Summer 8-15-2015

Spatial and Epigenetic Regulation of T-Cell Receptor Beta Gene Assembly

Kinjal Majumder
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences
Immunology

Dissertation Examination Committee:
Eugene Oltz, Chair
Paul Allen
Maxim Artyomov
Takeshi Egawa
Kenneth Murphy
Barry Sleckman

Spatial and Epigenetic Regulation of T-Cell Receptor Beta Gene Assembly

by
Kinjal Majumder

A dissertation presented to the
Graduate School of Arts & Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2015
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AgR	Antigen Receptor
ATM	Ataxia Telangiectasia Mutated
BAC	Bacterial Artificial Chromosome
C	Constant
CBE	CTCF Binding Element
Cer	Contracting Element for Recombination
Ccnd3	Cyclin D3
CD3	Cluster of Differentiation 3
ChIP	Chromatin Immunoprecipitation
CpG	Cytosine-phosphate-Guanine
CTCF	CCCTC-Binding Factor
D	Diversity
Ig	Immunoglobulin
J	Joining
DamID	DNA Adenine Methyltransferase Identification
DN	Double Negative
DP	Double Positive
Eα	*Tcra* enhancer
Eβ	*Tcrb* enhancer
Eμ	*IgH* enhancer
H19-ICR	H19- Imprint Control Region
IGCR1	Intergenic Control Region 1
Kb	Kilobase Pairs

LAT	Linker for Activation of T Cells
Mb	Megabase
NHEJ	Non Homologous End Joining
PDβ	Dβ associated Promoter
pTα	pre-T cell receptor alpha
qPCR	Quantitative Real-Time Polymerase Chain Reaction
RAD21	RAD21 homolog (*S. pombe*)
RAG	Recombinase Activating Gene
RC	Recombination Center
RR	Regulatory Region
RSS	Recombination Signal Sequence
Sis	Silencer In the Intervening Sequence
SMC	Structural Maintenanse of Chromosomes
TEA	T Early A Promoter
TCR	T Cell Receptor
V	Variable
ZAP70	Zeta-chain Associated Protein Kinase 70
3C	Chromosome Conformation Capture
3D-FISH	Three-Dimensional Fluorescence *In Situ* Hybridization
4C	Circular Chromosome Conformation Capture
5C	**..... Circular Chromosome Conformation Capture Carbon Copy**

Acknowledgments

I am indebted to my thesis mentor, Dr. Eugene Oltz, for his mentorship and the opportunity to work in his lab. He is always happy to point me in the right direction, and has encouraged (sometimes with reservation) my ideas for experiments, collaborations and future professional pursuits. He has been a wonderful mentor over the past five years.

I would like to thank all the members of the Oltz Lab, both past and present, for creating such a fun environment to do science. I would like to specially thank Olivia Koues who trained me to do experiments and for providing a lot of intellectual support over the years. I would also like to thank Katherine Kyle for taking such great care of our mouse colony. I am indebted to all my friends and fellow graduate students in the Immunology program for their support, sympathy and encouragement over the years.

I would like to thank my parents, Somajyoti and Nandita Majumder for their inspiration and support. They encouraged me to break societal expectations, strike out on my own and build my own life in the United States. As an eleven year old child in Sydney, I was inspired to pursue my PhD by my father, and now I can proudly say, "Challenge completed!" Despite the distance separating us, I know their love and support is always with me. In their absence, the role of parents in my daily life has been filled by my in-laws Chris and Cindy Wunderlich. It is an honor to be a part of their family. From cooking turducken to riding motorcycles to family adventures, Chris and Cindy always inspire me to remain young at heart and enjoy life.

Finally, I want to dedicate this dissertation and my deepest gratitude to my wife Erica, who has always been there for me. I would not have been able to complete my PhD without her. From the frustration of failed experiments and heartbreak of paper rejections to the ultimate elation of

graduating, she has been by my side through everything. I look forward to our future adventures and the rest of our lives together.

Kinjal Majumder

Washington University in St. Louis

May 2015

ABSTRACT OF THE DISSERTATION

Spatial and Epigenetic Regulation of T-Cell Receptor Beta Gene Assembly

by

Kinjal Majumder

Doctor of Philosophy in Biology and Biomedical Sciences

Immunology

Washington University in St. Louis, 2015

Dr. Eugene M. Oltz, Chair

The adaptive immune system endows mammals with a sophisticated mechanism to recognize foreign proteins via surface antigen receptors that are expressed on the surface of all lymphocytes. This defense network is generated by V(D)J recombination, a set of sequentially controlled DNA cleavage and repair events that assembles functional antigen receptor genes from distally located Variable (V), Diversity (D) and Joining (J) gene segments. However, the recombination process must be stringently regulated to prevent formation of chromosomal translocations, which can lead to tumors. The process of V(D)J recombination is controlled at the levels of tissue, stage and allele specificity by a collection of architectural and regulatory elements that are distributed throughout each antigen receptor locus. Our laboratory has characterized several genetic elements that regulate chromatin accessibility and recombination at the T cell receptor beta (*Tcrb*) locus. These elements include transcriptional promoters and enhancers, which interact with each other in conformational space to form a promoter-enhancer holocomplex, facilitating Dβ to Jβ recombination. Simultaneously, spatial apposition of the Vβ cluster to the DβJβ region (a phenomenon called locus contraction) increases the efficiency of

long-range Vβ recombination. Using extensive chromatin profiling of the *Tcrb* locus, we have discovered that selection of Vβ genes depend upon their association with transcriptionally active chromatin and high quality Recombination Signal Sequences, which serve as substrates for the V(D)J recombinase proteins RAG1/2. We further identify a bi-functional barrier-tethering region upstream of the DβJβ cluster that is essential for stabilizing its long-range interactions with distal Vβ gene segments in progenitor $CD4^{-}CD8^{-}$ double negative (DN) thymocytes. Following *Tcrb* rearrangement, progenitor thymocytes proliferate and differentiate into $CD4^{+}CD8^{+}$ Double Positive (DP) cells, where the Vβ genes are epigenetically silenced and the distal ends of *Tcrb* are spatially segregated (presumably to inhibit further rearrangements). However, we have found that the transcriptionally inactive proximal Vβ genes continue to interact with the DβJβ cluster in a proliferation independent manner. These findings divide the *Tcrb* locus into two architectural domains, of which only the distal part is spatially segregated in DP cells. The loss of distal Vβ interaction is also observed in DP thymocytes containing a rearranged *Tcrb* allele, suggesting this conformation is DP-intrinsic. Our results have unraveled new mechanisms that stabilize the long-range *Tcrb* conformation in DN cells, how the Vβ segments are selected to recombine and how *Tcrb* topology is retained by DP-intrinsic mechanisms. These studies pave the way for future investigations into the role of boundary elements and tissue specific transcription factors in sculpting AgR gene assembly and regulating genome topology.

Chapter 1: Long-range *Tcrb* Interactions: FISH-ing, ChIP-ing and C-ing the Epigenome

Review in preparation for Advances in Immunology

1.1 Introduction

Ordered organization of the genome is essential for proper packaging and function of the nuclear material. In addition to DNA, the nuclear environment is made up of many proteins and RNA. All of these factors are essential for the general function of a cell. At its simplest level, the DNA molecule is a long polymer that is packaged into nucleosomes by wrapping around histone proteins, forming chromatin. The role of chromatin in regulating genetic processes are dependent on its accessibility to the nuclear environment, and is controlled by chemical modifications of the histone proteins. The packaged and condensed chromatin forms the chromosome. There exists a void in our knowledge of the scaffolding process that packages chromatin into chromosomes. Early studies on the topology of the genome were restricted by technological constraints. However, the advances in high-throughput sequencing and imaging technologies have provided unprecedented insights into how the chromatin topology exists in the nucleus and regulates the function of the cell.

High-throughput sequencing studies of chromosome conformation, along with polymer modelling of DNA show that chromosomes adopt the configuration of a fractal globule, which was first described for polymers by Grosberg and colleagues in 1988 (Grossberg et al., 1988). Consistent with the mathematical properties of a fractal, genome folding physically exhibits a never-ending pattern. Current observations and discoveries have whittled the genome down into the building blocks of chromosomes, called Topologically Associated Domains (TADs) (Fudenberg and Mirny, 2012). These are megabase-sized domains contain high levels of intra-chromosomal interactions. In order to derive biologically meaningful information from the folded

genome, the past decade has witnessed extensive characterization of model gene loci, and how changes in the architecture of these loci influences gene expression.

The antigen receptor loci (AgR) loci serve as excellent model systems to mechanistically study the long-range regulation of gene expression. They span large distances on the linear genome, and in order for AgR loci to recombine, they must be activated in a tissue, stage and allele specific manner. This epigenetic regulation of AgR chromatin regulates the recruitment of the recombination machinery. Prior to recombination, the loci undergo undergo contraction, bringing distal elements into spatial proximity, presumably to facilitate rearrangement, or be transcriptionally regulated by the same cis-elements. However, a confounding array of variables can regulate the expression and architecture of these loci including regulatory elements, structural protein binding, non-coding RNA's and CpG methylation. The *Tcrb* locus is the smallest antigen receptor locus, spanning 700 kb of the linear mouse genome, and forms an ideal model system to mechanistically dissect the principles guiding genome topology.

Antigen receptor (AgR) genes in general are assembled by regulated DNA break and repair events known as V(D)J recombination. V(D)J recombination is a stepwise rearrangement process carried out by the RAG endonuclease complex. RAG recruitment to the AgR loci depends on the presence of accessible chromatin, mediated by the epigenetic landscape. Chromatin accessibility depends upon the function of transcription regulatory elements, which communicate with one another over long-range, generating a unique three-dimensional architecture. The *Tcrb* locus is a unique model system for dissecting these mechanisms of V(D)J recombination, containing well defined regulatory elements that influence local chromatin accessibility. It is made of 35 Vβ gene segments, two Dβ-Jβ gene clusters, a well enhancer and 22 binding sites for the architectural

protein CTCF. These epigenetic and topological mechanisms integrate together to sculpt the T-cell repertoire ensuring lymphocyte diversity while minimizing the chances of autoreactivity (by maintaining monospecificity of lymphocytes through allelic exclusion) or tumorigenesis (by forming translocations caused by RAG-mediated DNA breaks).

1.2 Linear structure of the *Tcrb* locus

The mouse *Tcrb* (m*Tcrb*) locus is located on chromosome 6 and spans 700 kilobases (kb) of the linear genome whereas the human *Tcrb* (h*Tcrb*) locus is located on chromosome 7 and spans 620 kb of the linear genome. As shown in Figure 1.1, both m*Tcrb* and h*Tcrb* exhibit similarities in their linear arrangement which suggest a common evolutionary ancestry and common mechanisms may regulate their expression (Glusman et al., 2001; Lane et al., 2010). The 3' end of both loci harbor the Dβ-Jβ clusters, which span 25 kb. This region is made up of two Dβ-Jβ gene cassettes that are independently regulated by germline promoters (Osipovich and Oltz, 2010). The end of the Dβ-Jβ cluster also contains the sole *Tcrb*-enhancer Eβ, which is located upstream of the final Vβ gene segment. The Vβ gene segment at the 3' end of *Tcrb* undergoes inversional rearrangement in both mice and humans.

The Dβ-Jβ cluster is separated from upstream Vβ genes by a series of inactive Trypsinogen (Trp) genes. In mice, there are two Trp clusters: the 3' Trp which separates Dβ-Jβ from *Trbv2-30* (spanning 250 kb), and the 5' Trp which separates *Trbv1* from *Trbv2* (spanning 150 kb). In contrast, h*Tcrb* contains inactive Trp genes spanning 50 kb. The large size of mouse 3' Trp cluster may be the result of segmental duplications of 50 kb genomic regions containing *Prss1* and *Prss3* genes. Similarly, the 5' Trp region also contains a 37 kb segmental duplication. Similar to the

m*Tcrb*-Trp region, the h*Tcrb*-Trp cluster also contains segmental duplications of the *Prss1* genes (Bailey et al., 2002. Indeed, single nucleotide polymorphisms (SNPs) have been detected at the *Prss1* gene, which increase susceptibility of patients to chronic pancreatitis (Derikx et al., 2014; Whitcomb et al., 2012). The presence of these inactive Trp genes within *Tcrb* imply that mechanisms exist which keep the the Trp cluster repressed in DN thymocytes. It is noteworthy that although a biological role for these segmental duplications in regulating *Tcrb* repertoire has not been discovered, such regions show poor alignment of sequencing reads, which may confound the interpretation of high-throughput sequencing data, design of primers and imaging assays using DNA-hybridising probes.

The presence of recombinationally inert pseudogenes are also observed in both m*Tcrb* and h*Tcrb* loci. m*Tcrb* contains a cluster of inactive pseudogenes in the 5' *Trbv* region, containing *Trbv*6-11. However, the pseudogenes are spaced throughout the Vβ cluster in h*Tcrb*. Studies on the three-dimensional architecture of mouse *Igk* have shown that Vκ pseudogenes are looped out of the Jκ interactome (Lin et al., 2012). This observation predicted that in the *Tcrb* locus adopts a looped conformation which spatially separates the Vβ pseudogenes from the DβJβ cluster (discussed below).

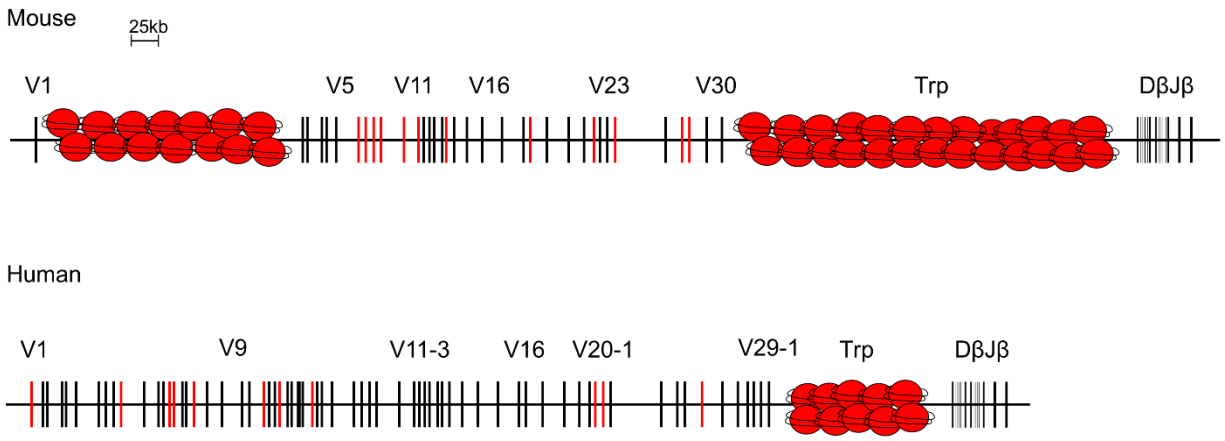


Figure 1.1: Comparison of the mouse and human T Cell Receptor Beta loci. The active genes are presented as black bars, and pseudogenes are presented as red bars. Inactive chromatin is shown as red nucleosomes.

1.3 Coordinating V(D)J Recombination by Controlling RAG

The generation of antigen receptor (AgR) diversity by V(D)J recombination is a cornerstone of adaptive immunity. Through this process, variable (V), diversity (D), and joining (J) gene segments are selected from large arrays on AgR loci and fused during lymphocyte development to generate the antigen recognition portion (V region) of immunoglobulin (Ig) and T cell receptors (TCRs). The repertoire of receptor combinations drives the production of functional lymphocyte subsets and sets up a frontline defense against common pathogens (Bassing et al., 2002; Cobb et al., 2006). The process of V(D)J recombination can be controlled by regulating RAG expression, its ability to bind DNA substrates or the survival of lymphocytes undergoing rearrangement.

One level of control exerted on the rearrangement process is via the expression of the recombinase machinery. $CD4^{-}CD8^{-}$ double negative (DN) thymocytes initiate recombination by

expressing the RAG proteins RAG1 and RAG2, which target the *Tcrb, Tcrg* and *Tcrd* loci. The focal zone of RAG deposition in AgR loci is the D-J region (on *Tcrb, Tcrg, Tcrd* and *Igh*) or J region (on *Tcra*, *Igk* or *Igl*). This feature of the AgR loci is referred to as the recombination center (RC). *Tcrb* assembly progresses in a stepwise manner, with Dβ to Jβ rearrangement at the RC preceding Vβ to DβJβ recombination. The generation of a functional VDJβ exon leads to the expression of TCRβ protein, downregulation of *Rag1/2* and assembly of the pre-TCR (Brady and Bassing, 2011). Signaling through the pre-TCR leads to proliferation of the thymocytes and differentiation to the $CD4^+CD8^+$ double positive (DP) stage. This upregulates *Rag1/2,* which leads to *Tcra* rearrangement. Assembly of Vα-Jα chains lead to cell surface expression of αβ TCR on DP thymocytes (Krangel et al., 2009). Subsequent selection events by MHC and self-antigens in the thymus lead to the generation of mature T cells. Each mature T cell expresses a unique T Cell Receptor capable of recognizing foreign antigen peptides (Brady and Bassing, 2011).

The mechanism of RAG action follows an ordered rearrangement process. Before this can happen, RAG must bind to its cognate DNA targets. These are known as Recombination Signal Sequences (RSSs), which flank AgR gene segments. RSSs are made up of conserved heptamer and nonamer segments that are separated by 12 or 23 basepair spacers. This generates 12RSS or 23RSS respectively. After being recruited to the RSS, RAG cleaves the DNA between the gene segment and its associated RSS. Distal RAG-RSS-bound gene segments are synapsed together following the 12/23 rule. According to the 12/23 rule, genes containing 23RSS and 12RSS are assembled together during V(D)J recombination. As shown in Figure 1.2, the Dβ1 gene segment is flanked by a 5' 12RSS and a 3' 23RSS whereas the Jβ gene segments contain a 5' 12RSS. This orientation of the RSS facilitates rearrangement between Dβ1 and Jβ. Following Dβ-Jβ

rearrangement, the 5’ 12RSS on the Dβ-Jβ exon rearranges with the 3’ 23RSS on upstream Vβ gene segment. However, the 12/23 rule predicts that Vβ should also recombine with Jβ due to compatible RSS. The occurrence of such rearrangements are inhibited, and the underlying principle has been dubbed the beyond 12/23 rule (B12/23, Figure 1.2) (Bassing et al., 2000; Sleckman et al., 2000). Molecular explanations of B12/23 restriction have attributed unique properties to the 5’Dβ1 RSS which enables rearrangement with Vβ in a position independent manner (Bassing et al., 2002). However, neither 12/23 nor B12/23 restrictions can explain the tissue-specific and ordered rearrangements of *Tcrb*.

The inherent properties of the subunits that make up RAG proteins impact its recombining capacity. Both RAG1 and RAG2 are made of a core and a non-core subunit (Schatz and Ji, 2011). RAG-mediated cleavage is carried out by the core subunits of RAG1 and RAG2. However, the non-core subunit of RAG1 promotes Vβ rearrangement by increasing the utilization of Vβ genes with poor RSS (Horowitz and Bassing, 2014). The non-core subunit of RAG2 is also thought to perform a similar function (Liang et al., 2002). Taken together, these findings suggest that the non-core RAG subunits enhance the binding and rearrangement potential of RAG to regions that don’t possess the ideal RSS consensus sequence.

In addition to controlling RAG recruitment and protein function, V(D)J rearrangement can be controlled at the level of DNA repair, with the end result being life or death for the developing lymphocyte (Bassing et al., 2002). RAG-mediated DNA cleavage occurs in two steps. First, RAG causes a nick between the gene segment and the heptamer. Following cleavage, the resulting 3’ hydroxyl group attacks the opposite strand to generate a DNA double-strand break, forming a

covalently sealed DNA hairpin. Distally located gene segments are paired together to form a synapse, which is then repaired by DNA repair proteins of the NHEJ machinery. This leads to the formation of a novel coding join and extrachromosomal signal join (Gellert, 2002; Schatz and Ji, 2011). Absence of the NHEJ repair machinery can significantly alter the survivability, or tumorigenesis, by these developing lymphocytes (Helmink and Sleckman, 2012).

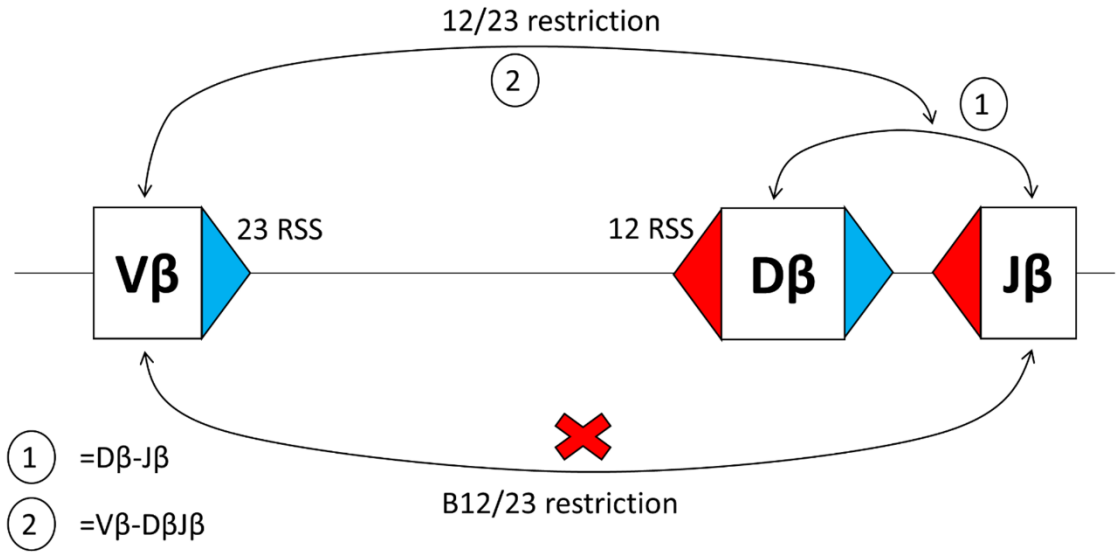


Figure 1.2: Schematic of *Tcrb* assembly. Simplified schematic of the imposition of 12/23 and B12/23 restriction. 23 RSS are colored in blue and 12 RSS are colored in red.

1.4 Accessibility control of *Tcrb* assembly

Despite the expression of a common recombinase enzyme and the presence of common RSS substrates on all AgR loci, RAG is selectively recruited to appropriate regions and carry out rearrangements. For example, *Tcrb* rearranges in DN thymocytes and not in proB cells. The mechanistic basis for this process lies in the expression of germline transcripts from gene segments that can subsequently undergo rearrangements (Van Ness et al., 1981; Yancopoulos and Alt, 1985).

This link between germline transcription and recombination suggested that AgR genes are packaged within chromatin that impedes RAG recruitment. However, modification of the associated chromatin by epigenetic modifiers confers recombinational and transcriptional potential to these genes. Indeed, actively transcribing promoters are associated with trimethylated lysine 4 on histones (H3K4), which is a direct substrate of RAG2 (Matthews et al, 2007; Liu et al., 2007). Therefore, transcriptional accessibility of chromatin plays an important role in regulating RAG recruitment and rearrangement. This led to the identification and characterization of cis- elements that regulate transcription such as promoters, enhancers and insulators, and their role in AgR assembly.

Enhancer- mediated assembly of Tcrb genes

Enhancers are intergenic elements that confer transcriptional competence to promoters. Tissue-specific activity of enhancer elements can function as rheostats that modulate gene expression. The *Tcrb* locus contains a single enhancer, Eβ, at its 3' end, which is activated early during thymocyte development (McDougall et al., 1988). Eβ conferred recombinational potential to the TCRβ minilocus in DN thymocytes (Ferrier et al., 1990). The deletion of endogenous Eβ in mice inhibited αβ T cell development by preventing *Tcrb* recombination, which abolished the expression of a cell-surface TCR (Bories et al., 1996; Bouvier et al., 1996). Subsequent studies showed that the loss of Eβ diminished RC-chromatin accessibility by erasing active histone marks and increasing CpG methylation. These changes in the epigenetic landscape reduced the expression of germline transcripts (Mathieu et al., 2000; Spicuglia et al., 2002; Oestreich et al.,

2006). Therefore, abolition of chromatin accessibility at the PDβ1 and PDβ2 germline promoters in the absence of Eβ inhibited the ability of *Tcrb* to rearrange.

Eβ contains binding sites for transcription factors including KLF, bHLH, ETS1, Runx1 and GATA3 (Bonnet et al., 2009), which are expressed during lineage commitment. These transcription factors, upon being expressed at the appropriate T-cell lineage, are recruited to Eβ and can serve as a platform for epigenetic modifiers such as p300 leading to the revision of the epigenetic landscape of the RC. This confers transcriptional accessibility to the germline promoters over short-range, forming a stable enhancer-dependent holocomplex. This modification of the RC architecture licenses promoter function (Oestreich et al., 2006). However, the stepwise mechanism of holocomplex formation remains unclear. The enhancer could first be activated before communicating with the promoters. Alternatively, the enhancer and promoters could communicate together to assemble a stable platform which then recruits the necessary epigenetic modifiers. This revised epigenetic landscape can render the associated RC- chromatin accessible to the transcriptional machinery and finally, RAG-mediated rearrangement.

Promoter-directed assembly of Tcrb genes

Promoters form genetic platforms for the assembly of the transcriptional machinery. Promoters are located in proximity to all AgR gene segments, and when activated, express sterile transcripts. Following successful V(D)J recombination, mature AgR gene expression is driven by the V-associated promoters. On *Tcrb*, the presence of individual promoters associated with germline gene segments and their dependence on Eβ for proper function suggest that promoters are involved in local modulation of chromatin accessibility.

The germline promoters on *Tcrb*-RC associated with the Dβ1 and Dβ2 gene segments are called PDβ1 and PDβ2 respectively, and are associated with high levels of sterile transcription, histone acetylation and trimethylated H3K4 in DN thymocytes. The role of PDβ1 in regulating local chromatin accessibility has been verified in mice and on model TCRβ substrates. Loss of PDβ1 abrogates germline transcription, active chromatin marks and rearrangement of the neighboring Dβ1-Jβ1.1-1.7 cluster (Sikes et al., 1999; Whitehurst et al., 2000). However, accessibility and rearrangement of the Dβ2-Jβ2.1-2.7 cluster remains unaltered due to its unique germline promoter, PDβ2, which remains active in the absence of PDβ1 (McMillan et al., 2008). The mutually independent regulation of PDβ1 and PDβ2 provides a unique model to study the range of influence of a germline promoter and how their expression can be modulated by the enhancer.

Loss of PDβ1 ablates the active acetylated histone marks on the associated Dβ1-Jβ1 cluster. The most prominent hypoacetylation of the Dβ1-Jβ1 cluster occurs at a region close to Dβ1, suggesting that the PDβ1 exerts a limited local influence. This was tested by repositioning PDβ1 400 bp away from its native location in a minilocus substrate, which was sufficient to impair Dβ-Jβ rearrangement despite retaining germline transcription through the neighboring Jβ gene segments. However, inversion of the PDβ1 promoter supported rearrangement adjacent to the Dβ1 gene segments despite the absence of transcription through the RAG targets (Sikes et al., 2002). This showed that promoters exert short-range control over locus assembly independent of germline transcription. Therefore, germline transcription may be a by-product of active promoters. Insertion of a transcription terminator in the Jα cluster of *Tcra* abolishes the rearrangement of Jα genes

downstream of the terminator while retaining upstream Jα rearrangement (Abarrategui and Krangel, 2006). Both promoter repositioning and terminator knock-in studies showed that active transcription disassembles DNA from nucleosomes, making the RSS more accessible to RAG1. Taken together, these observations indicate that transcription regulation, chromatin modification and H3K4me3 deposition are important for licensing recombination.

The epigenetic mechanisms that regulate V(D)J recombination were deciphered by targeted recruitment of chromatin modifiers that establish repressive histone marks (such as H3-K9 methylation) upstream of PDβ1 in minilocus substrates. In these studies, recruitment of the histone methyl-transferase G9a to PDβ1 suppressed Dβ-Jβ rearrangements (Osipovich et al., 2004). On the other hand, the targeted recruitment of BRG1 (the catalytic component of SWI/SNF chromatin remodeling complex) upstream of the RC induced Dβ-Jβ accessibility and rearrangements even in the absence of PDβ1 (Osipovich et al., 2007). These studies provided a common framework which suggested that revision of local chromatin is essential for RAG accessibility. In this regard, PDβ1/2 impart their influence over local chromatin while Eβ is essential for accessibility of the entire RC. The formation of a stable promoter-enhancer holocomplex may be the common mechanism which regulates RC-accessibility and subsequently, rearrangement (Oestreich et al., 2006).

Regulation of Tcrb chromatin accessibility by insulators

The modulation of promoter-enhancer communication and the spread of chromatin throughout the genome is essential for turning genes on and off, and to regulate the epigenetic landscape. This function is performed by insulators, which are DNA-protein complexes that can

block enhancer-promoter interactions or direct enhancers to their cognate promoters. Insulators also have additional roles in preventing the spread of heterochromatin into euchromatin (Yang and Corces, 2012). The *Tcrb* locus in DN thymocytes is divided into alternating active (*Trbv1*, *Trbv2-Trbv30* and the RC) and repressive chromatin compartments (5' and 3' Trp regions). This suggests that insulator elements at the chromatin boundaries cordon off the *Tcrb* genes from the Trp genes.

On the *Tcrb* locus, the RC-associated chromatin remains highly active in DN and DP thymocytes. A robust difference in active thymocyte-specific *Tcrb* chromatin is observed at the border between the 3' Trp genes and the RC. The 5' end of the RC contains a CTCF-bound insulator element (Majumder et al., 2015). Loss of this insulator in ΔPDβ1 thymocytes leads to the spread of active RC chromatin upstream, leading to the activation of the *Prss2* gene in DN thymocytes. A LINE element associated with the upstream-RC region has also been implicated as a barrier-type insulator separating active from inactive chromatin (Carabana et al., 2011). Deciphering the molecular mechanisms that commission these chromatin insulator properties warrant further investigation.

Insulators can have a robust effect on *Tcrb* rearrangement. Knocking in the strong chicken globin insulator H19-ICR in between the Dβ1-Jβ1 and the Dβ2-Jβ2 clusters diminishes rearrangements to Dβ1. This Dβ1 inhibition is presumably due to the chromatin barrier imposed in the middle of the RC by the insulator restricting the target of Eβ as far as the Dβ2 cluster (Shrimali et al., 2012). However, it remains to be tested whether knocking in H19-ICR alters the long-range architecture of *Tcrb*.

The differential expression of genes on either side of an insulator element may be caused by altered methylation of CpG islands. Differential DNA methylation plays a role regulating CTCF occupancy at *Cdkn2A*, *Bcl6* and *Bdnf* loci (Rodriguez et al., 2010; Lai et al., 2010; Chang et al., 2010). Indeed, the consensus sequence for CTCF binding sites contain CpG islands, raising the question of how DNA methylation and CTCF-function may be related (Engel et al., 2004; Wang et al., 2012). Genome-wide assays have implicated the role of CTCF and PARP in exerting the insulator control (Ong et al., 2013). CTCF and PARP form a complex with the DNA methyl transferase DNMT1 (Zampieri et al., 2012; Guastafierro et al. 2008). At the *Tcrb* locus, the CTCF-bound insulator element upstream of PDβ1 may function through this pathway, by recruiting PARP and DNMT, which then methylates the 5'RC border. However, the localization of PARP and DNMT on *Tcrb* has not been mapped. Future studies on PARP and DNMT recruitment will provide additional insights into the molecular mechanisms regulating the *Tcrb* insulator function.

1.5 Topological control of *Tcrb* expression

The mammalian genome is folded into distinct architectural subclasses. This generates a hierarchy of three dimensional conformations. Firstly, chromosomes occupy distinct sub-compartments within the nucleus (Guelen et al., 2008). Secondly, chromosomes are folded into discreet megabase-sized structures of high frequency cis-interacting regions called Topologically Associated Domains (TADs). Within a TAD, there exists long-range promoter-enhancer interactions that regulate gene expression (Dixon et al., 2012). The *Tcrb* locus occupies a TAD within mouse chromosome 6 (Zhang et al., 2012). However, the three-dimensional mechanisms that cordons *Tcrb* from the rest of nucleus, how the *Tcrb*-TAD is stabilized and the functional

importance of *Tcrb*-TAD interactions are unknown. Resolving these mechanisms regulating intra- and inter-TAD interactions will provide insights into how regulation of *Tcrb* topology impacts T-cell function.

The adoption of lineage-specific locus architecture was observed using 3D-FISH, where labelled BAC probes were used to hybridize (and subsequently visualize) the distal ends of AgR loci. This revealed the general principle of "locus contraction" of AgR loci during the stage when they undergo recombination (Shih and Krangel, 2013). Using 3D-FISH, *Tcrb* was discovered to undergo a lineage specific contraction, where the distal ends came together in DN thymocytes, and were subsequently separated spatially in DP cells (Skok et al., 2007). The advent of chromosome conformation capture technologies allowed the elucidation of how the fine structure of AgR loci is stabilized. Chromosome conformation capture technologies utilize the principle that long-range genome folding can be detected by the frequency with which novel ligation junctions are formed when nuclei are crosslinked, digested and intra-molecularly ligated. The frequency with which novel ligation junctions are formed can be detected using many different platforms. These include 3C-qPCR (which requires primers and probes for regions of interest), 4C (which utilizes sequencing to show the genome-wide interaction partners of a viewpoint), 5C (which utilizes primers throughout a region of interest, combined with sequencing to show all relative interactions) and HiC (which assays the interaction of every fragment with every other throughout the genome in an unbiased manner) (Dekker et al., 2013).

RAG- mediated control of topology and accessibility

All in-vivo studies on chromatin accessibility, transcription and AgR looping are performed in lymphocytes extracted from RAG-deficient animals, to preclude the confounding effects of rearranged loci in RAG-proficient setting. Although the *Tcrb* locus in RAG-deficient thymocytes is maintained in a poised germline configuration, the in-vivo binding of RAG to these targets were only surmised. The formal test for this phenomenon came about with the generation of a RAG-deficient mice expressing a nuclease-incompetent RAG transgene (Ji, Resch et al., 2010). In these mice, RAG binds to the *Tcrb*-RC but not upstream Vβ gene segments. However, in ΔEβ thymocytes, where the RC is epigenetically silenced, RAG does not bind to the RC. On the other hand, partial inactivation of the RC by deleting PDβ1 alters the RAG binding profile. In these thymocytes, RAG is bound only to the transcriptionally accessible Dβ2 cluster but not the Dβ1 cluster (Ji, Little et al., 2010). Since RAG-mediated synapsis of distal recombining elements are required for rearrangement, it is possible that RAG mediates long-range locus conformation.

Architectural protein dependent control

The mammalian nucleus contains a finite space within which the genome must be folded and packaged. The organization of the linear genome into distinct three dimensional folded compartments is mediated by different combinations of architectural proteins bound to DNA, forming DNA-protein complexes. These proteins can include CTCF, Cohesin and Mediator. CTCF binds to a GC-rich consensus sequence involved in transcriptional activation, repression and insulationand can also functions as a chromatin organizer by mediating long-range genome

interactions. The most fundamental unit of the genome interaction is the loop. These can be categorized into structural loops or functional loops (Ong and Corces, 2011).

Structural loops help to package the mammalian genome, and are generated over long-range with CTCF and Cohesin at its base. Tens of thousands of CTCF binding sites are spread throughout the genome. Distally located CTCF-bound regions communicate with one another by looping out the intervening DNA sequence. The appropriate homodimerization of CTCF-bound DNA elements require the CTCF proteins to be recruited in a convergent orientation. This convergent orientation of CTCF sites are observed in more than 90% of CTCF sites in human cells (Rao et al., 2014). The *Tcrb* locus contains 22 CTCF sites, which are deposited throughout the Vβ cluster and on either ends of the RC (Shih et al., 2012). Surprisingly, out of a total 400 kb stretch of Trypsinogen regions (both 5' and 3'), there is only one CTCF bound element (dubbed 5'PC in Majumder et al., 2015). With regard to the orientations of CTCF sites on *Tcrb*, we have found that all Vβ CTCF sites, and 3 out of 4 RC-proximal CTCF sites are in mutually convergent orientations (Figure 1.3). These findings predict a CTCF-dependent molecular mechanism that could explain lineage-specific contraction of *Tcrb*, where Vβ CTCF sites and RC-associated CTCF sites are in favorable orientations to homodimerize.

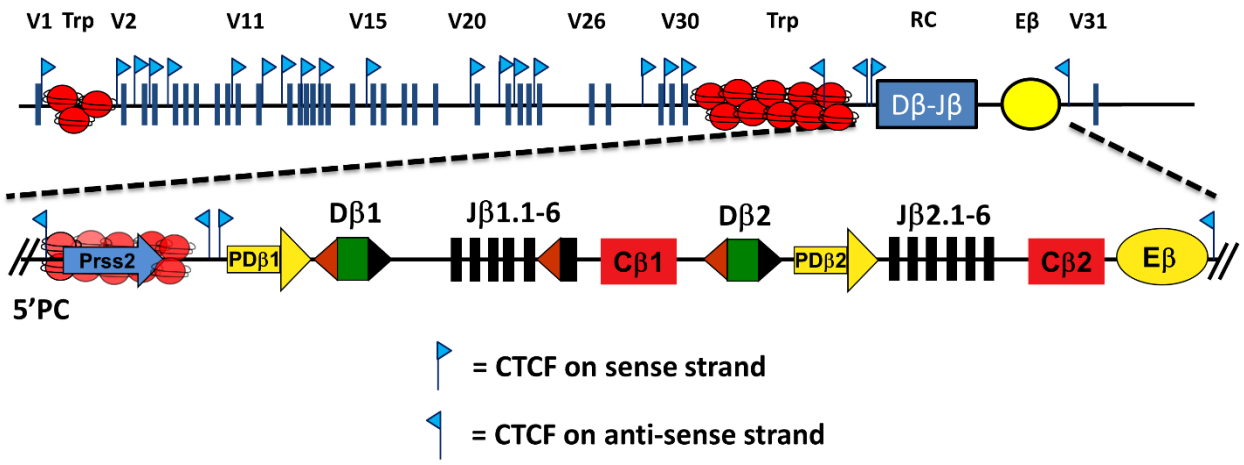


Figure 1.3: CTCF orientations on *Tcrb* locus. Schematic of the mouse Tcrb locus with a zoomed-in view of the RC. CTCF binding sites are demarcated by flags and their orientation is indicated at shown.

The homo-dimerization of distally bound CTCF sites are stabilized by the collar-like Cohesin complex, which forms a multi-subunit scaffold involved in sister chromatid cohesion. Cohesin complexes are made up of a core component (SMC1 and SMC3) and non-core component (RAD21). Cohesin forms a collar around homodimerized CTCF-bound sites, stabilizing the structural loop (Merkenschlager and Odom, 2013). In AgR loci, Cohesin has been shown to stabilize regulatory loops on *Tcra*, and influence thymocyte development (Seitan et al., 2011).

Regulatory loops reinforce promoter-enhancer communication, and are mediated by the Cohesin and Mediator proteins (Kagey et al., 2010). The assembly of Cohesin and Mediator complexes correlate with CTCF binding sites or transcriptional start sites. However, the stepwise mechanisms of cohesin recruitment and loop formation have not been worked out (Merkenschlager, 2010). On the *Tcrb* locus, the Eβ-mediated holocomplex at the RC region generates a regulatory loop. The CTCF sites bookending the *Tcrb*-RC, the germline promoters and the enhancer can serve as docking sites for cohesin and mediator subunits. Therefore, *Tcrb* is a

unique model for future studies on the underlying mechanisms that generate regulatory loops. In comparison, the *Tcra*-RC contains a TEAp-Eα regulatory loop that is stabilized by CTCF and RAD21 (Shih et al., 2012; Seitan et al., 2011). The first studies on the role of CTCF/Cohesin in stabilizing long range topology have performed in knockdown or knockout model systems. The observed impact on AgR topology can be an indirect effect due to the loss of these key regulatory proteins. This can be addressed by performing site-directed mutagenesis of CTCF binding sites (or the regulatory element of interest), as performed on the *Igh*-IGCR (Guo et al., 2011). Indeed, mutation of the *Igh*-IGCR sites diminished loops at the *Igh*-RC, between IGCR and 3'CBE, as well as their interactions with the Eμ enhancer.

Regulatory element mediated control

AgR loci attain their long-range conformations through a multi-step process to regulate recombination. The *Tcrb* locus adopts a unique architecture in DN thymocytes where the Vβ genes interact with the Dβ-Jβ array (Skok et al., 2007). However, the contact points mediating these long-range interactions remain unknown. One possible model to explain this *Tcrb* conformation came from studies of the *Igh* locus which suggested that pro-B cell specific contraction is mediated by Eμ, which can also impact the association of the RC with the VH gene cluster (Guo, Gerasimova et al., 2011). The molecular explanation of regulatory element mediated interactions may be the spatial apposition of the VH and RC regions into a shared transcription factory (Verma-Gaur et al., 2012). On the other hand, the *Tcra* enhancer Eα regulates long-range loops in a contraction independent manner (Shih et al., 2012). These discrepancies in AgR locus conformations may be due to the complex nature of the *Igh* and *Tcra* loci, which span long distances and contain multiple

regulatory elements. Studies on the *Tcrb* locus can help to resolve these issues due to the simpler and smaller arrangement of *Tcrb*. Regardless, the findings on both *Igh* and *Tcra* topology suggest that transcriptional regulatory elements mediate *Tcrb* topology.

It is possible that regulatory elements synergize with architectural proteins to generate AgR conformations. This mechanism was suggested by the presence of CTCF binding sites on Eα, which may influence the changes in Eα-dependent conformations (Shih et al., 2012). However, on *Tcrb*, Eβ does not exhibit CTCF binding. Instead, a CTCF binding site is located 1.5 kb downstream of Eβ. Loss of Eβ has a profound impact on the RC chromatin, but it is essential to know how the partially accessible RC communicates over long range. This led to analyses of the knockout models of germline promoters, where the RC remains partially active. The phenomenon by which CTCF-bound elements perform dual roles as long-range tethers or chromatin barriers is also observed on *Igh* by the IGCR element and on *Igk* by the Cer element (Guo et al., 2011; Xiang et al., 2013). The IGCR CTCF sites have been proposed to play a synergistic effect in regulating *Igh* architecture (Lin et al., 2015). In the only thymocyte-specific AgR locus studied so far, *Tcra*, it has been observed that loss of the TEA-promoter led to dissolution of contacts between its neighboring Jα gene segments with Eα. However, this did not influence *Tcra* contraction in DP thymocytes (Shih et al., 2012). It is possible that the PDβ1 CTCF function in the same way, but has not been tested.

Prior findings on the structure of the *Tcrb* locus have focused solely on the architecture of its distal elements by 3D-FISH assays, with supporting information using semi-quantitative 3C-

qPCR assays. However, the detailed architecture of *Tcrb*, the elements mediating this structure and what role it plays in T cell function have not been deciphered.

1.6 Sculpting the T-Cell Repertoire

The relative use of gene segments can influence the development of the primary repertoire in precursor lymphocytes. The use of V gene segments play an important role during the positive and negative selection processes, even influencing the development of unique lymphocyte subsets. For example, the use of *Trbv13-2* gene segment is a primary determinant of iNKT cells (Godfrey et al., 2000). The mechanisms selecting V gene segments must strike a balance between undergoing appropriate rearrangements and being able to recognize foreign and self-antigens. These considerations about the development of the adaptive immune system raised the question of what are the primary determinants that shape the pre-selection Vβ repertoire.

Studies of long-range locus contraction and looping have advanced the idea that perhaps spatial proximity is required for the Vβ gene segments to recombine with the RC. Furthermore, it raised the question whether relative proximity increased the likelihood of Vβ synapsis with the RC. In support of this model, Friedman and colleagues have shown through in-silico modelling of architecture that Jβ usage correlates with chromatin conformation at the RC (Ndifon et al., 2012). However, these findings were not extended to the Vβ repertoire. With regard to Vβ accessibility, deletion of the germline promoter associated with *Trbv14* abolishes the rearrangement of only this gene segment but not neighboring ones (Ryu et al., 2004). This finding suggests that Vβ

accessibility is locally regulated by their independent germline promoters, and they may play a role in subsequent RAG recruitment. However, the functional role played by the promoter on *Tcrb* repertoire selection is not known. Interestingly, promoter repositioning studies on the *Tcra* locus have shown that the poorly recombining *Trav12* genes are increased by knocking in the promoter of highly recombining *Trav15* genes, suggesting a direct role for promoter function on RAG recruitment and subsequent rearrangement (Naik et al., 2015). Regardless, since the focal zone of RAG binding is at the RC, the V and DJ regions must communicate spatially, which should also influence rearrangement potential. In contrast with *Tcrb*, the *Igh* locus spans 2.7 Mb, and is made up of 200 VH gene segments. Analyses of the VH repertoire divided *Igh* into four epigenetically and transcriptionally discreet domains. The VH genes proximal to the RC are not packaged in chromatin containing active histone modifications and noncoding RNA, but are close to CTCF sites suggesting they form the base of VH loops with the RC. On the other hand, the distally located VH genes are associated with strong histone marks and noncoding RNA, but are far away from CTCF sites. These findings suggest that the large size of *Igh* has forced the evolution of compensatory mechanisms that sculpt the VH repertoire (Choi et al., 2013).

The hallmarks of repertoire formation that remain to be tested include: the function of RSS qualities and how they can differentially recruit RAG proteins. Indeed, it has been shown that the non-core RAG1 complex can overcome inefficiencies in RSS quality (Horowitz et al., 2014). However, it is unclear whether this plays a dominant role over promoter quality in sculpting the pre-selection repertoire. Similarly, the efficiency of rearrangements and subsequent formation of the repertoire can be influenced by the DNA-repair machinery. ATM has been shown to play a role in this process (Hathcock et al., 2013). In support of these conclusion, AT-mutated patients

exhibit a skewed *Tcrb* repertoire with a decreased thymic output (Giovannetti et al., Blood). However, the factors that contribute to repertoire selection remain elusive. This can include transcription, chromatin accessibility, RSS quality or spatial proximity.

1.7 Developmental changes in *Tcrb* expression

During thymocyte development, several mechanisms regulate the expression of *Tcrb* after the stages in which it has undergone rearrangement. Cell surface expression of the TCRβ protein with the surrogate light chain (pTα), along with signaling components of the CD3 complex leads to pre-TCR signaling through the ZAP70-mediated signal transduction pathway (von Boehmer, 2005). These events lead to the phosphorylation and degradation of RAG proteins, terminating further rearrangements of the *Tcrb* locus. Subsequent proliferation and differentiation of DN cells leads to the expression of the CD4/CD8 co-receptors on the surface, generating DP thymocytes. DP thymocytes re-express RAG proteins which now rearrange the *Tcra* locus, but not *Tcrb*. Out of the two *Tcrb* alleles in this stage, one is fully rearranged, but no longer undergoes secondary rearrangement. Meanwhile, the second allele contains DβJβ but no Vβ rearrangements and is referred to as "allelically excluded". However, the molecular mechanisms exerting these silencing mechanisms on the rearranged and the excluded alleles remain enigmatic.

Allelic exclusion

The mono-allelic expression of *Tcrb* in DP thymocytes, where the second *Tcrb* allele is silenced, is referred to as allelic exclusion. The mechanisms of this transcriptional silencing has been attributed to changes in accessibility control, which influences RAG recruitment (Brady and Bassing, 2010). The signals for allelic exclusion emanate from the rearranged (and cell surface expressed) T-Cell Receptor. This is known as feedback inhibition, and are experimentally mimicked in RAG-deficient mice by expressing TCR transgenes or by anti-CD3ε injections, thus providing models that allow the study of mono-allelic silencing of the germline *Tcrb* allele (Chattopadhyay et al., 1998; Shinkai and Alt, 1994).

Multiple mechanisms play a role in enforcing the transcriptional silencing of Vβ genes in DP thymocytes. Firstly, transcriptional accessibility, as detected by active chromatin modifications such as histone acetylation are erased in the germline Vβ segments (Tripathi et al., 2002). This erasure of active epigenetic landscape correlates with the loss of germline transcripts (Chattopadhyay et al., 1998,). These phenomenon may be a result of repositioning the *Tcrb* locus towards the transcriptionally repressive nuclear periphery in DP cells, which has been proposed as a potential mechanism of allelic exclusion (Skok et al., 2007; Reddy et al., 2008). In support of this model, the nuclear periphery has been shown to contain a dearth of RAG2, which would form a recombinationally unfavorable nuclear environment (Chan et al., 2013).

The DNA upstream of PDβ1 is also rich in GAGA sequences. These sequences have been shown to target to the nuclear lamina. Indeed, DamID-sequencing assays have shown that this GAGA-rich region coincide with LAD borders in committed cells (Peric-Hupkes et al., 2010).

Association with the transcriptionally repressive compartment of the nuclear lamina may function as a spatial-segregation mechanism to isolate inactive chromatin. 3D-FISH studies have shown that the *Tcrb* locus moves from a central to peripheral location in the nucleus during the DN-to-DP transition (Skok et al., 2007). This movement places the locus in the nuclear periphery, thus preventing additional rearrangement and possibly exerting allelic exclusion (Schlimgen et al., 2008). A similar mechanism of allelic exclusion has been proposed for *Igh* locus (Roldan et al., 2008). However, further studies are needed to address how *Tcrb* is localized to the nuclear lamina during thymocyte development.

In addition to nuclear localization, feedback inhibition may exerted by spatial separation of the distal *Tcrb* ends. Indeed, early 3D-FISH and 3C studies suggested that *Tcrb* undergoes a "decontraction" in DP cells, where distal ends are spatially segregated from one another (Skok et al., 2007). However, fine mapping of the *Tcrb* loops during developmental transition are yet to be performed. In contrast, the distal ends of *Tcra* have been shown to decontract upon differentiation leaving the proximal ends contracted (Shih and Krangel, 2010). The molecular machinery forming and stabilizing these loops remain an enigma.

A possible mechanism of *Tcrb* decontraction is the proliferative burst that thymocytes undergo when differentiating from the DN to DP stage. During the DN to DP transition, thymocytes divide 10-11 times, during which their chromosome architecture is disassembled in pre-mitotic stages and reassembled in G0/G1 phase (Naumova et al., 2003). This raises the unique possibility of two independent mechanisms that modulate chromatin architecture: one to facilitate breakdown and another to perform reassembly. The cyclin dependent kinases that play a dominant

role in many cellular processes during cell division may play a role in modulating *Tcrb* architecture during the proliferative burst.

Alternate mechanisms may confer the DP cell-specific chromosome conformation. This may include tissue-specific transcription factors. Indeed, overexpression of E2A overrides the feedback inhibition signals at the Vβ cluster, increasing the accessibility of the *Trbv* gene segments (Agata et al., 2007). It would be interesting to test whether *Tcrb* adopts a contracted and looped conformation once again upon E2A overexpression in these DP thymocytes. Therefore, studies over the last decade have shown that regulation of *Tcrb* allelic exclusion is a multifactorial process that requires the synergism of chromatin accessibility, conformation, nuclear localization and transcription factors and cell cycle proteins. With the adoption of assays that can probe genome folding at higher resolution, current models are being revised to include the formation of distal versus proximal V-subdomains. These models suggest that long range topological changes associated with allelic exclusion are restricted to the distal V-gene segments but not proximal V-genes.

Regulating the rearranged Tcrb locus

In addition to allelic exclusion, successful rearrangements also require the inhibition of secondary Vβ to Dβ2 rearrangements in DP thymocytes on VDJβ rearranged alleles. Regulation of chromatin accessibility can influence this process. It has been shown in Vβ1NT mice containing a functional *Trbv5*-Dβ1-Jβ1.4 rearrangements that the region upstream of the rearranged gene is CpG methylated and repressed for active chromatin marks. These features of inaccessible chromatin inhibit RAG recruitment, and subsequent secondary rearrangements (Brady, Oropallo

et al., 2010). The molecular mechanisms exerting inaccessible chromatin on the rearranged locus remain unknown. It is possible that Eβ plays a role in this process through differential looping.

The roles of feedback inhibition in regulating allelic exclusion have led to the formation of the lateral inhibition model. This model posits that DNA damage responses generated at one allele of *Tcrb* may upregulate ATM-dependent pathways. These include phosphorylation of genes involved in transcription and epigenetic modification (Alt et al., 1980; Bredemeyer et al., 2008). Epigenetic modifiers, when upregulated, can alter chromatin accessibility and locus architecture, leading to the inhibition of further recombination from the second allele (Brady, Steinel et al., 2010).

1.8 Summary

In this dissertation, we have focused on the spatial and epigenetic mechanisms that sculpt the formation of the pre-selection T cell repertoire, developing a tiered model for Vβ selection during rearrangement. We have further investigated the role of transcriptional regulatory elements and architectural proteins that help to generate the tissue-specific conformation of *Tcrb*. Our findings show that an incompatibility exists between the formation of boundary and long-range tethering for CTCF-bound elements. Finally, we have revisited the structural changes in *Tcrb* topology in greater detail, which has been thought to help exert allelic exclusion and feedback inhibition. Our findings reveal that novel mechanisms beyond spatial proximity exist that alter *Tcrb* accessibility upon transition of DN thymocytes to DP stages of development.

1.9 References

Abarrategui, I., & Krangel, M. S. (2006). Regulation of T cell receptor-α gene recombination by transcription. *Nature immunology*, *7*(10), 1109-1115.

Agata, Y., Tamaki, N., Sakamoto, S., Ikawa, T., Masuda, K., Kawamoto, H., & Murre, C. (2007). Regulation of T cell receptor β gene rearrangements and allelic exclusion by the helix-loop-helix protein, E47. *Immunity*, *27*(6), 871-884.

Alt, F. W., Enea, V., Bothwell, A. L., & Baltimore, D. (1980). Activity of multiple light chain genes in murine myeloma cells producing a single, functional light chain. *Cell*, *21*(1), 1-12.

Bailey, J. A., Gu, Z., Clark, R. A., Reinert, K., Samonte, R. V., Schwartz, S., ... & Eichler, E. E. (2002). Recent segmental duplications in the human genome. *Science*, *297*(5583), 1003-1007.

Bassing, C. H., Alt, F. W., Hughes, M. M., D'Auteuil, M., Wehrly, T. D., Woodman, B. B., ... & Sleckman, B. P. (2000). Recombination signal sequences restrict chromosomal V (D) J recombination beyond the 12/23 rule.*Nature*, *405*(6786), 583-586.

Bassing, C. H., Swat, W., & Alt, F. W. (2002). The mechanism and regulation of chromosomal V (D) J recombination. *Cell*, *109*(2), S45-S55.

Bonnet, M., Huang, F., Benoukraf, T., Cabaud, O., Verthuy, C., Boucher, A., ... & Spicuglia, S. (2009). Duality of enhancer functioning mode revealed in a reduced TCRβ gene enhancer knockin mouse model. *The Journal of Immunology*, *183*(12), 7939-7948.

Bories, J. C., Demengeot, J., Davidson, L., & Alt, F. W. (1996). 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.*Proceedings of the National Academy of Sciences*, *93*(15), 7871-7876.

Bouvier, G., Watrin, F., Naspetti, M., Verthuy, C., Naquet, P., & Ferrier, P. (1996). Deletion of the mouse T-cell receptor beta gene enhancer blocks alphabeta T-cell development. *Proceedings of the National Academy of Sciences*, *93*(15), 7877-7881.

Brady, B. L., & Bassing, C. H. (2011). Differential Regulation of Proximal and Distal Vβ Segments Upstream of a Functional VDJβ1 Rearrangement upon β-Selection. *The Journal of Immunology*, *187*(6), 3277-3285.

Brady, B. L., Oropallo, M. A., Yang-Iott, K. S., Serwold, T., Hochedlinger, K., Jaenisch, R., ... & Bassing, C. H. (2010). Position-dependent silencing of germline Vβ segments on TCRβ alleles containing preassembled VβDJβCβ1 genes. *The Journal of Immunology*, *185*(6), 3564-3573.

Brady, B. L., Steinel, N. C., & Bassing, C. H. (2010). Antigen receptor allelic exclusion: an update and reappraisal. *The Journal of Immunology*, *185*(7), 3801-3808.

Bredemeyer, A. L., Helmink, B. A., Innes, C. L., Calderon, B., McGinnis, L. M., Mahowald, G. K., ... & Sleckman, B. P. (2008). DNA double-strand breaks activate a multi-functional genetic program in developing lymphocytes. *Nature*,*456*(7223), 819-823.

Carabana, J., Watanabe, A., Hao, B., & Krangel, M. S. (2011). A barrier-type insulator forms a boundary between active and inactive chromatin at the murine TCRβ locus. *The Journal of Immunology*, *186*(6), 3556-3562.

Chan, E. A., Teng, G., Corbett, E., Choudhury, K. R., Bassing, C. H., Schatz, D. G., & Krangel, M. S. (2013). Peripheral subnuclear positioning suppresses Tcrb recombination and segregates Tcrb alleles from RAG2. *Proceedings of the National Academy of Sciences*, *110*(48), E4628-E4637.

Chang, J., Zhang, B., Heath, H., Galjart, N., Wang, X., & Milbrandt, J. (2010). Nicotinamide adenine dinucleotide (NAD)–regulated DNA methylation alters CCCTC-binding factor (CTCF)/cohesin binding and transcription at the BDNF locus. *Proceedings of the National Academy of Sciences*, *107*(50), 21836-21841.

Chattopadhyay, S., Whitehurst, C. E., Schwenk, F., & Chen, J. (1998). Biochemical and functional analyses of chromatin changes at the TCR-β gene locus during CD4− CD8− to CD4+ CD8+ thymocyte differentiation. *The Journal of Immunology*, *160*(3), 1256-1267.

Choi, N. M., Loguercio, S., Verma-Gaur, J., Degner, S. C., Torkamani, A., Su, A. I., ... & Feeney, A. J. (2013). Deep sequencing of the murine Igh repertoire reveals complex regulation of nonrandom V gene rearrangement frequencies.*The Journal of Immunology*, *191*(5), 2393-2402.

Cobb, R. M., Oestreich, K. J., Osipovich, O. A., & Oltz, E. M. (2006). Accessibility control of V (D) J recombination. *Advances in immunology*, *91*, 45-109.

Dekker, J., Marti-Renom, M. A., & Mirny, L. A. (2013). Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nature Reviews Genetics*, *14*(6), 390-403.

Derikx, M. H., Kovacs, P., Scholz, M., Masson, E., Chen, J. M., Ruffert, C., ... & Mora, J. (2014). Polymorphisms at PRSS1–PRSS2 and CLDN2–MORC4 loci associate with alcoholic and non-alcoholic chronic pancreatitis in a European replication study. *Gut*, gutjnl-2014.

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., ... & Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, *485*(7398), 376-380.

Engel, N., West, A. G., Felsenfeld, G., & Bartolomei, M. S. (2004). Antagonism between DNA hypermethylation and enhancer-blocking activity at the H19 DMD is uncovered by CpG mutations. *Nature genetics*, *36*(8), 883-888.

Ferrier, P., Krippl, B., Blackwell, T. K., Furley, A. J., Suh, H. E. I. K. Y. U. N. G., Winoto, A., ... & Alt, F. W. (1990). Separate elements control DJ and VDJ rearrangement in a transgenic recombination substrate. *The EMBO journal*, *9*(1), 117.

Fudenberg, G., & Mirny, L. A. (2012). Higher-order chromatin structure: bridging physics and biology. *Current opinion in genetics & development*, *22*(2), 115-124.

Gellert, M. (2002). V (D) J Recombination: RAG Proteins, Repair Factors, and Regulation*. *Annual review of biochemistry*, *71*(1), 101-132.

Giovannetti, A., Mazzetta, F., Caprini, E., Aiuti, A., Marziali, M., Pierdominici, M., ... & Fiorilli, M. (2002). Skewed T-cell receptor repertoire, decreased thymic output, and predominance of terminally differentiated T cells in ataxia telangiectasia. *Blood*, *100*(12), 4082-4089.

Glusman, G., Rowen, L., Lee, I., Boysen, C., Roach, J. C., Smit, A. F., ... & Hood, L. (2001). Comparative genomics of the human and mouse T cell receptor loci. *Immunity*, *15*(3), 337-349.

Godfrey, D. I., Hammond, K. J., Poulton, L. D., Smyth, M. J., & Baxter, A. G. (2000). NKT cells: facts, functions and fallacies. *Immunology today*, *21*(11), 573-583.

Gopalakrishnan, S., Majumder, K., Predeus, A., Huang, Y., Koues, O. I., Verma-Gaur, J., ... & Oltz, E. M. (2013). Unifying model for molecular determinants of the preselection Vβ repertoire. *Proceedings of the National Academy of Sciences*, *110*(34), E3206-E3215.

Grosberg, A. Y., Nechaev, S. K., & Shakhnovich, E. I. (1988). The role of topological constraints in the kinetics of collapse of macromolecules. *Journal de physique*, *49*(12), 2095-2100.

Guastafierro, T., Cecchinelli, B., Zampieri, M., Reale, A., Riggio, G., Sthandier, O., ... & Caiafa, P. (2008). CCCTC-binding factor activates PARP-1 affecting DNA methylation machinery. *Journal of Biological Chemistry*, *283*(32), 21873-21880.

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M. B., Talhout, W., ... & van Steensel, B. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*, *453*(7197), 948-951.

Guo, C., Yoon, H. S., Franklin, A., Jain, S., Ebert, A., Cheng, H. L., ... & Alt, F. W. (2011). CTCF-binding elements mediate control of V (D) J recombination. *Nature*, *477*(7365), 424-430.

Hathcock, K. S., Bowen, S., Livak, F., & Hodes, R. J. (2013). ATM influences the efficiency of TCRβ rearrangement, subsequent TCRβ-dependent T cell development, and generation of the pre-selection TCRβ CDR3 repertoire. *PloS one*, *8*(4), e62188.

Helmink, B. A., & Sleckman, B. P. (2012). The Response to and Repair of RAG-Mediated DNA Double Stranded Breaks. *Annual review of immunology*,*30*, 175.

Horowitz, J. E., & Bassing, C. H. (2014). Noncore RAG1 Regions Promote Vβ Rearrangements and αβ T Cell Development by Overcoming Inherent Inefficiency of Vβ Recombination Signal Sequences. *The Journal of Immunology*, *192*(4), 1609-1619.

Ji, Y., Resch, W., Corbett, E., Yamane, A., Casellas, R., & Schatz, D. G. (2010). The in vivo pattern of binding of RAG1 and RAG2 to antigen receptor loci. *Cell*, *141*(3), 419-431.

Kagey, M. H., Newman, J. J., Bilodeau, S., Zhan, Y., Orlando, D. A., van Berkum, N. L., ... & Young, R. A. (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature*, *467*(7314), 430-435.

Lai, A. Y., Fatemi, M., Dhasarathy, A., Malone, C., Sobol, S. E., Geigerman, C., ... & Wade, P. A. (2010). DNA methylation prevents CTCF-mediated silencing of the oncogene BCL6 in B cell lymphomas. *The Journal of experimental medicine*, *207*(9), 1939-1950.

Lane, J., Duroux, P., & Lefranc, M. P. (2010). From IMGT-ONTOLOGY to IMGT/LIGMotif: the IMGT® standardized approach for immunoglobulin and T cell receptor gene identification and description in large genomic sequences.*BMC bioinformatics*, *11*(1), 223.

Liang, H. E., Hsu, L. Y., Cado, D., Cowell, L. G., Kelsoe, G., & Schlissel, M. S. (2002). The "dispensable" portion of RAG2 is necessary for efficient V-to-DJ rearrangement during B and T cell development. *Immunity*, *17*(5), 639-651.

Lin, Y. C., Benner, C., Mansson, R., Heinz, S., Miyazaki, K., Miyazaki, M., ... & Murre, C. (2012). Global changes in the nuclear positioning of genes and intra-and interdomain genomic interactions that orchestrate B cell fate. *Nature immunology*, *13*(12), 1196-1204.

Lin, S. G., Guo, C., Su, A., Zhang, Y., & Alt, F. W. (2015). CTCF-binding elements 1 and 2 in the Igh intergenic control region cooperatively regulate V (D) J recombination. *Proceedings of the National Academy of Sciences*, *112*(6), 1815-1820.

Liu, Y., Subrahmanyam, R., Chakraborty, T., Sen, R., & Desiderio, S. (2007). A plant homeodomain in RAG-2 that binds Hypermethylated lysine 4 of histone H3 is necessary for efficient antigen-receptor-gene rearrangement. *Immunity*,*27*(4), 561-571.

Majumder, K., Koues, O. I., Chan, E. A., Kyle, K. E., Horowitz, J. E., Yang-Iott, K., ... & Oltz, E. M. (2015). Lineage-specific compaction of Tcrb requires a chromatin barrier to protect the

function of a long-range tethering element. *The Journal of Experimental Medicine*, *212*(1), 107-120.

Mathieu, N., Hempel, W. M., Spicuglia, S., Verthuy, C., & Ferrier, P. (2000). Chromatin Remodeling by the T Cell Receptor (Tcr)-β Gene Enhancer during Early T Cell Development Implications for the Control of Tcr-β Locus Recombination. *The Journal of experimental medicine*, *192*(5), 625-636.

Matthews, A. G., Kuo, A. J., Ramón-Maiques, S., Han, S., Champagne, K. S., Ivanov, D., ... & Oettinger, M. A. (2007). RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V (D) J recombination. *Nature*, *450*(7172), 1106-1110.

McDougall, S., Peterson, C. L., & Calame, K. (1988). A transcriptional enhancer 3'of C beta 2 in the T cell receptor beta locus. *Science*, *241*(4862), 205-208.

McMillan, R. E., & Sikes, M. L. (2008). Differential activation of dual promoters alters Dβ2 germline transcription during thymocyte development. *The Journal of Immunology*, *180*(5), 3218-3228.

Merkenschlager, M. (2010). Cohesin: a global player in chromosome biology with local ties to gene regulation. *Current opinion in genetics & development*, *20*(5), 555-561.

Merkenschlager, M., & Odom, D. T. (2013). CTCF and cohesin: linking gene regulatory elements with their targets. *Cell*, *152*(6), 1285-1297.

Naik, A. K., Hawwari, A., & Krangel, M. S. (2015). Specification of Vδ and Vα Usage by Tcra/Tcrd Locus V Gene Segment Promoters. *The Journal of Immunology*, *194*(2), 790-794.

Naumova, N., Imakaev, M., Fudenberg, G., Zhan, Y., Lajoie, B. R., Mirny, L. A., & Dekker, J. (2013). Organization of the mitotic chromosome. *Science*, *342*(6161), 948-953.

Ndifon, W., Gal, H., Shifrut, E., Aharoni, R., Yissachar, N., Waysbort, N., ... & Friedman, N. (2012). Chromatin conformation governs T-cell receptor Jβ gene segment usage. *Proceedings of the National Academy of Sciences*, *109*(39), 15865-15870.

Oestreich, K. J., Cobb, R. M., Pierce, S., Chen, J., Ferrier, P., & Oltz, E. M. (2006). Regulation of TCRβ gene assembly by a promoter/enhancer holocomplex. *Immunity*, *24*(4), 381-391.

Ong, C. T., & Corces, V. G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nature Reviews Genetics*, *12*(4), 283-293.

Osipovich, O., Cobb, R. M., Oestreich, K. J., Pierce, S., Ferrier, P., & Oltz, E. M. (2007). Essential function for SWI-SNF chromatin-remodeling complexes in the promoter-directed assembly of Tcrb genes. *Nature immunology*, *8*(8), 809-816.

Osipovich, O., Milley, R., Meade, A., Tachibana, M., Shinkai, Y., Krangel, M. S., & Oltz, E. M. (2004). Targeted inhibition of V (D) J recombination by a histone methyltransferase. *Nature immunology*, *5*(3), 309-316.

Osipovich, O., & Oltz, E. M. (2010, December). Regulation of antigen receptor gene assembly by genetic–epigenetic crosstalk. In *Seminars in immunology* (Vol. 22, No. 6, pp. 313-322). Academic Press.

Peric-Hupkes, D., Meuleman, W., Pagie, L., Bruggeman, S. W., Solovei, I., Brugman, W., ... & van Steensel, B. (2010). Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation.*Molecular cell*, *38*(4), 603-613.

Rao, S. S., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., ... & Aiden, E. L. (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell*, *159*(7), 1665-1680.

Reddy, K. L., Zullo, J. M., Bertolino, E., & Singh, H. (2008). Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature*, *452*(7184), 243-247.

Rodriguez, C., Borgel, J., Court, F., Cathala, G., Forné, T., & Piette, J. (2010). CTCF is a DNA methylation-sensitive positive regulator of the INK/ARF locus.*Biochemical and biophysical research communications*, *392*(2), 129-134.

Roldán, E., Fuxa, M., Chong, W., Martinez, D., Novatchkova, M., Busslinger, M., & Skok, J. A. (2005). Locus' decontraction'and centromeric recruitment contribute to allelic exclusion of the immunoglobulin heavy-chain gene. *Nature immunology*, *6*(1), 31-41.

Ryu, C. J., Haines, B. B., Lee, H. R., Kang, Y. H., Draganov, D. D., Lee, M., ... & Chen, J. (2004). The T-cell receptor β variable gene promoter is required for efficient Vβ rearrangement but not allelic exclusion. *Molecular and cellular biology*, *24*(16), 7015-7023.

Schatz, D. G., & Ji, Y. (2011). Recombination centres and the orchestration of V (D) J recombination. *Nature Reviews Immunology*, *11*(4), 251-263.

Schlimgen, R. J., Reddy, K. L., Singh, H., & Krangel, M. S. (2008). Initiation of allelic exclusion by stochastic interaction of Tcrb alleles with repressive nuclear compartments. *Nature immunology*, *9*(7), 802-809.

Seitan, V. C., Hao, B., Tachibana-Konwalski, K., Lavagnolli, T., Mira-Bontenbal, H., Brown, K. E., ... & Merkenschlager, M. (2011). A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation. *Nature*, *476*(7361), 467-471.

Senoo, M., & Shinkai, Y. (1998). Regulation of Vbeta germline transcription in RAG-deficient mice by the CD3epsilon-mediated signals: implication of Vbeta transcriptional regulation in TCR beta allelic exclusion. *International immunology, 10*(5), 553-560.

Shih, H. Y., & Krangel, M. S. (2010). Distinct contracted conformations of the Tcra/Tcrd locus during Tcra and Tcrd recombination. *The Journal of experimental medicine*, *207*(9), 1835-1841.

Shih, H. Y., & Krangel, M. S. (2013). Chromatin architecture, CCCTC-binding factor, and V (D) J recombination: managing long-distance relationships at antigen receptor loci. *The Journal of Immunology*, *190*(10), 4915-4921.

Shih, H. Y., Verma-Gaur, J., Torkamani, A., Feeney, A. J., Galjart, N., & Krangel, M. S. (2012). Tcra gene recombination is supported by a Tcra enhancer-and CTCF-dependent chromatin hub. *Proceedings of the National Academy of Sciences*, *109*(50), E3493-E3502.

Shinkai, Y., & Alt, F. W. (1994). CD3ε-mediated signals rescue the development of CD4+ CD8+ thymocytes in RAG-2−/− mice in the absence of TCR β chain expression. *International immunology*, *6*(7), 995-1001.

Shrimali, S., Srivastava, S., Varma, G., Grinberg, A., Pfeifer, K., & Srivastava, M. (2012). An ectopic CTCF-dependent transcriptional insulator influences the choice of Vβ gene segments for VDJ recombination at TCRβ locus. *Nucleic acids research*, *40*(16), 7753-7765.

Sikes, M. L., Meade, A., Tripathi, R., Krangel, M. S., & Oltz, E. M. (2002). Regulation of V (D) J recombination: a dominant role for promoter positioning in gene segment accessibility. *Proceedings of the National Academy of Sciences*, *99*(19), 12309-12314.

Sikes, M. L., Suarez, C. C., & Oltz, E. M. (1999). Regulation of V (D) J recombination by transcriptional promoters. *Molecular and cellular biology*,*19*(4), 2773-2781.

Skok, J. A., Gisler, R., Novatchkova, M., Farmer, D., de Laat, W., & Busslinger, M. (2007). Reversible contraction by looping of the Tcra and Tcrb loci in rearranging thymocytes. *Nature immunology*, *8*(4), 378-387.

Sleckman, B. P., Bassing, C. H., Hughes, M. M., Okada, A., D'Auteuil, M., Wehrly, T. D., ... & Alt, F. W. (2000). Mechanisms that direct ordered assembly of T cell receptor β locus V, D, and J gene segments. *Proceedings of the National Academy of Sciences*, *97*(14), 7975-7980.

Spicuglia, S., Kumar, S., Yeh, J. H., Vachez, E., Chasson, L., Gorbatch, S., ... & Ferrier, P. (2002). Promoter activation by enhancer-dependent and-independent loading of activator and coactivator complexes. *Molecular cell*, *10*(6), 1479-1487.

Xiang, Y., Park, S. K., & Garrard, W. T. (2013). Vκ gene repertoire and locus contraction are specified by critical DNase I hypersensitive sites within the Vκ-Jκ intervening region. *The Journal of Immunology*, *190*(4), 1819-1826.

Tripathi, R., Jackson, A., & Krangel, M. S. (2002). A change in the structure of Vβ chromatin associated with TCR β allelic exclusion. *The Journal of Immunology*, *168*(5), 2316-2324.

Van Ness, B. G., Weigert, M., Coleclough, C., Mather, E. L., Kelley, D. E., & Perry, R. P. (1981). Transcription of the unrearranged mouse C κ locus: sequence of the initiation region and comparison of activity with a rearranged V κ-C κ gene. *Cell*, *27*(3), 593-602.

Verma-Gaur, J., Torkamani, A., Schaffer, L., Head, S. R., Schork, N. J., & Feeney, A. J. (2012). Noncoding transcription within the Igh distal VH region at PAIR elements affects the 3D structure of the Igh locus in pro-B cells.*Proceedings of the National Academy of Sciences*, *109*(42), 17004-17009.

Von Boehmer, H. (2005). Unique features of the pre-T-cell receptor α-chain: not just a surrogate. *Nature Reviews Immunology*, *5*(7), 571-577.

Wang, H., Maurano, M. T., Qu, H., Varley, K. E., Gertz, J., Pauli, F., ... & Stamatoyannopoulos, J. A. (2012). Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome research*, *22*(9), 1680-1688.

Whitcomb, D. C., LaRusch, J., Krasinskas, A. M., Klei, L., Smith, J. P., Brand, R. E., ... & Zarnescu, N. O. (2012). Common genetic variants in the CLDN2 and PRSS1-PRSS2 loci alter risk for alcohol-related and sporadic pancreatitis.*Nature genetics*, *44*(12), 1349-1354.

Whitehurst, C. E., Schlissel, M. S., & Chen, J. (2000). Deletion of germline promoter PDβ1 from the TCRβ locus causes hypermethylation that impairs Dβ1 recombination by multiple mechanisms. *Immunity*, *13*(5), 703-714.

Yancopoulos, G. D., & Alt, F. W. (1986). Regulation of the assembly and expression of variable-region genes. *Annual review of immunology*, *4*(1), 339-368.

Yang, J., & Corces, V. G. (2012). Insulators, long-range interactions, and genome function. *Current opinion in genetics & development*, *22*(2), 86-92.

Zampieri, M., Tiziana, G., Roberta, C., Fabio, C., Maria, G. B., Anna, R., ... & Paola, C. (2012). ADP-ribose polymers localized on Ctcf-Parp1-Dnmt1 complex prevent methylation of Ctcf target sites. *Biochemical Journal*, *441*(2), 645-652.

Zhang, Y., McCord, R. P., Ho, Y. J., Lajoie, B. R., Hildebrand, D. G., Simon, A. C., ... & Dekker, J. (2012). Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell*, *148*(5), 908-921.

Chapter 2: Unifying Model for Molecular Determinants of the Preselection Vβ Repertoire

This paper has been published in the Proceedings of the National Academy of Sciences:

Gopalakrishnan, S., Majumder, K., Predeus, A., Huang, Y., Koues, O. I., Verma-Gaur, J., ... & Oltz, E. M. (2013). Unifying model for molecular determinants of the preselection Vβ repertoire. *Proceedings of the National Academy of Sciences*, *110*(34), E3206-E3215.

2.1 Abstract

The primary antigen receptor repertoire is sculpted by the process of V(D)J recombination, which must strike a balance between diversification and favoring gene segments with specialized functions. The precise determinants of how often gene segments are chosen to complete variable region coding exons remain elusive. We quantified Vβ use in the preselection *Tcrb* repertoire and report relative contributions of 13 distinct features that may shape their recombination efficiencies, including transcription, chromatin environment, spatial proximity to their DβJβ targets, and predicted quality of recombination signal sequences (RSSs). We show that, in contrast to functional Vβ gene segments, all pseudo-Vβ segments are sequestered in transcriptionally silent chromatin, which effectively suppresses wasteful recombination. Importantly, computational analyses provide a unifying model, revealing a minimum set of five parameters that are predictive of Vβ use, dominated by chromatin modifications associated with transcription, but largely independent of precise spatial proximity to DβJβ clusters. This learned model-building strategy may be useful in predicting the relative contributions of epigenetic, spatial, and RSS features in shaping preselection V repertoires at other antigen receptor loci. Ultimately, such models may also predict how designed or naturally occurring alterations of these loci perturb the preselection use of variable gene segments.

2.2 Introduction

Gene activity is regulated at multiple levels to coordinate expression during development. At a most basic level, the collection of cis-acting elements for a genetic locus recruits transcription factors that alter its chromatin environment to either induce or repress gene activity. Emerging studies indicate that the 3D conformation of a locus also plays an important role in the regulation of its composite genes (Dekker, 2008). At most genes, many levels of control are integrated to achieve the requisite gene expression state. For example, transcriptional promoters interact with their cognate enhancers over considerable distances in the linear genome to generate "hubs" where the two *cis* elements are in spatial proximity (Dekker, 2008; Shih et al., 2012).

All of these regulatory strategies are used to generate functional Ig (*Ig*) and T-cell receptor (*Tcr*) genes during lymphocyte development (Bossen et al., 2012). Each antigen receptor (AgR) locus is composed of multiple variable (V), joining (J), and sometimes diversity (D) gene segments that are assembled by the process of V(D)J recombination, creating a potential variable region exon (Bassing et al., 2002). Recombination is mediated by the RAG-1/2 enzymatic complex, which is expressed in all developing lymphocytes and recognizes semiconserved recombination signal sequences (RSSs) flanking all AgR gene segments (Schatz and Ji, 2011). On selection of two compatible gene segments by RAG-1/2, recombination proceeds via a DNA break/repair mechanism, ultimately fusing the two selected segments (Bassing et al., 2002; Schatz and Ji, 2011).

The assembly of AgR genes is strictly regulated despite a common collection of genomic RSS targets and expression of recombinase in all resting (G0/G1) lymphocyte precursors (Cobb et al., 2006). The most obvious level of regulation is lineage specificity. The RAG-1/2 complex assembles *Tcr* genes in precursor T cells, whereas *Ig* genes are targeted in precursor B cells. Even

within an AgR locus, gene segment recombination is ordered, with D–J rearrangements preceding V–DJ. Numerous studies support a key role for chromatin accessibility in determining the recombination potential of gene segments (Feeney, 2009). The primary RAG-1/2 targets in a given cell type are transcriptionally active and DNase hypersensitive, two hallmarks of accessible chromatin. Indeed, RAG-2 binds directly to a histone modification that accompanies transcription [trimethylated histone H3 lysine 4 (H3K4me3)], providing a link between chromatin and recombinase targeting (Liu et al., 2009; Matthews et al., 2007). At all AgR loci, activation of (D)J clusters is dependent on communication between at least one distal enhancer and a proximal promoter, which triggers transcription of the unrearranged (D)J segments (Oestreich et al., 2006). Recent studies indicate that the high transcriptional activity focuses RAG-1/2 binding at (D)J clusters, forming "recombination centers" into which V gene segments must be brought (Ji et al., 2011).

Although chromatin accessibility explains most aspects of RAG-1/2 deposition at recombination centers, this feature is not sufficient to ensure rearrangement of the distant V segments. Insertion of a powerful *Tcra* enhancer (Ea) into *Tcrb* maintains chromatin accessibility at nearby Vβ gene segments but does not facilitate their recombination at a stage of thymocyte development in which only *Tcra* genes rearrange (Jackson et al., 2005). Subsequent studies have shown that long-range recombination of V segments requires changes in the 3D structure of an AgR locus, bringing the V cluster into spatial proximity with (D)J recombination centers located up to 3.2 Mb away (Guo, Gerasimova et al., 2011; Jhunjhunwala et al., 2008; Skok et al., 2007). Long-range interactions and locus conformations are determined in large part by CCCTC-binding factor (CTCF) and cohesin, factors that bind numerous sites throughout the mammalian genome

forming loops containing the intervening DNA (Rubio et al., 2008). With regard to AgR loci, deletion of CTCF, its binding sites, or essential cohesion subunits disrupt spatial interactions at *Igk*, *Igh*, and *Tcra*, respectively, and perturb V to (D)J recombination (Seitan et al., 2011; de Almeida et al., 2011; Guo, Yoon et al., 2011; Xiang et al., 2011).

In addition to lineage, stage, and allele specificity, it is also likely that the relative use of gene segments is regulated to shape the primary repertoire of V(D)J rearrangements in precursor lymphocyte populations. During subsequent stages of lymphocyte development, V gene segment use is an important component of positive/negative selection and, in some cases, is a primary determinant of functional subsets within a lineage (e.g., TRVB13-2 for iNKT cells) (Godfrey et al., 2000). As such, each species may have evolved toward a unique frequency profile for V use at each AgR locus, balancing requirements for receptor diversity, production of functional subsets, and efficacy of given V segments for antigens expressed by common pathogens. The mechanisms that sculpt preselection V repertoires likely incorporate a combination of the chromatin and spatial features described above. However, their relative contributions to the efficiency of long-range V to (D)J recombination at any AgR locus remain unknown.

We now address this basic question in adaptive immunity, beginning with the molecular determinants that shape Vβ use in preselection thymocytes. The *Tcrb* locus is an attractive starting point for building such models because it contains a manageable set of 35 Vβ segments for molecular analysis; the cis elements controlling recombination also are well defined (Figure 2.1A). New experimental data for chromatin profiles, spatial proximity, and transcription, as well as predictions of RSS quality, were incorporated into a computational analysis that weights each of these features in determining Vβ recombination frequencies. Our data and analyses indicate that

Tcrb adopts a 3D structure in which the relative proximity of each Vβ gene segment to DβJβ clusters is not a significant determinant in its recombination frequency. Instead, each Vβ gene segment has sufficient spatial access to the DβJβ recombination center, and use is fine-tuned by local Vβ chromatin environments, with a particular emphasis on transcription-dependent histone modifications. Indeed, these chromatin features are absent at nonfunctional Vβ gene segments regardless of their RSS quality or precise proximity to DβJβ clusters. This model-building approach should help unravel the primary determinants of preselection V use at other AgR loci and in predicting how natural alterations of large V clusters may impact immune receptor repertoires.

2.3 Materials and Methods

Cell Purification and Antibodies. Thymocytes from C57BL/6 mice (4–6 wk) were depleted of CD4+ and CD8+ cells using magnetic activates cell separation (MACS) (Miltenyi Biotec). The remaining DN cells were stained and sorted for the CD25hi/CD44low DN3 population, yielding a >95% purity. CD19+ bone marrow cells from RAG-deficient mice were purified using MACS in conjunction with CD19 microbeads (Miltenyi Biotec), providing a >90% pure population of pro-B cells. The list of antibodies used is given in Materials and Methods.

High-Throughput Sequencing of Tcrb Rearrangement. gDNA from sorted DN3 cells was amplified by multiplex PCR for Vβ-Dβ-Jβ rearrangements, and the amplicons were deep sequenced by Adaptive Biotechnologies. The gene segment use was analyzed using ImmunoSEQ Analyzer software.

5′ RACE. Total RNA (0.5 μg) from DN3 thymocytes was converted to cDNA, and 5′ RACE was performed using a Cβ primer (5′-AGCTCCACGTGGTCAGGGAAGAA- 3′) following the manufacturer's protocol (Ambion). The RACE product was blunted, concatemerized, and sonicated to an average size of 175 bp. The sheared fragments were ligated with Illumina adapters and sequenced using an Illumina HiSeq-2000 to provide paired-end reads extending 101 bases. Raw reads were de-multiplexed, and unique FASTA reads were obtained using the FASTX tool kit (http://hannonlab.cshl.edu/fastx_toolkit). For quality control, a portion of the 5′ RACE product was cloned, and individual clones were sequenced. Sequences were analyzed using IMGT High-V quest (www.imgt.org) (57)

Quantitative PCR for VβDβJβ Rearrangements. We designed a panel of Taqman PCR assays using probes and primers specific for either Jβ1.1 or Jβ2.1 gene segments in combination with a primer specific for each of the 35 Vβ segments. We also generated a collection of plasmids containing each Vβ cloned directly upstream of either Jβ1.1 or Jβ2.1 in an orientation that mimics the corresponding V-D-J rearrangement product. For this purpose, Jβ1.1 or Jβ2.1 segments were amplified by PCR from mouse gDNA and cloned into the NotI/BamHI sites of pBS-KSII. Subsequently, Vβ segments were amplified and cloned upstream of the Jβ region. The specificity of Vβ primers was confirmed by BLAST searches and a panel of PCR assays showing that amplification of control plasmids containing other Vβ segments was detected at <1% compared with the bona fide target. Template plasmids were used to generate standard curves, allowing us to correct for minor differences in PCR efficiency between each of the assays. Total Vβ-DbJb1.1 or Vβ-DbJb2.1 rearrangement product (alleles) was quantified relative to amounts of an unrearranged region within the genome (b2-microglobulin) using the formula $E^{-Ct(V\text{-}J\beta)}/E^{-Ct(B2M)}$, where E is the primer efficiency. The list of primers and probes used is given in Table T2.

Chromosome Conformation Capture. 3C assays were performed on 107 RAG1- deficient C57BL/6 DN thymocytes or CD19+ pro-B cells using HindIII as described in Hagège et al. (58). Primers and probes designed for HindIII fragments corresponding to each vantage point in the recombination center (Dβ1, Dβ2, and Eβ) were used in Taqman assays with primers specific for each Vβ gene-containing fragment. Standard curves were generated for these Taqman assays using

HindIII-digested bacterial artificial chromosomes (BACs) spanning the entire *Tcrb* locus, which were then ligated to yield a library of all possible products. Interaction between the nearest neighbor fragments in the ERCC3 gene was set as 1. Cross-linking frequencies were calculated as described in Hagège et al. (58). A list of primers, probe sequences, and BAC clones are provided in Table T3.

ChIP and FAIRE. ChIP experiments for H3K4me2, H3K27ac, and P300 were performed with chromatin from RAG-deficient thymocytes (C57BL/6) as described previously (59). The ChIP DNA was purified using a Qiagen DNA purification kit and subjected to whole genome amplification (Sigma), labeled, and hybridized to custom Nimblegen microarrays according to themanufacturer's protocol by Mogene. Total input DNA was used as the hybridization control. A subset of ChIP-Chip data was verified at various locations throughout *Tcrb* using quantitative PCR (qPCR; data not shown). FAIRE was performed on cross-linked nuclei from RAG-deficient DN thymocytes and purified pro-B cells using published methods (Giresi and Lieb, 2009). Purified FAIRE DNA was used for subsequent analyses by qPCRs or array hybridization. DNA from non–cross-linked cells, processed in parallel, was used as reference samples. Model-based analysis of 2-color arrays (MA2C, version 1.4.1) was used to normalize the microarray data, detect peaks, and generate University of California, Santa Cruz (UCSC) wiggle (WIG) files. ChIP-seq experiments were performed as above using chromatin from RAG-deficient thymocytes (C57BL/6) for H3ac, H3K4me3, and CTCF. ChIP-seq data for RNA Pol II, H3K4Me1, and ChIP-Chip data for H3K9me2 from RAG-deficient thymocytes were downloaded from www.comline.fr/ciml/ (Pekowska et al., 2011). The ChIP-seq raw data were aligned to the mouse reference genome

(mm9) using Bowtie 0.12.8. The resulting binary sequence alignment maps (BAM) files were used to generate UCSC wiggle (WIG) files and peaks using model-based analysis of ChIP-seq software (MACS, version 1.4.2). The list of antibodies used in ChIP experiments is given in SI Materials and Methods.

RNA-seq. Total RNA from RAG-deficient DN thymocytes was extracted using an Ambion Ribopure kit. Ribosomal RNA was removed using Ribo-ZERO (EpiCentre). mRNA was fragmented and reverse-transcribed to yield double stranded cDNA, which was sequenced on an Illumina HiSeq-2000 using paired-end reads extending 101 bp. Raw data were de-multiplexed and aligned to the mouse reference genome (mm9) using TopHat 1.4.1. Transcript abundances were estimated from the alignment files using Cufflinks.

Luciferase Assays. The Eβ enhancer was amplified and cloned into the BamHI site of pGL3 (Promega). Each tested upstream Vβ region (300–500 bp) was amplified and cloned into the XhoI/HindIII sites of the Eβ-containing vector. T3 cells (60) were transfected transiently with firefly (4 μg) and Renilla (40 ng) luciferase plasmids using electroporation. After 24 h, the transfected cells were assayed for firefly and Renilla activities. A list of primers is provided in Table T4.

V(D)J Recombination Substrates. A Dβ1-Jβ1.1 rearrangement that includes the 5′ Dβ1-RSS was amplified from thymus DNA and cloned into pCDNA3.1. Each recombination substrate includes the specified Vβ-RSS together with its upstream and downstream flanking sequences (80 and 130

bp, respectively), which were cloned 5′ to the DβJβ1.1 join (deletion substrates). An inert yellow fluorescent protein (YFP) coding sequence was inserted as a stuffer between the Vβ and Dβ1-Jβ1.1 elements. A list of Vβ-specific primers is provided in Table T5.

Recombination Substrate Assays. Human embryonic kidney 293T cells were transfected with an equimolar mixture of eight recombination substrates
(TRBV1, 15, 16, 18, 20, 23, 24, and 26), pEBB-RAG1, and pEBB-RAG2, using Trans-IT 293 (Mirus) (Lee et al., 2003). Plasmid substrates were recovered 48 h post-transfection and digested with NotI to minimize unrearranged PCR products and DpnI to cut untransfected substrates (Lee et al., 2003). The digested DNA mixture was amplified with primers that are common to all substrates—one that recognizes plasmid sequence upstream of the Vβs and one specific for Jβ1.1 (dsT7-CAAGCTGGCTAGCGTTTAAAC and J1.1TR-CTCGAATATGGACACGGAGGACATGC). PCR was performed for 30 cycles on serial fourfold dilutions of recovered substrates. The products were separated on 1% agarose gels, transferred to Zetaprobe (BioRad), and probed with labeled Vβ-specific oligonucleotides.

Computational Analysis. Regression analysis was performed following a two-step procedure that is a simplified version of the protocol described previously (Dong et al., 2012). Step 1. For each of the chromatin features analyzed, the region spanning Vβ segments was divided into three bins: jth Vj segment itself, 1 kb immediately upstream (Uj), and 1 kb immediately downstream of the V segment (Dj). The signal intensity of each bin (3 bins × 35 Vβs, 105 total bins) was measured from the UCSC WIG files containing either read counts (ChIP-seq) or MA2C scores (ChIP-chip) using

BEDtools. The signal intensities were then converted to the natural logarithm of their values. To eliminate any ln(0) values in the computational analyses, a pseudocount of 1 was added to the read counts. Pearson's correlation coefficients were then used to define which of the three bins (Vj, Uj, Dj) correlate best with V recombination frequencies. The bin for each feature with the highest correlation coefficient was used in further analyses. Recombination frequencies fj for Vj regions (expressed in percent of overall use) were transformed into their natural logarithm values [ln(fj + 0.01), where 0.01 is an added pseudocount]. The Vβ gene segments were then classified as rearranging or nonrearranging, and random forest classification was used to determine which of the features distinguish best between rearranging and inert Vβ gene segments (R package; RandomForest). Step 2. Linear regression analysis was performed for 13 variables using data corresponding to only the subset of 23 rearranging Vβ segments (nonzero recombination frequency) using R package (leaps) to identify the most important regressors for recombination levels. The analysis was further refined to determine a reduced set of variables that attains statistical significance (Tables T6–T8 and Dataset S1; online).

Antibodies used. CD4-FITC (561835), CD8-FITC (553031), CD4-biotin (553044), CD8a-biotin (553028), CD44-PE (553134), andCD25-APC(557192) antibodies were purchased from BD Biosciences and used for cell staining and sorting. H3K4me2 (07-030, Millipore), H3K27ac (ab4729), and P300 (C-20) (SC-585×) for H3ac (06-599; Millipore), H3K4me3 (39159; Active Motif), and CTCF (07-729; Millipore) antibodies were purchased and used for ChIP experiments.

2.4 Results

Preselection Tcrb Repertoire.

Recent deep sequencing studies of mRNA corresponding to VβDβJβ combinations expressed in peripheral CD4+ T lymphocytes have provided an approximation of the post-selection *Tcrb* repertoire (Ndifon et al., 2012). However, our goal is to understand variables that impact the efficiency of long-range Vβ to DβJβ recombination, which shapes the preselection *Tcrb* repertoire. Accordingly, these analyses must be performed on primary thymocytes before their positive or negative selection, which may alter the Vβ repertoire. Preferably, a DNA-based assay should be used to quantify Vβ use because mRNA expression of VβDβJβ rearrangements may be influenced by promoter strength or message stability. We developed the requisite assay (see below), which was applied to genomic DNA (gDNA) from sorted double negative (DN3) cells (>95% purity; CD4−, CD8−, CD25high, CD44low), a developmental stage in which Vβ to DβJβ recombination occurs at a high frequency, but the vast majority of cells have yet to undergo *Tcrb*-dependent selection (Cobb et al., 2006). We reasoned that the relative frequency of rearrangements in this cell population involving a particular Vβ segment, regardless of whether the joins are productive or out of frame, accurately reflects its recombination potential.

Initially, we deep sequenced products of a multiplex PCR amplification that incorporates primers for each mouse Vβ and Jβ gene segment, analogous to an approach described previously for analysis of human *Tcrb* repertoires (Robins et al., 2009). However, when applied to our DN3 thymocyte samples, a small subset of the mouse Vβ primers exhibit amplification biases in the multiplexing platform, limiting their usefulness for establishing relative Vβ frequencies. In contrast, this approach yields a relative Jβ use similar to that observed in prior studies, suggesting

no significant bias in the Jβ primers (Figure 2.1A) (Ndifon et al., 2012). In keeping with this, we noticed that the collection of VβDβJβ rearrangements for each Jβ segment has a nearly identical Vβ distribution. For example, TRBV16 is used in 8.6% of all rearrangements involving Dβ1Jβ1.1. A nearly identical percentage of Dβ1Jβ1.2 rearrangements, or any other Dβ-Jβ combination, use the TRBV16 gene segment (7.5–8.6%). The Jβ-independent frequency of Vβ use held true for all Vβ gene segments (Figure 2.2B; Figure 2.1B). Moreover, recent studies have reported similar Vβ use for rearrangements involving either Dβ1 or Dβ2 (Ndifon et al., 2012). Thus, an accurate depiction of Vβ use can be established from a simplified approach in which levels of Vβ rearrangements to a single Jβ gene segment are measured quantitatively.

Accordingly, we designed Taqman PCR assays to independently measure rearrangements between Jβ1.1 and each of the 35 Vβ gene segments that undergo V to DJ recombination (Figure 2.2A). We also prepared control plasmids containing each of the Vβ-Jβ1.1 combinations to serve as templates for standard curves. Initial experiments verified that all Vβ-Jβ1.1 plasmids amplified with comparable efficiencies (±5%) using Vβ-specific primers with a Jβ1.1 primer/probe combination. Control PCR assays revealed no significant cross-reactivity of Vβ-specific primers with off-target Vβ segments. Standard curves were used to quantify levels of each Vβ-Dβ1Jβ1.1 recombination product in gDNA from sorted DN3 thymocytes. The relative frequencies of Vβ use were consistent in three biological replicates and averaged values are shown in Figure 2.2C. Similar Vβ frequencies were observed in assays measuring a subset of Vβ-Dβ2Jβ2.1 rearrangements (Figure 2.2D), confirming the Dβ and Jβ independence of Vβ use. Consistent with previous observations, analysis of gDNA from DN-depleted thymocytes revealed only a few modest differences in Vβ use, indicating that the pre- and post-selection Vβ repertoires in mouse

thymocytes are largely comparable (Figure 2.1C) (Wilson et al., 2001). In contrast, deep sequencing of the 5′-RACE library from two DN3 samples yielded a distribution that differed at a subset of Vβ segments compared with our quantitative gDNAbased assay (Figure 2.2E). These findings suggest that mRNA levels corresponding to rearrangements involving some Vβ gene segments may not accurately reflect their recombination frequency in preselection thymocytes.

Overall, we observe a >10-fold range in relative Vβ use. Only TRBV13-2 (formerly Vβ8.2) and TRBV19 (formerly Vβ6) are significantly overrepresented in the primary repertoire of *Tcrb* rearrangements. The preponderance of TRBV13-2 is consistent with analyses using a restricted set of Vβ-specific antibodies from T-cell populations (Wilson et al., 2001). In contrast, rearrangements were undetectable for 11 of the 35 Vβ segments. Five of these 11 "inert" gene segments are predicted to have nonfunctional RSSs (Figure 2.2C, asterisks), crippling their recognition by the RAG-1/2 recombinase. Six of the remaining inert gene segments have functional RSSs, but are pseudogene segments due to disruptions in their coding potentials (ψ; Figure 2.2C). A lack of VβDβJβ rearrangements involving these six pseudogene segments flanked by functional RSSs indicates that other factors influence their recombination efficiencies (see below). Only two functional Vβs, TRBV15 and TRBV30, were underused compared with the remaining 22 functional segments, which displayed only a modest variability in their use (approximately threefold range). These repertoire data suggest that *Tcrb* has evolved to normalize use of nearly all functional Vβ segments, perhaps by modulating the three determinants of long-range recombination efficiency: RSS quality, spatial proximity, and chromatin environment.

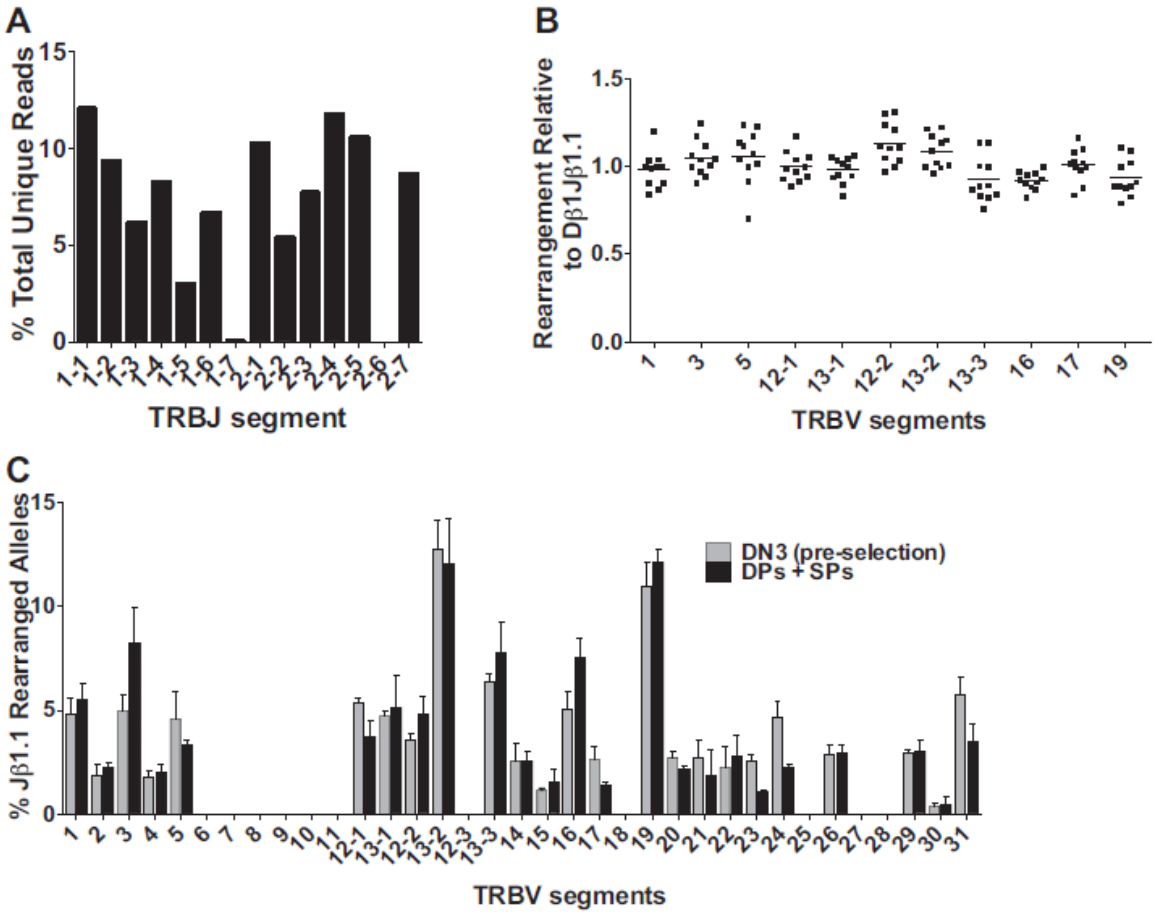


Figure 2.1: Variable (V)β repertoire comparisons. (A) Joining (J)β use profile from high-throughput sequencing [mean (n = 3),15,000–20,000 unique reads per sample]. (B) Distribution of rearrangements from high-throughput sequencing involving Vβ segments and each of the 11 functional Jβ segments. Shown are distributions for rearrangements of Vβ segments yielding at least 1,000 unique reads. Data are represented relative to the distribution of Vβ-Jβ1.1, where percent total Vβ-Jβ1.1 is set to a value of 1 (Fig. 1). Each circle represents a data point for a given Jβ segment. (C) Comparison of Vβ use in preselection and post-selection thymocytes measured by the genomic DNA (gDNA) assay described in Fig. 1C (mean ± SEM, n = 3).

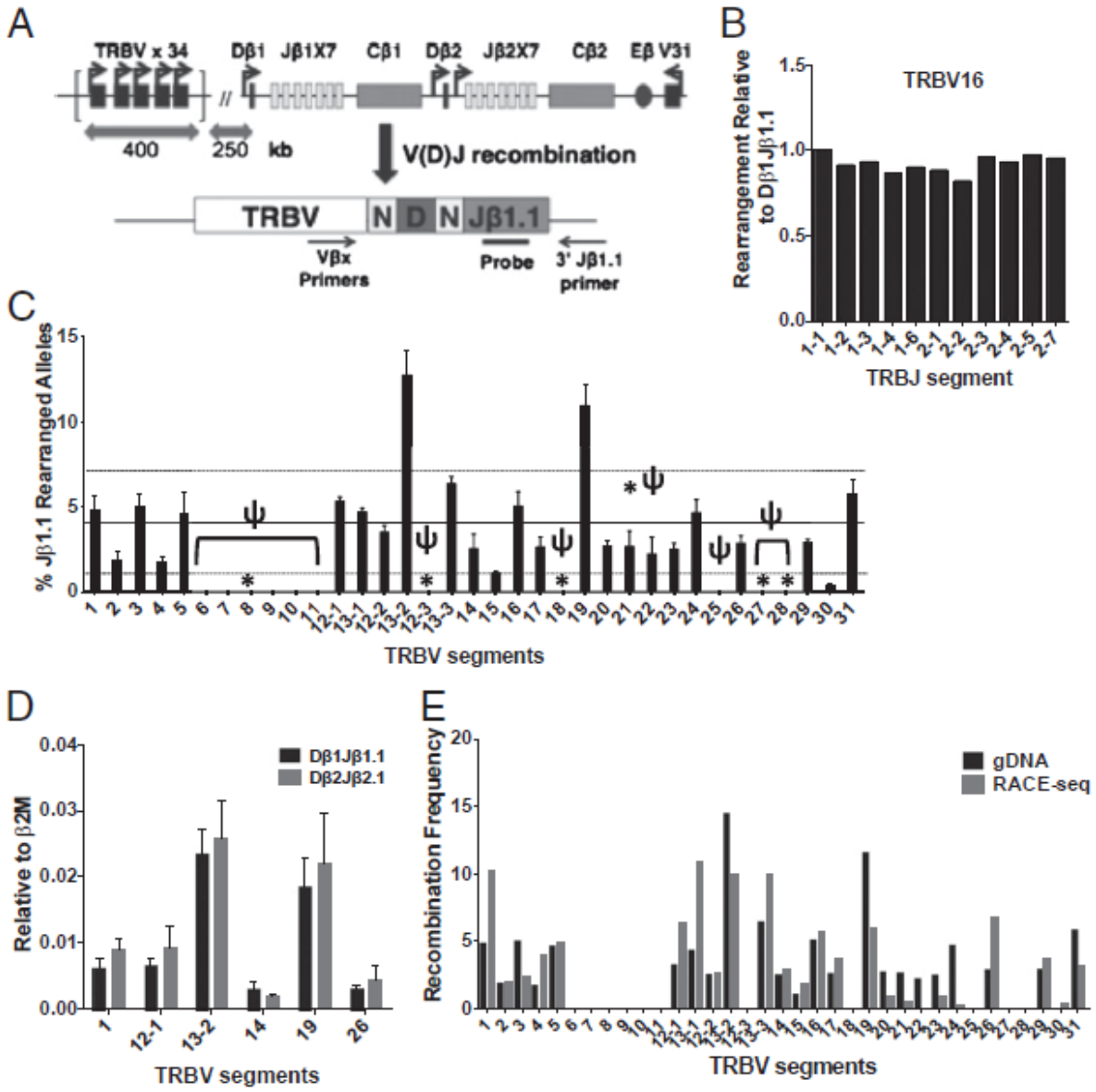


Figure 2.2: Preselection *Tcrb* V repertoire. (A) Schematic representation of the murine Tcrb locus (Upper) and Taqman assay (Lower) used to quantify VβDβ1Jβ1.1 recombination products. Bold arrows near gene segments denote promoters (Upper). N, N-regions (nontemplated regions of diversification); locations of primers and probes for Taqman assays are shown (Lower). (B) Distribution of V(D)J rearrangements from high-throughput sequencing involving select Vβ segments and each of the 11 functional Jβ segments. The distribution for a given Vβ-Jβ combination is calculated as the number of unique reads for that combination divided by the total number of unique reads for the corresponding Jβ element. Data are represented relative to the distribution of Vβ-Jβ1.1, where percent total Vβ-Jβ1.1 is set to a value of 1. (C) Preselection Vβ repertoire. Taqman real-time PCR quantification of VβDβ1Jβ1.1 rearrangements was performed on gDNA from DN3 thymocytes. Signals from each assay were normalized to values obtained from an assay for the invariant β2M gene. Average levels from three independent DN3 preparations are shown (n = 3, ± SEM). Recombination frequencies are shown as the percent contribution of a given Vβ segment to the total level of Jβ1.1 rearrangement. Pseudogenes are denoted by ψ and gene segments with nonfunctional RSSs are marked with an asterisk. The average Vβ use and SD are denoted by dotted black lines. (D)

Taqman real-time PCR assays measuring VβDβ1Jβ1.1 vs. VβDβ2Jβ2.1 rearrangements in DN3 thymocytes were quantified as described in C. (E) Comparison of Vβ use values in DN3 thymocytes using gDNA- vs. mRNA-based methods. Average values from gDNA assay (n = 3) and RNA-5′ RACE seq (n = 2) are shown.

Spatial Access of Vβ Gene Segments to the DβJβ Recombination Center.

Long-range recombination of V gene segments at all *Ig* and *Tcr* loci is facilitated by a contraction process, which places the V cluster into spatial proximity with distal (D)J targets located 0.1–3.2 Mb away in the linear genome (Bossen et al., 2012; Kosak et al., 2002). Deletion of transcription factors or cis elements that disrupt locus contraction significantly impair V to (D)J recombination, supporting a functional link between these processes (Guo, Gerasimova et al., 2011; Fuxa et al., 2004; Reynaud et al., 2008; Liu et al., 2008). Additional evidence indicates that V clusters fold into a compact rosette-like structure, which may permit extensive interactions between a recombination center and many or all of its upstream V segments (Jhunjhunwala et al., 2008). Alternatively, the spatial architecture of V clusters may sculpt the repertoire by positioning a subset of V segments closer to their (D)J targets (efficient rearrangement) while spatially excluding others (inefficient rearrangement). Indeed, emerging studies at *Igk* suggest that Vκ pseudogene segments may be spatially excluded from interactions with Jκ substrates, perhaps minimizing their recombination potential (Lin et al., 2012).

To test whether spatial proximity is a key determinant in shaping the preselection *Tcrb* repertoire, we measured interaction frequencies between restriction fragments spanning each Vβ segment and fragments spanning either of the two DβJβ clusters using chromosome conformation capture (3C) (Dekker, 2008). In the linear genome, the distance between these restriction fragments range from 250 to 700 kb (except for TRBV31, which is ~3 kb downstream of Eβ and rearranges by inversion). 3C assays were performed on cross-linked chromatin from RAG1-deficient thymocytes, a predominantly DN3 cell population in which *Tcrb* is in an active germ-

line conformation. The use of RAG-deficient thymocytes circumvents complications in data analysis that arise from active *Tcrb* rearrangement. Although we cannot rule out a role for RAG-1 in defining the precise 3D conformation of *Tcrb* (Chaumeil et al., 2013), prior studies demonstrate that RAG proteins are dispensable for locus contraction (Skok et al., 2007).

We measured the cross-linking efficiency of each Vβ-containing HindIII fragment to three downstream vantage points within the *Tcrb* recombination center. Specifically, we probed Vβ cross-linking to HindIII fragments containing either of its two substrates (Dβ1 or Dβ2), or the transcriptional enhancer Eβ, which generates active chromatin over the DβJβ clusters (Oestreich et al., 2006; Spicuglia et al., 2000). Regardless of the vantage point, nearly all Vβ gene segments interact more frequently with the DβJβ recombination center in DN thymocytes compared with CD19+ pro-B cells purified from RAG-deficient bone marrow (Figure 2.3A; Figure 2.4A and B). These data verify and extend previous analyses showing that *Tcrb* adopts a T cell–specific conformation, juxtaposing the Vβ cluster with its DβJβ targets (Skok et al., 2007).

Of particular note, interaction levels measured from a given vantage point (e.g., Dβ1) display significant differences across the collection of Vβ segments (Figure 2A). There were also differences in interactions between specific Vβ segments and two vantage points. For example, the fragments spanning TRBV1 or TRBV18/19 both interact with Dβ1 at a much higher frequency than with Dβ2 (Figure 2.3A; Figure 2.4A). Conversely, TRBV17 displays a greater interaction with Dβ2 (Figure 2.3A; Figure 2.4A). Despite these differences, the TRBV1 and TRBV19 segments are used with indistinguishable frequencies in recombination products involving either Dβ1 or Dβ2 (Figure 2.2D). In contrast to preliminary findings at *Igk* (Lin et al., 2012), a group of pseudogene segments spanning TRBV6–TRBV11 each interact with DβJβ clusters at a relatively

high frequency, but these gene segments are absent from the preselection *Tcrb* repertoire despite having functional RSSs. These findings suggest that relative Vβ use in the preselection *Tcrb* repertoire cannot be fully explained by differences in their spatial proximity to the DβJβ regions.

To more rigorously investigate the relationship between spatial proximity and long-range recombination, we performed Spearman ranking correlations for 3C and Vβ repertoire data. Because the absolute values of 3C data cannot be quantitatively compared between the three assays, we first ranked cross-linking efficiencies of the Vβ segments within each vantage point (Table T1). No significant correlations between 3C ranking and TRBV rearrangement are observed for any of the three individual viewpoints within the DβJβ recombination center. We also calculated the average ranking for each Vβ segment over the three assays (Dβ1, Dβ2, and Eβ) and compared these values with relative use in VβDβJβ joins (Table T1). As shown in Figure 2.3B, there is an absence of significant correlation between Vβ use and its average rank for interactions with the DβJβ recombination center. Consistent with this finding, we also observe no obvious correlation between the recombination frequency of a Vβ segment and its proximity to CTCF binding. We conclude that, although gross locus contraction is important to bring the entire Vβ cluster into spatial proximity with its Dβ substrates, the precise magnitude of each Vβ–Dβ interaction is not a primary determinant of recombination efficiency. Instead, our 3C and repertoire data indicate that once *Tcrb* is contracted in DN thymocytes, the large Vβ cluster adopts a conformation in which spatial access of Vβ segments to the recombination center is not limiting.

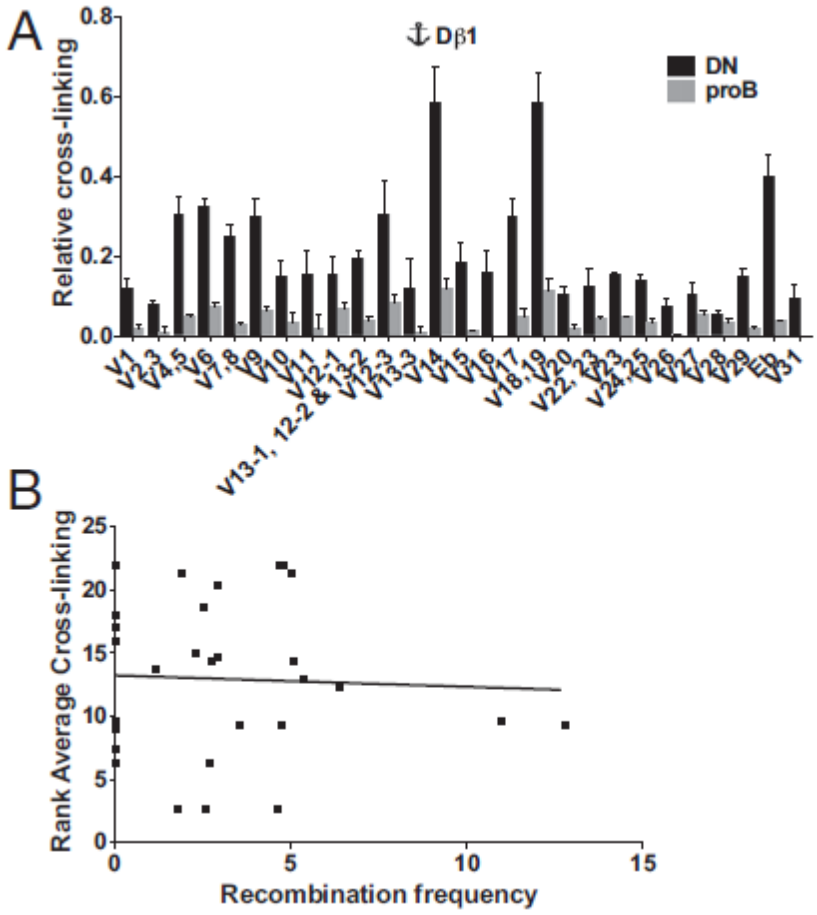


Figure 2.3: Role of Vβ spatial proximity in shaping the *Tcrb* repertoire. (A) 3C analysis of RAG-deficient thymocytes showing relative cross-linking frequencies between a Dβ1 anchor and HindIII fragments spanning Vβ gene segments. Data are presented as mean ± SEM (n = 3). (B) Spearman correlation of Vβ use and average ranked values for 3C cross-linking frequency from three viewpoints within the recombination center (Dβ1, Dβ2, and Eβ). The Spearman correlation coefficient shows no significance (rs = 0.035, P = 0.85).

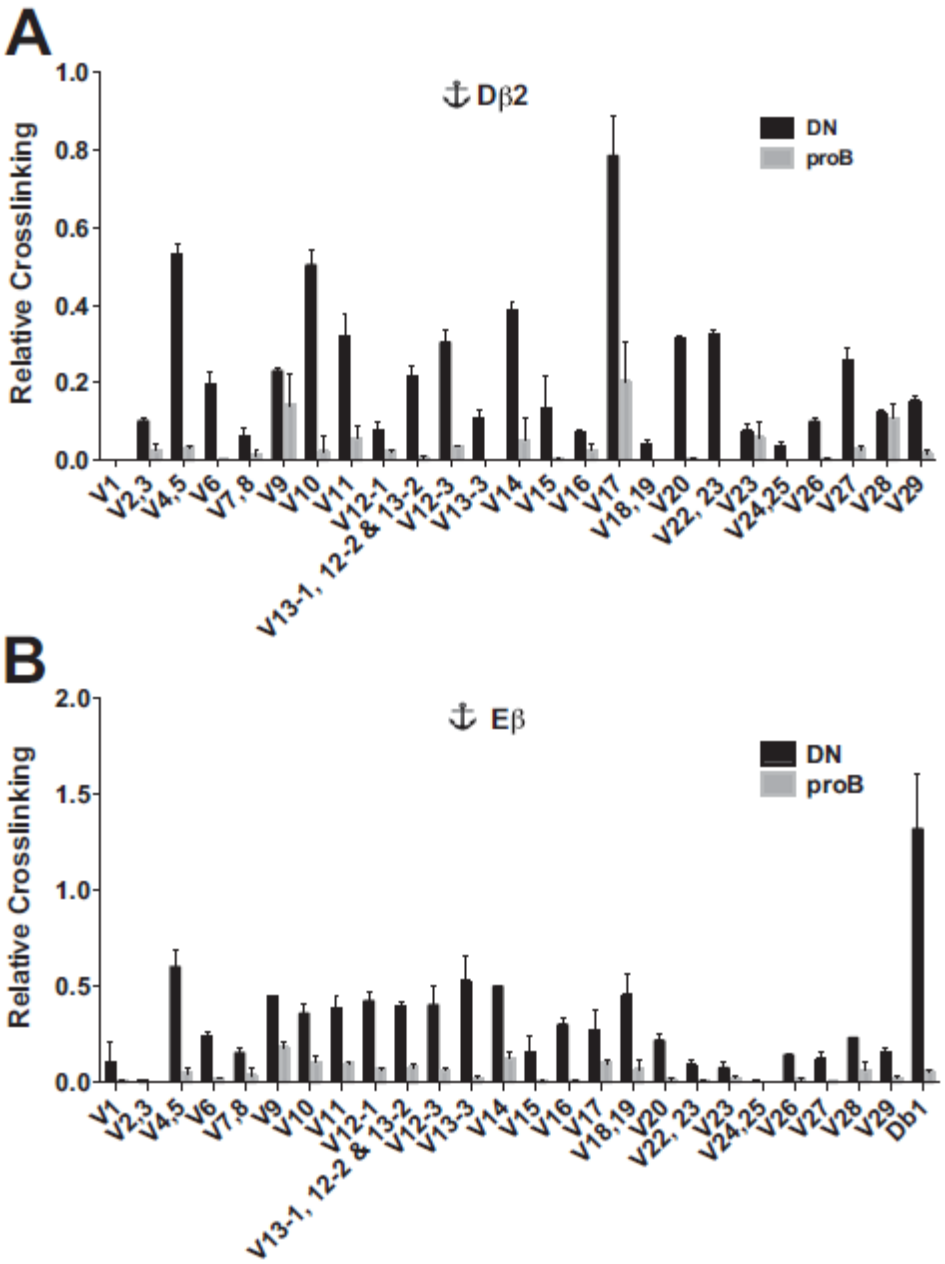


Figure 2.4: Role of spatial proximity in shaping the *Tcrb* repertoire. (A) Chromosome conformation capture (3C) analysis of Rag-deficient double negative (DN) thymocytes showing relative cross-linking frequencies between a diversity (D)β2 anchor and HindIII fragments spanning Vβ gene segments. Data are presented as mean ± SEM (n = 3). (B) 3C analysis using an Eβ anchor.

Role of RSS Quality in Determining Vβ Use.

Despite general conservation of the heptamer-spacer-nonamer configuration, RAG- 1/2 substrates exhibit substantial variation compared with the consensus RSS sequence: (CACAGTG)–12- or 23-bp spacer– (ACAAAAACC) (Hesse et al., 1989; Livák, 2003). In vivo replacement or natural variants of RSSs can alter the use of gene segments, including those within the *Tcrb* recombination center (Posnett et al., 1994; Wu et al., 2003; Nadel et al., 1998). In vitro studies using plasmid substrates have defined the effects of positional substitutions within the consensus RSS on recombination efficiency (Hesse et al., 1989; Feeney et al., 2000; Jung et al., 2003). Thus, one component of nonrandom Vβ use is likely the quality of its flanking RSS.

To examine this possibility, we took advantage of an algorithm (www.itb.cnr.it/rss/) that predicts the RSS quality of any given sequence (Cowell et al., 2003). In brief, this algorithm calculates the theoretical recombination potential of an RSS using a statistical model that assigns a score based on the contribution of each nucleotide within the heptamer-spacer-nonamer sequence. The algorithm output is a recombination signal information content (RIC) score, which predicts the quality of an input RSS with a reasonable degree of accuracy based on data from plasmid recombination substrates (Cuomo et al., 1996). For *Tcrb*, 6 of the 35 Vβ gene segments are flanked by nonfunctional RSSs with a RIC score of <−58.5, the threshold defined by Cowell et al., (TRBV8, 12-3, 18, 21, 27, and 28). The remaining 29 Vβ segments have a substantial range in predicted RSS quality, with RIC scores between −29 (TRBV4) and −58.2 (TRBV11). Recombination is undetectable for five of the six Vβ segments flanked by RSSs that score below the functional threshold (Figure 2.2C). The exception is TRBV21, which rearranges at a detectable

level, but is predicted to have a marginally nonfunctional RSS (RIC score, −58.6) consisting of a consensus heptamer and a 22-bp rather than 23- bp spacer.

The correlation between RIC scores and Vβ use is shown in Figure 2.5. Although a positive correlation is apparent, the magnitude of Vβ use diverges significantly from linearity compared with predicted RSS quality. In general, Vβ RSSs with lower quality (RIC scores, −45 to −58) are either inert or rearrange at a level below the average frequency. RSSs with RIC scores >−45 exhibit a broad range of Vβ recombination frequencies, as highlighted by the following examples: (i) TRBV13-2 is the most frequently used segment but shares a nearly identical RIC score with TRBV14, which rearranges at an average frequency; and (ii) six Vβ segments (TRBV7, 15, 16, 20, 24, and 26) have nearly indistinguishable RIC scores (−41 to −42), but one Vβ is recombinationally inert (TRBV7) and the remaining five display an eightfold range in their utilization. We cannot rule out the possible contribution of coding sequences adjacent to each RSS in altering its quality as a RAG-1/2 substrate. Inspection of coding flanks revealed only a small subset with features predicted to attenuate RAG cleavage (e.g., AT or pyrimidine stretches for TRBV12-1, 12–2, 14, 17, and 29) (Cuomo et al., 1996; Gerstein and Lieber, 1993; Olaru et al., 2003; Yu and Lieber, 1999). However, as shown below, the recombination frequency of these gene segments correlate best with features of associated chromatin. Together, our data indicate that, although predicted RSS qualities contribute to the formation of a preselection *Tcrb* repertoire, other levels of control clearly impact Vβ use.

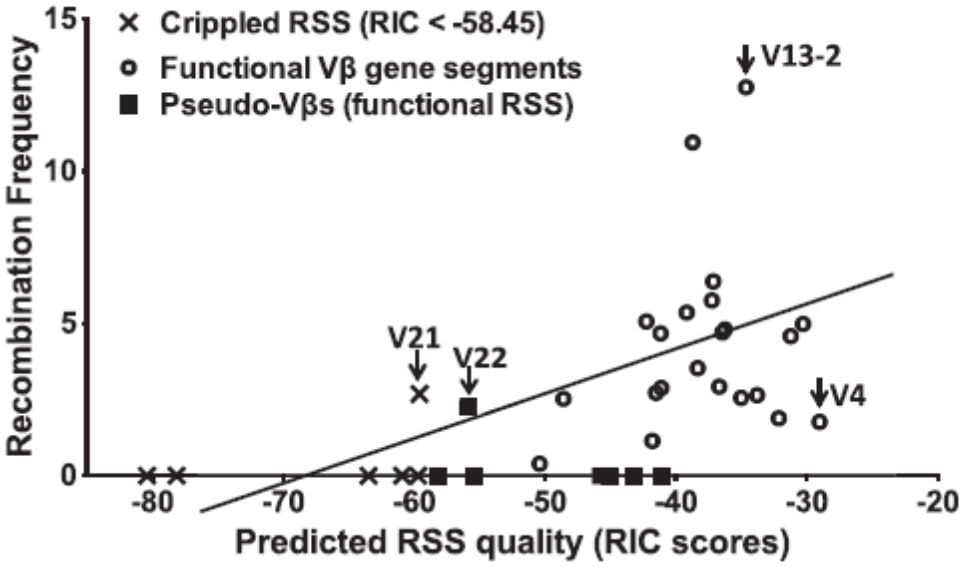


Figure 2.5: Correlation between Vβ utilization and predicted RSS quality. The correlation between predicted Vβ RIC23 scores and observed Vβ recombination frequencies (Fig. 1B), yielding a Spearman's rank correlation coefficient rs = 0.6456, P < 0.0001.

Role of Chromatin Environment in Determining Vβ Recombination Potential.

Chromatin accessibility at gene segments has been studied extensively as a determinant of the tissue- and stage-specific mechanisms controlling V(D)J recombination (Cobb et al., 2006; Feeney, 2009). Germ-line transcription of gene segments leads to the deposition of H3K4me3, a histone modification that is recognized by RAG2 and augments endonuclease function of the RAG complex (Liu et al., 2007; Matthews et al., 2007; Shimazaki et al., 2009). As such, levels of chromatin accessibility and transcription at each Vβ segment may help determine its use in the preselection *Tcrb* repertoire.

The emerging approach of "chromatin profiling" uses combinatorial patterns of histone modifications, nucleosome density, and factor binding to assess the epigenetic status of genomic regions (Ernst et al., 2011). To compare epigenetic landscapes at the 35 Vβ segments, we generated chromatin profiling data from RAG1- deficient thymocytes using ChIP assays in combination with *Tcrb* microarrays (ChIP-chip) or deep sequencing. We also performed formaldehyde-assisted isolation of regulatory elements (FAIRE), which identifies nucleosome-depleted regions in the genome (Giresi and Lieb, 2009). The new ChIP-chip (P300, H3K27ac, H3K4me2), ChIP-seq (H3ac, H3K4me3, and CTCF), and FAIRE-Chip data from RAG-deficient thymocytes were combined with epigenomic data available in public repositories (H3K4me1, RNA Pol II, and H3K9me2) from RAG-deficient thymocytes (Pekowska et al., 2011). We used a published methodology to integrate cross-platform data derived from ChIP-chip and Chip-seq (Chen et al., 2011). In addition to nucleosome depletion (FAIRE), the analyzed features characterize active promoter regions (transcription, RNA Pol II, H3K4me3, and H3ac), active regulatory elements

(H3K4me1, H3K27ac, and P300), poised chromatin (H3K4me2), insulators (CTCF), and silent chromatin (H3K9me2).

Relative intensities for each feature at the 35 Vβ segments (±1 kb) are represented as a heat map in Figure 2.6A. Examples of several features for selected gene segments in chromatin environments ranging from highly active to silent are depicted in Figure 2.6B. Overall, most of the Vβ segments that participate in Vβ to DβJβ recombination exhibit higher levels of active chromatin features than the inert Vβ elements (H3K4me, RNA Pol II/ transcription, and histone acetylation). In contrast, the repressive H3K9me2 modification was enriched over many of the inert Vβ segments. One region within the Vβ cluster containing the TRBV12-2 and 13-2 gene segments is conspicuously active (Figure 2.6B), with high levels of germ-line transcripts and other features associated with open chromatin, including one of the few discernible P300 peaks. As noted above, TRBV13-2 is also the most frequently rearranged gene segment in DN3 thymocytes, suggesting a dominant correlation between open chromatin and long-range recombination efficiency. Consistent with this possibility, many of the pseudogene segments, even those containing functional RSSs, are expressed at a low level and are associated with chromatin that lacks activating histone marks (Figure 2.6A, asterisks). In silico analysis of Vβ upstream sequences (−1 kb to leader) for predicted transcription factor binding profiles (TRASFAC/ JASPAR databases) revealed no distinguishable differences between functional and pseudo-Vβ gene segments. Promoter activity as measured by luciferase assays in a transfected pre–T-cell line show that all tested upstream Vβ regions from recombinationally active gene segments (11/11) are functional promoters. In contrast, only some of the tested regions upstream of pseudogene segments (4/8) exhibit promoter activity (ψ; Figure 2.7), indicating no clear correlation between Vβ utilization

and promoter strength. Thus, it appears that the mouse Vβ cluster has evolved multiple strategies to silence chromatin at nonfunctional gene segments.

A reasonable concordance was observed between chromatin environments and recombination efficiencies when comparing Vβ segments with equivalent RIC scores. For example, TRBV15 and TRBV16 are predicted to have RSSs of nearly identical qualities but reside in distinct chromatin environments. The elevated levels of transcription and activating histone marks at TRBV16 correspond to an elevated level of recombination (Figure 2.6C). In some cases, both the predicted RSS quality and chromatin environment apparently contribute to Vβ use. For example, TRBV23 and TRBV24 are both transcriptionally active and have comparable chromatin features (see heatmap in Figure 2.6A); however, the lower predicted RSS quality for TRBV23 (−48.6) compared with TRBV24 (−41.2) correlates with an attenuated level of recombination. We also noted that contributions of chromatin to rearrangement frequencies may derive from different combinations of features. TRBV20 and TRBV26 exhibit nearly identical use (2.7% and 2.9%) and RIC scores (−41.5 and −41.1), but patterns of specific chromatin features at these gene segments differ significantly (see heat map in Figure 2.6A). To further validate these comparisons, we performed semi-quantitative assays to measure the qualities of eight Vβ-RSSs using plasmid-based substrates (including the six Vβ-RSSs mentioned above). The relative qualities of these RSSs, tested in conjunction with a natural target (5′Dβ1-RSS), are in line with predictions from RIC scores (Figure 2.6D), further supporting our conclusions. Together, these profiling studies indicate a strong contribution of chromatin environment to Vβ recombination frequencies but also suggest that individual parameters of chromatin accessibility may affect substrate use in a weighted manner.

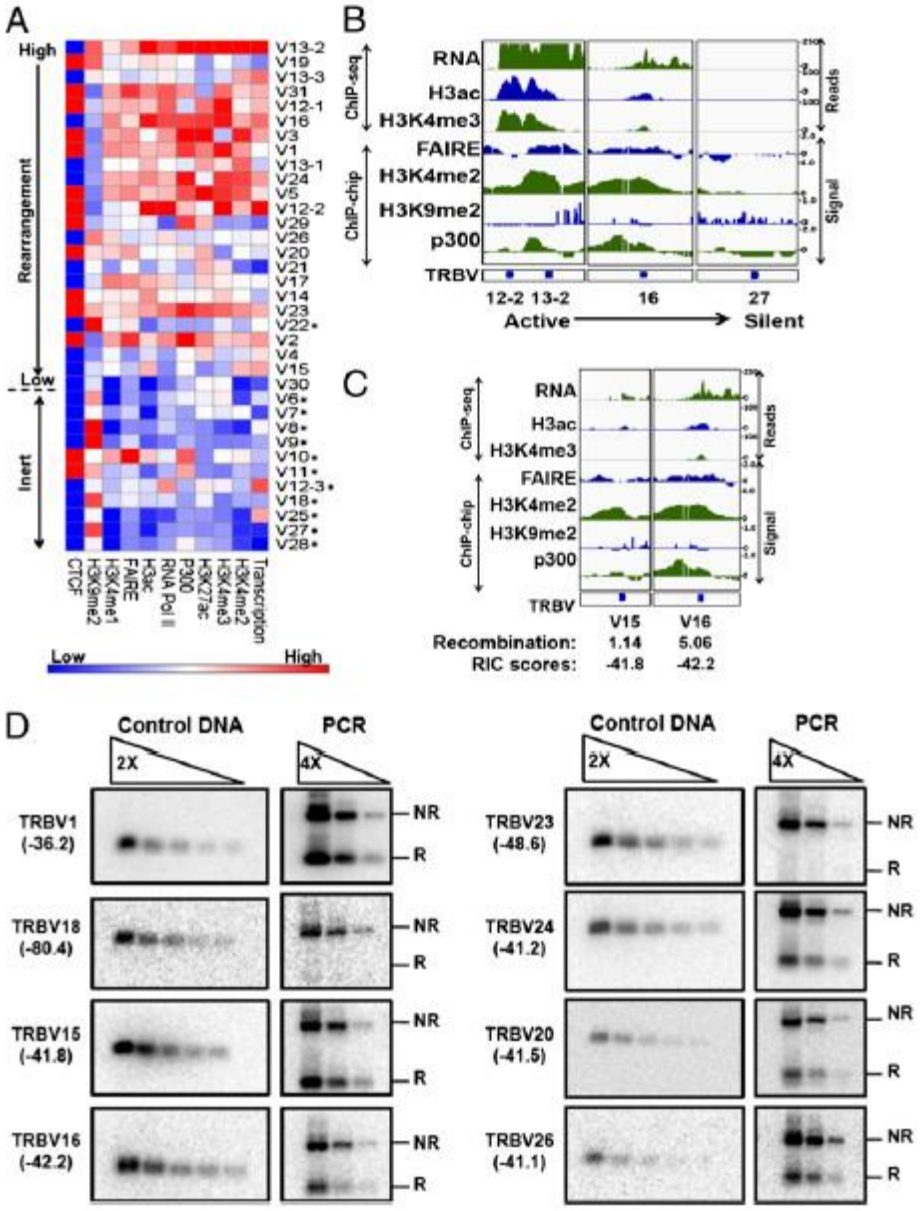


Figure 2.6: Role of chromatin landscape in Vβ use. (A) Relative intensities of various chromatin features (transcription, RNA Pol II, P300, histone modification signals, and proximal CTCF sites) at the 35 Vβ segments are represented as a heatmap. The log2 values of ChIP-Seq or ChIP-Chip signal intensities at the Vβ segment (±1 kb) for each of the above features were quantified using BEDtools, and the relative intensity for each feature was plotted as a heatmap. CTCF intensities are represented as binary values of 1 or 0 assigned for presence or absence of CTCF within 1 kb of the Vβ segment. Asterisks denote pseudo-V gene segments. (B) Profiles for transcription (RNA), nucleosome depletion (FAIRE), P300, and indicated histone modifications are shown at select Vβ segments. RNA-seq data for transcription, ChIP-seq data for H3ac and H3K4me3, and ChIP-chip data (signal = log2 ratio of ChIP DNA/input DNA) for H3K4me2, P300, and FAIRE are displayed. See Materials and Methods for sources of epigenomic data. (C) Epigenetic profiles at Vβ segments highlighting the influence of chromatin landscapes on gene segment use. (D) An equimolar mixture of the eight indicated Vβ 23-RSS deletion substrates was assayed for rearrangement in conjunction with the 5′Dβ1 12-RSS following transfection into 293T cells with RAG-1/2 expression vectors (40).

Rearrangements were detected by PCR using primers shared by all of the substrates (NR, not rearranged; R, Vβ rearranged to Dβ1). RIC scores for each TRBV-RSS are shown in parentheses. Rearrangements for each substrate were detected using probes specific to the given Vβ segment. A semi-quantitative measure of rearrangement efficiencies was obtained by comparing twofold dilutions of Vβ plasmid inserts (3 ng–500 ng, Left) with fourfold dilutions of the PCR product (Right). Shown are data from one representative PCR amplification of four independent transfections. Control DNA and PCR products for each Vβ substrate are on the same blot. The TRBV15, 16, 20, 24, and 26 RSSs exhibit similar recombination efficiencies based on this semi-quantitative assay (RIC scores all approximately −42), whereas the TRBV18 and 23 RSSs exhibit minimal rearrangement (lower RIC scores) and TRVB1 rearranges most efficiently (best RIC score).

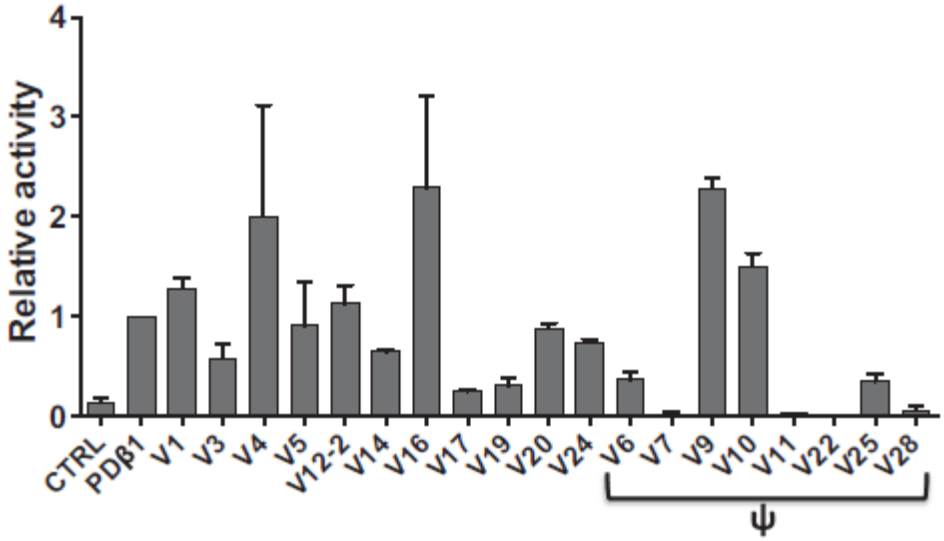


Figure 2.7: Luciferase assays. Promoter activity assay for upstream regions of Vβ regions. Relative promoter strengths for select upstream Vβ regions were assessed using luciferase reporter constructs. pGL3 constructs containing the Eβ enhancer and the respective Vβ promoter regions were transfected into a pre–Tcell line (T3). Promoter activities were assessed at 24 h post-transfection and normalized to Renilla (RLU). Data are represented as averages ± SEM (n = 3) relative to the control PDβ1 promoter. ψ, pseudo-Vβ gene segments; CTRL, promoterless construct.

Computational Analysis of Vβ Use Determinants.

Our data indicate that predicted RSS qualities and chromatin landscapes likely contribute in a combinatorial manner to the efficiency of long-range *Tcrb* assembly. To examine these combinatorial relationships, we used classification and regression analyses comparing chromatin features and predicted RSS quality with Vβ use. These analyses were guided by recent computational strategies devised to predict gene expression levels based on patterns of histone modifications (Dong et al., 2012; Karlić et al., 2010). We applied one validated approach (Dong et al., 2012) to study whether chromatin features, predicted RSS quality, and spatial proximity are predictive of the observed Vβ repertoire.

The chosen computational approach takes into account (i) the signal intensity of each chromatin feature, (ii) levels of germ-line transcription, (iii) RIC scores, and (iv) spatial proximity based on the average 3C rank score. With regard to chromatin features, distinct positional profiles are observed for various histone marks. For example, H3K4me3 is enriched over active promoters and progressively wanes along gene bodies. Accordingly, we divided the regions spanning each Vβ segment into three bins: the Vβ segment itself (leader to RSS), its upstream promoter region (1 kb 5′ of leader), and its downstream region (1 kb 3′, including the RSS). For each feature, we computed Pearson correlation coefficients for the three bins vs. Vβ recombination frequencies (Figure 2.8). We find the best correlation for a majority of histone modifications in the upstream/promoter bin (H3K4me1, H3K4me2, H3K4me3, P300, H3ac, and H3K27ac). In contrast, repression by H3K9me2 was most correlative in the bin that contains Vβ segments. FAIRE and RNA Pol II signals have very similar predictive abilities over both the Vβ and its downstream bins. These findings are strikingly similar to correlations observed between chromatin

features and gene expression (Dong et al., 2012; Karlić et al., 2010), further underscoring the relationship between transcriptional activity and Vβ recombination frequencies. A particularly satisfying outcome of this analysis is the correlation between FAIRE signals and the bins flanking RSSs, presumably reflecting a requirement for nucleosome depletion at RAG-1/2 targets (Kwon et al., 2000; Osipovich et al., 2007).

Next, we identified features that are most predictive of whether a Vβ segment will rearrange at any frequency or will remain inert. For this and the remaining analyses, we used signal intensities only from bins exhibiting the highest correlation between each chromatin mark and Vβ use (Figure 2.8, asterisks). A computational approach called random forest was used (Dong et al., 2012), which randomly tests combinations of binned features for their predictive abilities to classify gene segments as active or inert (Figure 2.9A). This analysis revealed that three features—predicted RSS quality, FAIRE, and RNA Pol II signals—are sufficient to classify the recombination potential of a given Vβ segment with a high level of confidence. The classifications are also evident from linear regression analysis on these three features relative to Vβ recombination frequencies (Figure 2.9B; 30/35 segments predicted correctly). When we used the random forest algorithm, but focused only on values for RIC score, FAIRE, and RNA Pol II signals, 32/35 Vβ segments classified correctly as active vs. inert (Materials and Methods). The three exceptions common to both random forest– and linear regression–based classifications are TRBV15, 21, and 22; segments predicted to be inert but exhibiting detectable levels of recombination. These outliers could reflect partial compensation by chromatin features other than the factors determined by our algorithms. Notwithstanding, the most important predictive features of recombinational

competency are linked mechanistically to RAG substrate quality (RIC score), substrate accessibility (nucleosome depletion), and RNA Pol II association.

We next moved beyond black and white classifications to analyze the relative importance of Vβ features in fine-tuning recombination frequencies of the 23 active gene segments. For this purpose, we performed linear regression on the selected bins for each feature vs. frequency values. As shown in Figure 2.9C, the features that correlate most significantly with Vβ use are H3K4 methylation, H3Ac, and RNA Pol II occupancy, which normally associate with transcriptionally active regions. The repressive H3K9me2 mark correlates negatively with levels of Vβ recombination. In contrast to its dominant role as a determinant for recombinational competence, RIC scores for the 23 active Vβ gene segments correlate poorly with their relative levels of rearrangement. A similar discordance between recombination frequencies and RSS qualities for a limited set of mouse VH and Vκ gene segments has been described previously (Williams et al., 2001; Aoki-Ota et al., 2012). These findings suggest that chromatin environment, rather than predicted RSS quality, is the dominant feature for fine-tuning Vβ use in long-range recombination.

We next investigated whether various combinations of the 13 features included in this study are predictive of Vβ recombination efficiencies. As a starting point, we examined the predictive capacity of all 13 features using linear regression (Figure 2.10A). This analysis yielded a correlation coefficient for best fit of 0.78, which was statistically insignificant ($P > 0.05$). We next tested whether a subset of these 13 features correlate in a significant manner with observed frequencies of Vβ use. For this purpose, we examined various subsets of features, ranging from a single feature to 12 of the 13 variables in all possible combinations. This combinatorial analysis yielded a set of five features that correlate significantly with Vβ use (Figure 2.9D; Pearson

correlation coefficient = 0.69, P = 0.03). In descending order of contribution to the fitted model, the identified features were H3K4me3, H3K4me2, transcription, P300, and CTCF. The first four features largely determine the efficiency for most TCRBV segments, whereas the remaining feature, CTCF proximity, improves the fit for several outliers that are poorly predicted by H3K4me3, H3K4me2, transcription, and P300. When further analyzed by clustering, we found that the four chromatin features (H3K4me3, H3K4me2, P300, and CTCF) in this set of five core parameters represent four classes of related marks that share a significant portion of epigenetic information (Figure 2.9E). For example, H3K4me3 correlates strongly with H3ac and RNA Pol II occupancy, three features enriched near active promoters, in essence encapsulating the information content of the entire class. The relative contributions of the five core features to the accuracy of fit and the corresponding linear regression formula are provided in Figure 2.10B.

Together, the computational analyses derive a two-tiered model for predicting Vβ use in the preselection *Tcrb* repertoire. First, RIC scores in combination with nucleosome and RNA Pol II densities discriminate active from inert substrates. The recombination frequency of the active Vβ set can be discerned from values for the five core parameters identified by statistical correlations. Moreover, this basal set of five parameters may be useful in future studies to predict the impact on preselection Vβ repertoires of naturally occurring or engineered perturbations at *Tcrb*.

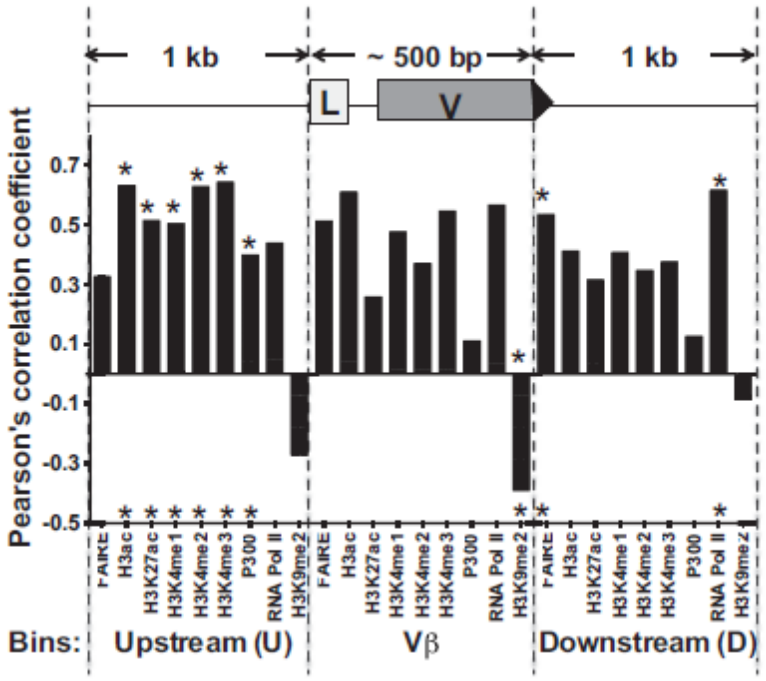


Figure 2.8: Spatial distribution of chromatin features and predictive potential for Vβ use. The regions surrounding each Vβ segment were divided into three bins (see schematic); U, upstream (1 kb); V, Vβ gene body; D, downstream (1 kb). Signal densities for each chromatin feature in the spatial bins were correlated with recombination frequencies, yielding a Pearsons's correlation coefficient for each bin. The coefficients were used to determine the best bin, which are denoted by asterisks.

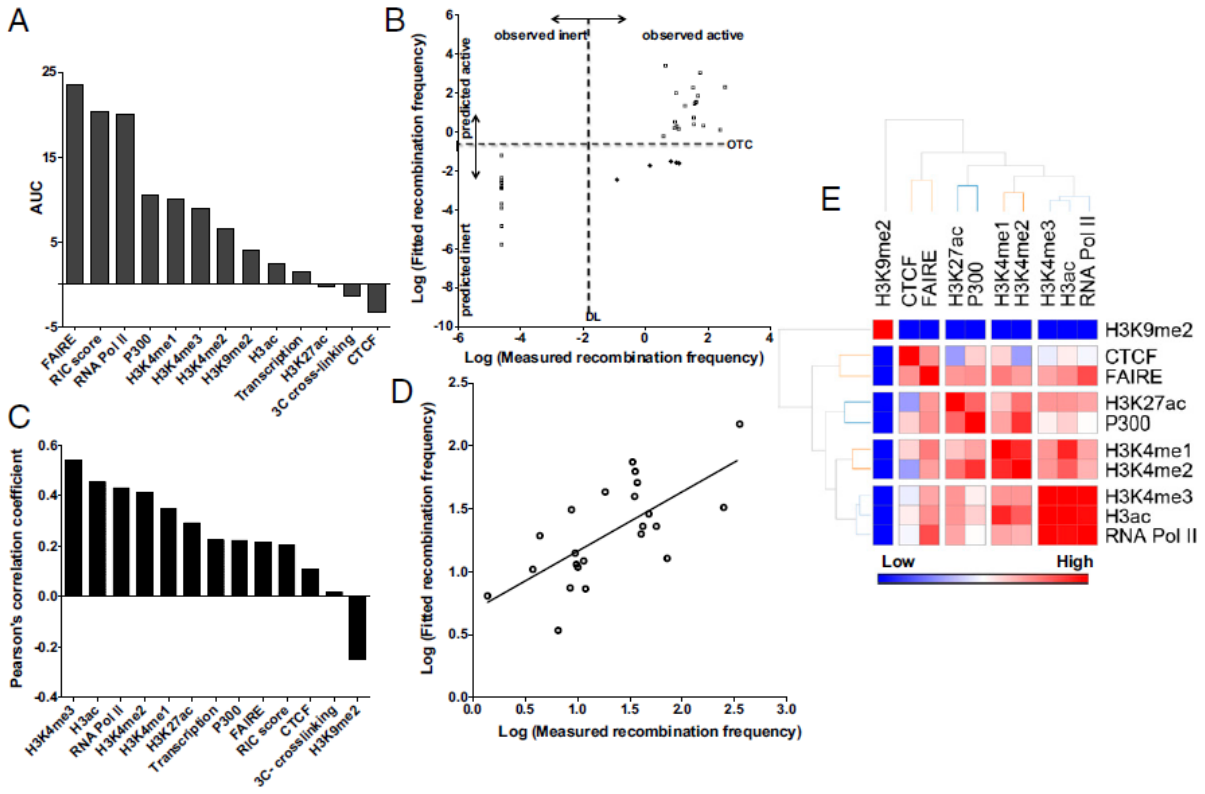


Figure 2.9: Computational analysis of Vβ use determinants. (A) Features that distinguish rearranging from inert Vβ segments (classifier step; Materials and Methods). Random forest analysis was performed on the shown features to classify Vβ segments. AUC, area under the curve, which represents the relative contribution of each feature to the learned classification scheme. (B) Scatter plot representing the classifier step in the two-step model. Linear regression between observed and fitted frequencies using the three most discriminative features for recombining vs. inert Vβ gene segments (RIC scores, FAIRE signal, and RNA Pol II occupancy). Each symbol represents a Vβ gene segment. Data were generated from the natural logarithm values of recombination frequencies (observed and fitted). The dashed horizontal line represents the optimal threshold for classifying (OTC) rearranging from nonrearranging segments based on the linear combination of the three features. The dashed vertical line represents the detection limit (DL) of Taqman assays used for measuring recombination. Open circles correspond to Vβ segments predicted accurately; black diamonds correspond to outliers. Two of these five exceptions were resolved when the random forest algorithm was applied using the three classification features (RIC score, FAIRE, RNA Pol II). (C) Pearson correlation to rank factors that fine tune Vβ use in the two-step model (regressor step; Materials and Methods). (D) Scatter plot of overall correlation between natural log values of observed and fitted (predicted) frequencies using the five core parameters (H3K4me3, H3K4me2, transcription, P300, and CTCF). Each circle represents one rearranging Vβ segment. The line indicates the best fit between measured and fitted rearrangement frequencies and reflects a strong correlation (Pearson correlation coefficient, 0.69; P = 0.03). (E) Cluster analysis highlights similarities in epigenetic information provided by individual chromatin features.

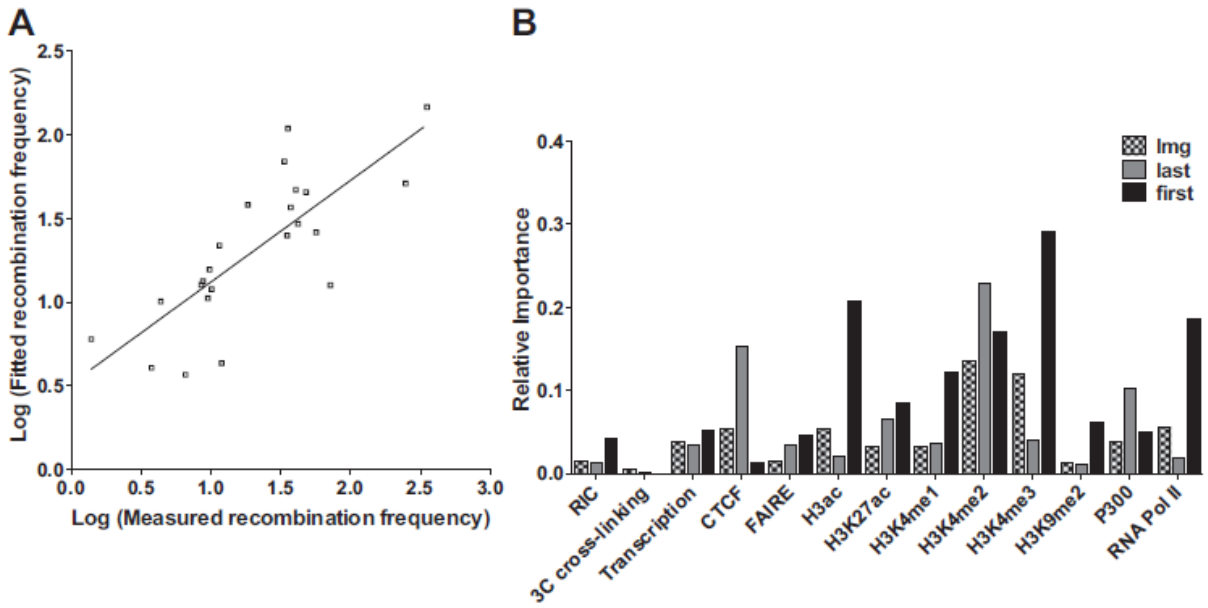


Figure 2.10: Computational analysis of Vβ use determinants. (A) Scatter plot of overall correlation between natural log values of observed and fitted frequencies using the complete set of 13 features. Each circle represents one rearranging Vβ segment. The line indicates the best fit between measured and fitted rearrangement frequencies reflect a strong correlation (Pearson correlation coefficient, 0.779; P = 0.47). (B) Relative contribution of the minimal eight features to the accuracy of fit as computed by three different approaches (lmg, last, first) and the corresponding linear regression coefficients. The best fit formula is as follows:

$$\sum_{i=1}^{8} \text{Coefficient(i)} \times \text{Feature(i)}$$

The raw values and coefficients corresponding to each feature are provided in Tables T5–T8.

2.5 Discussion

We took an integrative approach to define the molecular determinants of Vβ recombination frequencies, an important component of the preselection *Tcrb* repertoire. Prior studies have examined the independent effects of RSS quality, 3D architecture, transcription, or chromatin accessibility on recombination of specified gene segments. However, our unified analysis shows how these features impact the efficiency of long-range V to (D)J recombination at an endogenous AgR locus. Using several independent computational approaches, we find that (i) RSS quality and nucleosome density are the major determinants of whether a given Vβ segment will participate in *Tcrb* gene assembly, (ii) the relative use of a Vβ segment is fine-tuned by its chromatin environment, (iii) the optimal epigenetic landscape for Vβ recombination is a blend of transcriptional activation marks, nucleosome depletion, and a lack of the repressive H3K9me2 mark, and (iv) the precise magnitude of spatial proximity between a Vβ segment and the DβJβ recombination center does not significantly influence its relative utilization. Collectively, we find that a minimum set of five features can be measured to predict the recombination frequency of a competent Vβ segment with a high degree of accuracy.

A critical component of our study was a determination of the preselection Vβ repertoire. The relative use of Vβ segments may have important consequences with regard to AgR-mediated thymic selection, the production of functional T-cell subsets that use specific Vβ segments, or the baseline antigenic profile recognized by emerging T lymphocytes. We used a DNA-based approach to directly quantify rearrangement levels of the 35 Vβ segments in sorted DN3 thymocytes. This approach avoids two caveats of prior repertoire analyses, biases introduced by thymocyte selection or by mRNA expression differences, both of which were observed in our

companion assays. We find that only a few functional Vβ segments are either over- or underused in the preselection *Tcrb* repertoire. One of the overused Vβ segments, TRBV13-2 (formerly Vβ8.2), is enriched in invariant natural killer (iNKT) cells, a subset of lymphocytes that respond to lipid antigens and produce a robust cytokine response. We postulate that the ideal chromatin environment encompassing TRBV13-2 has evolved to augment its rearrangement efficiency, ensuring a sufficient production of iNKT cells, which provide a rapid cellular immune response to numerous foreign antigens. Notwithstanding, rearrangement levels for the vast majority of functional Vβ gene segments (18/22) fall within a threefold range. The relatively limited range of distribution likely reflects a requirement to maximize *Tcrb* diversity before its pairing with *Tcra* for subsequent selection by MHC-peptide complexes.

As shown here, the normalization of Vβ use results predominantly from the chromatin environment encompassing each gene segment, with perhaps a minor contribution from its RSS quality. The dominance of chromatin in fine-tuning Vβ use was evident from several outlier gene segments. The TRBV15 and TRBV30 segments are underused compared with all of the other functional Vβ elements, likely because they are poorly transcribed or lack most features of active chromatin. Likewise, nearly all of the pseudogene segments that are flanked by functional RSSs reside in a repressive chromatin environment. For the latter category, we provide evidence that some, but not all, germline promoters associated with pseudo-Vβ segments have been incapacitated, despite their retention of potential factor binding sites found in functional Vβ promoters. Another potential mechanism for pseudogene suppression could be their localization to the nuclear periphery or lamina (Reddy et al., 2008). However, the precise underlying

mechanisms that sequester these pseudogene segments in repressive chromatin, preventing wasteful recombination, remain to be defined.

With regard to the collection of rearranging Vβ segments, the dominant chromatin features in determining their relative use are associated with active transcription. The strongest correlations exist between recombination efficiencies, histone acetylation (H3ac), H3K4 methylation, nucleosome depletion, and RNA Pol II occupancy. Although a link between this transcriptional epigenetic state and recombination has long been appreciated, its dominant role in sculpting the primary repertoire of antigen receptors is a unique finding of our study. One likely mechanism for this relationship is the affinity of RAG complexes for chromatin bearing the H3K4me3 mark. Prior ChIP-seq studies demonstrate that RAG-1/2 is bound to the DβJβ recombination center in DN thymocytes but is relatively absent from the Vβ cluster (Ji et al., 2010). This reflects the extremely high levels of H3K4me3 on DβJβ chromatin compared with Vβ segments (~10-fold difference) (Ji et al., 2010). Based on our integrative model, we suggest that after *Tcrb* contracts, pre-bound RAG-1/2 complexes at the DβJβ recombination center may preferentially target Vβ segments that are most enriched for transcription-associated marks, including H3K4me3. Thus, the strength of each Vβ promoter within its native chromosomal context may be a dominant feature for shaping the preselection *Tcrb* repertoire.

One important aspect of our study is that the precise magnitude of association between a Vβ segment and DβJβ clusters, as measured by 3C, does not contribute discernibly to its level of use. Clearly, general locus contraction is an important mechanism for bringing V segments into spatial proximity with their distant (D)J substrates (Bossen et al., 2012). However, the spatial architecture adopted by the large Vβ cluster in DN thymocytes must provide sufficient access to

all of its composite gene segments by RAG-1/2 bound at the DβJβ recombination center. Recent studies of *Igk* suggest that most V segments within this locus also may have similar spatial access to their target J segments (Lin et al., 2012). Given the 10-fold range in cross-linking efficiencies between various Vβ segments and the two DβJβ clusters, we conclude that spatial constraints on long-range Vβ to DβJβ recombination are binary rather than digital, requiring only that target gene segments cross a threshold of spatial proximity. Presumably, this spatial threshold is surpassed via a combination of locus contraction and folding of the Vβ cluster into a more compact structure.

In conclusion, a combination of epigenetic, spatial, transcriptional, and RSS features were used to identify the dominant determinants for sculpting the preselection Vβ repertoire. We concede that a model for Vβ use may not completely apply to all other AgR loci. Indeed, pseudo-Vκ segments interact inefficiently with their target Jκ cluster, perhaps suppressing their recombination (Lin et al., 2012). In contrast, pseudo- and functional Vβ segments interact indistinguishably with their DβJβ substrates. Recombination of pseudo-Vβ segments is, instead, suppressed by sequestration into inactive chromatin. This distinction may reflect a more dominant role for spatial constraints at the much larger *Igk* locus. Notwithstanding, much of the relevant epigenetic and RSS quality data necessary to build predictive models for other AgR loci are available publicly. In most cases, the lacking features are reliable DNA-based analysis of V use and complete sets of 3C data covering V clusters. We suspect that as multiplex PCR approaches improve, eliminating primer bias, comprehensive preselection repertoires for all AgR loci will emerge. Current methods for quantifying spatial proximity on a global scale lack the resolution of focused 3C assays; however, technical improvements and increased sequencing depths may soon overcome these obstacles. The learned model-building strategy used here should be a valuable

guide for defining relative contributions of epigenetic, spatial, and RSS features in shaping preselection V repertoires. Ultimately, these models should also be valuable for predicting how designed or naturally occurring alterations of AgR loci perturb the preselection V repertoire. These alterations could range from targeted RSS and promoter substitutions to natural variant AgR alleles that lack portions of the large V clusters, creating "holes" in the immune repertoire. Indeed, a striking parallel exists between the use of several mouse and human Vβ orthologs (Livák, 2003), underscoring the potential utility of our model to predict the effects of human TCRB polymorphisms on primary repertoire formation.

2.6 Refeernces

Aoki-Ota, M., Torkamani, A., Ota, T., Schork, N., & Nemazee, D. (2012). Skewed Primary Igκ Repertoire and V–J Joining in C57BL/6 Mice: Implications for Recombination Accessibility and Receptor Editing. *The Journal of Immunology*, *188*(5), 2305-2315.

Bassing, C. H., Swat, W., & Alt, F. W. (2002). The mechanism and regulation of chromosomal V (D) J recombination. *Cell*, *109*(2), S45-S55.

Bossen, C., Mansson, R., & Murre, C. (2012). Chromatin topology and the regulation of antigen receptor assembly. *Annual review of immunology*, *30*, 337-356.

Chaumeil, J., Micsinai, M., Ntziachristos, P., Deriano, L., Wang, J. M. H., Ji, Y., ... & Skok, J. A. (2013). Higher-order looping and nuclear organization of Tcra facilitate targeted rag cleavage and regulated rearrangement in recombination centers. *Cell reports*, *3*(2), 359-370.

Chen, Y., Meyer, C. A., Liu, T., Li, W., Liu, J. S., & Liu, X. S. (2011). MM-ChIP enables integrative analysis of cross-platform and between-laboratory ChIP-chip or ChIP-seq data. *Genome Biol*, *12*(2), R11.

Cobb, R. M., Oestreich, K. J., Osipovich, O. A., & Oltz, E. M. (2006). Accessibility control of V (D) J recombination. *Advances in immunology*, *91*, 45-109.

Cowell, L. G., Davila, M., Yang, K., Kepler, T. B., & Kelsoe, G. (2003). Prospective estimation of recombination signal efficiency and identification of functional cryptic signals in the genome by statistical modeling. *The Journal of experimental medicine*, *197*(2), 207-220.

Cuomo, C. A., Mundy, C. L., & Oettinger, M. A. (1996). DNA sequence and structure requirements for cleavage of V (D) J recombination signal sequences.*Molecular and Cellular Biology*, *16*(10), 5683-5690.

Degner, S. C., Verma-Gaur, J., Wong, T. P., Bossen, C., Iverson, G. M., Torkamani, A., ... & Feeney, A. J. (2011). CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the Igh locus and antisense transcription in pro-B cells. *Proceedings of the National Academy of Sciences*,*108*(23), 9566-9571.

Dekker, J. (2008). Gene regulation in the third dimension. *Science*, *319*(5871), 1793-1794.

Dong, X., Greven, M. C., Kundaje, A., Djebali, S., Brown, J. B., Cheng, C., ... & Weng, Z. (2012). Modeling gene expression using chromatin features in various cellular contexts. *Genome Biol*, *13*(9), R53.

Ernst, J., Kheradpour, P., Mikkelsen, T. S., Shoresh, N., Ward, L. D., Epstein, C. B., ... & Bernstein, B. E. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, *473*(7345), 43-49.

Feeney, A. J. (2009). Genetic and epigenetic control of V gene rearrangement frequency. In *V (D) J Recombination* (pp. 73-81). Springer New York.

Feeney, A. J., Tang, A., & Ogwaro, K. M. (2000). B-cell repertoire formation: role of the recombination signal sequence in non-random V segment utilization.*Immunological reviews*, *175*(1), 59-69.

Ferrier, P., Krippl, B., Blackwell, T. K., Furley, A. J., Suh, H. E. I. K. Y. U. N. G., Winoto, A., ... & Alt, F. W. (1990). Separate elements control DJ and VDJ rearrangement in a transgenic recombination substrate. *The EMBO journal*,*9*(1), 117.

Fuxa, M., Skok, J., Souabni, A., Salvagiotto, G., Roldan, E., & Busslinger, M. (2004). Pax5 induces V-to-DJ rearrangements and locus contraction of the immunoglobulin heavy-chain gene. *Genes & development*, *18*(4), 411-422.

Gerstein, R. M., & Lieber, M. R. (1993). Coding end sequence can markedly affect the initiation of V (D) J recombination. *Genes & development*, *7*(7b), 1459-1469.

Giresi, P. G., & Lieb, J. D. (2009). Isolation of active regulatory elements from eukaryotic chromatin using FAIRE (Formaldehyde Assisted Isolation of Regulatory Elements). *Methods*, *48*(3), 233-239.

Godfrey, D. I., Hammond, K. J., Poulton, L. D., Smyth, M. J., & Baxter, A. G. (2000). NKT cells: facts, functions and fallacies. *Immunology today*, *21*(11), 573-583.

Guo, C., Gerasimova, T., Hao, H., Ivanova, I., Chakraborty, T., Selimyan, R., ... & Sen, R. (2011). Two forms of loops generate the chromatin conformation of the immunoglobulin heavy-chain gene locus. *Cell*, *147*(2), 332-343.

Guo, C., Yoon, H. S., Franklin, A., Jain, S., Ebert, A., Cheng, H. L., ... & Alt, F. W. (2011). CTCF-binding elements mediate control of V (D) J recombination.*Nature*, *477*(7365), 424-430.

Hagège, H., Klous, P., Braem, C., Splinter, E., Dekker, J., Cathala, G., ... & Forné, T. (2007). Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nature protocols*, *2*(7), 1722-1733.

Hesse, J. E., Lieber, M. R., Mizuuchi, K., & Gellert, M. (1989). V (D) J recombination: a functional definition of the joining signals. *Genes & development*, *3*(7), 1053-1061.

Jackson, A., Kondilis, H. D., Khor, B., Sleckman, B. P., & Krangel, M. S. (2005). Regulation of T cell receptor β allelic exclusion at a level beyond accessibility. *Nature immunology*, *6*(2), 189-197.

Jhunjhunwala, S., van Zelm, M. C., Peak, M. M., Cutchin, S., Riblet, R., van Dongen, J. J., ... & Murre, C. (2008). The 3D structure of the immunoglobulin heavy-chain locus: implications for long-range genomic interactions. *Cell*,*133*(2), 265-279.

Ji, Y., Resch, W., Corbett, E., Yamane, A., Casellas, R., & Schatz, D. G. (2010). The in vivo pattern of binding of RAG1 and RAG2 to antigen receptor loci. *Cell*, *141*(3), 419-431.

Jung, D., Bassing, C. H., Fugmann, S. D., Cheng, H. L., Schatz, D. G., & Alt, F. W. (2003). Extrachromosomal recombination substrates recapitulate beyond 12/23 restricted V (D) J recombination in nonlymphoid cells. *Immunity*, *18*(1), 65-74.

Karlić, R., Chung, H. R., Lasserre, J., Vlahoviček, K., & Vingron, M. (2010). Histone modification levels are predictive for gene expression. *Proceedings of the National Academy of Sciences*, *107*(7), 2926-2931.

Kosak, S. T., Skok, J. A., Medina, K. L., Riblet, R., Le Beau, M. M., Fisher, A. G., & Singh, H. (2002). Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science*, *296*(5565), 158-162.

Kwon, J., Morshead, K. B., Guyon, J. R., Kingston, R. E., & Oettinger, M. A. (2000). Histone acetylation and hSWI/SNF remodeling act in concert to stimulate V (D) J cleavage of nucleosomal DNA. *Molecular cell*, *6*(5), 1037-1048.

Lee, A. I., Fugmann, S. D., Cowell, L. G., Ptaszek, L. M., Kelsoe, G., & Schatz, D. G. (2003). A functional analysis of the spacer of V (D) J recombination signal sequences. *PLoS biology*, *1*(1), e1.

Lefranc, M. P., Giudicelli, V., Ginestoux, C., Jabado-Michaloud, J., Folch, G., Bellahcene, F., ... & Duroux, P. (2009). IMGT®, the international ImMunoGeneTics information system®. *Nucleic acids research*, *37*(suppl 1), D1006-D1012.

Lin, Y. C., Benner, C., Mansson, R., Heinz, S., Miyazaki, K., Miyazaki, M., ... & Murre, C. (2012). Global changes in the nuclear positioning of genes and intra-and interdomain genomic interactions that orchestrate B cell fate. *Nature immunology*, *13*(12), 1196-1204.

Liu, H., Schmidt-Supprian, M., Shi, Y., Hobeika, E., Barteneva, N., Jumaa, H., ... & Shi, Y. (2007). Yin Yang 1 is a critical regulator of B-cell development.*Genes & development*, *21*(10), 1179-1189.

Liu, Y., Subrahmanyam, R., Chakraborty, T., Sen, R., & Desiderio, S. (2007). A plant homeodomain in RAG-2 that binds Hypermethylated lysine 4 of histone H3 is necessary for efficient antigen-receptor-gene rearrangement. *Immunity*,*27*(4), 561-571.

Livák, F. (2003). Evolutionarily conserved pattern of gene segment usage within the mammalian TCRβ locus. *Immunogenetics*, *55*(5), 307-314.

Matthews, A. G., Kuo, A. J., Ramón-Maiques, S., Han, S., Champagne, K. S., Ivanov, D., ... & Oettinger, M. A. (2007). RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V (D) J recombination. *Nature*, *450*(7172), 1106-1110.

Nadel, B., Tang, A., Lugo, G., Love, V., Escuro, G., & Feeney, A. J. (1998). Decreased frequency of rearrangement due to the synergistic effect of nucleotide changes in the heptamer and nonamer of the recombination signal sequence of the Vκ gene A2b, which is associated with increased susceptibility of Navajos to Haemophilus influenzae type b disease. *The Journal of Immunology*, *161*(11), 6068-6073.

Ndifon, W., Gal, H., Shifrut, E., Aharoni, R., Yissachar, N., Waysbort, N., ... & Friedman, N. (2012). Chromatin conformation governs T-cell receptor Jβ gene segment usage. *Proceedings of the National Academy of Sciences*, *109*(39), 15865-15870.

Oestreich, K. J., Cobb, R. M., Pierce, S., Chen, J., Ferrier, P., & Oltz, E. M. (2006). Regulation of TCRβ gene assembly by a promoter/enhancer holocomplex. *Immunity*, *24*(4), 381-391.

Olaru, A., Patterson, D. N., Villey, I., & Livák, F. (2003). DNA-Rag protein interactions in the control of selective D gene utilization in the TCRβ locus. *The Journal of Immunology*, *171*(7), 3605-3611.

Osipovich, O., Cobb, R. M., Oestreich, K. J., Pierce, S., Ferrier, P., & Oltz, E. M. (2007). Essential function for SWI-SNF chromatin-remodeling complexes in the promoter-directed assembly of Tcrb genes. *Nature immunology*, *8*(8), 809-816.

Pekowska, A., Benoukraf, T., Zacarias-Cabeza, J., Belhocine, M., Koch, F., Holota, H., ... & Spicuglia, S. (2011). H3K4 tri-methylation provides an epigenetic signature of active enhancers. *The EMBO journal*, *30*(20), 4198-4210.

Posnett, D. N., Vissinga, C. S., Pambuccian, C., Wei, S., Robinson, M. A., Kostyu, D., & Concannon, P. (1994). Level of human TCRBV3S1 (V beta 3) expression correlates with allelic polymorphism in the spacer region of the recombination signal sequence. *The Journal of experimental medicine*, *179*(5), 1707-1711.

Reddy, K. L., Zullo, J. M., Bertolino, E., & Singh, H. (2008). Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature*,*452*(7184), 243-247.

Reynaud, D., Demarco, I. A., Reddy, K. L., Schjerven, H., Bertolino, E., Chen, Z., ... & Singh, H. (2008). Regulation of B cell fate commitment and immunoglobulin heavy-chain gene rearrangements by Ikaros. *Nature immunology*, *9*(8), 927-936.

Ribeiro de Almeida, C., Stadhouders, R., de Bruijn, M. J., Bergen, I. M., Thongjuea, S., Lenhard, B., ... & Hendriks, R. W. (2011). The DNA-binding protein CTCF limits proximal Vκ recombination and restricts κ enhancer interactions to the immunoglobulin κ light chain locus. *Immunity*, *35*(4), 501-513.

Robins, H. S., Campregher, P. V., Srivastava, S. K., Wacher, A., Turtle, C. J., Kahsai, O., ... & Carlson, C. S. (2009). Comprehensive assessment of T-cell receptor β-chain diversity in αβ T cells. *Blood*, *114*(19), 4099-4107.

Rubio, E. D., Reiss, D. J., Welcsh, P. L., Disteche, C. M., Filippova, G. N., Baliga, N. S., ... & Krumm, A. (2008). CTCF physically links cohesin to chromatin. *Proceedings of the National Academy of Sciences*, *105*(24), 8309-8314.

Schatz, D. G., & Ji, Y. (2011). Recombination centres and the orchestration of V (D) J recombination. *Nature Reviews Immunology*, *11*(4), 251-263.

Seitan, V. C., Hao, B., Tachibana-Konwalski, K., Lavagnolli, T., Mira-Bontenbal, H., Brown, K. E., ... & Merkenschlager, M. (2011). A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation. *Nature,476*(7361), 467-471.

Shih, H. Y., Verma-Gaur, J., Torkamani, A., Feeney, A. J., Galjart, N., & Krangel, M. S. (2012). Tcra gene recombination is supported by a Tcra enhancer-and CTCF-dependent chromatin hub. *Proceedings of the National Academy of Sciences*, *109*(50), E3493-E3502.

Shimazaki, N., Tsai, A. G., & Lieber, M. R. (2009). H3K4me3 stimulates the V (D) J RAG complex for both nicking and hairpinning in trans in addition to tethering in cis: implications for translocations. *Molecular cell*, *34*(5), 535-544.

Skok, J. A., Gisler, R., Novatchkova, M., Farmer, D., de Laat, W., & Busslinger, M. (2007). Reversible contraction by looping of the Tcra and Tcrb loci in rearranging thymocytes. *Nature immunology*, *8*(4), 378-387.

Spicuglia, S., Payet, D., Tripathi, R. K., Rameil, P., Verthuy, C., Imbert, J., ... & Hempel, W. M. (2000). TCRα enhancer activation occurs via a conformational change of a pre-assembled nucleo-protein complex. *The EMBO journal*, *19*(9), 2034-2045.

Williams, G. S., Martinez, A., Montalbano, A., Tang, A., Mauhar, A., Ogwaro, K. M., ... & Feeney, A. J. (2001). Unequal VH gene rearrangement frequency within the large VH7183 gene family is not due to recombination signal sequence variation, and mapping of the genes shows a bias of rearrangement based on chromosomal location. *The Journal of Immunology*, *167*(1), 257-263.

Wilson, A., Maréchal, C., & MacDonald, H. R. (2001). Biased Vβ usage in immature thymocytes is independent of DJβ proximity and pTα pairing. *The Journal of Immunology*, *166*(1), 51-57.

Wu, C., Bassing, C. H., Jung, D., Woodman, B. B., Foy, D., & Alt, F. W. (2003). Dramatically increased rearrangement and peripheral representation of Vβ14 driven by the 3′ Dβ1 recombination signal sequence. *Immunity*, *18*(1), 75-85.

Xiang, Y., Zhou, X., Hewitt, S. L., Skok, J. A., & Garrard, W. T. (2011). A multifunctional element in the mouse Igκ locus that specifies repertoire and Ig loci subnuclear location. *The Journal of Immunology*, *186*(9), 5356-5366.

Yu, K., & Lieber, M. R. (1999). Mechanistic basis for coding end sequence effects in the initiation of V (D) J recombination. *Molecular and cellular biology*, *19*(12), 8094-8102.

Table T1: 3C ranks and rearrangement frequencies

3C HindIII fragments	Rank Eβ	Rank Dβ1	Rank Dβ2	Average rank	Percent recombination
V1	21	20	25	22	4.80392
V2	24	23	17	21.3	1.890142
V3	24	23	17	21.3	4.983436
V4	1	5	2	2.67	1.766591
V5	1	5	2	2.67	4.590954
V6	13	3	12	9.33	0
V7	18	8	22	16	0
V8	18	8	22	16	0
V9	5	7	10	7.33	0
V10	10	16	3	9.67	0
V11	9	12	6	9	0
V12-1	6	14	19	13	5.360448
V13-1	8	9	11	9.33	4.710525
V12-2	8	9	11	9.33	3.534854
V13-2	8	9	11	9.33	12.76473
V12-3	7	4	8	6.33	0
V13-3	2	19	16	12.3	6.387671
V14	3	1	4	2.67	2.558942
V15	17	10	14	13.7	1.14465
V16	11	11	21	14.3	5.066513
V17	12	6	1	6.33	2.652322
V18	4	2	23	9.67	0
V19	4	2	23	9.67	10.95472
V20	15	21	7	14.3	2.722579
V22	22	18	5	15	2.254725
V23	23	13	20	18.7	2.528768
V24	25	17	24	22	4.680041
V25	25	17	24	22	0
V26	19	24	18	20.3	2.878938
V27	20	22	9	17	0
V28	14	25	15	18	0
V29	16	15	13	14.7	2.927078

Table T2 : Primers and probes for Vβ utilization assay

Taqman probes (5′FAM and 3' TAMRA from Sigma Life Sciences)	Sequences
Jβ 1.1 probe	5'FAM-TGTGAGTCTGGTTCCTTTACCAA-3′TAMRA
Jβ 2.1 Probe	5'HEX-TAGGACGGTGAGTCGTGTCC-3′TAMRA
Primers for cloning VβJβ template plasmids	Sequences
Jβ 1.1 F	5′-GACAGACGGATCCTGGCACTGTGCAAACACAGAAGTC-3′
Jβ 1.1 R	5′-TACATCGCGGCCGCACTCGAATATGGACACGGAGGACA-3′
Jβ2.1F	5′-GACAGACGGATCCGTAACTATGCTGAGCAGTTCTTCGGACC-3′
Jβ2.1R	5′-TACATCGCGGCCGCAGTCCTGGAAATGCTGGCACAAAC-3′
V1-F	5′-TATCTCGAGCTGGAGCAAAACCCAAGGTGG-3′
V1-R	5′-CGAGAAGCTTTGCAGTACAAGGTTCTGCCCT-3′
V2-F	5′-TATCTCGAGCGAAAATTATCCAGAAACCAA-3′
V2-R	5′-CGAGAAGCTTGCACAGAAGTATGTGGCCGAG-3′
V3-F	5′-TATCTCGAGCAGATGGTGACCCTCAATTGT-3′
V3-R	5′-TAGCGAAGCTTTAAGCTGCTGGCACAGAAG-3′
V4-F	5′-TATCTCGAGGACGGCTGTTTTCCAGACTC-3′
V4-R	5′-CGAGAAGCTTTGGCACAGAGATACACAGCAG-3′
V5-F	5′-TATCTCGAGGATATCTAATCCTGGGAAGAGC-3′
V5-R	5′-CGAGAAGCTTCTGCCGTGGATCCAGAAGACT-3′
V6-F	5′-TATCTCGAGGTTACAGACATGGGACAGAATGTCA-3′
V6-R	5′-CGAGAAGCTTAGCTGCTGGCATACATAGTGGAGT-3′
V7-F	5′-TATCTCGAGAGCAGGCTCTGTCTTCTGACTTGT-3′
V7-R	5′-CGAGAAGCTTAGAACAGTGCAGAGTCCTTTGGCT-3′
V8-F	5′-TAGCCTCGAGCATTCAGACTCCCAAATCAT-3′
V8-R	5′-TAGCGAAGCTTTCTGTGCATGATCTGGAGAC-3′
V9-F	5′-TATCTCGAGGTGACACAATTTCTGGTCCTACTGG-3′
V9-R	5′-CGAGAAGCTTCTTCTGGCACAGAGATAGATGCCT-3′
V10-F	5′-TATCTCGAGGGTGGAATCACCCAGACACCTAGATA-3′
V10-R	5′-CGAGAAGCTTAGTACATGGAGGTCTGGTTGGAACTG-3′
V11-F	5′-TATCTCGAGAGGCACTTCTGATATGTGGCCTCT-3′

V11-R	5′-CGAGAAGCTTAGTTAGAAACCATGGCTCTTGCCC-3′
V12-1-F	5′-TATCTCGAGCTGACGTGTATTCCCATCTCT-3′
V12-1-R	5′-TAGCGAAGCTTTCCAGTTCCAAGGCACTCATG-3′
V13-1-F	5′-TATCTCGAGTGGTTAGCCCAAGTGTGCTTCTCT-3′
V13-1-R	5′-CGAGAAGCTTAAGCCAATTCCAGCAGGAGGAAGA-3′
V12-2-F	5′-TATCTCGAGCATTGCTGCTGCTGCTGCTGC-3′
V12-2-R	5′-TAGCGAAGCTTACACGGCAGAGTCCTCTAG-3′
V13-2-F	5′-TATCTCGAGTCCTGTGTTCAAGTGAGTGCTGGT-3′
V13-2-R	5′-CGAGAAGCTTTTGGTCTGGAGGCCTTGTATCCAT-3′
V12-3-F	5′-TAGCCTCGAGCCTTCTCCCCAGGTTCAGC-3′
V12-3-R	5′-TAGCGAAGCTTCACAGTAAAGTCCTCTAGGTCC-3′
V13-3-F	5′-TATCTCGAGGACGATATGATCAGGCTTTG-3′
V13-3-R	5′-TAGCGAAGCTTAGAAATATACAGCTGTCTGAG-3′
V14-F	5′-TATCTCGAGTATGCAGTCCTACAGGAAGGGCAA-3′
V14-R	5′-CGAGAAGCTTAAACTGCTGGCACAGAGATAGGTG-3′
V15-F	5′-TATCTCGAGCAGACACCCAGACATGAGGT-3′
V15-R	5′-CGAGAAGCTTACAGCTGAGTCCTTGGGTTCTG-3′
V16-F	5′-TATCTCGAGCACCTAGGCACAAGGTGACA-3′
V16-R	5′-TAGCGAAGCTTCAGGACTCAGCGGTGTATCT-3′
V17-F	5′-TATCTCGAGGGATACTACGGTTAAGCAGAAC-3′
V17-R	5′-TAGCGAAGCTTAGCACAGAGGTACATGGCAG-3′
V18-F	5′-TAGCCTCGAGGCTGGTGTCACCACGAACCT-3′
V18-R	5′-TAGCGAAGCTTCTCTGCATCTTCCAGATCTGC-3′
V19-F	5′-TATCTCGAGCTCAGACACCCAAATTCCTGA-3′
V19-R	5′-CGAGAAGCTTGCTATACTGCTGGCACAGAGA-3′
V20-F	5′-TATCTCGAGCGTCTATCAATATCCCAGAAG-3
V20-R	5′-CGAGAAGCTTAGCACCACAGAGATATAAGCC-3′
V21-F	5′-TAGCCTCGAGGTTGTCCAGAATCCTAGACAT-3′
V21-R	5′-TAGCGAAGCTTGTACACAGCTGAATCTGTTAG-3′
V22-F	5′-TATCTCGAGCCAAGTTATCCAGACTCCAT-3′
V22-R	5′-TAGCGAAGCTTATAACACTGAGTCTCCAGCCTC-3′
V23-F	5′-TATCTCGAGGAAAGGCCAGGAAGCAGAGAT-3′
V23-R	5′-CGAGAAGCTTGCTGGAGCACAAGTACAGTGC-3′
V24-F	5′-TATCTCGAGGAGTAACCCAGACTCCACGAT-3′
V24-R	5′-CGAGAAGCTTGACTGCTGGCACAGAGCTACA-3′
V25-F	5′-TAGCCTCGAGCTAGCTTCAAGGCTCTTCTA-3′
V25-R	5′-TAGCGAAGCTTATGTAGAATCTCCTGCTTCT-3′
V26-F	5′-TATCTCGAGCAGACTCCAAGATATCTGGTG-3′
V26-R	5′-CGAGAAGCTTCTGCTGGCACAGAGGTACAGT-3′

V27-F	5′-TAGCCTCGAGCTCCAAAGTACTCTATTATG-3′
V27-R	5′-TAGCGAAGCTTGAGGTAGGATTCATTCTCTG-3′
V28-F	5′-TAGCCTCGAGCATCCAAATCGCAAGACACC-3′
V28-R	5′-TAGCGAAGCTTAGGTGCACACATGCCTGGTCG-3′
V29-F	5′-TATCTCGAGCTGATCAAAAGAATGGGAGAG-3′
V29-R	5′-CGAGAAGCTTCTAGCACAGAAGTACACAGATG-3′
V30-F	5′-TATCTCGAGTGCTTGCCTCATGGATCTCTGTCT-3′
V30-R	5′-CGAGAAGCTTGAACTACAGAAATAGATACTGC-3′
V31-F	5′-TAGCCTCGAGCTGAGACTGATTACATGTAA-3′
V31-R	5′-TAGCGAAGCTTAGAAGCCAGAGTGGCTGAGA-3′
qPCR primers for Taqman assay	Sequences
qV1F	5′-GCCACACGGGTCACTGATAC-3′
qV2F	5′-GTTCAAAGAAAAACCATTTAG-3′
qV3F	5′-GATGGTTCATATTTCACTCT-3′
qV4F	5′-CAGATAAAGCTCATTTGAAT-3′
qV5F	5′-GCCCAGACAGCTCCAAGCTAC-3′
qV6F	5′-CAGAGATGCCTGATGGATTGTT-3′
qV7F	5′-CAGCACACCAATTTGGTGACT-3′
qV8F	5′-GAGGTCTCTAAGGGGTAC-3′
qV9F	5′-CTTCTCCATGTTGAAGAGCCAA-3′
qV10F	5′-AGAAATGAGATACAGAGCTTTCC-3′
qV11F	5′-AGTTAGAAACCATGGCTCTTGC-3′
qV12-1F	5′-’TAGCAATGTGGTCTGGTACCAG-3′
qV13-1F	5′-GGTACAAGGCCACCAGAACA-3′
qV12-2F	5′-TCTCTCTGTGGCCTGGTATCAA-3′
qV13-2F	5′-GCTGGCAGCACTGAGAAAGGA-3′
qV12-3F	5′-CCTGAGTGCCTTGGACCT-3′
qV13-3F	5′-TTCCCTTTCTCAGACAGCTGTA-3′
qV14F	5′-TATCAGCAGCCCAGAGACCAG-3′
qV15F	5′-CACTCTGAAGATTCAACCT-3′
qV16F	5′-CTCAGCTCAGATGCCCAAT-3′
qV17F	5′-CAATCCAGTCGGCCTAACA-3′
qV18F	5′-CCACGAACCTAAGATACAT-3′
qV19F	5′-CTCGAGAGAAGAAGTCATCT-3′
qV20F	5′-CAGTCATCCCAACTTATCCT-3′
qV21F	5′-GCTAAGAAACCATGTACCAT-3′
qV22F	5′-CAGTTCCTCTGAGGCTGGA-3′

qV23F	5′-CTGTGTGCCCCTCCAGCTCA-3′
qV24F	5′-CTCAGCTAAGTGTTCCTCGA-3′
qV25F	5′-CTATGTGGCATATTACTGGT-3′
qV26F	5′-CCTTCAAACTCACCTTGCAGC-3′
qV27F	5′-CATTGTTCATATGGCATT-3′
qV28F	5′-CTCTGATAGATATATCAT-3′
qV29F	5′-CTGATTCTGGATTCTGCTA-3′
qV30F	5′-CAATGCAAGGCCTGGAGACA-3′
qV31F	5′-AAATCAAGCCCTAACCTCTAC-3′
qJβ1.1R	5′-CTCGAATATGGACACGGAGGACATGC-3′
qJβ2.1R	5′-CCTGATACAGGGCCTTGGATAGTTA-3′

Table T3: Primers and probes for 3C assay

3C anchor primers and Taqman probes (5′FAM and 3′ TAMRA from Sigma Life Sciences)	Sequences
Dβ_1 HindIII probe	5′-AAGGCATTGTTGCATGATCCT-3′
Dβ_2 HindIII probe	5′-AAATGCTGGGCCTCTGTAGA-3′
Eβ_ HindIII probe	5′-CATAAGCATTGTCATGTTTGTGACA-3′
ERCC3 HindIII probe	5′-AAAGCTTGCACCCTGCTTTAGTGGCC-3′
Dβ_1 HindIII primer	5′-TGAAATTTTTCTGCCGAAAGGAC-3′
Dβ_2 HindIII primer	5′-GCGGGATCCAAGAGAACTCA-3′
Eβ_ HindIII primer	5′-GAAAATTGGCATCGGTTTGC-3′
HindIII primers	Sequences
V1	5′-TATCTCTGTGGGGCATGCAG-3′
V2	5′-TTTCATTCACAGCCGACCAG-3′
V3	5′-TTTCATTCACAGCCGACCAG-3′
V4	5′-AGCTCGACACAGAAAGCAAGTT-3′
V5	5′-AGCTCGACACAGAAAGCAAGTT-3′
V6	5′-GGTTCCCTTCACTTCCCACA-3′
V7	5′-GTCCGCTAGCAGCCAGAGTT-3′
V8	5′-GTCCGCTAGCAGCCAGAGTT-3′
V9	5′-ACCAGAGGGCAGCTGAAAAT-3′
V10	5′-GTGCCTGTACCATGCTGTGG-3′
V11	5′-TTCAGCAAGTAGGTGCGAAGA-3′
V12-1	5′-TGGTGGGATCCTGACAGCTTATA-3′
V13-1	5′-CCATCTGCATGAACACCTTCTT-3′
V12-2	5′-CCATCTGCATGAACACCTTCTT-3′

V13-2	5′-CCATCTGCATGAACACCTTCTT-3′
V12-3	5′-GGATCTTGGTCTCGGGAGGT-3′
V13-3	5′-CTCAGCTGCACCCTCACAAC-3′
V14	5′-CAGGCTTTTGAGTGGCATGT-3′
V15	5′-AGGCAGGAGGTGAGTCTTGG-3′
V16	5′-TATCATGCCCAGCTGCATTC-3′
V17	5′-GTTAGGCCGACTGGATTGGA-3′
V18	5′-GGCAGTGTTACAGAACCCAGTG-3′
V19	5′-GGCAGTGTTACAGAACCCAGTG-3′
V20	5′-TGTGATGGGTTGTCATCTGGA-3′
V22	5′-CCAAGGGATGATGTCACAGG-3′
V23	5′-TACACCGGCCAGGAGAGACT-3′
V24	5′-ACTAGGCCAGCAGAGGATGC-3′
V25	5′-ACTAGGCCAGCAGAGGATGC-3′
V26	5′-AGCATAGGATTGGGCCTCAG-3′
V27	5′-CATCACTGCGCCTAGCAATC-3′
V28	5′-GCGTGTGCCACGTTTTTGTA-3′
V29	5′-CTCTAGCAATCCCCCTGTGC-3′
V31	5′-AAGGAGAGAGCAGGCCACAG-3′
Dβ_1	5′-AAGGCATTGTTGCATGATCC-3′
Dβ_2	5′-TGGGGCCCTCACTTTTCTTA-3′
Eβ	5′-TCCTAAGGAGAGGCAGAGTGG-3′
ERCC3	5′-GACTTCTCACCTGGGCCTACA-3′

Table T4: Luciferase assay cloning primers

Primer name	Sequences
Eβ-F	5′-ATTGGATCCGTTAACCAGGCACAGTAGGACC-3′
Eβ-R	5′-ATTGGATCCCCATGGTGCATACTGAAGGCTTC-3′
Pro-V1F	5′-TAGCCTCGAGGAGTGACTAGTTACTTCTGC-3′
Pro-V1R	5′-TAGCGAAGCTTCTCTGAGACCTCAGGTTCTC-3′
Pro-V3-F	5′-TATCTCGAGGGGACTCAGTTCAGTAGTC-3′
Pro-V3-R	5′-CGAGAAGCTTAGTAGGGTCACGGCAGGAA-3′
Pro-V4F	5′-TAGCCTCGAGTGTGCTAAGGGCACCAATGAAT-3′
Pro-V4R	5′-TAGCGAAGCTTGTTGGGTCAAGGCAGGGCAAAT-3′
Pro-V5-FX	5′-TAGCCTCGAGTATCCATTGTATGCTCTGTTTG-3′
Pro-V5-RH	5′-TAGCGAAGCTTGGTGGAATCAGGCTCCAGACG-3′
Pro-V6-FX	5′-TAGCCTCGAGCTACAAGCTCCCAAGAGAGAG-3′
Pro-V6-RH	5′-TAGCGAAGCTTCTCTGGAGAAGACAGAGGAC-3′

Pro-V7F	5′-TAGCCTCGAGGCTGCTGAATAGCAAGTTTCCAG-3′
Pro-V7R	5′-TAGCGAAGCTTTTGGAGGTTTGGATCTGTAGTCT-3′
Pro-V9-FX	5′-TAGCCTCGAGGGAACTTTCATGTGAGGAGA-3′
Pro-V9-RH	5′-TAGCGAAGCTTCTGCAAAAATATAAGTTGTGAACAG-3′
Pro-V10-FX	5′-TAGCCTCGAGGGGATATCTCTATGCTTTAATG-3′
Pro-V10-RH	5′-TAGCGAAGCTTCTGGAGAAGGAGGCATAAGGA-3′
Pro-V11F	5′-TAGCCTCGAGTTCCCTACAGTGTCAAGGGCTG-3′
Pro-V11R	5′-TAGCGAAGCTTTGTACCCACAGGGTTGTTCTCA-3′
Pro-V12-2-F	5′-TAGCCTCGAGCAACTGACTCAGAGAAAAAC-3′
Pro-V12-2-F	5′-TAGCGAAGCTTTCCTCTCAGGATACTGGTCTCT-3′
Pro-V14F	5′-TACATCGCTAGCCATTTATGTGTACCATAATAAT-3′
Pro-V14R	5′-TAGCCTCGAGGGCAGATTGAGGGCAGAGGAG-3′
Pro-V16F	5′-TAGCCTCGAGTTGCAATCTACCTCTGCTGCTC-3′
Pro-V16R	5′-TAGCGAAGCTTTTGTGATGACACCACTGTCTCCG-3′
Pro-V17-FX	5′-TAGCCTCGAGGCAGGTGTGACCTACGATAAC-3′
Pro-V17-RH	5′-TAGCGAAGCTTGGATGGTCCAGAACAGGAAA-3′
Pro-V19-F	5′-TATCTCGAGCATTTGAGAAAGACAACAA-3′
Pro-V19-R	5′-CGAGAAGCTTAGTTTGGAGGGACTTTCTT-3′
Pro-V20F	5′-TAGCCTCGAGGATAAGGTAACTGAAGCGGGA-3′
Pro-V20R	5′-TAGCGAAGCTTCTTCAGTGTTGACTTCACACC-3′
Pro-V22F	5′-TAGCCTCGAGGATGAAATATGGTAACAAGG-3′
Pro-V22R	5′-TAGCGAAGCTTAGGAGATAAAGGGCTACATA-3′
Pro-V24F	5′-TACATCGCTAGCCCAATGATATGTGCAGAGATGA-3′
Pro-V24R	5′-TAGCCTCGAGGATCACACTAGGCCAGCAGAG-3′
Pro-V25F	5′-TAGCCTCGAGCAATTGGGCCATCTTCTGCCAC-3′
Pro-V25R	5′-TAGCGAAGCTTCAGGTGGATACTTCATTCC-3′
Pro-V28F	5′-TAGCCTCGAGAGTTGTCTTGTGGGCAACTCTG-3′
Pro-V28R	5′-TAGCGAGATCTGCTAGATAGCCTCAAGGCTGCAAA-3′

Table T5: Recombination substrate oligos

Primer name	Sequences
RS V1F	TAGCCTCGAGATACGGAGCTGAGGCTGCAAG
RS V1R	TACATCGCGGCCGCAGTCACCTTATAACTCATGCA
RS V15F	TAGCCTCGAGCCTTCTCCACTCTGAAGATTC
RS V15R	TACATCGCGGCCGCTTCCACCCAAGATTTCTTAA
RS V16F	TAGCCTCGAGACTCAACTCTGAAGATCCAGA
RS V16R	TACATCGCGGCCGCTAATGTAATACTCGTTACCAT
RS V18F	TAGCCTCGAGCCCAACATCCTAAAGTGGG

RS V18R	TACATCGCGGCCGCTTCCTCCGTAAGCATGGTG
RS V20F	TAGCCTCGAGCAGTCATCCCAACTTATCCT
RS V20R	TACATCGCGGCCGCCTCCTGGGTACCCTCCCATTTC
RS V23F	TAGCCTCGAGCACTCTGCAGCCTGGGAATC
RS V23R	TACATCGCGGCCGCTGACTTGGTCTGGGTGTGCTG
RS V24F	TAGCCTCGAGAGTGCATCCTGGAAATCCTAT
RS V24R	TACATCGCGGCCGCAGACCTGGCCTGTTTCTCATG
RS V26F	TAGCCTCGAGCAAGAAGTTCTTCAGCAAATA
RS V26R	TACATCGCGGCCGCGATACAGGTTTCAGTTAGTT

Table T6: Computational analysis coefficients for determinants of Vβ frequencies (all *Tcrb* V gene segments): Classifer step, three features

Features	Estimate	SE	t	Pr(>jtj)
Intercept	1.09059	1.52205	0.717	0.47903
Recombination signal information content (RIC) score	0.08803	0.02619	3.362	0.00207
Formaldehyde-assisted isolation of regulatory elements (FAIRE)	0.03185	0.01639	1.944	0.06105
RNA Pol II	0.65913	0.26654	2.473	0.01909

Table T7: Computational analysis coefficients for determinants of Vβ frequencies (all *Tcrb* V gene segments): Combinatorial analysis of 13 features and their correlation to recombination frequency.

Number of features	Pearson correlation coefficient	P value
13	0.77954	0.4707
8	0.74191	0.1015
7	0.72604	0.07434
6	0.71277	0.04925
5	0.68818	0.03779
4	0.66405	0.0265
3	0.64982	0.01359
2	0.60304	0.01089
1	0.53998	0.00782

Table T8: Coefficients for determinants of Vβ frequencies (rearranging Vβ segments)

Features	Estimate	SE	t	Pr(>jtj)
All *Tcrb* V gene segments (regressor step, 13 features)				
Intercept	0.08707	5.81E+00	0.015	0.9882
RIC	0.08817	3.72E−02	2.373	0.0273
3C score	−2.17745	3.14E+00	−0.693	0.4961
Transcription	−0.1299	1.56E−01	−0.83	0.4156
CCCTC-binding factor (CTCF)	0.9394	1.24E+00	0.756	0.4579
FAIRE	0.01538	2.46E−02	0.625	0.5384
H3ac	−0.31847	5.05E−01	−0.63	0.5353
H3K27ac	0.03124	3.86E−02	0.81	0.4271
H3K4me1	0.16488	6.88E−01	0.24	0.813
H3K4me2	0.0194	1.67E−02	1.159	0.2595
H3K4me3	−0.08483	3.68E−01	−0.231	0.8197
H3K9me2	0.74873	1.27E+00	0.59	0.5618
P300	−0.03168	3.03E−02	−1.047	0.3069
RNA PolII	1.10351	5.21E−01	2.119	0.0462
All *Tcrb* V gene segments (Regressor step, 5 features)				
Intercept	0.62866	0.35047	1.794	0.0907
Transcription	−0.0613	0.0467	−1.313	0.2066
CTCF	0.31418	0.25634	1.226	0.2371
H3K4me2	0.00779	0.00365	2.137	0.0475
H3K4me3	0.16139	0.0666	2.423	0.0268
P300	−0.0066	0.00646	−1.027	0.319

Chapter 3: Lineage-Specific Compaction of *Tcrb* Requires a Chromatin Barrier to Protect the Function of a Long-range Tethering Element

This paper has been published in the Journal of Experimental Medicine:

Majumder, K., Koues, O. I., Chan, E. A., Kyle, K. E., Horowitz, J. E., Yang-Iott, K., ... & Oltz, E. M. (2015). Lineage-specific compaction of Tcrb requires a chromatin barrier to protect the function of a long-range tethering element. *The Journal of Experimental Medicine*, *212*(1), 107-120.

3.1 Abstract

Gene regulation relies on dynamic changes in three-dimensional chromatin conformation, which are shaped by composite regulatory and architectural elements. However, mechanisms that govern such conformational switches within chromosomal domains remain unknown. We identify a novel mechanism by which cis-elements promote long-range interactions, inducing conformational changes critical for diversification of the TCRβ antigen receptor locus (*Tcrb*). Association between distal Vβ gene segments and the highly expressed DβJβ clusters, termed the recombination center (RC), is independent of enhancer function and recruitment of V(D)J recombinase. Instead, we find that tissue-specific folding of *Tcrb* relies on two distinct architectural elements located upstream of the RC. The first, a CTCF-containing element, directly tethers distal portions of the Vβ array to the RC. The second element is a chromatin barrier that protects the tether from hyperactive RC chromatin. When the second element is removed, active RC chromatin spreads upstream, forcing the tether to serve as a new barrier. Acquisition of barrier function by the CTCF element disrupts contacts between distal Vβ gene segments and significantly alters *Tcrb* repertoires. Our findings reveal a separation of function for RC-flanking regions, in which anchors for long-range recombination must be cordoned off from hyperactive RC landscapes by chromatin barriers.

3.2 Introduction

The packaging of mammalian genomes into chromatin and its folding into discrete topological domains can be altered dynamically to regulate gene expression. In many cases, these processes are linked mechanistically. For example, conversion of repressive to active chromatin is usually preceded by changes in locus topology that facilitate long-range contacts between gene promoters and their regulatory elements, including transcriptional enhancers (de Laat et al., 2013; Sanyal et al., 2012). Deciphering the regulatory logic that sets active and inactive conformations within a genomic space to control expression of its composite genes remains an important goal.

In this regard, antigen receptor (AgR) loci serve as models to study the relationships between regulatory elements and developmental alterations of chromatin, three-dimensional conformation, and gene activity (Cobb et al., 2006; Jackson et al., 2006; Jhunjhunwala et al., 2008; Steinel et al., 2010). In precursor lymphocytes, specific regions within AgR loci are activated then repressed at distinct stages of development (Osipovich et al., 2010). Dynamic changes in chromatin and locus topology direct the ordered assembly of immunoglobulin (*Ig*) and T cell receptor (*Tcr*) genes from large arrays of Variable (V), Diversity (D) and Joining (J) segments. Although each step in the assembly process is executed by a common enzymatic machinery, composed of the RAG1 and RAG2 proteins, recombination is initiated only within regions of AgR loci marked by accessible chromatin (Cobb et al., 2006). Moreover, recombination between distant gene segments requires their spatial apposition via locus contraction (Kosak et al., 2002; Skok et al., 2007).

The general architecture of AgR loci and the mechanisms employed to control their assembly share many similarities (Shih and Krangel, 2013). As an example, thymocytes first

activate an enhancer, termed Eβ, situated at the 3' terminus of the 700 kb *Tcrb* locus (Bories et al., 1996; Bouvier et al., 1996). Once activated, Eβ interacts with promoters flanking two clusters of DβJβ gene segments, forming stable loops and triggering transcription of the unrearranged segments (Oestreich et al., 2006). The germline transcription is accompanied by covalent modification and opening of chromatin, which attracts RAG-1/2 binding and mediates Dβ to Jβ recombination (Ji, Resch et al., 2010). Indeed, robust germline transcription at (D)J clusters is an initial activation event at all AgR loci, which generates a focal zone of RAG binding, termed the recombination center (RC) (Schatz and Ji, 2011). At *Tcrb*, DβJβ joins serve as substrates for long-range recombination with an array of 30 *Trbv* segments that are separated from the RC by 250 to 500 kb. Analogous to other AgR loci, long-range *Tcrb* recombination requires lineage-specific changes in locus topology. Upon commitment to the T cell lineage, the entire locus contracts, bringing distal *Trbv* segments into spatial proximity with the RC (Skok et al., 2007). In addition to the global "contraction", which brings the locus ends together, the *Trbv* cluster itself adopts a more densely packed configuration in thymocytes. This more compact configuration likely facilitates efficient sampling of Vβ gene segments by the RC following locus contraction, ensuring a diverse *Trbv* repertoire.

Recent studies have begun to reveal the *cis*-elements and *trans*-acting factors that underlie some topological changes at AgR loci. A common theme is the involvement of CTCF and the cohesin complex, which together play a major role in sculpting the three-dimensional (3D) architecture of eukaryotic genomes (Phillips et al., 2009). CTCF binds directly to DNA at thousands of genomic sites, which can interact through space via CTCF-CTCF dimerization. These contacts are stabilized by CTCF-mediated recruitment of cohesin, which forms a collar around the

base of resultant chromatin loops (Nasmyth et al., 2009). In developing lymphocytes, ablation of CTCF or RAD21, a critical cohesin subunit, impairs promoter-enhancer interactions and perturbs the repertoire of distant V segments used in long-range V(D)J recombination (Ribeiro de Almeida et al., 2011; Seitan et al., 2011). In addition to its structural role, CTCF regulates AgR assembly via its insulator function, forming boundaries between active and repressive chromatin domains. At both *Igh* and *Igk*, CTCF-bound insulators prevent the spread of active chromatin from the RC to the most proximal V gene segments (Guo, Yoon et al., 2011; Xiang et al., 2013). Inactivating mutation of these elements augments germline transcription and recombination of the most RC-proximal V segments, presumably by extending the reach of powerful enhancers situated in the RC. In what may be a related finding, all of these CTCF-binding elements associate by contact with the collection of enhancers that decorate the 3' end of the *Igh*- and *Igk*-RCs (Guo, Yoon et al., 2011; Xiang et al., 2013).

Although the roles of CTCF-bound regions in AgR locus conformation are emerging, the requirements for transcriptional regulatory elements in these lineage-specific processes remain murky. Conflicting data exist for whether *Ig* and *Tcr* enhancers are required for contraction of their corresponding loci (Guo, Gerasimova et al., 2011; Medvedovic et al., 2013; Shih et al., 2012). With regard to the more intricate aspects of AgR locus topology, enhancer deletions consistently disrupt their associations with distal promoters and other enhancers (Shih et al., 2013). However, existing data derive from the perspective of regulatory elements rather than monitoring specific interactions between V and (D)J clusters. We now probe multiple perspectives to determine how promoters and enhancers within the *Tcrb*-RC shape its active, lineage-specific conformation. In thymocytes, we find that the large *Trbv* array is juxtaposed with the RC independent of enhancer

function, RAG binding, and germline transcription. Instead, the active *Tcrb* conformation depends on an RC-flanking region, which harbors a chromatin barrier function, but is not the major contact point for *Trbv* segments. Loss of the RC–proximal region activates a nearby CTCF-binding site to become a new chromatin barrier, disarming it as the major contact point for distal *Trbv* segments. Our findings indicate a separation of function for RC-flanking regions, which require that long-range contact points be insulated from the hyperactive landscape of the RC.

3.3 Materials and Methods

Mouse strains. ΔPDβ1, ΔEβ and ΔminPDβ1 mice were maintained on a Rag1$^{-/-}$/C57BL/6 background (Bories et al., 1996; Whitehurst et al., 2000). DP thymocytes were generated in *Rag1*$^{-/-}$ mice by anti-CD3ε injections as described (Shinkai et al., 1994). The mEβ mouse, which harbors crippling mutations at both Runx binding sites in Eβ, was generated by homologous recombination in ES cells and will be described elsewhere (I.T., manuscript in preparation). The endogenous Runx binding sequences TGTGGTT and TGCCACA were mutated to TGTCCAT and TTGGACA, respectively. The mEβ allele was backcrossed onto the Rag1$^{-/-}$/C57BL/6 background. D708A mice were obtained from the Schatz lab (Ji et al., 2010). Rag1$^{-/-}$/C57BL/6 mice were used as positive control for 3C, ChIP and germline transcription assays, and are labelled as "WT" in the figures. Developmental stages in RAG-deficient thymocytes harboring different *Tcrb* genotypes were assessed by CD44:CD25 staining. The majority (>94%) of cells were DN3 in each of the genotypes, as expected (Yannoutsos et al., 2001).

Chromosome Conformation Capture assays. 3C assays were performed and analyzed as described (Gopalakrishnan et al., 2013; Hagège et al., 2007). Refer to Supplemental Tables T9 and T10 for primer and probe combinations.

Chromatin Immunoprecipitation. ChIP assays were performed as described (Gopalakrishnan et al., 2013). The following antibodies were used: CTCF (Rockland, 600-401-C42), Rad21 (Abcam, ab992), H3ac (Millipore, 06-599), H3K4me2 (Abcam, ab32356), H3K4me3 (Abcam, ab8580), H3K9me2 (Abcam, ab1220), H3K27me3 (Abcam, ab6147) and IgG (Santa Cruz, sc-2027). ChIPs

were analyzed by qPCR using SYBR Green and primer combinations shown in Supplemental Table T11. The LTR region between *Prss2* and *Tcrb*-RC was assayed with primers 7.4 UDB and 5.5 UDB published (Carabana et al., 2011).

3D-FISH. Hybridizations were performed with bacterial artificial chromosomes (BACs) that recognize the *Trbv1* (RP23-75P5), *Trbv2-Trbv12* (RP23-306O13), Trypsinogen region (RP23-203H5) and the *Tcrb*-RC (RP23-421M9). To generate probes, BACs were nick translated with Biotin and Digoxigenin using Roche kits (11-745-824-910 and 11-745-8816-910). The FISH probes were hybridized to slides of fixed, permeabilized thymocytes, then incubated with anti-biotin (Jackson, 200-162-211), anti-digoxigenin (200-602-156) and DAPI (Invitrogen) stains. Hybridized slides were imaged on a Nikon A1 confocal microscope and analyzed using ImageJ to measure 3D distances between foci as described (Shih and Krangel, 2010).

Germline *Tcrb* Transcription. cDNA generated from 2µg total thymocyte or pro-B cell RNA (iScript Supermix, Biorad) was analyzed by qPCR using the primer combinations provided in Supplemental Table T11.

Recombination Assays. Genomic DNA was extracted from 10^6 total thymocytes using the Qiagen DNeasy Blood and Tissue Kit (69504). Taqman qPCR assays to measure Jβ2 rearrangement frequencies were performed as described (Gopalakrishnan et al., 2013).

3.4 Results

RC activation is dispensable for its long-range interactions with Trbv

The molecular determinants for spatial apposition of distal *Trbv* segments with their DβJβ targets remain unknown. A key RC feature is its robust, Eβ-dependent transcriptional activity, which decorates the DβJβ clusters with H3K4me3 and RAG-1/2 (Ji, Little, et al., 2010; Ji, Resch, et al., 2010). As proposed by others, this molecular landscape may be a prerequisite for capturing distant *Trbv* segments into a transcription factory occupied by the highly expressed RC, forming long-range *Tcrb* loops (Verma-gaur et al., 2012). Accordingly, inactivation of the RC should exclude it from transcription factories and disrupt long-range V-DJ interactions. Prior studies at *Igh* and *Igk* suggest that distant V-RC interactions are enhancer-independent (Hewitt et al., 2008; Medvedovic et al., 2013), but these conclusions are complicated by residual RC transcription and potential redundancies between multiple enhancers. In contrast, deletion of Eα cripples transcription of the *Tcra*-RC and perturbs its interactions with proximal *Trav* segments (Shih et al., 2012). As such, the validity of the transcription factory co-occupancy model remains unresolved.

Mouse *Tcrb* harbors a single known enhancer that is essential for transcription and recombination of its RC in DN thymocytes (Bories et al., 1996; Bouvier et al., 1996). When transcriptionally active, the *Tcrb*-RC samples Vβ segments by adopting a thymocyte-specific conformation, in which these distal elements are brought into spatial proximity (Gopalakrishnan et al., 2013). To directly test causal relationships between RC activation and *Trbv*-DβJβ associations, we measured their spatial proximity in DN thymocytes containing transcriptionally active or inactive versions of DβJβ clusters. Thymocytes with a transcriptionally inactive RC

derive from mice in which two critical Runx binding sites in Eβ were destroyed by targeted mutagenesis (Figure 3.1A, mEβ). The mutant Eβ maintains linear spacing within the RC, but recapitulates all aspects of *Tcrb* inactivation observed with a complete Eβ deletion, termed ΔEβ (Mathieu et al., 2000a). The defects resulting from enhancer inactivation include, ablation of germline DβJβ transcription (Figure 3.1B), diminished levels of H3K4me3 deposition (Figure 3.1C), and loss of looping between the enhancer region and both Dβ-associated promoters (Figure 3.1D). Unless indicated otherwise, DN thymocytes for all experiments were from mice bred into a RAG1-deficient background (C57BL/6) to preclude *Tcrb* rearrangements, which would confound interpretation of looping data.

We measured *Trbv*-RC association in wild-type (WT) versus mEβ alleles using chromosome conformation capture (3C), which quantifies crosslinking efficiency of a given genomic viewpoint with other restriction fragments (Dekker et al., 2002). As shown in Figure 3.1E, the Eβ region associates more efficiently with *Trbv* segments in DN thymocytes compared with pro-B cells, confirming its cell type-specific interactome (Gopalakrishnan et al., 2013). No significant differences are observed for long-range *Tcrb* interactions in DN thymocytes from RAG1-deficient mice compared with those expressing a catalytically inactive, but binding-competent version of RAG1 (D708A, Figure 3.1F) (Ji, Resch, et al., 2010). Thus, DN-specific looping between the Vβ cluster and *Tcrb*-RC is independent of RAG1 binding.

Interactions between the enhancer region and *Trbv* segments are mostly diminished in DN thymocytes from mEβ animals (Figure 3.1E). However, the inactive Eβ maintains a subset of contacts with the central *Trbv12-Trbv16* cluster (see Discussion). Surprisingly, associations between *Trbv* segments and both DβJβ clusters within the RC are unaffected by deletional or

mutational inactivation of the enhancer when monitored from either Dβ viewpoints (Figures 3.1G, H). Thus, when Eβ is functional, it interacts with RC promoters and incorporates into the *Trbv*-DJ interactome; but when this enhancer is disabled, it separates from the thymocyte-specific aggregation of VDJ gene segments. We conclude that *Tcrb* adopts a thymocyte-specific conformation, which facilitates long-range *Trbv*-DJ interactions, independent of Eβ function, RC transcription, and RAG deposition. Importantly, these findings formally preclude the transcription factory co-occupation model for *Tcrb* looping.

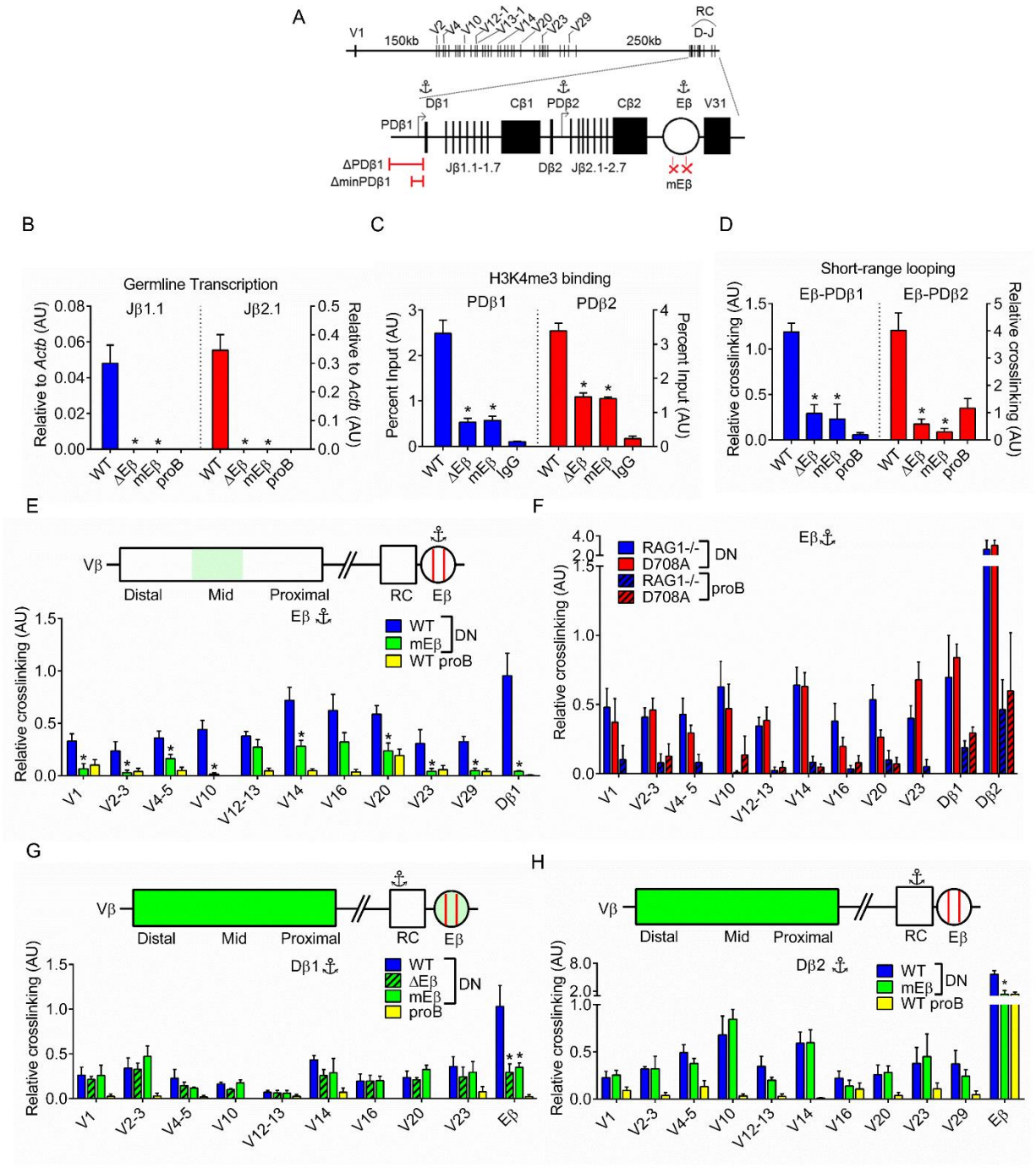


Figure 3.1: Long-range *Trbv*–RC interactions are Eβ-independent. (A) Schematic depiction of the entire mouse *Tcrb* locus (upper) and a magnified version of 30 kb spanning the RC (lower). Promoter deletions (ΔPDβ1 and ΔminPDβ1) and enhancer mutations (mEβ) are shown at the bottom. Viewpoints used in 3C assays are designated as anchor symbols. (B) Germline transcription was measured relative to *Actb* in RAG-deficient thymocytes (WT, mEβ or ΔEβ alleles) and pro-B cells as described (Osipovich et. al, 2007). (C) H3K4me3 deposition was measured by ChIP

at PDβ1 and PDβ2 in RAG-deficient thymocytes (WT, mEβ or ΔEβ alleles). ChIP using a non-specific isotype control is shown (IgG). (D) 3C analysis was performed to test the crosslinking between Eβ and Dβ1 (left) or Dβ2 (right) in RAG-deficient thymocytes (WT, mEβ or ΔEβ alleles) and pro-B cells (background levels). (E) Long-range interactions were tested by 3C using the Eβ viewpoint (anchor symbol). Relative crosslinking between Hind III fragments spanning Eβ and each indicated gene segment was calculated as described (Gopalakrishnan et al., 2013). The data are summarized as a cartoon in the top panel. Green shading indicates whether crosslinking in mEβ relative to WT alleles is unchanged (darkest green), reduced significantly (lighter green), or reduced to background levels in pro-B cells (white). (F) 3C assays were performed with the Eβ viewpoint in DN thymocytes and pro-B cells from RAG-deficient mice, which either lack or express a D708A RAG transgene (Ji et al., 2010). (G) 3C assays were performed with the Dβ1 viewpoint in DN thymocytes (WT, Eβ or ΔEβ alleles) and pro-B cell controls. Results are summarized in the schematic on top as described in Figure 1E. (H) 3C interactions were monitored using the Dβ2 viewpoint. Data are presented as mean values from at least three independent experiments (± SEM). Significant differences between WT and mEβ samples are denoted as * $p < 0.05$ (t-test).

Eβ function in Trbv topology and transcription

By comparison with cells from other lineages, the *Trbv* cluster adopts a more compact conformation in DN thymocytes (Skok et al., 2007), which likely facilitates sampling of *Trbv* segments by the RC and diversifies their usage in the primary TCRβ repertoire. However, recombination of *Trbv* segments is not completely normalized; instead, it is influenced significantly by relative levels of Vβ germline transcription (Gopalakrishnan et al., 2013). Thus, the primary TCRβ repertoire is determined by both topological and transcriptional properties of the *Trbv* cluster.

To assess whether Eβ is required for these repertoire-sculpting features, we measured intra-Vβ association using 3C. When examined from viewpoints in either the distal (*Trbv5*) or proximal (*Trbv23*) portion of the cluster, intra-*Trbv* crosslinking is unaffected by the mEβ mutation (Figures 3.2A, B). However, in keeping with data presented in Figure 3.1, long-range association of *Trbv5* and *Trbv23* with the enhancer is reduced. In mEβ thymocytes, both CTCF and RAD21 remain bound to sites within *Tcrb* at levels well above background; their binding differed statistically at only one tested site in the *Trbv* cluster, *Trbv10*, where CTCF decreased modestly (Figures 3.2C and D). However, inactivation of Eβ diminishes transcription at a subset of *Trbv* segments that are most highly expressed in DN thymocytes (Figure 3.2E). Attenuated expression of these germline segments may reflect either a requirement for association with a transcriptionally active RC or with the functional Eβ element (see discussion). We conclude that Eβ is dispensable for compaction of the *Trbv* cluster, but augments the transcriptional activity of specific Vβ segments, which could influence the primary *Tcrb* repertoire. A definitive test is precluded because Eβ is essential for DβJβ recombination, a prerequisite for subsequent rearrangement of *Trbv* segments.

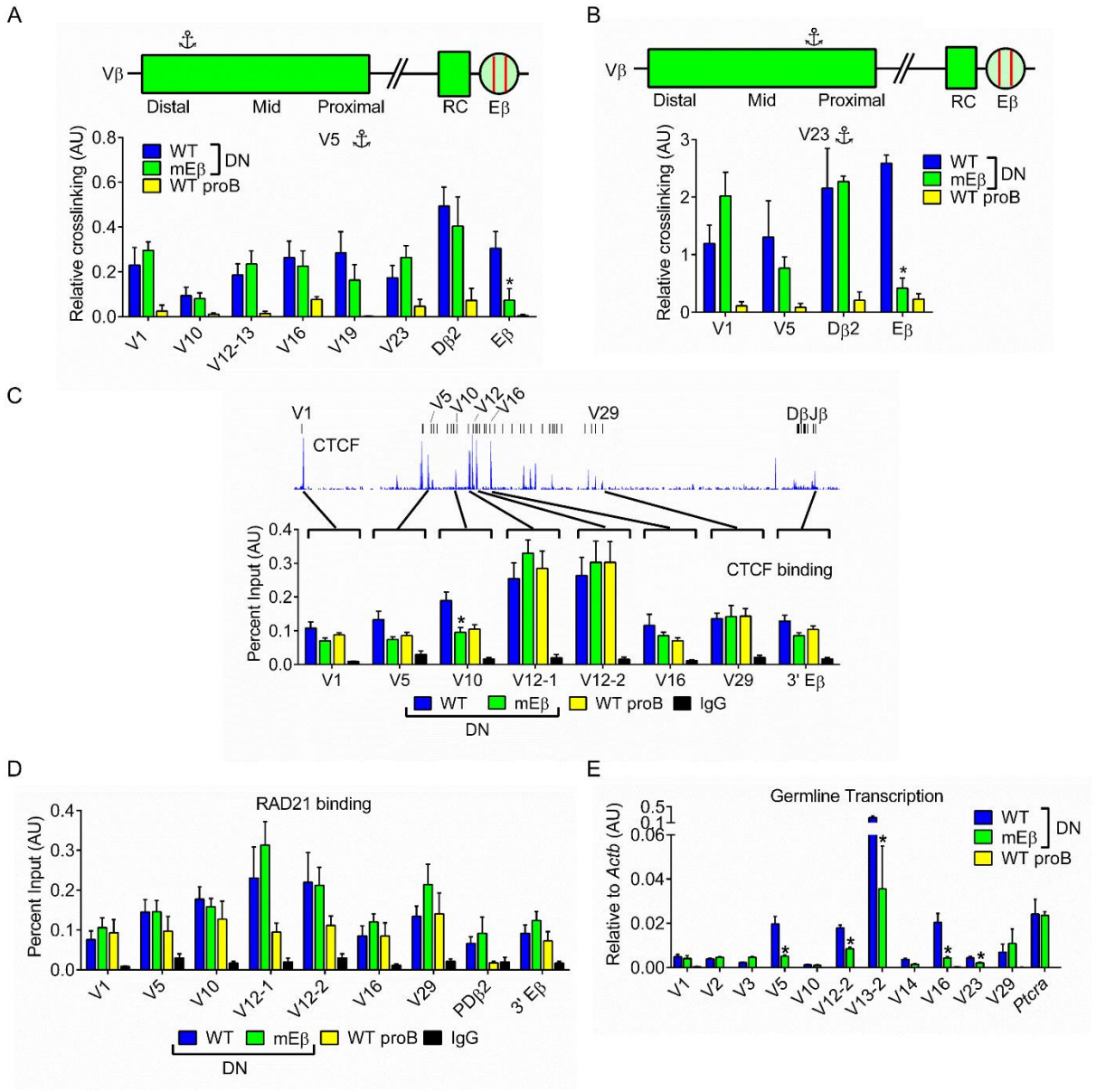


Figure 3.2: Impact of Eβ on topology, structural protein deposition, and transcription of Vβ segments. 3C data for the (A) *Trbv5* and (B) *Trbv23* viewpoints in RAG-deficient DN thymocytes or pro-B cells (see Figure 1A for details). (C) Published ChIP-seq profile for CTCF in RAG-deficient DN thymocytes (upper panel, (Shih et al, 2012)). ChIP-qPCR for (C) CTCF and (D) RAD21 binding at the indicated sites in WT or mEβ thymocytes versus RAG-deficient pro-B cells. Data are presented as mean values for percent input signal from at least three independent experiments (± SEM). (E) Germline transcription of *Trbv* segments as monitored by RT-qPCR assays in the indicated cell types. Mean values from three independent experiments after normalization to signals for β-actin are shown (± SEM). Significant differences between WT and mEβ samples are denoted as * $p<0.05$ (t-test).

RC promoter deletion reveals two Trbv interaction domains

In addition to Eβ, transcription and rearrangement of the RC is controlled by two promoters, termed PDβ1 and PDβ2, situated within their respective DβJβ clusters (Figure 3.1A) (Sikes et al., 1998; Sikes et al., 2002). Activation of the Dβ1Jβ, but not Dβ2Jβ, cluster is crippled in thymocytes harboring a 3.5 kb deletion spanning PDβ1 (ΔPDβ1 allele, Figure 3.1A) (Whitehurst et al., 1999). To test whether activities associated with the promoter region contribute to folding of *Tcrb* into its active conformation, we performed 3C analyses on DN thymocytes from ΔPDβ1/*Rag1*$^{-/-}$ mice. Because ΔPDβ1 removes one relevant restriction site near Dβ1, we focused RC interactome studies on Dβ2 and Eβ. As shown in Figure 3.3A (upper panel), Dβ2 interactions with the most proximal portion of the *Trbv* cluster are unaffected by the ΔPDβ1 mutation (*Trbv16-30*). However, we observe a significant reduction in Dβ2 crosslinking with distal portions of the *Trbv* array (*Trbv1-14*). Precisely the same bifurcation in long-range interactions is observed when Eβ is used as the 3C viewpoint (Figure 3.3B). The ΔPDβ1 mutation also reduced CTCF levels at sites in the distal *Trbv* array (Figure 3.3D), which may be a consequence of disrupting their association with CTCF-rich elements near the RC (see discussion). However, RAD21 binding and germline *Trbv* transcription throughout *Tcrb* are unaffected in ΔPDβ1 thymocytes (Figure 3.3E and F).

To gain more insight into its putative bidomainal structure, we probed interactomes of the *Trbv* array using a distal and a proximal Vβ segment as viewpoints. The distal *Trbv5* segment exhibits tissue-specific, enhancer-independent association with other gene segments in the *Trbv* array, as well as a robust interaction with the RC (Figure 3.3A, lower panel). Crosslinking of this region with other distal Vβ segments is unaffected by the ΔPDβ1 deletion. However, its

associations with the proximal half of *Trbv* and with the RC are significantly diminished in ΔPDβ1 thymocytes. Thus, the more distal *Trbv* segments form a higher-order structure independent of PDβ1, but require this promoter region for its interaction with the 3' half of the *Trbv* cluster. Conversely, the more proximal *Trbv23* region associates with the RC and another 3' segment, *Trbv29*, independent of PDβ1, but requires this promoter region for its association with more distal *Trbv* segments (Figure 3.3C).

A primary function of the region deleted from ΔPDβ1 alleles is promoter activity, which drives transcription and remodels the Dβ1Jβ chromatin landscape (Whitehurst et al., 1999). To explore whether promoter function is the primary determinant of long-range interactions between distal *Trbv* segments and the RC, we revived a mouse strain that harbors a deletion spanning only the minimal promoter upstream of Dβ1 (ΔminPDβ1) (Whitehurst et al., 2000). Only residual levels of germline Dβ1 transcription are detected in thymocytes from ΔminPDβ1/*Rag1*$^{-/-}$ mice (Figure 3.4A, (Whitehurst et al., 2000)). Despite this dramatic transcriptional defect, long-range *Trbv*-RC interactions are unaffected by the ΔminPDβ1 deletion (Figures 3.4B and C).

Together, these data indicate that the *Trbv* array is topologically divided into two domains. The more proximal half of *Trbv*, which still lies >250 kb upstream of the DβJβ clusters, associates with the RC in thymocytes via mechanisms that are independent of PDβ1 and Eβ. The distal half of *Trbv* forms tissue-specific contacts with both the RC and the proximal *Trbv* domain. Although these interactions are independent of PDβ1 promoter activity, they require a 3 kb region upstream of this minimal control element. Importantly, we find that the most distal *Trbv* segments are significantly underutilized in Vβ-Dβ2Jβ rearrangements when comparing ΔPDβ1 with ΔminPDβ1 thymocytes on RAG-sufficient backgrounds (Figure 3.4D). The underutilization of distal *Trbv*

gene segments in *Tcrb* repertoire is observed when assaying rearrangements in whole-thymocyte populations as well as in sorted $CD44^{-}CD25^{+}$ (DN3) subsets (Figure 3.5). In contrast, *Trbv* segments in the proximal domain are utilized at comparable or higher frequencies in ΔminPDβ1 thymocytes. Thus, mechanisms that ensure tethering of distal *Trbv* domains are important for generating maximal diversity in the TCRβ repertoire.

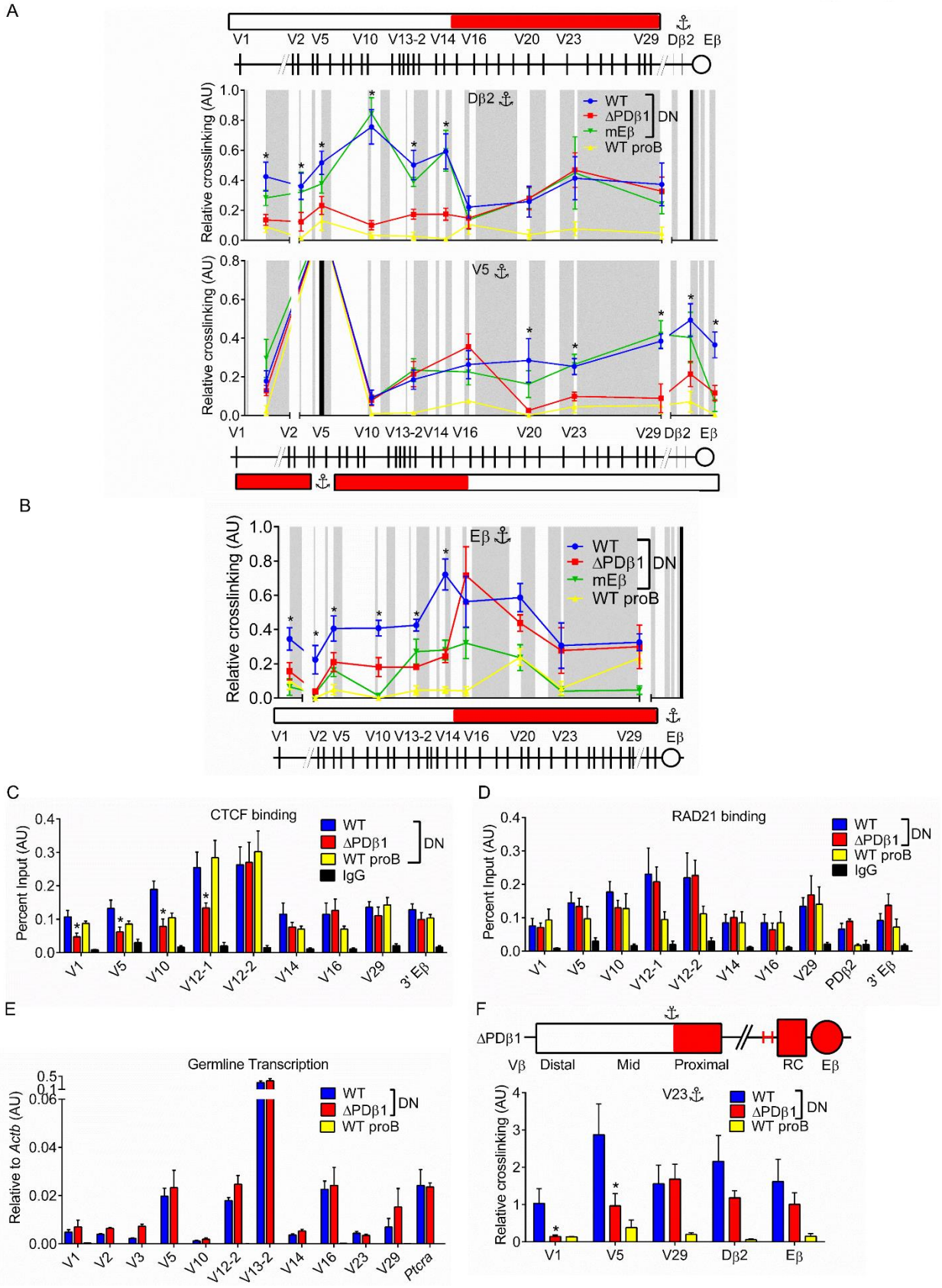

A
V1 V2 V5 V10 V13-2 V14 V16 V20 V23 V29 Dβ2 Eβ
Dβ2
WT
ΔPDβ1
mEβ
WT proB
DN
Relative crosslinking (AU)
V5
B
Eβ
C
CTCF binding
Percent Input (AU)
IgG
V1 V5 V10 V12-1 V12-2 V14 V16 V29 3' Eβ
D
RAD21 binding
PDβ2
E
Germline Transcription
Relative to Actb (AU)
V1 V2 V3 V5 V10 V12-2 V13-2 V14 V16 V23 V29 Ptcra
F
ΔPDβ1
Vβ Distal Mid Proximal RC Eβ
V23
V1 V5 V29 Dβ2 Eβ

Figure 3.3: Deletion of the 5' RC flank resolves two *Trbv* interaction domains. (A) 3C analysis of RAG-deficient thymocytes (WT, ΔPDβ1 or mEβ alleles) and pro-B cells using the Dβ2 (upper), *Trbv5* (lower) and (B) Eβ viewpoints. Individual Hind III fragments are represented by alternating white and grey bars. Bold black bars indicate viewpoint locations. See Figure 1 for details of cartoon data summaries. Red shading indicates whether *Trbv*–Dβ2 crosslinking in ΔPDβ1 relative to WT alleles is unchanged (darkest red) or reduced to background levels in pro-B cells (white). (C) 3C assays were performed with the *Trbv23* viewpoint. Data are presented as mean values (± SEM) from at least three independent experiments. Significant differences between WT and ΔPDβ1 samples are denoted as * $p<0.05$ (t-test). ChIP-qPCR assay for (D) CTCF and (E) RAD21 binding at sites near the indicated *Trbv* segments. Refer to Figure 2C for details. (F) *Trbv* germline transcription was quantified relative to *Actb* as shown. Data are presented as mean percent input (± SEM) from at least three independent experiments with statistically significant differences denoted as * $p<0.05$ (t-test).

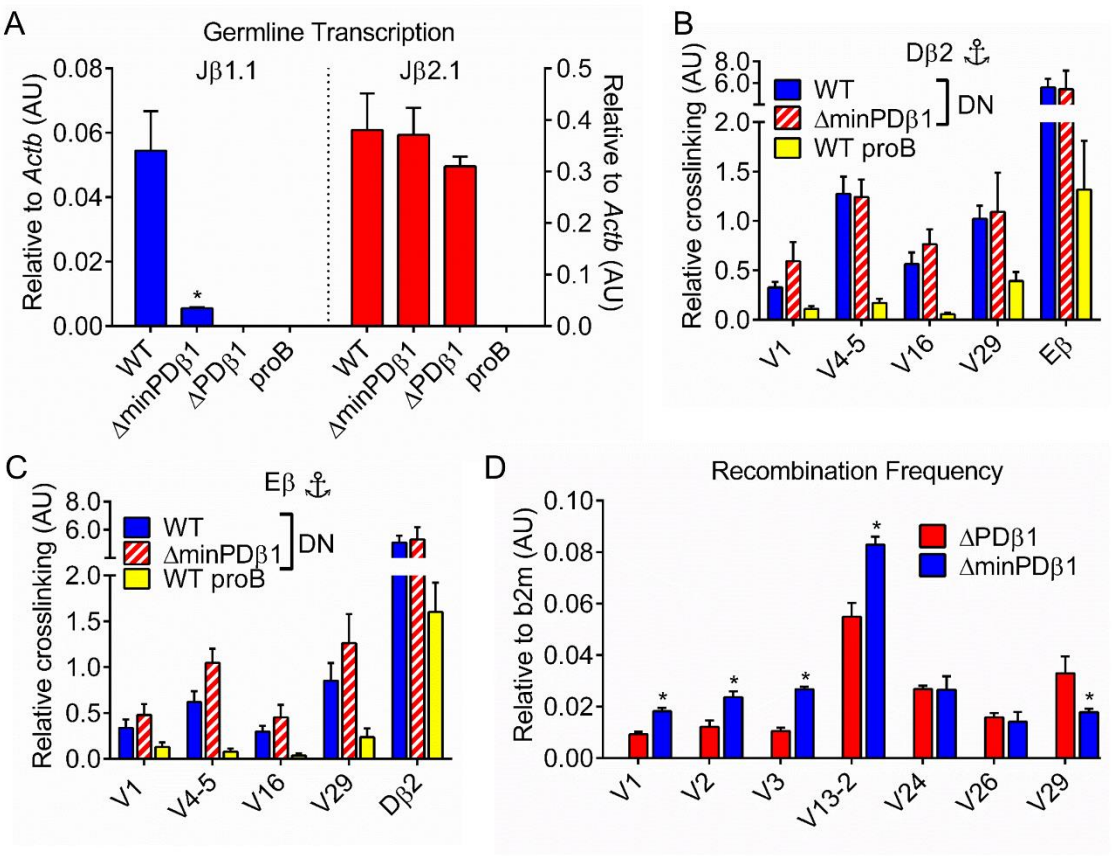


Figure 3.4: *Tcrb* looping is independent of Dβ1 promoter function. (A) Spliced germline transcripts traversing Jβ1.1 or Jβ2.1 to their respective Cβ exons were quantified relative to *Actb* in DN thymocytes from the indicated genotypes and WT pro-B cells. 3C assays were performed with (B) Dβ2 and (C) Eβ viewpoints in the indicated genotypes. (D) Quantification of *Trbv* usage in total thymocytes from ΔPDβ1 and ΔminPDβ1 mice on a RAG-sufficient background. Relative levels of joins between the indicated Vβ segments and Dβ2Jβ2.1 were assayed and normalized as described previously (Gopalakrishnan et al., 2013). Data are represented as mean of three independent experiments (± SEM) with statistically significant differences indicated as * $p<0.05$ (t-test).

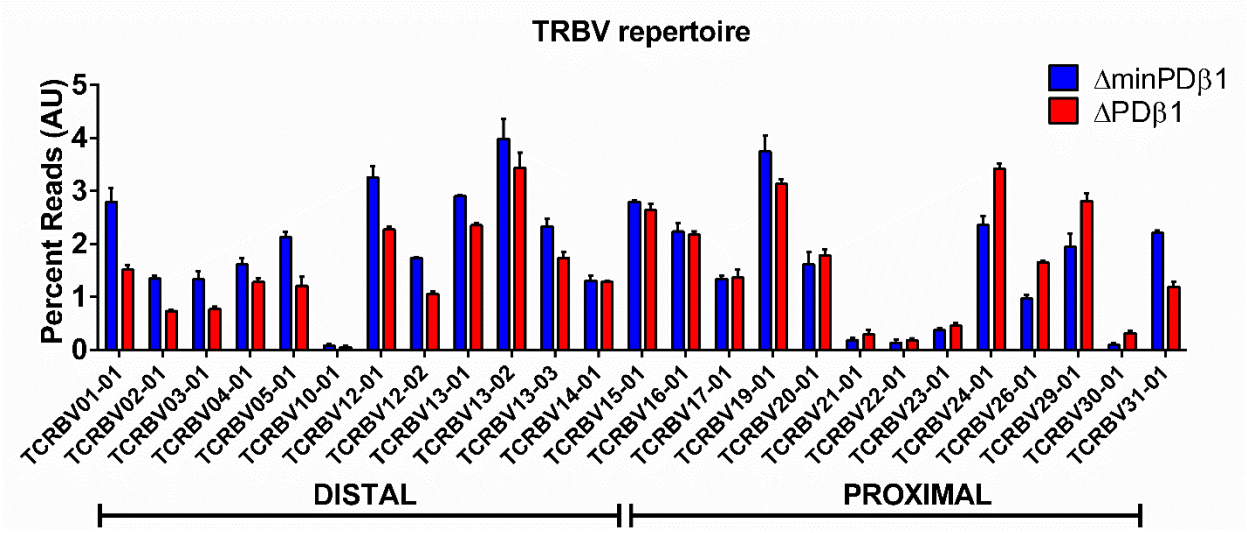


Figure 3.5: The pre-selection *Tcrb* repertoire underutilizes distal *Trbv* gene segments. CD44^{-}CD25^{+} (DN3) thymocytes were sorted from mice of the indicated genotypes, DNA was extracted and subjected to repertoire analysis by sequencing (Adaptive Biotechnologies). ΔPDβ1 thymocytes showed a robust underutilization of distal *Trbv* genes. Data are presented as mean±SEM of three independent experiments.

Tcrb contraction is PDβ1-dependent but Eβ-independent

Tcrb undergoes a large-scale spatial reconfiguration, termed contraction, upon differentiation of progenitors into DN thymocytes (Skok et al., 2007). As monitored by 3D-FISH, contraction brings opposing termini of *Tcrb* -- the distal 5' *Trbv* region and RC -- into proximity, to facilitate long-range V-DJ recombination. Upon assembly of a productive *Tcrb* allele and transition to the DP stage of development, locus contraction is reversed, segregating the *Trbv* and DJ clusters, presumably enforcing allelic exclusion (Skok et al., 2007). However, functional relationships between AgR locus contraction and long-range V-RC looping remain unclear.

To test whether known REs contribute to *Tcrb* contraction, we performed 3D-FISH analyses on thymocytes from RAG-deficient mice harboring WT, mEβ, and ΔPDβ1 alleles. RAG1$^{-/-}$:D708A thymocytes were also assayed to test whether the deposition of RAG1 influences *Tcrb* contraction. Representative primary data for FISH experiments are shown in Figure 3.6A. As expected, distances between the V1 and trypsinogen probes (Figure 3.6B, upper) are significantly greater in *Rag1*$^{-/-}$ DP versus DN thymocytes, reflecting the contracted nature of *Tcrb* in the latter (Figure 3.6B, lower). *Tcrb* contraction is unaffected in DN thymocytes upon inactivation of the RC (mEβ and ΔEβ thymocytes), RAG1:D708A binding, or loss of the minimal Dβ1 promoter (Figure 3.6B). In contrast, the locus adopts an intermediate conformation in ΔPDβ1/*Rag1*$^{-/-}$ thymocytes, significantly more extended than in DN cells harboring a WT-*Tcrb*, but significantly more contracted than in their DP counterparts.

These conclusions are supported by FISH data using two additional probe sets that measure distances between the RC and either the most distal *Trbv* segment (*Trbv1*, Figure 3.6C) or the main portion of the distal domain (*Trbv2-12*, Figure 3.6D). Thus, consistent with 3C data, folding of the

most distal *Trbv* portion into the RC–3'*Trbv* aggregate is independent of transcriptional activity at DβJβ clusters. Instead, full contraction of the locus requires a region directly upstream of the RC, which includes PDβ1.

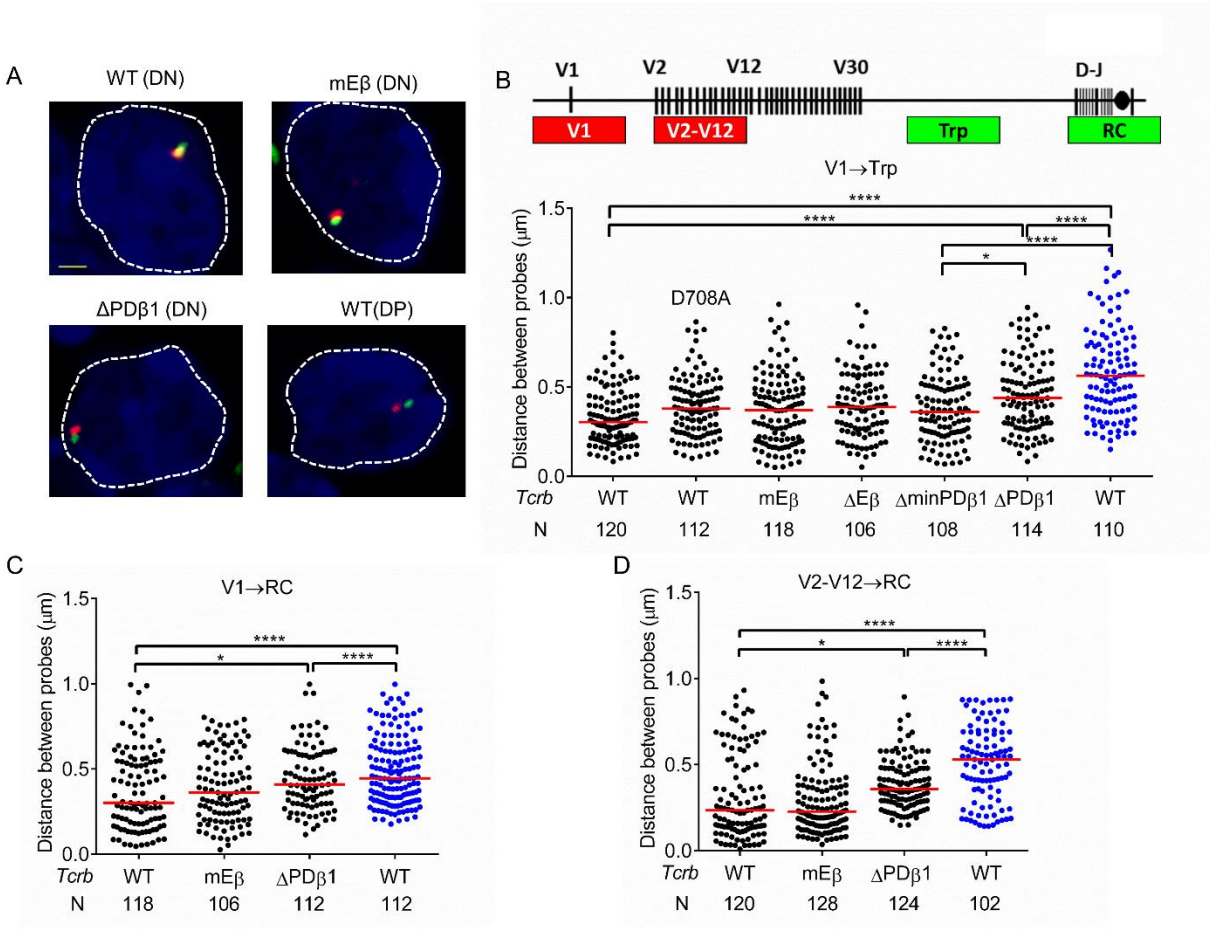
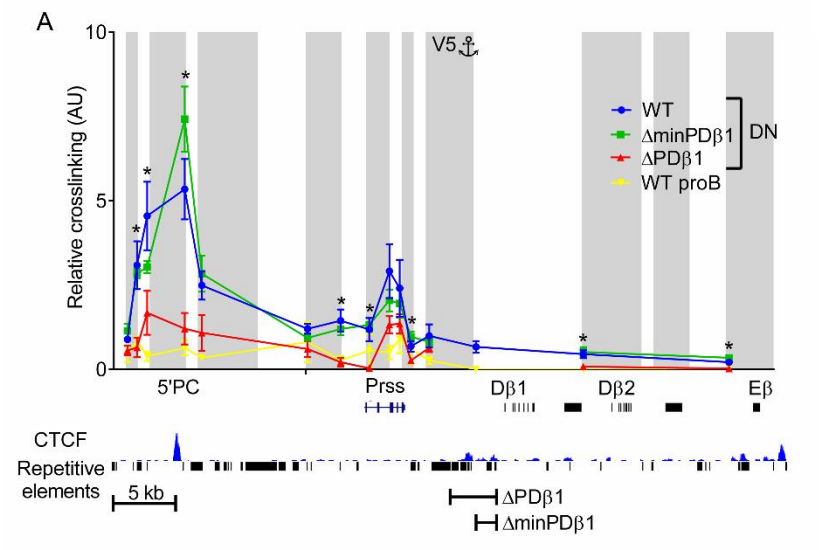
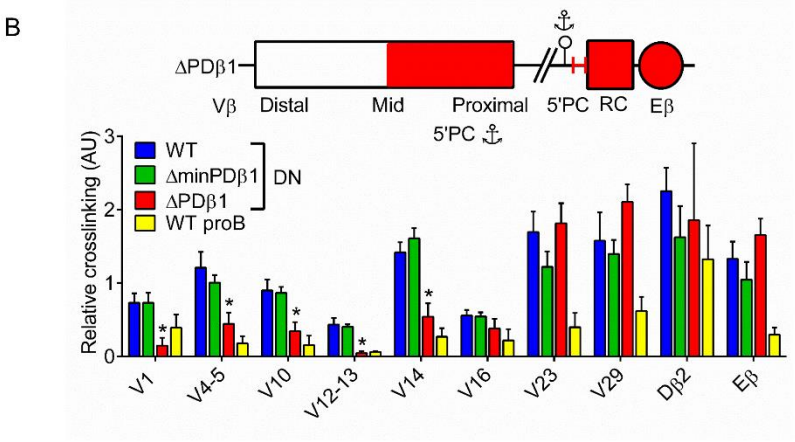


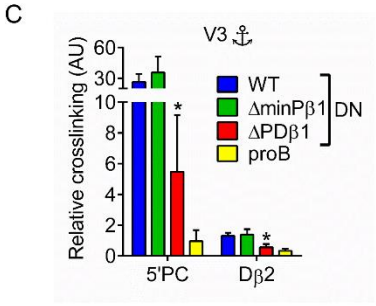
Figure 3.6: Partial decontraction of *Tcrb* in ΔPDβ1 thymocytes. (A) Representative 3D-FISH data of *Tcrb* contraction for the V1 (red) and trypsinogen region probes (green) shown in panel 5B. Blue corresponds to DAPI staining (Scale bars, 1μm). (B-D) Distances between the indicated regions of *Tcrb* were measured using BAC probes spanning (A) *Trbv1* (red) and trypsinogen (green), (B) *Trbv1* and the recombination center (RC, green), and (C) Trbv2-12 (red) and the RC (green). Contraction was measured in RAG-deficient DN thymocytes (shown as black dots) for the indicated *Tcrb* genotypes or in DP thymocytes (blue dots). Results are presented as scatter plots of distances between probe foci for each *Tcrb* allele and represent total data from at least three independent preparations of slides. Statistical analyses revealed no significant differences between independent experiments performed on the same genotype or cell type. Median values are indicated by red horizontal lines. Significant differences are denoted as * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, and **** $p \leq 0.0001$ (One way ANOVA, Tukey's posthoc test).

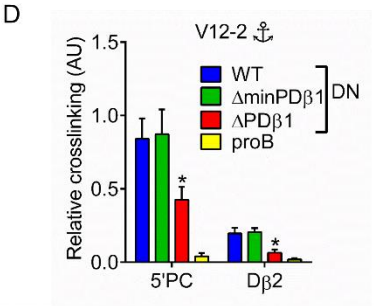
A CTCF-binding region serves as the focal point for distal Trbv-RC interactions

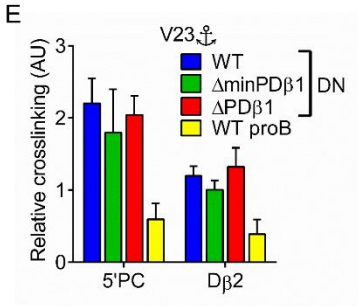
In an attempt to understand how the region upstream of minPDβ1 impacts long-range *Tcrb* looping, we surveyed its interactions with a distal portion of the *Trbv* cluster. Using *Trbv5* as a viewpoint, we scanned interactions with a series of restriction fragments upstream of PDβ1 (Figure 3.7A). Compared with pro-B cells, *Trbv5* crosslinks more efficiently with this region in DN thymocytes at nearly all tested locations. The most robust *Trbv5* interaction occurs upstream of a silent trypsinogen gene, termed *Prss2*, which coincides with a prominent site for CTCF binding (Figure 3.7A, lower) (Shih et al., 2012). Association between *Trbv5* and this region, which we call the 5'*Prss2*-CTCF site (5'PC), is even greater than its interaction with the RC. Importantly, this prominent contact is disrupted in *Tcrb* loci with the large (ΔPDβ1), but not the minimal, Dβ1 promoter deletion. These findings are completely consistent with 3C data obtained with either 5'PC (Figure 3.7B) or two other distal *Trbv* segments as viewpoints for interactome analyses (Figure 3.7C and 3.7D). In contrast, robust interactions between 5'PC and proximal *Trbv* segments are unaffected by the ΔPDβ1 deletion (Figures 3.7E). We conclude that 5'PC is a focal point for long-range interactions between the distal *Trbv* domain and the RC, a process that depends on a region upstream of minimal PDβ1.


A
Relative crosslinking (AU)
V5
WT
ΔminPDβ1
ΔPDβ1
WT proB
DN
5'PC
Prss
Dβ1
Dβ2
Eβ
CTCF
Repetitive elements
5 kb
ΔPDβ1
ΔminPDβ1


B
ΔPDβ1
Vβ
Distal
Mid
Proximal
5'PC
RC
Eβ
5'PC
Relative crosslinking (AU)
WT
ΔminPDβ1
ΔPDβ1
WT proB
DN
V1
V4-5
V10
V12-13
V14
V16
V23
V29
Dβ2
Eβ


C
V3
Relative crosslinking (AU)
WT
ΔminPβ1
ΔPDβ1
proB
DN
5'PC
Dβ2


D
V12-2
Relative crosslinking (AU)
WT
ΔminPDβ1
ΔPDβ1
proB
DN
5'PC
Dβ2


E
V23
Relative crosslinking (AU)
WT
ΔminPDβ1
ΔPDβ1
WT proB
DN
5'PC
Dβ2

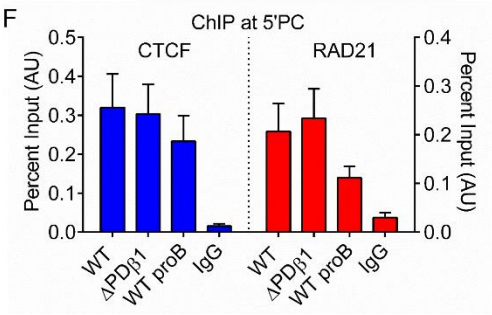

F
ChIP at 5'PC
CTCF
RAD21
Percent Input (AU)
Percent Input (AU)
WT
ΔPDβ1
WT proB
IgG
WT
ΔPDβ1
WT proB
IgG

Figure 3.7: Identification of a *Trbv* tethering point in the RC flank. 3C data for (A) *Trbv5*, (B) 5' PC, (C) *Trbv3*, (D) *Trbv12-2*, and (E) *Trbv23* viewpoints in RAG-deficient DN thymocytes (WT, ΔPDβ1 or ΔminPDβ1 mice) or pro-B cells (see Figure 1A for details). The lower part of panel A shows a ChIP-seq track for CTCF in DN thymocytes (Shih et al., 2012) as well as locations of repetitive elements. Data for the 5'PC viewpoint are summarized as a cartoon (see Figure 1E). (F) ChIP-qPCR for CTCF and RAD21 at 5'PC in the indicated cell types. All data are represented as means (± SEM) of three independent experiments. Significant differences are denoted as * $p \leq 0.05$ (t test between WT and ΔPDβ1 genotypes).

An RC barrier element is required for long-range Trbv looping to 5'PC

Although 5'PC tethers the distal *Trbv* domain, the mechanisms by which ΔPDβ1, but not ΔminPDβ1, disrupt thymocyte-specific contacts were unclear. In this regard, the 5'PC region remains completely intact on ΔPDβ1 alleles; deleted sequences are restricted to a region at least 20 kb downstream (Figure 3.7A lower). Furthermore, ChIP experiments reveal no significant differences in CTCF or RAD21 binding at 5'PC when comparing WT and ΔPDβ1 alleles (Figure 3.7F). These findings imply an activity associated with the 3 kb region upstream of minPDβ1 that impacts the ability of 5'PC to form long-range interactions with distal portions of *Trbv*.

The region of interest has several distinguishing characteristics, including a repetitive tract at its 5' end and a pair of low intensity CTCF/RAD21 binding sites (Figure 3.7A, lower). These features are reminiscent of insulators that form boundaries between active and repressive chromatin domains (Wendt et al., 2008). In keeping with this possibility, a gene situated upstream of the putative chromatin barrier, *Prss2,* is transcriptionally active in ΔPDβ1 thymocytes, but is completely silent in the context of WT, ΔminPDβ1 or mEβ alleles (Figure 3.8A). *Prss2* activation in ΔPDβ1 thymocytes is mirrored by an acquisition of H3K4me3 at its promoter region (Figure 3.8B).

To further define how the ΔPDβ1 deletion impacts neighboring chromatin domains, we performed ChIP experiments for activating histone modifications within and upstream of the *Tcrb*-RC. As shown in Figure 3.8C (lower), the H3K4me2 mark for accessible chromatin spreads throughout the RC in DN thymocytes, continuing to a CTCF site upstream of minPDβ1, after which it drops dramatically (Carabana et al., 2011). As expected, this modification is nearly absent in mEβ thymocytes, which harbor inactive *Tcrb*-RCs. Strikingly, H3K4me2 spreads much further

upstream in thymocytes from the ΔPDβ1, but not ΔminPDβ1 mice, indicating disruption of a chromatin boundary in the former. Instead, a new chromatin boundary is established at or near 5'PC in the ΔPDβ1 thymocytes. A similar profile is observed for a second active chromatin mark, H3ac (Figure 3.8C, upper).

Conversely, the repressive modifications H3K9me2 and H3K27me3, drop significantly near the boundary region upstream of the RC in DN thymocytes with either WT or ΔminPDβ1 alleles (Figure 3.8D). When enhancer function is disrupted (mEβ), the H3K9me2 and K3K27me3 marks also cover the inactivated RC, as expected. When the border region is removed (ΔPDβ1), there is a modest, but significant loss of these modifications directly upstream, likely reflecting the invasion of active chromatin into this normally repressed region. Similarly, there is a modest invasion of the two repressive marks into the most proximal end of the RC. Thus, the most significant impact of removing the 5'PDβ1 boundary region is the invasion of active chromatin (H3K4me2 and H3ac) for a substantial distance upstream of the RC, resulting in the transcriptional activation of *Prss2*.

Taken together, our results demonstrate that the region upstream of PDβ1 serves as a chromatin barrier, which is required to preserve the function of 5'PC as a tether for distal regions of the *Trbv* cluster. When the normal boundary separating active from inactive chromatin is disrupted by the ΔPDβ1 deletion, a barrier function for 5'PC is unmasked, impairing its ability to maintain distal *Trbv*-RC contacts.

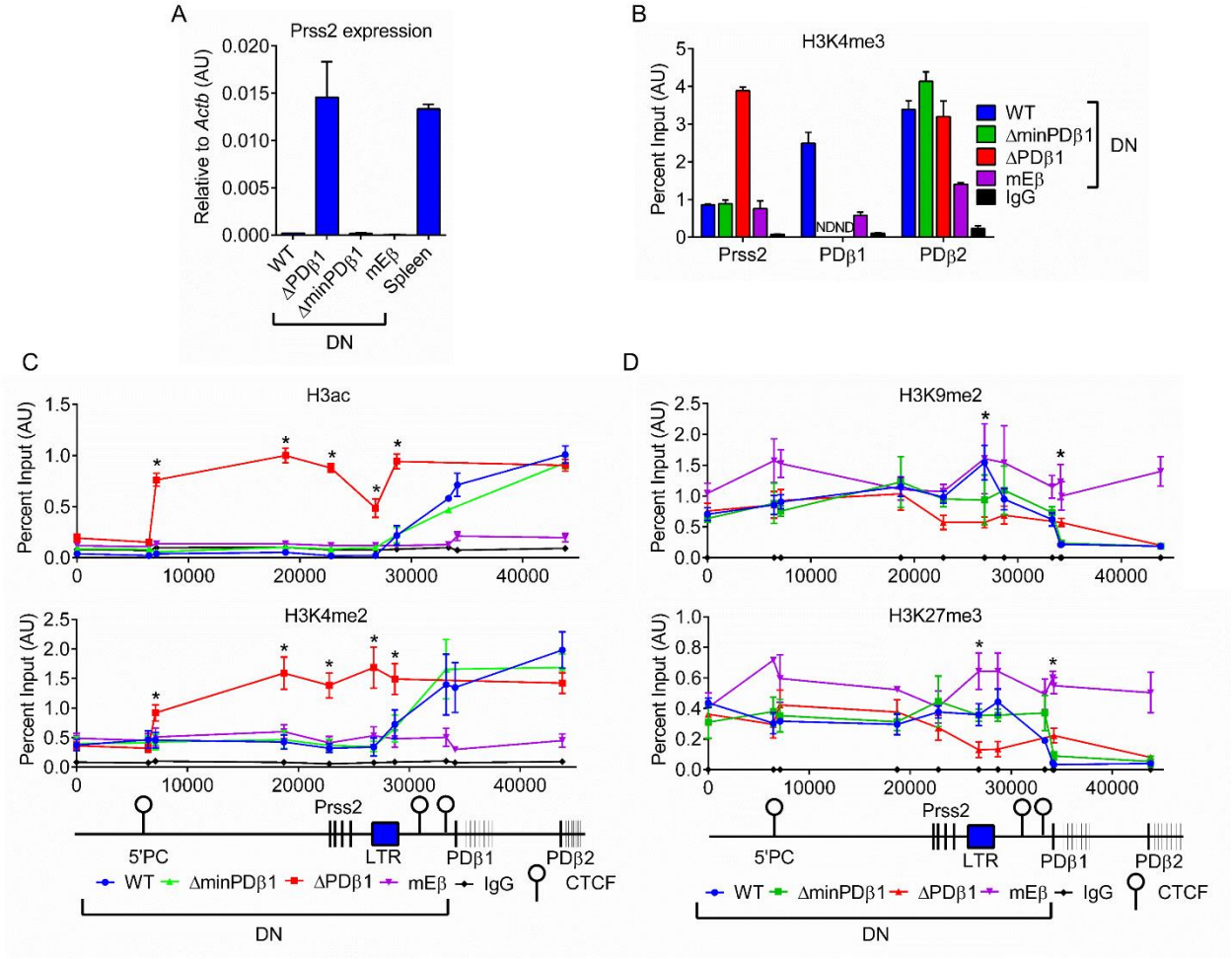


Figure 3.8: Long-range *Trbv* looping to 5'PC requires an RC barrier element. (A) Expression of *Prss2* transcripts relative to *Actb* in DN thymocytes (WT, ΔPDβ1, ΔminPDβ1, and mEβ mice) and in spleen from C57BL/6 mice, as a positive control. (B) Levels of the H3K4me3 modifications at the indicated promoters in RAG-deficient thymocytes harboring the listed *Tcrb* genotypes. ChIP-qPCR assays for: (C) active histone marks H3ac (top) and H3K4me2 (bottom), and (D) repressive histone marks H3K9me2 (top) and H3K27me3 (bottom) at the indicated sites upstream or within the RC. All data are represented as means (± SEM) of at least two independent experiments. Significant differences are denoted as * $p \leq 0.05$ (t-test between WT and ΔPDβ1 genotypes).

3.5 Discussion

Lineage- and stage-specific assembly of AgR genes requires whole-scale changes in locus structure and extensive revisions to their chromatin landscapes, which are largely directed by regulatory elements flanking RCs. Here, we shed light on the complex function of these regulatory elements in both aspects of *Tcrb* assembly. As discussed below, our findings have implications not only for regulatory strategies employed by other AgR loci, but also for the spatial mechanisms that control gene expression programs.

Tcrb adopts a thymocyte-specific conformation that, surprisingly, is independent of RC activity, including its transcription and binding of RAG proteins. Instead, the fully active *Tcrb* conformation requires a region directly flanking the RC, which functions as a barrier element to block the spread of active RC chromatin into a repressive upstream region. Disruption of the barrier relocates the active–inactive chromatin boundary to the nearest upstream CTCF site (5'PC), which normally serves as a major tethering point for distal *Trbv* segments. Our findings suggest that forcing 5'PC to become an insulator decommissions its tethering function, partially unspools the active *Tcrb* conformation, and skews the primary repertoire to favor more proximal *Trbv* segments.

Although Eβ function is essential for RC activation, it is dispensable for long-range association between *Trbv* segments and the two DβJβ clusters. Similarly, *Tcrb* contraction is Eβ-independent, an observation that is consistent with data from other AgR loci harboring enhancer deletions (Shih and Krangel, 2013). These findings preclude several proposed mechanisms for the folding of AgR loci, or at least *Tcrb*, into their active conformations, including: (i) a requirement for accessible RC chromatin, (ii) RAG-mediated interactions between RC and V domains, and (iii) co-occupancy of the RC and distal V segments in a transcription factory. Instead, we find that the

crippled enhancer either protrudes from the Vβ–DβJβ interactome, or is potentially sequestered into the central *Trbv12-16* gene cluster, resulting in transcriptional attenuation of the most active *Trbv* segments. Suppression of these *Trbv* segments is unlikely to result directly from loss of enhancer contact, but rather is an indirect effect of their continued association with a repressed RC. In support of this possibility, contacts between many of these Vβ segments and the RC are disrupted on the ΔPDβ allele, which retains robust expression of the Dβ2Jβ cluster, as well as normal level of germline Vβ transcription. Likely, germline transcription of the *Trbv* segments is mostly due to the activity of their associated promoters, but when juxtaposed with a repressive chromatin environment in the mEβ alleles, the promoters are silenced.

The general relevance of enhancer-independent V-RC association at other AgR loci is uncertain given available data, in part because *Ig* loci, unlike *Tcrb*, are decorated with multiple enhancers that form interaction networks and could have redundant functions in generating an active conformation (Degner-Leisso and Feeney, 2010). Of equal importance, many of the prior studies have probed locus-wide interactions only from the enhancer perspective, but based on our findings, viewpoints within the (D)J cluster itself may yield more relevant data for long-range V-RC interactions (Changying Guo et al., 2011; Medvedovic et al., 2013). At *Tcra*, a single enhancer (Eα) is tethered to the Jα germline promoter (TEA), generating an active chromatin hub for tertiary interactions with proximal *Trav* segments (Shih et al., 2012). Deletion of either Eα or TEA perturbs the proximal Vα to Jα contacts or redistributes enhancer interactions to include the intervening *Tcrd* locus. Thus, in contrast to *Tcrb*, interactions between proximal V segments and their RC targets are enhancer-dependent at *Tcra*, suggesting that certain aspects of topological control are AgR locus-specific. Conformational requirements likely are tailored to the unique

architectures of *Ig* and *Tcr* loci, and may reflect the broad range of spatial mechanisms that can be employed to control gene expression in eukaryotes.

A surprising aspect of our study was that removal of the 5' RC flank, which includes PDβ1, disrupts long-range *Tcrb* interactions, resolving the *Trbv* cluster into distal and proximal domains, each with unique spatial determinants. The bidomainal architecture of *Trbv* is apparent from effects of the ΔPDβ1 deletion on long-range associations in a cell population (3C assays), or by probing locus contraction in single cells (3D-FISH). The protrusion of distal *Trbv* segments from the Vβ–DβJβ interactome is independent of promoter function since a more specific disruption of the core PDβ1 element has no impact on distal *Trbv*–RC juxtaposition. Based on our extensive 3C data, we map the approximate border between proximal and distal Vβ interaction domains to within the *Trbv14-16* region, a 16 kb stretch. Although precise border mapping and underlying mechanisms for its establishment remain to be resolved, we point out that the boundary coincides well with a transition between robust CTCF binding within the distal *Trbv* portion and more modest binding of these structural factors in the proximal domain (Figure 3.2C) (Shih et al., 2012). We have been unable to identify other distinguishing characteristics of this region, including unique chromatin landscapes or predicted transcription factor sites. In what may be a related issue, determinants for tethering the proximal *Trbv* domain to its RC target, 250 kb away, remain unknown. Like the distal domain, proximal *Trbv* segments form major contacts with 5'PC; however, these interactions are unaffected by the ΔPDβ1 deletion. In contrast with the distal domain, proximal *Trbv* segments generally form equally robust associations with 5'PC and the RC. Based on these observations, we propose that the distal *Trbv* cluster relies on CTCF-dominant contacts with 5'PC to bring it into proximity with the RC. Disruption of these contacts may also

explain the partial loss of CTCF binding near distal *Trbv* segments in ΔPDβ1 thymocytes. In contrast, the proximal region of *Trbv* could also bridge to the RC by CTCF-independent mechanisms, which may be analogous to transcription factor-mediated looping at *Igh* (Medvedovic et al., 2013).

In our quest to decipher how the 5'RC flank impacts its association with distal *Trbv* segments, we found that the ΔPDβ1 deletion disrupts a chromatin boundary. As a result, hyperactive RC chromatin spreads upstream, leading to inappropriate expression of the silent *Prss2* gene. Although the deleted region exhibits two modest peaks of CTCF–RAD21 in DN thymocytes, the precise determinants of its insulator function remain unclear. In this regard, the region between PDβ1 and *Prss2* is repetitive and contains a viral LTR element that is expressed at low levels in DN thymocytes and has insulator properties (Carabana et al., 2011). A closer inspection of chromatin data for this region suggests that it contains a bimodal insulator consisting of the LTR, which blocks the spread of repressive chromatin downstream into the RC (Carabana et al., 2011), and the PDβ1-associated CTCF sites, which prevents the spread of hyperactive RC chromatin upstream into the *Prss2* region (shown here).

Notwithstanding these mechanistic uncertainties, deletion of the 5'RC flank disrupts an active chromatin barrier, which allows it to spread upstream until reaching the next CTCF region, 5'PC. When 5'PC becomes the dominant RC chromatin barrier, it is decommissioned as a long-range tether for distal *Trbv* segments. Several potential underlying mechanisms for this functional switch can be envisioned, including the major revision of local epigenetic landscapes when the RC-flanking insulator is disarmed. In this regard, cohesin mediates long-range chromatin looping not only through its association with CTCF, but also when it is recruited to the transcriptional

mediator complex (Kagey et al., 2010). Emerging studies indicate that CTCF-cohesin bridges are predominantly structural in nature, similar to distal *Trbv*-5'PC interactions, whereas cohesin-mediator largely bridges loops between regulatory elements (Kagey et al., 2010). Perhaps the activation of transcription near 5'PC converts it into a region that favors participation in regulatory, rather than structural loops.

Our finding that distal *Trbv*-RC interactions depend on a bifunctional insulator–tethering element upstream of the RC, is likely relevant to the architectural determinants of other AgR loci. For example, *Igh* enhancers interact with a CTCF-rich region, called the IGCR, which clearly serves as a chromatin boundary between its RC and proximal *Ighv* segments (Guo, Yoon et al., 2011). Similarly, two CTCF regions in *Igk*, termed Cer and Sis, contribute to the insulation of proximal *Igkv* segments from the enhancer-rich *Igkj* cluster (Xiang et al., 2013; Xiang et al., 2011). Based on our discovery of a bifunctional element in the *Tcrb*-RC flank, we would hypothesize that at *Ig* loci, the most RC-proximal CTCF site(s) serves as an insulator (e.g., CBE2 in IGCR; Sis at *Igk*) to protect the tethering function of the more distal CTCF site(s) (e.g., CBE1 in IGCR; Cer at *Igk*). Resolution of these issues in the topological regulation of AgR loci will lend important insights into the menu of mechanisms that can be deployed to control gene expression programs in response to developmental cues or physiologic agonists.

3.6 References

Bories, J. C., Demengeot, J., Davidson, L., & Alt, F. W. (1996). 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. *Proceedings of the National Academy of Sciences*, *93*(15), 7871-7876.

Bouvier, G., Watrin, F., Naspetti, M., Verthuy, C., Naquet, P., & Ferrier, P. (1996). Deletion of the mouse T-cell receptor beta gene enhancer blocks alphabeta T-cell development. *Proceedings of the National Academy of Sciences*, *93*(15), 7877-7881.

Carabana, J., Watanabe, A., Hao, B., & Krangel, M. S. (2011). A barrier-type insulator forms a boundary between active and inactive chromatin at the murine TCRβ locus. *The Journal of Immunology*, *186*(6), 3556-3562.

Cobb, R. M., Oestreich, K. J., Osipovich, O. A., & Oltz, E. M. (2006). Accessibility control of V (D) J recombination. *Advances in immunology*, *91*, 45-109.

De Laat, W., & Duboule, D. (2013). Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature*, *502*(7472), 499-506.

Degner-Leisso, S. C., & Feeney, A. J. (2010, December). Epigenetic and 3-dimensional regulation of V (D) J rearrangement of immunoglobulin genes. *Seminars in immunology* (Vol. 22, No. 6, pp. 346-352). Academic Press.

Dekker, J., Rippe, K., Dekker, M., & Kleckner, N. (2002). Capturing chromosome conformation. *Science*, *295*(5558), 1306-1311.

Gopalakrishnan, S., Majumder, K., Predeus, A., Huang, Y., Koues, O. I., Verma-Gaur, J., ... & Oltz, E. M. (2013). Unifying model for molecular determinants of the preselection Vβ repertoire. *Proceedings of the National Academy of Sciences*, *110*(34), E3206-E3215.

Guo, C., Gerasimova, T., Hao, H., Ivanova, I., Chakraborty, T., Selimyan, R., ... & Sen, R. (2011). Two forms of loops generate the chromatin conformation of the immunoglobulin heavy-chain gene locus. *Cell*, *147*(2), 332-343.

Guo, C., Yoon, H. S., Franklin, A., Jain, S., Ebert, A., Cheng, H. L., ... & Alt, F. W. (2011). CTCF-binding elements mediate control of V (D) J recombination. *Nature*, *477*(7365), 424-430.

Hagège, H., Klous, P., Braem, C., Splinter, E., Dekker, J., Cathala, G., ... & Forné, T. (2007). Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nature protocols*, *2*(7), 1722-1733.

Hewitt, S. L., Farmer, D., Marszalek, K., Cadera, E., Liang, H. E., Xu, Y., ... & Skok, J. A. (2008). Association between the Igk and Igh immunoglobulin loci mediated by the 3′ Igk enhancer induces' decontraction'of the Igh locus in pre–B cells. *Nature immunology*, *9*(4), 396-404.

Jackson, A. M., & Krangel, M. S. (2006). Turning T-cell receptor β recombination on and off: more questions than answers. *Immunological reviews*, *209*(1), 129-141.

Jhunjhunwala, S., van Zelm, M. C., Peak, M. M., Cutchin, S., Riblet, R., van Dongen, J. J., ... & Murre, C. (2008). The 3D structure of the immunoglobulin heavy-chain locus: implications for long-range genomic interactions. *Cell*,*133*(2), 265-279.

Ji, Y., Little, A. J., Banerjee, J. K., Hao, B., Oltz, E. M., Krangel, M. S., & Schatz, D. G. (2010). Promoters, enhancers, and transcription target RAG1 binding during V (D) J recombination. *The Journal of experimental medicine*,*207*(13), 2809-2816.

Ji, Y., Resch, W., Corbett, E., Yamane, A., Casellas, R., & Schatz, D. G. (2010). The in vivo pattern of binding of RAG1 and RAG2 to antigen receptor loci. *Cell*, *141*(3), 419-431.

Kagey, M. H., Newman, J. J., Bilodeau, S., Zhan, Y., Orlando, D. A., van Berkum, N. L., ... & Young, R. A. (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature*, *467*(7314), 430-435.

Kosak, S. T., Skok, J. A., Medina, K. L., Riblet, R., Le Beau, M. M., Fisher, A. G., & Singh, H. (2002). Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science*, *296*(5565), 158-162.

Mathieu, N., Hempel, W. M., Spicuglia, S., Verthuy, C., & Ferrier, P. (2000). Chromatin Remodeling by the T Cell Receptor (Tcr)-β Gene Enhancer during Early T Cell Development Implications for the Control of Tcr-β Locus Recombination. *The Journal of experimental medicine*, *192*(5), 625-636.

Medvedovic, J., Ebert, A., Tagoh, H., Tamir, I. M., Schwickert, T. A., Novatchkova, M., ... & Busslinger, M. (2013). Flexible Long-Range Loops in the V H Gene Region of the Igh Locus Facilitate the Generation of a Diverse Antibody Repertoire. *Immunity*, *39*(2), 229-244.

Nasmyth, K., & Haering, C. H. (2009). Cohesin: its roles and mechanisms.*Annual review of genetics*, *43*, 525-558.

Oestreich, K. J., Cobb, R. M., Pierce, S., Chen, J., Ferrier, P., & Oltz, E. M. (2006). Regulation of TCRβ gene assembly by a promoter/enhancer holocomplex. *Immunity*, *24*(4), 381-391.

Osipovich, O., & Oltz, E. M. (2010, December). Regulation of antigen receptor gene assembly by genetic–epigenetic crosstalk. In *Seminars in immunology*(Vol. 22, No. 6, pp. 313-322). Academic Press.

Osipovich, O., Cobb, R. M., Oestreich, K. J., Pierce, S., Ferrier, P., & Oltz, E. M. (2007). Essential function for SWI-SNF chromatin-remodeling complexes in the promoter-directed assembly of Tcrb genes. *Nature immunology*, *8*(8), 809-816.

Phillips, J. E., & Corces, V. G. (2009). CTCF: master weaver of the genome.*Cell*, *137*(7), 1194-1211.

Ribeiro de Almeida, C., Stadhouders, R., de Bruijn, M. J., Bergen, I. M., Thongjuea, S., Lenhard, B., ... & Hendriks, R. W. (2011). The DNA-binding protein CTCF limits proximal Vκ recombination and restricts κ enhancer interactions to the immunoglobulin κ light chain locus. *Immunity*, *35*(4), 501-513.

Sanyal, A., Lajoie, B. R., Jain, G., & Dekker, J. (2012). The long-range interaction landscape of gene promoters. *Nature*, *489*(7414), 109-113.

Schatz, D. G., & Ji, Y. (2011). Recombination centres and the orchestration of V (D) J recombination. *Nature Reviews Immunology*, *11*(4), 251-263.

Seitan, V. C., Hao, B., Tachibana-Konwalski, K., Lavagnolli, T., Mira-Bontenbal, H., Brown, K. E., ... & Merkenschlager, M. (2011). A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation. *Nature*,*476*(7361), 467-471.

Shih, H. Y., & Krangel, M. S. (2010). Distinct contracted conformations of the Tcra/Tcrd locus during Tcra and Tcrd recombination. *The Journal of experimental medicine*, *207*(9), 1835-1841.

Shih, H. Y., & Krangel, M. S. (2013). Chromatin architecture, CCCTC-binding factor, and V (D) J recombination: managing long-distance relationships at antigen receptor loci. *The Journal of Immunology*, *190*(10), 4915-4921.

Shih, H. Y., Verma-Gaur, J., Torkamani, A., Feeney, A. J., Galjart, N., & Krangel, M. S. (2012). Tcra gene recombination is supported by a Tcra enhancer-and CTCF-dependent chromatin hub. *Proceedings of the National Academy of Sciences*, *109*(50), E3493-E3502.

Shinkai, Y., & Alt, F. W. (1994). CD3ε-mediated signals rescue the development of CD4+ CD8+ thymocytes in RAG-2−/− mice in the absence of TCR β chain expression. *International immunology*, *6*(7), 995-1001.

Sikes, M. L., Gomez, R. J., Song, J., & Oltz, E. M. (1998). A developmental stage-specific promoter directs germline transcription of DβJβ gene segments in precursor T lymphocytes. *The Journal of Immunology*, *161*(3), 1399-1405.

Sikes, M. L., Meade, A., Tripathi, R., Krangel, M. S., & Oltz, E. M. (2002). Regulation of V (D) J recombination: a dominant role for promoter positioning in gene segment accessibility. *Proceedings of the National Academy of Sciences*, *99*(19), 12309-12314.

Skok, J. A., Gisler, R., Novatchkova, M., Farmer, D., de Laat, W., & Busslinger, M. (2007). Reversible contraction by looping of the Tcra and Tcrb loci in rearranging thymocytes. *Nature immunology*, *8*(4), 378-387.

Steinel, N. C., Brady, B. L., Carpenter, A. C., Yang-Iott, K. S., & Bassing, C. H. (2010). Posttranscriptional silencing of VβDJβCβ genes contributes to TCRβ allelic exclusion in mammalian lymphocytes. *The Journal of Immunology*,*185*(2), 1055-1062.

Verma-Gaur, J., Torkamani, A., Schaffer, L., Head, S. R., Schork, N. J., & Feeney, A. J. (2012). Noncoding transcription within the Igh distal VH region at PAIR elements affects the 3D structure of the Igh locus in pro-B cells. *Proceedings of the National Academy of Sciences*, *109*(42), 17004-17009.

Wendt, K. S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., ... & Peters, J. M. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature*, *451*(7180), 796-801.

Whitehurst, C. E., Chattopadhyay, S., & Chen, J. (1999). Control of V (D) J recombinational accessibility of the Dβ1 gene segment at the TCRβ locus by a germline promoter. *Immunity*, *10*(3), 313-322.

Whitehurst, C. E., Schlissel, M. S., & Chen, J. (2000). Deletion of germline promoter PDβ1 from the TCRβ locus causes hypermethylation that impairs Dβ1 recombination by multiple mechanisms. *Immunity*, *13*(5), 703-714.

Xiang, Y., Zhou, X., Hewitt, S. L., Skok, J. A., & Garrard, W. T. (2011). A multifunctional element in the mouse Igκ locus that specifies repertoire and Ig loci subnuclear location. *The Journal of Immunology*, *186*(9), 5356-5366.

Xiang, Y., Park, S. K., & Garrard, W. T. (2013). Vκ gene repertoire and locus contraction are specified by critical DNase I hypersensitive sites within the Vκ-Jκ intervening region. *The Journal of Immunology*, *190*(4), 1819-1826.

Yannoutsos, N., Wilson, P., Yu, W., Chen, H. T., Nussenzweig, A., Petrie, H., & Nussenzweig, M. C. (2001). The role of recombination activating gene (RAG) reinduction in thymocyte development in vivo. *The Journal of experimental medicine*, *194*(4), 471-480.

Table T9: Taqman bait primers and probes for 3C

Region	Probe (5'FAM and 3'TAMRA)	Primer
Eβ	CATAAGCATTGTCATGTTTGTGACA	GAAAATTGGCATCGGTTTGC
Dβ1	AAGGCATTGTTGCATGATCCT	TGAAATTTTTCTGCCGAAAGGAC
Dβ2	AAATGCTGGGCCTCTGTAGA	GCGGGATCCAAGAGAACTCA
5' PC	CAGTGGGGAATCAGACTTTCA	TGTGTTGAAGATTGGGGTGA
V3	CCAATGCCCTAATTAACATATTTTCA	CCAGATCTTAGATTTCTGGCCAAC
V5	CAGTCGTTCTTTATGTCTGATACTGTG	TCCCTCAGCGGTTCAGTAGTC
V12-2	TGGTTGAGTAGCAACTTTCTCTTTG	TCTGGAAAATACCCTTATTCCATTG
V23	TACACCGGCCAGGAGAGACT	GGCTTCTGTGTAACTGCAGCAT
ERCC3	AAAGCTTGCACCCTGCTTTAGTGGCC	GCCCTCCCTGAAAATAAGGA

Table T10: Taqman 3C-capture primers

Region	Primer
V1	ACCCATGTCCTCAGGGTTTC
V2-3	TTTCATTCACAGCCGACCAG
V4-5	AGCTCGACACAGAAAGCAAGTT
V10	GTGCCTGTACCATGCTGTGG
V12-13	CCATCTGCATGAACACCTTCTT
V14	CAGGCTTTTGAGTGGCATGT
V16	TATCATGCCCAGCTGCATTC
V20	TGTGATGGGTTGTCATCTGGA

V23	TACACCGGCCAGGAGAGACT
V29	CTCTAGCAATCCCCCTGTGC
Dβ1	AAGGCATTGTTGCATGATCC
Dβ2	TGGGGCCCTCACTTTTCTTA
5’ PC	CCAACTTGCAGTGTGGTCCT
u/s of 5’PC (1XH3)	TCACGCCAAAATACCTGTGA
u/s of 5’PC (2XH3)	GACCAGCAATGGTTAGACTGAA
u/s of 5’PC (3XH3)	TTGTTGTTCACTCTCCTTTCTGA
d/s of 5’PC (1XH3)	TTGCAAGTACCATTTCATGTCAA
u/s Prss2 (2XH3)	CCTCTGATGGAAGGAATTTGC
u/s Prss2 (1XH3)	GCACAGGGAAGTGAGCAGAC
w/ Prss2 promoter	AAATGAGCCTGCATGTCCAC
Prss2 exon2	CAGAGCCACTCCTGAGCAAG
Prss2 exon3	GAGTGGCATGTGAGTGTCCA
d/s Prss2 exon4	GTCCGATGCCCTCTTCTGAT
LTR region	AGGCTCATTTGGGTTGGAGA

Table T11: ChIP-and RT-qPCR primers

Region	Primer
J1 GLT (F)	GAACCAGACTCACAGTTGTAGAGG
Cβ1 GLT (R)	GCTCTCCTTGTAGGCCTGAG
J2 GLT (F)	ACGACTCACCGTCCTAGAGG

Cβ2 GLT (R)	CATTCACCCACCAGCTCAG
V1 (F)	TCAAGCTGTGAACCTACGCTGCAT
V1 (R)	AGGTAATCAGCACCGGGAAGAGAT
V2 (F)	ACAATCAGACTGCCTCAAGTCGCT
V2 (R)	TATGTGGCCGAGTCATCAGGCTTT
V3 (F)	AGGACAGCAGATGGAGTTTCTGGT
V3 (R)	AAGCTGCTGGCACAGAAGTACACA
V5 (F)	TGGAATGTGAGCAACATCTGGGAC
V5 (R)	GGGCACCGTCTCATTTCGAATCAA
V10 (F)	TCTGGTATCAACAAGATGCAGGGC
V10 (R)	AGGTCTGGTTGGAACTGGTTGACT
V12-2 (F)	TCTGTGGCCTGGTATCAACAGACT
V12-2 (R)	GAATCTGCTGGGCAGGTTTCCTTT
V14 (F)	TCCTACAGGAAGGGCAAGCTGTTT
V14 (R)	ATCGATCCGAGGGCAACTGTGAAT
V16 (F)	TGCTGGTGTCATCCAAACACCTAG
V16 (R)	TTGGGCATCTGAGCTGAGAATCGT
V23 (F)	AAGGAGAGATTCTCAGCTGTGTGC
V23 (R)	TGACTGCTGGAGCACAAGTACAGT
V29 (F)	TGCTGGAATGTGGACAGGACATGA
V29 (R)	AGGGATGTCTCCTTCGCTGTTACT

Ptcra (F)	GTCAGGAGCACATCGAGCAGAAG
Ptcra (R)	CACACGCTGGTAGATGGAAGGC
Prss2 (exon1)	ACCATGAGTGCACTTCTGATCC
Prss2 (exon 2)	GGCAGGTGTATCCTCCAACA
Actb (F)	GGCTGTATTCCCCTCCATCG
Actb (R)	CCAGTTGGTAACAATGCCATGT
V1 CTCF (F)	AGGAAGATTGTGGGCAACTG
V1 CTCF (R)	AACCAAATAAACGGCAGCAC
V5 CTCF (F)	GCACTGCCAATCTCTGCAT
V5 CTCF (R)	CATTTCTTTCCCGTTCTCCA
V12-1 CTCF (F)	CAACGGGCAAAATTTGAGAT
V12-1 CTCF (R)	CTGCTCTGTTCTGGGTCTCC
V12-2 CTCF (F)	CCCCAGAAGCCTTATTTTGA
V12-2 CTCF (R)	GGGCTGCATATCAAAGCACT
V14 CTCF (F)	TCACCTATGGCCTCCTTGTC
V14 CTCF (R)	CCTGCTTGGCAAACTCTAGG
V29 CTCF (F)	AACCCTCCATCCCTTTCACT
V29 CTCF (R)	CTGGTTCCGTTTTTAATGGG
5’ PC (F)	CAGTGTTTGCCGACAGCTTA
5’ PC (R)	CACGCCTGGGTTTGTTTACT
u/s 5’PC (F)	CCATGAAGGGTGGAGTCAGT

u/s 5’PC (R)	CATAGCACCATGTCCACCAC
d/s 5’PC (F)	GGTGTAGTGGGTGGGTTTTG
d/s 5’PC (R)	GGCCCTAAGTGTGTTTGCTT
u/s Prss2 pro (F)	GGGGGAAAGACAGAAAAAGG
u/s Prss2 pro (R)	TTCCATGCCTATGTCCAACA
Prss2 promoter (F)	GGGAACTATAAAGACAGGCACTC
Prss2 promoter (R)	AGTGAAACTCACCAGCAGCTC
PDβ1 (F)	TCACCTTCCTTATCTTCAACTCCC
PDβ1 (R)	TCCCATAGAATTGAATCACCGTGG
Dβ1 (F)	AAGCTGTAACATTGTGGGGACAGG
Dβ1 (R)	CAATCTTGGCCTAGCAGGCTGCAG
PDβ2 (F)	TATGCTGAGCAGTTCTTCGGACCA
PDβ2 (R)	AGTCCTGGAAATGCTGGCACAAAC
3’ Eβ CTCF (F)	GTGTTTGGTGCCAGGAACAGA
3’ Eβ CTCF (R)	TGGTTACCTTGGCAACTGAGA

Chapter 4: Domain-Specific and Stage-Intrinsic Changes in *Tcrb* Conformation During Thymocyte Development

This paper is in review at the Journal of Immunology

4.1 Abstract

Considerable crosstalk exists between mechanisms controlling genome architecture and gene expression. Antigen receptor loci are excellent models for these processes because they are regulated at both conformational and transcriptional levels to facilitate their assembly by V(D)J recombination. Upon commitment to the DN stage of T cell development, *Tcrb* adopts a compact conformation that promotes long-range recombination between Vβ gene segments (*Trbv*s) and their DβJβ targets. Formation of a functional VβDβJβ join signals for robust proliferation of DN thymocytes and their differentiation into DP cells, where *Trbv* recombination is squelched (allelic exclusion). DP differentiation also is accompanied by decontraction of *Tcrb*, which has been thought to separate the entire *Trbv* cluster from DβJβ segments (spatial segregation-based model for allelic exclusion). However, DP cells also repress transcription of unrearranged *Trbv*s, which may contribute to allelic exclusion. We performed a more detailed study of developmental changes in *Tcrb* topology and found that only the most distal portion of the *Trbv* cluster separates from DβJβ segments in DP thymocytes, leaving most *Trbv*s spatially available for rearrangement. Preferential dissociation of distal *Trbv*s is independent of robust proliferation or changes in transcription, chromatin, or architectural factors, which are coordinately regulated across the entire *Trbv* cluster. Segregation of distal *Trbv*s also occurs on alleles harboring a functional VβDβJβ join, suggesting this process is independent of rearrangement status and is DP-intrinsic. Our finding that most *Trbv*s remain associated with DβJβ targets in DP cells revises allelic exclusion models from their current conformation-dominant to a transcription-dominant formulation.

4.2 Introduction

The assembly and expression of antigen receptor (AgR) genes is controlled at multiple levels, including chromatin accessibility, transcription, and changes in three-dimensional conformation (Bassing et al., 2002; Osipovich and Oltz, 2010; Krangel, 2009; Bossen et al., 2012). Because AgR genes span large regions of mammalian genomes, which are independently activated and repressed during lymphocyte development, these loci serve as excellent models to study long-range mechanisms that coordinate gene expression (Roy et al., 2011). In addition to gene expression, regulatory mechanisms shared among AgR loci orchestrate the process of V(D)J recombination (Bossen et al., 2012), which assembles exons encoding their variable regions from large arrays of Variable (V), Diversity (D), and Joining (J) gene segments. The availability of gene segments for recombination within each AgR locus is modulated during lymphocyte development to guide their ordered and tissue-specific assembly, as well as to enforce allelic exclusion, a process that ensures the production of only a single, functional allele for each AgR gene in most B and T cells (Brady, Steinel et al., 2010). Proper regulation of V(D)J recombination is essential for generating a diverse repertoire of AgRs, driving lymphocyte development, and avoiding chromosomal translocations that characterize many forms of leukemia and lymphoma (Gopalakrishnan et al., 2013; Choi et al., 2013; Zhang et al., 2012).

All AgR loci have at least one powerful enhancer element that activates promoters associated with the most proximal clusters of gene segments, driving high levels of their transcription and an extensive revision of their chromatin landscape (Cobb et al., 2006). The revised landscape includes deposition of the histone modification H3K4me3, an epigenetic mark recognized by RAG2 (Matthews et al., 2007; Liu et al., 2007), which, together with RAG1, are

lymphocyte-specific components of the V(D)J recombinase. Because the enhancer-proximal gene segments are decorated with such a high density of RAG proteins, these regions have been coined recombination centers (RCs) (Ji et al., 2010). As an example, the *Tcrb* locus contains a single enhancer, called Eβ, which is activated upon commitment of multipotent progenitors to the T cell lineage in the thymus (Yang et al., 2010; Bouvier et al., 1996). In turn, Eβ associates with promoters located in each of the proximal DβJβ clusters, triggering (i) robust transcription, (ii) chromatin accessibility at the un-rearranged gene segments, (iii) RAG-1/2 deposition, and (iv) Dβ-to-Jβ recombination, which occurs over short distances (Oestreich et al., 2006).

Complete assembly of variable region exons for *Tcrb*, *Tcrd*, and *Igh* requires a second round of recombination between joined DJ elements in the RC and one of many V segments splayed out over large genomic distances. Numerous studies have shown that the second, long-range V-to-DJ recombination event is facilitated by conformational changes at these loci (Kosak et al., 2002; Hewitt et al., 2010; Shih and Krangel, 2013). As an example, upon commitment to the double-negative (DN) stage of T cell development, the most distal ends of *Tcrb*, which are separated by >500 kb in the linear genome, come together in three-dimensional space, a process coined locus contraction (Skok et al., 2007; Schlimgen et al., 2008). *Tcrb* contraction coincides with the folding of its Vβ cluster into two spatially distinct domains, spanning proximal and distal portions of the *Trbv* array (Majumder et al., 2015). Each of the more compact *Trbv* domains also folds into the RC, presumably via the process of locus contraction, endowing at least most of the Vβ segments spatial access to DβJβ substrates (Gopalakrishnan et al., 2013; Majumder et al., 2015). Indeed, we have shown that *Trbv* usage is largely limited by the activities of their associated promoters rather than by their absolute proximity to the RC (Gopalakrishnan et al., 2013),

suggesting that all Vβ gene segments have crossed a spatial threshold required for RAG-mediated recombination.

Interactions between the more distal *Trbv* domain and the RC appear focused on a site called 5'PC, which binds the architectural protein CTCF and is located ~25 kb upstream of the Dβ1Jβ cluster (Majumder et al., 2015). Genome-wide studies have revealed that when CTCF is bound to pairs of sites with convergent orientations, CTCF-CTCF dimerization can generate structural loops (Phillips et al., 2009; Kim et al., 2007). In many cases, such chromosomal loops are stabilized via the association of CTCF dimers with cohesin, a ring-like complex that "locks" the loop bases into place (Parelho et al., 2008). Of note, the numerous CTCF sites scattered throughout both *Trbv* domains are all in the same orientation, which favors their association with the 5'PC site near the RC. A similar mechanism of long-range tethering appears to be at play for other AgR loci, with V segments forming distinct domains that harbor multiple CTCF sites in a convergent orientation relative to those near the RC (Guo et al., 2011; Lin et al., 2015). In what may be a related finding, ablation of CTCF or its key binding sites in AgR loci disrupts spatial interactions and long-range V(D)J recombination (Guo et al., 2011; Degner et al., 2015; Degner et al., 2011; Xiang et al., 2013; Seitan et al., 2011).

Although locus contraction promotes long-range recombination at nearly all AgR loci, this process is developmentally dynamic. For example, when DN thymocytes generate a productive *Tcrb* allele, pre-TCR signaling induces at least ten rounds of rapid cell division (Sicinska et al., 2003). These proliferating cells ultimately differentiate into the resting $CD4^+CD8^+$ (double-positive, DP) subset, in which distal ends of *Tcrb* separate spatially, presumably reverting to their

original “decontracted” state found in multipotent progenitors (Skok et al., 2007; Schlimgen et al., 2008). Spatial segregation of the *Trbv* cluster from the RC is thought to help enforce allelic exclusion (Skok et al., 2007; Schlimgen et al., 2008), disfavoring further long-range *Tcrb* recombination, which could generate two functional antigen receptor chains. Similar changes in contraction status have been observed at some (Roldán et al., 2005; Shih et al., 2010), but not all (Stadhouders et al., 2014), AgR loci during developmental transitions between precursor lymphocyte subsets.

Despite these advances, developmental changes in conformation have not been characterized for any AgR locus at a sufficient resolution to understand the precise nature of locus decontraction and its implications for allelic exclusion. We now use chromosome conformation capture technologies to probe architectural remodeling of *Tcrb* conformations during transition from the DN (contracted, V-to-DJ recombination active) to the DP stage of thymocyte development (decontracted, V-to-DJ recombination excluded). Remarkably, we show that decontracted *Tcrb* loci in DP thymocytes retain a close association between the RC and most of the *Trbv* cluster, with only the distal portion of the Vβ array dissociating from this interactome. Therefore, transcriptional repression of *Trbv*s, rather than their spatial separation from the RC, must be a dominant mechanism ensuring allelic exclusion of most Vβ segments in DP thymocytes. Dissociation of the distal *Trbv* domain, as well as locus decontraction, are independent of gross changes in CTCF/cohesin deposition and the massive proliferative burst that accompanies DN cell differentiation to the DP stage. Importantly, we find that, in DP thymocytes, dissociation of distal *Trbv* segments also occurs on alleles harboring a functional VβDβJβ rearrangement, which delete regions within *Tcrb* implicated as architectural determinants. Together, these findings indicate that

developmental changes in *Tcrb* conformation are Vβ domain-specific and are governed by DP-intrinsic mechanisms.

4.3 Materials and Methods

Mice. Thymocytes were harvested from $Rag1^{-/-}$/C57BL/6 mice directly (DN), after injection of α-CD3ε (DP) (Shinkai et al., 1994), or those expressing a *Tcrb* transgene (DP) (Shih et al., 2012; Shinkai et al., 1993). For certain experiments (indicated in Results), thymocytes were isolated from mice of the following genotypes, all of which were on a mixed 129SvEv/C57BL/6 background: $RAG1^{-/-}$:*Tcrb*: $Ccnd3^{-/-}$ (DN and DP), $V\beta1^{NT/NT}$:$Rag1^{-/-}$ (DP) (Brady, Oropallo et al., 2010), or $V\beta1^{NT/NT}$:$Lat^{-/-}$:$Rag1^{-/-}$ (DN) (Brady et al., 2011). All experiments were conducted on mice that were between 4 and 6 weeks of age. Animal procedures and experimental protocols were approved by the Institutional Animal Care and Use Committees at Washington University School of Medicine in St. Louis and Children's Hospital of Philadelphia.

Chromosome Conformation Capture (3C) assays. Chromosome conformation capture assays were performed precisely as described (Majumder et al., 2015) using 10^7 thymocytes or pro-B cells from $Rag1^{-/-}$ mice or from cultured 3T3 fibroblasts. Cross-linking efficiencies were measured using Taqman-qPCR assays with primers and probes shown in Table T12.

Circular Chromosome Conformation Capture coupled with high-throughput sequencing (4C-seq). 4C assays were performed with Hind III as the primary restriction enzyme for cutting cross-linked chromatin. The Hind III-digested DNA was ligated using the 3C protocol and resuspended in Buffer EB (100 µl, Qiagen). Secondary enzyme digestion was performed with Nla III (100U, overnight). After heat inactivation, the digested DNA was religated to generate circularized products of interaction partners, purified by phenol:chloroform extraction,

precipitated in isopropanol, resuspended in Buffer EB (100 µl), and quantified as described previously (Medvedovic et al., 2013). Inverse PCR was performed on the circularized DNA using primers within Hind III-Nla III fragments at Eβ or at *Trbv5* (see Table T12). Inverse PCR products were diluted 1:100 in TE buffer and used as templates for nested inverse PCRs (see Table T12), yielding the 4C DNA libraries.

Purified 4C DNA (100ng, PCR Purification Kit, Qiagen) was used for indexed library preparation. On average, eight indexed libraries were pooled, and subjected to 42 bp single-end sequencing according to the manufacturer's protocol (Illumina HiSeq2000, San Diego, CA). Sequence tags were aligned to the reference genome (build MM9) with Bowtie (Langmead et al., 2012). The r3C-seq package (Thongjuea et al., 2013) was used to calculate reads per million (RPM) for each sample and identify anchor interaction regions. EMBOSS (Rice et al., 2000) was used to generate a genome-wide map of the Hind III restriction fragments for assignment of reads. To compare between samples, RPM values for each fragment were quantile normalized. For visualization of the 4C-seq data, a running mean was calculated using a window size of three contiguous Hind III fragments (Medvedovic et al., 2013).

Public Data Sources. ChIP-seq data were obtained from the GEO repository for CTCF (accession number GSE41739, Shih et al., 2012), as well as H3K4me1 and H3K4me3 (accession number GSE55635, Pekowska et al., 2011).

3D-FISH. 3D-FISH assays were performed with *Tcrb* BAC probes spanning *Trbv1* (RP23-75P5) and Trypsinogen (RP23-203H5) precisely as described (Majumder et al., 2015). Imaged data were analyzed using ImageJ as described (Majumder et al., 2015).

Chromatin Immunoprecipitation (ChIP) assays. ChIP assays were performed as described previously (Majumder et al., 2015) using the following antibodies: CTCF (Rockland, 600-401-C42), RAD21 (Abcam, ab992), H3Ac (EMD Millipore, 06-599) H3K4me2 (Abcam, ab32356) and IgG (Santa Cruz Biotechnology, Inc., sc-2027). ChIP assays were analyzed by q-PCR with primer combinations shown in Table T12.

RNA extraction and germline transcription. RNA was prepared from 0.5-1 million cells using TRIzol (Invitrogen). cDNA was generated from 1 µg of RNA using MuLV Reverse Transcriptase (NEB), and analyzed by q-PCR with primer combinations provided in Table T12.

4.4 Results

Only distal Trbv segments dissociate from the recombination center in DP thymocytes

The folding of *Tcrb* into its active conformation in DN thymocytes is accompanied by overall compaction of the locus (Skok et al., 2007; Schlimgen et al., 2008; Majumder et al., 2015), as measured by 3D-FISH, and a robust association of the RC with the 500 kb *Trbv* cluster (Gopalakrishnan et al., 2013). Expression of TCRβ protein and subsequent pre-TCR signaling drives the DN to DP thymocyte transition, as well as repression of additional Vβ-to-DβJβ recombination, enforcing allelic exclusion (Brady, Steinel et al., 2010). Repression of *Tcrb* gene assembly in DP cells correlates with significantly attenuated expression of un-rearranged *Trbv* segments (Senoo and Shinkai, 1998; Tripathi et al., 2002) and spatial segregation of the extreme 5' and 3' ends of the locus; i.e., decontraction (Skok et al., 2007; Schlimgen et al., 2008; Majumder et al., 2015). The latter cell imaging data have been extrapolated into models suggesting that *Tcrb* reverts to an extended conformation in DP thymocytes, with a loss of all RC-*Trbv* interactions (Cobb et al., 2006; Skok et al., 2007).

To study developmental changes in *Tcrb* conformation at a higher resolution, we performed 4C-seq, an approach that allows us to measure the cross-linking efficiency of a given restriction fragment (viewpoint) with the entire genome. The 4C-seq assays were performed with cross-linked chromatin from RAG-deficient DN thymocytes or RAG-deficient animals injected with α-CD3ε antibody, which drives differentiation into the DP stage (Shinkai et al., 1994). Both DN and DP thymocytes from these mice retain *Tcrb* gene segments in their germline configuration due to a lack of recombinase activity (Shinkai et al., 1994). As shown in Figure 4.1A, the *Tcrb* enhancer

region (Eβ) associates with the entire *Trbv* cluster in DN thymocytes, which is consistent with prior findings (Gopalakrishnan et al., 2013; Skok et al., 2007; Majumder et al., 2015). Surprisingly, many of these long-range interactions are maintained in DP thymocytes, despite *Tcrb* decontraction, as ascertained by 3D-FISH using probes for the 5' and 3' ends of the locus (Figure 4.1B). Notably, only the distal region of the *Trbv* cluster appears to dissociate uniformly from this RC viewpoint upon transition to the DP stage, suggesting that decontraction does not involve the entire *Tcrb* locus in DP thymocytes.

To further validate the compartmentalized changes in *Tcrb* conformation during thymocyte development, we performed 3C-qPCR to focus on specific interactions between pairs of restriction fragments. We find that Eβ associates with *Trbv* segments spread over the proximal half of the cluster equally in DN and DP thymocytes (Figure 4.1C). In contrast, Eβ association with the more distal *Trbv* segments diminishes significantly in DP cells (Figure 4.1C). The changes in long-range *Trbv*-RC interactions, as well as reversal of locus contraction, are recapitulated in RAG-deficient DP thymocytes when their development is driven by the expression of a *Tcrb* transgene (Figures 4.1B and 4.1D) (Shih et al., 2012; Shinkai et al., 1993). The latter data preclude the possibility that super-physiologic signaling by α-CD3ε is responsible for our observations.

3C analyses using complementary viewpoints in the *Trbv* cluster further confirmed our findings from 4C-seq. For example, robust cross-linking between the distal *Trbv5* region and two restriction fragments in the DβJβ cluster is observed in DN thymocytes, but is reduced to near background levels in DP cells (Figure 4.1E, left). As expected, *Trbv5* association with 5'PC, the CTCF binding element that serves as a tether for long-range interactions between the RC and distal

Vβ-region (Majumder et al., 2015), is also lost in DP thymocytes (Figure 4.1E, left). In sharp contrast, associations between the more proximal *Trbv23* and the 5'PC-RC portion of the locus are unaffected by the DN to DP transition (Figure 4.1E, right), even though these regions are separated by greater than 200 kb in the linear genome. Preferential dissociation of the distal *Trbv* region from the RC does not correlate with domain-specific changes in transcription, which is reduced in DP thymocytes for Vβ segments within both the proximal and distal domains (Figure 4.1F). Similarly, histone modifications associated with active promoters (H3K4me3), or simply open chromatin (H3K4me1) (Pekowska et al., 2011), are coordinately regulated when comparing ChIP-seq data from DN and DP thymocytes (Figure 4.1G). We conclude that dissociation of Vβ segments from the RC is highly restricted to the distal region of the *Trbv* cluster and is independent of changes in transcription and chromatin, which occur over the entire *Trbv* cluster.

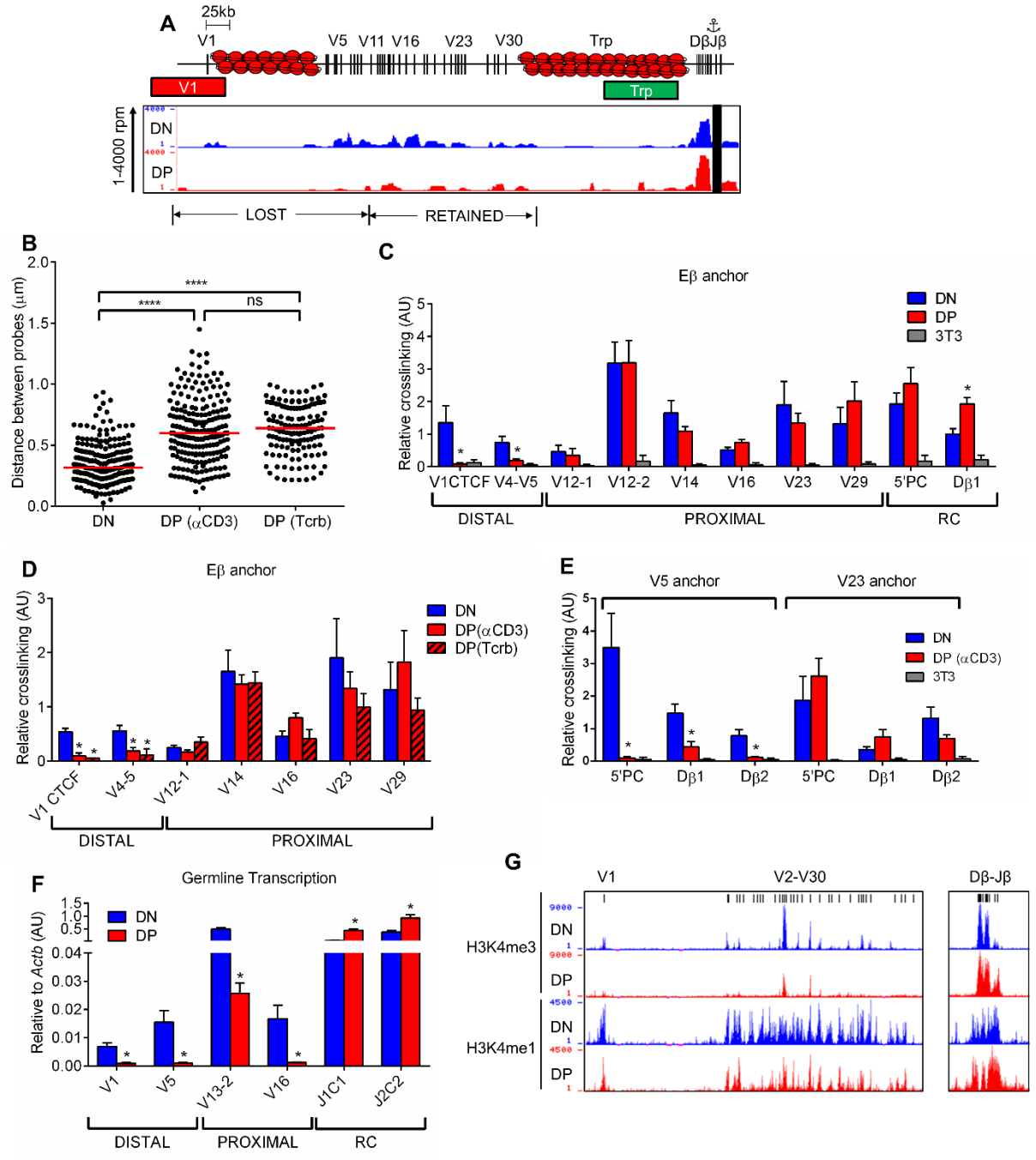


Figure 4.1: Specific dissociation of distal *Trbv* segments in DP thymocytes. (A) Top: Schematic of mouse *Tcrb* showing inactive trypsinogen regions as red nucleosomes. Bottom: Long-range interaction of Eβ viewpoint in RAG1$^{-/-}$ DN thymocytes (blue) compared with RAG1$^{-/-}$ DP thymocytes (red), which were generated by α-CD3ε injections. These data represent an average of three independent experiments with values reflecting running means of three adjacent fragments as described (40). Location of the Eβ viewpoint is denoted by a black bar, as well as an anchor symbol above the cartoon. The y-axis represents normalized reads per million (rpm, see Materials and Methods for

details) with a scale of 1 to 4000. **(B)** 3D-FISH assays for *Tcrb* contraction using probes spanning the *Trbv1* and trypsinogen gene cluster (shown as red and green bars, respectively, in Fig. 1A). Data are presented for thymocytes from RAG1$^{-/-}$ animals (DN) or RAG1$^{-/-}$ mice treated with anti-CD3ε or expressing a *Tcrb* transgene (DP). Results are presented as scatter plots of distances between probe foci and represent three independent preparations of slides. Significant differences in the 3D-FISH assays are denoted as ****, P<0.00001 (one-way ANOVA, Tukey's post hoc test), and ns represents non-significant differences in median distances. **(C)** 3C analysis was performed with the Eβ viewpoint in DN and DP thymocytes (α-CD3ε). NIH-3T3 cells (grey) serve as controls for background levels of cross-linking to the indicated upstream *Trbv* gene segments. For all 3C panels, data are presented as average of three independent experiments ± SEM and relative cross-linking values were calculated as described (Majumder et al., 2015). Significant differences between DN and DP thymocytes in all 3C-qPCR assays are denoted as *, P<0.05 (Student's t test). **(D)** 3C analysis using the Eβ viewpoint to compare DP thymocytes from RAG1$^{-/-}$ mice generated with either α-CD3ε treatment (solid red) or expression of a *Tcrb* transgene (red with black diagonals). **(E)** 3C-analysis was performed in DN, DP, and 3T3 cells using a *Trbv5* (distal domain) or a *Trbv23* viewpoint (proximal domain). **(F)** Germline transcription of the indicated regions was quantified relative to mRNA levels for *Actb* in DN and DP thymocytes as described previously (Majumder et al., 2015). All qPCR values are presented as mean ± SEM of at least three independent experiments. Significant differences between DN and DP thymocytes are denoted as *, P<0.05 (Student's t test). **(G)** Genome browser tracks of ChIP-seq data showing levels of H3K4me3 and H3K4me1 on *Tcrb* in DN (blue) and DP (red) thymocytes. ChIP-seq data were obtained from a published study (Pekowska et al., 2011).

Trbv domains remain folded in DP thymocytes

In DN thymocytes, distal and proximal regions of the *Trbv* cluster fold into separate spatial domains that associate with the RC in 5'PC-dependent and -independent manners, respectively (Majumder et al., 2015). Additional 4C experiments using the distal *Trbv5* region as a viewpoint revealed that its association with other distal Vβ segments is retained in DP thymocytes (Figure 4.2A). However, *Trbv5* becomes spatially segregated from *Trbv* segments in the RC-proximal portion of the cluster during this developmental transition (Figure 4.2A).

3C data using distal and proximal *Trbv* viewpoints confirmed that each domain retains its internal, folded conformation, but only the distal Vβ domain strays from the RC in DP thymocytes (Figure 4.1E and 4.2B). For example, association between *Trbv5* and *Trbv1* within the distal domain is unchanged in DN versus DP cells, while interaction between *Trbv5* (distal) and *Trbv12-2* (proximal) diminishes significantly (Figure 4.2B, left), despite a nearly equal linear distance between *Trbv5* and each of these two gene segments. Conversely, *Trbv23*, which is in the proximal domain, cross-links with equal efficiencies to the proximal *Trbv12-2* segment in DN and DP cells, but with a reduced efficiency to the distal *Trbv1* segment in DP thymocytes (Figure 4.2B, right).

These conformational data indicate that the developmental transition of DN to DP thymocytes disrupts long-range interactions between the RC and the distal Vβ domain, which remains in a folded conformation. In contrast, RC interactions with the folded cluster of proximal *Trbv* segments are fully retained upon differentiation to the DP stage. We validated our 4C-seq and 3C-qPCR data using 3D-FISH assays using probes which recognized the proximal (*Trbv12-30*), distal (*Trbv1*) and the RC region of *Tcrb*. Consistent with our 3C analyses, *Trbv1* spatially

dissociated from *Trbv12-30* upon DN to DP transition (Figure 4.3, left) but *Trbv12-30* remained spatially proximal to the RC (Figure 4.3, right). Thus, we conclude that repression of recombination at proximal *Trbv* segments in DP cells is independent of spatial dissociation from their target substrates in the DβJβ cluster.

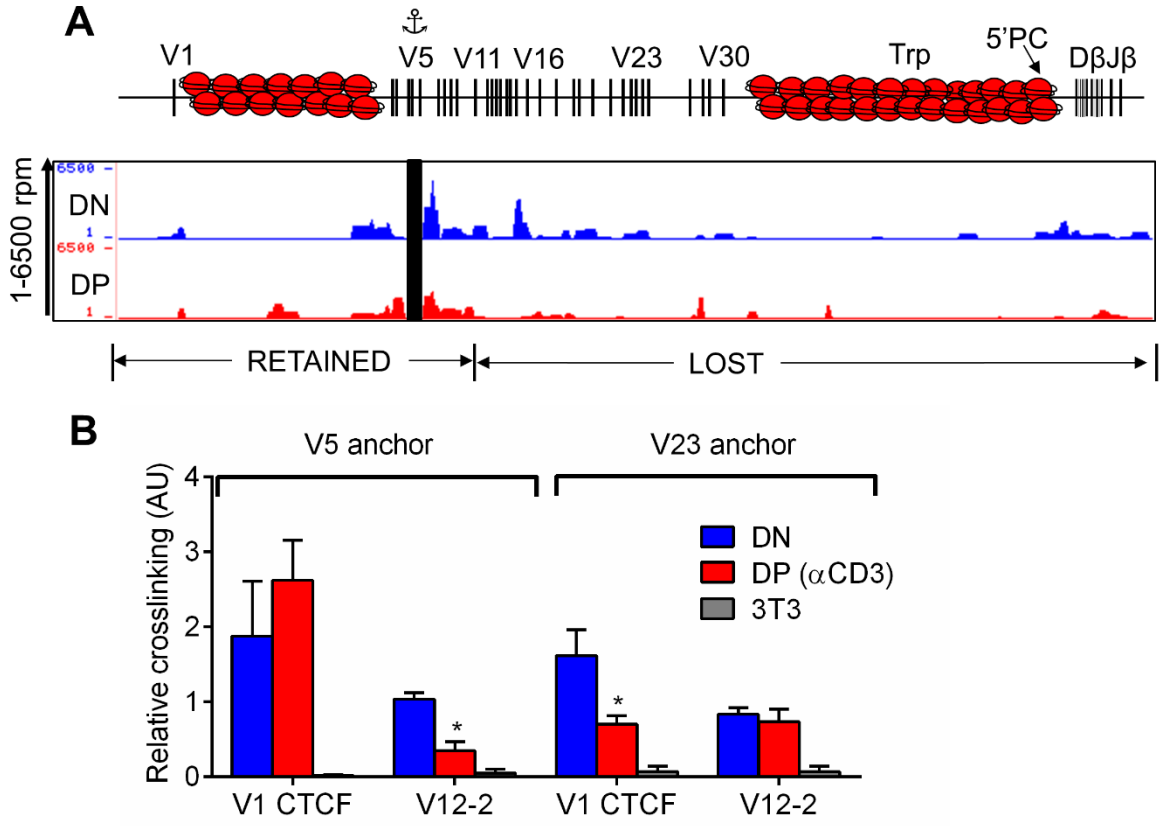


Figure 4.2: Interactions within each *Trbv* domain are unchanged in DN and DP thymocytes. (A) Top: Schematic of *Tcrb*, as described for Fig. 1A. Bottom: Long-range interactions using a *Trbv5* viewpoint (anchor symbol in *Tcrb* schematic). The data are presented as an average of three independent experiments, with y axes (rpm) ranging from 1 to 6500. **(B)** 3C analysis was performed with *Trbv5* and *Trbv23* viewpoints in DN and DP thymocytes, using 3T3 fibroblasts as negative controls.

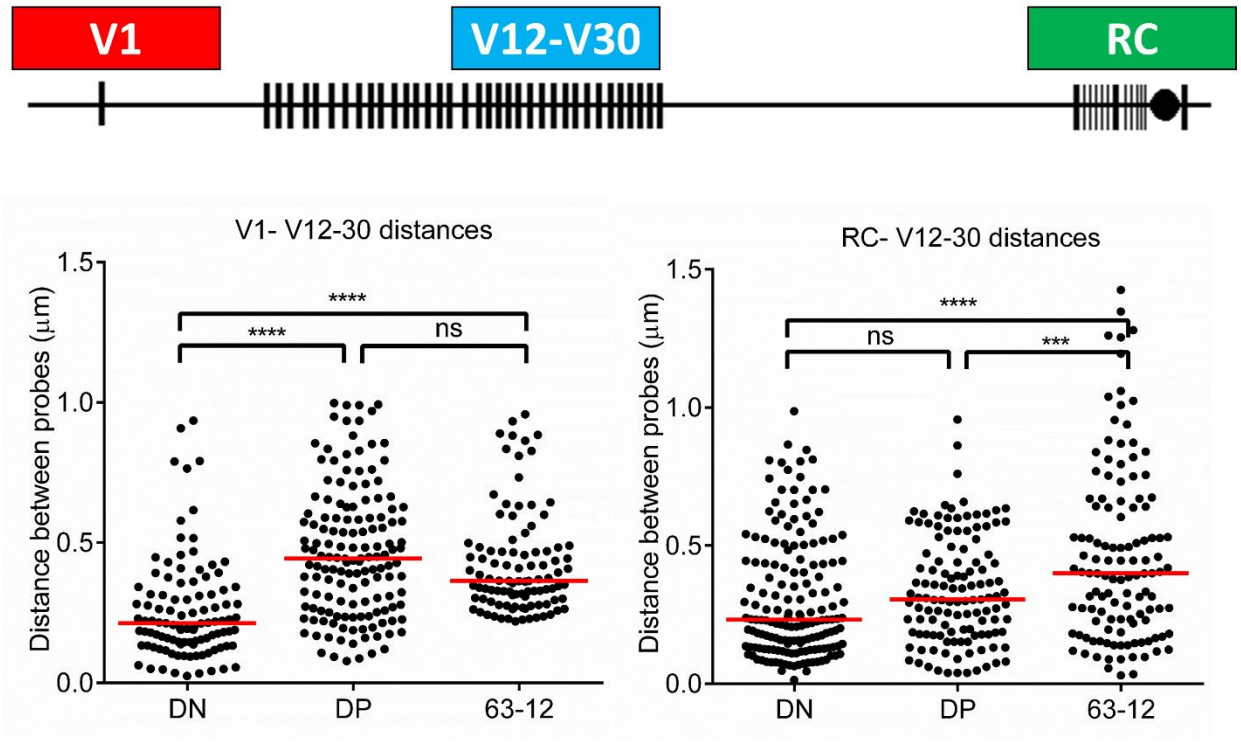


Figure 4.3: *Tcrb* segregates into two *Trbv* domains upon DN to DP transition. 3D-FISH assays were performed to compute the spatial distance between *Trbv1* and the distal Vβ region, spanning *Trbv12-30* (left), and from the distal Vβ to the RC (right), as shown in the top schematic, in the indicated cell types. The RAG2-deficient pro-B cell line, 63-12, was used as a negative control for *Tcrb* contraction. Results are presented as scatter plots of distances between probe foci and represent two independent preparations of slides. Significant differences in the 3D-FISH assays are denoted as ****, P<0.00001, ***, P<0.0001, (one-way ANOVA, Tukey's post hoc test), and ns represents non-significant differences in median distances.

Binding of architectural proteins and chromatin boundaries are largely unaltered in DN and DP thymocytes

The architectural protein CTCF facilitates the formation of structural loops in metazoan genomes, including those found in AgR loci (Phillips et al., 2009; Kim et al., 2007). Many loops are stabilized via the association of CTCF with the ring-like cohesin complex, which is thought to ensnare the bases of chromatin loops (Parelho et al., 2008). To determine whether the loss of distal *Trbv*–RC contacts in DP thymocytes is mechanistically related to reduced binding of supporting architectural complexes, we performed ChIP-qPCR assays using chromatin from DN ($RAG1^{-/-}$) and DP subsets ($RAG1^{-/-}$:α-CD3ε). CTCF binding sites in *Tcrb* are shown in Figure 4.4A (top) as established by ChIP-seq data from RAG-deficient DN thymocytes (Shih et al., 2012). Remarkably, levels of CTCF and the cohesin subunit RAD21 are not altered significantly in DN and DP thymocytes at the vast majority of tested sites, including those in distal and proximal *Trbv* domains, as well as RC flanking sites (Figure 4.4A, middle and bottom). The one exception is a modest loss of CTCF at the 5'PC site, which functions as a tether for the distal *Trbv* domain in DN cells. However, levels of RAD21 at this tethering site remain unchanged in DP when compared with DN thymocytes.

We have recently shown that the tethering function of 5'PC can be compromised, despite retention of CTCF and RAD21 binding, by deletion of a chromatin boundary element (BE) located ~25 kb upstream of the RC (ΔPDβ1, Figure 4.4C, top) (Majumder et al., 2015). Removal of the natural BE permits a spread of highly active chromatin from the RC to 5'PC, which becomes a new boundary and, in some manner, acquisition of this function compromises its ability to serve

as a long-range tether (Majumder et al., 2015). Thus, one potential mechanism for specific dissociation of distal *Trbv* segments would be inactivation of the RC-proximal chromatin boundary in DP thymocytes, which would in turn disarm the 5'PC tether.

Initially, we tested this possibility by monitoring RNA expression from the normally silent trypsinogen gene, *Prss2*, which is activated in DN thymocytes upon deletion of the RC-proximal BE (Figure 4.4B). However, *Prss2* remains transcriptionally silent in DP thymocytes, suggesting that RC-proximal boundary function remains intact in these cells. This conclusion is bolstered by ChIP assays that monitored the spread of active chromatin upstream of the RC. The invasion of active chromatin is significant in DN cells lacking the native BE (ΔPDβ1), but not in DP thymocytes (Figure 4.4C). Thus, dissociation of distal *Trbv* segments in DP thymocytes is not simply due to a loss of CTCF-cohesin within this structural domain, nor is it due to a disruption of chromatin boundary function upstream of the RC, which would impair distal *Trbv* tethering to 5'PC.

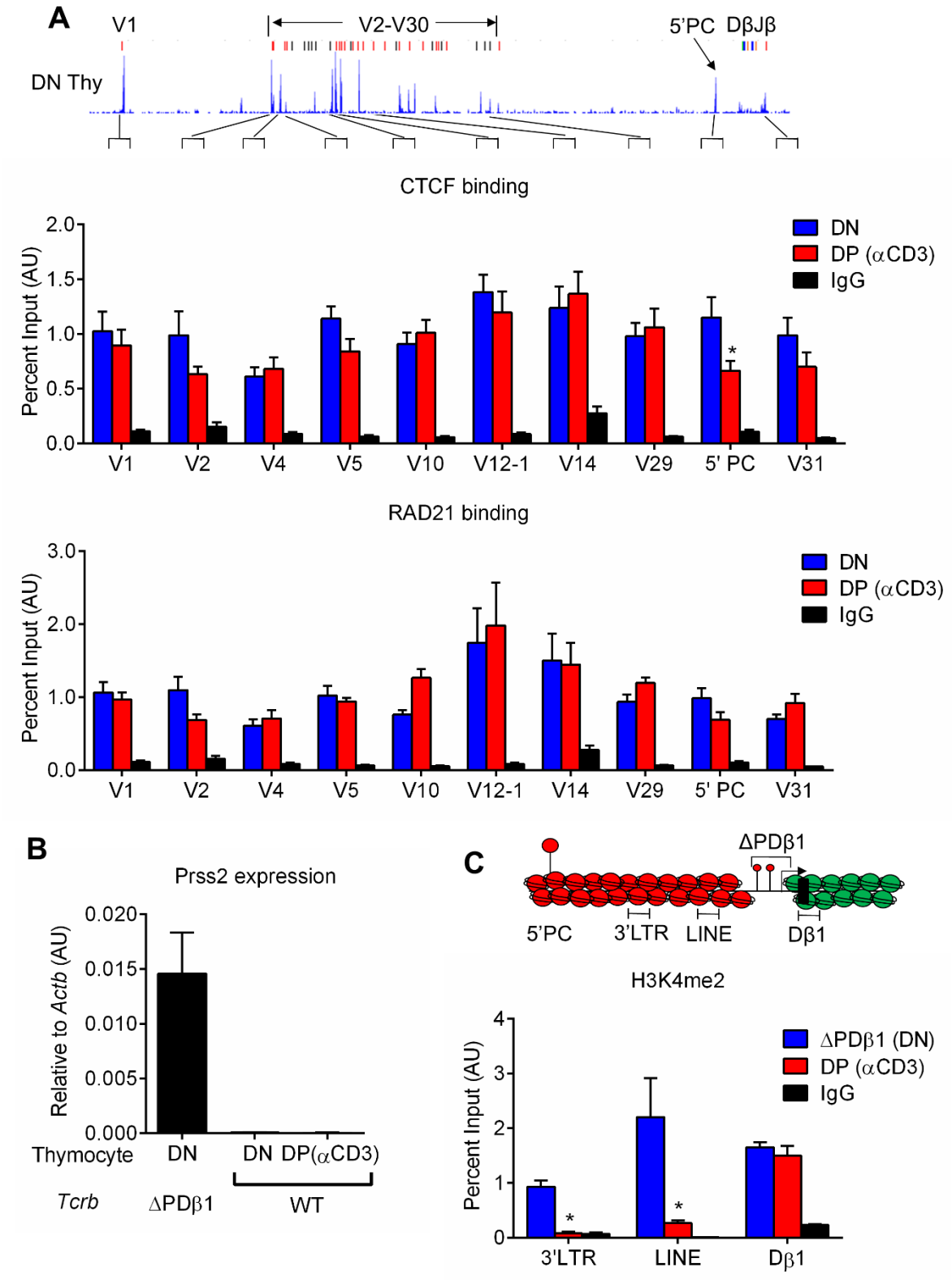

A
V1
V2-V30
5'PC
DβJβ
DN Thy
CTCF binding
Percent Input (AU)
DN
DP (αCD3)
IgG
V1
V2
V4
V5
V10
V12-1
V14
V29
5' PC
V31
RAD21 binding
B
Prss2 expression
Relative to Actb (AU)
Thymocyte
DN
DN DP(αCD3)
Tcrb
ΔPDβ1
WT
C
ΔPDβ1
5'PC
3'LTR
LINE
Dβ1
H3K4me2
ΔPDβ1 (DN)

Figure 4.4: Architectural proteins and chromatin boundaries are retained in DN and DP thymocytes. (A) Top: Published ChIP-seq profile for CTCF binding in RAG-deficient DN thymocytes (Shih et al., 2012). ChIP-qPCR data for CTCF and RAD21 occupancy (middle and bottom, respectively) are shown for DN versus DP thymocytes. Background levels of signal were determined using total IgG as a ChIP antibody (black bar). Data are presented as mean ± SEM of percent input signal from three independent experiments. Statistically significant differences between DN and DP samples are represented as *, $P<0.05$ (Student's t test). **(B)** *Prss2* expression was quantified relative to *Actb* using RT-qPCR assays in the indicated cell types (Majumder et al., 2015). Mean values from three-independent experiments are shown (±SEM). **(C)** Top: Schematic of the barrier region upstream of *Tcrb*-RC is shown with red nucleosomes denoting repressive chromatin and green nucleosomes representing active chromatin. Locations of qPCR primers used for H3K4me2 ChIP are shown under the cartoon. Bottom: ChIP-qPCR data for H3K4me2, presented as percent input chromatin (±SEM) from three independent experiments. Statistically significant differences between ΔPDβ1 and DP samples are represented as *, $P<0.05$ (Student's t test).

Stage-specific changes in Tcrb conformation are independent of the DN to DP proliferative burst

Prior studies have shown that a majority of structural loops within the genome are disrupted during mitosis and reform in resting daughter cells (Naumova et al., 2013). Developmental progression of DN thymocytes to the DP stage is associated with a robust proliferative burst. Indeed, a TCRβ^{+} DN thymocyte, on average, undergoes 10-11 rounds of division before coming to rest at the DP stage (Sicinska et al., 2003). Accordingly, we hypothesized that changes in *Tcrb* conformation during the DN to DP transition may require this robust proliferation, which would dissolve the DN architecture and allow its reconfiguration into the DP conformation.

To test this hypothesis, we used mice that lacked the gene encoding a CDK4/6 regulatory subunit, Cyclin D3 (*Ccnd3*$^{-/-}$), a defect that severely compromises thymocyte proliferation (Sicinska et al., 2003). When RAG1$^{-/-}$:*Tcrb* transgenic mice are crossed into the *Ccnd3*-deficient background (RAG1$^{-/-}$:*Ccnd3*$^{-/-}$:*Tcrb*), thymocytes progress to the DP stage of development without the normal proliferative burst (Sicinska et al., 2003). These cellular defects are highlighted in Figure 4.5A. Compared with RAG1$^{-/-}$:*Tcrb* mice, in which the TCRβ-driven proliferative burst generates large numbers of DP thymocytes (1.2 X 10^{8}) (Sicinska et al., 2003, Shih et al., 2012), RAG1$^{-/-}$:*Ccnd3*$^{-/-}$:*Tcrb* thymuses also contain primarily DP cells, but at dramatically reduced numbers (1.5 X 10^{7} cells). When probing endogenous *Tcrb* loci, RAG1$^{-/-}$:*Ccnd3*$^{-/-}$:*Tcrb* thymocytes exhibit no defects in short-range interactions between Eβ and the Dβ1 region compared with their proliferation-competent RAG1$^{-/-}$:*Tcrb* counterparts (Figure 4.5B). Importantly, preferential dissociation of the distal *Trbv* domain is unaffected in *Ccnd3*$^{-/-}$

thymocytes when these interactions are measured from two independent viewpoints within the RC (Figures 4.5C and 4.5D).

We also examined global conformational changes at *Tcrb* in RAG1$^{-/-}$:*Ccnd3*$^{-/-}$:*Tcrb* thymocytes using 3D-FISH. We first sorted DN and DP cells from these animals to remove contaminating CD8^{+} cells (Figure 4.5A), which represent an intermediate between the DN and DP stages (Serwold et al., 2007). Fixed cells were hybridized to fluorescent probes for regions at the very 5' end of *Tcrb* and near the RC (Figure 4.5B, top), with probe separation used as a metric for locus contraction. As shown in Figure 4.5E, *Tcrb* is contracted in RAG1$^{-/-}$:*Ccnd3*$^{-/-}$:*Tcrb* DN thymocytes compared with DP cells. In fact, locus decontraction is statistically indistinguishable in DP thymocytes derived from *Ccnd3*-deficient and –sufficient animals. Together, the low- (FISH) and high-resolution (3C) data indicate that dissociation of the distal *Trbv* and RC domains is independent of the massive proliferative burst that precedes DP thymocyte differentiation and, instead, may be mediated by DP-intrinsic mechanisms.

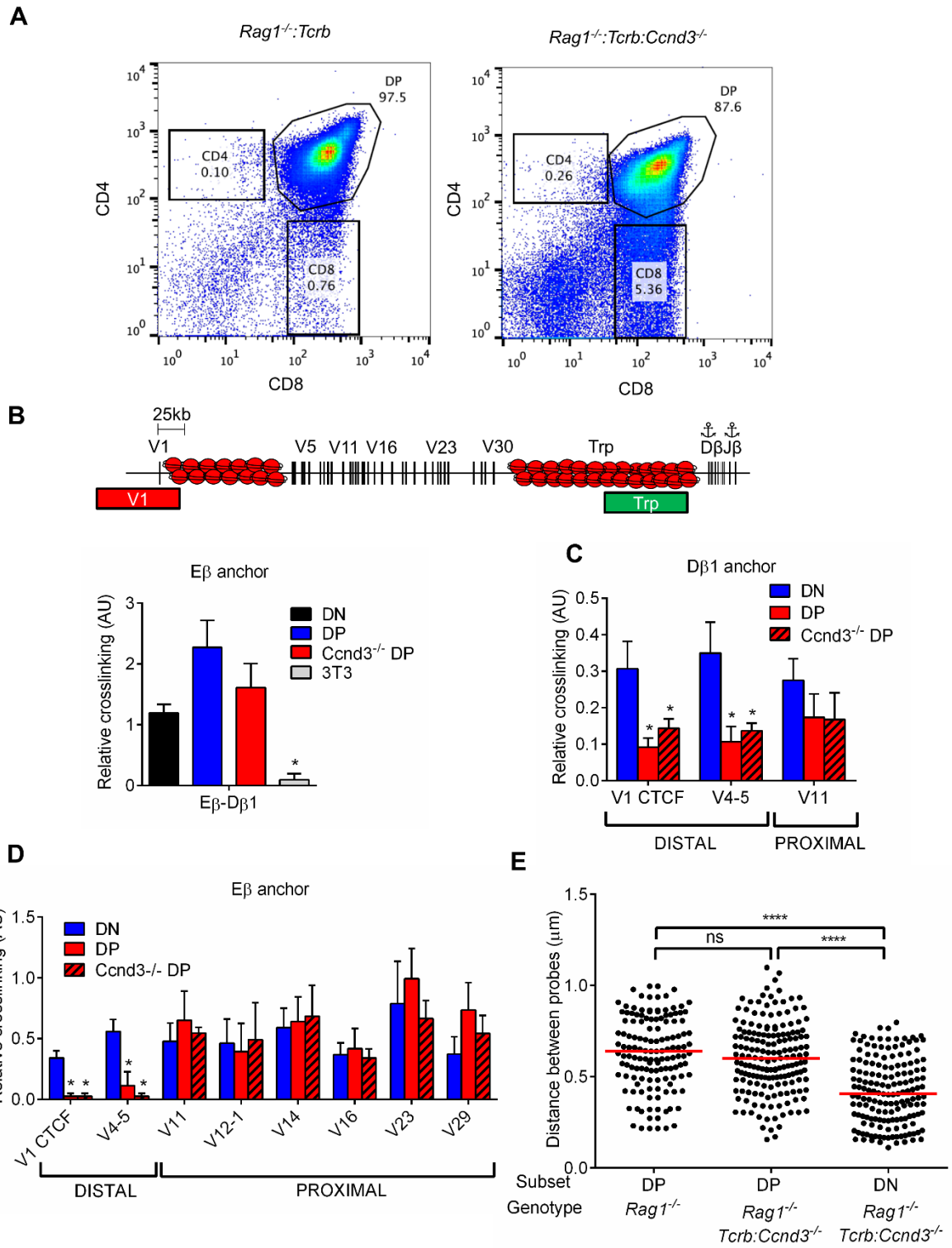

A
Rag1-/-:Tcrb
Rag1-/-:Tcrb:Ccnd3-/-
DP 97.5
CD4 0.10
CD8 0.76
DP 87.6
CD4 0.26
CD8 5.36
CD4
CD8
B
25kb
V1
V5 V11 V16
V23
V30
Trp
DβJβ
V1
Trp
Eβ anchor
Relative crosslinking (AU)
DN
DP
Ccnd3-/- DP
3T3
Eβ-Dβ1
C
Dβ1 anchor
Relative crosslinking (AU)
DN
DP
Ccnd3-/- DP
V1 CTCF
V4-5
V11
DISTAL
PROXIMAL
D
Eβ anchor
Relative crosslinking (AU)
DN
DP
Ccnd3-/- DP
V1 CTCF
V4-5
V11
V12-1
V14
V16
V23
V29
DISTAL
PROXIMAL
E
Distance between probes (μm)
ns
****
****
Subset
Genotype
DP
Rag1-/-
DP
Rag1-/-
Tcrb:Ccnd3-/-
DN
Rag1-/-
Tcrb:Ccnd3-/-

Figure 4.5: Alterations in *Tcrb* conformation do not require extensive thymocyte proliferation. (A) Representative FACS analysis of thymocytes from mice with the indicated genotypes. Cells were stained with CD4-PE and CD8-FITC. A representative of duplicate experiments is shown with percentages of thymocyte subsets highlighted. **(B)** 3C analysis for short-range interactions between the Eβ viewpoint and Dβ1 region. **(C, D)** Long-range interactions between either the Dβ1 (panel C) or Eβ (panel D) viewpoint and gene segments within the proximal (V11) and distal (V1 and V4-5) domains of the *Trbv* cluster. Thymocytes from RAG1-deficient (DN), *Rag1*$^{-/-}$:*Tcrb* (DP), and *Rag1*$^{-/-}$:*Tcrb*:*Ccnd3*$^{-/-}$ (DP) mice were used for this analysis. **(E)** 3D-FISH with *Trbv1* and *Trp* probes (see Fig. 4B, schematic) was used to determine the contraction status of *Tcrb* in DN or DP cells from the indicated genotypes (see Fig. 1B for details on statistical analyses). All mice were deficient for RAG1.

The hinge region for dissociation of distal Trbvs is unaltered epigenetically and transcriptionally during thymocyte development

To gain a better understanding of mechanisms that control specific dissociation of the distal *Trbv* domain in DP thymocytes, we used 3C to map the inflection point between retained and lost RC interactions. Our initial 3C analyses (Figure 4.1C) suggested that the transition occurred in a 40 kb window upstream of *Trbv12-1* (retained RC interaction) and *Trbv5* (lost RC interaction). A more refined 3C "walk", using Eβ as the viewpoint, revealed a transition between two adjacent restriction fragments, which have similar (*Trbv11*) or substantially diminished (5'V11) interaction with Eβ when comparing DN to DP thymocytes (Figure 4.5A, left panel). All additional restriction fragments tested upstream of 5'V11, including those spanning distal *Trbv* segments, exhibit diminished association with Eβ in DP relative to DN cells (Figures 4.1A, 4.1C and 4.6A). The topological transition in DP thymocytes is also evident when two independent viewpoints are used near the RC (Dβ2 and 5'PC, Figure 4.6B, right panel). Importantly, complementary 3C experiments, using viewpoints at this topological transition within the *Trbv* cluster, supported our conclusions. Association between a restriction fragment spanning *Trbv11* and a fragment within the RC is indistinguishable in DN and DP thymocytes (Figure 4.6C, right), whereas 3C data from the 5'V11 viewpoint reflects its dissociation from the RC during this developmental transition (Figure 4.6C, left). Therefore, the region located directly upstream of *Trbv11* serves as a "hinge" for developmentally regulated dissociation of distal Vβ gene segments from the RC in DP thymocytes.

We next explored potential mechanisms controlling this developmental hinge, focusing on specific changes to chromatin or function that may occur in this region during the DN to DP transition. First, binding of CTCF and RAD21 to a site nearest the hinge region is unaltered in DP compared with DN thymocytes (Figure 4.6D). Second, transcription of two LINE repetitive elements located 3' of *Trbv11* is repressed in DP thymocytes, which is similar to reduced expression of all germline *Trbv* segments in these cells (Figure 4.6E). Third, like transcription, histone acetylation at sites within the hinge region also diminishes during transition to the DP subset (Figure 4.6F). Thus, the hinge region exhibits no obvious epigenetic or transcriptional characteristics distinguishing it from developmental changes that occur over the entire *Trbv* cluster.

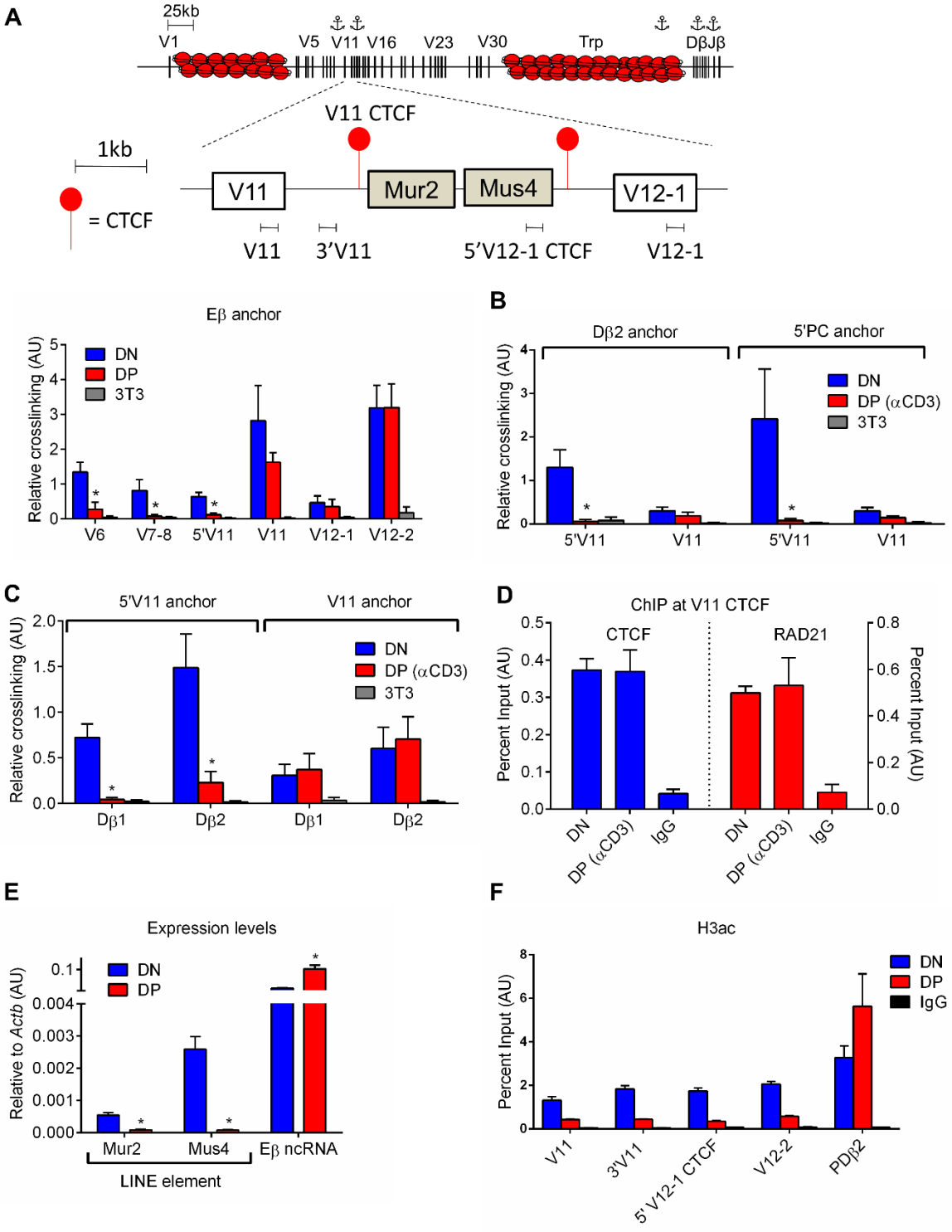

A
25kb
V1
V5
V11
V16
V23
V30
Trp
DβJβ
V11 CTCF
1kb
= CTCF
V11
Mur2
Mus4
V12-1
V11
3'V11
5'V12-1 CTCF
V12-1
Eβ anchor
Relative crosslinking (AU)
DN
DP
3T3
V6
V7-8
5'V11
V11
V12-1
V12-2
B
Dβ2 anchor
5'PC anchor
Relative crosslinking (AU)
DN
DP (αCD3)
3T3
5'V11
V11
5'V11
V11
C
5'V11 anchor
V11 anchor
Relative crosslinking (AU)
DN
DP (αCD3)
3T3
Dβ1
Dβ2
Dβ1
Dβ2
D
ChIP at V11 CTCF
CTCF
RAD21
Percent Input (AU)
Percent Input (AU)
DN
DP (αCD3)
IgG
DN
DP (αCD3)
IgG
E
Expression levels
Relative to Actb (AU)
DN
DP
Mur2
Mus4
Eβ ncRNA
LINE element
F
H3ac
Percent Input (AU)
DN
DP
IgG
V11
3'V11
5' V12-1 CTCF
V12-2
PDβ2

Figure 4.6: Identification and analysis of the hinge region between distal and proximal *Trbv* domains. (A-C) 3C data for Eβ (panel A, bottom), Dβ2 (panel B, left), 5'PC (panel B, right), 5'V11 (panel C, left), and V11 (panel C, right) viewpoints using the indicated cell types. See Fig. 1C for details. **(D)** ChIP-qPCR data for CTCF and RAD21 deposition at a site positioned closest to the *Trbv11* hinge region as indicated in Fig. 5A schematic. **(E)** Expression of LINE elements (Mur2 and Mus4) situated between *Trbv11* and *Trbv12-1* were calculated relative to *Actb* using RT-qPCR in DN and DP thymocytes. Non-coding transcripts generated near the Eβ enhancer were used as an independent control. **(F)** Histone acetylation at the indicated regions as measured by ChIP-qPCR.

Stage-specific dissociation of distal Trbv segments on functionally rearranged alleles

Our data suggest that, on germline *Tcrb* alleles in RAG-deficient mice, the distal *Trbv* domain separates spatially from the RC in DP thymocytes via mechanisms independent of 5'PC function. During normal T cell development, the DN to DP transition is driven by functional *Tcrb* rearrangements, which are usually restricted to a single allele in each cell (Brady, Steinel et al., 2010). All long-range recombination events between upstream Vβ segments and the RC will delete 5'PC. Moreover, a large subset of these rearrangements also delete the inflection point for dissociation of distal Vβ segments from the RC, positioned upstream of *Trbv11*. Thus, examination of *Tcrb* conformation on a more physiologic allele, harboring a distal *Trbv* rearrangement, would provide an independent test for whether 5'PC or the 5'V11 hinge region are involved in stage-specific separation of distal *Trbv* segments.

For this purpose, we performed 3C analyses on thymocytes from a mouse strain that harbors two functional Vβ5Dβ1Jβ1.4 alleles in their germline (Figure 4.7A), termed Vβ1NT mice (Brady, Oropallo et al., 2010). While the remaining *Trbv*s on this allele can rearrange to Dβ2Jβ segments in DN thymocytes, recombination of these upstream Vβ segments is repressed in DP cells (Brady and Bassing, 2011). In the $RAG1^{-/-}$ genetic background, the Vβ1NT allele drives developmental progression of thymocytes to the DP stage (Brady and Bassing, 2011). Thus, to assess *Tcrb* conformation in DN thymocytes, pre-TCR signaling was crippled in $RAG1^{-/-}$:Vβ1NT mice by making them homozygous for a null mutation in the LAT adaptor molecule (Brady, Oropallo et al., 2010, Brady and Bassing, 2011). As shown in Figure 4.7A, the Eβ enhancer associates not only with the rearranged Vβ5Dβ1Jβ1.4 exon in DN cells, but also with the

remaining un-rearranged *Trbv* segments, including *Trbv1* situated ~150 kb upstream. Enhancer interactions with the *Trbv* segments are cell type-specific since they are diminished significantly in pro-B cells from the RAG1$^{-/-}$:LAT$^{-/-}$:Vβ1NT mice (Figure 4.7A). In DP thymocytes from RAG1$^{-/-}$:Vβ1NT mice, short-range interactions between Eβ and Dβ2 are retained (Figure 4.7A, right panel), as well as those with the rearranged *Trbv4/5* region (Figure 4.6A, left panel). However, the enhancer no longer associates with more distal *Trbv1-3* gene segments. Dissociation is not solely due to enhancer capture by the promoter of the rearranged *Trbv5*, because we find a parallel loss of interactions between the distal *Trbv1* segment and a second viewpoint within the RC spanning the Dβ2Jβ cluster (Figure 4.7B).

Diminished RC–*Trbv* looping upstream of the pre-assembled VβDβJβ exon in DP thymocytes does not correlate with reduced deposition of CTCF at any site along the Vβ1NT allele (Figure 4.7C). Similarly, RAD21 levels are not significantly altered at any of these CTCF sites, with the exception of a site situated 5' to the rearranged *Trbv5* segment, where RAD21 binding is reduced (Figure 4.7D). However, the functional significance of reduced cohesin deposition at this single site remains unclear. Notwithstanding, we show that spatial segregation of distal *Trbv* segments from the RC in DP thymocytes is a common feature of germline, as well as rearranged *Tcrb* alleles. Dissociation of the distal *Trbv* domain in DP thymocytes occurs regardless of whether the 5'PC tether or the 5'V11 inflection regions are present. These data strongly suggest that separation of distal Vβ segments is independent of specific cis-acting elements and is a process inherent to thymocytes transitioning from the DN to DP stage.

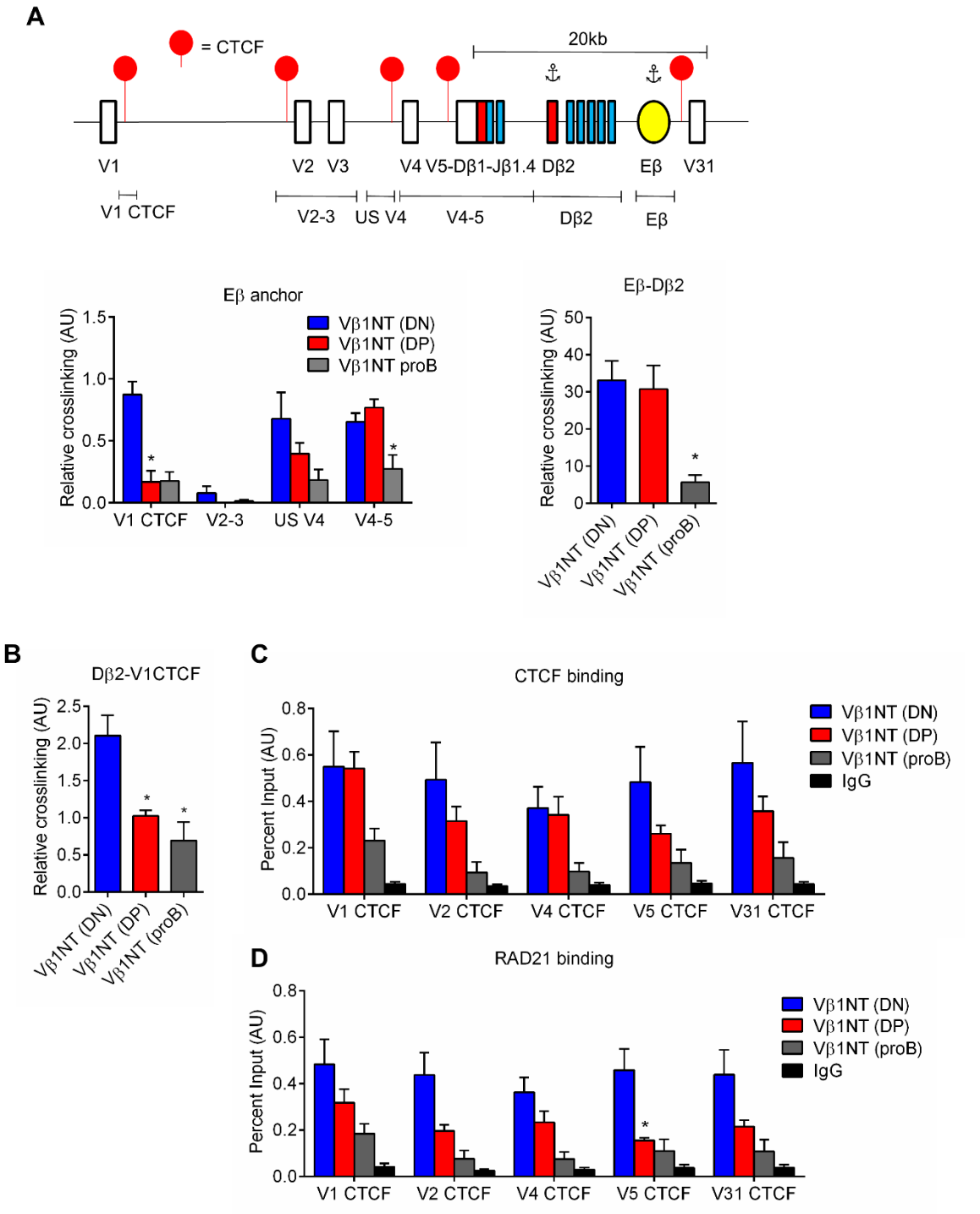


Figure 4.7: DP-specific dissociation of distal *Trbv* segments on a rearranged *Tcrb* allele. (A) Schematic of the rearranged *Tcrb* locus in Vβ1NT mice. Location of CTCF binding sites within the region are designated as red lollipops. Hind III fragments used for 3C-qPCR analysis are shown below the cartoon. 3C analysis was performed

using the Eβ viewpoint, probing its interactions with either the upstream *Trbv* region (left panel) or with the nearby Dβ2Jβ cluster (right panel). Chromatin for these assays was obtained from purified pro-B cells or DP thymocytes harvested from RAG-deficient mice with the Vβ1NT rearrangement on both of their germline alleles. For corresponding DN thymocytes, the Vβ1NT allele was introduced into a $RAG1^{-/-}$:$LAT^{-/-}$ background. **(B)** 3C-qPCR assays that measure interactions between the Dβ2 viewpoint and the most distal *Trbv1* region. **(C, D)** ChIP-qPCR assays were performed to quantify CTCF (panel C) and RAD21 deposition (panel D) at the indicated sites.

4.5 Discussion

The assembly of large AgR loci during lymphocyte development is controlled by coordinated changes in transcription, chromatin, and conformation (Bassing et al., 2002; Osipovich and Oltz, 2010; Krangel, 2009; Bosen et al., 2012; Roy et al., 2011). At the appropriate developmental stage for assembly, recombinase targets within each AgR locus become transcriptionally active and long-range recombination is facilitated by locus contraction (Matthews et al., 2007). For *Igh* and *Tcrb*, these features are reversed at subsequent stages of development to enforce allelic exclusion, despite continued expression of the V(D)J recombinase (Kosak et al., 2012; Skok et al., 2007; Senoo and Shinkai, 1998; Tripathi et al., 2002). With regard to locus conformation, it has been thought that *Igh* and *Tcrb* decontraction is accompanied by the loss of spatial associations between their respective RCs and V clusters (Skok et al., 2002; Roldán et al., 2005). In this work, we provide the most rigorous examination to date of this model, using high-resolution 3C to probe spatial associations within *Tcrb* as thymocytes pass from the permissive DN stage of development to the DP subset, in which its allelic exclusion is enforced.

A major finding from our 3C studies is that most of the *Trbv* cluster remains associated with the RC after transition of thymocytes to the DP stage and after *Tcrb* locus decontraction (Figure 4.8). Only the most distal portion of the *Trbv* cluster, upstream from the *Trbv11* gene segment, spatially segregates from the RC. These conformational and cell imaging data remain consistent with one another, because all previous 3D-FISH assays, including our own, employed probes for the most distal *Trbv* domain, which separates to a greater average distance from RC probes in DP thymocytes (Skok et al., 2007; Majumder et al., 2015). Although only distal Vβ segments dissociate from the RC, both the proximal and distal *Trbv* domains appear to remain

folded in their thymocyte-specific conformations, which promotes domain-specific association of gene segments (Figure 4.8). Of note, our conclusions differ significantly from those of a published study that also used 3C to probe *Tcrb* conformations in DN and DP thymocytes (Skok et al., 2007). The authors of the previous study concluded that both *Trbv*-RC and *Trbv*-*Trbv* interactions are disrupted in DP cells over the entire Vβ cluster. The specific source of this discrepancy remains unclear; however, the previous study used DN thymocytes cultured on stromal cells and older methods for 3C analyses that are, at best, only semi-quantitative.

Notwithstanding, our data clearly preclude a model in which *Tcrb* allelic exclusion is enforced primarily by topological dissociation of the *Trbv* cluster from its DβJβ targets in DP thymocytes. Although this spatial mechanism may be dominant for *Trbv1-5*, the only functional Vβ segments that separate from the RC, distinct factors must prevent long-range recombination of the other 16 Vβ segments, which remain associated with the RC in DP thymocytes. Thus, for this latter set of Vβ segments, transcriptional suppression in DP cells is the most likely mechanism for enforcing allelic exclusion. In this regard, we have shown that, in DN thymocytes, the relative levels of spatial proximity for each *Trbv* segment do not contribute in a measurable way to their usage in the pre-selection *Tcrb* repertoire (Gopalakrishnan et al., 2013). Instead, the level of transcription for each *Trbv* segment provides an excellent correlate for its usage in Vβ-to-DβJβ recombination. We conclude that Vβ promoter activity, rather than spatial constraints, is a dominant mechanism for both *Tcrb* gene assembly in DN cells and its subsequent allelic exclusion in DP thymocytes.

We also assessed possible mechanisms that may drive *Tcrb* conformational changes during thymocyte development. Our analyses discount domain-specific changes in transcription or general chromatin features as a force for the preferential dissociation of distal *Trbv* segments from the RC. All of these features appear to be regulated coordinately over the entire *Trbv* cluster, including the mapped inflection point for RC dissociation near *Trbv11*. Moreover, loss of distal *Trbv*-RC interactions in DP cells is independent of CTCF/cohesin deposition at specific sites within the Vβ cluster or at their tether near the RC. We also found that *Tcrb* decontraction, as well as distal *Trbv* dissociation from the RC, are both independent of the massive proliferative burst preceding DN to DP transition.

The latter finding suggests that *Tcrb* conformational remodeling occurs via a DP-intrinsic process. Indeed, this model is strongly supported by our topological analyses of a functionally rearranged *Tcrb* allele. In DN thymocytes, upstream *Trbv* segments remain associated with an RC that harbors a functional Vβ5Dβ1Jβ1.4 rearrangement (Figure 4.8). Similar to germline *Tcrb* loci, the distal *Trbv* segments become transcriptionally repressed (Brady, Oropallo et al., 2010; Brady and Bassing, 2011; Senoo and Shinkai, 1998) and dissociate from the RC interactome in DP thymocytes. The spatial segregation occurs on a rearranged allele that lacks both the normal long-range tether for distal Vβ segments (5'PC) and the mapped inflection point for *Trbv*-RC associations in DP cells (5'V11). These findings indicate that, rather than specific cis-elements orchestrating *Tcrb* conformational changes, perhaps spatial segregation of distal *Trbv* segments occurs via a process that is intrinsic to DP thymocytes. In fact, a switch from RC association with entire V clusters at the earliest precursor stages, to a loss of RC-distal V interactions at subsequent stages, may be a general feature of lymphocyte development. In DN thymocytes, distal and

proximal V segments associate with RCs at both *Tcrb* (this study) and *Tcrad* (Shih and Krangel, 2010). In the latter case, a global V-RC interactome likely facilitates more diverse usage of V segments in assembled *Tcrd* genes. However, upon transition to the DP stage, distal portions of both the *Trbv* and *Trav* clusters dissociate from the RC, which limits initial *Tcra* recombination to the most proximal Vα segments (Shih and Krangel, 2010). For *Tcrb*, DP-intrinsic dissociation of distal Vβ segments might curb their secondary rearrangement to the downstream Dβ2Jβ cluster, if present, which would delete an existing VβDβ1Jβ join. Likewise, in pro-B cells, distal V segments at both the *Igh* and *Igk* loci associate with their RCs (Stadhouders et al., 2014; Lin et al., 2012); however, whether these long-range interactions are maintained in pre-B cells remains an open question.

In the case of thymocyte development, future studies must focus on mechanisms driving the stage-specific changes in *Tcrb* conformation. One attractive possibility for an underlying mechanism is stage-specific alteration of transcriptional status over the entire *Trbv* cluster. Our prior studies have shown that the *Trbv* domains fold and interact with distant DβJβ cluster in DN thymocytes, independent of RC transcription (Majumder et al., 2015). Because many Vβ segments become transcriptionally active upon their differentiation to the DN stage, we speculate that the induced expression dictates conformational changes that drive RC-V association upon T cell commitment. Likewise, dissociation of the distal *Trbv* domain is accompanied by a widespread loss of Vβ transcription (Senoo and Shinkai, 1998; Tripathi et al., 2002). Perhaps both of these expression-induced changes in conformation are governed by the activation or suppression of a

key transcription factor, such as E2A, which has been implicated in the control of Vβ-to-DβJβ recombination (Agata et al., 2007).

Even so, a critical question would be why only the most distal Vβ segments dissociate in DP cells, despite the loss of transcription over the entire *Trbv* cluster. We suspect that clues can be derived from our finding that 5'V11 serves as a hinge for the loss of RC interactions on germline *Tcrb* alleles in DP thymocytes. The *Trbv11* region is the transition from a large stretch of silent pseudogene segments (*Trbv6-11*) to the most highly active cluster of *Trbv* segments (*Trbv12-13*) in DN thymocytes, the latter of which remains partially active in DP cells (see Fig. 1G). In this regard, genome-wide conformation studies have shown that regions of similar transcriptional status tend to form topologically associated domains (Dixon et al., 2012). Thus, in DP thymocytes, silencing of the distal Vβ segments may drive their preferential association with neighboring regions of repressed transcription, including the silent trypsinogen genes between *Trbv1* and *Trbv2* and pseudogenes that encompass the 5' portion of the locus (Figure 4.8). These changes would spatially segregate *Trbv1-11* from the modestly transcribed *Trbv12-13* and other downstream segments, which retain their association with the RC via thymocyte-intrinsinc mechanisms. Future investigations of such conformational changes at AgR loci will continue to shed light on the mechanisms controlling genome topologies and how these changes, in turn, regulate many aspects of gene function.

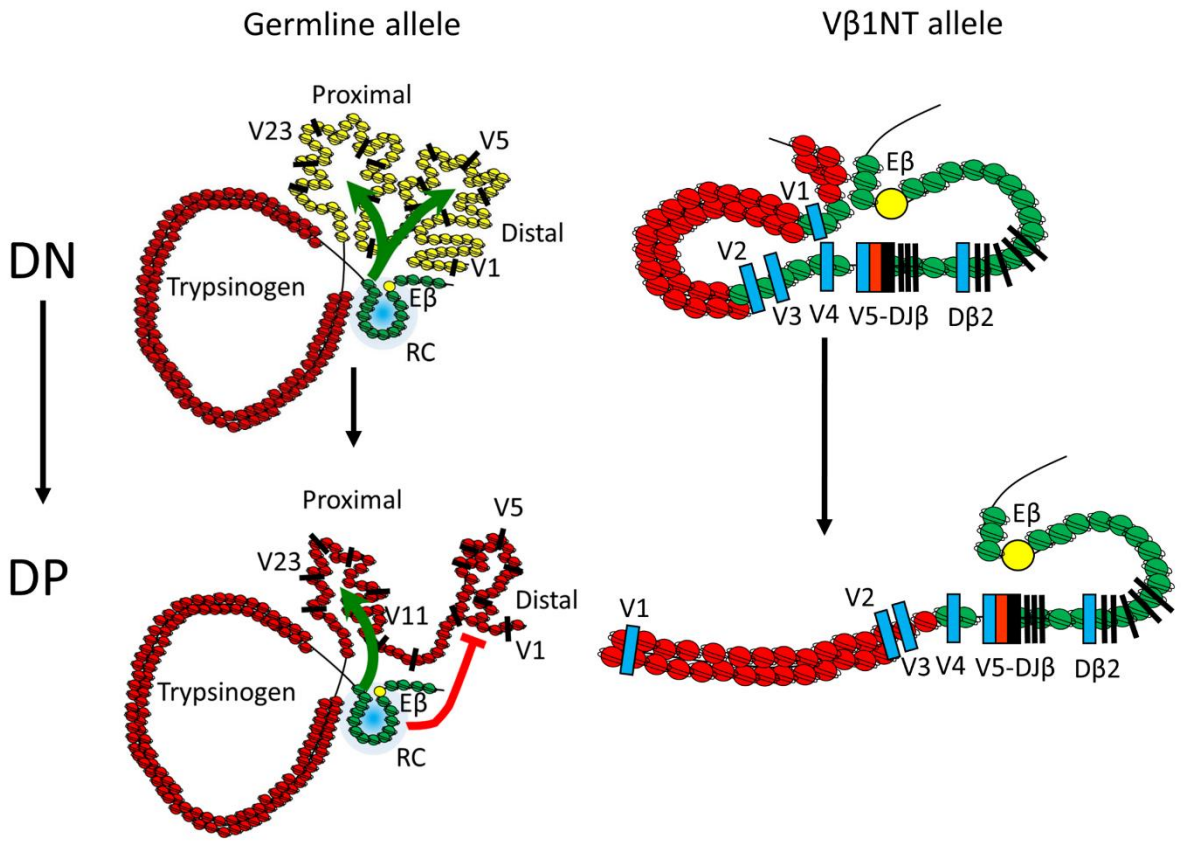


Figure 4.8: Model for developmental reconfiguration of *Tcrb*. Cartoon depictions of germline (left) and rearranged (right) *Tcrb* loci in either DN (top) or DP (bottom) thymocytes. Distinct types of chromatin and transcriptional activity are represented by colored nucleosomes: repressive, transcriptionally inert (red); highly transcribed, accessible (green); modestly transcribed, accessible (yellow). Left: Folding of the distal (containing V1 and V5) and proximal (containing V23) *Trbv* domains, as well as their interactions with the RC/Eβ region on a germline *Tcrb* allele, are highlighted as green arrows. In DP thymocytes with a germline *Tcrb*, both *Trbv* domains remain folded but only the distal domain dissociates from the RC (red block). Right: Similarly, distal *Trbv* segments that remain on a rearranged *Tcrb* allele (e.g., V1) spatially segregate from the Eβ interactome in DP thymocytes.

4.6 References

Agata, Y., Tamaki, N., Sakamoto, S., Ikawa, T., Masuda, K., Kawamoto, H., & Murre, C. (2007). Regulation of T cell receptor β gene rearrangements and allelic exclusion by the helix-loop-helix protein, E47. *Immunity*, *27*(6), 871-884.

Bassing, C. H., Swat, W., & Alt, F. W. (2002). The mechanism and regulation of chromosomal V (D) J recombination. *Cell*, *109*(2), S45-S55.

Bossen, C., Mansson, R., & Murre, C. (2012). Chromatin topology and the regulation of antigen receptor assembly. *Annual review of immunology*, *30*, 337-356.

Bouvier, G., Watrin, F., Naspetti, M., Verthuy, C., Naquet, P., & Ferrier, P. (1996). Deletion of the mouse T-cell receptor beta gene enhancer blocks alphabeta T-cell development. *Proceedings of the National Academy of Sciences*, *93*(15), 7877-7881.

Brady, B. L., & Bassing, C. H. (2011). Differential Regulation of Proximal and Distal Vβ Segments Upstream of a Functional VDJβ1 Rearrangement upon β-Selection. *The Journal of Immunology*, *187*(6), 3277-3285.

Brady, B. L., Oropallo, M. A., Yang-Iott, K. S., Serwold, T., Hochedlinger, K., Jaenisch, R., ... & Bassing, C. H. (2010). Position-dependent silencing of germline Vβ segments on TCRβ alleles containing preassembled VβDJβCβ1 genes. *The Journal of Immunology*, *185*(6), 3564-3573.

Brady, B. L., Steinel, N. C., & Bassing, C. H. (2010). Antigen receptor allelic exclusion: an update and reappraisal. *The Journal of Immunology*, *185*(7), 3801-3808.

Choi, N. M., Loguercio, S., Verma-Gaur, J., Degner, S. C., Torkamani, A., Su, A. I., ... & Feeney, A. J. (2013). Deep sequencing of the murine Igh repertoire reveals complex regulation of nonrandom V gene rearrangement frequencies.*The Journal of Immunology*, *191*(5), 2393-2402.

Cobb, R. M., Oestreich, K. J., Osipovich, O. A., & Oltz, E. M. (2006). Accessibility control of V (D) J recombination. *Advances in immunology*, *91*, 45-109.

Degner, S. C., Verma-Gaur, J., Wong, T. P., Bossen, C., Iverson, G. M., Torkamani, A., ... & Feeney, A. J. (2011). CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the Igh locus and antisense transcription in pro-B cells. *Proceedings of the National Academy of Sciences,108*(23), 9566-9571.

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., ... & Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, *485*(7398), 376-380.

Gopalakrishnan, S., Majumder, K., Predeus, A., Huang, Y., Koues, O. I., Verma-Gaur, J., ... & Oltz, E. M. (2013). Unifying model for molecular determinants of the preselection Vβ repertoire. *Proceedings of the National Academy of Sciences*, *110*(34), E3206-E3215.

Guo, C., Yoon, H. S., Franklin, A., Jain, S., Ebert, A., Cheng, H. L., ... & Alt, F. W. (2011). CTCF-binding elements mediate control of V (D) J recombination.*Nature*, *477*(7365), 424-430.

Hewitt, S. L., Chaumeil, J., & Skok, J. A. (2010). Chromosome dynamics and the regulation of V (D) J recombination. *Immunological reviews*, *237*(1), 43-54.

Ji, Y., Resch, W., Corbett, E., Yamane, A., Casellas, R., & Schatz, D. G. (2010). The in vivo pattern of binding of RAG1 and RAG2 to antigen receptor loci. *Cell*, *141*(3), 419-431.

Kim, T. H., Abdullaev, Z. K., Smith, A. D., Ching, K. A., Loukinov, D. I., Green, R. D., ... & Ren, B. (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell*, *128*(6), 1231-1245.

Kosak, S. T., Skok, J. A., Medina, K. L., Riblet, R., Le Beau, M. M., Fisher, A. G., & Singh, H. (2002). Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science*, *296*(5565), 158-162.

Krangel, M. S. (2009). Mechanics of T cell receptor gene rearrangement. *Current opinion in immunology*, *21*(2), 133-139.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods*, *9*(4), 357-359.

Lin, Y. C., Benner, C., Mansson, R., Heinz, S., Miyazaki, K., Miyazaki, M., ... & Murre, C. (2012). Global changes in the nuclear positioning of genes and intra-and interdomain genomic interactions that orchestrate B cell fate. *Nature immunology*, *13*(12), 1196-1204.

Lin, S. G., Guo, C., Su, A., Zhang, Y., & Alt, F. W. (2015). CTCF-binding elements 1 and 2 in the Igh intergenic control region cooperatively regulate V (D) J recombination. *Proceedings of the National Academy of Sciences*, *112*(6), 1815-1820.

Liu, Y., Subrahmanyam, R., Chakraborty, T., Sen, R., & Desiderio, S. (2007). A plant homeodomain in RAG-2 that binds Hypermethylated lysine 4 of histone H3 is necessary for efficient antigen-receptor-gene rearrangement. *Immunity*,*27*(4), 561-571.

Majumder, K., Koues, O. I., Chan, E. A., Kyle, K. E., Horowitz, J. E., Yang-Iott, K., ... & Oltz, E. M. (2015). Lineage-specific compaction of Tcrb requires a chromatin barrier to protect the function of a long-range tethering element. *The Journal of Experimental Medicine*, *212*(1), 107-120.

Matthews, A. G., Kuo, A. J., Ramón-Maiques, S., Han, S., Champagne, K. S., Ivanov, D., ... & Oettinger, M. A. (2007). RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V (D) J recombination. *Nature*, *450*(7172), 1106-1110.

Medvedovic, J., Ebert, A., Tagoh, H., Tamir, I. M., Schwickert, T. A., Novatchkova, M., ... & Busslinger, M. (2013). Flexible Long-Range Loops in the V H Gene Region of the Igh Locus Facilitate the Generation of a Diverse Antibody Repertoire. *Immunity*, *39*(2), 229-244.

Naumova, N., Imakaev, M., Fudenberg, G., Zhan, Y., Lajoie, B. R., Mirny, L. A., & Dekker, J. (2013). Organization of the mitotic chromosome. *Science*,*342*(6161), 948-953.

Oestreich, K. J., Cobb, R. M., Pierce, S., Chen, J., Ferrier, P., & Oltz, E. M. (2006). Regulation of TCRβ gene assembly by a promoter/enhancer holocomplex. *Immunity*, *24*(4), 381-391.

Osipovich, O., & Oltz, E. M. (2010, December). Regulation of antigen receptor gene assembly by genetic–epigenetic crosstalk. In *Seminars in immunology* (Vol. 22, No. 6, pp. 313-322). Academic Press.

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H. C., ... & Merkenschlager, M. (2008). Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell*, *132*(3), 422-433.

Pekowska, A., Benoukraf, T., Zacarias-Cabeza, J., Belhocine, M., Koch, F., Holota, H., ... & Spicuglia, S. (2011). H3K4 tri-methylation provides an epigenetic signature of active enhancers. *The EMBO journal*, *30*(20), 4198-4210.

Phillips, J. E., & Corces, V. G. (2009). CTCF: master weaver of the genome.*Cell*, *137*(7), 1194-1211.

Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. *Trends in genetics*, *16*(6), 276-277.

Roldán, E., Fuxa, M., Chong, W., Martinez, D., Novatchkova, M., Busslinger, M., & Skok, J. A. (2005). Locus' decontraction'and centromeric recruitment contribute to allelic exclusion of the immunoglobulin heavy-chain gene. *Nature immunology*, *6*(1), 31-41.

Roy, A. L., Sen, R., & Roeder, R. G. (2011). Enhancer–promoter communication and transcriptional regulation of Igh. *Trends in immunology*,*32*(11), 532-539.

Schlimgen, R. J., Reddy, K. L., Singh, H., & Krangel, M. S. (2008). Initiation of allelic exclusion by stochastic interaction of Tcrb alleles with repressive nuclear compartments. *Nature immunology*, *9*(7), 802-809.

Seitan, V. C., Hao, B., Tachibana-Konwalski, K., Lavagnolli, T., Mira-Bontenbal, H., Brown, K. E., ... & Merkenschlager, M. (2011). A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation. *Nature*,*476*(7361), 467-471.

Senoo, M., & Shinkai, Y. (1998). Regulation of Vbeta germline transcription in RAG-deficient mice by the CD3epsilon-mediated signals: implication of Vbeta transcriptional regulation in TCR beta allelic exclusion. *International immunology*, *10*(5), 553-560.

Serwold, T., Hochedlinger, K., Inlay, M. A., Jaenisch, R., & Weissman, I. L. (2007). Early TCR expression and aberrant T cell development in mice with endogenous prerearranged T cell receptor genes. *The Journal of Immunology*,*179*(2), 928-938.

Shih, H. Y., & Krangel, M. S. (2010). Distinct contracted conformations of the Tcra/Tcrd locus during Tcra and Tcrd recombination. *The Journal of experimental medicine*, *207*(9), 1835-1841.

Shih, H. Y., & Krangel, M. S. (2013). Chromatin architecture, CCCTC-binding factor, and V (D) J recombination: managing long-distance relationships at antigen receptor loci. *The Journal of Immunology*, *190*(10), 4915-4921.

Shinkai, Y., & Alt, F. W. (1994). CD3ε-mediated signals rescue the development of CD4+ CD8+ thymocytes in RAG-2−/− mice in the absence of TCR β chain expression. *International immunology*, *6*(7), 995-1001.

Sicinska, E., Aifantis, I., Le Cam, L., Swat, W., Borowski, C., Yu, Q., ... & Sicinski, P. (2003). Requirement for cyclin D3 in lymphocyte development and T cell leukemias. *Cancer cell*, *4*(6), 451-461.

Skok, J. A., Gisler, R., Novatchkova, M., Farmer, D., de Laat, W., & Busslinger, M. (2007). Reversible contraction by looping of the Tcra and Tcrb loci in rearranging thymocytes. *Nature immunology*, *8*(4), 378-387.

Stadhouders, R., de Bruijn, M. J., Rother, M. B., Yuvaraj, S., de Almeida, C. R., Kolovos, P., ... & Hendriks, R. W. (2014). Pre-B Cell Receptor Signaling Induces Immunoglobulin κ Locus Accessibility by Functional Redistribution of Enhancer-Mediated Chromatin Interactions. *PLoS biology*, *12*(2), e1001791.

Xiang, Y., Park, S. K., & Garrard, W. T. (2013). Vκ gene repertoire and locus contraction are specified by critical DNase I hypersensitive sites within the Vκ-Jκ intervening region. *The Journal of Immunology*, *190*(4), 1819-1826.

Thongjuea, S., Stadhouders, R., Grosveld, F. G., Soler, E., & Lenhard, B. (2013). r3Cseq: an R/Bioconductor package for the discovery of long-range genomic interactions from chromosome conformation capture and next-generation sequencing data. *Nucleic acids research*, *41*(13), e132-e132.

Tripathi, R., Jackson, A., & Krangel, M. S. (2002). A change in the structure of Vβ chromatin associated with TCR β allelic exclusion. *The Journal of Immunology*, *168*(5), 2316-2324.

Yang, Q., Jeremiah Bell, J., & Bhandoola, A. (2010). T-cell lineage determination. *Immunological reviews*, *238*(1), 12-22.

Zhang, Y., McCord, R. P., Ho, Y. J., Lajoie, B. R., Hildebrand, D. G., Simon, A. C., ... & Dekker, J. (2012). Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell*, *148*(5), 908-921.

Table T12: List of primers and probes used in 4C, 3C, ChIP-qPCR and RT-qPCR assays

4C inverse PCR primers	
Eβ inverse HindIII	CCTAAGGAGAGGCAGAGTGG
Eβ inverse NlaIII	AGGAAATTGATTTCAGCTCTCAT
Eβ nested inverse HindIII	GTGGTAGGAATTGTTAGGAAAAG
Eβ nested inverse NlaIII	AAGAGTGCCTTGATAGGGAAA
V5 inverse HindIII	TGTATGCCATCTTGTCTGGT
V5 inverse NlaIII	AATATGAGCCTGGCATATAGG
V5 nested inverse HindIII	CTTTTTTCTCTACTTTTAGCTAGTCT
V5 nested inverse NlaIII	AGCCTGGCATATAGGGTTAC
Taqman probe sequences (5'FAM and 3'TAMRA) for 3C assays	
Eβ probe	CATAAGCATTGTCATGTTTGTGACA
Dβ1 probe	AAGGCATTGTTGCATGATCCT
Dβ2 probe	AAATGCTGGGCCTCTGTAGA
5'PC probe	CAGTGGGGAATCAGACTTTCA
5' V11 probe	TCCCTCAATACAGACCGAAAG
V11 probe	CTTGGCTGCCATTTTTCCTA
V5 probe	CAGTCGTTCTTTATGTCTGATACTGTG
V23 probe	TACACCGGCCAGGAGAGACT
Primers sequences with respective Taqman probes for 3C assays	
Eβ primer	GAAAATTGGCATCGGTTTGC
Dβ1 primer	TGAAATTTTTCTGCCGAAAGGAC
Dβ2 primer	GCGGGATCCAAGAGAACTCA
5'PC primer	TGTGTTGAAGATTGGGGTGA
5' V11 primer	AGCATCAATTATGACAATGCAGA
V11 primer	TTTTCCCAATCTCTTGGCTG
V5 primer	TCCCTCAGCGGTTCAGTAGTC
V23 primer	GGCTTCTGTGTAACTGCAGCAT
Primer sequences for 3C assays	
V1 CTCF	ACCCATGTCCTCAGGGTTTC
V2-3	TTTCATTCACAGCCGACCAG
usV4	TGGAAAATTTTAGCAAAGAACTTGAA
V4-V5	AGCTCGACACAGAAAGCAAGTT
V6	GGTTCCCTTCACTTCCCACA
V7-8	GTCCGCTAGCAGCCAGAGTT
5'V11	TGCATGTGGCTTAGGAGTCAA
V11	TCTTGGCTGCCATTTTTCCT
V12-1	TCCACCATTTCCCTTCCAAC
V12-2	CCATCTGCATGAACACCTTCTT
V14	CAGGCTTTTGAGTGGCATGT
V16	TATCATGCCCAGCTGCATTC
V23	TACACCGGCCAGGAGAGACT

V29	CTCTAGCAATCCCCCTGTGC
5’PC	CCAACTTGCAGTGTGGTCCT
Dβ1	AAGGCATTGTTGCATGATCC
Dβ2	TGGGGCCCTCACTTTTCTTA
Primer sequences for ChIP-qPCR and RT-qPCR assays	
Mur2 fwd	TTAGGTGCTTCTTGGCCAGT
Mur2 rev	TAGGAAAAATGGCAGCCAAG
Mus4 fwd	TGAAGGAAAGGTCCCTCAGA
Mus4 rev	GACCCTGTGCTCAGTCCAAT
Eβ ncRNA fwd	TGGGATCATGTCGCCTTATT
Eβ ncRNA rev	CCAGGCAGAGTTTGGAGAAC
Prss2 fwd	ACCATGAGTGCACTTCTGATCC
Prss2 rev	GGCAGGTGTATCCTCCAACA
Jβ1-Cβ1 fwd	GAA CCA GAC TCA CAG TTG TAG AGG
Jβ1-Cβ1 rev	GCT CTC CTT GTA GGC CTG AG
Jβ2-Cβ2 fwd	ACG ACT CAC CGT CCT AGA GG
Jβ2-Cβ2 rev	CAT TCA CCC ACC AGC TCA G
V1 CTCF fwd	AGGAAGATTGTGGGCAACTG
V1 CTCF rev	AACCAAATAAACGGCAGCAC
V2 CTCF fwd	TCTCAAGCTCATGAAGGGAAA
V2 CTCF rev	TCACAGCTCCCCCTAGAGAA
V4 CTCF fwd	TTCATTCCACTGGCCACAAG
V4 CTCF rev	CTGAATCTCATTGCCACAGC
V5 CTCF fwd	GCACTGCCAATCTCTGCAT
V5 CTCF rev	CATTTCTTTCCCGTTCTCCA
V10 CTCF fwd	CTGGAATTTTCCGGTTCCTT
V10 CTCF rev	TGGCTCCATCCTCAGTCTCT
V11 CTCF fwd	CTTTGGTTTGGAGGCACAAT
V11 CTCF rev	CGGAGGCTTTAGATCACCAA
V12-1 CTCF fwd	CAACGGGCAAAATTTGAGAT
V12-1 CTCF rev	CTGCTCTGTTCTGGGTCTCC
V14 CTCF fwd	TCACCTATGGCCTCCTTGTC
V14 CTCF rev	CCTGCTTGGCAAACTCTAGG
V29 CTCF fwd	AACCCTCCATCCCTTTCACT
V29 CTCF rev	CTGGTTCCGTTTTTAATGGG
5’PC fwd	CAGTGTTTGCCGACAGCTTA
5’PC rev	CACGCCTGGGTTTGTTTACT
V31 CTCF fwd	GTGTTTGGTGCCAGGAACAGA
V31 CTCF rev	TGGTTACCTTGGCAACTGAGA
3’LTR fwd	TTCCTCTGGAACCTGCAGCT
3’LTR rev	AAGCCACAGTGTGTGGTCTG
LINE fwd	TAAGAGTCCACAAAGACTGAAC
LINE rev	TGCTCCAGCTGTTAGGAGTAT

Dβ1 fwd	AAGCTGTAACATTGTGGGGACAGG
Dβ1 rev	CAATCTTGGCCTAGCAGGCTGCAG
PDβ2 fwd	GTTTCTGAGGCATGTGTCTCTGCG
PDβ2 rev	TCCTCTTTGTCACAGTGCCCACC
Actb fwd	GGCTGTATTCCCCTCCATCG
Actb rev	CCAGTTGGTAACAATGCCATGT
V1 fwd	TCAAGCTGTGAACCTACGCTGCAT
V1 rev	AGGTAATCAGCACCGGGAAGAGAT
V5 fwd	TGGAATGTGAGCAACATCTGGGAC
V5 rev	GGGCACCGTCTCATTTCGAATCAA
V11 fwd	GGGCAAGAGCCATGGTTTCTAACT
V11 rev	TGAAAGAGAGGTGCTCTGTCGAGT
V12-1 fwd	AGGAAGGTCCGTTCTGACGTGTAT
V12-1 rev	TGTTGGACTGAGAATCTGCTGGGT
V13-2 fwd	ACAAGGTGGCAGTAACAGGAGGAA
V13-2 rev	TTGGTCTGGAGGCCTTGTATCCAT
V16 fwd	TGCTGGTGTCATCCAAACACCTAG
V16 rev	TTGGGCATCTGAGCTGAGAATCGT
3’V11 fwd	GGATGACCCAGAAGGATGAA
3’V11 rev	TGTGATTGCTGGGATTTGAA
5’V12-1 CTCF fwd	TGAAGGAAAGGTCCCTCAGA
5’V12-1 CTCF rev	GACCCTGTGCTCAGTCCAAT

Chapter 5: Conclusions and Future Directions

The molecular mechanisms governing folding of our genomes into the nucleus, and how this influences tissue-epecific gene expression, remain largely an enigma. In this regard, the T-Cell Receptor Beta locus provides a good model system to uncover the spatial and epigenetic features of genome topology. Some of these folding mechanisms are essential for the generation of antigen receptor diversity while others are essential for eukaryotic gene regulation. The small size of *Tcrb* locus with one enhancer element makes it a good model system to dissect using qPCR based techniques. To date, our knowledge of *Tcrb* topology was limited, including the formation of a promoter-enhancer holocomplex at the recombination center (Oestreich et al., 2006) and reversible contraction of *Tcrb* during thymocyte development (Skok et al., 2007). Our studies have expanded on the existing knowledge, provided novel insights into how the *Tcrb* repertoire is sculpted by spatial and epigenetic mechanisms, what factors maintain tissue-specific conformation and how *Tcrb* topology is changed throughout development.

Generation of the pre-selection Vβ repertoire

Our studies show that the *Tcrb* locus adopts a tissue-specific conformation in DN thymocyte, where all the Vβ gene segments have spatial access to the RC. The relative proximity of Vβ genes over one another does not confer an advantage in recombination potential. However, the determinants that drive selection of Vβ gene recombination are chromatin accessibility, germline transcription and RSS quality (Figure 5.1). These factors serve as substrates for RAG recruitment and subsequent rearrangement. Our predictive model ranks the factors that are essential and dispensable for Vβ rearrangement. These factors can now be mechanistically tested in model systems of thymocyte development.

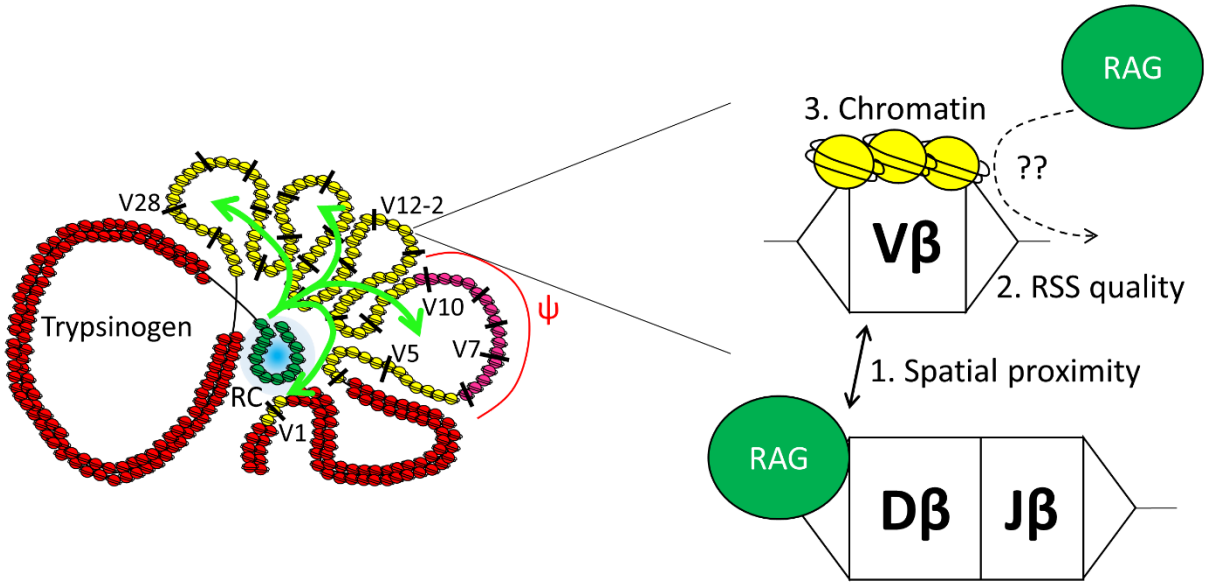


Figure 5.1: Model for molecular mechanisms that sculpt pre-selection *Tcrb* repertoire. In the left panel, transcriptionally active recombination center is colored in blue, transcriptionally active chromatin is presented as green nucleosomes, inactive chromatin as red nucleosome, poised chromatin as yellow nucleosomes and pseudogene chromatin as pink nucleosomes. The right panel shows the relative contributions of different factors in the pre-selection repertoire.

Trbv13-2 is one of the most efficiently recombining gene segments with its germline promoter upstream of *Trbv12-2*. The high recombination potential of *Trbv13*-2 may be due to high levels of germline transcripts emanating throughout this region. Alternatively, this high recombination potential may be a function of the highly efficient *Trbv13-2* RSS, because both instances would favor robust recombination. We can test this hypothesis by knocking in the crippled RSS of a *Trbv* pseudogene, like that of *Trbv8*, which would allow us to mechanistically test the role played by transcription on RSS quality. Along the same logic, switching the *Trbv12-2* promoter for the weak *Trbv30* promoter will enable us to test the relative roles played by transcription and H3K4me3 modification in regulating rearrangement.

The insulator function of CTCF proteins in maintaining genome accessibility can be mechanistically tested given our current knowledge of *Tcrb* chromatin. The bonafide test for this would be to mutate the strong CTCF site upstream of *Trbv12-2*. If the V12-2-CTCF site prevents the active chromatin from spreading upstream to *Trbv12-1* and *Trbv13-1*, then its mutation would increase their chromatin marks, associated germline transcripts and recombination potential. Alternatively, if V12-2-CTCF has a strict tethering function to the RC, its mutation would diminish long-range looping. These assays would allow us to test the synergism between chromatin, transcription, RSS quality and looping in sculpting the repertoire.

Regulation of tissue-specific *Tcrb* topology

We have shown that the regulation of lineage-specific *Tcrb* architecture (both contraction and looping) is independent of RAG binding, Eβ function and function of the minimal promoter upstream of Dβ1. However, we identify two unique CTCF-containing elements required for long-range tethering. One element within the inactive trypsinogen region (dubbed 5'PC) is essential for spatially harnessing the distal Vβ cluster. However, in order to perform its looping role, 5'PC must be protected by the CTCF-bound barrier element upstream of PDβ1. Loss of the upstream PDβ1 barrier leads to spread of active RC chromatin into the trypsinogen cluster, forcing 5'PC to become a new boundary and decommissioning its long-range looping potential. These studies identified the focal zone of long-range *Tcrb* interaction in DN thymocytes (Figure 5.2).

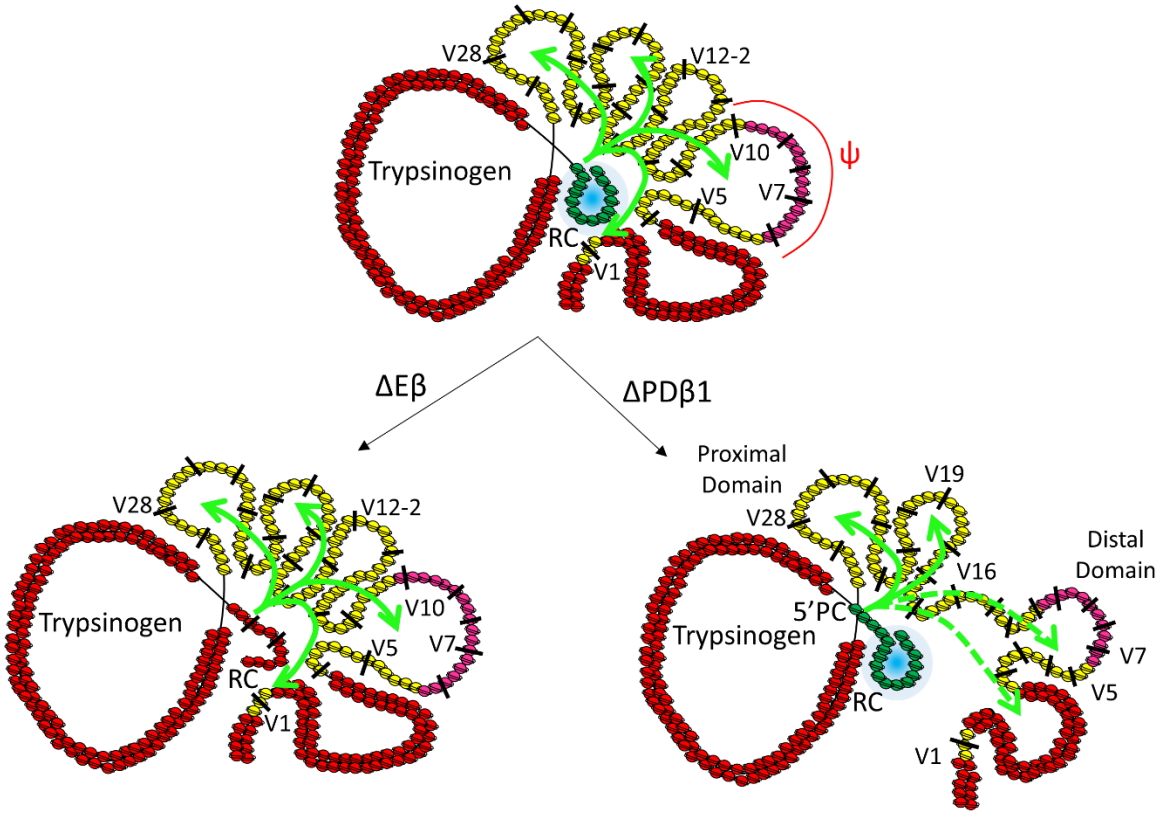


Figure 5.2: Regulatory element mediated control of *Tcrb* architecture. Changes in *Tcrb* architecture are shown in WT versus ΔEβ and ΔPDβ1 thymocytes. Transcriptionally active recombination center is colored in blue, transcriptionally active chromatin is presented as green nucleosomes, inactive chromatin as red nucleosome, poised chromatin as yellow nucleosomes and pseudogene chromatin as pink nucleosomes.

The mechanisms that stabilize long-range loop formation of *Tcrb* in DN thymocytes have remained enigmatic until now. Our studies using the ΔPDβ mice unexpectedly revealed a bi-domainal structure of the Vβ cluster, which is required for proper folding of the Vβ cluster to the RC as well and gross contraction of the locus. These studies also unearthed the important CTCF site in Trp gene cluster called 5'PC, which functions in distal Vβ tethering (Figure 5.3). We predict that mutation of the 5'PC site will spatially isolate the Vβ gene segments from the gene cluster, similar to the observations made on *Igh*-IGCR and *Igk*-Cer regions. Moreover, mutating the 5'PC site in ΔPDβ1 background should provide additional insights into how a barrier element functions.

Loss of the 5'RC boundary leads to spread of active chromatin to 5'PC (in ΔPDβ1 thymocytes), which now becomes a new boundary element. The next CTCF site upstream of 5'PC is located at *Trbv30*. It is possible that in Δ5'PC-ΔPDβ1 thymocytes the active chromatin would spread upstream throughout the entire 250 kb trypsinogen cluster to the V30-CTCF site. If not, these results will provide us with an in-vivo tool to map how far active chromatin can spread without inhibition by barrier elements.

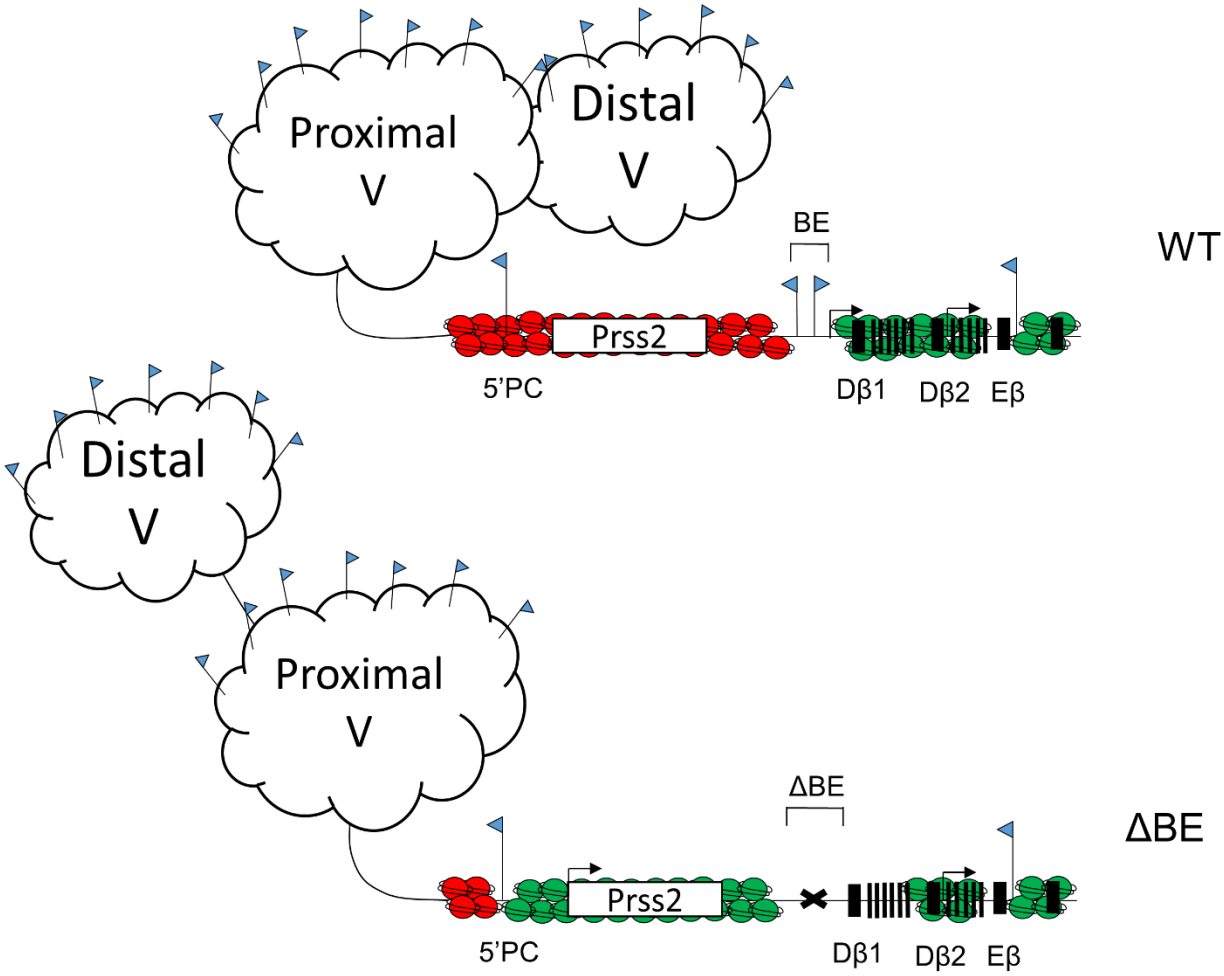


Figure 5.3: Regulation of *Tcrb* architecture by bifunctional barrier-tethering elements. The proposed mechanism for changes in long-range interactions by altered boundary elements function at the *Tcrb*-RC. Transcriptionally active chromatin is presented as green nucleosomes and transcriptionally repressive chromatin is presented as red

nucleosomes. The Vβ domains are illustrated as clouds. CTCF binding and orientation are presented as described in Figure 1.3.

The transition zone for distal-versus-proximal Vβ-domain interactions to the RC is present at the *Trbv14-Trbv16* region. This transition zone contains a strong CTCF binding site downstream of *Trbv14* (V14-CTCF). We propose that V14-CTCF is an important target for future mutagenesis studies, as it would allow the characterization of an intra-Vβ looping element, shifting the focus of *Tcrb* interactome studies from the RC. Mutating V14-CTCF (in WT or ΔPDβ1 backgrounds) may reveal unexpected roles for how the distal *Trbv* genes are tethered. The mutated V14-CTCF could alter bidomainal folding of *Trbv* tethering, shifting the transition to the next CTCF site.

The most distal Vβ gene segment *Trbv1* exists in a poised chromatin compartment, and contains a strong downstream CTCF site (V1-CTCF). Being downstream of *Trbv1*, V1-CTCF can perform a barrier role, separating the repressive chromatin of the 5'trp cluster from the active chromatin associated with *Trbv1*. If V1-CTCF is indeed a barrier, mutating it would repress *Trbv1*–associated chromatin, inhibiting its rearrangement. However, if V1-CTCF functions as a tethering element, then its mutation would abrogate long-range locus contraction and *Trbv1* looping. Since V1-CTCF is the only CTCF site in the 5'end of *Tcrb*, it is likely to be a bifunctional barrier-tether. Therefore, it may also play a role in locus contraction by bringing the distal 5' end of *Tcrb* close to the RC. The role of locus contraction in bringing the *Trbv2-Trbv30* cluster close to the RC can be tested in mice that are mutated in V1-CTCF. It would allow us to define what the underlying biological implications of locus contraction.

Maintenance of DP-intrinsic and proliferation independent topology

One of the existing models of the regulation of *Tcrb* topology has been the exertion of reversible DN thymocyte-specific contraction. As thymocytes progress to the DP stage of development, the RC and Vβ cluster were thought to spatially segregate from each other. Using high-throughput sequencing combined with 3D-FISH and qPCR to map the fine changes in *Tcrb* topology, we discovered that the DP-specific conformation does not spatially segregate the entire Vβ cluster. Although all the Vβ genes are transcriptionally inactive, the proximal Vβ gene segments continue to interact with the RC in a *Ccnd3*-independent manner. These findings suggest a DP-intrinsic mechanism exists that regulates long-range *Tcrb* loops. This principle is further validated by the loss of long-range loops in DP cells containing a rearranged *Tcrb* allele in Vβ1NT thymocytes (Figure 4.7).

Future studies on developmental control of *Tcrb* architecture will focus on two important principles: (i) How is the tissue-specific conformation in DN thymocytes induced when CLP's differentiate into the DN lineage, and (ii) What are the DP-specific mechanisms that maintain proximal *Trbv* looping. Work by the Blobel lab has shown that forced tethering of Ldb1 to the β-globin locus is sufficient to recapitulate loop formation between the β-globin promoter and LCR enhancer element (Deng et al., 2012). A similar strategy can be employed by artificially recruiting lineage-specific transcription factors such as E2A to the *Tcrb* locus in CLPs to test its effect on locus architecture. Alternatively, transcriptional upregulation may be sufficient to induce long-range folding. This phenomenon can be tested in DP thymocytes. The long-range DP-intrinsic loops terminate at the pseudo-Vβ gene segments. If transcriptional activity is required for loop formation, inducing the transcription of the upstream-Vβ cluster (*Trbv1-Trbv11*) will show the

retention of distal Vβ interaction in DP cells. This can be carried out using nuclease-deficient Cas9 fused with VP64, in combination with guide RNAs (Konermann et al., 2014) that target the upstream Vβ cluster. If transcription is required for the long-range loop formation, inducing distal Vβ expression will be sufficient to induce long-range looping of *Tcrb*. In this regard, the induction of trypsinogen transcription using CRISPR-Cas9 system can also be used to test whether the *Tcrb* architecture is altered. Active transcription from the trypsinogen region may alter *Tcrb* regulatory loop formation, thus altering its architecture.

Lineage-specific transcription factors can also play a role in the transcriptional induction process. Indeed, E2A is a favorite candidate for exerting such control. Overexpression of E2A in DP cells has been shown to subvert the epigenetic silencing of Vβ accessibility in DP cells possibly by altering the *Tcrb* topology (Agata et al., 2007). However, it remains unknown whether changes in Vβ accessibility are directly caused by E2A binding or by the induction of epigenetic factors that influence topology.

Recent genome-wide studies on the correlation of CTCF binding with long-range interactions have turned the attention of the field towards the orientation of CTCF sites. If two distal CTCF-bound regions of the genome were to interact by homodimerization of its associated CTCF proteins, then there are four possible permutations (2 possibilities per site multiplied by 2 sites) for CTCF-CTCF orientations. However, genome-wide analysis of Hi-C, alongside CTCF ChIP-seq data show a high probability of only one such orientation. This is the mutually convergent orientation (Rao et al., 2014). In support of this model for structural loop formation, we have found that the Vβ CTCF sites are in the forward, and three out of four RC-associated CTCF sites are in the reverse orientation. Therefore, looping principles of genome folding gleaned

from high-throughput data can be mechanistically tested on the *Tcrb* locus. In future studies, the orientation of the 5'PC, Eβ CTCF or PDβ1-associated CTCF sites can be inverted to determine the role played by mutually convergent CTCF orientations in exerting topological control of *Tcrb* expression and development of the Vβ repertoire. The RC-associated CTCF sites may play a synergistic role in maintaining *Tcrb* topology, as the IGCR-associated CTCF sites have been proposed to perform on *Igh* (Lin et al., 2015). These principles can be tested by performing single mutations of the RC-associated CTCF sites.

In conclusion, the *Tcrb* locus is an ideal model system to study the molecular mechanisms regulating genome topology by the interplay of regulatory elements and architectural proteins. Our studies on *Tcrb* have unraveled several general principles of stabilizing chromosomal topology within the nucleus, how these topological scaffolds regulate gene expression and how the three-dimensional configuration helps to assemble the adaptive immune repertoire. As we acquire more sophisticated and sensitive tools for analysis, we will resolve more mysteries behind the regulation of gene expression and antigen receptor gene assembly.

References

Agata, Y., Tamaki, N., Sakamoto, S., Ikawa, T., Masuda, K., Kawamoto, H., & Murre, C. (2007). Regulation of T cell receptor β gene rearrangements and allelic exclusion by the helix-loop-helix protein, E47. *Immunity*, *27*(6), 871-884.

Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P. D., ... & Blobel, G. A. (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell*, *149*(6), 1233-1244.

Konermann, S., Brigham, M. D., Trevino, A. E., Joung, J., Abudayyeh, O. O., Barcena, C., ... & Zhang, F. (2014). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*.

Lin, S. G., Guo, C., Su, A., Zhang, Y., & Alt, F. W. (2015). CTCF-binding elements 1 and 2 in the Igh intergenic control region cooperatively regulate V (D) J recombination. *Proceedings of the National Academy of Sciences*, *112*(6), 1815-1820.

Oestreich, K. J., Cobb, R. M., Pierce, S., Chen, J., Ferrier, P., & Oltz, E. M. (2006). Regulation of TCRβ gene assembly by a promoter/enhancer holocomplex. *Immunity*, *24*(4), 381-391.

Rao, S. S., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., ... & Aiden, E. L. (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell*, *159*(7), 1665-1680.

Skok, J. A., Gisler, R., Novatchkova, M., Farmer, D., de Laat, W., & Busslinger, M. (2007). Reversible contraction by looping of the Tcra and Tcrb loci in rearranging thymocytes. *Nature immunology*, *8*(4), 378-387.