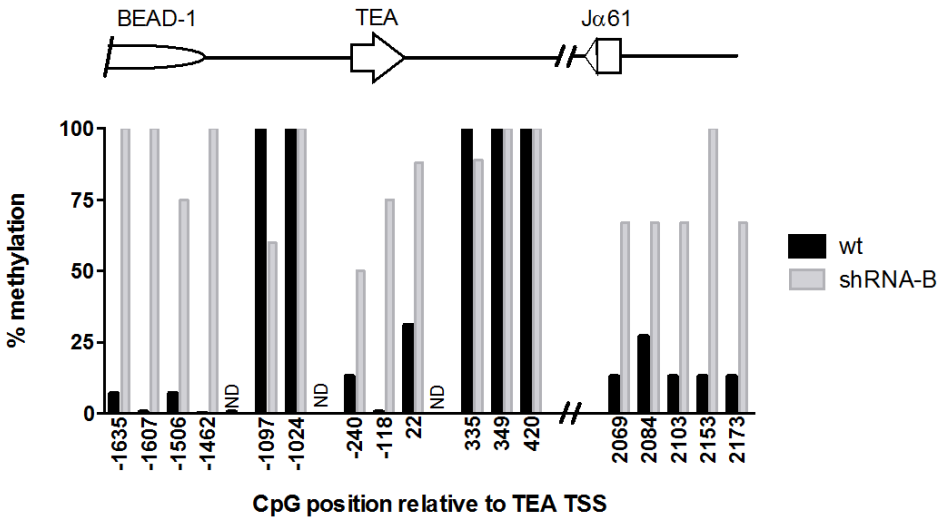
Location	siControl		CTCF RNAi 1-15		CTCF RNAI 2-12	
Relative to TEA TSS	%	Ratio	%	Ratio	%	Ratio
-240	10%	1/10	29%	2/7	50%	4/8
-118	0%	0/10	25%	2/8	75%	6/8
22	56%	5/9	71%	5/7	88%	7/8

% CpG Methylation 0 100

D

Location	-1635	-1607	-1506	-1462	-1097	-1024	-240	-118	22	335	349	420	2069	2084	2103	2153	2173
EF1	7%	0%	7%	0%	100%	100%	13%	0%	31%	100%	100%	100%	13%	27%	13%	13%	13%
shRNA-B	100%	100%	75%	100%	60%	100%	50%	75%	88%	89%	100%	100%	67%	67%	67%	100%	67%

% CpG Methylation 0 100

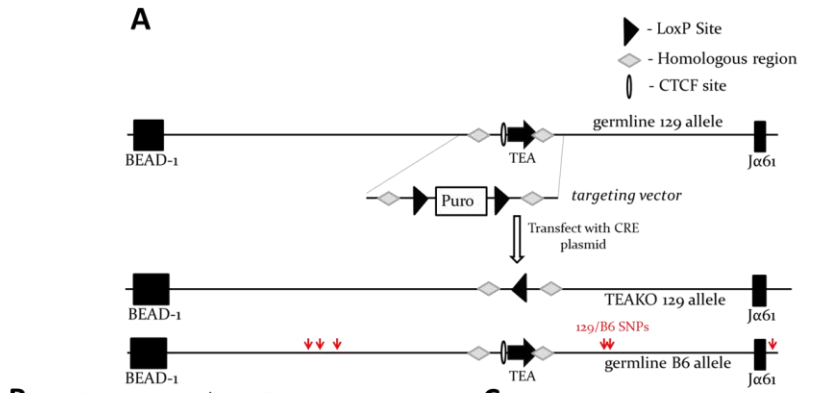


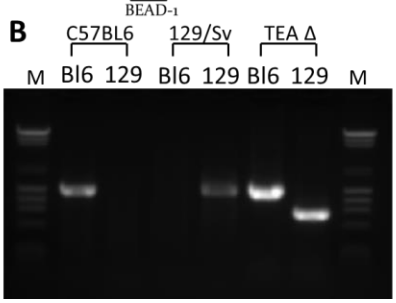
Additional analysis of one clone (CTCF shRNA B) revealed that full methylation of the sequence immediately surrounding TEA was unaffected by CTCF knock down, while BEAD-1 and Jα61 sequences outside of the hypermethylation peaks mirrored the strong increase in methylation shown for TEA. As such, our findings suggest that CTCF is required to prevent the peaks of hypermethylation that flank TEA in ESCs from spreading not only across the promoter, but also across the region that separates *Tcrd* from the Jα array.

2.3.5 Monoallelic TEA knockout induces biallelic ground-state hypermethylation of the Jα region.

CTCF binds a wide array of targets across the ESC genome [70], and its knockdown would be predicted to impact a variety of genes. To test whether CTCF acts directly to protect against hypermethylation across the BEAD-1-Jα61 region, we assessed the impact on DNA methylation from deleting TEA in ground-state ESCs. Cre-*loxP* targeting was previously used to remove the Jα49 promoter from the 129 allele of the EF1 ES line, and then to retarget a 668 bp fragment that includes TEA (-514 to +154, relative to the TEA TSS), while leaving the C57BL6 allele unaltered [39]. Using these cells, we isolated Cre-transfectants that carried a TEA deletion from the 129 allele (Fig. 6) and carried them through 6-10 passages in 2i medium to restore ground-state epigenetic patterning [71].

We used single-nucleotide polymorphisms (SNPs) between the 129 and B6 sequence of *Tcra* to determine the effects of TEA deletion on DNA methylation and histone marker deposition across the BEAD-1-Jα61 region. As with CTCF K/D, we found no significant differences in H3K9 acetylation or H3K4 or K27 methylation between the 129 and B6 alleles of either parental EF1 cells or $TEA^{+/-}$ cells (data not shown). DNA methylation analysis of the region, however, revealed significant differences between the wildtype B6 allele and the mutated 129 allele (Fig. 7). Beyond the peaks of





C

GAGCCTGCCAATCATGGATAAGGACAGATCTCC.....CGGGTTTAAACTAGGG
Germline 129 Allele Vector MCS

CTCTAGAACTAGTGGATCCGGAACCCTTAATATAACTTCGTATAATGTATGCTA
LoxP Site

TACGAAGTTATTAGGTCCCTCGACCTGCAGCCCAAG.....CTAGCTATCTTGTG
Vector MCS

GAATACTCAGATATGTGTCTAAGATGGAAATCTCAGAAGCATTTGGAATGTG
Germline 129 Allele

D

CpG Location (Relative to TEA TSS)

	EF-1 (embryonic)		TEAKO (embryonic)		TEAKO DP (DP)		RxB (DP)	
	%	Rati	%	Ratio	%	Ratio	%	Ratio
75251	20%	1/5	0%	0/4	0%	0/5	0%	0/4
75295	0%	0/5	0%	0/4	0%	0/5	0%	0/4
75347	20%	1/5	0%	0/4	80%	4/5	50%	2/4
75364	40%	2/5	50%	2/4	80%	4/5	75%	3/4
75452	20%	1/5	0%	0/4	40%	2/5	25%	1/4

Percent Methylation
0 - 20%
20 - 40%
40 - 60%
60 - 80%
80 - 100%

Figure 2.6: Generation of monoallelic TEA knockout ESC lines. A) Schematic representation of strategy for Cre-lox-mediated deletion of TEA on the 129 Sv allele of EF1 ESCs. Single nucleotide polymorphisms exploited for allele-specific analyses of DNA methylation are indicated (red arrows). Deletion of the 650 bp TEA promoter from the 129Sv allele of EF1 cells was confirmed by PCR (B) and sequencing (C). Initial bisulfite analysis revealed that deletion of TEA did not affect DNA methylation in the vicinity of Eα, regardless of TEA:Eα intrachromosomal interactions (D).

heavy methylation flanking the wildtype promoter (and which remain heavily methylated in $TEA^{+/-}$), upstream BEAD-1 and downstream Jα61 sequences show significant hypermethylation on both the mutated and wildtype alleles. Despite this biallelic hypermethylation of flanking sequences, which mirrored the hypermethylation observed in CTCF K/D cells, the promoter itself remained hypomethylated on the wildtype B6 allele of $TEA^{+/-}$ cells (data not shown), and downstream Eα sequences remained hypomethylated on both alleles (Fig. 6D). Together with data from the K/D lines, these analyses suggest that: (a) CTCF binding likely directly protects a narrow window containing the 300 bp TEA promoter against hypermethylation targeted to the promoter's flanking sequences, while (b) protection of more distal sequences in the BEAD-1-Jα61 region of ground-state ESCs may require promoter-dependent interactions between the two *Tcra* alleles. The role of CTCF in maintaining hypomethylation across the region remains to be determined. However, since BEAD-1 and Jα61 are more heavily methylated in DP thymocytes despite CTCF binding at TEA, interallelic protection against methylation appears limited to earlier stages of development.

2.3.6 CTCF knockdown results in increased DNA methylation at the *Ptcra* enhancer.

Previous work by Smale and colleagues showed that the enhancer for the PreTα gene (*Ptcra*), another gene specifically activated in early thymocytes and required for progression to the DP cell stage, also remained hypomethylated throughout hematopoietic development [72, 73]. In their studies, the authors found that premethylation of the *Ptcra* enhancer silenced reporter genes stably introduced into DN cells, while methylation marks were erased when reporter genes were introduced into ES cells [74]. We analyzed ChIP-seq data from the Schübeler laboratory [56] to assess CTCF binding in the vicinity of the *Ptcra* enhancer (Fig. 8A). Though we observed no direct CTCF binding at the *Ptcra* enhancer, a strong CTCF binding was found about 4 kb upstream in ESCs,

DN and DP cells. We also noted that additional CTCF peaks at both the *Ptcra* enhancer and promoter were induced in DP cells, where PreTα expression is downregulated [75]. MeDIP-seq analysis corroborated the window of unmethylated CpGs over the Ptcra enhancer in ESCs, but revealed strong methylation of the *Ptcra* promoter (Fig. 8B). In contrast, little methylation was observed across the *Ptcra* regulatory region in mouse embryonic fibroblasts.

Though CTCF did not appear to bind the *Ptcra* enhancer, we wished to determine if CTCF might play an indirect role in methylation protection. Smale and colleagues proposed that Sp1 acts as a pioneer factor to promote the ESC-specific priming of the *Ptcra* enhancer through unknown mechanisms [72]. As CTCF is noted to collaborate with Sp1 at a number of genes [76, 77] and has been implicated in regulating other pioneer factors [31, 56, 65, 78, 79], we determined the effect of CTCF K/D at the *Ptcra*. As seen at TEA, knockdown of CTCF significantly increased methylation across the *Ptcra* enhancer (Fig. 8C). The increase in methylation varied widely across the enhancer region, suggesting perhaps that CTCF may not play as central a role in protecting *Ptcra* against methylation as it does at TEA. Indeed, targeted mutagenesis of the *Ptcra* reporter also suggested a potential role for an unidentified E protein in methylation protection [72, 74]. Using ChIA-PET, we also identified multiple CTCF-mediated interchromosomal interactions involving the *Prcra* enhancer (data not shown). Whether methylation protection of the *Ptcra* enhancer requires CTCF binding distal to the enhancer, or requires CTCF-mediated formation of a larger protein:DNA complex remains to be determined. Regardless, our findings at TEA and *Ptcra,* coupled with the genome-wide correlation between CTCF binding and local hypomethylation, suggest that the novel role of CTCF in ESC priming of differentiation-specific genes may be a more general mechanism of epigenetic developmental regulation.

A

	-1506	-1462	-1288	-1097	-1024	-240	-118	22	335	349	420	2069	2084	2103	2153	2173
EF1	7%	0%	nd	100%	100%	13%	0%	31%	100%	100%	100%	13%	27%	13%	13%	13%
TEAΔ B6 Allele	33%	67%	83%	100%	100%	17%	0%	33%	88%	50%	100%	80%	40%	80%	80%	100%
TEAΔ 129 Allele	82%	36%	91%	64%	91%	xx	xx	xx	100%	100%	nd	80%	80%	100%	80%	80%

B

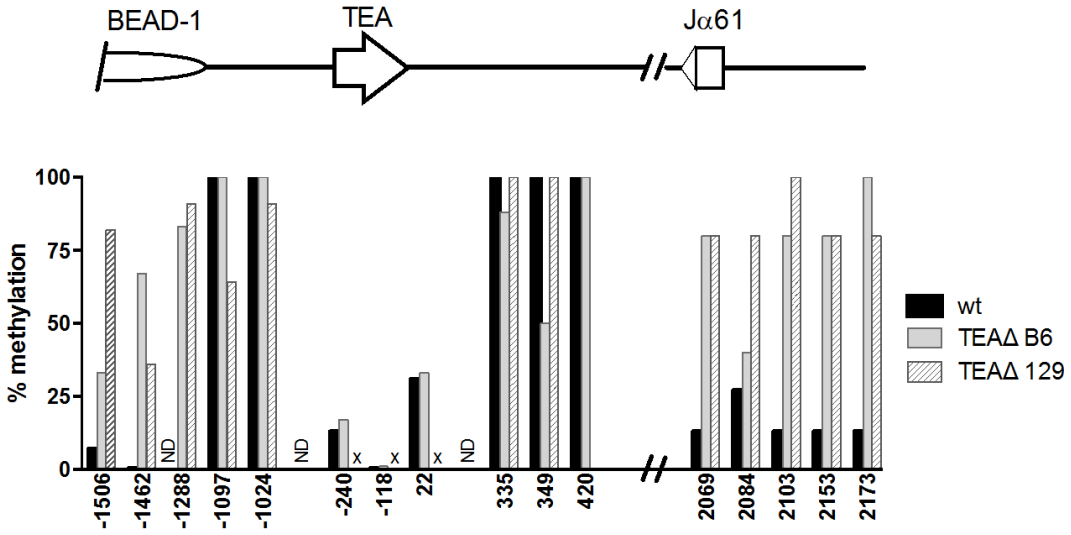
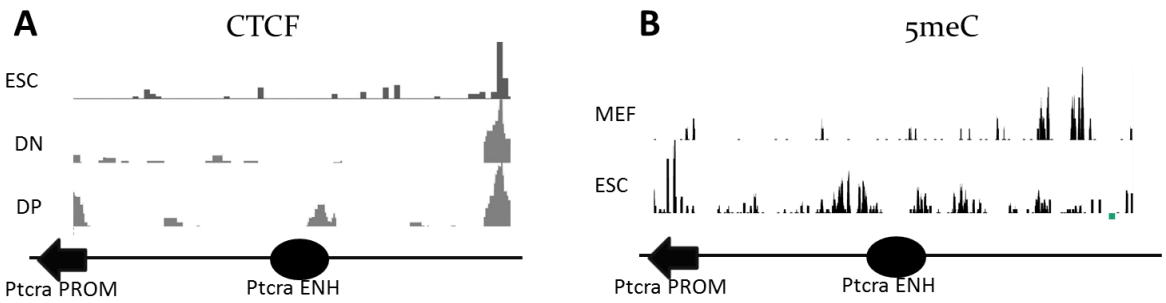


Figure 2.7: Allele-specific knockout of TEA results in a biallelic increase in localized DNA methylation immediately outside of the promoter. A) Bisulfite mapping was performed using allele-specific primers to identify the effects of TEA deletion. For ratios, see Table S2. nd – not determined, xx – contained within deletion. B) A graphical representation of DNA methylation patterns of EF1 and both the knockout and wild-type allele of a TEAKO clone, roughly aligned to the genetic landmarks BEAD-1, TEA and Jα61 (ND – methylation status of intervening CpGs not determined, X – CpG deleted from TEAΔ 129 allele). CpGs are identified based on distance from the TEA transcription start site.



C

PtcraE	EF1		siControl		shRNA-B	
chr17:	%	Ratio	%	Ratio	%	Ratio
46905024	33%	1/3	20%	1/5	43%	3/7
46905074	0%	0/3	20%	1/5	43%	3/7
46905112	33%	1/3	0%	0/5	57%	4/7
46905157	0%	0/3	0%	0/5	14%	1/7
46905189	0%	0/3	0%	0/5	57%	4/7
46905207	33%	1/3	20%	1/5	71%	5/7

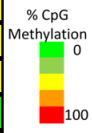


Figure 2.8: Knockout of TEA results in increased DNA methylation at an unmethylated window in the *Ptcra* enhancer . A) CTCF binding analysis performed using ChIP-Seq datasets compiled from the Torkamani lab (DN, DP [37]) and the Schuebeler lab (ESC [56]) and analyzed for the Ptcra region (reference genome mm9, chr17:46,900,500-46,908,500). B) DNA methylation analysis performed using MeDIP-Seq datasets compiled from the Beck lab [68] and analyzed for the Ptrca region (same coordinates as in Fig. 8A). Shown MEF/ESC are representative of triplicate MeDIP-seq experiments. All datasets are freely available from GEO and visualization was performed using the UCSC Genome Browser. (http://genome.ucsc.edu/) C) Bisulfite mapping of the Ptcra enhancer in wild-type EF1 cells and upon stable integration of anti-CTCF or scrambled shRNA constructs.

2.4 Discussion

Though *Tcra* is not transcriptionally or recombinationally active until the DP stage of thymocyte development, our findings suggest that *Tcra* regulatory elements are primed for activation in very early embryonic stem cells, and that CTCF specifically protects the T early alpha promoter in ESCs against repressive methylation. Specifically, we found that the CTCF binding pattern at TEA and Eα that is critical for *Tcra* germline transcription and recombination in DP cells is present in early epiblast ground-state cells, and correlates with strong hypomethylation of DNA surrounding both regions throughout development. Indeed, the only cell type analyzed that shows hypermethylation of CpGs directly over TEA is murine sperm, which lacks CTCF expression [80] (Fig. S1). Loss of either CTCF or its *cis* target resulted in hypermethylation of the BEAD-1-TEA-Jα61 region, and in the case of heterozygous TEA deletion this hypomethylation was not confined to the targeted allele, strongly suggesting that CTCF binding protects the TEA region against methylation, perhaps through formation of a biallelic regulatory complex. We also report evidence of a similar protective role for CTCF at the *Ptcra* gene, suggesting CTCF's role as a protector of differentiation-specific genes against methylation during early embryonic development may be more widespread.

While >85% of all CpG dinucleotides are methylated in the genome of terminally differentiated mammalian cells [81], >90% of CpGs within so-called CpG islands are hypomethylated [82]. This hypomethylation at CpG islands, which are predominantly found in the promoter regions of genes - particularly housekeeping genes - has long been known to be essential for promoter activity [28, 83, 84]. However, much less is known about the regulatory impact of CpG methylation on genes that are not associated with CpG islands. *Tcra* does not have sufficient GC content or CpG dinucleotide percentage to be considered either a CpG island or a CpG island shore [82, 85-87]. In

fact, the nearest CpG island identified by the UCSC Genome Browser to the TEA promoter is located far downstream, at the transcription start site of *Dad1*, and there are no CpG islands upstream of TEA for over a megabase (http://genome.ucsc.edu/cgi-bin/hgTracks). Outside of CpG islands, low levels of methylation at promoters correlate with transcriptional activation and developmental potency [88, 89], while within gene bodies, high levels of DNA methylation correlate with active transcription [89, 90]. Consistent with this pattern, we find that the Jα region appears hypomethylated in the EF-1 cell line (where it is not transcribed), but is heavily methylated in Rxβ thymocytes (where it is transcribed), while TEA remains hypomethylated throughout development.

Genome-wide, CTCF binding inversely correlates with CpG methylation. A causal relationship between the two remains unclear, though the ability of CTCF to direct demethylation of premethylated reporter transgenes argues that CTCF binding can initiate local demethylation [56]. In this study, we illustrate that CTCF binding at TEA, in a manner independent from its recently discovered role in *Tcra* recombination, is retained in a variety of differentiated embryonic and extraembryonic tissues. The embryo begins in a globally demethylated state as a result of gamete fusion and goes through a process of *de novo* methylation soon after uterine implantation [26, 29, 91, 92]. Detection of similar CTCF binding patterns in the TEA region of cells isolated from different compartments of the early blastocyst (placenta from the trophectoderm and fetal tissues from the epiblast) suggests that CTCF binding occurs prior to the peri-implantation global methylation event. Ground-state ESC culture (i.e., minimal defined serum-free medium supplemented with LIF and inhibitors of MEK and GSK3) supports the development of a totipotent state similar to pre-implantation embryonic cells [93]. So-called 2i cultures are also marked by dramatic genome-wide hypomethylation relative to traditional serum-cultured ESCs [94]. Yet, loss of either CTCF or the TEA

promoter led the methylation domains that normally flank TEA to expand across the promoter region in our 2i cultures, suggesting that CTCF binding is essential to maintain the promoter's hypomethylation from the ground state of embryonic development onward.

Irrespective of the role played by CTCF in protecting TEA from methylation, it is noteworthy that this protective role may be necessitated by the hypermethylated environment that surrounds TEA. Though DNA methylation patterns are reversible between 2i and serum ESCs [94], hypermethylation of the 3 kb stretch that separates BEAD-1 from Jα61 (which includes the TEA promoter) remains relatively constant in both our 2i ESCs and in 2i ESCs examined by Habibi et al. ([94], Fig. S1). Indeed, only in the genome of oocytes was the TEA flanking region entirely unmethylated [95], suggesting that it is targeted for methylation in the earliest stages of embryogenesis. In contrast, analysis of the methylome in naturally CTCF-repressed sperm [80] reveals full methylation of the entire region, including the normally unmethylated TEA promoter itself (Fig. S1), further affirming a role for CTCF in this protection. What might be the functional significance of this hypermethylation? One possibility is suggested by the positioning of *Tcrd* between the V and J segments of *Tcra.* Given that TEA sits less than 10 kb downstream of the *Tcrd* enhancer and less than 4 kb downstream of the TRVD5 promoter, constitutive hypermethylation of the intervening region may aid in segregating the Jα gene segments from *Tcrd* gene regulation. Such a role might explain why mice deficient for the BEAD-1 insulator element originally hypothesized to play that role showed no apparent perturbation in either *Tcrd* or *Tcra* gene regulation [96]. The 2 kb differentially methylated region (DMR) that governs *H19/Igf2* imprinting is similarly hypermethylated in 2i and serum ESCs [94, 97], though unlike the region at TEA, the DMR is CpG-rich, is only methylated on the paternal allele, and this methylation includes 4 unoccupied CTCF

binding sites [98]. Notably however, loss of CTCF leads to hypermethylation of the maternal DMR similar to the effect we observe at TEA [99]. Alternatively, hypermethylation of the TEA region, together with the equally stable binding of CTCF at TEA, may play a more structural role in chromosomal organization or maintenance. At present, the functional significance of the methylated sequence surrounding TEA as well as the factors responsible for its establishment and maintenance remain to be determined.

The density of SNPs between the 129 and B6 alleles of *Tcra/d* in EF-1 ESCs allowed us to observe that while TEA itself remained unmethylated on wildtype B6 alleles of heterozygous 2i $TEA^{+/-}$ cells, methylation was dramatically increased across more distal BEAD-1 and Jα61 sequences on both the B6 and mutant 129 alleles. Hypomethylation of the B6 TEA and homozygous hypomethylation of Eα in these cells precluded an indirect effect of TEA deletion on *Tcra* methylation in general. At the same time, the Jα region shows increasing methylation in RAG-deficient DN thymocytes, and full methylation in RAG-deficient DP thymocytes, despite the presence of CTCF bound to TEA (data not shown). As such, the biallelic impact of TEA deletion that we observe on flanking sequence methylation in 2i cells may indicate an inter-allelic interaction that is limited to early development. Homologous pairing of chromosomes has been noted previously in embryonic stem cells, where transient association of the two X chromosomes leads to monoallelic X inactivation [100]. CTCF is vital to X-inactivation, as CTCF knockdown precludes endogenous homologous association and impairs the efficiency of X inactivation [101]. Similarly, biallelic methylation protection across the BEAD-1-Jα61 region of 2i ESCs may contribute to establishing the functional boundary that separates *Tcra* and *Tcrd,* and is eventually replaced by activities that do not require hypomethylation. Further analysis will be necessary to determine the molecular nature

of the biallelic impact on BEAD-1-Jα61 methylation we observe in TEA$^{+/-}$ cells, and its significance to developmental regulation of *Tcra/d.*

In summation, our findings illustrate that a defined DNA methylation phenotype for the BEAD-1-TEA-Jα region is established in ground-state embryonic cells, long before commitment to the hematopoietic lineage. This methylation phenotype is dependent on CTCF expression as well as a CTCF site located within the TEA promoter, and monoallelic elimination of TEA appears to dramatically alter proximal DNA methylation patterns in ground-state ESCs in a biallelic fashion. How this alteration correlates with nuclear association of chromosomes and transcriptional competence in ESCs, as well as the effect of this deletion on hematopoietic lineage commitment and subsequent antigen receptor assembly potential, will increasingly shed light on the complex epigenetic regulation of this and other differentiation-specific genes in early development.

Table 2.S1. *PCR primer sets*

Amplicon	Strand	Sequence	Tm
KO Screen			
B6 Allele	S	CTA AAA ACC AGG CCT GAC TTT CC	63°C
	A	AGG AGT CAG TCA CAG TCC CAT CAA	
129 Allele	S	CTA AAA ACC AGG ACT GGC TTT CT	63°C
	A	AGG AGT CAG TCA CAG TCC CAT CGC	
ChIP			
TEApro2	S	TCT GTC TCC ATT CCC CAT CAT CC	63°C
	A	TCT CTT ATC CTT GAA GCC CGT CC	
Chr1	S	GTG GAT AGG GGG TCT GAG G	63°C
	A	CCA CTA ACC TGA CTA TCA AGC C	
3C			
	Eα anchor	AAG TCA AGG CAC AGA CAG TC	
	TEA probe	TCC GCA AGG TCT TTC TTC AAA GCA GGA	
	Eδ probe(Test1)	TCG ACC CTG CTA TAG TCA TC	
RT-qPCR			
βactin	S	AGA GCT ATG AGC TGC CTG ACG GCC	59°C
	A	AGT AAT CTC CTT CTG CAT CCT GTC	
CTCF rtPCR	S	CCT GCT GTA GAA ATT GAA CCT GAG CC	66°C
	A	ATA GTC CTG GTG CCG AGC AAG GCC CC	
Bisulfite			
5'Vδ5bsf	S	TTT TGT TGG TTT TGT TGT GT	57°C
	A	TCT CTA CAC ACT TAA ATT TC	
3'Vδ5bsf	S	TTT AAA AGT AGA GGT GGG GT	55°C
	A	AAC CCT ATC RAA CAT TAC AAT AAA C	
3'BEADbsf	S	TGG GTG TAA TAG TTT TTT TTG	56°C
	A	TAT TCT TTT ACT ACA CAT CTA TTT AC	
5'TEAab3bsf	S	AAA GGG TAT TTT TGA GAA GGT TA	57°C
	A	AAA CTA CTT ACT ACA CAC AA	
5'TEA-Bbsf	S	GGT TTT TAG TTT AGT TTT AG	53°C
	A	CAA CCA TTT TTT CCT ATT TAA C	
5'TEA-a1bsf	S	ATT TGG GGA TAA AAT TTT GG	53°C
	A	AAA CTT TAA ACT CAC TCA ATC	
TEAbsf	S	AGG TAA AGT GAA GGA AAT TAG AA	53°C
	A	AAC TCA TAA AAC CAA TAC AAA AA	
3'TEAbsf	S	TTT TTG TTT TTT TGT GGT ATT TT	53°C
	A	AAT TCT TCA TCC TTA TTC TTA CC	
Jα61bsf	S	TGG AAG TGT TTT TTT AGG AAT AG	53°C
	A	ACA CAT TTT TTC TTT CTA AAA C	
Eαbsf	S	AGG GGG TAT TGT TTR TTT AG	59°C
	A	RAA TTC CAA ACC CTT ATT CT	
TEAKObsf	S	GTT GTA GTT GTT TGT AAT GAG TTT	56°C
	A	CTT TAC ATT ATC TCT ACC TA	
PreTαEbsf	S	GGT ATG GGT TAT TAA GTT AGT A	55°C
	A	CAA TTT TAA TAA CCA CTT TCC	

Table 2.S2. *Percent methylation of CpG dinucleotides surrounding the TEA promoter*

Cell line	CpG positions*																	
	-4388	-4352	-4309	-4278	-4269	-4243	-4206	-4154	-4110	####	-3598	-3507	-2568	-2523	-2501	-2463	-1635	-1607
EF-1	0% (0:7)	42.9% (3:7)	28.6% (2:7)	28.6% (2:7)	0% (0:7)	28.6% (2:7)	28.6% (2:7)	0% (0:7)	0% (0:7)	28.6% (2:7)	10% (1:10)	10% (1:10)	40% (2:5)	0% (0:5)	0% (0:5)	20% (1:5)	9.1% (1:14)	0% (0:14)
CTCF RNAi 2-12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100% (4:4)	100% (4:4)

*numbering relative to the first base of the TEA transcription start site

Table 2.S2. *Percent methylation of CpG dinucleotides surrounding the TEA promoter*

Cell line	CpG positions*																			
	-1506	-1462	-1288	-1097 or -1073	-1024	-240	-118	22	335	349	420	2069	2084	2103	2153	2173	75251	75295	75347	75364	75452
EF-1	9.1% (1:14)	0% (0:14)	-	100% (9:9)	100% (9:9)	12.5% (2:16)	0% (0:16)	31% (4:13)	100% (7:7)	100% (7:7)	100% (6:6)	13% (2:15)	27% (4:15)	13% (2:15)	13% (2:15)	13% (2:15)	20% (1:5)	0% (0:5)	20% (1:5)	40% (2:5)	20% (1:5)
CTCF RNAi 2-12	75% (3:4)	100% (4:4)	-	60% (3:5)	100% (5:5)	50% (4:8)	75% (6:8)	87.5% (7:8)	89% (8:9)	100% (9:9)	100% (9:9)	67% (2:3)	67% (2:3)	67% (2:3)	100% (3:3)	67% (3:3)	-	-	-	-	-
CTCF RNAi 1-15	-	-	-	-	-	28.6% (2:7)	25% (2:8)	71.4% (5:7)	-	-	-	-	-	-	-	-	-	-	-	-	-
siRNA Control	-	-	-	-	-	10% (1:10)	0% (0:10)	56% (5:9)	-	-	-	-	-	-	-	-	-	-	-	-	-
TEAKO (B6)	33% (2:6)	67% (4:6)	83% (5:6)	100% (6:6)	100% (6:6)	16.7% (1:6)	0% (0:6)	33.3% (2:6)	88% (7:8)	50% (4:8)	100% (8:8)	80% (4:5)	40% (2:5)	80% (4:5)	80% (4:5)	100% (5:5)	0% (0:4)*	0% (0:4)*	0% (0:4)*	50% (2:4)*	0% (0:4)*
TEAKO (129)	82% (9:11)	36% (4:11)	91% (10:11)	64% (7:11)	91% (10:11)	xx	xx	xx	100% (5:5)	100% (5:5)	-	80% (4:5)	80% (4:5)	100% (5:5)	80% (4:5)	80% (4:5)					
TEAKO DP	-	-	-	83.3% (5:6)	100% (6:6)	xx	xx	xx	100% (10:10)	100% (10:10)	100% (6:6)	100% (7:7)	100% (7:7)	100% (7:7)	100% (7:7)	100% (7:7)	0% (0:5)	0% (0:5)	80% (4:5)	80% (4:5)	40% (2:5)
RxB	-	-	-	100% (5:5)	80% (4:5)	0% (0:3)	0% (0:3)	0% (0:2)	75% (3:4)	75% (3:4)	100% (4:4)	100% (3:3)	100% (3:3)	100% (3:3)	67% (2:3)	33% (1:3)	0% (0:4)	0% (0:4)	50% (2:4)	75% (3:4)	25% (1:4)

*numbering relative to the first base of the TEA transcription start site

% Methylation (# times CpG was methylated : # times CpG was sequenced). Example cell: 42.9% (3:7) 42.9% methylation, 7 clones analyzed, CpG was methylated 3 times and unmethylated 4 times

*: unable to differentiate B6 allele from 129 allele -: Not tested xx: CpG deleted in knockout formation

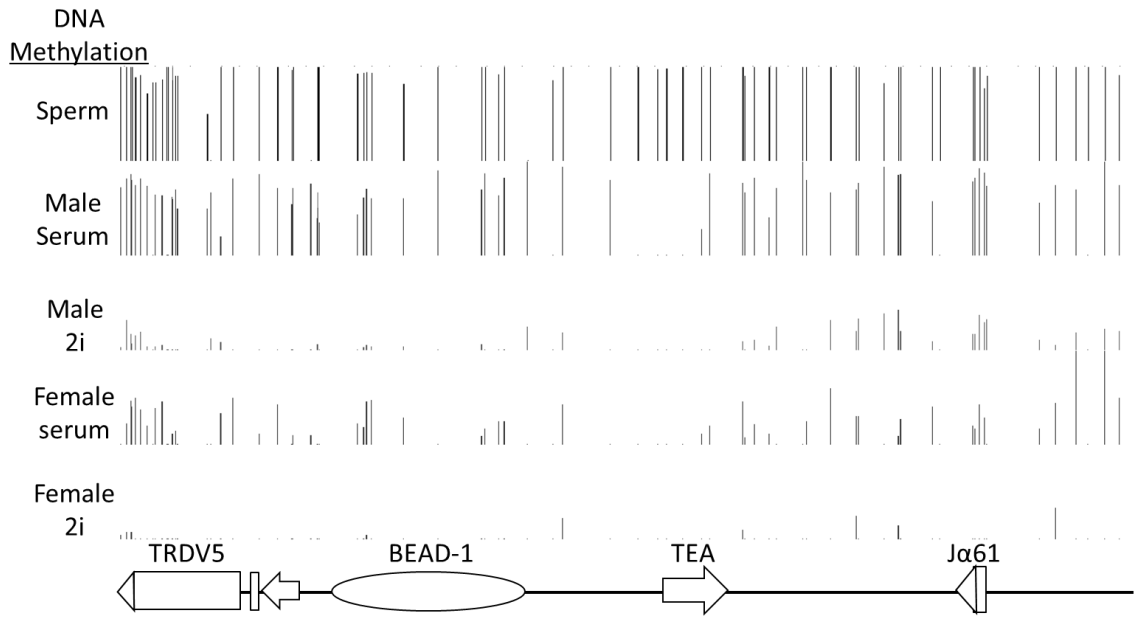


Figure 2.S1: Regions upstream and downstream of TEA remain relatively hypermethylated in both serum and 2i conditions. Global CpG methylation analyses performed using MethylC-seq (sperm) and whole-genome bisulfite sequencing (ESCs) were compiled from the Kono lab (sperm [95]) and the Marks lab (ESCs [94]) were examined in the vicinity of TEA (reference genome mm9, chr14:54,768,400-54,776,110).

Cell line	CpG					
	N1	N2	N3	N4	N5	N6
TEAKO (129)	60% (3:5)	100% (5:5)	100% (5:5)	100% (5:5)	80% (4:5)	100% (5:5)
TEAKO DP	100% (4:4)	100% (4:4)	100% (4:4)	100% (4:4)	100% (4:4)	100% (4:4)

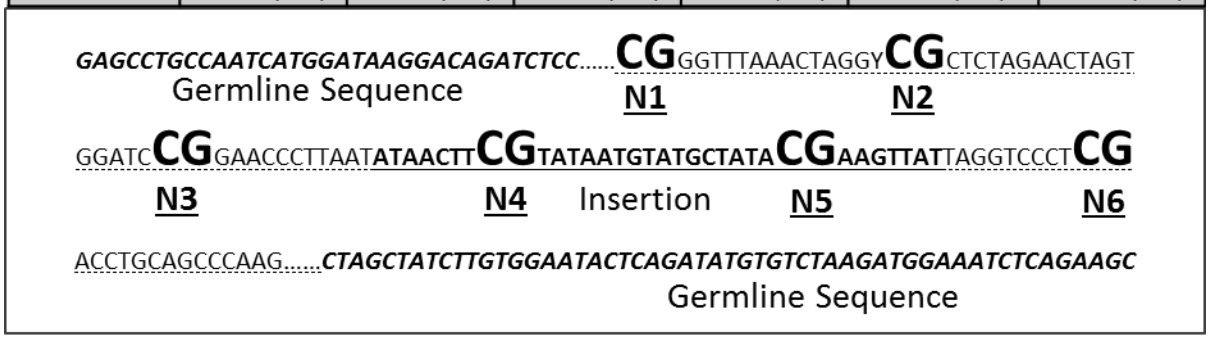


Figure 2.S2: CpG dinucleotides inserted during creation of TEAKO are fully methylated in both embryonic stem cells and DP thymocytes. Bisulfite mapping was performed using an upstream primer within the deleted region and a downstream primer within the germline sequence to selectively amplify alleles containing the TEA knockout for both the embryonic stem cell line TEAKO and DP thymocytes from a biallelic knockout mouse. The sequence information below identifies the location of each CpG within the inserted sequence.

2.5 References

1. Hemberger, M., W. Dean, and W. Reik, *Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal.* Nat Rev Mol Cell Biol, 2009. **10**(8): p. 526-37.

2. Fouse, S.D., et al., *Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation.* Cell Stem Cell, 2008. **2**(2): p. 160-9.

3. Szutorisz, H., et al., *The proteasome restricts permissive transcription at tissue-specific gene loci in embryonic stem cells.* Cell, 2006. **127**(7): p. 1375-88.

4. Efroni, S., et al., *Global transcription in pluripotent embryonic stem cells.* Cell Stem Cell, 2008. **2**(5): p. 437-47.

5. Bernstein, B.E., et al., *A bivalent chromatin structure marks key developmental genes in embryonic stem cells.* Cell, 2006. **125**(2): p. 315-26.

6. Liber, D., et al., *Epigenetic priming of a pre-B cell-specific enhancer through binding of Sox2 and Foxd3 at the ESC stage.* Cell Stem Cell, 2010. **7**(1): p. 114-26.

7. Atlasi, Y., et al., *Wnt Signaling Regulates the Lineage Differentiation Potential of Mouse Embryonic Stem Cells through Tcf3 Down-Regulation.* PLoS Genet, 2013. **9**(5): p. e1003424.

8. Meissner, A., et al., *Genome-scale DNA methylation maps of pluripotent and differentiated cells.* Nature, 2008. **454**(7205): p. 766-70.

9. Rideout, W.M., 3rd, K. Eggan, and R. Jaenisch, *Nuclear cloning and epigenetic reprogramming of the genome.* Science, 2001. **293**(5532): p. 1093-8.

10. Richly, H., L. Aloia, and L. Di Croce, *Roles of the Polycomb group proteins in stem cells and cancer.* Cell Death Dis, 2011. **2**: p. e204.

11. Su, Y., B. Deng, and R. Xi, *Polycomb group genes in stem cell self-renewal: a double-edged sword.* Epigenetics, 2011. **6**(1): p. 16-9.

12. De Gobbi M, G.D., Lynch M, Vernimmen D, Hughes JR, Goardon N, Luc S, Lower KM, Sloane-Stanley JA, Pina C, Soneji S, Renella R, Enver T, Taylor S, Jacobsen SE, Vyas P, Gibbons RJ, Higgs DR, *Generation of bivalent chromatin domains during cell fate decisions.* Epigenetics & Chromatin, 2011. **4**(1): p. 9.

13. Margueron, R. and D. Reinberg, *The Polycomb complex PRC2 and its mark in life.* Nature, 2011. **469**(7330): p. 343-9.

14. Cao, Y., *Regulation of germ layer formation by pluripotency factors during embryogenesis.* Cell Biosci, 2013. **3**(1): p. 15.

15. Ng, H.H. and M.A. Surani, *The transcriptional and signalling networks of pluripotency.* Nat Cell Biol, 2011. **13**(5): p. 490-6.

16. Takahashi, K. and S. Yamanaka, *Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.* Cell, 2006. **126**(4): p. 663-76.

17. Balakrishnan, S.K., et al., *Functional and molecular characterization of the role of CTCF in human embryonic stem cell biology.* PLoS One, 2012. **7**(8): p. e42424.

18. Bell, A.C., A.G. West, and G. Felsenfeld, *The protein CTCF is required for the enhancer blocking activity of vertebrate insulators.* Cell, 1999. **98**(3): p. 387-96.

19. Ohlsson, R., V. Lobanenkov, and E. Klenova, *Does CTCF mediate between nuclear organization and gene expression?* Bioessays, 2010. **32**(1): p. 37-50.

20. Zlatanova, J. and P. Caiafa, *CTCF and its protein partners: divide and rule?* J Cell Sci, 2009. **122**(Pt 9): p. 1275-84.

21. Moore, J.M., et al., *Loss of maternal CTCF is associated with peri-implantation lethality of Ctcf null embryos.* PLoS One, 2012. **7**(4): p. e34915.

22. Wan, L.B., et al., *Maternal depletion of CTCF reveals multiple functions during oocyte and preimplantation embryo development.* Development, 2008. **135**(16): p. 2729-38.

23. Fedoriw, A.M., et al., *Transgenic RNAi reveals essential function for CTCF in H19 gene imprinting.* Science, 2004. **303**(5655): p. 238-40.

24. Li, T., et al., *CTCF regulates allelic expression of Igf2 by orchestrating a promoter-polycomb repressive complex 2 intrachromosomal loop.* Mol Cell Biol, 2008. **28**(20): p. 6473-82.

25. Li, E., *Chromatin modification and epigenetic reprogramming in mammalian development.* Nat Rev Genet, 2002. **3**(9): p. 662-73.

26. Santos, F., et al., *Dynamic reprogramming of DNA methylation in the early mouse embryo.* Dev Biol, 2002. **241**(1): p. 172-82.

27. Mikkelsen, T.S., et al., *Dissecting direct reprogramming through integrative genomic analysis.* Nature, 2008. **454**(7200): p. 49-55.

28. Bird, A., *DNA methylation patterns and epigenetic memory.* Genes Dev, 2002. **16**(1): p. 6-21.

29. Wu, H., et al., *Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes.* Science, 2010. **329**(5990): p. 444-8.

30. Mukhopadhyay, R., et al., *The binding sites for the chromatin insulator protein CTCF map to DNA methylation-free domains genome-wide.* Genome Res, 2004. **14**(8): p. 1594-602.

31. Kanduri, C., et al., *Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive.* Curr Biol, 2000. **10**(14): p. 853-6.

32. Han, L., D.H. Lee, and P.E. Szabo, *CTCF is the master organizer of domain-wide allele-specific chromatin at the H19/Igf2 imprinted region.* Mol Cell Biol, 2008. **28**(3): p. 1124-35.

33. Nishana, M. and S.C. Raghavan, *Role of recombination activating genes in the generation of antigen receptor diversity and beyond.* Immunology, 2012. **137**(4): p. 271-81.

34. Alt, F.W., et al., *Ordered rearrangement of immunoglobulin heavy chain variable region segments.* EMBO J, 1984. **3**(6): p. 1209-19.

35. Wilson, A., W. Held, and H.R. MacDonald, *Two waves of recombinase gene expression in developing thymocytes.* J Exp Med, 1994. **179**(4): p. 1355-60.

36. Fehling, H.J., et al., *Crucial role of the pre-T-cell receptor alpha gene in development of alpha beta but not gamma delta T cells.* Nature, 1995. **375**(6534): p. 795-8.

37. Shih, H.Y., et al., *Tcra gene recombination is supported by a Tcra enhancer- and CTCF-dependent chromatin hub.* Proc Natl Acad Sci U S A, 2012. **109**(50): p. E3493-502.

38. Degner-Leisso, S.C. and A.J. Feeney, *Epigenetic and 3-dimensional regulation of V(D)J rearrangement of immunoglobulin genes.* Semin Immunol, 2010. **22**(6): p. 346-52.

39. Hawwari, A., C. Bock, and M.S. Krangel, *Regulation of T cell receptor alpha gene assembly by a complex hierarchy of germline Jalpha promoters.* Nat Immunol, 2005. **6**(5): p. 481-9.

40. Seitan, V.C., et al., *A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation.* Nature, 2011. **476**(7361): p. 467-71.

41. Ying, Q.L., et al., *The ground state of embryonic stem cell self-renewal.* Nature, 2008. **453**(7194): p. 519-23.

42. Goecks, J., et al., *Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences.* Genome Biol, 2010. **11**(8): p. R86.

43. Blankenberg, D., Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J, *Galaxy: a web-based genome analysis tool for experimentalists.* Current Protocols in Molecular Biology, 2010. **Chapter 19**(Unit 19.10.1-21).

44. Giardine, B., et al., *Galaxy: a platform for interactive large-scale genome analysis.* Genome Res, 2005. **15**(10): p. 1451-5.

45. Goecks, J., et al., *NGS analyses by visualization with Trackster.* Nat Biotechnol, 2012. **30**(11): p. 1036-9.

46. Kent, W.J., et al., *The human genome browser at UCSC.* Genome Res, 2002. **12**(6): p. 996-1006.

47. Liu, Z., et al., *Control of embryonic stem cell lineage commitment by core promoter factor, TAF3.* Cell, 2011. **146**(5): p. 720-31.

48. Josh Haimes, M.K., *Demonstration of a ddCq Calculation Method to Compute Relative Gene Expression from qPCR Data.* Thermo Fisher Scientific Tech Note, 2010.

49. Herold, M., M. Bartkuhn, and R. Renkawitz, *CTCF: insights into insulator function during development.* Development, 2012. **139**(6): p. 1045-57.

50. Chaumeil, J. and J.A. Skok, *The role of CTCF in regulating V(D)J recombination.* Curr Opin Immunol, 2012. **24**(2): p. 153-9.

51. Shih, H.Y. and M.S. Krangel, *Chromatin architecture, CCCTC-binding factor, and V(D)J recombination: managing long-distance relationships at antigen receptor loci.* J Immunol, 2013. **190**(10): p. 4915-21.

52. Hernandez-Munain, C., B.P. Sleckman, and M.S. Krangel, *A developmental switch from TCR delta enhancer to TCR alpha enhancer function during thymocyte maturation.* Immunity, 1999. **10**(6): p. 723-33.

53. Wilson, A., J.P. de Villartay, and H.R. MacDonald, *T cell receptor delta gene rearrangement and T early alpha (TEA) expression in immature alpha beta lineage thymocytes: implications for alpha beta/gamma delta lineage commitment.* Immunity, 1996. **4**(1): p. 37-45.

54. Wang, H., et al., *Widespread plasticity in CTCF occupancy linked to DNA methylation.* Genome Res, 2012. **22**(9): p. 1680-8.

55. Kim, T.H., et al., *Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome.* Cell, 2007. **128**(6): p. 1231-45.

56. Stadler, M.B., et al., *DNA-binding factors shape the mouse methylome at distal regulatory regions.* Nature, 2011. **480**(7378): p. 490-5.

57. Shinkai, Y., et al., *Restoration of T cell development in RAG-2-deficient mice by functional TCR transgenes.* Science, 1993. **259**(5096): p. 822-5.

58. Nishida, H., et al., *Histone H3 acetylated at lysine 9 in promoter is associated with low nucleosome density in the vicinity of transcription start site in human cell.* Chromosome Res, 2006. **14**(2): p. 203-11.

59. McMurry, M.T. and M.S. Krangel, *A role for histone acetylation in the developmental regulation of VDJ recombination.* Science, 2000. **287**(5452): p. 495-8.

60. Shih, H.Y. and M.S. Krangel, *Distinct contracted conformations of the Tcra/Tcrd locus during Tcra and Tcrd recombination.* J Exp Med, 2010. **207**(9): p. 1835-41.

61. Lee, B.K. and V.R. Iyer, *Genome-wide studies of CCCTC-binding factor (CTCF) and cohesin provide insight into chromatin structure and regulation.* J Biol Chem, 2012. **287**(37): p. 30906-13.

62. Handoko, L., et al., *CTCF-mediated functional chromatin interactome in pluripotent cells.* Nat Genet, 2011. **43**(7): p. 630-8.

63. Tsumura, A., et al., *Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b.* Genes Cells, 2006. **11**(7): p. 805-14.

64. Trowbridge, J.J., et al., *DNA methyltransferase 1 is essential for and uniquely regulates hematopoietic stem and progenitor cells.* Cell Stem Cell, 2009. **5**(4): p. 442-9.

65. Bell, A.C. and G. Felsenfeld, *Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene.* Nature, 2000. **405**(6785): p. 482-5.

66. Xie, X., et al., *Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites.* Proc Natl Acad Sci U S A, 2007. **104**(17): p. 7145-50.

67. Yu, D.H., et al., *Developmentally programmed 3' CpG island methylation confers tissue- and cell-type-specific transcriptional activation.* Mol Cell Biol, 2013. **33**(9): p. 1845-58.

68. Wilson, G.A., et al., *Resources for methylome analysis suitable for gene knockout studies of potential epigenome modifiers.* Gigascience, 2012. **1**(1): p. 3.

69. Zhong, X.P. and M.S. Krangel, *An enhancer-blocking element between alpha and delta gene segments within the human T cell receptor alpha/delta locus.* Proc Natl Acad Sci U S A, 1997. **94**(10): p. 5219-24.

70. Chen, X., et al., *Integration of external signaling pathways with the core transcriptional network in embryonic stem cells.* Cell, 2008. **133**(6): p. 1106-17.

71. Marks, H., et al., *The transcriptional and epigenomic foundations of ground state pluripotency.* Cell, 2012. **149**(3): p. 590-604.

72. Xu, J., et al., *Pioneer factor interactions and unmethylated CpG dinucleotides mark silent tissue-specific enhancers in embryonic stem cells.* Proc Natl Acad Sci U S A, 2007. **104**(30): p. 12377-82.

73. Reizis, B. and P. Leder, *Expression of the mouse pre-T cell receptor alpha gene is controlled by an upstream region containing a transcriptional enhancer.* J Exp Med, 1999. **189**(10): p. 1669-78.

74. Xu, J., et al., *Transcriptional competence and the active marking of tissue-specific enhancers by defined transcription factors in embryonic and induced pluripotent stem cells.* Genes Dev, 2009. **23**(24): p. 2824-38.

75. Tremblay, M., et al., *Regulation of pT alpha gene expression by a dosage of E2A, HEB, and SCL.* J Biol Chem, 2003. **278**(15): p. 12680-7.

76. Kang, Y., et al., *Dynamic transcriptional regulatory complexes including BORIS, CTCF and Sp1 modulate NY-ESO-1 expression in lung cancer cells.* Oncogene, 2007. **26**(30): p. 4394-403.

77. Ruiz-Llorente, S., et al., *Genome-wide analysis of Pax8 binding provides new insights into thyroid functions.* BMC Genomics, 2012. **13**: p. 147.

78. Zhang, Y., et al., *CCCTC-binding factor acts upstream of FOXA1 and demarcates the genomic response to estrogen.* J Biol Chem, 2010. **285**(37): p. 28604-13.

79. Hark, A.T., et al., *CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus.* Nature, 2000. **405**(6785): p. 486-9.

80. Loukinov, D.I., et al., *BORIS, a novel male germ-line-specific protein associated with epigenetic reprogramming events, shares the same 11-zinc-finger domain with CTCF, the insulator protein involved in reading imprinting marks in the soma.* Proc Natl Acad Sci U S A, 2002. **99**(10): p. 6806-11.

81. Ehrlich, M., et al., *Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells.* Nucleic Acids Res, 1982. **10**(8): p. 2709-21.

82. Illingworth, R.S. and A.P. Bird, *CpG islands--'a rough guide'.* FEBS Lett, 2009. **583**(11): p. 1713-20.

83. Graves, J.A., *5-azacytidine-induced re-expression of alleles on the inactive X chromosome in a hybrid mouse cell line.* Exp Cell Res, 1982. **141**(1): p. 99-105.

84. Bird, A.P. and A.P. Wolffe, *Methylation-induced repression--belts, braces, and chromatin.* Cell, 1999. **99**(5): p. 451-4.

85. Brinkman, A.B., et al., *Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk.* Genome Res, 2012. **22**(6): p. 1128-38.

86. Gardiner-Garden, M. and M. Frommer, *CpG islands in vertebrate genomes.* J Mol Biol, 1987. **196**(2): p. 261-82.

87. Irizarry, R.A., et al., *The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores.* Nat Genet, 2009. **41**(2): p. 178-86.

88. Farthing, C.R., et al., *Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes.* PLoS Genet, 2008. **4**(6): p. e1000116.

89. Jones, P.A., *Functions of DNA methylation: islands, start sites, gene bodies and beyond.* Nat Rev Genet, 2012. **13**(7): p. 484-92.

90. Hellman, A. and A. Chess, *Gene body-specific methylation on the active X chromosome.* Science, 2007. **315**(5815): p. 1141-3.

91. Medvedev, S.P., A.I. Shevchenko, and S.M. Zakian, *Molecular basis of Mammalian embryonic stem cell pluripotency and self-renewal.* Acta Naturae, 2010. **2**(3): p. 30-46.

92. Li, X. and X. Zhao, *Epigenetic regulation of mammalian stem cells.* Stem Cells Dev, 2008. **17**(6): p. 1043-52.

93. Morgani, S.M., et al., *Totipotent embryonic stem cells arise in ground-state culture conditions.* Cell Rep, 2013. **3**(6): p. 1945-57.

94. Habibi, E., et al., *Whole-Genome Bisulfite Sequencing of Two Distinct Interconvertible DNA Methylomes of Mouse Embryonic Stem Cells.* Cell Stem Cell, 2013.

95. Kobayashi, H., et al., *Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks.* PLoS Genet, 2012. **8**(1): p. e1002440.

96. Sleckman, B.P., et al., *Assessing a role for enhancer-blocking activity in gene regulation within the murine T-cell receptor alpha/delta locus.* Immunology, 2001. **104**(1): p. 11-8.

97. Ficz, G., et al., *FGF Signaling Inhibition in ESCs Drives Rapid Genome-wide Demethylation to the Epigenetic Ground State of Pluripotency.* Cell Stem Cell, 2013.

98. Engel, N., J.L. Thorvaldsen, and M.S. Bartolomei, *CTCF binding sites promote transcription initiation and prevent DNA methylation on the maternal allele at the imprinted H19/Igf2 locus.* Hum Mol Genet, 2006. **15**(19): p. 2945-54.

99. Kurukuti, S., et al., *CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2.* Proc Natl Acad Sci U S A, 2006. **103**(28): p. 10684-9.

100. Bacher, C.P., et al., *Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation.* Nat Cell Biol, 2006. **8**(3): p. 293-9.

101. Xu, N., et al., *Evidence that homologous X-chromosome pairing requires transcription and Ctcf protein.* Nat Genet, 2007. **39**(11): p. 1390-6.

CONCLUSION

In this study we shed light on powerful and novel regulatory mechanisms governing rearrangement of the murine T cell receptor gene loci. Within *Tcra*, we have identified a stable region of DNA hypomethylation at the TEA promoter during embryonic and hematopoietic development resulting from stable binding of the transcription factor CTCF at a site ~100 bp upstream of the TEA transcription start site. Removal of CTCF from this site via CTCF knockdown or monoallelic deletion of TEA resulted in a dramatic increase in DNA methylation over 1 kb both upstream and downstream of TEA, while stable fully methylated DNA regions located 800 bp upstream and 300 bp downstream of TEA were unaffected. In addition, the TEA-distal increases in DNA methylation upstream and downstream of TEA occurred on both alleles, suggesting protection from DNA methylation at these regions may involve CTCF-mediated allelic cross-talk.

Recombination of *Tcra* and *Tcrd* loci occur during different stages of thymocytes development (DP and DN, respectively) despite their location in the same genomic region, with less than 4 kb separating the TEA promoter from the upstream TRDV5 gene segment. The Eδ enhancer, which lies 10 kb upstream of TRDV5, activates TRDV5 transcription during the DN stage and is capable of supporting promoter activity up to 55 kb away, an insulation mechanism between these two regions seems a likely necessity. As insulation mechanisms that keep transcription from *Tcra*-associated promoters silent during DN T cell development remain elusive, we propose that our extended region of hypermethylation surrounding the TEA promoter, which stretches from 400 bp downstream of purported insulator BEAD-1 to 600 bp upstream of proximal Jα subunit Jα61, may help segregate Jα promoters from *Tcrd* transcription control elements. Further studies regarding functional significance of the hypermethylated region to Jα transcriptional control are underway,

using transgenic reporters that harbor targeted deletions within this region, and assessing reporter activity using an *in vitro* thymocyte differentiation system.

Vertebrate development requires lineage-specific restriction of a wide swath of genes. Recent findings from the ENCODE project powerfully illustrate the roles of noncoding regulatory elements and their epigenetic modulation in orchestrating developmental gene regulation. Harnessing the power of the epigenome to develop novel therapies will require a much better understanding of how it contributes to both development and disease. Much focus has centered of late on generation of pluripotent cell lines from established somatic cell types through reversion (induced pluripotent stem cells or iPSCs) or somatic cell nucleus transfer into oocytes. Such cells offer astounding medical implications, from growing new organs *in vitro* to providing fundamental insights on cancer establishment and progression. However, both of these techniques remain problematic due in part to partial retention of somatic epigenetic features in the resultant stem cells.

In this study, we've identified a defined and stable epigenetic signature established during early embryonic development and maintained through terminal differentiation. Whether CTCF binding post-fertilization serves to demethylate TEA or binds prior to genome methylation and insulates the promoter remains to be seen. Likewise, future studies will be necessary to determine the extensibility of this priming mechanism to the regulation of other tissue-specific genes, though preliminary analyses of the *Ptcra* enhancer already suggest a similar role for CTCF-mediated insulation there. Our ongoing characterization of the *Ptcra* enhancer, as well as analyses of other CTCF-bound tissue-specific elements, is expected to shed new light on the epigenetic regulation of tissue-specific genes in early development.

Beyond gene priming, my research has generated a new tool for analysis of recombination patterns during thymocytes development, in the form of a battery of T cell lines that have stably generated collections of Dβ-Jβ and Vβ-DJ recombinations. Indeed, these cell lines have helped our laboratory characterize a novel mechanism for promoter regulation within the TCRβ locus, in which repression of the 5'PDβ2 promoter is removed through sufficient accumulation of dsDNA breaks specifically at the DJβ2 gene segment cassette. Recombinase-dependent DNA breaks trigger de-repression by inducing loss of the stress-response factor USF-1 from its binding site within the 12-RSS of Dβ2. Both USF-1 binding and transcriptional repression can be restored through inhibition of DNA-PKcs, an essential component of the DNA repair machinery that completes V(D)J recombination, and a known regulator of USF-1. The physiological ramifications of 5'PDβ2 de-repression have yet to be demonstrated, though work to that end is underway in the laboratory. However, our previous research has shown that promoter location relative to Dβ1 has a dramatic effect on the ability to form DJβ1 joints. Specifically, when promoter activity was redirected downstream of Dβ1 in transgenic reporters, DJβ1 recombination was severely attenuated. We expect a reversal of this phenomenon at endogenous Dβ2. We believe the slow accumulation of DJβ2 joints in mouse and human thymocytes forces initial V-to-DJ recombination to target more rapidly accumulating DJβ1 joints, likely stemming from DJβ2 promoter activity that is located downstream of Dβ2. Since the ability of DNA-PKcs recruited to DJβ2 breaks to redirect promoter activity upstream of Dβ2 would be expected to enhance DJβ2 recombinational accessibility for Vβ assembly, our findings suggest a novel positive feedback loop in which recombination activity augments recombinational accessibility and subsequent usage of Dβ2 in the repertoire. Further

characterization of USF-1 function at *Tcrb* and formal testing of this hypothesis is underway in the laboratory. Additionally, USF-1 has previously been shown to regulate transcription in *Tcrd* and in IgH. We have also begun projects in the laboratory to more carefully assess the potential impact of recombination-mediated USF-1 feedback at each of the other antigen receptor loci. These studies are expected to clarify the mechanics and generalizability of USF-1-mediated dsDNA break response to development of the antigen receptor repertoire.

Antigen receptor gene recombination and expression is a fundamental part of our capacity to respond to colonization by pathogenic organisms. The highly orchestrated manner in which antigen receptor genes are developmentally regulated, and the diverse array of genetic tools already amassed to study lymphocyte development, make the antigen receptor genes powerful models for dissecting the epigenetic programs that govern gene activation during embryogenesis and differentiation. This document presents evidence supporting two novel regulatory mechanisms: gene priming by CTCF-mediated hypomethylation, and DNA damage-mediated feedback of chromatin accessibility to further enzymatic targeting. It is our hope that these exciting findings will help shed light specifically on antigen receptor gene assembly and, more broadly, on the epigenetic programs that govern developmental gene regulation.

APPENDIX

CHAPTER A1

RAG-DSB Feedback Control of Dβ2 Germline Transcription in a CD4/CD8 Double Negative T cell line acts through USF-1-mediated Promoter Repression

ABSTRACT

The adaptive immune repertoire is generated through somatic rearrangement of the genes that encode our antigen receptor proteins. This rearrangement process, termed V(D)J recombination, occurs in a stepwise manner during lymphogenesis, and is rigorously controlled through transcriptional modulation of each antigen receptor gene's chromatin conformation, increasing or reducing the accessibility of discrete intragenic regions to recombinase. During αβ T cell development, *Tcrb* assembly involves simultaneous recombination of two DβJβ cassettes, followed closely by recombination of a more distal upstream Vβ segment to the first assembled DJβ joint. Initial Vβ-to-DJβ joints predominantly involve the upstream DJβ (DJβ1) cassette. However, usage of DJβ1 and downstream DJβ2 in the circulating repertoire is roughly equal, as secondary Vβ-to-DJβ2 rearrangements serve to rescue frameshifted primary joints involving DJβ1. We previously showed that recombinational accessibility of DJβ1 is governed by transcriptional activation at the PDβ1 promoter positioned immediately 5' of Dβ1, while Dβ2 is flanked by a downstream promoter (3'PDβ2) and a 5'PDβ2 that is repressed prior to recombination. As part of a larger study to define the molecular mechanism of 5'PDβ2 repression and its release upon recombination, I constructed a series of T cell lines with distinct DJβ recombination profiles. I found that transcriptional activity of 5'PDβ2 increased in cells that specifically harbored DJβ2 rearrangements. This increase correlated with reduced binding by the stress-response transcription factor USF-1 at the 5'PDβ2 repressor site localized to the Dβ2 12-RSS. Finally, USF-1 binding and 5'PDβ2 repression was restored in the rearranged cell lines upon addition of the DNA-PKcs inhibitor, Nu7026, mirroring results our laboratory obtained in gamma-irradiated thymocytes. Together, these data elucidate a novel

feedback response whereby DJβ2 rearrangement that is initially delayed due to transcriptional activity 3' of Dβ2 triggers derepression of 5'PDβ2. This mechanism allows DJβ2 to rescue nonfunctional primary rearrangements and maximize the diversity of the αβ T cell repertoire.

A1.1 Introduction

Adaptive immunity requires antigen receptor gene assembly through a somatic rearrangement process called V(D)J recombination. The recombinase holoenzyme contains an array of ubiquitous DNA repair proteins, together with the lymphocyte-specific recombination activating gene proteins 1 and 2 (RAG 1 and RAG2), which target conserved recombination signal sequences (RSSs) that flank each V, D or J coding segment [1]. Each antigen receptor gene is assembled in a regulated, stepwise manner that is intrinsically linked to lymphocyte developmental progression [2]. Hierarchical regulation of V(D)J recombination is achieved through the combined impact of enzyme-substrate specificity provided by RAG-1/2 and their target RSSs and by regulating the chromatin accessibility of individual RSSs. So-called recombinational accessibility remains poorly understood [3]. However, evidence to date suggests that transcriptional promoters interspersed among antigen receptor gene segment clusters, complexed with distal enhancers, modulate local chromatin architecture and accessibility to recombinase via transcription of germline or unrearranged gene segments [3-5].

αβ T lymphocytes require sequential recombination of *Tcrb* then *Tcra* genes during CD4/CD8 double negative (DN) and double positive (DP) stages, respectively. *Tcrb* assembly proceeds in a stepwise manner involving independent D-to-J recombination at two D-Jβ cassettes, followed by Vβ segment recombination to the newly-formed DJ joint. Recombinational accessibility of the upstream DJβ1 cassette is governed in part by the promoter PDβ1, located upstream of Dβ1 [6]; when this promoter is deleted, nucleosomal phasing across the Dβ1 RSS is altered and recombination involving this cassette is impaired [7-9]. Accessibility of the downstream DJβ2

cassette involves a more complex array of genetic elements, containing promoters both upstream and downstream of Dβ2 and a putative repressor element centered on the Dβ2 12RSS [4, 10, 11].

Germline transcription also differs significantly between the two DJβ cassettes. While PDβ1 drives Dβ1 transcription in the presence or absence of DJβ1 recombination, evidence suggests transcription at DJβ2 is initiated from the downstream PDβ2 in the absence of DJ2 recombination but begins from the upstream PDβ2 upon deletion of the downstream promoter by excision during DJβ2 rearrangement [4, 11]. However, efforts to demonstrate the sensitivity of Dβ2 promoter usage to locus recombination have been limited in our laboratory by the transient nature of DJβ joints in developing thymocytes (functional DJβ joints are rapidly resolved by V-to-DJβ recombination). Toward this end, we have constructed a series of RAG-deficient T cell lines that possess stable profiles of Dβ-Jβ recombinations. These cell lines, confirmed by both PCR and sequencing, were used to clarify the role of recombination in the differential activation of Dβ2 germline promoters.

A1.2 Materials and Methods

A1.2.1 Cells and Antibodies.

The $Rag1^{-/-}$ $P53^{-/-}$ pro-T cell line, P5424, was originally derived from C57BL/6 mice [12]. P5424 cells were cultured at 37°C, 5% CO_2 in RPMI 1640 medium supplemented with 10% fetal calf serum, 2 mM L-glutamine, penicillin/streptomycin (50 units/ml), and 50 μM β-mercaptoethanol. Balb 3T3 fibroblasts were maintained in DMEM supplemented with 10% fetal calf serum, HEPES (N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid), 2 mM L-glutamine, and penicillin/streptomycin (50 units/ml). Antibodies used in the following studies included: H3K9ac (Ab10812; Abcam), H3K4me2 (07-030; Millipore), H3K4me3 (39159; Active Motif), USF-1 (sc-229; Santa Cruz Biotechnology) and isotype-matched rabbit IgG (10-4102; Rockland Immunochemicals).

A1.2.2 Stable recombination through transient transfection.

7×10^6 cells were electroporated (300 μl serum free RPMI, 260 V/950 mF) with 5 μg p-phosphoglycerate kinase (PGK)–RAG1, PGK-RAG2 and 5 μg pMACS4.1 (Miltenyi Biotech) and allowed to recover overnight. Transfectants were enriched using Dynabeads FlowComp Mouse CD4 magnetic beads (Invitrogen) and retransfected as above prior to subcloning. Successfully rearranged subclones were screened by PCR. The integrity of all constructs and DJβ rearrangement profiles of all transfectants were confirmed by sequencing (Eton Biosciences).

A1.2.3 Ionizing radiation treatments.

To model RAG-induced DSBs, P5424 cells (10^6/ml RPMI growth medium) were exposed to increasing doses of ionizing radiation using a Gammacell 220 cobalt-60 irradiator (MDS Nordion) or plated in RPMI supplemented with 3 μM etoposide for 4 h. After genotoxic insult, cells were washed in PBS, replated in normal medium, and allowed to recover for 18–24 h prior to chromatin

and RNA extraction. For kinase inhibition studies, cells were pretreated with 0.1% DMSO, SB203508 (10 μM; Calbiochem), Nu7026 (10 μM; Cayman Chemical), or Ku55933 (15 μM; Selleck Chemicals) for 40 min at 37°C before genotoxic insult.

A1.2.4 CpG methylation.

Sodium bisulfite modification of DNA was performed using EpiTect Bisulfite Kits (Qiagen) according to the manufacturer's instructions. Methylation was quantitated by the Sequenom MassARRAY platform with EpiTYPER analysis software (Sequenom). EpiDesigner software (Sequenom) was used to design T7-tagged and matched primers to CpG-deficient targets across Dβ2. PCR was performed using HotStarTaq (Qiagen) and products were processed using MassCLEAVE as per the manufacturer's protocol (Sequenom). Resulting fragmented transcripts were spotted onto SpectroCHIPS for mass spectrometry analysis on a MassARRAY instrument (Sequenom) to quantify the methylated fraction in each amplicon.

A1.2.5 Chromatin immunoprecipitation.

Chromatin was prepared from formaldehyde cross-linked P5424 as described [13]. Specifically, protein-DNA complexes were cross-linked using formaldehyde (Sigma, 1% final concentration) added drop-wise to cells in growth medium followed by gentle shaking for 10 minutes at room temperature. Drop-wise addition of glycine (125 mM final concentration) followed by gentle shaking for 5 minutes at room temperature served to stop cross-linking. Cross-linked cells were pelleted (1100 RPM, 4^{o}C, 5 minutes), washed with 10 mL PBS, and final cell pellets were supplemented with 5 μl PMSF (100 mM stock) and 5 μl 100X protease inhibitor cocktail (PIC, Sigma P8340) and resuspended in 1 mL of lysis buffer (10 mM Tris-HCl (pH 7.5), 10 mM NaCl, 3 mM $MgCl_2$, 0.5% NP-40). Cells were lysed for 30 minutes on ice and then dounce homogenized for 10 strokes.

Nuclei were pelleted by microcentrifugation (5000 x G, 10 minutes, 4^{o}C), and gently but thoroughly resuspended in 500 uL MNase digestion buffer (50 mM Tris-HCl (pH 7.5), 1 mM $CaCl_2$, 4 mM $MgCl_2$, 0.33 M Sucrose, 5 mM Sodium Butyrate) supplemented with 1 μM PMSF and 1X PIC (1.04 mM AEBSF, 800 nm aprotinin, 40 μm bestatin, 14 μm E-64, 20 μm leupeptin, 15 μm pepstatin A). DNA content in the preparation was assessed in a 1:10 dilution of nuclei resuspension in 0.1% SDS using a Nanodrop 1000 Spectrophotometer (ThermoScientific).

Nuclei were resuspended in MNase digestion buffer at a concentration of 500 μg/mL and incubated at 37^{o}C for 5 minutes followed by addition of Micrococcal Nuclease (Sigma) at a concentration of 10 U/200 μg nuclei. Digestion proceeded for 7 minutes at 37^{o}C following MNase addition, with periodic quick vortexing to facilitate homogenous digestion. Digestion was stopped through addition of EDTA (10mM final) and incubation on ice for 10 minutes. Nuclear debris were pelleted by centrifugation (10 minutes at 13,000 RPM in a microcentrifuge), and supernatant containing sheared chromatin was stored at -20 or -80^{o}C

Chromatin was subsequently assayed by chromatin immunoprecipitation (ChIP) as described [13]. To elaborate, paramagnetic Dynabeads (Dynal) coupled to recombinant Staphylococcal Protein A (5 μL/IP) were added to a 1.5 mL microcentrifuge tube, and washed twice using 1x RD buffer, consisting of 1x RIPA (10 mM Tris-HCl (pH7.5), 150 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1% Triton X-100, 0.1 % SDS, and 0.1% NaDeoxycholate) with supplementary addition of sheared salmon sperm DNA (ssDNA) (2.5 mg/mL) and BSA (50 mg/mL). During washes, beads were agitated via alternating placement against a strong magnet. Beads were resuspended in RD (50 μL/IP) containing PMSF and PIC, transferred to sterile PCR strip tubes, and incubated 1 hr to overnight with antibody of interest (1-5 μg) with rotation at 4^{o}C. Conjugated antibody-bead complexes were washed twice in RD,

resuspended in 90 μl RD containing PMSF/PIC and mixed with 10 μL of sheared chromatin (~10^6 cell equivalents). Immunoprecipitations were incubated at 4°C with rotation overnight and complexes were captured using a strong magnet and washed (via rapid movement of the strip tubes back and forth across a bar magnet to ensure bead agitation) as follows: 2 washes with 1x RIPA, 2 washes with a high salt wash (500 mM NaCl, 0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl at pH 8.0), 2 washes with 250 mM LiCl, 1% NP40, 1% NaDeoxycholate, 1 mM EDTA, 10 mM Tris-HCl at pH 8.0, and 2 washes with TE (Tris HCl-EDTA) at pH 8.0. Note: Washes were conducted using 180 μL of the respective wash buffer per tube.

To elute protein-DNA complexes after washing, beads were resuspended in 100 μL of 100 mM $NaHCO_3$ (37°C) and mixed by gentle vortex for 15 min at room temperature. After magnetic capture, protein-DNA eluates were transferred to fresh PCR strip tubes containing 2 μL 5 M NaCl for crosslink reversal (15 min at 95°C). Untreated input controls were prepared as follows: 10 μL of input chromatin was diluted in 90 μL 100 mM $NaHCO_3$, 2 μL 5 M NaCl, and carried through the remainder of the process starting with crosslink reversal. Following crosslink reversal, Proteinase K (Ambion, 10 μg/mL final) was added to each sample and protein digestion was carried for 60 minutes at 37°C. Proteinase K digestion was then stopped through the addition of PMSF (2 mM final) and DNA was purified using UPrep columns (Genesee) according to the manufacturer's instructions, with subsequent elution in 100 μL H_20.

Samples (4 μl) were subjected to QPCR in a MyIQ real-time thermal cycler (Bio-RAD) using 1x SensiMix *Plus* (Quantace) in triplicate reactions. Primers and annealing temperatures for chromatin immunoprecipitation are shown (Table I). Cycling parameters for 20 μl reactions were

95°C for 10 min, followed by 50 cycles of 95°C for 20 s; annealing for 30 s; and 72°C for 30 s, followed by a final 3 minutes at 72°C and melt curve analysis. Average fold enrichment in bound fractions was calculated for triplicate amplifications as previously described [14]. To elaborate: during amplification, fluorescence signal intensity was plotted as a function of cycle number. Threshold cycle (Ct) was automatically determined and Ct values were screened for reproducibility across triplicate reactions using 1-cycle fluctuations as an appropriate cutoff. Assuming a polymerase amplification rate of 100%, fold enrichment was then determined using the following equation: enrichment = $2^{(Ctinput- CtIP)}$. Enrichment signals were further normalized to that obtained for isotype-matched control antisera.

A1.2.6 RNA extraction and RT-PCR.

RNA was extracted using RNAzol RT (MRC) according to the manufacturer's instructions. 1–3 μg RNA was reverse transcribed using Moloney murine leukemia virus reverse transcriptase (MMuLV RT, New England Biolabs) and oligo d(T) primers (2 μM). The resultant cDNAs were amplified using intron-spanning primer pairs and melting temperatures described in Table 1 by quantitative PCR (as described above). The relative abundance of cDNAs was quantified following QPCR by $\Delta\Delta C_T$ normalization to matched untreated controls and standardized for loading variations by comparison with values obtained for β-actin.

A1.3 Results

A1.3.1 Isolation of mouse T cell lines possessing an array of Dβ-to-Jβ recombination patterns.

To determine the effect of different recombination profiles on transcription across the TCRβ DβJβ cassettes, we utilized a CD4/CD8 double negative T cell line called P5424. P5424 was derived from a mouse (C57Bl6) pro-T cell with both recombinase subunit RAG-1 and tumor suppressor gene p53 removed by homologous transformation, which renders these cells both immortal and incapable of independent recombination [12]. To generate different stable patterns of Dβ-Jβ recombination, it was necessary to transiently provide the missing recombinase subunit. The natural properties of P5424 and lymphocytes in general (small, non-adherent, and round) make them inherently difficult to transfect. To overcome these issues, we used a transient transfection and selection methodology that temporarily provided RAG-1 along with CD4 (Figure 1). Transfection efficiencies proved to be very low. However, since our goal was to isolate cells with only a minimum number of recombinations, the relative inefficiency worked in our favor. After nine rounds of serial transfection, the enriched population was subcloned and screened for Dβ-Jβ recombination by PCR. Using primers located upstream of both Dβ segments and downstream primers within Jβ clusters of each cassette, we identified a number of stable recombination patterns, including cells that carried monoallelic rearrangements of Dβ1Jβ1 but not Dβ2Jβ2, Dβ2Jβ2 but not Dβ1Jβ1, or both Dβ1Jβ1 and Dβ2Jβ2 (Figure 2). A single clone was identified that carried biallelic Dβ1Jβ1 rearrangements in the absence of any Dβ2Jβ2 joints. Further PCR analysis revealed that many of these clones also carried TCRγ and TCRδ recombinations as well (summarized in Table 2). Finally, DJβ joints were confirmed by sequence analysis (representative sequences for clones c-20 and c-22 are shown in Table 3).

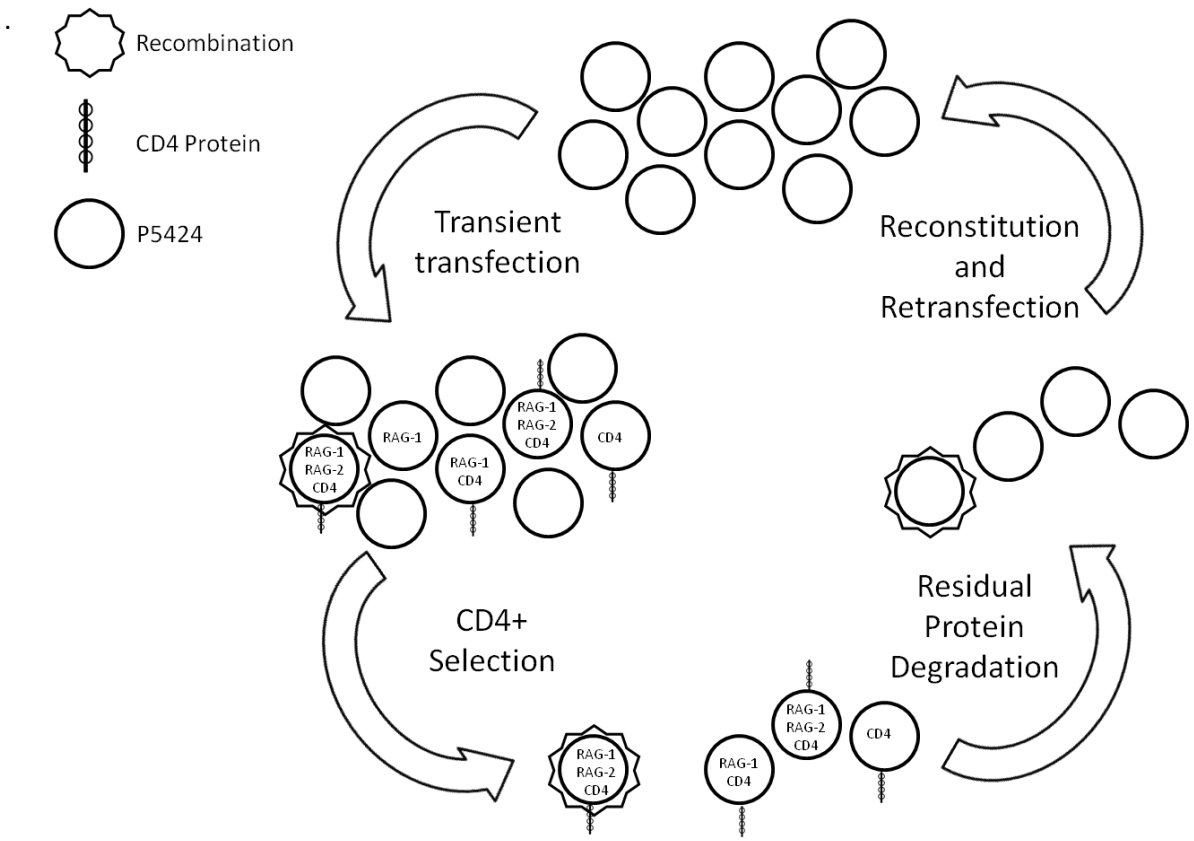


Figure A1.1: Serial transfection and selection of RAG-competent P5424. P5424 cells were transiently transfected with plasmids containing full-length RAG-1 and RAG-2, along with an expression vector for human CD4 and allowed to recover overnight. Magnetic anti-hCD4 beads were used to enrich for transfected cells. After 2-4 days, cells were again transfected and selected for hCD4 expression.

Table A1.1 PCR primer sets.

Amplicon	Strand	Sequence	Tm
Dβ1	S	GAT CCA GAA TGC TTT CAC G	63°C
	A	CTG CAT CCT TTG CTG CTA	
Dβ2	S	CCC TCT CAG TCA GAC AAA CCT CTC	63°C
	A (gl)	CTT CCA GTT GAA TCA TTG TGG CCC C	
	A (DJ2.5)	CAA AGT ACT GGG TGT CTT GGT TAC	
Jβ2.2	S	CAG GAC TGT GCA AAC ACC GGG CAG	63°C
	A	TCC CAG CAT CCA ATC CAG AAA TCC	
Jβ2.4	S	GCA TTT CCC AGG CTA AGG	56°C
	A	CCC CAG AAC CCA ACA CAA	
3'Dβ1 RNA	S	GGC TAC CTC ACT TTG ATG	63°C
	A	CCC CAG GCC TCT GCA TG ATG TTC TGT GTG	
3'Dβ2 RNA	S	GGT CTT ATA ACA TCT ATG CAT C	63°C
	A	CTG TGT GAC AGG TTT GGG TGA GCC CTC TG	
βactin	S	GTG GGC CGC CCT AGG CAC CAG	56°C
	A	CTC TTT GAT GTC ACC CAC GAT TTC	
DJβ2 REC	S	GTA GAC ACC TGT GGA GAA GAA ACT	63°C
	A	TGA GAG CTG TCT CCT ACT ATC GAT	
USF-1 RNA	S	ACAAAGTCCTCCGAAAAAAGACCG	63°C
	A	GGGTCCTCTCCAGTAGCCACTGCGC	

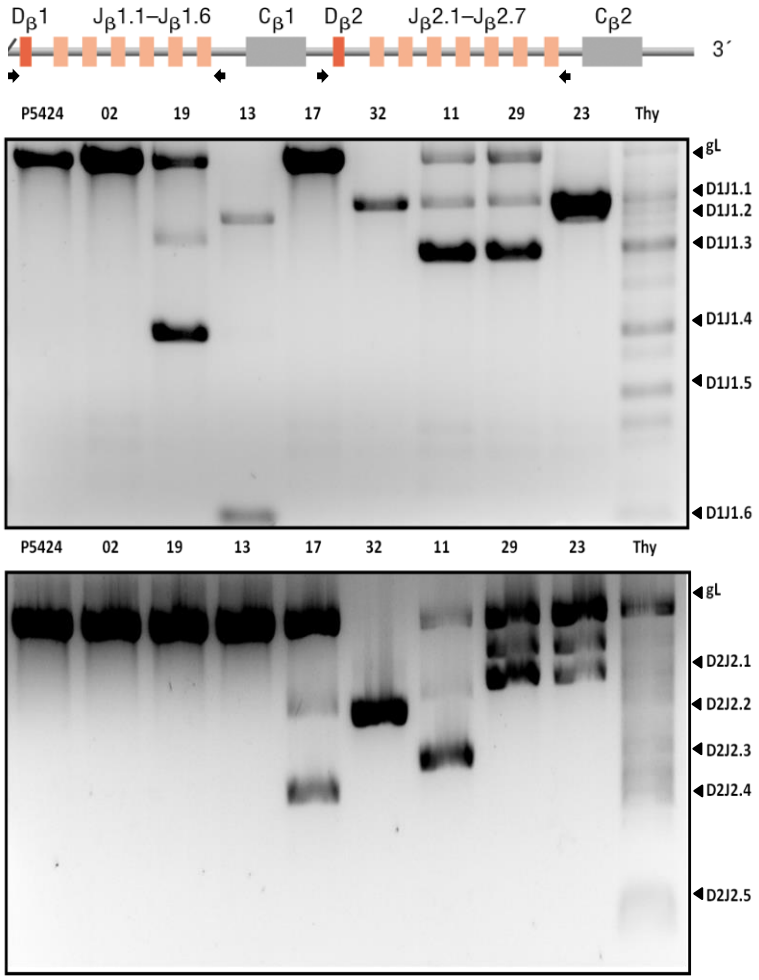


Figure A1.2: ***Tcrb*** **rearrangements in individual RAG-P5424 clones. Using forward PCR primers located upstream of Dβ1 and Dβ2 and reverse primers downstream of the final associated J for each cassette (illustrated in top panel), we were able to identify each possible Dβ1-Jβ1 and Dβ2-Jβ2 rearrangement in P5424 subclones following serial transient transfection with RAG expression plasmids. Rearrangement profiles of the D1J1 (upper panel) and the D2J2 cassettes (lower panel) in each of the indicated clones are shown. Results were confirmed by sequence analysis.**

Table A1.2. Tcr rearrangements in P5424-RAG clones.

Clone	*Tcrb* rearrangements	Dβ1	Dβ2	*Tcrd* rearrangements	*Tcrg* rearrangements
P5424	D1gl D2gl	Gl	gl	gl	gl
O2	D1gl D2gl	Gl	gl	gl	gl
36	D1gl D2gl	Gl	gl	gl	gl
13	D1rec D2gl	J1.2/J1.6	gl	V1:J1	V5:J1
19	D1rec D2gl	J1.4	gl	V4:J1, V4:J2	gl
c20	D1rec D2gl	J1.2/J1.4	gl	gl	gl
17	D1gl D2rec	Gl	J2.4	V1:J1	V3:J1, V5:J1
32	D1rec D2rec	J1.1	J2.2	V4:J1	gl
20	D1rec D2rec	J1.5/J1.6	J2.4	gl	V2:J1
29	D1rec D2rec	J1.3	J2.1	V4:J1, V4:J2	V2:J1, V3:J1, V5:J1
11	D1rec D2rec	J1.3	J2.3	gl	gl
23	D1rec D2rec	J1.1/J1.2	J2.1	V4:J2	V2:J1, V3:J1
c22	D1rec D2rec	J1.2/J1.3	J2.5	gl	gl

Table A1.3. D-to-J recombinant sequence in clones c20 and c22.

Dβ1Jβ1 cluster:	
P5424	Dβ1—GGGACAGGGGGCCACGGTGATTCAATTCTATGG
P5424-c20	a) Dβ1—GGGACAGGGGG* ---t--- TTTCCAACGAAA—Jβ1.4
	b) Dβ1—GG********** ------- **AACTCCGACT—Jβ1.2
P5424-c22	a) Dβ1—GGGACAGGG*** --agg-- ***ACTCCGACT—Jβ1.2
	b) Dβ1—GGGACAGGGG** ---tc-- TTCTGGAAATAC—Jβ1.3
Dβ2Jβ2 cluster:	
P5424	Dβ2—GGGACTGGGGGGGCCACAATGATTCAACTGGAA
P5424-c20	a) Dβ2—GGGACTGGGGGGGCCACAATGATTCAACTGGAA
	b) Dβ2—GGGACTGGGGGGGCCACAATGATTCAACTGGAA
P5424-c22	a) Dβ2—GGGACTGGGGGGGCCACAATGATTCAACTGGAA
	b) Dβ2—GGGACTGGGG** ---t--- AACCAAGACACC— Jβ2.5

*Dβ coding sequence in parental P5424 is underlined.

A1.3.2 Transcriptional activation of 5'PDβ2 is upregulated on both germline and rearranged alleles in cells possessing at least one Dβ2-Jβ2 joint.

After confirming and characterizing the DβJβ recombination pattern of each RAG-P5424 clone via PCR and sequencing, we prepared RNA samples to examine Dβ2 germline transcription. As noted above, Dβ2 is flanked by two promoters: one immediately upstream (5'PDβ2) and one downstream between Dβ2 and Jβ2.1 (3'PDβ2) [4]. Prior to *Tcrb* recombination, only 3'PDβ2 is active, while analyses of DP thymocytes that carry DJβ2 joints (and consequently have deleted 3'PDβ2) show strong activation of 5'PDβ2 [4]. This led us to speculate that repression of 5'PDβ2 is mediated by elements deleted by D-to-J joining. However, luciferase reporter assays and EMSA analyses both pointed to a single repressor element that binds the pleiotropic stress-response protein Upstream Stimulatory Factor-1 (USF-1) within the 5' Dβ2 12-RSS (used for V-to-DJ recombination, and not deleted by D-to-J joining) [11]. Previous studies have shown that USF-1 activity in skin keratinocytes is modulated in response to UV light-induced DNA lesions. We speculated that a similar response to RAG-mediated DNA damage might underlie the association between DJβ2 recombination and 5'PDβ2 activation. If so, we would expect to find in our panel of differentially rearranged P5424 clones that 5'PDβ2 activation could be uncoupled from DJβ2 recombination of the same allele. To that end, we designed an RT-PCR strategy that would only amplify message that read across the unrearranged Dβ2 coding segment and its flanking RSSs.

As shown in Figure 3, transcription initiating immediately 5' of Dβ2 dramatically increased upon accumulation of a single Dβ2 recombination in clone 17 (50- and 100-fold over levels observed for constant regions 2 (Cβ2) and 1(Cβ1), respectively, neither of which was significantly affected by

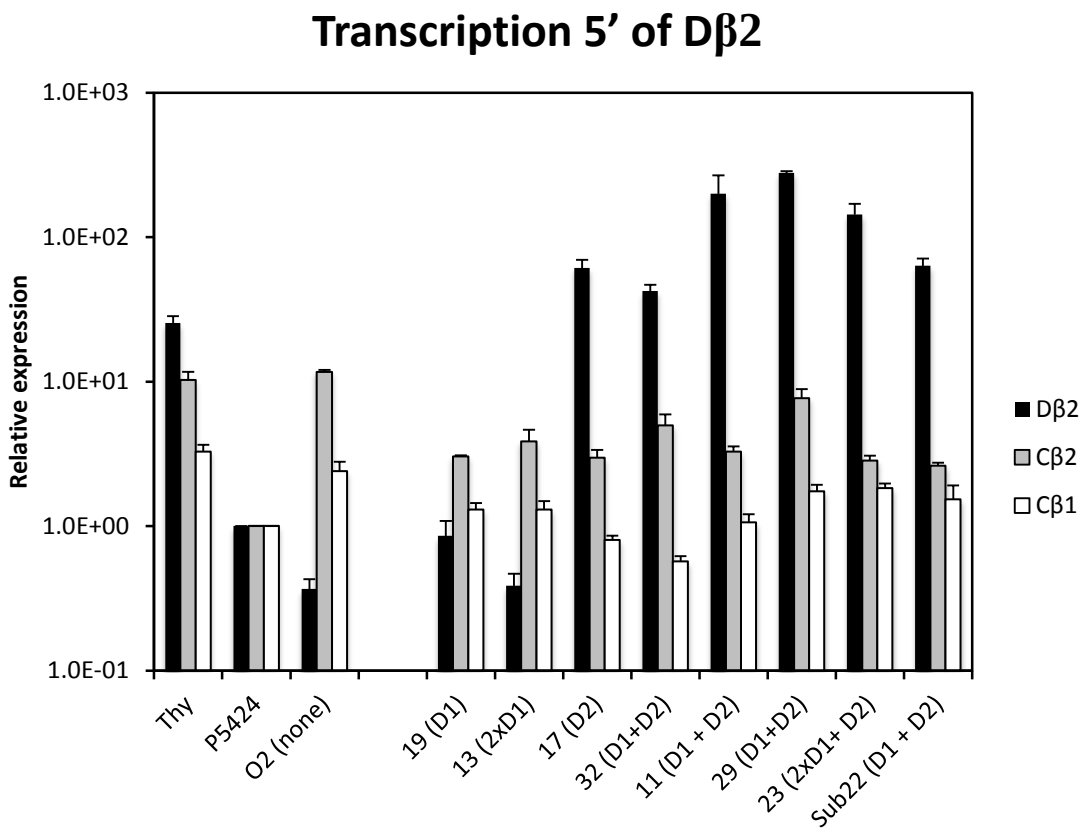


Figure A1.3: Transcription upstream of Dβ2 increases upon D2J2 recombination. Quantitative RT-PCR of germline transcripts initiating immediately 5' of Dβ2 (black bars) was compared with levels of total (germline and rearranged) transcription across Cβ1 (white bars) and Cβ2 (gray bars) in wild-type thymocytes, parental P5424 and the RAG-P5424 clones. The rearrangement status of each clone is indicated in parentheses. Amplification signals for each target are normalized to β-actin, and expressed relative to that in parental P5424. Bars represent mean relative expression (± SEM), n = 3.

Tcrb assembly). In sharp contrast, 5'Dβ2 transcription was not induced by either mono- (clone 19) or bi-allelic (clone 13) generation of DJβ1. Likewise, the presence of DJβ1 joints in addition to DJβ2 joints did not significantly augment 5'Dβ2 transcription levels beyond that obtained in the presence of DJβ2 joints alone (compare clone 17 to clones 32, 11, 29, 23 and sub22), suggesting that recombination outside of the DJβ2 cassette may in and of itself be insufficient to remove repression of the 5'PDβ2 promoter [4, 11]. Cells from wild-type C57Bl6 thymii showed an intermediate (20-fold) increase in transcription 5' of Dβ2, consistent with a mixed population in which some percentage of the cells have not undergone Dβ2-Jβ2 recombination. Clone O2, a cell subcloned from the same transfection repeat as the others but lacking *Tcrb* recombination, displayed a pattern of transcription similar to parental P5424. This confirmed that 5'Dβ2 transcription was unlikely to derive from the process of repeated transient transfections and selections we employed to isolate the rearranged clones.

A1.3.3 Differences in transcription at Dβ2 are not due to epigenetic alterations.

Previous reports from our laboratory showed that 5'Dβ2 is bound by a variety of transcription factors in its repressed state in DN thymocytes and P5424, which suggests at least some level of accessibility prior to promoter activation [10]. To measure the effect of recombination on epigenetic modifications associated with accessibility, we used bisulfite conversion to map the methylation of CpG dinucleotides in the vicinity of Dβ2 for P5424 clones c20 and Sub22 (c22). Both of these clones carry bi-allelic DJβ1 joints; while c20 lacks any DJβ2 rearrangements, c22 harbors one germline DJβ2 and one Dβ2-Jβ2.5 joint (Table 3). Of eight CpGs within 600 bp of Dβ2, all were largely unmethylated (<15% methylation) in parental P5424, c20 and c22 (Table 4). In contrast, methylation levels were significantly elevated in Balb 3T3 fibroblasts at all but the 2 CpGs most

proximal to Dβ2, which remained unmethylated. Protection of the Dβ2 CpGs in such divergent cell lineages suggests that, unlike surrounding sequence, the Dβ2 coding segment may be universally protected against DNA methylation, consistent with its critical regulatory role. Regardless, the data clearly demonstrate that, unlike 5'Dβ2 transcription, CpG methylation patterns upstream of Dβ2 are fixed prior to locus assembly and consequently do not contribute to 5'PDβ2 induction upon DJβ2 recombination.

Beyond CpG methylation, covalent histone modifications have been tightly linked to the chromatin opening and transcriptional status of genes. In particular, acetylation of histone H3 at lysine 9 (H3K9ac) and di and tri-methylation of H3 lysine 4 (H3K4me2 and H3K4me3) strongly correlate with chromatin opening and transcriptional activity, respectively, while tri-methylation of H3 lysine 27 (H3K27me3) strongly correlates with transcriptional silencing [15]. Allele-specific PCR analyses of c22, with one rearranged Dβ2 joint and an unrearranged Dβ2 cassette, presented a unique opportunity to look at histone modification patterns upon Dβ2 recombination. Using chromatin immunoprecipitation (ChIP) and primer sets that selectively amplified either germline Dβ2 or the Dβ2-Jβ2.5 joint, we assessed c22 for levels of H3K9ac, H3K4me and H3K27me enrichment at Dβ2 of each *Tcrb* allele (Fig. 4). As with CpG methylation levels, histone modifications did not differ significantly at Dβ2 between the two alleles, which both showed elevated levels of H3K9ac, H3K4me2 and me3, but lacked detectable tri-methylation of H3K27. The "transcriptionally active" pattern of high H3K4 methylation and low H3K27 methylation even on the germline allele appears at first blush to be inconsistent with our transcriptional findings. However, we note that low levels of 5'Dβ2 transcription are indeed detected in parental cells (Fig. 3). Additionally, because ChIP resolution is limited to the shearing size of the chromatin template, and repressed 5'PDβ2

Table A1.4. Percent methylation of CpG dinucleotides surrounding Dβ2.

Cell line	CpG positions*							
	-453	-374	-15	+178	+251	+422	+464	+529
B3T3	0.33	0.48	0.07	0.05	0.18	0.44	0.80	0.21
P5424	0.07	0.04	0	0.02	0.04	0	0.08	0.04
P5424-c20	0.09	0.05	0.08	0.02	0.02	0.02	0.07	0.01
P5424-c22	0.14	0.05	0	0.01	0.05	0.05	0.13	0.05

*numbering relative to the first coding nucleotide of Dβ2

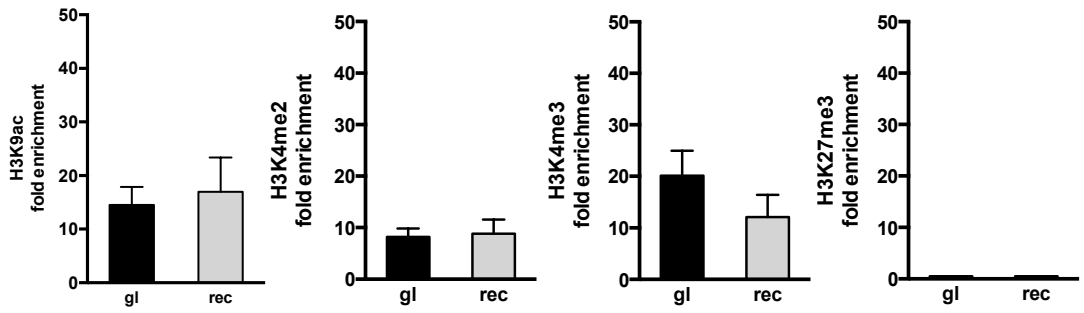


Figure A1.4: Repression does not impact histone modifications at Dβ2. Chromatin from the c22 subclone of P5424 was immunoprecipitated with the indicated antibodies against covalently modified histone H3. Resultant DNAs were analyzed by QPCR for histone modifications at proximal to Dβ2 on the germline (gl, black) allele or the allele that carries a Dβ2Jβ2.5 recombination (rec, grey). Enrichment was calculated relative to pre-IP input control levels and was normalized against signals obtained with non-specific IgG control antibodies. Bars indicate means (± SD, n = 3), and are representative of 2 experiments with independent chromatin preparations.

sits only 600 bp upstream of transcriptionally active 3'PDβ2 in the germline cassette, our ChIP analyses at Dβ2 likely reflect the modification patterns at both promoters. Taken together, our findings provide strong evidence that repression of 5'PDβ2 in germline DJβ2 cassettes is not mediated by epigenetic silencing.

A1.3.4 USF-1 binding at the 5'PDβ2 repressor element is eliminated upon Dβ2 recombination

Work of others in our laboratory revealed that the element of 5'PDβ2 repression localized to an E-box within the Dβ2 12-RSS spacer (bases -17 to -8, relative to the start of the Dβ2 coding sequence) and that the protein Upstream Stimulatory Factor – 1 (USF-1) bound strongly and specifically to this E-box *in vitro* and *in vivo* [11]. To determine if USF-1 binding correlates with Dβ2 recombination, we conducted ChIP on both parental P5424 and c22 (Fig. 5). While pronounced USF-1 enrichment at 5'PDβ2 was present in P5424, this enrichment was completely absent from the D2J2.5 joint (data not shown), as well as from the germline Dβ2 segment of c22 (Fig. 5A). To exclude the possibility that reduced USF-1 binding in c22 stems from a downregulation in USF-1 expression, we assessed USF-1 transcription levels in P5424, c20 and c22 by RT-qPCR (Fig. 5B). Unlike our ChIP findings, we observed no significant difference in USF-1 mRNA levels between P5424 and the c20 and c22 subclones. This observation is consistent with USF's ubiquitous distribution in mammalian tissues [16] and provides evidence against down-regulation of USF-1 expression as a factor in 5'Dβ2 repression or removal thereof.

A1.3.5 Inhibition of DNA-PKcs restores USF-1 binding and 5'Dβ2 repression in c22

USF-1 is best known for gene regulation in response to stresses such as UV radiation and starvation [17]. In lipogenesis, USF-mediated activation of the fatty acid synthase promoter is dependent on USF-1 phosphorylation by DNA-dependent protein kinase (DNA-PK). Phosphorylation

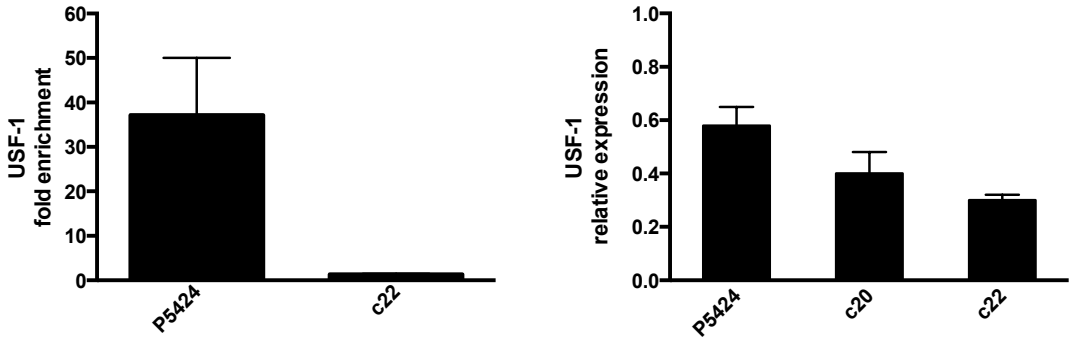


Figure A1.5: USF-1 binding at germline Dβ2 is diminished in c22 despite abundant USF-1 transcription. (Left panel) USF-1 binding at Dβ2 was analyzed by ChIP of chromatin from the P5424 and P5424-c22 cell lines. Antibody-dependent enrichment over input control is expressed relative to nonspecific IgG as mean ± SD (n = 3), and is representative of 2 independent experiments. (Right panel) QRT-PCR of USF-1 mRNA in RAG-P5424 subclones. Bars represent means (± SD, n = 3). Relative signals were calculated by $\Delta\Delta C_T$ and normalized to β actin controls.

allows USF-1 to recruit a variety of binding partners including p300 [18]. The catalytic subunit of DNA-PK, called DNA-PKcs, also plays an instrumental role in coding joint formation and protein phosphorylation during V(D)J recombination, to the extent that absence of DNA-PKcs leads to severe combined immunodeficiency (scid) [19]. Work by others in our laboratory demonstrated that DNA-PKcs-deficient mice show persistent USF-1 binding at Dβ2, despite an accumulation of RAG-mediated double stranded DNA breaks (DSBs) in the *Tcr* loci [11]. Conversely, DSBs induced by ionizing radiation treatment of either freshly isolated thymocytes or P5424 cells induced loss of USF-1 binding from the Dβ2 repressor sequence in these cells, and relieved repression of transcription at 5'PDβ2 [11]. Both USF-1 binding and 5'PDβ2 repression were restored by addition of the DNA-PKcs inhibitor Nu7026, directly implicating DNA-PKcs in regulation of 5'PDβ2 repression [11].

DNA-PKcs activity is downregulated following DSB repair; however, we found that USF-1 displacement and 5'PDβ2 activation persist in the rearranged P5424 clones long after transient RAG expression has been lost. If loss of repression is regulated in *trans* by DNA-PKcs-induced USF-1 phosphorylation, then the possibility remains that c22's constitutive de-repression of 5'PDβ2 is due to long-term recruitment of DNA-PKcs initiated by recombinational accumulation and/or the repeated transfection cycle. To determine if DNA-PKcs-mediated relief of 5'PDβ2 repression was reversible, we treated c22 with a variety of kinase inhibitors and assayed 5'PDβ2 for transcription and USF-1 enrichment. While treatment with inhibitors for alternate kinases p38 MAPK and ATM as well as a DMSO-only control had no effect, inhibition of DNA-PKcs by Nu7026 both restored USF-1 binding and dramatically reduced transcription at 5'PDβ2. This ability to restore repression suggests that our cloning process may have led to the enrichment of persistent DNA breaks or otherwise aberrantly regulated DNA-PKcs activity. Either case illustrates the limits to which we can exploit

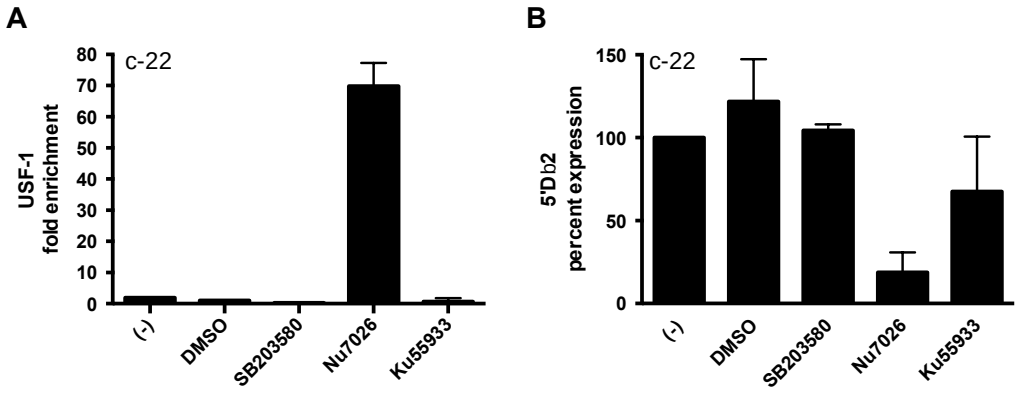


Figure A1.6: 5'PDβ2 repression is restored in c22 by inhibition of DNA-PKcs. Shown are representative ChIP QPCR of USF-1 binding at Dβ2 (A) and Dβ2 mRNA levels (B) in P5424-c22 one day after treatment with the indicated kinase inhibitors. Bars represent means ± SD (n = 3) for each sample. Fold enrichment of USF-1 and relative gene expression were calculated as described in Fig. A1.5.

transformed cell lines in our analyses of 5'PDβ2 regulation. Nonetheless, our data collectively support a model in which Dβ2 expression (and likely its recombinational accessibility) is regulated by DSB-sensitive binding of USF-1 at the Dβ2 12-RSS.

A1.4 Discussion

Transcription is required for recombinational accessibility in all studied immunoglobulin and T cell receptor loci, allowing interaction between the restriction signal sequences (RSS) associated with these gene segments and the recombinase complex. Our data suggest that in *Tcrb*, recombination between Dβ2 and an associated J serves as a trigger for transcriptional activation of the 5'Dβ2 promoter. Alternation of promoter usage during recombination would not be unique to *Tcrb*; the *Tcra* locus is a well-characterized case of such alternation. It has also been shown that Jα usage generally coincides with proximity to an active promoter and the distribution of activating histone mark H3K4me3, deposited by transcription [20]. In the unrearranged *Tcra*, transcription across the Jα region proceeds from the T early alpha (TEA) promoter, located immediately upstream of the foremost Jα segment. *Tcra* rearrangements, like most AgR rearrangements, are deletional, meaning that DNA like the TEA promoter that sits between the upstream Vα segments and downstream Jα segments are excised from the genome during V-to-J joining. In the case of frameshifted primary joints, *Tcra* assembly is rescued by secondary rearrangements between progressively more distal V and J segments, each of which must also have promoter activity to ensure their recombinational accessibility. A host of other promoters have been identified within the Jα array that are derepressed by the elimination of TEA transcription across the region, including those located around Jα49, Jα47, Jα45, Jα42 and Jα37 [21-23]. Since all known antigen receptor

genes engage in non-coding transcription in a highly conserved way, it is possible most D or J gene segments are flanked by such germline promoters.

Promoter regulation of rearrangement at Dβ1 depends on promoter activity immediately upstream of the Dβ1 coding sequence, but does not require that this activity be mediated by a specific promoter or its binding factors [5, 7]. During transcriptional elongation, RNA polymerase tracking requires transient histone displacement from the transcription bubble. Occlusion, displacement and repositioning of nucleosomes by polymerase-associated factors also induces RAG accessibility to the transcribed RSS [8]. Presumably then, 5'PDβ2 de-repression would indirectly regulate Dβ2 recombination by shifting transcriptional elongation through the Dβ2 segment and its flanking RSSs. Work is underway in the laboratory to directly test this indirect regulatory mechanism.

Previous studies of the DJβ2 cassette of *Tcrb* revealed interaction between this upstream promoter, a downstream repressor and a downstream promoter that facilitates Jβ2 transcription in the absence of Dβ2 recombination [4]. This downstream promoter appears to be largely regulated by transcription factor NFκB, but the upstream Dβ2 promoter involves a complex interplay between Runx1, GATA-3 and E47 as well as repressive effects from an element located in the upstream 12-RSS of Dβ2 and a 110 bp distally homopurine stretch [4, 10]. The greater complexity of Dβ2 regulation likely leads to the canonized observation that while both DJβ cassettes are amenable to recombination simultaneously, DJβ1 is prone to greater recombinational accumulation than DJβ2 [24]. Control of the repressor proximal to Dβ2 has recently been linked with binding of Upstream Stimulatory Factor 1 (USF-1), which is displaced from the Dβ2 repressor upon DSB-induced activation of DNA-PKcs. Non-specific DSBs generated through high but sub-lethal doses of chemical

or ionizing radiation of cultured RAG-deficient thymocytes were sufficient to induce such de-repression [11]. However, I found that a single recombination at Dβ2, as identified in clone 17, was enough to dramatically increase transcription from 5'PDβ2 despite the continued presence of both repressive elements and little apparent recombination within *Tcrg* or *Tcrd*, suggesting that de-repression of 5'PDβ2 need not involve extensive DSB induction *in trans*, but can instead involve very limited DSBs targeted to the Dβ2 cassette. Given that DNA-PKcs is directly recruited to RAG synaptic complexes, its presence at Dβ2 in response to D-to-J recombination would be expected to dramatically enhance its likelihood to find and phosphorylate USF-1 bound less than 30 bp upstream of the 3'Dβ2 cleavage site than DNA-PKcs activated by randomized genome-wide DSBs following irradiation.

USF-1 has been linked to immune system development, regulating β2-microglobulin, *Igh* and *Igλ* expression as well as germline Dδ2 promoter activity in *Tcrd* [17, 25]. We have also observed USF-1 regulation of TEA in *Tcra* (Bradshaw, Stone and Sikes, unpublished data). In ChIP-seq studies in human HepG2 cells, USF-1 was found to bind in excess of 2000 genomic targets, and has been linked to the stress-response transcriptional regulation of a wide array of genes associated with metabolic, immunologic, neurologic, and DNA damage response pathways [17]. Though much work has gone into dissecting its regulation by the MAP kinase P38, considerably less is known about its regulation by DNA-PKcs beyond that found for lipogenesis [18, 26, 27] and for Dβ2 transcriptional regulation [11]. Given its wide array of potential genomic targets and the unique demands on lymphocytes to accommodate RAG-induced DSBs as normal physiological events, dissecting the mechanism by which USF-1 regulates lymphocyte development in general and *Tcrb* assembly in particular will be of critical importance.

RAG-mediated double-stranded DNA breaks are thought to regulate a range of genes in developing lymphocytes, primarily through the kinase Ataxia Telangiectasia Mutated (ATM) which can up- or down-regulate a variety of transcription factors through tyrosine phosphorylation [28, 29]. Likewise, DSBs generated as recombinational intermediates have long been proposed to impact lymphocyte progression. Whereas most cells only encounter DSBs as a consequence of replication aberrations or environmental insult, RAG-DSBs are generated as part of the normal physiology of developing lymphocytes. As such, these cells must be uniquely programmed such that RAG-DSBs do not induce the overriding stress-response/apoptosis pathways seen in other cells. Microarray analyses of developing B lymphocytes found that the generation of RAG-DSBs correlated with altered expression in 364 genes (of which less than 20% were dependent on ATM:NFκB regulation) [30]. How these additional genes are regulated by DSBs and their roles in lymphocyte homeostasis, DNA repair, and development are unknown.

Our finding that DSBs induce an alteration in promoter activity at Dβ2 suggests the possibility that initial recombination events could trigger responses that influence subsequent steps in antigen receptor assembly. Specifically, the role of germline transcription in augmenting the recombinational accessibility of transcribed gene segments is now well established [3, 7, 23, 31]. In both *Tcra* and *Tcrb*, cell-promoter proximity directly correlates with recombinational frequency of a given gene segment (particularly when upstream of the targeted gene segment) [5, 21, 23]. By extension, it is predicted that while initially poor Dβ2 rearrangement efficiency stems from the 3' positioning of its germline promoter (3'PDβ2), de-repression of the upstream promoter in response to DSBs strongly enhances Dβ2 recombinational efficiency. In the preceding experiments, we have provided evidence of such a positive feedback system. *In vivo*, both DβJβ1 and DβJβ2 joining of *Tcrb*

begins in DN1 cells and continues until *Tcrb* recombination is completed in the DN3 stage of development. While the delayed activation of 5'PDβ2 would limit initial Dβ2-Jβ rearrangement, the consequence of this delay is that initial Vβ rearrangements (which are limited to recombined DJβ segments) would preferentially target the upstream DJβ1 cassette. Since individual V-to-DJ rearrangements have only a 1-in-3 chance of resulting in an in-frame joint, and since the DJβ1 cassette is deleted when Vβ segments target the DJβ2 cassette, limiting the initial recombinational accessibility of the DJβ2 cassette is advantageous in that it allows for a secondary V-to-DJβ2 joint per allele should the primary V-to-DJβ1 joint be frame-shifted. This not only dramatically increases the probability that an individual cell will generate an in-frame *Tcrb* assembly (increasing probability from 56% to near >90%), but it also ensures that both DJβ cassettes (which code for the CDR3 loop of the antigen binding domain) are relatively equally represented in the final *Tcrb* repertoire to maximize TCR diversity.

Because the USF-1 binding site at Dβ2 sits within the 5'Dβ2 12-RSS spacer, its binding during initial repression of 5'PDβ2 may not only indirectly limit Dβ2 recombinational accessibility by repressing transcriptional read-through of the Dβ2 coding sequence, it may also directly hinder recruitment of RAG-1/2 to the Dβ2 12-RSS. Such a steric hindrance may contribute to the absence of Vβ-Dβ2 rearrangement, and may also explain the absence of Dβ1-to-Dβ2 recombination in wild-type cells. In contrast, the Dβ1 23-RSS does support rearrangement to both the Jβ1 and Jβ2 12-RSSs with roughly equal efficiency [32], while significant D-to-D and V-to-D joining has been reported in mice lacking the non-core domain of RAG2, a mutation that disturbs the ordered use of RSSs throughout the antigen receptor genes [33]. Given the variety of ways in which DSB-mediated regulation of

USF-1 binding at Dβ2 may contribute to proper *Tcrb* assembly, deciphering the mechanism of this novel feedback response will be critical to our understanding of lymphocyte and repertoire development.

A1.5 References

1. Swanson, P.C., *The bounty of RAGs: recombination signal complexes and reaction outcomes.* Immunol Rev, 2004. **200**: p. 90-114.

2. Oltz, E.M., *Regulation of antigen receptor gene assembly in lymphocytes.* Immunol Res, 2001. **23**(2-3): p. 121-33.

3. Cobb, R.M., et al., *Accessibility control of V(D)J recombination.* Adv Immunol, 2006. **91**: p. 45-109.

4. McMillan, R.E. and M.L. Sikes, *Differential activation of dual promoters alters Dbeta2 germline transcription during thymocyte development.* J Immunol, 2008. **180**(5): p. 3218-28.

5. Sikes, M.L., et al., *Regulation of V(D)J recombination: a dominant role for promoter positioning in gene segment accessibility.* Proc Natl Acad Sci U S A, 2002. **99**(19): p. 12309-14.

6. Sikes, M.L., et al., *A developmental stage-specific promoter directs germline transcription of D beta J beta gene segments in precursor T lymphocytes.* J Immunol, 1998. **161**(3): p. 1399-405.

7. Sikes, M.L., C.C. Suarez, and E.M. Oltz, *Regulation of V(D)J recombination by transcriptional promoters.* Mol Cell Biol, 1999. **19**(4): p. 2773-81.

8. Kondilis-Mangum, H.D., et al., *Transcription-dependent mobilization of nucleosomes at accessible TCR gene segments in vivo.* J Immunol, 2010. **184**(12): p. 6970-7.

9. Whitehurst, C.E., S. Chattopadhyay, and J. Chen, *Control of V(D)J recombinational accessibility of the D beta 1 gene segment at the TCR beta locus by a germline promoter.* Immunity, 1999. **10**(3): p. 313-22.

10. McMillan, R.E. and M.L. Sikes, *Promoter activity 5' of Dbeta2 is coordinated by E47, Runx1, and GATA-3.* Mol Immunol, 2009. **46**(15): p. 3009-17.

11. Stone, J.L., et al., *DNA double-strand breaks relieve USF-mediated repression of Dbeta2 germline transcription in developing thymocytes.* J Immunol, 2012. **188**(5): p. 2266-75.

12. Mombaerts, P., et al., *Characterization of immature thymocyte lines derived from T-cell receptor or recombination activating gene 1 and p53 double mutant mice.* Proc Natl Acad Sci U S A, 1995. **92**(16): p. 7420-4.

13. Sikes, M.L., et al., *A streamlined method for rapid and sensitive chromatin immunoprecipitation.* J Immunol Methods, 2009. **344**(1): p. 58-63.

14. Ciccone, D. and M. Oettinger, *Chromatin modifications as clues to the regulation of antigen receptor assembly.* Novartis Found Symp, 2004. **259**: p. 146-58; discussion 158-69.

15. Li, B., M. Carey, and J.L. Workman, *The role of chromatin during transcription.* Cell, 2007. **128**(4): p. 707-19.

16. Sirito, M., et al., *Ubiquitous expression of the 43- and 44-kDa forms of transcription factor USF in mammalian cells.* Nucleic Acids Res, 1994. **22**(3): p. 427-33.

17. Corre, S. and M.D. Galibert, *Upstream stimulating factors: highly versatile stress-responsive transcription factors.* Pigment Cell Res, 2005. **18**(5): p. 337-48.

18. Wong, R.H., et al., *A role of DNA-PK for the metabolic gene regulation in response to insulin.* Cell, 2009. **136**(6): p. 1056-72.

19. Blunt, T., et al., *Defective DNA-dependent protein kinase activity is linked to V(D)J recombination and DNA repair defects associated with the murine scid mutation.* Cell, 1995. **80**(5): p. 813-23.

20. Abarrategui, I. and M.S. Krangel, *Noncoding transcription controls downstream promoters to regulate T-cell receptor alpha recombination.* EMBO J, 2007. **26**(20): p. 4380-90.

21. Hawwari, A. and M.S. Krangel, *Regulation of TCR delta and alpha repertoires by local and long-distance control of variable gene segment chromatin structure.* J Exp Med, 2005. **202**(4): p. 467-72.

22. Hawwari, A. and M.S. Krangel, *Role for rearranged variable gene segments in directing secondary T cell receptor alpha recombination.* Proc Natl Acad Sci U S A, 2007. **104**(3): p. 903-7.

23. Abarrategui, I. and M.S. Krangel, *Regulation of T cell receptor-alpha gene recombination by transcription.* Nat Immunol, 2006. **7**(10): p. 1109-15.

24. Born, W., et al., *Rearrangement of T-cell receptor beta-chain genes during T-cell development.* Proc Natl Acad Sci U S A, 1985. **82**(9): p. 2925-9.

25. Carabana, J., E. Ortigoza, and M.S. Krangel, *Regulation of the murine Ddelta2 promoter by upstream stimulatory factor 1, Runx1, and c-Myb.* J Immunol, 2005. **174**(7): p. 4144-52.

26. Meulle, A., et al., *Positive regulation of DNA double strand break repair activity during differentiation of long life span cells: the example of adipogenesis.* PLoS One, 2008. **3**(10): p. e3345.

27. Wong, R.H. and H.S. Sul, *DNA-PK: relaying the insulin signal to USF in lipogenesis.* Cell Cycle, 2009. **8**(13): p. 1977-8.

28. Bredemeyer, A.L., et al., *ATM stabilizes DNA double-strand-break complexes during V(D)J recombination.* Nature, 2006. **442**(7101): p. 466-70.

29. Matsuoka, S., et al., *ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage.* Science, 2007. **316**(5828): p. 1160-6.

30. Hystad, M.E., et al., *Characterization of early stages of human B cell development by gene expression profiling.* J Immunol, 2007. **179**(6): p. 3662-71.

31. Corcoran, A.E., *The epigenetic role of non-coding RNA transcription and nuclear organization in immunoglobulin repertoire generation.* Semin Immunol, 2010. **22**(6): p. 353-61.

32. Livak, F., et al., *Genetic modulation of T cell receptor gene segment usage during somatic recombination.* J Exp Med, 2000. **192**(8): p. 1191-6.

33. Curry, J.D. and M.S. Schlissel, *RAG2's non-core domain contributes to the ordered regulation of V(D)J recombination.* Nucleic Acids Res, 2008. **36**(18): p. 5750-62.