Regulation of the T Cell Receptor Beta Locus by Nuclear Lamina Association

by

Shiwei Chen

Department of Immunology
Duke University

Date:_______________________

Approved:

___________________________
Michael S. Krangel, Supervisor

___________________________
You-Wen He

___________________________
Yuan Zhuang

___________________________
Qi-Jing Li

___________________________
Blanche Capel

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Immunology in the Graduate School of Duke University

2018
ABSTRACT

Regulation of the T Cell Receptor Beta Locus by Nuclear Lamina Association

by

Shiwei Chen

Department of Immunology
Duke University

Date: _______________________

Approved:

___________________________
Michael S. Krangel, Supervisor

___________________________
You-Wen He

___________________________
Yuan Zhuang

___________________________
Qi-Jing Li

___________________________
Blanche Capel

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Immunology in the Graduate School of Duke University

2018
Abstract

T lymphocytes of the adaptive immune system recognize antigens using T cell receptors, with each lymphocyte bearing a receptor of unique specificity. T cell receptors are generated through V(D)J recombination, a process in which gene segments are assembled for the generation of a functional receptor protein via somatic recombination.

The T cell receptor β (Tcrb) locus encodes the β chain of the T cell receptor αβ heterodimer. Among antigen receptor loci, the Tcrb locus is unique in that it frequently associates with the nuclear periphery during the developmental stage in which the locus undergoes recombination. Previous work showed that the association of the Tcrb locus with the nuclear periphery was a stochastic process and that the recombination of Tcrb alleles at the nuclear periphery was suppressed compared to Tcrb alleles in the interior of the nucleus. However, the mechanisms that instruct the frequent association of Tcrb alleles with the nuclear periphery remained unknown.

We characterized the association of the Tcrb locus with the nuclear lamina (NL) at high resolution using DamID. DamID analysis revealed that the association of the Tcrb locus with the NL was heterogeneous, and that a lamina-associated domain (LAD) border within the locus segregated the Tcrb recombination center (RC) from RC-proximal chromatin. This LAD border constrains the activity of the Tcrb enhancer (Eβ) to the RC. Accordingly, deletion of the LAD border caused the Eβ-dependent activation of RC-proximal chromatin, resulting in a loss of NL association, increased chromatin looping to the RC, and increased transcription and histone acetylation of genes within the affected
300 kb region. Activated gene segments underwent recombination at higher frequencies, causing a substantial alteration of the Tcrb repertoire. Therefore, our studies identified a LAD border in the Tcrb locus that served to limit the Eβ-dependent activation of RC-proximal gene segments in order to maintain the diversity of the Tcrb repertoire.
## Contents

Abstract ........................................................................................................................................ iv  
List of Tables .................................................................................................................................. x  
List of Figures .................................................................................................................................. xi  
List of Abbreviations ...................................................................................................................... xiii  
Acknowledgements ........................................................................................................................ xviii  
1. Introduction ................................................................................................................................... 1  
   1.1. The development of T lymphocytes ....................................................................................... 1  
      1.1.1. The innate and adaptive immune systems ................................................................. 1  
      1.1.2. Hematopoiesis in the bone marrow ........................................................................... 2  
      1.1.3. Thymocyte development ............................................................................................ 5  
   1.2. V(D)J recombination ............................................................................................................. 7  
      1.2.1. Regulation of V(D)J recombination ......................................................................... 8  
      1.2.2. Mechanism of V(D)J recombination ......................................................................... 10  
      1.2.3. Structure and regulation of antigen receptor loci ................................................... 12  
         1.2.3.1. Structure and regulation of the Tcrb locus ......................................................... 15  
      1.2.4. Allelic exclusion .......................................................................................................... 18  
      1.2.5. Locus contraction and chromatin looping ................................................................. 22  
   1.3. Regulation of V(D)J recombination by CTCF ................................................................. 27  
      1.3.1. The structure of CTCF ............................................................................................... 27  
      1.3.2. CTCF and cohesin ..................................................................................................... 27  
      1.3.3. CTCF-mediated looping in antigen receptor loci ................................................... 29
1.3.4. Intergenic CBEs diversify antigen receptor repertoires .......................... 30
1.4. Organization of the mammalian nucleus ...................................................... 34
1.4.1. Nuclear compartments ............................................................................. 34
1.4.2. Topological domains ............................................................................... 35
1.4.3. Lamina-associated domains at the nuclear periphery ............................ 37
1.4.4. Regulation of LADs ............................................................................... 40
1.4.5. Tcrb recombination is suppressed at the nuclear periphery .................. 42
2. Thesis prospectus ............................................................................................ 44
  2.1. Specific aim 1: To determine how Tcrb is localized to the nuclear periphery ... 44
  2.2. Specific aim 2: To characterize the regulatory impact of the usRC LAD border 45
3. Materials and methods .................................................................................... 47
  3.1. Mice ........................................................................................................... 47
  3.2. Cell culture ................................................................................................ 47
  3.3. Preparation of DNA-FISH probes ............................................................ 48
  3.4. DNA-FISH .............................................................................................. 48
  3.5. Microscopy imaging and analysis .............................................................. 50
  3.6. Flow cytometry and cell sorting ............................................................... 50
  3.7. Western blot ............................................................................................ 51
  3.8. CRISPR/Cas9 plasmid construction and transfection ............................ 52
  3.9. Retroviral packaging and transduction .................................................... 53
  3.10. RNA isolation and analysis .................................................................. 54
  3.11. DamID .................................................................................................... 55
  3.12. DamID-seq analysis .............................................................................. 57
3.13. Ctcf deletion efficiency .......................................................... 58
3.14. CTCF ChIP ........................................................................... 59
3.15. Histone ChIP ......................................................................... 61
3.16. ChIP-seq ............................................................................... 63
3.17. 3C ......................................................................................... 64
3.18. Determination of preexisting Tcrb rearrangements ....................... 66
3.19. Tcrb repertoire analysis ............................................................. 68
3.20. Statistical methods ................................................................. 70
3.21. Data deposition ...................................................................... 70

4. Specific aim 1: To determine how Tcrb is localized to the nuclear periphery .......... 71
   4.1. Introduction .......................................................................... 71
   4.2. Identification of a cell line suitable for DamID ................................ 75
   4.3. Determination of preexisting Tcrb rearrangements in VL3-3M2 cells ....... 78
   4.4. DamID on P5424 and VL3-3M2 cells ........................................ 79
   4.5. High throughput DamID sequencing in VL3-3M2 cells ...................... 81
   4.6. Deletion of LAD border regions by CRISPR/Cas9 ............................ 82
   4.7. Deletion of Eβ by CRISPR/Cas9 ................................................ 87
   4.8. Dissection of the usRC LAD border ............................................. 90
   4.9. Discussion ............................................................................ 99

5. Specific aim 2: To characterize the regulatory impact of the usRC LAD border ...... 101
   5.1. Introduction .......................................................................... 101
   5.2. Conditional knockout of Ctcf in DN thymocytes ............................... 102
   5.3. Analysis of transcription in VL3-3M2 lines ..................................... 108
5.4. A Tcrb recombination system in the VL3-3M2 cell line.......................... 114
5.5. Analysis of chromatin looping by 3C .............................................. 121
5.6. H3K27ac profile of the Tcrb locus .................................................. 124
5.7. H3K9me2 profile of the Tcrb locus .................................................. 128
5.8. Histone modifications in primary thymocytes .................................... 131
5.9. Discussion ...................................................................................... 134
6. Conclusions ...................................................................................... 139
   6.1. The generation of antigen receptor diversity ................................. 139
   6.2. NL association of the Tcrb locus ................................................... 140
   6.3. Mechanisms of LAD border regulation .......................................... 142
   6.4. Antigen receptor diversification by intergenic CBEs ....................... 145
   6.5. Targeted deletions of regulatory elements ...................................... 147
   6.6. Concluding remarks .................................................................... 148
Appendix A. R code used in high throughput sequencing analysis.............. 150
References .......................................................................................... 160
Biography ............................................................................................ 181
List of Tables

Table 1. Mutations in VL3-3M2 derivative cell lines........................................................... 53
Table 2. List of transcription primers. .................................................................................. 55
Table 3. List of DamID-qPCR primers.................................................................................. 56
Table 4. List of Ctef deletion primers .................................................................................. 59
Table 5. List of CTCF ChIP primers .................................................................................... 60
Table 6. List of H3K27me3 ChIP primers ........................................................................... 62
Table 7. List of 3C primers and probes............................................................................... 66
Table 8. List of recombination (PCR) primers ................................................................. 67
Table 9. List of recombination (Taqman) primers ............................................................. 68
Table 10. List of recombination (SYBR) primers .............................................................. 69
List of Figures

Figure 1. Hematopoiesis in the bone marrow. ................................................................. 3
Figure 2. Thymocyte development. .................................................................................. 6
Figure 3. Mechanism of V(D)J recombination. ................................................................. 12
Figure 4. Organization of antigen receptor loci. ............................................................... 14
Figure 5. Organization of the Tcrb locus. ........................................................................ 17
Figure 6. Allelic exclusion of the Tcrb locus................................................................. 20
Figure 7. 3C and related technologies............................................................................... 24
Figure 8. Regulation of Tcrb architecture by intergenic CBEs. ..................................... 33
Figure 9. Organization of LADs. .................................................................................... 38
Figure 10. Overview of DamID. ....................................................................................... 75
Figure 11. Subnuclear localization of Tcrb in candidate cell lines. ............................. 78
Figure 12. VL3-3M2 Tcrb loci. ......................................................................................... 79
Figure 13. DamID analysis of VL3-3M2 and P5424 cells. ........................................... 80
Figure 14. DamID-seq in VL3-3M2 cells ....................................................................... 82
Figure 15. LAD border deletions in VL3-3M2 cells. ..................................................... 84
Figure 16. DamID-qPCR analysis of LAD border-deleted lines. .................................... 85
Figure 17. DamID-seq in wild-type and ΔusRC lines .................................................... 87
Figure 18. DamID-qPCR analysis of Eβ-deleted lines. ................................................ 89
Figure 19. PDβ1 transcription in VL3-3M2 derivative lines ......................................... 92
Figure 20. DamID-qPCR analysis of usRC-dissected lines ....................................... 93
Figure 21. DamID-qPCR analysis of Int-dissected lines ............................................. 94
Figure 22. DamID-qPCR analysis of PDβ1-dissected lines................................. 95
Figure 23. DamID-qPCR analysis of compound-deleted lines............................. 96
Figure 24. Effect sizes of deletions in VL3-3M2 derivative lines............................ 98
Figure 25. Effect of CTCF deletion on thymocyte development............................ 105
Figure 26. Tcrb repertoire in wild-type and CTCF-deleted thymocytes............... 106
Figure 27. Efficiency of CTCF deletion in Ctf^fl/Cd2-Cre mice......................... 108
Figure 28. Tcrb locus transcription in usRC- and Eβ-deleted VL3-3M2 lines........ 110
Figure 29. Tcrb locus transcription in usRC-dissected lines................................ 113
Figure 30. Initial Tcrb recombination testing in VL3-3M2 lines......................... 115
Figure 31. Tcrb recombination in synchronized VL3-3M2 populations................. 116
Figure 32. Vβ repertoire in synchronized VL3-3M2 populations.......................... 118
Figure 33. Vβ-to-Jβ replacement rearrangements on the rearranged allele.......... 120
Figure 34. 3C profiles of the Tcrb locus using a HindIII digest......................... 122
Figure 35. 3C profiles of the Tcrb locus using a BglII digest............................. 124
Figure 36. H3K27ac ChIP-seq profile in VL3-3M2 lines.................................. 126
Figure 37. H3K27me3 enrichment in VL3-3M2 lines....................................... 127
Figure 38. H3K9me2 ChIP-seq profile in VL3-3M2 lines.................................. 129
Figure 39. Genome-wide features of LADs and LAD borders.............................. 130
Figure 40. H3K9me2 ChIP-seq profile in primary thymocytes............................. 132
Figure 41. H3K27ac ChIP-seq profile in primary thymocytes............................. 133
Figure 42. Model of usRC LAD border function in VL3-3M2 cells..................... 135
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3′RR</td>
<td>3′ regulatory region</td>
</tr>
<tr>
<td>3C</td>
<td>Chromosome conformation capture</td>
</tr>
<tr>
<td>4C</td>
<td>Circular chromosome conformation capture</td>
</tr>
<tr>
<td>53BP1</td>
<td>p53-binding protein 1</td>
</tr>
<tr>
<td>5C</td>
<td>Chromosome conformation capture carbon copy</td>
</tr>
<tr>
<td>5′PC</td>
<td>5′ Prss2 CTCF site</td>
</tr>
<tr>
<td>BAC</td>
<td>Bacterial artificial chromosome</td>
</tr>
<tr>
<td>BCR</td>
<td>B cell receptor</td>
</tr>
<tr>
<td>bp</td>
<td>Base pair</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine serum albumin</td>
</tr>
<tr>
<td>C</td>
<td>Constant</td>
</tr>
<tr>
<td>CBE</td>
<td>CTCF binding element</td>
</tr>
<tr>
<td>CD</td>
<td>Cluster of differentiation</td>
</tr>
<tr>
<td>Cer</td>
<td>Contracting element for recombination</td>
</tr>
<tr>
<td>ChIA-PET</td>
<td>Chromatin interaction analysis by paired-end tag</td>
</tr>
<tr>
<td>ChIP</td>
<td>Chromatin immunoprecipitation</td>
</tr>
<tr>
<td>CLP</td>
<td>Common lymphoid progenitor</td>
</tr>
<tr>
<td>CMP</td>
<td>Common myeloid progenitor</td>
</tr>
<tr>
<td>CTCF</td>
<td>CCCTC-binding factor</td>
</tr>
<tr>
<td>D</td>
<td>Diversity</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Dam</td>
<td>DNA adenine methyltransferase</td>
</tr>
<tr>
<td>DamID-qPCR</td>
<td>DamID analyzed by qPCR</td>
</tr>
<tr>
<td>DamID-seq</td>
<td>DamID coupled with high throughput sequencing</td>
</tr>
<tr>
<td>DC</td>
<td>Dendritic cell</td>
</tr>
<tr>
<td>DN</td>
<td>CD4^-CD8^- double negative</td>
</tr>
<tr>
<td>DNA-PKcs</td>
<td>DNA-dependent protein kinase catalytic subunit</td>
</tr>
<tr>
<td>DP</td>
<td>CD4^+CD8^+ double positive</td>
</tr>
<tr>
<td>DSB</td>
<td>Double-stranded DNA break</td>
</tr>
<tr>
<td>DZNep</td>
<td>3-deazaneplanocin A</td>
</tr>
<tr>
<td>ETP</td>
<td>Early thymic progenitor</td>
</tr>
<tr>
<td>EZH2</td>
<td>Enhancer of zeste homolog 2</td>
</tr>
<tr>
<td>FDG</td>
<td>Fluorescein di-b-galactoside</td>
</tr>
<tr>
<td>FISH</td>
<td>Fluorescence <em>in situ</em> hybridization</td>
</tr>
<tr>
<td>gDNA</td>
<td>Genomic DNA</td>
</tr>
<tr>
<td>GFP</td>
<td>Green fluorescent protein</td>
</tr>
<tr>
<td>GMP</td>
<td>Granulocyte/macrophage progenitor</td>
</tr>
<tr>
<td>H3ac</td>
<td>Histone H3 acetylation</td>
</tr>
<tr>
<td>H3K4me2</td>
<td>Histone H3 lysine 4 dimethylation</td>
</tr>
<tr>
<td>H3K4me3</td>
<td>Histone H3 lysine 4 trimethylation</td>
</tr>
<tr>
<td>H3K9me2</td>
<td>Histone H3 lysine 9 dimethylation</td>
</tr>
<tr>
<td>H3K27ac</td>
<td>Histone H3 lysine 27 acetylation</td>
</tr>
<tr>
<td>Acronym</td>
<td>Definition</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>H3K27me3</td>
<td>Histone H3 lysine 27 trimethylation</td>
</tr>
<tr>
<td>HBSS</td>
<td>Hank's balanced salt solution</td>
</tr>
<tr>
<td>HDAC3</td>
<td>Histone deacetylase 3</td>
</tr>
<tr>
<td>HMGB1</td>
<td>High mobility group box 1</td>
</tr>
<tr>
<td>HP1</td>
<td>Heterochromatin protein 1</td>
</tr>
<tr>
<td>HSC</td>
<td>Hematopoietic stem cell</td>
</tr>
<tr>
<td>IGCR1</td>
<td>Intergenic control region 1</td>
</tr>
<tr>
<td>Igh</td>
<td>Immunoglobulin heavy chain</td>
</tr>
<tr>
<td>Igk</td>
<td>Immunoglobulin kappa light chain</td>
</tr>
<tr>
<td>Igl</td>
<td>Immunoglobulin lambda light chain</td>
</tr>
<tr>
<td>INT</td>
<td>Intergenic CTCF binding element</td>
</tr>
<tr>
<td>ISP</td>
<td>Immature CD8(^+) single positive</td>
</tr>
<tr>
<td>J</td>
<td>Joining</td>
</tr>
<tr>
<td>kb</td>
<td>Kilobase</td>
</tr>
<tr>
<td>lacO</td>
<td>lac operator</td>
</tr>
<tr>
<td>LacI</td>
<td>lac repressor</td>
</tr>
<tr>
<td>LAD</td>
<td>Lamina-associated domain</td>
</tr>
<tr>
<td>LAP</td>
<td>Lamina-associated peptide</td>
</tr>
<tr>
<td>LAT</td>
<td>Linker for activation of T cells</td>
</tr>
<tr>
<td>LBR</td>
<td>Lamin B receptor</td>
</tr>
<tr>
<td>Lck</td>
<td>Lymphocyte protein tyrosine kinase</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Term</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>LCR</td>
<td>Locus control region</td>
</tr>
<tr>
<td>Mb</td>
<td>Megabase</td>
</tr>
<tr>
<td>MeCP2</td>
<td>Methyl CpG binding protein 2</td>
</tr>
<tr>
<td>MEP</td>
<td>Megakaryocyte/erythroid progenitor</td>
</tr>
<tr>
<td>MHC</td>
<td>Major histocompatibility complex</td>
</tr>
<tr>
<td>MPP</td>
<td>Multipotent progenitor</td>
</tr>
<tr>
<td>NHEJ</td>
<td>Nonhomologous end joining</td>
</tr>
<tr>
<td>NIPBL</td>
<td>Nipped-B-like protein</td>
</tr>
<tr>
<td>NK</td>
<td>Natural killer</td>
</tr>
<tr>
<td>NL</td>
<td>Nuclear lamina</td>
</tr>
<tr>
<td>PAMP</td>
<td>Pathogen-associated molecular pattern</td>
</tr>
<tr>
<td>PAX5</td>
<td>Paired box protein 5</td>
</tr>
<tr>
<td>PHD</td>
<td>Plant homeodomain</td>
</tr>
<tr>
<td>Pol II</td>
<td>RNA polymerase II</td>
</tr>
<tr>
<td>pre-Tα</td>
<td>Pre-T cell receptor α</td>
</tr>
<tr>
<td>PRR14</td>
<td>Proline rich protein 14</td>
</tr>
<tr>
<td>PRR</td>
<td>Pattern recognition receptor</td>
</tr>
<tr>
<td>qPCR</td>
<td>Quantitative PCR</td>
</tr>
<tr>
<td>RAG</td>
<td>Recombination activating gene</td>
</tr>
<tr>
<td>RB</td>
<td>Retinoblastoma protein</td>
</tr>
<tr>
<td>RC</td>
<td>Recombination center</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>RSS</td>
<td>Recombination signal sequence</td>
</tr>
<tr>
<td>RUNX3</td>
<td>Runt related transcription factor 3</td>
</tr>
<tr>
<td>SA</td>
<td>Stromal antigen</td>
</tr>
<tr>
<td>SE</td>
<td>Standard error</td>
</tr>
<tr>
<td>Sis</td>
<td>Silencer in the intervening sequence</td>
</tr>
<tr>
<td>SMC</td>
<td>Structural maintenance of chromosomes</td>
</tr>
<tr>
<td>SP</td>
<td>Single positive</td>
</tr>
<tr>
<td>SSC</td>
<td>Saline-sodium citrate</td>
</tr>
<tr>
<td>TAD</td>
<td>Topologically associating domain</td>
</tr>
<tr>
<td>TCR</td>
<td>T cell receptor</td>
</tr>
<tr>
<td>TdT</td>
<td>Terminal deoxynucleotidyl transferase</td>
</tr>
<tr>
<td>TEA</td>
<td>T early-α</td>
</tr>
<tr>
<td>ThPOK</td>
<td>T-helper inducing POZ/Kruppel-like factor</td>
</tr>
<tr>
<td>UTR</td>
<td>Untranslated region</td>
</tr>
<tr>
<td>XRCC4</td>
<td>X-ray repair cross-complementing protein 4</td>
</tr>
<tr>
<td>V</td>
<td>Variable</td>
</tr>
<tr>
<td>WAPL</td>
<td>Wings apart-like homolog</td>
</tr>
<tr>
<td>YY1</td>
<td>Yin Yang 1</td>
</tr>
<tr>
<td>ZF</td>
<td>Zinc finger</td>
</tr>
</tbody>
</table>
Acknowledgements

I would like to thank my mentor, Mike, for all of the guidance throughout the course of my research. He taught me extensively about critical thinking, experimental design and performing good science.

Many thanks to all current and former lab members that I have been privileged to work alongside. Thank you Beth, Han-Yu, Ebeth, Abani, Zanchun, Liang, Zach, Danielle, Aaron and Marie. All of your companionship and advice have made my time in graduate school enjoyable.

Several individuals directly contributed in performing the research outlined here. DamID-seq was performed in collaboration with the Reddy Lab at Johns Hopkins University. Aaron helped in creating and screening some of the VL3-3M2 derivative cell lines generated. Europe, a rotation student, performed some of the ChIP experiments.

This work would not have been possible without the support of non-research staff and core facilities at Duke. The Immunology Business Office went above and beyond to help with any request, big or small. Finally, I would like to thank the Duke Cancer Institute Flow Cytometry Shared Resource, Duke Light Microscopy Core Facility and the Duke Center for Genomic and Computational Biology for the consultation sessions and teaching of techniques that were instrumental in completing my research.
1. Introduction

1.1. The development of T lymphocytes

1.1.1. The innate and adaptive immune systems

The mammalian immune system can be broadly divided into innate and adaptive arms. The innate immune system forms the first layer of defense against invading pathogens, mounting an immediate response at the site of infection. If the pathogen is not successfully cleared, the adaptive immune system acts as a second layer of defense, creating a slower but directed response against antigens expressed by pathogens or pathogen-infected host cells (Chaplin, 2010; Warrington et al., 2011).

The innate immune system is comprised of physical barriers, non-cellular humoral components and also a plethora of cell types which include phagocytes, granulocytes, innate lymphoid cells and innate-like lymphocytes. These innate immune cells not only directly act on pathogens in order to clear them, but also activate and amplify the effects of other components of the immune system (Iwasaki and Medzhitov, 2015). The adaptive immune system is split into cellular and humoral arms, which are mediated by T and B lymphocytes respectively. T lymphocytes can be divided functionally and phenotypically based on the surface expression of cluster of differentiation (CD) -4 and -8. CD4+ T lymphocytes are helper cells that direct the correct type of immune response against a given pathogen, while CD8+ T lymphocytes are cytotoxic cells that directly mediate the killing of infected cells. B lymphocytes secrete antibodies which bind to their target antigen, mediating an array of downstream responses like neutralization, complement activation or activation of other effector cell
types. After a pathogenic infection is cleared, long-lived memory T and B lymphocytes remain in the host, and are capable of mounting a faster and stronger response if the same pathogen is encountered later in life (Chaplin, 2010; Warrington et al., 2011).

A major difference between innate and adaptive immune cells is the nature of the receptors expressed on the component cell types. Innate immune cells utilize pattern recognition receptors (PRRs) to recognize evolutionarily-conserved pathogen-associated molecular patterns (PAMPs) which are expressed on broad groups of pathogens. PRRs are encoded in the germline, and innate immune cells can express multiple types of PRRs on the same cell. In contrast, T and B lymphocytes recognize antigens using T cell receptors (TCRs) and B cell receptors (BCRs) respectively. TCRs and BCRs are not encoded in the germline but are instead generated by a process of somatic recombination termed V(D)J recombination, which allows for the generation of unique receptors that recognize a single antigen on each T and B lymphocyte (Brubaker et al., 2015; Chaplin, 2010; Jimenez-Dalmaroni et al., 2016; Yin et al., 2015).

1.1.2. Hematopoiesis in the bone marrow

The majority of immune cells are derived from the process of hematopoiesis. During fetal development, hematopoiesis occurs in the yolk sac, liver and spleen until the bone marrow develops, after which the bone marrow takes over as the primary organ for hematopoiesis. Hematopoiesis occurs by the differentiation of progenitor cells into cells with declining lineage potential, eventually giving rise to terminally differentiated cells (Figure 1).
Hematopoiesis in the bone marrow.

The differentiation hierarchy begins with HSCs and gives rise to all hematopoietic lineages. HSC, hematopoietic stem cell; MPP, multipotent progenitor; CMP, common myeloid progenitor; CLP, common lymphoid progenitor; MEP, megakaryocyte/erythroid progenitors; GMP, granulocyte/macrophage progenitor; NK, natural killer.

Hematopoietic stem cells (HSCs) give rise to all hematopoietic lineages and are the only cell type in the bone marrow that exhibits long-term self-renewal and capacity for differentiation. HSCs first differentiate into multipotent progenitors (MPPs), which retain the lineage potential but not the unlimited self-renewal capability of HSCs. The first bifurcation point in the hematopoietic lineage occurs during the differentiation of...
MPPs into either common myeloid progenitors (CMPs) or common lymphoid progenitors (CLPs). CMPs give rise to two further precursor cell types: megakaryocyte/erythroid progenitors (MEPs) and granulocyte/macrophage progenitors (GMPs). MEPs develop into erythrocytes and platelets, which are non-leukocyte components of blood. GMPs generate most of the innate immune cell types which include neutrophils, basophils, eosinophils, mast cells, monocytes, macrophages and dendritic cells (DCs). In the other lineage pathway, CLPs mostly give rise to B, T and natural killer (NK) lymphoid cell types, although some myeloid potential remains (Hofer et al., 2016; Rieger and Schroeder, 2012). T cells do not develop in the bone marrow but instead in the thymus, which will be described in the next section.

The study of hematopoiesis has primarily been driven by the sorting of cells based on surface markers and transplantation into myeloablated recipient animals (Hofer et al., 2016; Rieger and Schroeder, 2012). Recent improvements in single cell fate mapping and sequencing technologies have shown that the characterization of progenitor populations based on cell surface markers obscures the underlying heterogeneity within these populations. In particular, the intermediate oligopotent cell types are actually comprised of multiple independent lineages with determined fates (Alberti-Servera et al., 2017; Karamitros et al., 2018; Paul et al., 2015; Perie et al., 2015; Rodriguez-Fricelli et al., 2018).
1.1.3. Thymocyte development

The thymus is the organ in which thymocytes undergo development to form mature T lymphocytes. The stages of thymocyte development are commonly tracked by the expression of a panel of surface markers. The thymus is seeded with CLPs that migrate from the bone marrow. Thymocyte development begins at the CD4<sup>-</sup>CD8<sup>-</sup> double negative (DN) stage, which can be further subdivided into four stages going from DN1 to DN4, distinguished by the expression of c-kit, CD44 and CD25 (Figure 2). The DN1 stage contains early thymic progenitors (ETPs), which are the earliest known intrathymic T cell progenitors. The DN2 stage is when irreversible commitment to the T lineage occurs, with pre-committed and post-committed stages divided into DN2a and DN2b respectively. The DN3 stage marks the completion of recombination at the T<sub>crb</sub>, T<sub>crg</sub> and T<sub>crd</sub> loci that encode the TCR-β, -γ and -δ chains respectively. TCRβ ultimately pairs with TCRα, which undergoes recombination at a later stage, to form TCRαβ heterodimers, while TCRγ and TCRδ pair to form TCRγδ heterodimers. Cells with functional TCRγδ receptors exit the thymus as mature γδ T cells, while the remaining cells undergo a checkpoint step, termed β selection, in which TCRβ pairs with pre-Tα to allow selection of cells bearing a functional T<sub>crb</sub> rearrangement. Cells that pass the β selection checkpoint progress into the DN4 stage, where they undergo multiple rounds of proliferation (Carpenter and Bosselut, 2010; Rothenberg, 2014).

Cells progressing from the DN4 stage begin to express the characteristic T cell surface markers, CD4 and CD8, to form CD4<sup>+</sup>CD8<sup>+</sup> double positive (DP) cells. This process includes a transitional immature CD8<sup>+</sup> single positive (ISP) stage, where CD8 is
first expressed before CD4. DP cells undergo recombination of the Tcra locus that encodes the TCRα chain, the second half of the TCRαβ heterodimer. DP thymocytes undergo positive selection, which is the process of selecting for thymocytes with TCRs that can productively interact with major histocompatibility complexes (MHC) bearing self-peptides. This is supported by the capacity of the Tcra locus to undergo multiple rounds of rearrangement in order to generate a functional TCR. DP thymocytes that fail to generate a TCR that can interact with self-peptide loaded MHC die of neglect, while positively-selected thymocytes progress to the next stage of thymocyte development (Carpenter and Bosselut, 2010).

**Figure 2. Thymocyte development.**

The process of thymocyte development in the thymus, starting from the ETP/DN1 stage. Cells with successful rearrangements of Tcrg and Tcrd exit the thymus as γδ T cells. Important transition checkpoints are indicated by dotted lines. Stages that express RAG are indicated by the filled grey arrows, with antigen receptor loci undergoing recombination indicated below. ETP, early thymic progenitor; DN, double negative; ISP, intermediate single positive; SP, single positive; RAG, recombination activating gene.
Positively-selected DP thymocytes downregulate the expression of either CD4 or CD8 to become single positive (SP) thymocytes. The downregulation of CD4 or CD8 is dependent on the class of MHC that the TCR binds to. Class I MHCs are co-bound by TCR and CD8, and cells bearing such TCRs downregulate CD4. The converse is true of class II MHCs which are co-bound by TCR and CD4, resulting in the downregulation of CD8. The expression of CD4 or CD8 in SP thymocytes is regulated by the master transcription factors T-helper inducing POZ/Kruppel-like factor (ThPOK) and Runx related transcription factor 3 (RUNX3) respectively (Luckey et al., 2014; Woolf et al., 2003).

Thymocytes undergo clonal deletion, a process of eliminating cells bearing TCRs with high avidity for self MHCs in order to prevent the escape of cells predisposed to autoimmunity. CD4+ SP cells with intermediate affinities for self MHCs develop into regulatory T cells, a class of immunosuppressive T cells that modulate the immune response (Carpenter and Bosselut, 2010; Josefowicz et al., 2012). Thymocytes that are not clonally deleted exit the thymus into the circulatory system as mature T cells. Similar to SP thymocytes, the transcriptional program of mature CD4 and CD8 T cells is regulated and enforced by the master transcription factors ThPOK and RUNX3 respectively (He et al., 2010).

1.2. V(D)J recombination

V(D)J recombination is a process where variable (V), diversity (D) and joining (J) gene segments undergo somatic recombination and are assembled alongside a constant
(C) gene segment to form a functional antigen receptor gene. This is a highly-regulated process that involves the formation and resolution of targeted double-stranded DNA breaks (DSBs) and is critically dependent on recombination activating gene (RAG) -1 and -2 proteins, collectively referred to as RAG. RAG1 directly binds and cleaves DNA while RAG2 stabilizes RAG-DNA interactions and also targets the RAG complex to regions of open chromatin (Schatz and Ji, 2011; Teng and Schatz, 2015).

1.2.1. Regulation of V(D)J recombination

Due to the ability of the RAG recombinase to bind and cleave DNA, it is important to ensure that RAG activity is tightly regulated. Inappropriate rearrangements resulting from off-target RAG activity can lead to chromosomal translocations that mediate oncogene activation, resulting in the development of lymphomas and leukemias (Lieber et al., 2006; Teng and Schatz, 2015).

The expression of the Rag genes is tightly controlled in developing lymphocytes (Kuo and Schlissel, 2009). Thymocytes express Rag genes at the DN stage for the recombination of Tcrb, Tcrd and Tcrg, and at the DP stage for the recombination of Tcra. In developing B lymphocytes, Rag is expressed at the pro-B cell stage for the recombination of Igh, and at the pre-B cell stage for the recombination of Igk and Igλ.

RAG recruitment and cleavage at antigen receptor loci is directed by short DNA elements known as recombination signal sequences (RSSs). V segments possess an RSS at their 3’ ends, D segments possess RSSs at both ends and J segments possess an RSS at their 5’ ends. RSSs are comprised of conserved heptamer and nonamer DNA sequences
separated by a spacer that is usually 12 or 23 base pairs (bp) in length. The length of the spacer is important for the pairing of RSSs, with an RSS containing a 12 bp spacer pairing with an RSS with a 23 bp spacer. This is termed the 12/23 rule (Schatz and Ji, 2011; Schatz and Swanson, 2011).

RAG recruitment to chromatin is not only limited to antigen receptor loci, but was also shown to occur at thousands of sites across the genome (Ji et al., 2010; Maman et al., 2016; Teng et al., 2015). The genome-wide binding of RAG2 is directed by its plant homeodomain (PHD), which recruits RAG2 to histone H3 lysine 4 trimethylation (H3K4me3) marks (Ji et al., 2010). The genome-wide binding of RAG1 is directed to regions of active chromatin, primarily by targeting H3K4me3-enriched regions bound by RAG2, and secondarily by direct binding to RSS-like sequences, termed cryptic RSSs, that are enriched for histone H3 lysine 27 acetylation (Maman et al., 2016; Teng et al., 2015).

At antigen receptor loci the remodeling of chromatin from a repressive to an accessible state is an important prerequisite for RAG binding. Germline transcription from promoters removes repressive histone modifications like histone H3 lysine 9 dimethylation (H3K9me2) in favor of activating modifications like acetylated histones and H3K4me3 (Abarrategui and Krangel, 2006, 2007; Yancopoulos and Alt, 1985). Nucleosome remodeling at RSSs also serve to uncover the RSSs in order to make them accessible for binding and cleavage by RAG (Bevington and Boyes, 2013; Kondilis-Mangum et al., 2010; Kwon et al., 1998). The focal enrichment of RAG at the highly
accessible RSSs of D and J gene segments forms a structure known as the recombination center (RC) (Schatz and Ji, 2011).

1.2.2. Mechanism of V(D)J recombination

The process of V(D)J recombination is ordered, with antigen receptor loci containing D gene segments undergoing D-to-J recombination before V-to-DJ recombination (Helmink and Sleckman, 2012).

V(D)J recombination begins with the direct binding of RAG to one RSS (Figure 5). The capture of the second RSS is termed synapsis, and results in the formation of the paired complex which contains a tetramer of two RAG1 and two RAG2 proteins (Jones and Gellert, 2002). The heptamers of the RSSs are stacked in a parallel orientation bound by the heptamer binding domain of RAG1, while the nonamers are stacked in an antiparallel orientation bound by the nonamer binding domain of RAG, facilitated by bending of the 12- and 23-spacers bound by high mobility group box 1 (HMGB1) (Kim et al., 2018; Yin et al., 2009). RAG1 creates DSBs immediately adjacent to the heptamer, by first creating a nick on one DNA strand by hydrolysis of the DNA phosphodiester backbone. This causes the flipping of the first base of the heptamer, which promotes the attack of the opposite DNA strand by the liberated 3’ hydroxyl group in a transesterification reaction, creating a hairpin loop termed the coding end. This leaves the RSS with a blunt-ended DSB termed the signal end (Kim et al., 2018; Ru et al., 2015; Yin et al., 2009).
The coding and signal ends are repaired by the nonhomologous end joining (NHEJ) DNA repair pathway. The hairpin loops at the coding ends are bound by KU70/80 and DNA-dependent protein kinase catalytic subunit (DNA-PKcs), following which Artemis cleaves a short distance away from the hairpin dyad to form a palindromic sticky end (Ma et al., 2002). Terminal deoxynucleotidyl transferase (TdT) next catalyzes the addition of random nucleotides to the sticky ends (Gilfillan et al., 1993; Komori et al., 1993). The two coding ends anneal by the formation of small regions of microhomology, which can be facilitated by the activity of exonucleases and the insertion of additional nucleotides by the X family of DNA polymerases. Finally, the coding ends are joined together by X-ray repair cross-complementing protein 4 (XRCC4) and DNA ligase IV, forming the coding joint (Gu et al., 2007; Mahajan et al., 2002). The resolution of the coding ends results in the formation of a highly variable junction that increases the diversity of antigen receptors. Meanwhile, the two blunt-ended signal ends are ligated precisely to form the signal joint.
12

Figure 3. Mechanism of V(D)J recombination.

Overview of the phases of V(D)J recombination. RAG is depicted as a blue oval. Two unrearranged gene segments with compatible RSSs are depicted in different colors. The coding joint contains variable non-templated sequences represented in a third color.

1.2.3. Structure and regulation of antigen receptor loci

The seven antigen receptor genes are encoded by six antigen receptor loci (Figure 3) (Jhunjhunwala et al., 2009). Three loci encode the TCR chains utilized by T cells: the Tcra/Tcrod locus encodes both the TCRα and -δ chains, the Tcrb locus encodes the TCRβ
chain and the $Tcr\gamma$ locus encodes the TCR$\gamma$ chain. The $Tcra/Tcrd$ locus is unique in that the $Tcr\delta$ locus is fully nested within the $Tcra$ locus, and recombination of the $Tcra$ locus causes the deletion of the $Tcr\delta$ locus. Furthermore, some of the V gene segments are utilized in both $Tcra$ and $Tcr\delta$ recombination, while other V segments are exclusive to either $Tcra$ or $Tcr\delta$. The other three loci encode immunoglobulin chains utilized by B cells: the $Igh$ locus encodes the immunoglobulin heavy chain, the $Igk$ locus encodes the immunoglobulin $\kappa$ light chain and the $Igl$ locus encodes the immunoglobulin $\lambda$ light chain. All loci possess V, J and C gene segments, while only the $Tcr\delta$, $Tcrb$ and $Igh$ loci possess D gene segments.

Of the six antigen receptor loci, four of these ($Tcra/Tcrd$, $Tcrb$, $Igh$ and $Igk$) are large and span hundreds of kilobases (kb) to megabases (Mb) in size. These four loci have a similar structure, where the V gene segments are distributed across a large fraction of the locus, while the D (if present), J and C gene segments are found towards one end of the locus. The $Tcr\gamma$ and $Igl$ loci are smaller and only about 200 kb in size, containing limited numbers of V, J and C gene segments. The positioning of the V gene segments in the V gene array is generally described relative to the position of the J gene segments, with proximal V gene segments found closer to the J gene segments while distal V gene segments are farther away.
Figure 4. Organization of antigen receptor loci.

Diagram of the six murine antigen receptor loci, not drawn to scale. Names of important regulatory elements are shown above the respective elements. CBE, CTCF binding element.

Multiple regulatory elements are scattered throughout the large and complex antigen receptor loci (Bossen et al., 2012). Promoters are found upstream of each V gene segment and also upstream of, or interspersed within the D or J gene segment arrays. These promoters direct germline transcription of their respective gene segments, and the V promoters are also required for the expression of fully-recombined genes. Each antigen
receptor locus also contains at least one enhancer element, which mediates stage-specific transcriptional activation and also changes in the landscape of histone modifications. Enhancers can also regulate other aspects of antigen receptor locus control. For example, the 3′ regulatory region (3’RR) enhancer in the Igh locus controls the process of class switch recombination, which swaps the constant region in the immunoglobulin heavy chain without altering its antigen binding specificity (Pinaud et al., 2011). Several antigen receptor loci also possess intergenic binding elements for the architectural protein CCCTC-binding factor (CTCF), which will be further described in section 1.3.4.

1.2.3.1. Structure and regulation of the Tcrb locus

The murine Tcrb locus resides on chromosome 6 and spans approximately 700 kb (Figure 4). The Tcrb locus consists of 35 Vβ, 2 Dβ, 13 Jβ and 2 Cβ gene segments. Of these, 12 Vβ and 1 Jβ gene segments are non-functional pseudogenes. 33 of the 35 Vβ gene segments are scattered across a 250 kb region in the middle of the locus, flanked by regions of trypsinogen genes of 150 kb and 250 kb in size at the 5′ and 3′ ends respectively. These trypsinogen genes are normally silent in the T cell lineage. The two remaining Vβ gene segments are found on the two extreme ends of the Tcrb locus. The Dβ, Jβ and Cβ gene segments are arranged into two clusters, with promoters associated with each cluster. The first cluster contains a promoter upstream of the Dβ gene segment, named PDβ1, while the second cluster has two promoters, named 5′ and 3′-PDβ2 (Majumder et al., 2015a; McMillan and Sikes, 2008, 2009). These promoters drive transcription of their respective clusters (Oestreich et al., 2006; Whitehurst et al., 1999;
Whitehurst et al., 2000), and are preferentially active in thymocytes but not mature T cells (Doty et al., 1999; Sikes et al., 1998).

The Tcrb locus contains a single enhancer E_β, which is located 6 kb downstream of the DJC_β clusters. E_β is specifically active in the T cell lineage and is activated at the earliest stages of T lineage commitment (McDougall et al., 1988). E_β is critical for Tcrb regulation as its deletion completely blocks Tcrb rearrangement and thus thymocyte development (Bories et al., 1996; Bouvier et al., 1996). In developing thymocytes, E_β mediates chromatin remodeling, histone acetylation and germline transcription of the PD_β1 and PD_β2 promoters, but has no effect or only a partial effect on V_β gene segment promoters in unrearranged Tcrb loci (Bouvier et al., 1996; Majumder et al., 2015b; Mathieu et al., 2000; McDougall et al., 1988; Spicuglia et al., 2002). E_β contains binding sites for several transcription factors including ETS1, RUNX1, E2A, GATA3 and Kruppel-like factors (KLF) (Bonnet et al., 2009; Tripathi et al., 2000). Of these transcription factor binding sites, a pair of composite ETS/RUNX motifs has been shown to be essential for E_β functionality, with a partial loss of E_β functionality upon the deletion of one ETS/RUNX motif, and a complete loss of functionality upon the deletion of both motifs (Bonnet et al., 2009; Zhao et al., 2017).

The nomenclature used to describe Tcrb gene segments has undergone several changes in order to remove unnecessary complexity in describing these gene segments (Arden et al., 1995; Bosc and Lefranc, 2000; Wilson et al., 1988). Tcrb genes are named “Trb” followed by the type of gene segment “v”, “d”, “j” or “c” followed by a number that corresponds to the position along the Tcrb locus. Closely-related V_β gene segments
belonging to the same family share a same number but also have a secondary identifier indicating the relative localization within the locus. For example, the first V\(_{\beta}\) gene segment is named \(Trbv1\), while the three genes belonging to the \(Trbv13\) family are named \(Trbv13-1\), \(Trbv13-2\) and \(Trbv13-3\). J\(_{\beta}\) gene segments are numbered based on their location within the two DJC\(_{\beta}\) clusters. Protein products of the gene segments retain the original numbering system. For example, the protein product of \(Trbv1\) is V\(_{\beta}2\). Gene segments of the other TCR genes follow a similar nomenclature.

**Figure 5. Organization of the Tcrb locus.**

Diagram of the murine Tcrb locus, drawn to scale. The entire locus is represented on top, with an expanded view of the recombination center (RC) and proximal elements below. Arrow direction represents the orientation of the CTCF binding element.

Recombination of the Tcrb locus occurs in DN thymocytes, with D\(_{\beta}\)-to-J\(_{\beta}\) recombination preceding V\(_{\beta}\)-to-DJ\(_{\beta}\) recombination. The order of recombination is maintained by the 12/23 rule of RSS pairing, with 5’ 12 RSSs on J\(_{\beta}\) segments pairing with
3′ 23 RSSs on Dβ segments. During Vβ-to-DJβ recombination, 5′ 12 RSSs on Dβ segments pair with 3′ 23 RSSs on Vβ segments. Interestingly, despite the fact that Vβ and Jβ gene segments have compatible RSSs, direct Vβ-to-Jβ recombination is not observed in vivo. This restriction is known as the beyond 12/23 rule, and the mechanism behind it has yet to be elucidated (Bassing et al., 2000; Sleckman et al., 2000).

The factors that influence Tcrb recombination have been characterized in great detail (Gopalakrishnan et al., 2013). This study took into account a multitude of factors, including transcription, binding of transcriptional mediators, histone modifications, nucleosome occupancy, gene segment location, association with a nearby CTCF binding element (CBE) and RSS quality. The authors showed that non-pseudogene Dβ and Jβ gene segments were utilized equally efficiently. Vβ gene segment usage varied, and was be predicted by a two-step model where the first step is a binary predictor of whether the gene segment is utilized or inert based on RSS quality and nucleosome occupancy. The second step predicted the relative usage of Vβ gene segments based on five features: histone H3 lysine 4 di/trimethylation (H3K4me2/3), transcription, binding of the transcriptional activator p300 and presence of a nearby CBE.

1.2.4. Allelic exclusion

Allelic exclusion is a regulatory mechanism in which active antigen receptor genes in lymphocytes are only expressed from a single allele, resulting in the generation of antigen receptors with a single specificity. The extent of allelic exclusion varies between antigen receptor genes, with the Igh and Tcrb loci exhibiting the tightest allelic
exclusion with 99.99% of B cells and 97-99% of T cells exhibiting allelic exclusion of their respective loci. In contrast, allelic exclusion at the \( Tcra \) locus is less stringent; only 90% of T cells exhibit \( Tcra \) allelic exclusion (Brady et al., 2010b).

Allelic exclusion is partly caused by the random nature of coding joint formation during V(D)J recombination. One-third of rearrangements on a given allele are in-frame and can produce a functional protein product. Assuming both alleles undergo rearrangement, this leaves one-ninth of cells with in-frame rearrangements on both alleles, four-ninths of cells with in-frame rearrangements on one allele, and the remaining four-ninths of cells with no in-frame rearrangements. As cells with no functional antigen receptor genes fail to survive selection, this leaves a theoretical maximum of one-fifth of mature T or B cells with in-frame rearrangements on both alleles. Therefore, this implies that additional mechanisms beyond random probability are required for allelic exclusion.

\( Tcrb \) allelic exclusion is initiated by asynchronous recombination of \( Tcrb \) loci (Figure 6). Both \( Tcrb \) alleles undergo \( D\beta \)-to-\( J\beta \) recombination, but only one \( Tcrb \) allele undergoes \( V\beta \)-to-\( DJ\beta \) recombination at a time. The generation of an in-frame \( Tcrb \) rearrangement allows its assembly with pre-T cell receptor \( \alpha \) (pre-T\( \alpha \)) and CD3 to create the pre-TCR. Pre-TCR signaling allows the thymocyte to pass the \( \beta \) selection checkpoint and to start the process of feedback inhibition, which suppresses further \( Tcrb \) recombination by activation and repression of the transcription factors ETS1 and E47 respectively (Agata et al., 2007; Eyquem et al., 2004). If the first \( Tcrb \) allele does not produce a productive rearrangement, then rearrangement of the second allele can commence.
Figure 6. Allelic exclusion of the Tcrb locus.

Representation of Tcrb allelic exclusion. Both alleles undergo Dβ-to-Jβ recombination, while Vβ-to-DJβ recombination occurs on one allele at a time. If the first allele undergoes a functional recombination, then Vβ-to-DJβ recombination of the second allele does not commence.

In order to maintain allelic exclusion, Tcrb rearrangement on unrearranged alleles has to be suppressed during the DP thymocyte stage where RAG is expressed. During the transition from DN to DP thymocytes, most unrearranged Vβ gene segments downregulate accessibility and lose activating histone modifications, a reversal of the features that allowed for the initiation of recombination (Tripathi et al., 2002). However, this is complicated by the fact that the VDJβ-rearranged alleles cause activation of upstream Vβ gene segments by removing chromatin separating the Vβ gene segments from Eβ (Jackson and Krangel, 2005; Jia et al., 2007). In theory, the allele with an out-of-frame rearrangement to the first DJCβ cluster has a chance to perform a second
rearrangement using an upstream $V_\beta$ and the second DJC$_\beta$ cluster, which could result in a second in-frame $Tcrb$ rearrangement (Brady et al., 2010a). This implies that additional mechanisms are in place to prevent such rearrangements from occurring.

The idea that reduced RAG accessibility is key to maintaining allelic exclusion was tested by the creation of mouse models that perturbed $Tcrb$ locus accessibility in DP thymocytes. The knock-in of the $Tcra$ enhancer $E_a$ enabled $V_\beta$ gene segments near the knocked-in enhancer to be maintained in an active state in DP thymocytes. While allelic exclusion was not disrupted in the DP stage, there was a partial disruption of allelic exclusion that was presumed to occur in the DN stage through unknown mechanisms (Jackson et al., 2005). A further mouse model was created where a pre-rearranged DJ$_\beta$ gene segment was introduced proximal to the $E_a$ knock-in. This mouse model exhibited rearrangements of nearby $V_\beta$ gene segments to the knocked-in DJ$_\beta$ segments, showing that allelic exclusion was successfully broken (Kondilis-Mangum et al., 2011). This led to the conclusion that allelic exclusion was enforced by locus accessibility and also the physical separation of $V_\beta$ and DJ$_\beta$ gene segments over long distances of chromatin. However, another mouse model that deleted 475 kb of the $Tcrb$ locus spanning $Trbv5$ to $Trbj1\text{-}6$, inclusive, contradicted this hypothesis. Even though this mouse model also exhibited increased $V_\beta$ gene segment accessibility and greatly shortened distances between $V_\beta$ and DJ$_\beta$ gene segments, cells from these mice could not break allelic exclusion nor showed a measurable increase in the fraction of cells with biallelic VDJ$_\beta$ rearrangements (Senoo et al., 2003). Therefore, the break of allelic exclusion in the
previous model was primarily mediated by the knock-in of Eα and not observed under natural conditions.

1.2.5. Locus contraction and chromatin looping

The large size of antigen receptor loci necessitates mechanisms to bring distal gene segments scattered over hundreds of kb or even several Mb along the linear chromosome into spatial proximity in order for recombination to occur. Changes in locus conformation that bring together or separate distal regions of chromatin are termed contraction or decontraction respectively. DNA fluorescence in situ hybridization (DNA-FISH) has typically been used to study locus contraction and decontraction, by hybridizing fluorescent DNA probes to different sections of the antigen receptor locus and quantifying the three-dimensional distance between hybridized foci.

Locus contraction at antigen receptor loci is developmentally-regulated and linked to recombination. The Igh locus undergoes contraction in pro-B cells when the Igh locus undergoes recombination, and decontracts in pre-B cells when Igh recombination has completed (Jhunjhunwala et al., 2008; Kosak et al., 2002; Roldan et al., 2005; Sayegh et al., 2005). The other antigen receptor loci also undergo similar stage-specific changes in contraction and decontraction (Kondilis-Mangum et al., 2011; Majumder et al., 2015c; Roldan et al., 2005; Shih and Krangel, 2010). Locus decontraction is hypothesized to be one of the factors that facilitate allelic exclusion by reversing the changes in locus conformation necessary for recombination to take place (Roldan et al., 2005).
The development of chromosome conformation capture (3C) and 3C-derived technologies has enabled the study of the spatial organization of chromatin at resolutions not possible with DNA-FISH (Dekker et al., 2002). 3C and its derived technologies detect long-range chromatin interactions by digesting crosslinked chromatin and performing proximity ligation on the digested chromatin (Figure 7) (Bonev and Cavalli, 2016; Davies et al., 2017; Schmitt et al., 2016). These techniques vary in terms of resolution and coverage. All 3C-based technologies follow a similar workflow that begins by crosslinking chromatin with formaldehyde. Chromatin is fragmented by a restriction enzyme digest to yield compatible sticky ends, which are then re-ligated so sequences in spatial proximity become linked to each other. This process results in the formation of a 3C library that contains restriction fragments joined to a large variety of fragments, with the distribution of fragments representing the probability of spatial interactions between said fragments in a population of cells. This 3C library can be directly assayed by quantitative PCR (qPCR) by the use of primers in “viewpoint” and “target” sequences or further processed for use in 3C derivative technologies.
Figure 7. 3C and related technologies.

3C and related technologies detect DNA fragments found in spatial proximity within the nucleus. These technologies differ in genome coverage and resolution.
The first advancement of 3C technology was the development of circular chromosome conformation capture (4C), which allows for the detection of all interacting partners of a given fragment. 4C utilizes a second digestion and ligation step in order to create small circularized fragments. An inverse PCR reaction using a pair of primers in the viewpoint fragment allows for the amplification of the interacting partners of the viewpoint fragment, which can be analyzed by high throughput sequencing. Similar variants include chromosome conformation capture carbon copy (5C) which utilizes multiplexed primer pairs annealing to ligation junctions in order to only amplify junctions of interest, or capture-C which utilizes oligonucleotide capture in order to enrich for fragments of interest.

Unlike the abovementioned derivative 3C technologies that assay the interactions of a subset of fragments, Hi-C generates interaction maps between all parts of the genome. Hi-C utilizes a modified 3C library preparation protocol that involves the biotinylation of ligated junctions. These biotinylated junctions can be enriched by streptavidin pulldown following sonication, resulting in the formation of a Hi-C library highly enriched in junction sequences. A variant called chromatin interaction analysis by paired-end tag (ChIA-PET) uses a similar library preparation approach to Hi-C but combines it with chromatin immunoprecipitation (ChIP) to identify interacting partners bound by a protein of interest. Hi-C and ChIA-PET allow for the profiling of genomic interactions at an unprecedented scale, although the resolution is compromised unless the libraries are sequenced at extreme depth, requiring billions of reads per library.
Chromatin loops have shown to be important in many regulatory circuits. A notable example is the control of β-globin expression by the locus control region (LCR) of the β-globin cluster. LCR-promoter loops direct the expression of specific globin genes, and the artificial tethering of the LCR to a developmentally-silenced globin gene is sufficient to cause activation of the target gene (Deng et al., 2012; Deng et al., 2014; Palstra et al., 2003).

The structure of the Igh locus is determined by the formation of several layers of loops to mediate the ordered recruitment of regulatory elements and gene segments for recombination. One set of loops is dependent on the E_µ enhancer, which brings together regulatory elements across a 300 kb domain. The formation of this loop domain is dependent on the transcription factor Yin Yang 1 (YY1) (Guo et al., 2011a). The V_H gene segments form another set of loops, which are more dynamic in nature and also independent of the E_µ and 3′RR enhancers. These loops are dependent on YY1, paired box protein 5 (PAX5) and CTCF (Gerasimova et al., 2015; Guo et al., 2011a; Medvedovic et al., 2013; Montefiori et al., 2016). Looping at many other antigen receptor loci has been determined to be highly dependent on CTCF, which will be discussed in the next section.
1.3. Regulation of V(D)J recombination by CTCF

1.3.1. The structure of CTCF

CTCF is a ubiquitously expressed 11 zinc finger (ZF) transcription factor that has many known functions. CTCF was initially discovered in 1990 as a factor that represses the expression of chicken c-myc by binding to a region upstream of the promoter that contains multiple CCCTC core sequences (Lobanenkov et al., 1990). Since then, numerous studies have characterized the role of CTCF as a transcriptional activator, repressor, insulator and chromatin looping factor (Ong and Corces, 2014).

CBEs are large and non-palindromic, conferring directionality to CTCF binding. CBEs universally contain a 15 bp core binding motif and may also possess upstream and downstream binding motifs separated by short spacers of variable lengths (Nakahashi et al., 2013). The core motif is bound by CTCF ZFs 4-7, while the upstream and downstream motifs are bound by ZFs 9-11 and ZF3 respectively. ZF8 confers flexibility in binding, allowing CTCF to bind to CBEs with different lengths of spacers between the core and upstream motif (Hashimoto et al., 2017; Yin et al., 2017).

1.3.2. CTCF and cohesin

Cohesin is a ring-like protein complex that was originally identified to be involved in the adhesion of sister chromatids during mitosis (Funabiki et al., 1996). Cohesin has since been discovered to be a very important regulator of interphase genomic architecture (Mehta et al., 2013; Merkenschlager and Nora, 2016). Cohesin is part of the family of structural maintenance of chromosomes (SMC) proteins. This family includes
condensin, which is involved in chromosome condensation at the beginning of mitosis, and the SMC5/6 complex, which functions in replication fork resolution, homologous recombination and chromosome segregation (Gligoris and Lowe, 2016).

Cohesin is comprised of a trimer of SMC1, SMC3 and RAD21 proteins (Gligoris et al., 2014; Haering et al., 2002; Huis in ’t Veld et al., 2014). SMC1/3 have helical structures formed by the protein folding back on itself at the hinge region, with the N- and C-termini forming a globular nucleotide binding domain. These two proteins interact with each other at the hinge region. RAD21 is a winged-helix protein that completes the cohesin trimer, and interacts with the nucleotide binding domains of SMC1/3 to form a ring-like structure that can trap DNA. Loading and unloading of cohesin onto DNA is dependent on the action of accessory proteins. Cohesin loading is carried out by nipped-B-like protein (NIPBL) and MAU2 (Ciosk et al., 2000; Hu et al., 2011), while unloading is carried out by a complex of wings apart-like homolog (WAPL), stromal antigen (SA) and PDS5, which cleaves RAD21 to cause the cohesin ring to dissociate from DNA (Gandhi et al., 2006; Kueng et al., 2006; Ouyang et al., 2013).

In interphase nuclei, the binding of cohesin to chromatin often coincides with CTCF (Parelho et al., 2008; Rubio et al., 2008). Cohesin is postulated to mediate CTCF-dependent looping between CBEs, as the depletion of cohesin disrupts CTCF-mediated looping. CTCF is postulated to mediate localization of cohesin, as the depletion of CTCF reduces cohesin enrichment near CBEs but not the overall enrichment of cohesin on chromatin (Parelho et al., 2008; Splinter et al., 2006). A mechanistic explanation for how cohesin mediates CTCF-dependent looping is through loop extrusion (Barrington et al., 2014).
Loop extrusion is the spooling of chromatin fibers through cohesin rings, with the cohesin rings not only serving as a physical anchors to hold chromatin together, but also bringing together distal regions of chromatin. Biophysical evidence of loop extrusion has been shown for the related condensin protein complex, (Ganji et al., 2018). Loop extrusion stops at pairs of CBEs with convergent orientation, accounting for the overlap of CTCF and cohesin binding sites across the genome. The loop extrusion model has been tested in several different systems, including disrupting cohesin loading or release factors to affect chromosome loop sizes (Gassler et al., 2017; Haarhuis et al., 2017), and ultra-high resolution visualization of chromosome looping (Vian et al., 2018).

1.3.3. CTCF-mediated looping in antigen receptor loci

A large number of CBEs are found scattered throughout the antigen receptor loci. Most of these CBEs are found in the V gene array, although the positions of the CBEs relative to the V gene segments differ between loci. In the Tcra/Tcrd locus, most of the CBEs are found within the promoters of the V gene segments (Shih et al., 2012). In the Tcrb locus, the CBEs are distributed with no clear pattern relative to Vβ gene segments (Gopalakrishnan et al., 2013). In the Igh locus, CBEs in the RC-proximal VH region are located 3’ of the RSS, while CBEs in the RC-distal VH region are distributed in intergenic regions (Degner et al., 2011). In the Igk locus, the CBEs are distributed unevenly across the Vκ region, with CBE density correlating with Vκ repertoire usage (Ribeiro de Almeida et al., 2011).
In both Tcra/Tcrd and Igh loci, the loss of CTCF causes a reduction in looping between V gene segments and the RC (Degner et al., 2011; Shih et al., 2012). These loops are important for the diversification of antigen receptor repertoires, as the Igk repertoire becomes biased towards the usage of RC-proximal Vκ gene segments in CTCF-deleted cells (Ribeiro de Almeida et al., 2011). These experiments point to a general role of CTCF in regulating antigen receptor locus conformation and recombination. Two recent studies have shown that Igh recombination is highly dependent on CBEs near VH gene segments, where the addition or deletion of CBEs modify the Igh repertoire to be highly biased towards the most RC-proximal VH gene segment bearing a CBE (Jain et al., 2018; Qiu et al., 2018). Intergenic CBEs have also been shown to be essential antigen receptor regulatory elements, and will be discussed in the next section.

1.3.4. Intergenic CBEs diversify antigen receptor repertoires

In the Tcra/Tcrd, Igh and Igk loci, intergenic CBEs downstream of the V gene arrays have been characterized to be crucial in promoting V gene usage diversity. The most well-studied of these intergenic CBEs is intergenic control region 1 (IGCR1) found in the Igh locus. IGCR1 comprises of a pair of divergently-oriented CBEs that form a chromatin loop with distal CBEs downstream of 3’RR, constraining the activity of enhancers to DH and JH gene segments within this loop (Guo et al., 2011b). Deletion of IGCR1 causes the distal CBEs to loop to the next CBE found upstream of IGCR1, establishing a new loop domain that causes the DH-proximal VH gene segments VH7183 and VHQ52 to exhibit upregulated germline transcription and enrichment of activating
histone modifications (Qiu et al., 2018). This results in a $V_H$ repertoire heavily biased towards the activated $D_H$-proximal $V_H$ gene segments and also the appearance of aberrant $V_H$-to-$D_H$ rearrangements (Guo et al., 2011b; Jain et al., 2018; Lin et al., 2015; Qiu et al., 2018).

Intergenic CBEs (INT) -1 and -2 are found downstream of the bulk of the V gene array in the Tcra/Tcrd locus but upstream of the fetal $V_\delta$ gene segment Trdv4. INT1 and INT2 have different chromatin looping profiles, with INT1 exhibiting dynamic, low frequency interactions across the nearby 0.5 Mb of the locus, while INT2 interacts almost entirely within a smaller loop domain marked by INT2 and the T early-$\alpha$ (TEA) promoter. Deletion of INT1/2 causes a change in the loop organization of the Tcra/Tcrd locus, with a new chromatin loop formed between TEA and an upstream intergenic CBE termed INT3. INT1/2 deletion also causes repertoire changes, with a large increase in the recombination of the functional $V_\delta$ gene segment Trdv2-2, the pseudogene Trdv3, and also the appearance of aberrant $V_\delta$-to-$D_\delta$ rearrangements. Due to the nested nature of the Tcra/Tcrd locus, INT1/2 deletion also alters the Tcra repertoire, causing a decrease in distal $V_\alpha$ to proximal $J_\alpha$ recombination (Carico et al., 2017; Chen et al., 2015; Zhao et al., 2016).

The Igk locus contains two pairs of CBEs termed contracting element for recombination (Cer) and silencer in the intervening sequence (Sis). Cer and Sis are located in the intergenic space between the $V_\kappa$ and $J_\kappa$ gene segments. The deletion of Cer or Sis causes a heavy bias in $V_\kappa$ gene segment usage towards RC-proximal gene segments, and the deletion of Cer but not Sis causes a loss of locus contraction between
the 5′ and 3′ ends of the Igk locus, which are approximately 3 Mb apart in linear distance (Xiang et al., 2013; Xiang et al., 2011).

In the Tcrb locus, a set of three of intergenic CBEs (Figure 4) upstream of the RC has been characterized to play important roles in the regulation of Tcrb locus conformation and insulation (Figure 8). One CBE is located 28 kb upstream of Trbd1 and is named 5′ Prss2 CTCF site (5′PC). 5′PC functions as a tether that promotes locus contraction, with both RC-proximal and –distal Vβ gene segments interacting strongly with 5′PC when measured by 3C. The other two CBEs are divergently-oriented and found within PDβ1 (PDβ1CBE). PDβ1CBE serves as an insulator that blocks the spread of active chromatin marks like histone H3 acetylation (H3ac) and H3K4me3 from Eβ (Majumder et al., 2015b). Another study also mapped an insulator boundary to this region of chromatin, although the insulator was attributed to a trio of repetitive elements (LTR-LINE-LTR) as opposed to a CBE or a set of CBEs. This insulator was hypothesized to prevent the spread of repressive chromatin, rather than block the effects of Eβ (Carabana et al., 2011). The deletion of PDβ1 (which includes PDβ1CBE) causes 5′PC to assume insulator functions, with a spreading of active histone marks towards but not beyond 5′PC. This also causes 5′PC to lose its function as a chromatin tether (Majumder et al., 2015b).
Figure 8. Regulation of $Tcrb$ architecture by intergenic CBEs.

$Tcrb$ locus architecture in wild-type and PD$_{\beta}1$-deleted thymocytes are shown. Red arrows depict long-range interactions, and the dotted arrow indicates reduced interaction.
1.4. **Organization of the mammalian nucleus**

1.4.1. **Nuclear compartments**

The mammalian nucleus is highly organized and can be divided into many subnuclear compartments (Spector, 2001; Van Bortle and Corces, 2012). Unlike cytoplasmic organelles, these compartments are not delineated by lipid membranes but instead by the aggregation of nuclear proteins within parts of the nucleus. The largest nuclear compartment is occupied by the nucleolus, which is easily identified by light microscopy as a densely-staining organelle. The nucleolus is the site of ribosome biogenesis, where synthesis and processing of ribosomal RNA and the assembly of ribosomes occur. Other notable subnuclear compartments are transcription factories marked by the aggregation of RNA polymerase II (Pol II) in discrete foci, and the nuclear periphery that is composed of the inner nuclear membrane and nuclear lamin proteins. The nuclear periphery will be discussed in greater detail in section 1.4.3.

Chromatin within the nucleus can also be divided into two major varieties based on staining intensity by DNA-binding dyes: euchromatin and heterochromatin (Saksouk et al., 2015). Euchromatin is composed of lightly-packed chromatin and is thus less intensely stained than heterochromatin. Actively-transcribed genes are generally found in the euchromatin compartment. Heterochromatin is composed of densely-packed chromatin and can be subdivided into constitutive and facultative heterochromatin. Regions of the genome included in constitutive heterochromatin are cell type invariant and generally found at pericentromeric regions. Constitutive heterochromatin is generally composed of repetitive DNA elements and is almost devoid of genes. Facultative
heterochromatin varies between cell types and usually contains developmentally-silenced genes.

1.4.2. Topological domains

Eukaryotic interphase chromosomes are highly ordered and compartmentalized. Chromosomes occupy distinct regions of the nucleus, which are described as chromosome territories (Nagano et al., 2017; Stevens et al., 2017). Each chromosome can be further segregated into smaller units of chromatin organization known as topologically associating domains (TADs), which can range from tens of kb to several Mb in size. TADs are typically studied by the Hi-C technique first pioneered in 2009 (Lieberman-Aiden et al., 2009) and since then, a multitude of technical and technological improvements have allowed for the study of TADs at the single cell level (Nagano et al., 2013; Nagano et al., 2017; Stevens et al., 2017).

TADs are self-associating units defined by the high frequency of contact interactions between loci in the same TAD compared to loci outside of the TAD (Bonev and Cavalli, 2016; Yu and Ren, 2017). TADs appear to be a general feature of eukaryotic chromosomes, and have been observed in not only mammals, but also in the zebrafish Danio rerio (Gomez-Marin et al., 2015), in invertebrates like Drosophila melanogaster (Hou et al., 2012; Sexton et al., 2012) and Caenorhabditis elegans (Crane et al., 2015), and also in the budding yeast Saccharomyces cerevisiae (Hsieh et al., 2015).

TAD borders are frequently marked by CBEs, with TADs contained within CBEs of convergent orientation (Dixon et al., 2012; Rao et al., 2014; Tang et al., 2015). This
observation led to the hypothesis that TADs are formed by cohesin-mediated loop extrusion, described previously in section 1.3.2 (Sanborn et al., 2015; Vian et al., 2018). In fact, the disruption of CBEs at TAD borders has been shown to cause a disruption in TAD architecture (de Wit et al., 2015; Guo et al., 2015; Sanborn et al., 2015).

Principal component analysis of Hi-C contact maps reveal that TADs can be partitioned into alternating A and B compartments, which can be further divided into A1, A2, B1, B2 and B3 compartments (Lieberman-Aiden et al., 2009; Rao et al., 2014; Stevens et al., 2017). The A compartments correspond to euchromatin, with an enrichment of activating histone modifications and marked by early replication timing. The B compartments correspond to heterochromatin, with the B1 subcompartment enriched in histone H3 lysine 27 trimethylation (H3K27me3) associated with polycomb-repressed facultative heterochromatin, while the B2 and B3 subcompartments are depleted in activating histone modifications and marked by late replication timing. The partitioning of TADs into compartments is maintained even at a single cell level, showing that TADs are functional units of organization and not simply an artefact of population-level analysis (Stevens et al., 2017).

TADs exhibit varying degrees of plasticity and conservation. While the relative positioning of TAD boundaries is generally conserved across cell types and species (Dixon et al., 2012), entire TADs have been observed to flip between A and B compartments, with one study observing over a third of compartments flipping during embryonic stem cell differentiation (Dixon et al., 2015). Sub-TADs, which are smaller self-associating units within TADs that are visible in higher resolution genome maps,
display less conservation between cell types (Phillips-Cremins et al., 2013; Rao et al., 2014).

1.4.3. Lamina-associated domains at the nuclear periphery

The nuclear periphery is the region of the nucleus that is adjacent to the inner nuclear membrane. The structure of the nuclear periphery is maintained by the nuclear lamina (NL), composed of lamin B and lamin A/C intermediate filament proteins, which form distinct but interacting mesh frameworks (Shimi et al., 2015; Shimi et al., 2008; Xie et al., 2016). The NL associates with several inner nuclear membrane proteins including emerin, nesprin, lamin B receptor (LBR) and the various isoforms of lamina-associated peptide (LAP) -1 and -2 (Burke and Stewart, 2002; Gruenbaum et al., 2005; Holmer and Worman, 2001). The NL not only interacts with chromatin, but is key in signal transduction between the cytoskeleton and the nucleus. The importance of the NL is evident by the fact that mutations in NL-related genes cause a host of rare genetic disorders broadly termed laminopathies (Burke and Stewart, 2002; Dobrzynska et al., 2016; Luperchio et al., 2014).

Regions of chromatin in molecular contact with the NL are termed lamina-associated domains (LADs) (Figure 9) (van Steensel and Belmont, 2017). LADs have primarily been identified by a technique known as DamID, a profiling technology that uses bacterial DNA adenine methyltransferase (Dam) tethered to the NL to label chromatin localized to the NL (Guelen et al., 2008; van Steensel and Henikoff, 2000). LADs can also be identified by ChIP using antibodies against NL components, although
studies are fewer due to technical challenges regarding the fragmentation of heterochromatin (Becker et al., 2017; Gesson et al., 2016; Lund et al., 2015; Poleshko et al., 2017). LADs are generally associated with the heterochromatin compartment, with visible foci of heterochromatin often found near the NL (Wijchers et al., 2015), and have been shown to be tethered to the NL by LBR and lamin A/C (Solovei et al., 2013). It has been argued that all heterochromatin compartments, which include pericentromeric heterochromatin, LADs and a compartment of chromatin localized around the nucleolus termed nucleolar-associated domains, are functionally redundant (Kind et al., 2013; Ragoczy et al., 2014; Stevens et al., 2017).

**Figure 9. Organization of LADs.**

Overview of chromatin organization. A LAD is shown on the left, with genes within LADs in a repressed or inactive state. LAD chromatin is marked by H3K9me2. A non-LAD is shown on the right, with a higher density of active genes. LAD borders are marked by CBEs and active promoters oriented away from the LAD.
LADs encompass large regions of chromatin about 0.1 to 10 Mb in size, and cover approximately a third of the genome (Guelen et al., 2008). As indicated by the overlap of LADs with the heterochromatin compartment, LADs are repressive in nature, as genes located within LADs are either transcriptionally inactive or expressed at lower levels compared to non-LAD genes, although exceptions are known to exist (Park et al., 2014). LADs also exhibit a greatly reduced level of Pol II enrichment compared to non-LAD genes (Guelen et al., 2008; Peric-Hupkes et al., 2010). The borders between LADs and non-LADs have been characterized to be enriched for CTCF binding, CpG islands and active promoters transcribing away from the LAD (Guelen et al., 2008).

The transcriptionally repressive environment of LADs is also reflected in its histone modification profile, with LADs enriched in the repressive modification H3K9me2 and depleted in the activating modification H3K4me2 (Guelen et al., 2008; Kind et al., 2013; Peric-Hupkes et al., 2010; Wen et al., 2009). The overlap of LADs with H3K9me2 markings can also be observed by microscopy (Kind et al., 2013; Poleshko et al., 2017). Additionally, H3K27me3 has been found to be enriched within LADs but only in the 50-100 kb region adjacent to the border between a LAD and a non-LAD (Guelen et al., 2008; Harr et al., 2015); nevertheless, another a study found an inverse correlation between LADs and H3K27me3 (Kind et al., 2015).

LADs do not form a static subnuclear compartment, but instead exhibit varying degrees of plasticity. The LAD profile of the genome has been shown to undergo selective changes upon cell differentiation and activation, with most of the changes associated with the activation or repression of developmentally-regulated genes (Hewitt
et al., 2004; Meuleman et al., 2013; Peric-Hupkes et al., 2010; Robson et al., 2017; Williams et al., 2006) and dependent on progression through the cell cycle (Kind et al., 2013). This allowed for the classification of LADs into constitutive and facultative LADs, using a similar terminology to that used to classify heterochromatin. Constitutive LADs are typically composed of B2 and B3 TAD compartments, highly repressive, and display the highest degrees of NL contact frequency (Kind et al., 2015; Robson et al., 2017). Facultative LADs vary between cell types depending on activation status and exhibit lower levels of NL contact frequencies compared to constitutive LADs. Furthermore, genes in facultative LADs undergoing activation enter the A2 TAD compartment (Robson et al., 2017; Stevens et al., 2017).

1.4.4. Regulation of LADs

The transcriptionally suppressive environment of the nuclear periphery is mediated by the localization of transcriptional repressors and exclusion of transcriptional activators. H3K9me2 marks that are enriched in LADs are bound by heterochromatin protein 1 (HP1), which directly interacts with LBR and indirectly interacts with lamin A/C via proline rich protein 14 (PRR14) (Poleshko et al., 2013; Polioudaki et al., 2001). Another histone-modifying enzyme that is localized to the nuclear periphery is histone deacetylase 3 (HDAC3), which binds to LAP2β (Poleshko et al., 2017; Somech et al., 2005; Zullo et al., 2012). Transcriptional repressors like methyl CpG binding protein 2 (MeCP2) and retinoblastoma protein (RB) interact with LBR and lamin A/C respectively (Guarda et al., 2009; Ozaki et al., 1994). Another characterized mechanism of
transcriptional suppression is the exclusion of transcription factors and components of transcriptional machinery from the nuclear periphery (Yao et al., 2011). H3K9me2-HP1 interactions are critical in maintaining the suppressive environment of the nuclear periphery, as multiple studies have documented that perturbing H3K9me2-HP1 function by inhibiting H3K9me2 methyltransferases or by the usage of small molecule inhibitors caused LADs to dissociate from the nuclear periphery (Bian et al., 2013; Harr et al., 2015; Kind et al., 2013; Towbin et al., 2012).

Localization of genes at the nuclear periphery is sufficient to cause transcriptional suppression, as reporter genes forcibly localized to the nuclear periphery are repressed (Finlan et al., 2008; Reddy et al., 2008). In these studies, a reporter gene is randomly integrated into the genome alongside a large array of lac operator (lacO) repeats which bind the lac repressor protein (LacI). When a fusion protein of LacI with either LAP2β or emerin was expressed in these cells, the reporter gene was observed to localize to the nuclear periphery and expression of the reporter gene was inhibited (Finlan et al., 2008; Reddy et al., 2008). Conversely, forcing transcriptional activation by tethering of the herpes simplex virus VP16 transcriptional activator protein causes targeted genes to move away from the nuclear periphery (Kind et al., 2013; Therizols et al., 2014). Interestingly, the induction of chromatin decondensation in the absence of transcription is sufficient to cause movement of targeted genes away from the nuclear periphery (Therizols et al., 2014).

Studies have been undertaken to reveal the underlying principles that guide the subnuclear localization of chromatin towards, or away from, the NL. One technique used
was to randomly integrate DNA fragments into the genome in order to determine the localization of these fragments (Harr et al., 2015; Zullo et al., 2012). By characterizing the localization of progressively-smaller DNA fragments, minimal regions that could independently localize to the nuclear periphery were identified, and proteins binding to these regions were characterized by bioinformatics analysis followed by direct assays. These studies identified the transcription factors ThPOK and YY1 as capable of peripheral targeting, although the context required for peripheral targeting remains unclear as these transcription factors are also known to function as transcriptional activators in different systems (Harr et al., 2015; Zullo et al., 2012).

### 1.4.5. Tcrb recombination is suppressed at the nuclear periphery

The subnuclear positioning of the *Tcra* and *Tcrb* loci are strikingly different. Whereas *Tcra* localizes to the nuclear interior in DN and DP thymocytes but not in pro-B cells, *Tcrb* is frequently found at the nuclear periphery in all these cell types. The association with the nuclear periphery is stochastic, with each *Tcrb* allele associating independently of the other allele (Schlimgen et al., 2008). *Tcrb* recombination is dependent on subnuclear positioning, as *Tcrb* alleles positioned at the nuclear periphery undergo recombination at lower rates than centrally-positioned alleles. Of the alleles that recombined at the periphery, it was observed that these alleles were not tightly associated with the NL, but were instead partially detached. Furthermore, RAG2 is abundant in the nuclear interior but found at reduced levels at the nuclear periphery. These results imply a mechanism where NL association reduces the efficiency of *Tcrb* recombination and
diminishes the likelihood of both alleles attempting recombination in a similar time frame in order to promote allelic exclusion (Chan et al., 2013).
2. Thesis prospectus

The Tcrb locus is frequently localized to the nuclear periphery, a transcriptionally suppressive subnuclear compartment. The localization of Tcrb at the nuclear periphery was shown to suppress the recombination of Tcrb and was hypothesized to contribute to allelic exclusion (Chan et al., 2013; Schlimgen et al., 2008). However, the mechanisms that direct the high Tcrb-NL association remain unknown.

2.1. Specific aim 1: To determine how Tcrb is localized to the nuclear periphery

The goal of this aim is to characterize Tcrb-NL contacts at high resolution in order to identify elements involved in the regulation of locus subnuclear positioning. Previous studies of Tcrb subnuclear localization utilized DNA FISH, a technique with limited usefulness due to poor resolution and coverage. Therefore, we desired a technique that is capable of analyzing the association of the Tcrb locus with the NL at high resolution in order to accurately pinpoint the determinants of Tcrb subnuclear positioning.

We performed DamID to obtain a high resolution map of Tcrb-NL interactions in the DP thymocyte cell line VL3-3M2. We observed that the Tcrb locus was not homogenously associated with the NL, but had non-LAD regions that were positioned away from the NL, one of which being the RC. The NL association profile of the RC was clearly distinct from the upstream, RC-proximal trypsinogen region, which was highly NL-associated. The RC and RC-proximal trypsinogen region were clearly separated by a
LAD border that we have termed upstream of the RC (usRC). Deleting usRC caused a loss of NL association that extended across the RC-proximal trypsinogen region and affected the most RC-proximal V_\beta genes, and this phenotype was dependent on the activity of E_\beta. Dissection of usRC revealed the presence of three elements which constitute the LAD border activity of usRC. We conclude that the usRC LAD border separates the V_\beta gene segments and trypsinogen genes from the RC in order to suppress the effect of E_\beta to position chromatin away from the NL.

2.2. Specific aim 2: To characterize the regulatory impact of the usRC LAD border

In the previous aim, we identified how usRC regulates the conformation of the Tcrb locus at the NL. However, it is unknown how the unique subnuclear localization of Tcrb conferred by usRC contributes to the regulation of Tcrb.

Using the VL3-3M2 and derivative cell lines that we created in the previous aim, we characterized the effects of LAD border deletion on Tcrb locus conformation, histone modification, transcription and recombination. We found that the deletion of usRC caused long-distance interactions between the RC and RC-proximal genes to increase at the expense of long-distance interactions between the RC and RC-distal genes. The RC-proximal region also exhibited an increase in the presence of activating histone modifications, which correlated with an increase in transcription of trypsinogen genes and V_\beta gene segments contained within. Activated V_\beta gene segments exhibited a massive increase in recombination. We conclude that the disruption of NL association caused by
the deletion of usRC caused an activation of the RC-proximal trypsinogen and \( V_\beta \) genes, proving that the usRC LAD border is a core regulatory element of the \( Tcrb \) locus.
3. Materials and methods

3.1. Mice

*Ctcf*<sup>f/f</sup>*Cd2-Cre* mice were bred by crossing *Ctcf*<sup>f/f</sup>*Lck-Cre* mice (Shih et al., 2012) with *Yy1*<sup>f/f</sup>*Cd2-Cre* mice (Chen et al., 2016) and inbred to obtain the desired genotype. The *Ctcf*<sup>f/f</sup>*Cd2-Cre* mice were also bred onto a *Rag2*<sup>/−</sup> background. ChIP-seq studies were performed on *Rag2*<sup>/−</sup> mice. All mice were of strain 129 background. Mice were housed in a specific-pathogen-free facility managed by the Duke University Division of Laboratory Animal Resources. Mice of both sexes were included in all experiments; no differences on the basis of sex were noted. Mice were generally sacrificed at 4 weeks of age. For studies on a *Rag2*<sup>/−</sup> background, DN thymocytes were obtained directly while DP thymocytes were obtained by performing an intraperitoneal injection with 150 μg of anti-CD3ε antibody (145-2C11, Biolegend) 10 days prior to sacrifice. All mice were handled in accordance with protocols approved by the Duke University Institutional Animal Care and Use Committee.

3.2. Cell culture

VL3-3M2, P5424, A70.2 and M12 cells were maintained in RPMI 1640 (Gibco) with L-glutamine (Gibco) supplemented with 10% fetal bovine serum (FBS) (Gemini Bio-Products), 50 U/mL penicillin-streptomycin (Gibco) and 55 μM 2-mercaptoethanol (Gibco). BOSC23 cells were maintained in DMEM (Gibco) supplemented with 10% FBS
(Gemini Bio-Products) and 50 U/mL penicillin-streptomycin. Cells were cultured at 37°C in 5% CO₂ in a humidified incubator.

3.3. Preparation of DNA-FISH probes

Bacterial artificial chromosomes (BACs) RP23-457D7 and RP23-75P5 (BACPAC Resources Center, Children’s Hospital Oakland Research Institute) were used to generate DNA-FISH probes. 1 µg of BAC was labeled with either biotin or digoxigenin in a volume of 20 µL for 2 h at 16°C using Biotin- or DIG-Nick Translation Mix (Roche). 30 µg of mouse Cot-1 DNA (Invitrogen) and 300 µg of salmon sperm DNA (Invitrogen) were mixed with the labeled probe, and the mixture ethanol precipitated and resuspended in 50 µL of Hybrisol VII (MP Biomedicals). The probe mixture was stored at -20°C. Before hybridization, the DNA probe cocktail was made by mixing 5 µL each of labeled RP23-457D7 and RP23-75P5 probes together, and denatured by heating to 100°C for 5 min, placing on ice for 5 min and incubating at 37°C for 1 h.

3.4. DNA-FISH

5×10⁵ cells in a volume of 100 µL were deposited onto poly-L-lysine coated microscopy slides and left to adhere for 20 min. Cells were fixed in Hank’s balanced salt solution (HBSS) supplemented with 4% v/v paraformaldehyde (Electron Microscopy Sciences) for 10 min at 20°C. Fixed cells were permeabilized in HBSS supplemented with 0.5% w/v saponin (Sigma-Aldrich) and 0.5% v/v Triton X-100 for 1 h at 20°C, followed by incubation in 0.1 M HCl for 10 min at 20°C. Slides were kept for short term
storage in HBSS supplemented with 20% w/v glycerol at 4°C. Prior to hybridization, cells were subjected to four freeze/thaw cycles in liquid N₂, followed by treatment with 50 μg/mL RNAse A (Sigma-Aldrich) for 20 min at 37°C. For probe hybridization, cells were first denatured in 70% v/v formamide (Amresco) with 2X saline-sodium citrate (SSC) buffer (300 mM NaCl, 30 mM sodium citrate, pH 7.0) for 3 min at 77.8°C followed by a second denaturation in 50% v/v formamide with 2X SSC for 1 min at 77.8°C. 200 ng of the denatured DNA probe cocktail was added onto the cells and the slide was sealed by placing a cover slip on top of the cells and sealing the edges with rubber cement (Electron Microscopy Sciences). Hybridization was left to proceed overnight at 37°C. After removal of the cover slip, excess probes were removed by three washes in 50% v/v formamide with 2X SSC for 5 min at 42°C, followed by three washes in 0.2X SSC for 5 min at 60°C. Blocking was performed by incubation of the cells in 4% v/v bovine serum albumin (BSA) in 2X SSC for 30 min at 20°C. The NL was stained using a polyclonal goat anti-lamin B1 antibody (sc-6217, Santa Cruz Biotechnology) dissolved in 4% BSA in 2X SSC for 1 h at 20°C. Secondary antibody staining was performed with FITC-conjugated anti-goat antibody, Cy3-conjugated anti-biotin antibody and Cy5-conjugated anti-digoxigenin antibody (Jackson ImmunoResearch) dissolved in 4% BSA in 2X SSC for 1 h at 20°C. After both antibody staining steps, excess antibodies were removed by three washes in 2X SSC supplemented with 0.1% v/v Triton X-100. Slides were mounted in VectaShield (Vector Laboratories).
3.5. Microscopy imaging and analysis

Slides were imaged on a Leica SP5 confocal microscope using a 100X objective lens and 2X digital zoom. Each voxel measured 0.15 µm in the x and y dimensions and 0.13 µm in the z dimension. Image analysis was performed using Fiji (Schindelin et al., 2012). In brief, images were imported into Fiji and a manual threshold set to distinguish between signals from background. The thresholds were set to ensure staining intensity between images were similar. BAC probe foci were considered to colocalize with lamin B1 if at least two adjacent voxels on an xy-plane overlapped, with the overlap detected across at least two adjacent z planes.

3.6. Flow cytometry and cell sorting

The following monoclonal antibodies and reagents from Biolegend were used for staining: B220 (RA3-6B2), CD4 (GK1.5), CD8α (53-6.7), CD11b (M1/70), CD11c (N418), CD25 (3C7), CD44 (IM7), Gr-1 (RB6-8C5), TCRβ (H57-597), Ter-119 (TER-119) and 7-amino-actinomycin D (7AAD). Ctsf<sup>Cre</sup>/Cd2-Cre thymocytes were stained with fluorescein di-V-galactoside (FDG) using the FluoReporter lacZ Flow Cytometry Kit (Life Technologies). Tcrb repertoire analysis was performed by staining with the following FITC-conjugated monoclonal antibodies from the Mouse V<sub>β</sub> TCR screening panel (Pharmingen): V<sub>β</sub>3 (JK25), V<sub>β</sub>4 (KT4), V<sub>β</sub>7 (TR310), V<sub>β</sub>8.1/8.2 (MR5-2) and V<sub>β</sub>14 (14-2). Flow cytometric data acquisition was performed on a FACSCanto II (Becton Dickinson). Cell sorting was performed on an Astrios (Beckman Coulter), MoFlo XDP (Beckman Coulter) or FACSDiVa (Becton Dickinson).
3.7. Western blot

$10^6$ FDG$^+$ DN3 thymocytes from $Ctcf^{ff}Cd2$-Cre mice were lysed in 50 µL of radioimmunoprecipitation assay buffer (25 mM Tris-HCl pH 8.0, 150 mM NaCl, 1% v/v Triton X-100, 0.5% w/v sodium deoxycholate, 0.1% w/v SDS, protease inhibitor cocktail (Sigma-Aldrich)) for 10 min at 4°C. After centrifugation at 18000 g for 10 min at 4°C, the supernatant was transferred to a new tube and 50 µL of 0.125 M Tris-HCl pH 6.8, 20% v/v glycerol, 10% v/v 2-mercaptoethanol (Sigma-Aldrich), 4% w/v SDS and 0.004% w/v bromophenol blue (Sigma-Aldrich) was added. The sample was boiled at 95°C for 5 min and subsequently centrifuged. 10 µL of sample was loaded into an SDS-PAGE gel for gel electrophoresis in a running buffer containing 25 mM Tris-HCl pH 6.8, 190 mM glycine, 0.1 w/v SDS. The proteins were transferred onto a nitrocellulose membrane in buffer containing 25 mM Tris-HCl pH 6.8, 190 mM glycine, 20% v/v methanol using a Trans-Blot Semi-Dry Transfer Cell (Bio-Rad) running at 20 V for 1 h. Blocking was performed by rocking the membrane in 2% w/v fish gelatin dissolved in TBST buffer (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.1% v/v Tween 20) for 1 h at 20°C. Primary antibody staining was performed by rocking the membrane with CTCF (Millipore) or LAT (Cell Signaling) antibody dissolved in TBST buffer overnight at 4°C. After washing in TBST buffer, secondary antibody staining was performed by rocking the membrane with goat anti-rabbit antibody conjugated with Alexa Fluor 680 (Invitrogen). After washing with TBST buffer, the membrane was imaged on a LI-COR Odyssey (LI-COR).
3.8. CRISPR/Cas9 plasmid construction and transfection

CRISPR/Cas9 guides were designed using the Zhang lab CRISPR design tool available at http://crispr.mit.edu/ (Ran et al., 2013). Guides were cloned into pX458 according to the Zhang lab general cloning protocol. For transient transfection into VL3-3M2 cells, 2 μg of plasmid was mixed with 10⁶ cells in 100 μL of Cell Line Nucleofector Kit V solution (Amaxa) and electroporation was performed using program X-001. Transfected cells were allowed to expand for three days before single cell sorting into 96-well plates. Clones were expanded for nine days before genomic DNA (gDNA) was harvested for screening by PCR. Initial screening was performed on column and row pools, after which prospective clones were expanded for a second round of PCR screening followed by Sanger sequencing of the PCR product. Cell lines bearing compound deletions were generated by two rounds of this process. ΔusRC+ΔEβ was created using ΔEβ as a base, Δ5'PC+ΔPDβ1CBE was created using ΔPDβ1CBE as a base, and Δ5'PC+ΔPDβ1 was created using ΔPDβ1 as a base. Mutated sequences (with indicated guides) are listed in Table 1.
For each line, mutant alleles are shown on top with deleted nucleotides indicated (---) and inserted nucleotides indicated in lowercase. Unmutated sequences are shown below with target sequences in bold and protospacer adjacent motifs shown in italics.

### 3.9. Retroviral packaging and transduction

The BOSC23 cell line (Pear et al., 1993), a derivative of the HEK293T cell line, was grown to 80-90% confluence in 10 cm tissue culture-treated dishes and transfected with 10 μg of retroviral vector and 5 μg of packaging plasmid pkat2ampac (Finer et al., 1993).
1994) using calcium phosphate. Media was replaced one day later and viral supernatant harvested the following day. VL3-3M2 and P5424 cells were transduced by spin-infection. In brief, $10^6$ VL3-3M2 cells were resuspended in 1 mL of viral supernatant and 10 μg of hexadimethrine bromide (Sigma-Aldrich) and placed into a 48-well tissue culture-treated plate. Spin-infection was performed by centrifugation at 1500 g for 90 min at 32°C. Following spin-infection, the viral supernatant was removed and replaced with fresh culture medium.

3.10. RNA isolation and analysis

RNA was isolated using Trizol (Ambion). Following purification, RNA was treated with DNase I (New England Biolabs) and cDNA was synthesized using the iScript cDNA Synthesis Kit (Bio-Rad). qPCR was performed using the QuantiFast SYBR Green PCR kit (Qiagen) on a Roche Lightcycler 480 using the following program: 5 min at 95°C, followed by 45 cycles of 10 s at 95°C and 30 s at 62°C. qPCR primers detecting $V_\beta$ gene segment transcription were designed to span the RSS to only detect transcripts originating from unrearranged gene segments. Experimental values were normalized to Actb in each sample. qPCR primers are listed in Table 2.
Table 2. List of transcription primers.

<table>
<thead>
<tr>
<th>Target</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trbv3</td>
<td>CAGCAGCTGGAGGACTCAGC</td>
<td>CTCTTGGGAAGACAGGCTT</td>
</tr>
<tr>
<td>Trbv5</td>
<td>GCCGTGGATCCAGAAGACTC</td>
<td>TCTTCAATGAGCCCTCAGTT</td>
</tr>
<tr>
<td>Trbv13-2</td>
<td>TGGCAGCCACTGAAAGGGAG</td>
<td>CCGTGGGGAAGACAGGGAG</td>
</tr>
<tr>
<td>Trbv26</td>
<td>TCACCTTGCAGGCTTGAAGT</td>
<td>TGGGTTATGCAGAGGGAG</td>
</tr>
<tr>
<td>Trbv29</td>
<td>GTCTCAGGAAAGGAGCCAGGGAAGGAG</td>
<td>ACATGAAAGCCAGCATTAGGC</td>
</tr>
<tr>
<td>Trbv30</td>
<td>GGGCCAAACCTAACATTCT</td>
<td>GAAGGAGATGAGGTTCATCGG</td>
</tr>
<tr>
<td>Trv5</td>
<td>AGGACCCCTGAACACGCAGCAT</td>
<td>GGGGCATCCAGGCAATTGGA</td>
</tr>
<tr>
<td>Prss1</td>
<td>TCATCCCTGTGACCTCTCAAA</td>
<td>TCCAGGATAGGAGACCCTCA</td>
</tr>
<tr>
<td>Actb</td>
<td>GGCTGTATTTCCCTTCATCG</td>
<td>CCAGTTGTAACAAATGCCATGT</td>
</tr>
<tr>
<td>DJβ</td>
<td>GGGTAGTCAAAACACCTTGACT</td>
<td>CACGTGGTCAAGGAGAAGC</td>
</tr>
</tbody>
</table>

3.11. DamID

VL3-3M2 cells were transduced with pSMGV Dam-V5 (unfused Dam control) or pSMGV Dam-V5-LMNB1 (Dam-lamin B1) and grown for three days before sorting and expansion of GFP+ cells. gDNA was harvested and purified by phenol-chloroform extraction and isopropanol precipitation. Subsequent steps were based on the protocol described in (Vogel et al., 2007). 2.5 µg of gDNA was digested with 10 U of DpnI (New England Biolabs) in 10 µL overnight at 37°C, followed by heat inactivation at 80°C for 20 min. The DpnI-digested product was ligated to 40 pmol of double-stranded adapter dsAdR, consisting of the annealed oligonucleotides AdRt: CTAAATAGCAGTGGCGGCCAGAAGA and AdRb: TCCTCGGCGG, with 400 U of T4 DNA ligase (New England Biolabs) in 20 µL for 2 h at 16°C, followed by heat inactivation at 65°C for 10 min. Unmethylated fragments were digested with 10 U of DpnII (New England Biolabs) in 50 µL for 1 h at 37°C, followed by heat inactivation at 65°C for 20 min. 20% of the sample was utilized for PCR amplification using Advantage 2 polymerase mix (Takara) with 62.5 pmol of AdR-PCR primer: GGTCGCGGGCCGAG...
GATC using the following program: 68°C for 10 min, 94°C for 1 min, 65°C for 5 min, 68°C for 15 min, 3 cycles of 94°C for 1 min, 65°C for 1 min, 68°C for 10 min, and 18 cycles of 94°C for 1 min, 65°C for 1 min, 68°C for 2 min. The PCR product was purified using the QIAquick PCR Purification Kit (Qiagen). For DamID-qPCR, 10 ng of the DamID PCR product was used with the Quantifast SYBR Green PCR kit (Qiagen) on a Roche Lightcycler 480 using the following program: 5 min at 95°C, followed by 45 cycles of 10 s at 95°C and 30 s at 62°C. DamID values were calculated by taking the log₂ ratio of Dam-lamin B1 signals over unfused Dam. qPCR primers are listed in Table 3.

**Table 3. List of DamID-qPCR primers.**

<table>
<thead>
<tr>
<th>Target</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trbv1</td>
<td>CCCAGTCATCGCAGTCAGTT</td>
<td>TTGGGCTGAGGGTCCCAAAG</td>
</tr>
<tr>
<td>T₃₁</td>
<td>GTCAAGGAGAGGGACTTCAGC</td>
<td>CCCAAGTGGACTGGTGGGAC</td>
</tr>
<tr>
<td>T₃₂</td>
<td>AGTGCATCTGGAACCTTAACCAT</td>
<td>GATGGGCTGAGGGTCCCAAAG</td>
</tr>
<tr>
<td>T₃₃</td>
<td>CGTCTTTGTGTCTCAGATGGG</td>
<td>CTGGGCTGAGGGTCCCAAAG</td>
</tr>
<tr>
<td>Trbv2</td>
<td>GTTCAGTGCGAGGCAAAAT</td>
<td>CAAAAGGCAAGGTCCCAAG</td>
</tr>
<tr>
<td>Trbv5</td>
<td>TTATCCCTCCAGGGCAGCT</td>
<td>ATCCAGGCTGAGGGGTAT</td>
</tr>
<tr>
<td>Trbv9</td>
<td>CTACCAGGCAAGGACGAGCA</td>
<td>TGCCAGGACACTGGTGGC</td>
</tr>
<tr>
<td>Trbv13-3</td>
<td>CTCCTGAGCCGACCTTCCTTGGG</td>
<td>CAGAGGCAAGGACGAGCA</td>
</tr>
<tr>
<td>Trbv15</td>
<td>ATCCCTCTGCTCCCTGACTCG</td>
<td>ACAAATTGGCAACACTCACCA</td>
</tr>
<tr>
<td>Trbv17</td>
<td>ACACCTGCTGAGGCAAGTGAGGAA</td>
<td>ATAGAGGCTGACCTCCAGCAC</td>
</tr>
<tr>
<td>Trbv20</td>
<td>CTACCCAGGGAAAAGTTGAGGA</td>
<td>ATAGACAGGCTGACCTCCAGCAC</td>
</tr>
<tr>
<td>Trbv23</td>
<td>TATCCTCAGGAAAGGGGCACAC</td>
<td>TGACGCTCAGGCTGACCTCCAG</td>
</tr>
<tr>
<td>Trbv26</td>
<td>TAACTCTCCAGGACTGGGC</td>
<td>TACAGGCAAGGGGAAATGGGGC</td>
</tr>
<tr>
<td>Trbv30</td>
<td>AATCACCGGCGCAACTTAC</td>
<td>AGGTGCTGAGGGAAATGGGGC</td>
</tr>
<tr>
<td>Try5</td>
<td>AGGTGTAAGGAGAGGGAAGT</td>
<td>GAGATGGGAGAGTGGGAT</td>
</tr>
<tr>
<td>T₄₁</td>
<td>AAGTCATATTCTCTCCAGGGAAGT</td>
<td>ATGGGAGGATTTTACAGCT</td>
</tr>
<tr>
<td>T₄₂</td>
<td>GCTGAAATTGGAGGACTGAAAGG</td>
<td>GCCATGACTTTTGAAGGCCC</td>
</tr>
<tr>
<td>T₄₃</td>
<td>CAATAACCAGAGCTGCTCCA</td>
<td>TAAACTCTGAGGCCCCACAG</td>
</tr>
<tr>
<td>T₄₄</td>
<td>ATCCGGTGGACCCCATACGCAACGGC</td>
<td>TCCCTGGTCCCTTTGGAT</td>
</tr>
<tr>
<td>Trbj2-7</td>
<td>GAGGATATGCAATGACTGGGT</td>
<td>GACAGCAGCTCCTCTGGAAT</td>
</tr>
<tr>
<td>E₆</td>
<td>CCAATGGGCGGCTCTATGGA</td>
<td>CAAGGGGAGGCCCAGCTTT</td>
</tr>
<tr>
<td>Chr10Dip</td>
<td>CTCAATGCTAGATTGAGGAACC</td>
<td>CCAATCTCCCTGAGCTTGCG</td>
</tr>
<tr>
<td>Actb</td>
<td>TCCCTGCTCAGGAGGTGTCG</td>
<td>CGGGTCGCTGAGGCTGTT</td>
</tr>
<tr>
<td>Igh J558</td>
<td>TTCTGGTACCTCCTGCAGCTC</td>
<td>ACTGCATCTGACTGAGGGT</td>
</tr>
<tr>
<td>Or1a1</td>
<td>TCACATCTTGTGACATGGGA</td>
<td>TTTCAAAACACAATTTCTGCC</td>
</tr>
</tbody>
</table>
For DamID-seq, the DamID PCR product was randomized by sonication following end repair and ligation. Specifically, 1-2 μg DamID material was subjected to end repair using the NEBNext End Repair Module (New England Biolabs) and purified using DNA Clean and Concentrator -5 (Zymo Research). Purified DNA was ligated using 5U T4 DNA ligase (Roche) per 1 μg material in 20 μL to generate a library of large DNA fragments. This DNA library was diluted to 200 μL in H2O, sonicated for 1 h in 1.5 mL DNA LoBind tubes (Eppendorf) using a Bioruptor (Diagenode), then transferred to 1.5 mL TPX tubes (Diagenode) for four rounds of 10 min sonication to yield fragments of 150-300 bp in size. Samples were transferred to new TPX tubes between each round of sonication to minimize loss of DNA fragments. The sonicated DNA was transferred to 1.5mL DNA LoBind tubes, ethanol precipitated and resuspended in 10 μL of 10 mM Tris-HCl pH 8.0. Samples were quantified using Qubit and sequencing libraries were prepared using the NEBNext Ultra DNA Library Prep Kit for Illumina (New England Biolabs). In independent experiments, we obtained 100 bp single end reads using an Illumina HiSeq 2500 or 150 bp paired end reads (with one end discarded) using an Illumina HiSeq 4000.

3.12. DamID-seq analysis

DamID-seq reads were processed using LADetector (available at https://github.com/thereddylab/pyLAD), an updated Python implementation of LADetector described in (Harr et al., 2015) with incorporated sequence mapping. pyLAD was run with the parameters “--multimapping --quality-trim --unalignable --count-
overlaps --seed 5”. Reads were quality trimmed using a sliding window quality score average over 3 bases and a minimum score cutoff of 30. This was followed by trimming any matching overlap between read-ends and sequencing or DamID adapter-primer sequence. Reads containing a DamID adapter-primer sequence were split and adapter-primer sequence removed. All resulting reads greater than 20 bp were aligned to mm9 using Bowtie (Langmead et al., 2009) with parameters “--tryhard --best --m 1”. Unaligned reads had 10 bases trimmed from the 5’ end and were remapped, and the resultant unmapped reads were trimmed 10 bases from the 3’ end and remapped. Total aligned reads were assigned to DpnI bins, with reads straddling bin boundaries counting towards both. Prior to scoring, a value of 0.5 was added to bins with no reads. Bins falling in unaligned regions were removed prior to analysis. DamID scores were calculated for all non-zero bins as the log$_2$ ratio of Dam-lamin B1 over unfused Dam. Scores were partitioned using circular binary segmentation using the DNAcopy package in R (Seshan and Olshen, 2018). LADs were classified as regions >100 kb in size of positive signal, allowing for smaller regions of negative signal <10 kb in size. The annotated R code is listed in Appendix A.

### 3.13. Ctcf deletion efficiency

gDNA from FDG$^+$ DN3 thymocytes of Ctcf$^{+/\text{Cd2-Cre}}$ mice was purified by phenol-chloroform extraction and isopropanol precipitation. qPCR analysis was performed using the QuantiFast SYBR Green PCR kit (Qiagen) on a Roche Lightcycler 480 using the following program: 5 min at 95°C, followed by 45 cycles of 10 s at 95°C
and 30 s at 62°C. Experimental values were normalized to Cd14. qPCR primers are listed in Table 4.

Table 4. List of Ctcf deletion primers

<table>
<thead>
<tr>
<th>Target</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ctcf</td>
<td>CTAGGAGTGTAGTTTCAGTGAGGCC</td>
<td>GCTCTAAAGAAGGTTGTGAGTTTC</td>
</tr>
<tr>
<td>Cd14</td>
<td>GCTCAAACTTTTCAGAATCTACCACGAC</td>
<td>AGTCAGTTCTGGAGCCGGAAATC</td>
</tr>
</tbody>
</table>

3.14. CTCF ChIP

10^6 FDG^+ DN3 thymocytes from Rag2^−/− Ctcf^−/− Cd2-Cre mice were crosslinked in 1 ml of RPMI 1640 (Gibco) containing 10% v/v FBS (Gemini Bio-Products) and 1% w/v paraformaldehyde (Electron Microscopy Sciences) for 10 min at 20°C. Crosslinking was terminated by the addition of 100 μL 1.25M glycine (Sigma-Aldrich). Fixed cells were pelleted by centrifugation and cytoplasm removed by incubation in 1 mL 5 mM PIPES pH 8.0, 85 mM KCl, 0.5% v/v NP-40 for 10 min on ice. Nuclei were pelleted by centrifugation and lysed in 100 μL 50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS and volume increased to 1 mL by addition of 900 μL 10 mM Tris-HCl pH 8.0, 10 mM EDTA. Chromatin was sheared with a Sonicator 3000 (Qsonica) at a power level of 2 for 10 cycles of 15 s on and 30 s off. After centrifugation at 18000 g for 10 min at 4°C, the supernatant was diluted to 1.5 mL with a final concentration of 16.7 mM Tris-HCl pH 8.0, 167 mM NaCl, 1.2 mM EDTA, 1.1% v/v Triton X-100, 0.03% w/v SDS. Pre-clearing was performed by addition of 100 μl protein A agarose/salmon sperm DNA slurry (Millipore) and mixing for 3 h at 4°C. 500 μL of pre-cleared supernatant was incubated overnight at 4°C with 5 μg of antibody specific for CTCF (Millipore) or control IgG (R&D Systems). Pulldown was performed by adding 75 μl protein A
agarose/salmon sperm DNA slurry (Millipore) and mixing for 45 min at 4°C. The slurry was washed twice with 1 mL of each of the following wash buffers: 167 mM NaCl, 16.7 mM Tris-HCl pH 8.0, 1.2 mM EDTA, 1.1% v/v Triton X-100, 0.01% w/v SDS; 500 mM NaCl, 20 mM Tris-HCl pH 8.0, 2 mM EDTA, 1% v/v Triton X-100, 0.1% w/v SDS; 100 mM Tris-HCl pH 8.0, 0.5 M LiCl, 1% v/v NP-40, 1% w/v sodium deoxycholate; 10 mM Tris-HCl pH 8.0, 1 mM EDTA. Elution was performed by resuspending the slurry in 250 µl of 50 mM NaHCO₃, 1% w/v SDS and rocked for 15 min at 20°C. The elution is performed twice and eluates combined. Crosslinks were reversed by addition of NaCl to a final concentration of 200 mM and overnight incubation at 65°C. The eluate was treated with RNAse A (Sigma-Aldrich) at a concentration of 400 µg/mL for 30 min at 37°C followed by proteinase K (VWR) at a concentration of 40 µg/mL for 1 h at 65°C. DNA was purified by phenol-chloroform extraction and isopropanol precipitation and resuspended in 10 mM Tris-HCl pH 8.0, 0.1 mM EDTA. In addition to the listed components, all buffers used before elution also contained 100 µM PMSF (Sigma-Aldrich) and 100 µM benzamidine (Sigma-Aldrich). qPCR analysis was performed using the QuantiFast SYBR Green PCR kit (Qiagen) on a Roche Lightcycler 480 using the following program: 5 min at 95°C, followed by 45 cycles of 10 s at 95°C and 30 s at 62°C. Experimental values were normalized to input. qPCR primers are listed in Table 5.

Table 5. List of CTCF ChIP primers

<table>
<thead>
<tr>
<th>Target</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Trbv1</em></td>
<td>CGCTGGAGGAAATGAGGGA</td>
<td>CAGGGTGAGCCAGGAAAG</td>
</tr>
<tr>
<td>5′PC</td>
<td>TAAAACTGCTATGCTTTCGC</td>
<td>AGGGTAGGAAGTTACACAGT</td>
</tr>
<tr>
<td>E₅</td>
<td>AGGAAAGTGGCGAGAACCTGAA</td>
<td>GAGGGAGAAAGCCCTTGT</td>
</tr>
<tr>
<td><em>Dad1</em></td>
<td>TACACTTACCTGGGCCTTGG</td>
<td>ATGCACTTTCCTATGCTTG</td>
</tr>
</tbody>
</table>
3.15. Histone ChIP

$10^7$ VL3-3M2 cells or primary thymocytes were washed in phosphate buffered saline supplemented with 5 mM EDTA, 500 µM spermidine and 150 µM spermine, followed by the same solution lacking EDTA. Cells were lysed by incubation for 5 min on ice in 250 mM sucrose, 80 mM NaCl, 10 mM Tris-HCl pH 8.0, 6 mM MgCl$_2$, 1 mM CaCl$_2$, 500 µM spermidine, 150 µM spermine, containing either 0.2% v/v NP-40 (for VL3-3M2 cells) or 0.02% v/v NP-40 (for primary thymocytes). Samples were centrifuged at 750 g for 5 min at 4°C and the pellet was washed in 200 µl of digestion buffer (250 mM sucrose, 10 mM NaCl, 10 mM Tris-HCl pH 8.0, 3 mM MgCl$_2$, 1 mM CaCl$_2$). The pellet was then resuspended in 200 µl of digestion buffer and 200 gel units of micrococcal nuclease (New England Biolabs) was added for a 10 min incubation at 37°C. The reaction was terminated by addition of 300 µl 10 mM Tris-HCl pH 8.0, 5 mM EDTA. After centrifugation at 18000 g for 10 min at 4°C, supernatants were harvested and Triton X-100 added to a final concentration of 1% v/v. After checking digestion efficiency on a gel, chromatin was pre-cleared by addition of 50 µl protein A agarose/salmon sperm DNA slurry (Millipore) and mixing for 1 h at 4°C. Supernatant corresponding to 5 x $10^6$ cells was incubated overnight at 4°C with 5 µg of antibody specific for H3K9me2 (Abcam), H3K27ac (Abcam), H3K27me3 (Millipore) or control IgG (R&D Systems). Pulldown was performed by adding 50 µl protein A agarose/salmon sperm DNA slurry (Millipore) and mixing for 1 h at 4°C. The slurry was washed twice with 1 mL of each of the following wash buffers: 167 mM NaCl, 16.7 mM Tris-HCl pH 8.0, 1.2 mM EDTA, 1.1% v/v Triton X-100, 0.01% w/v SDS; 300 mM NaCl, 20 mM
Tris-HCl pH 8.0, 2 mM EDTA, 1% v/v Triton X-100, 0.1% w/v SDS; 50 mM Tris-HCl pH 8.0, 0.25 M LiCl, 0.5% v/v NP-40, 0.5% w/v sodium deoxycholate; 10 mM Tris-HCl pH 8.0, 1 mM EDTA. Elution was performed by resuspending the slurry in 100 µl of 10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.3% w/v SDS and 1 mg/mL proteinase K (VWR) overnight at 65°C. Elution buffer was set aside and a second elution performed using 0.5 M NaCl, 10 mM Tris-HCl pH 8.0, 1 mM EDTA. Eluates were combined and volumes were adjusted to 500 µl with 10 mM Tris-HCl pH 8.0, 1 mM EDTA. DNA was purified by phenol-chloroform extraction and isopropanol precipitation and resuspended in H$_2$O.

In addition to the listed components, all buffers used before elution also contained 10 mM sodium butyrate (Sigma-Aldrich), 100 µM PMSF (Sigma-Aldrich) and 100 µM benzamidine (Sigma-Aldrich). qPCR analysis of H3K27me3 was performed using the QuantiFast SYBR Green PCR kit (Qiagen) on a Roche Lightcycler 480 using the following program: 5 min at 95°C, followed by 45 cycles of 10 s at 95°C and 30 s at 62°C. Experimental values were normalized to input. qPCR primers are listed in Table 6.

**Table 6. List of H3K27me3 ChIP primers.**

<table>
<thead>
<tr>
<th>Target</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Trbv19</em></td>
<td>ACGGAGATGGCCGTTTTTTCTC</td>
<td>ATGCATGGAGAGGGGTAGC</td>
</tr>
<tr>
<td><em>Trbv26</em></td>
<td>CTTTCAACCTCAGCTTGCAG</td>
<td>CTGCTAGCACAGAAGATACAG</td>
</tr>
<tr>
<td><em>Trbv29</em></td>
<td>ATACAGGTCTTCACGGAAGAA</td>
<td>CTGCTAGCACAGAAGATACACA</td>
</tr>
<tr>
<td><em>Trbv30</em></td>
<td>AATCAGCGGCAAGAACCTA</td>
<td>CTGCTAGCACAGAAGATACACTGTC</td>
</tr>
<tr>
<td><em>Try5</em></td>
<td>GTTTGGCGCAATGGGTGTC</td>
<td>ACAGCTCCAGGGGCAATGGT</td>
</tr>
<tr>
<td><em>Pne1</em></td>
<td>CCATCCAGGGCTCAGTACCT</td>
<td>AGCCCTGGTGGGAAGTTCATGG</td>
</tr>
<tr>
<td><em>Sph</em></td>
<td>TGCCCTGTAGTTGTTTTGCC</td>
<td>CCACGTGCAATGATGTGA</td>
</tr>
<tr>
<td><em>Ea</em></td>
<td>GGAAGGGTGGAAAGCTATTCTC</td>
<td>GCCAGGGCCATTTCTTACT</td>
</tr>
<tr>
<td><em>Actb</em></td>
<td>TCTCCTATGCTGCTCCAGTTG</td>
<td>CGGCTGCTAAAGGAGGCTT</td>
</tr>
<tr>
<td><em>Hprt</em></td>
<td>GCCCTGACTTGTGCCCCTAAT</td>
<td>CGGCTAGTTTTCTGAGGTC</td>
</tr>
<tr>
<td><em>Ins1</em></td>
<td>AAGCCTGTTGATAAAAACCCAAAGGATA</td>
<td>CTTGCTAGTGCCCTTACAGG</td>
</tr>
<tr>
<td><em>Olr446</em></td>
<td>CCACGTCGTTTCCTCCACCTTGT</td>
<td>CTGTTCTTGGGAATGCTTTGATTT</td>
</tr>
</tbody>
</table>

62
3.16. ChIP-seq

Library preparation and high throughput sequencing were performed by the Duke Center for Genomic and Computational Biology Core Facility. 50 bp single end reads were obtained using an Illumina HiSeq 4000. Analysis was primarily performed using the Bioconductor set of packages in R (Huber et al., 2015; R Core Team, 2018). The annotated code is listed in Appendix A. In brief, demultiplexed .fastq files were trimmed using the fastqFilter command in the dada2 package (Callahan et al., 2016) with the parameters “truncLen = 0, trimLeft = 0, maxN = 0, minQ = 0, rm.phix = TRUE”. The trimmed files were aligned to the mm9 genome using the qAlign command in the QuasR package (Gaidatzis et al., 2015; Langmead et al., 2009), a wrapper for Rbowtie with the parameters “aligner = “Rbowtie”, maxHits = 1, paired = NULL”. .bam files were merged using the mergeBam command in the Rsamtools package (Morgan et al., 2018). To generate ChIP-seq tracks, reads were placed into 200 bp bins using the constructBins command in the mosaics package (Chung et al., 2018) with the parameters “fragLen = 150, binSize = 200, capping = 50, PET = FALSE”. Reads that corresponded to chromosome Y, chromosome M or unmapped parts of the chromosome were excluded. Bins were assembled into tracks in the .bedgraph format using a custom script. For H3K27ac tracks, bins were normalized to reads per million. H3K9me2 tracks were normalized for read count relative to the input sample and a value of 1 was added to bins with no reads. Log$_2$ ratios of the H3K9me2 sample over input were then plotted. Correlation graphs between DamID-seq and H3K9me2 ChIP-seq were constructed by placing reads into 10 kb bins using the constructBins command as described above,
except changing “binSize = 10000”. Matching bins were plotted and the Pearson’s correlation coefficient obtained using the cor function in R. To align the features surrounding LAD borders, we first defined the regions extending 50 kb upstream and downstream of the 5’ LAD borders. These regions were then aligned using the plotMeta command in the genomation package (Akalin et al., 2015).

3.17. 3C

5 x 10⁶ VL3-3M2 cells were crosslinked in 8 ml of RPMI 1640 (Gibco) containing 10% v/v FBS (Gemini Bio-Products) and 2% w/v paraformaldehyde (Electron Microscopy Sciences) for 10 min at 20°C. Crosslinking was terminated by the addition of 1 mL 1.25M glycine (Sigma-Aldrich). Fixed cells were pelleted by centrifugation and cytoplasm removed by incubation in 5 mL 10 mM Tris-HCl pH 8.0, 10 mM NaCl, 0.2% v/v NP-40, 1 mM PMSF (Sigma-Aldrich) and 1 mM benzamidine (Sigma-Aldrich) for 10 min on ice. Nuclei were pelleted by centrifugation and resuspended in 450 µL 1.1x restriction enzyme buffer (NEB buffer 2 for HindIII or NEB buffer 3 for BglII) and 7.5 µL of 20% w/v SDS was added for 1 h incubation at 37°C. 50 µL of 20% v/v Triton X-100 was then added for 1 h at 37°C, and 200 U HindIII or BglII (New England Biolabs) was added for overnight digest at 37°C. Reactions were then supplemented with an additional 200 U of restriction enzyme for a 6 h incubation. After heat inactivation and sequential incubation with SDS and Triton X-100 as above, the samples were diluted to 7 mL in 1x T4 DNA ligase buffer, and 4000 U T4 DNA ligase (New England Biolabs) was added for overnight incubation at 16°C. Reactions were then supplemented with an
additional 4000 U T4 DNA ligase for a 6 h incubation. Samples were then digested by addition of proteinase K (VWR) to 1 mg/mL for overnight incubation at 65°C, and DNA was purified by phenol/chloroform extraction followed by isopropanol precipitation. 3C products were quantified by a Taqman-based qPCR strategy using the LightCycler 480 probes master mix (Roche) on a Roche Lightcycler 480 using the following program: 5 min at 95°C, followed by 48 cycles of 10 s at 95°C and 30 s at 62°C. Values were normalized to a BAC standard generated from RP23-416M23, RP23-31E15 and RP23-238C12 (BACPAC Resources Center, Children’s Hospital Oakland Research Institute) and a second normalization step was performed by normalizing values to the nearest neighbor fragment. Primers and probes are listed in Table 7.
### Table 7. List of 3C primers and probes

<table>
<thead>
<tr>
<th>HindIII primers and probes</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Eβ F</td>
<td>GAAATTTGGCATCGGTTTGC</td>
</tr>
<tr>
<td>Eβ probe</td>
<td>TGAGACAAATTTGGAGGAGCTACAGTC</td>
</tr>
<tr>
<td>Eβ NN R</td>
<td>CCGAGGCTGCTGTTGTGAG</td>
</tr>
<tr>
<td>Trbv5 R</td>
<td>AGCTGACACAGAAGGAAGCATTT</td>
</tr>
<tr>
<td>Trbv13-1 R</td>
<td>CCACTGAGCATGACCCCTTCTT</td>
</tr>
<tr>
<td>Trbv13-3 R</td>
<td>GTCTCAGGTCACCCTCACCA</td>
</tr>
<tr>
<td>Trbv14 R</td>
<td>CAGCTTTTGAGTGGCATGGT</td>
</tr>
<tr>
<td>Trbv16 R</td>
<td>TATCATGCAAGCTGGATTC</td>
</tr>
<tr>
<td>Trbv23 R</td>
<td>TACACCGGCAAGGAGACT</td>
</tr>
<tr>
<td>Trbv26 R</td>
<td>GTAGCTCTCAGCATGGAGATG</td>
</tr>
<tr>
<td>Trbv29 R</td>
<td>CTTAGATCCCCCTTTGAGTC</td>
</tr>
<tr>
<td>Try5 R</td>
<td>CAGAGACGCTTAAGCTGGAC</td>
</tr>
<tr>
<td>Trbv29 F</td>
<td>CACCCTCCAGGAGAAGTC</td>
</tr>
<tr>
<td>Trbv29 probe</td>
<td>AGTCACACTACGTACGCTCGTGC</td>
</tr>
<tr>
<td>Trbv29 NN R</td>
<td>CAGCTTTAGCTGCCCATTC</td>
</tr>
<tr>
<td>5′PC R</td>
<td>GAAGTGTTGTGAGGATGTGGGTTG</td>
</tr>
<tr>
<td>DJβ R</td>
<td>CAGTGAACATCATTTGTGGTG</td>
</tr>
<tr>
<td>Eβ R</td>
<td>TCCTAAGGAGGAGGAGGTG</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BglII primers and probes</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Eβ F</td>
<td>GTCAATGCCATTATGGCTAGATCC</td>
</tr>
<tr>
<td>Eβ probe</td>
<td>TGCTGCTTCAGGACGCTGGT</td>
</tr>
<tr>
<td>Eβ NN R</td>
<td>AGATGGGATCATGAGCTGCTTA</td>
</tr>
<tr>
<td>Trbv29 R</td>
<td>CAGTCATACCATAGCCTGAGT</td>
</tr>
<tr>
<td>Trbv30 R</td>
<td>TGTCCTGCTTGTGGG</td>
</tr>
<tr>
<td>Trv5 R</td>
<td>ATGATAGGCTCCTGAGAG</td>
</tr>
<tr>
<td>Trbv30 F</td>
<td>ATCGAAGAGTGGTG</td>
</tr>
<tr>
<td>Trbv30 probe</td>
<td>AGAGTCCAGCAGTAGCAGGCG</td>
</tr>
<tr>
<td>Trbv30 NN R</td>
<td>AACGTCAGTGAAGCTAGC</td>
</tr>
<tr>
<td>5′PC R</td>
<td>GCCGTGAGATAGTGGAGAGG</td>
</tr>
<tr>
<td>DJβ R</td>
<td>CAGGCCAATGCTATCTGAC</td>
</tr>
<tr>
<td>Eβ R</td>
<td>CCCTAAGGATAGTATAGGAG</td>
</tr>
</tbody>
</table>

### 3.18. Determination of preexisting Tcrb rearrangements

Rearrangements were amplified by Taq-based PCR of gDNA using the following program: 5 min at 95°C, followed by 35 cycles of 30 s at 95°C, 15 s at 62°C and 1 min at 72°C. VDJβ rearrangements were determined by using a Vβ-specific forward primer and a reverse primer located downstream of either Trbv1-6 or Trbv2-7. DJβ rearrangements
were determined by using a forward primer located upstream of \textit{Trbd1} or \textit{Trbd2} and a reverse primer downstream of either \textit{Trbv1-6} or \textit{Trbv2-7}. Primers are listed in Table 8.

**Table 8. List of recombination (PCR) primers**

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{Trbv1} F</td>
<td>GCCACACGGGTCACTGATAC</td>
</tr>
<tr>
<td>\textit{Trbv2} F</td>
<td>AGTCGGTTCCAACCCTCAAGGT</td>
</tr>
<tr>
<td>\textit{Trbv3} F</td>
<td>CAGTTGAAGACCCAGATGTTC</td>
</tr>
<tr>
<td>\textit{Trbv4} F</td>
<td>CAAGGCGCTTCTCACTCTCA</td>
</tr>
<tr>
<td>\textit{Trbv5} F</td>
<td>GCCCAAGACAGCTCCACGCTAC</td>
</tr>
<tr>
<td>\textit{Trbv12-1} F</td>
<td>GACAAAGGATCTCCCTACGAGC</td>
</tr>
<tr>
<td>\textit{Trbv13-1} F</td>
<td>GTACAAGGCCACAGAAACA</td>
</tr>
<tr>
<td>\textit{Trbv12-2} F</td>
<td>AAGATGATGCGCTTGGGAGCTA</td>
</tr>
<tr>
<td>\textit{Trbv13-2} F</td>
<td>AGCCAAAGAACCTCCCTCCCTC</td>
</tr>
<tr>
<td>\textit{Trbv13-3} F</td>
<td>ATATACTGTCGTGCCAGCACG</td>
</tr>
<tr>
<td>\textit{Trbv14} F</td>
<td>GCTGTGAGGCTAAAGGAAACT</td>
</tr>
<tr>
<td>\textit{Trbv15} F</td>
<td>CGATTTCAAGGCTGAGATGC</td>
</tr>
<tr>
<td>\textit{Trbv16} F</td>
<td>CTCAGCTCAGTGGCCCAAT</td>
</tr>
<tr>
<td>\textit{Trbv17} F</td>
<td>CAATCCAGCTGCGCTAACCA</td>
</tr>
<tr>
<td>\textit{Trbv19} F</td>
<td>GAAGGCTATGATGCGTCTCG</td>
</tr>
<tr>
<td>\textit{Trbv20} F</td>
<td>TCCCCATGATCACTCCGACTT</td>
</tr>
<tr>
<td>\textit{Trbv23} F</td>
<td>CTGTGTCGCCCTCCAGTCA</td>
</tr>
<tr>
<td>\textit{Trbv24} F</td>
<td>CTCACTCAAGTGCTTGGTCTGA</td>
</tr>
<tr>
<td>\textit{Trbv26} F</td>
<td>CCTTCAAAGCTCCTTGACGC</td>
</tr>
<tr>
<td>\textit{Trbv29} F</td>
<td>CCCGTATTCTGAGTGCCTCT</td>
</tr>
<tr>
<td>\textit{Trbv30} F</td>
<td>GAGTTGGATTCACCAAGGAA</td>
</tr>
<tr>
<td>\textit{Trbv31} F</td>
<td>TACTTTGGGAGGTAGTAGT</td>
</tr>
<tr>
<td>\textit{Trbd1} F</td>
<td>GAGGACGAGCTTATCTGGGAGGTT</td>
</tr>
<tr>
<td>\textit{Trbd2} F</td>
<td>GTAGGCACTCGTTGGGAAAAGACT</td>
</tr>
<tr>
<td>\textit{Trbj1-6} R</td>
<td>CACCAACCCTCAGTGCAAAATG</td>
</tr>
<tr>
<td>\textit{Trbj2-7} R</td>
<td>TGAGAGCTGCTTCTCCTACTATGGAT</td>
</tr>
</tbody>
</table>

**Table 8. List of recombination (PCR) primers**

PCR products were run on an agarose gel and any bands were excised and purified using the QIAquick Gel Extraction Kit (Qiagen). The purified PCR products were then cloned into pCR4-TOPO-TA using the TOPO TA Cloning Kit for Sequencing (Invitrogen). Plasmid DNA from transformed bacteria was purified using the QIAprep Spin Miniprep Kit (Qiagen) and sent to the Duke DNA Analysis Facility for sequencing.
3.19. Tcrb repertoire analysis

High throughput sequencing of the Tcrb repertoire was performed by using the Immunoseq service (Adaptive Biotechnologies). In brief, gDNA was harvested from the target cell population and multiplex PCR performed to amplify Tcrb rearrangements. The multiplex PCR reaction is optimized to prevent any primer amplification bias. The resulting PCR product is then processed for high throughput sequencing to obtain the sequence of the VDJβ junction, which provides information on Vβ, Dβ and Jβ gene segment usage and junctional diversity.

Initial VL3-3M2 repertoire assessment was performed by Taqman-based qPCR of gDNA isolated from unsynchronized cells using the LightCycler 480 probes master mix (Roche) on a Roche Lightcycler 480 using the following program: 5 min at 95°C, followed by 48 cycles of 10 s at 95°C and 30 s at 62°C. Readings were normalized to Cd14. Primers and probes are listed in Table 9.

Table 9. List of recombination (Taqman) primers

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>DJβ F</td>
<td>GGTGTCTTGGTTCCACAGC</td>
</tr>
<tr>
<td>DJβ probe</td>
<td>CCTTGTACTTGTTGGCAGGACACCCCCGAC</td>
</tr>
<tr>
<td>Trbv13-2 R</td>
<td>AGCCAAAGAGACTTCTCCCTCC</td>
</tr>
<tr>
<td>Trbv19 R</td>
<td>GAAGGCTATGATGCTTGTCCTC</td>
</tr>
<tr>
<td>Trbv26 R</td>
<td>CTTTCAACTCTACCTGGAGAC</td>
</tr>
<tr>
<td>Trbv29 R</td>
<td>CCCTGATTCTGGATTTCTGCT</td>
</tr>
<tr>
<td>Trbv30 R</td>
<td>GAGTGATCCACCAAGGACAA</td>
</tr>
<tr>
<td>Cd14 F</td>
<td>GCTCAAATTTCCAATCTTACCACG</td>
</tr>
<tr>
<td>Cd14 R</td>
<td>AGTCAGTCTCTGGAGGCGCGAATG</td>
</tr>
<tr>
<td>Cd14 probe</td>
<td>CGCACCCTAGCCGGCTTTAAGGACAGA</td>
</tr>
</tbody>
</table>
For the timed recombination system, each VL3-3M2 line was transfected with a pX458 derivative targeting the VDJ\(_{\beta}\) junction on the rearranged allele and three days later, TCR\(\beta^-\)GFP\(^+\) cells were single cell sorted into 96-well plates. For each cell line, twelve clones were cultured for 21 days and recombination was scored by flow cytometry to quantify TCR\(\beta^+\) cells. To enrich for cells with functional rearrangements, we chose three clones from each line which displayed recombination frequencies closest to the median and sorted TCR\(\beta^+\) cells. The Tcrb repertoire was assessed by flow cytometric staining using V\(_{\beta}\)-specific mAbs and by high throughput sequencing of rearrangement events using the Immunoseq service (Adaptive Biotechnologies). Verification of the direct V\(\beta\)-to-J\(\beta\) rearrangements was performed by SYBR-based qPCR using the QuantiFast SYBR Green PCR kit (Qiagen) on a Roche Lightcycler 480 using the following program: 5 min at 95°C, followed by 45 cycles of 10 s at 95°C and 30 s at 62°C. Readings were normalized to Cd14. qPCR primers are listed in Table 10.

**Table 10. List of recombination (SYBR) primers**

<table>
<thead>
<tr>
<th>Target</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trbv2-Trbj2-5</td>
<td>AGTCGCTTCCAACCTCAAAGT</td>
<td>TGCCCTGGCCCAAAGTACTGG</td>
</tr>
<tr>
<td>Trbv3-Trbj2-4</td>
<td>CAGTTGAAAGACCAGATGGTTC</td>
<td>CCCGCACCAAAGTACAAGGTG</td>
</tr>
<tr>
<td>Cd14</td>
<td>GCTCAAACTTTCCAGAATCTACGAC</td>
<td>AGTCAGTTCTGGAGCCGGAAATC</td>
</tr>
</tbody>
</table>
3.20. Statistical methods

Unless otherwise stated, data were analyzed by two-way ANOVA with adjustment for multiple comparisons using GraphPad Prism software. Differences with adjusted p values of <0.05 were considered significant. For DamID-qPCR, effect sizes and significance were calculated by fitting a model of the form:

\[ y_{ijk} = \mu + \alpha_i + \beta_j + b_k + e_{ijk} \]

Where \( y_{ijk} \) is the DamID signal for the \( i \)-th locus, \( j \)-th VL3-3M2 line and \( k \)-th run. The model explains variation in expression levels relative to a baseline \( \mu \), which represents the average expression level at the \( T_p4 \) locus in wild-type cells. The terms \( \alpha_i \) show the difference in expression at other loci (relative to baseline) in wild-type cells. The terms \( \beta_j \) show the effect (average difference in expression) for other lines relative to wild-type. The assumption is that this effect is constant across all loci. The terms \( b_k \) represent the random effect of run/batch, assumed to have a zero mean Gaussian distribution. Finally \( e_{ijk} \) is unexplained/measurement error, also assumed to have a zero mean Gaussian distribution but independent of the batch effect. The model was fit to observed data for all runs but restricted to loci lying between \( T_p4 \) and \( Trbv30 \), inclusive, using the method of restricted maximum likelihood (REML).

3.21. Data deposition

High throughput sequencing data described herein has been deposited under the accession number GEO: GSE116954.
4. Specific aim 1: To determine how Tcrb is localized to the nuclear periphery

Some of the work presented and discussed in this chapter has been modified from the following publication:


Luperchio, T.R., Wong, X. and Reddy, K.L contributed by performing and analyzing DamID-seq; Byrd, A.T. contributed by making some of the CRISPR/Cas9-derived cell lines used; Roy Choudhury, K. contributed by performing the analysis of DamID effect sizes.

4.1. Introduction

The subnuclear localization of the Tcrb locus is unusual in that it is frequently found at the NL in DN thymocytes when Tcrb undergoes recombination (Schlimgen et al., 2008). This is in contrast to most known examples of developmentally-regulated genes, which frequently localize to the NL when inactive and relocate to the nuclear interior upon activation. It was observed that Tcrb alleles marked by p53-binding protein 1 (53BP1), a protein which is recruited to DSBs and utilized as a marker of actively-recombining Tcrb alleles, were more frequently found in the interior compared to the
nuclear periphery. Additionally, the abundance of RAG2 is lower at the nuclear periphery compared to the interior (Chan et al., 2013). These observations gave rise to the hypothesis that the nuclear periphery contributes to allelic exclusion by reducing the rate at which Tcrb alleles recombine.

Although it had been shown that the Tcrb locus is frequently associated with the NL, we did not have a map of Tcrb-NL interactions in high detail. This is important because of the apparent contradiction that transcription is required for recombination to occur (Abarrategui and Krangel, 2006, 2007; Yancopoulos and Alt, 1985), yet the Tcrb locus is frequently associated with the transcriptionally suppressive environment of the nuclear periphery. Furthermore, the trypsinogen regions found within the Tcrb locus are transcriptionally silent in T lineage cells. Therefore, it is essential to determine the association of the Tcrb locus with the NL at high resolution in order to identify and characterize regulatory regions and mechanisms that mediate the high association of the Tcrb locus with the NL.

To study Tcrb-NL association, we required a high resolution map of chromatin-NL interactions. Previous studies of Tcrb subnuclear localization were performed using DNA-FISH (Chan et al., 2013; Schlimgen et al., 2008) and were greatly limited by the poor resolution of the technique. The maximum resolution of light microscopy is limited to 0.2 μm due to diffraction limits resulting from the wavelengths of light used for imaging (Thorn, 2016). This diffraction limit makes it difficult to observe fine details in chromatin structure. A second limitation that reduces the resolution of microscopy is caused by the usage of BACs approximately 200 kb in size for probe construction. DNA-
FISH studies typically generate fluorescent probes from BACs in order to obtain a robust signal that is detectable above background. However, the large size of the BACs prevents high resolution mapping of the Tcrb locus. Thus, we required a different technique in order to analyze the association of the Tcrb locus with the NL at high resolution.

We selected DamID as the technique of choice to analyze Tcrb-NL interactions due to the high resolution of the technique and the wide body of substantiating literature (Guelen et al., 2008; van Steensel and Belmont, 2017; Vogel et al., 2007). DamID utilizes Dam to methylate adenine residues on GATC sequences in close proximity to the protein, generating a modification that is not naturally present on eukaryotic chromatin (Figure 10). By expressing a fusion protein of Dam and lamin B1, Dam can be used to selectively label chromatin at the NL. Dam-methylated DNA can be discriminated by the restriction endonuclease DpnI, allowing for the selective enrichment and amplification of Dam-methylated DNA. This is accomplished by the ligation of adapters onto the DpnI-digested blunt ends. A DpnII digest is then used to cut unmethylated GATC sequences in order to reduce background (not shown in Figure 10). The product is then subjected to PCR amplification, using the ligated adapters as primer binding sites, generating DamID-enriched DNA which can be used in direct analysis or further processed for high throughput sequencing. DamID values are expressed as a ratio of signals from Dam-lamin B1 over unfused Dam, the latter serving as a control for differences in methylation due to genomic accessibility rather than localization at the NL. High DamID signals indicate that a given region is frequently associated with the NL, while low DamID signals indicate that the region is positioned away from the NL.
Figure 10. Overview of DamID.

Adenine-methylated GATC sequences (in red) can be discriminated by the restriction endonuclease DpnI. Adapter sequences (in purple) are added to the ends of the digested fragments and act as primer binding sites for subsequent PCR amplification of Dam-methylated DNA.

The objective of this aim is to determine the mechanism by which the Tcrb locus associates with the NL. By obtaining a high resolution map of Tcrb-NL interactions using DamID, we will identify and characterize regulatory regions that are essential in conferring the NL association profile of the Tcrb locus.

4.2. Identification of a cell line suitable for DamID

Previous DamID studies were mostly conducted in non-lymphoid cell lines where Tcrb is inactive. Therefore, we had to select an appropriate lymphoid cell line to perform DamID. An important requirement was that the subnuclear localization of Tcrb in the chosen cell line had to be similar to that in primary thymocytes.

Four candidate cell lines were selected for testing. The first was the P5424 cell line, which originated from a spontaneous Rag1<sup>−/−</sup>Tp53<sup>−/−</sup> DN thymic lymphoma on a mixed 129/Sv x C57BL6 background (Mombaerts et al., 1995). The second was the VL3-3M2 cell line, a DP cell line induced using radiation leukemia virus on a C57BL6/Ka background (Groves et al., 1995). The third was the A70.2 cell line, a pre-B cell line induced using Abelson murine leukemia virus from an E<sub>μ</sub>-Bcl2 transgenic mouse on a mixed 129/Sv x C57BL6 background (Bredemeyer et al., 2006). The fourth was the M12...
cell line, which originated from a spontaneous B cell lymphoma on a BALB/c background (Kim et al., 1979).

We used DNA-FISH to determine the subnuclear localization of the Tcrb locus in the tested cell lines. BAC probes used for detecting the Tcrb locus were RP23-75P5, which marks the 5′ RC-distal end of the locus and overlaps with Trbv1, and RP23-457D7, which marks the 3′ end of the locus (Figure 11A). In all four cell lines, subnuclear localization of the Tcrb locus was similar to control Rag2−/− DN thymocytes, although the P5424 cell line exhibited a lower but not significantly different peripheral localization of both ends of the Tcrb locus (Figure 11B to 11F). We also observed that the P5424 and M12 cell lines had aberrations in copy number, with three Tcrb alleles present per cell (Figure 11B, 11E and 11G). In spite of the irregularity in Tcrb copy number in P5424 cells, we decided to proceed with performing DamID in both P5424 and VL3-3M2 cells, as these two cell lines are T lineage cell lines.
Figure 11. Subnuclear localization of Tcrb in candidate cell lines.

(A) Relative locations of BAC probes used. (B-E) Representative DNA-FISH images of (B) P5424, (C) VL3-3M2, (D) A70.2 and (E) M12 cell lines stained with lamin B1 (blue), RP23-75P5 (red) and RP23-457D7 (green). The M12 cell line was not stained with RP23-75P5. The scale bar indicates 2 μm. (F) Colocalization of indicated BAC probes with lamin B1. *Rag2*−/− DN thymocytes were included as a reference. (G) Number of Tcrb alleles per cell. In (F) and (G), data were combined from 1-4 independent experiments. Total number of cells analyzed: P5424: 457D7 = 123, 75P5 = 39; VL3-3M2: 457D7 = 253, 75P5 = 184; A70.2: 457D7 = 135, 75P5 = 151; M12: 457D7 = 23; *Rag2*−/− DN Thy: 457D7 = 31, 75P5 = 159.

4.3. Determination of preexisting Tcrb rearrangements in VL3-3M2 cells

Before assaying the NL association of the Tcrb locus in the P5424 and VL3-3M2 cell lines, it was imperative to understand the rearrangement status of the Tcrb loci, as existing rearrangements either confound data interpretation or even render the cell line unsuitable. As the P5424 cell line originated from a *Rag1*−/− mouse (Mombaerts et al., 1995), the Tcrb loci are in the germline configuration. For the VL3-3M2 cell line, we performed PCRs on gDNA using Vβ- or Dβ-specific forward primers, and reverse primers downstream of either Trbj1-6 or Trbj2-7. PCR products were cloned and subjected to Sanger sequencing. We identified a complete Trbv3-Trbd1-Trbj2-4 rearrangement on first allele, and a partial Trbd1-Trbj2-4 rearrangement on the second allele (Figure 12). Hereafter, the first allele will be referred to as the rearranged allele while the second allele will be referred to as the experimental allele. Thus, in VL3-3M2 cells, the 500kb
segment of the Tcrb locus between Trbv3 and Trbd1 is represented in a monoallelic fashion on the experimental allele.

Figure 12. VL3-3M2 Tcrb loci.

(A) Schematic of Tcrb loci. The rearrangements on each allele are indicated. (B) Sequence and alignment of Tcrb loci junctions. N, non-templated nucleotides.

4.4. DamID on P5424 and VL3-3M2 cells

To perform DamID on P5424 and VL3-3M2 cells, we transduced the cell lines with retroviruses containing a green fluorescent protein (GFP) reporter gene and either unfused Dam or Dam-lamin B1. GFP+ cells were sorted and gDNA extracted for DamID enrichment. We performed qPCR (DamID-qPCR) in order to quantify the NL association of the Tcrb locus and four control loci that showed consistently low or high signals. Tcrb
qPCR primers were designed to provide representative coverage of the locus, including V_β, RC-distal (T_d) and RC-proximal (T_p) trypsinogen regions, and the RC (Figure 13A). DamID values were scored as the log_2 ratio of signals from Dam-lamin B1 over the unfused Dam control.

![Diagram of TCRb locus and DamID-qPCR profiles](image)

**Figure 13. DamID analysis of VL3-3M2 and P5424 cells.**

(A) Location of DamID-qPCR primers. (B) DamID-qPCR profiles of the Tcrb locus (left) and non-Tcrb control loci (right). The data represent the mean and standard error (SE) of 3 independent experiments.

We observed that both P5424 and VL3-3M2 cells had a similar pattern of NL association at the Tcrb locus and control loci (Figure 13B). The Tcrb locus was not homogenously associated with the NL, but instead contained regions of high and low NL association. 
association. The regions of highest NL association were at both trypsinogen regions, while the regions of lowest NL association were the RC and the RC-distal Vβ gene segments spanning \textit{Trbv2} to \textit{Trbv13-3}. The differences between the two cell lines was in the dynamic range of signals, with P5424 cells exhibiting a smaller dynamic range that capped at a 1.8 fold difference (0.84 on a log_2 scale) between Dam-lamin B1 and unfused Dam at T_p2. In contrast, the difference at the corresponding point in VL3-3M2 cells was 5.7 fold (2.5 on a log_2 scale). The small dynamic signal range in P5424 cells was also evident at the non-\textit{Tcrb} control loci. Due to the larger signal range in VL3-3M2 cells, we decided to perform all subsequent experiments in this cell line.

\section*{4.5. High throughput DamID sequencing in VL3-3M2 cells}

The DamID-qPCR conducted in section 4.4 was a great improvement over existing knowledge about \textit{Tcrb} peripheral localization gathered by DNA-FISH. However, the DamID-qPCR coverage was still limited as it only covered 21 sites, with an average of 33 kb between sites.

To generate a higher resolution, unbiased map of \textit{Tcrb}-NL association, we performed DamID coupled with high throughput sequencing (DamID-seq) on VL3-3M2 cells (Figure 14). We observed an enrichment of DamID signals throughout the \textit{Tcrb} locus and these signals were highest at the trypsinogen regions, consistent with these regions constituting LADs. The Vβ gene segments in the middle of the \textit{Tcrb} locus exhibited a mixed character with interspersed regions of high and low DamID signals,
although DamID signals were generally higher at RC-proximal \( V_\beta \) gene segments than RC-distal \( V_\beta \) gene segments. The RC exhibited the lowest DamID signals.

![Diagram of DamID-seq and CTCF ChIP-seq profiles](image)

**Figure 14. DamID-seq in VL3-3M2 cells.**

DamID-seq profile of the \( Tcrb \) locus in VL3-3M2 cells. \( Tcrb \) locus schematic is indicated above, with functional gene segments colored in black and pseudogenes colored in red. DamID-seq experimental values are expressed as the log\(_2\) ratio of Dam-Lamin B1 over unfused Dam. Reads are merged from two independent replicates. CTCF ChIP-seq data are from GSE41743 (Shih et al., 2012). CTCF ChIP-seq data are expressed as reads per million.

The \( Tcrb \)-NL association map obtained by DamID-seq was broadly similar to that obtained by DamID-qPCR. Both techniques showed that there were smaller portions of the \( Tcrb \) locus that were not highly-associated with the NL, a result not observed by DNA-FISH due to the lower resolution of the latter technique. The alternating NL association of the \( Tcrb \) locus will be analyzed in subsequent sections.

### 4.6. Deletion of LAD border regions by CRISPR/Cas9

LAD borders positioned within the \( Tcrb \) locus could serve as potential regulatory elements that control the activation of selected \( Tcrb \) gene segments. Two LAD borders
are of particular interest. The first is the sharp LAD border upstream of the RC (usRC) that segregates the RC from the RC-proximal trypsinogen region. The second is found in the middle of the Vβ array spanning Trbv11 to Trbv14, where a gradual transition from non-LAD to LAD chromatin occurs. It is notable that both LAD borders are found around clusters of CBEs. usRC is located near 5’PC and PDβ1CBE, which were previously characterized to function as a chromatin tether and insulator respectively, while the region spanning Trbv11 to Trbv14 contains four CBEs that exhibit high CTCF binding.

To determine the function of the LAD borders described above in regulating NL association of the Tcrb locus, we utilized the CRISPR/Cas9 system to delete the regions containing the LAD border. We decided to delete large regions for two reasons: first, to capture any possible element involved in LAD border function, and second, prior work indicated that the deletion of a single CBE can result in a transfer of function to an adjacent CBE, partially compensating for loss-of-function (Chen et al., 2015; Jain et al., 2018; Majumder et al., 2015b; Qiu et al., 2018). Therefore, the targeted regions span all CBEs found proximal to the LAD border.

To generate derivative VL3-3M2 lines, pairs of guides flanking the target regions were transfected into VL3-3M2 cells, and single cells were expanded to identify clones bearing the desired deletion. The ΔusRC line bears a 28 kb deletion in the usRC region that covers 5’PC through PDβ1CBE, leaving core PDβ1 promoter elements intact (Figure 15, Table 1). The ΔTrbv11-Trbv14 line bears a 32 kb deletion that deletes the six functional and two pseudogene Vβ gene segments from Trbv11 to Trbv14, inclusive, and the four CBEs found within this region (Figure 15, Table 1).
Figure 15. LAD border deletions in VL3-3M2 cells.

Schematic of portions of the Tcrb locus selected for CRISPR/Cas9 targeting. A dual guide strategy was employed to delete the region in between the pair of guides. Location of guides relative to locus elements is shown.
Figure 16. DamID-qPCR analysis of LAD border-deleted lines.

(A) Location of DamID-qPCR primers. (B) DamID-qPCR profiles of the Tcrb locus (left) and non-Tcrb control loci (right). (C) DamID-qPCR difference maps relative to wild-type VL3-3M2. Data were transformed from (B). The data represent the mean and SE of 3 (ΔTrbv11-Trbv14), 6 (ΔusRC) or 16 (wild-type) independent experiments. * p < 0.05; ** p < 0.01; *** p < 0.001; **** p < 0.0001, by two-way ANOVA with Holm-Sidak’s multiple comparisons test. Δ, region deleted.
DamID-qPCR was performed on wild-type, ΔusRC and ΔTrbv11-Trbv14 lines. We plotted raw DamID values (Figure 16A and 16B) and also differences in DamID values in the derivative VL3-3M2 lines relative to wild-type (Figure 16A and 16C). ΔusRC exhibited a large reduction in the NL association of RC-proximal Vβ and trypsinogen regions while the rest of the Tcrb locus resembled wild-type. ΔTrbv11-Trbv14 also exhibited a reduction in the NL association of RC-proximal Vβ and trypsinogen regions, although this was modest compared to the change observed in ΔusRC. Therefore, we decided to focus our efforts on ΔusRC.

We performed DamID-seq on ΔusRC cells in order to obtain a high resolution map of the Tcrb locus in this cell line (Figure 17). The DamID-seq profiles between wild-type and ΔusRC were strikingly different at the RC-proximal Vβ and trypsinogen regions, but were very similar for the rest of the genome, indicating that the effect of usRC deletion is highly localized. The RC-proximal trypsinogen region and the two most RC-proximal Vβ gene segments, Trbv29 and Trbv30, exhibited substantial reductions in DamID signals, while the region spanning Trbv23 to Trbv26 had modest reductions in DamID signals.

The reduction in NL association of the RC-proximal regions in ΔusRC could be caused by Eβ-dependent activation of this region. The RC is transcriptionally active in VL3-3M2 cells, and is presumed to mediate positioning of the RC away from the NL. We hypothesized that the deletion of the usRC LAD border allowed the influence of Eβ to spread beyond the border and into the RC-proximal region, causing activation and thus dissociation from the NL.
**Figure 17. DamID-seq in wild-type and ΔusRC lines.**

DamID-seq profile of the Tcrb locus. Tcrb locus schematic is indicated above, with functional gene segments colored in black and pseudogenes colored in red. DamID-seq experimental values are expressed as the log₂ ratio of Dam-Lamin B1 over unfused Dam. Reads are merged from two independent replicates. CTCF ChIP-seq data are from GSE41743 (Shih et al., 2012). CTCF ChIP-seq data are expressed as reads per million.

### 4.7. Deletion of Eβ by CRISPR/Cas9

To test the hypothesis that the RC is positioned away from the NL due to the activity of Eβ and that this activity is normally confined to the RC by the usRC LAD border, we assessed the effects of Eβ deletion by DamID-qPCR (Figure 18A). ΔEβ exhibited an increase in NL association at the RC (as indicated by Trbj2-7) and at all Vβ
gene segments, but minimal changes were observed at both trypsinogen regions (Figure 18B and 18C). The effects of Eβ deletion on NL association were generally opposite in nature to the effects of usRC deletion, although the regions affected only exhibited a partial overlap.

To determine the functional relationship between usRC and Eβ, we created an additional line with both elements deleted (ΔusRC+ΔEβ) and observed an overall Tcrb-NL association pattern that was similar to the ΔEβ line (Figure 18B and 18C). Because usRC only appeared to have an effect on NL association on an Eβ-sufficient background, we concluded that the LAD border function of usRC acts to modulate the activity of Eβ by preventing the spreading of non-LAD chromatin into the RC-proximal trypsinogen and Vβ regions.
Figure 18. DamID-qPCR analysis of Eβ-deleted lines.

(A) Schematic showing regions deleted in the indicated lines. (B) Location of DamID-qPCR primers. (C) DamID-qPCR difference maps of the Tcrb locus (left) and non-Tcrb control loci (right) relative to wild-type VL3-3M2. Wild-type and ΔusRC data are identical to Figure 16. The data represent the mean and SE of 2 (ΔusRC+ΔEβ) or 3 (ΔEβ) independent experiments. * p < 0.05; ** p < 0.01; *** p < 0.001; **** p < 0.0001, by two-way ANOVA with Holm-Sidak’s multiple comparisons test. Δ, region deleted.
The DamID-qPCR results imply that the activity of E<sub>β</sub> is propagated over short and long ranges by two separate mechanisms. Short range E<sub>β</sub> activity is restricted from spreading upstream of the RC by the usRC LAD border. Long range E<sub>β</sub> activity affects the V<sub>β</sub> region, evidenced by the increase in NL association of V<sub>β</sub> region in the ΔE<sub>β</sub> line. These long-range effects appear to be distinct from the linear spread of E<sub>β</sub> activity that is restricted by usRC, as usRC prevents the short range effects of E<sub>β</sub> on NL association but not the distal effects of E<sub>β</sub> on NL association.

4.8. Dissection of the usRC LAD border

The region deleted in ΔusRC is 28 kb in size and contains multiple elements that may participate in LAD border function (Figure 15). Based on the previous bioinformatics analysis that showed enrichment of CBEs at LAD borders (Guelen et al., 2008), we hypothesized that the CBEs at 5′PC and PD<sub>β<sub>1</sub></sub><sup>CBE</sup> contribute, at least in part, to LAD border function. To dissect the functional elements that contribute to usRC LAD border activity, we created VL3-3M2 lines bearing smaller deletions. Due to the large number of cell lines described in the subsequent figures together with the large number of sites analyzed by DamID-qPCR, an unbiased approach was required to determine the overall changes in DamID-qPCR observed in each of the derivative VL3-3M2 cell lines. This approach will be described after all the derivative cell lines have been introduced. It is important to note that with the exception of the two E<sub>β</sub>-deleted lines, all derivative VL3-3M2 cell lines described retain PD<sub>β<sub>1</sub></sub> transcriptional activity (Figure 19).
The first set of cell lines divided usRC into three regions, each containing known regulatory elements (Figure 20). Δ5’PC deleted the 1 kb region surrounding 5’PC. ΔPDβ1 deleted the 4 kb region covering PDβ1CBE and PDβ1700, the latter containing a binding site for the transcription factor Sp1 (Sikes et al., 1998). ΔInt deleted the 23 kb intervening region between Δ5’PC and ΔPDβ1, which contains the trypsinogen gene Prss2 and LTR-LINE-LTR, a cluster of repetitive elements that was shown to function as a barrier-type insulator (Carabana et al., 2011).

The second set of cell lines was generated to subdivide the 23 kb region deleted in ΔInt (Figure 21). Within this 23 kb intervening region, we deleted the 4.9 kb region containing Prss2 (ΔPrss2) and the 3.1 kb region containing the LTR-LINE-LTR repetitive element (ΔLTR).

The third set of cell lines divided ΔPDβ1 into two components (Figure 22). The first line (ΔPDβ1CBE) deleted a 3.2 kb region encompassing both CBEs within PDβ1. The second line (ΔPDβ1700) deleted a 700 bp region containing the Sp1 binding element.

The final set of cell lines deleted a combination of elements in order to determine if NL association was regulated by multiple elements scattered across a larger region (Figure 23). These cell lines were designed based on the previous finding that the deletion of PDβ1 (which included PDβ1CBE) caused the insulator activity of PDβ1CBE to be transferred to 5’PC (Majumder et al., 2015b). We thus created two additional VL3-3M2 lines, the first bearing deletions in 5’PC and PDβ1CBE (Δ5’PC+ΔPDβ1CBE) and the second bearing deletions in 5’PC, PDβ1CBE and PDβ1700 (Δ5’PC+ΔPDβ1).
Figure 19. PDβ1 transcription in VL3-3M2 derivative lines.

Transcription originating from PDβ1 on the experimental allele was analyzed by qPCR. Expression is normalized to Actb, with Actb expression set to 4 on a log$_{10}$ scale. The data represent the mean and SE of 2 ($\Delta5'$PC+$\Delta$PDβ1$^{CBE}$) 3 ($\Delta5'$PC; ΔInt; $\Delta$PDβ1$^{700}$; $\Delta5'$PC+$\Delta$PDβ1; ΔusRC+$\Delta$Eβ), 4 ($\Delta$PDβ1$^{CBE}$; ΔEβ), 5 ($\Delta$PDβ1), 6 (wild-type) or 8 (ΔusRC) independent experiments. * p < 0.05, **** p < 0.0001, by one-way ANOVA with Holm-Sidak’s multiple comparisons test. #, not detectable.
Figure 20. DamID-qPCR analysis of usRC-dissected lines.

(A) Schematic showing regions deleted in the indicated lines. (B) Location of DamID-qPCR primers. (C) DamID-qPCR difference maps of the Tcrb locus (left) and non-Tcrb control loci (right) relative to wild-type VL3-3M2. Wild-type and ΔusRC data are identical to Figure 16. The data represent the mean and SE of 3 (Δ5′PC), 4 (ΔInt) or 5 (ΔPDβ1) independent experiments. * p < 0.05; ** p < 0.01; *** p < 0.001; **** p < 0.0001, by two-way ANOVA with Holm-Sidak’s multiple comparisons test.
Figure 21. DamID-qPCR analysis of Int-dissected lines.

(A) Schematic showing regions deleted in the indicated lines. (B) Location of DamID-qPCR primers. (C) DamID-qPCR difference maps of the Tcrb locus (left) and non-Tcrb control loci (right) relative to wild-type VL3-3M2. Wild-type and ΔusRC data are identical to Figure 16. ΔInt data are identical to Figure 20. The data represent the mean (ΔPrss2) or mean and and SE 2 (ΔLTR) independent experiments. * p < 0.05; *** p < 0.001; **** p < 0.0001, by two-way ANOVA with Holm-Sidak’s multiple comparisons test.
Figure 22. DamID-qPCR analysis of PDβ1-dissected lines.

(A) Schematic showing regions deleted in the indicated lines. (B) Location of DamID-qPCR primers. (C) DamID-qPCR difference maps of the Tcrb locus (left) and non-Tcrb control loci (right) relative to wild-type VL3-3M2. Wild-type and ΔusRC data are identical to Figure 16. ΔPDβ1 data are identical to Figure 20. The data represent the mean and SE of 3 (ΔPDβ1700) or 5 (ΔPDβ1CBE) independent experiments. * p < 0.05; ** p < 0.01; *** p < 0.001; **** p < 0.0001, by two-way ANOVA with Holm-Sidak’s multiple comparisons test.
Figure 23. DamID-qPCR analysis of compound-deleted lines.

(A) Schematic showing regions deleted in the indicated lines. (B) Location of DamID-qPCR primers. (C) DamID-qPCR difference maps of the Tcrb locus (left) and non-Tcrb control loci (right) relative to wild-type VL3-3M2. Wild-type and ΔusRC data are identical to Figure 16. The data represent the mean and SE of 3 (Δ5′PC+ΔPDβ1CBE) or 4 (Δ5′PC+ΔPDβ1) independent experiments. # p < 0.05; ** p < 0.01; *** p < 0.001; **** p < 0.0001, by two-way ANOVA with Holm-Sidak’s multiple comparisons test.
The DamID-qPCR results were highly diverse, with some cell lines showing localized changes in NL association while others showed a broader profile of changes stretching across multiple analyzed sites. To quantitatively determine the importance of the deleted elements in the derivative cell lines, we calculated overall effect sizes and p values from the DamID-qPCR data covering the six sites spanning Trbv30 to Tp4, inclusive. The selected sites span the region which exhibited the greatest changes between wild-type and ΔusRC in the DamID-seq data (Figure 17). Effect sizes were calculated using a restricted maximum likelihood model that took into account differences in genotype, variation at sites and also potential batch effects (see Materials and Methods, section 3.20). Wild-type VL3-3M2 cells were set as the baseline of 0. The ΔPrss2 and ΔLTR lines were not included in the model as they exhibited no significant changes at any site analyzed (Figure 21).

The two Eβ-deleted lines, ΔEβ and ΔusRC+ΔEβ, were the only lines that exhibited a positive effect size, indicating that the region analyzed became more associated with the NL (Figure 24). As Eβ deletion mainly affected the RC and Vβ regions but not the trypsinogen regions, effect sizes were small and was not statistically significant in the case of ΔEβ because the analysis was heavily biased towards the RC-proximal trypsinogen region. The remainder of the derived VL3-3M2 lines exhibited negative effect sizes of varying magnitude, with ΔusRC showing the largest effect size (Figure 24). While the deletion of Δ5’PC alone had no effect on NL association, the compound deletion of 5’PC with either PDβ1 (Δ5’PC+ΔPDβ1) or PDβ1CBE (Δ5’PC+ΔPDβ1CBE) resulted in a greater effect size compared to any of the individual deletions, showing that
these elements act synergistically to regulate functionality of the usRC LAD border (Figure 24). From the abovementioned result, it appears as though PDβ1700 does not contribute to LAD border functionality. However, the effect size in ΔPDβ1 appears to be additive of both PDβ1CBE and PDβ1700 (Figure 24). Therefore, it is likely that the lack of a difference between Δ5′PC+ΔPDβ1CBE and Δ5′PC+ΔPDβ1 is due to the fact that the maximal effect size has been reached, and the deletion of additional regulatory elements in the usRC region did not cause any further loss of NL association.

Figure 24. Effect sizes of deletions in VL3-3M2 derivative lines.

Effect sizes were calculated using a restricted maximum likelihood model for the sites spanning Trbv30 to Tp4, inclusive (see Materials and Methods, section 3.20). Cell lines and schematics of the deletions are shown on the left. * p < 0.05; ** p < 0.01; **** p < 0.0001.
4.9. Discussion

In this study, we performed DamID in the VL3-3M2 cell line to generate a high resolution map of the Tcrb locus. We observed that the Tcrb locus had an alternating pattern of LAD and non-LAD regions, with both trypsinogen regions most highly associated with the NL, the RC positioned away from the NL and the V_β gene segments exhibiting a mixed NL association pattern. The RC was separated from the RC-proximal region by a LAD border that we have named usRC. By performing CRISPR/Cas9-mediated deletion of the usRC LAD border, we observed a loss of NL association of the RC-proximal portion of the Tcrb locus, with greatest effects at the RC-proximal trypsinogen region and the two most proximal V_β genes Trbv29 and Trbv30, and weaker effects from Trbv23 to Trbv26. The positioning of the RC in wild-type VL3-3M2 cells, and the RC-proximal region in ΔusRC cells is E_β-dependent, as the deletion of E_β caused a massive increase in NL association of these regions. By creating a set of VL3-3M2 derivative lines containing deletions of smaller elements within usRC, we identified three core regulatory elements: 5’PC, PD_β_1^{CBE} and PD_β_1^{700}. These three elements function cooperatively to enforce a LAD border that blocks E_β from positioning the RC-proximal region into the nuclear interior.

Previous work used DNA-FISH to show that the Tcrb locus was frequently associated with the nuclear periphery (Chan et al., 2013; Schlimgen et al., 2008). These studies used fluorescent probes constructed from BACs that span approximately 200 kb. The low resolution of these studies were unable to identify the discontinuous NL association pattern we have identified here using DamID. One caveat is that DamID was...
performed in a cell line system due to toxicity issues in vivo caused by the production of Dam protein. Furthermore, the VL3-3M2 cell line is a DP cell line, which may not be the most suitable for studying the Tcrb locus as recombination has already completed. However, our DNA-FISH experiments (Figure 11), together with prior work (Schlimgen et al., 2008) show that the frequency of NL positioning of Tcrb alleles is high in VL3-3M2 cells and primary DN and DP thymocytes, which suggests that the NL association of the Tcrb locus observed in VL3-3M2 cells may be applicable to primary thymocytes.

It is noteworthy that the changes in NL association due to the usRC LAD border deletion were confined to the RC-proximal region, and had no effects on the RC-distal region. We believe that these proximal effects are $E_\beta$-dependent, as they were apparent when usRC was deleted on an $E_\beta$-containing allele (comparing ΔusRC to wild-type) but these effects were lost when usRC was deleted on an $E_\beta$-deficient allele (comparing ΔusRC+Δ$E_\beta$ to Δ$E_\beta$) (Figure 19). The distal effects of $E_\beta$ (comparing Δ$E_\beta$ to wild-type) were unaffected by usRC deletion. From these results, we conclude that in wild-type cells, short-range $E_\beta$ activity is confined to the RC by the usRC LAD border, while long-range $E_\beta$ activity that extends to the RC-distal V$\beta$ region is independent of the LAD border.
5. Specific aim 2: To characterize the regulatory impact of the usRC LAD border

Some of the work presented and discussed in this chapter has been modified from the following publications:


Doan, E.B. contributed by performing H3K9me2 and H3K27ac ChIP. Luperchio, T.R. and Reddy, K.L. contributed by performing the computational determination of LAD borders.

5.1. Introduction

In the previous section, we identified a LAD border that we named usRC which functions to segregate the RC-proximal LAD region from the non-LAD RC. We found that usRC is constituted of three important elements: 5'PC, PDβ1^{CBE} and PDβ1^{700}. Maximal disruption of NL association required the deletion of both 5'PC and PDβ1^{CBE}. In primary DN thymocytes, 5'PC functioned as a tether that interacts strongly with Vβ gene segments across the entire Tcrb locus, while PDβ1^{CBE} acted as an insulator that blocks the
spread of active chromatin marks from \( E_\beta \) (Majumder et al., 2015b). The deletion of \( PD_\beta 1 \) caused a spread of active chromatin marks towards, and terminating at 5′PC, which also resulted in the loss of the tether function of 5′PC. This led to the conclusion that in \( PD_\beta 1 \)-deleted primary DN thymocytes, 5′PC loses its tether function in order to act as an insulator (Majumder et al., 2015b).

We have extensively characterized the LAD structure of the \( Tcrb \) locus conferred by the usRC LAD border, but did not show the link to \( Tcrb \) locus regulation. Here, we used the VL3-3M2 and derivative cell lines to study the effects of usRC LAD border deletion on \( Tcrb \) locus conformation, histone modification, transcription and recombination. We found that usRC balanced the interactions of the \( V_\beta \) array with the RC by reducing interactions of RC-proximal genes in favor of RC-distal genes. We also showed that the usRC LAD border acted as an insulator that blocked the spread of active chromatin from the RC into the RC-proximal region. Upon deletion of usRC, genes within the RC-proximal region were increased in activating histone modifications, transcription and recombination. Therefore, we concluded that usRC prevents the activation of RC-proximal genes by keeping chromatin at the NL.

### 5.2. Conditional knockout of Ctcf in DN thymocytes

Given that CBEs feature prominently in the usRC LAD border, we wanted to determine the role of CTCF in regulating the \( Tcrb \) locus using a mouse model where CTCF is ablated in thymocytes. The lab had previously shown that CTCF deficiency in DP thymocytes reduced \( Tcra \) rearrangement due to lower CTCF-dependent long distance
interactions (Shih et al., 2012). In that study, conditional knockout mice were used where floxed $Ctf$ alleles ($Ctf^{fl}$) were crossed with a Cre recombinase transgene driven by a lymphocyte protein tyrosine kinase (Lck) promoter ($Lck$-$Cre$), resulting in the deletion of $Ctf$ starting from the DN2 stage of thymocyte development (Shih et al., 2012). As $Tcrb$ recombination has already begun at the DN2 stage, this mouse model is unsuitable for the study of the $Tcrb$ locus. Therefore, we crossed $Ctf^{fl}$ alleles (Figure 25A) with a Cre recombinase transgene driven by a human $Cd2$ minigene ($Cd2$-$Cre$) (de Boer et al., 2003), which is active from the ETP stage of development to allow time for $Ctf$ alleles to be completely deleted before $Tcrb$ recombination commences. $Ctf^{fl}Cd2$-$Cre$ mice were also bred onto a $Rag2^{-/-}$ background for assays requiring unrearranged $Tcrb$ alleles.

Thymocyte development in $Ctf^{fl}Cd2$-$Cre$ mice was blocked between the DN to DP stages, which manifested as a small increase in the number of DN thymocytes and a severe decrease in the number of DP and CD4SP thymocytes and overall thymic cellularity (Figure 25B-C). We observed an increase in the number of CD8SP thymocytes, which most likely represent CD8$^+$ ISP cells, consistent with a block at the DN and DP transition. The pattern of thymocyte development we observed is similar to the one seen in $Ctf^{fl}Lck$-$Cre$ mice, where the developmental blockade was attributed to a cell cycle defect (Heath et al., 2008; Shih et al., 2012).
Figure 25. Effect of CTCF deletion on thymocyte development.

(A) Schematic of Ctcf^f alleles. Exons are numbered. loxP sites are indicated as filled triangles. Puro<sup>R</sup>, puromycin resistance gene; lacZ, lac operon β-galactosidase. (B) Representative flow cytometry profiles of wild-type Ctcf<sup>ff</sup> (top) and Ctcf<sup>ff</sup>Cd2-Cre (bottom) thymocytes. The left set of panels is pre-gated on live thymocytes with other lineage markers gated out. Subsequent gating strategy is indicated by red arrows. (C) Cellularity of total thymus and thymocyte subpopulations as shown in (B). The data represent the mean and SE from 4 mice. * p < 0.05, by unpaired t-test with Holm-Sidak correction. (D) Representative flow cytometry profiles of FDG staining in Ctcf<sup>ff</sup>Cd2-Cre thymocytes.

The Ctcf<sup>f</sup> alleles were designed to turn on expression of a lacZ reporter gene following Cre-mediated deletion. This allows cells with at least one deleted allele to be detectable by flow cytometry following incubation with fluorescein di-β-D-galactopyranoside (FDG), which gets hydrolyzed by LacZ into fluorescein (Figure 25A). The proportion of DN thymocytes that stained positive for FDG exhibited a gradual increase from DN2 to DN4 (Figure 25D).

To analyze the effects of Ctcf deletion on the Tcrb repertoire, we sorted FDG<sup>+</sup> DN3 thymocytes from Ctcf<sup>ff</sup>Cd2-Cre mice and DN3 thymocytes from control Ctcf<sup>ff</sup> mice to perform high throughput sequencing. We did not observe a dramatic alteration in the Tcrb repertoire of Ctcf-deleted thymocytes (Figure 26A). However, the relative usage of four V<sub>β</sub> gene segments were decreased by at least 1.5-fold and the relative usage of five V<sub>β</sub> gene segments were increased by at least 1.5-fold. Notably, the four V<sub>β</sub> segments with decreased usage were located proximal to a CBE, while the five V<sub>β</sub> segments with increased usage were not. Thus, to make an overall assessment of the effects of Ctcf deletion on the Tcrb repertoire, we divided V<sub>β</sub> gene segments according to the presence
or absence of a nearby CBE as previously defined (Gopalakrishnan et al., 2013). The 12 CBE-bearing V\(\beta\) segments comprised approximately half of the \(Tcrb\) repertoire in control cells, but only 37% in \(Ctcf\)-deleted cells. Conversely, the other 12 V\(\beta\) segments lacking a nearby CBE increased in usage to 63% in \(Ctcf\)-deleted cells (Figure 26B).

![Figure 26. \(Tcrb\) repertoire in wild-type and CTCF-deleted thymocytes.](image)

(A) \(Tcrb\) repertoire analyzed by high throughput sequencing of gDNA isolated from DN3 thymocytes. \(V\beta\) usage is plotted as the percentage of unique reads obtained. #, \(V\beta\) gene segment associated with a CBE. Data obtained from one experiment where samples were pooled from 2 (wild-type) or 5 (\(Ctcf^{-/-}\)) mice. (B) \(V\beta\) usage in (A) grouped by proximity to a nearby CBE as described in (Gopalakrishnan et al., 2013).
We hypothesized that the weak Tcrb rearrangement phenotype in Ctcf^{+/f}Cd2-Cre mice was caused by residual levels of CTCF remaining in cells. To determine the efficiency of Cre-mediated deletion of Ctcf, we performed qPCR on gDNA from sorted Ctcf^{+/f}Cd2-Cre FDG^+ DN3 thymocytes and from control Ctcf^{+/f} DN3 thymocytes. Approximately 95% of Ctcf alleles were deleted (Figure 27A). We performed western blots to assess the amount of CTCF protein present in cells. CTCF protein levels were reduced but not completely ablated in Ctcf^{+/f}Cd2-Cre FDG^+ DN3 thymocytes (Figure 27B). As the residual levels of CTCF might be sufficient to enable normal regulation of the Tcrb locus, we determined the levels of CTCF binding by ChIP and assayed enrichment by qPCR. CTCF binding was reduced at two Tcrb CBEs and also the Tcra enhancer E_{α1}, but were still substantially above the negative control Dad1 (Figure 27C). The inefficient ablation of Ctcf in this mouse model made further study difficult, therefore we discontinued this line of research.
Figure 27. Efficiency of CTCF deletion in Ctf\textsuperscript{eff}/Cd2-Cre mice.

(A) Deletion of Ctf\textsuperscript{eff} alleles at the genomic level. qPCR values were initially normalized to Cd14, and subsequently normalized to wild-type which was set to 1. The data represent the mean and SE from 4 mice. (B) Western blot of CTCF with a LAT loading control. (C) CTCF ChIP in the Tcrb locus and control loci. The E\textalpha enhancer from the Tcra locus is used as a positive control while Dad1 is a negative control. ChIP values are expressed relative to input. Data obtained from 1-2 replicates.

5.3. Analysis of transcription in VL3-3M2 lines

To determine whether the changes in NL association upon deletion of the LAD border caused transcriptional activation of the affected portions of the Tcrb locus, we assayed the transcription of V\beta gene segments and trypsinogen genes. Because the VL3-3M2 cell line expresses RAG and can undergo recombination (Williams et al., 2001), we ensured that we only assayed the transcription of unrearranged V\beta gene segments. This was accomplished using a primer design where one primer in each primer pair is located within the 3’ untranslated region (UTR), which would be deleted during recombination of the gene segment.

In wild-type VL3-3M2 cells, we observed low or undetectable levels of transcription in RC-proximal trypsinogen genes and V\beta gene segments, whereas RC-
distal V\(_{\beta}\) gene segments (Trbv3, Trbv5 and Trbv13-2) exhibited substantially higher levels of transcription (Figure 28). The deletion of usRC caused the two most RC-proximal V\(_{\beta}\) gene segments (Trbv29 and Trbv30) and two RC-proximal trypsinogen genes (Try5 and Prss1) to become transcriptionally active (Figure 28). Similar to the results obtained by DamID, we did not observe any changes in the transcription of RC-distal V\(_{\beta}\) gene segments. Tcrb locus transcription is entirely dependent on E\(_{\beta}\), as transcription was completely abrogated in \(\Delta E_{\beta}\) and \(\Delta usRC+\Delta E_{\beta}\) (Figure 28). Thus, similar to NL association, usRC blocks the effects of E\(_{\beta}\) on RC-proximal promoters but does not control more distal effects.
Figure 28. Tcrb locus transcription in usRC- and Eβ-deleted VL3-3M2 lines.

(A) Location of genes and gene segments analyzed. (B) Transcription of genes and gene segments on the experimental allele were analyzed by qPCR. Expression values are normalized to Actb, with Actb expression set to 4 on a log_{10} scale. The data represent the mean and SE of 3 (ΔusRC+ΔEβ), 4 (ΔEβ), 6 (wild-type) or 8 (ΔusRC) independent experiments. **** p < 0.0001, by two-way ANOVA with Holm-Sidak’s multiple comparisons test. #, not detectable.

We also analyzed transcription in the VL3-3M2 lines bearing smaller deletions within usRC. Among the three lines bearing deletions of individual elements, Δ5’PC and ΔInt had transcriptional profiles that were similar to wild-type (Figure 29A and 29B). Only ΔPDβ1 exhibited a significant transcriptional activation of RC-proximal genes, although this was limited to the trypsinogen gene Try5 (Figure 29A and 29B). We next analyzed the lines bearing smaller deletions in PDβ1, and observed that the transcriptional profile of ΔPDβ1^{700} was similar to wild-type, while ΔPDβ1^{CBE} was similar to ΔPDβ1.
(Figure 29A and 29C). This showed that the transcriptional activation in ΔPDβ1 can be entirely attributed to the region containing CBEs. Finally, when we analyzed transcription in the VL3-3M2 lines bearing compound deletions in usRC, Δ5′PC+ΔPDβ1CBE exhibited greater transcriptional activation compared to ΔPDβ1CBE, but was still lower than ΔusRC (Figure 29A and 29D). In contrast, Δ5′PC+ΔPDβ1 exhibited transcriptional activation of RC-proximal genes to a similar extent as ΔusRC (Figure 29A and 29D). In the lines with smaller usRC deletions, we often observed a reduction in RC-distal transcription. We are unsure of the cause of this reduction, but we note that we observed high variation in the data, as exemplified by the large error bars.
**Figure 29. Tcrb locus transcription in usRC-dissected lines.**

(A) Location of genes and gene segments analyzed. (B-D) Transcription of genes and gene segments on the experimental allele were analyzed by qPCR. Expression values are normalized to Actb, with Actb expression set to 4 on a log10 scale. Wild-type and ΔusRC data are identical to Figure 29. The data represent the mean and SE of 2 (Δ5′PC+ΔPDβ1CBE) 3 (Δ5′PC; ΔInt; ΔPDβ1700; Δ5′PC+ΔPDβ1), 4 (ΔPDβ1CBE) or 5 (ΔPDβ1) independent experiments. * p < 0.05, ** p < 0.01, *** p < 0.001, **** p < 0.0001, by two-way ANOVA with Holm-Sidak’s multiple comparisons test. #, not detectable.

In primary DN thymocytes, Eβ has partial or no effects on Vβ gene segment promoters (Bouvier et al., 1996; Majumder et al., 2015b; Mathieu et al., 2000; McDougall et al., 1988; Spicuglia et al., 2002). However, we observed that in the Eβ-deleted lines, Vβ gene segment transcription was completely abrogated. Because VL3-3M2 is a DP thymocyte cell line, this result may indicate that Vβ gene segment transcription may naturally transition from an Eβ-independent mode of regulation to an Eβ-dependent mode of regulation upon transition from the DN to the DP stage of thymocyte development.

The RC-distal Vβ gene segments exhibited robust transcription in all Eβ-containing lines, which correlated with the low DamID ratios observed in this region. This agreed with the previous observation that RC-proximal and -distal Vβ gene segments have different modes of regulation (Majumder et al., 2015b; Majumder et al., 2015c), although the boundary that segregates proximal and distal Vβ gene segments is unclear.
5.4. A Tcrb recombination system in the VL3-3M2 cell line

VL3-3M2 cells contain one functionally-rearranged Tcrb allele, which should suppress rearrangements on the second, experimental allele due to allelic exclusion. However, we were able to detect rare rearrangement events on the experimental allele, which implied that Tcrb allelic exclusion was spontaneously broken at a low frequency. We utilized this unexpected feature in order to test the impact of usRC deletion on Tcrb recombination. In a preliminary experiment, we analyzed recombination in genomic DNA (gDNA) from wild-type, ΔusRC, ΔPDβ1 and Δ5'PC+ΔPDβ1 lines by qPCR. Relative to wild-type, we detected 15- to 150-fold increases in Trbv29 and Trbv30 rearrangement on the experimental allele in ΔusRC and Δ5'PC+ΔPDβ1 (Figure 30). We were unable to perform a more complete analysis of the Tcrb repertoire using this technique as most rearrangements were below the detection threshold. Furthermore, these lines had been maintained in culture for variable lengths of time, which may skew the interpretation of results.
Figure 30. Initial Tcrb recombination testing in VL3-3M2 lines.

(A) Location of genes and gene segments analyzed. (B) Relative Tcrb rearrangement in unsynchronized populations of VL3-3M2 cell lines analyzed by qPCR of gDNA. qPCR values were initially normalized to Cd14 in each sample, and normalized values for each Vβ rearrangement in VL3-3M2 mutant cell lines were expressed relative to wild-type, which was set to 1. Data generated from one sample for each genotype.

To avoid the aforementioned problems, we created a culture system in which cells bearing an initial DJβ rearrangement on the experimental allele were cultured for a defined period of time prior to assaying Vβ-to-DJβ recombination (Figure 31A). VL3-3M2 cells have a functional Tcra rearrangement, which pairs with the rearranged Tcrb allele to give rise to TCRαβ heterodimers. These TCRαβ heterodimers are expressed on the cell surface and can be detected using a pan-TCRβ monoclonal antibody. By using the CRISPR/Cas9 system to disrupt the coding sequence on the rearranged Tcrb allele, we generated TCRβ− derivatives of the VL3-3M2 lines. In these lines, any functional
rearrangement on the experimental Tcrb allele can be detected by staining for TCRβ+ cells. After expanding single-cell sorted TCRβ− cells for 21 days in culture, we assayed the rate of recombination in wild-type, ΔusRC, ΔPDβ1 and Δ5′PC+ΔPDβ1 lines by flow cytometric analysis. We observed a 2.5- to 3-fold increase in Tcrb recombination frequency in ΔusRC and Δ5′PC+ΔPDβ1 compared to wild-type (Figure 31B).

**Figure 31. Tcrb recombination in synchronized VL3-3M2 populations.**

(A) Timeline for the disruption and analysis of the rearranged allele in VL3-3M2 and derivative lines. (B) Flow cytometric analysis of Tcrb recombination in 12 clonal TCRβ− cultures grown for 21 days. Horizontal lines denote the mean. *** p < 0.001, **** p < 0.0001, by non-parametric one-way ANOVA with Dunn’s multiple comparisons test.

To analyze the Tcrb repertoire, we sorted TCRβ+ cells that arose from our culture system and analyzed Vβ usage by flow cytometry (Figure 32A). Compared to wild-type, we observed a 2.5-fold increase in the usage of Vβ7 (encoded by Trbv29) in ΔPDβ1 and a
4.5-fold increase in ΔusRC and Δ5′PC+ΔPDp1 (Figure 32B). We also observed an increase in Vβ3 (encoded by Trbv26) usage, but these changes were smaller in magnitude and not statistically significant. We were unable to assay recombination of Vβ18 (encoded by Trbv30) as a specific monoclonal antibody was unavailable. We also performed high throughput sequencing on wild-type and ΔusRC cultures to analyze the entire Tcrb repertoire. Unique rearrangements on the experimental allele were strongly biased toward the usage of Trbv29 and Trbv30 in ΔusRC (Figure 32C). Because the number of unique sequences represented a substantial proportion of the number of TCRβ+ cells seeded, the RC-proximal bias in Vβ usage reflected multiple unique recombination events, rather than the outgrowth of a limited number of clones that had rearranged early and expanded throughout the 21 day time-course.
Figure 32. Vβ repertoire in synchronized VL3-3M2 populations.

(A) Timeline for the disruption and analysis of the rearranged allele in VL3-3M2 and derivative lines. (B) Flow cytometric analysis of Tcrb repertoire in TCRβ+ cells. Horizontal lines denote the mean. (C) Tcrb repertoire analyzed by high throughput sequencing of gDNA isolated from TCRβ+ cells analyzed in (B). Vβ usage is plotted as the percentage of unique reads representing rearrangement events on the experimental allele. The data represent the mean and SE from 2 (wild-type) or 3 (ΔusRC) clonal populations. In (C), the number of unique reads obtained ranged from 162 to 240. * p < 0.05; ** p < 0.01; *** p < 0.001; **** p < 0.0001, by two-way ANOVA with Holm-Sidak’s multiple comparisons test.
The high throughput sequencing also revealed an unexpected class of \textit{Tcrb} recombination events. These recombination events appeared to be direct V\textsubscript{β}-to-J\textsubscript{β} rearrangements that paired \textit{Trbv1} or \textit{Trbv2} with either \textit{Trbj2-5} or \textit{Trbj2-7}. Based on the usage of \textit{Trbv1} and \textit{Trbv2}, these recombination events were likely to have occurred on the rearranged allele, replacing the CRISPR/Cas9-inactivated rearrangement (Figure 33A). Direct V\textsubscript{β}-to-J\textsubscript{β} rearrangements are not observed \textit{in vivo} despite having compatible RSSs, a restriction known as the beyond 12/23 rule (Bassing et al., 2000; Sleckman et al., 2000). To independently verify these rearrangement events, we performed qPCRs to detect the relative frequencies of the original \textit{Trbv3-Trbd1-Trbj2-4} rearrangement and the most abundant (\textit{Trbv2-Trbj2-5}) V\textsubscript{β}-to-J\textsubscript{β} “replacement rearrangement” in pre-sorted (mostly TCR\textsubscript{β}−) and sorted TCR\textsubscript{β}+ populations. In two wild-type and three ΔusRC samples, the original rearrangement was highly abundant both pre- and post-sort, with the replacement rearrangement detected at low levels. However, in the remaining wild-type sample, we observed a substantial depletion of the original rearrangement with a proportional increase in the replacement rearrangement in the post-sorted population (Figure 33B). Consistent with this result, two-thirds of the sorted TCR\textsubscript{β}+ cells in this sample expressed surface V\textsubscript{β}4 (encoded by \textit{Trbv2}) (Figure 33C). Further analysis of the high throughput sequencing data revealed that there was a single \textit{Trbv2-Trbj2-5} replacement rearrangement that made up the vast majority of sequences obtained from this particular sample. This indicated that the observed replacement rearrangement had occurred early during the culture period, compromising the selection for functional
rearrangements on the experimental allele. Hence, we excluded this highly unusual sample from the analyses in Figures 33B and 33C.

Figure 33. Vβ-to-Jβ replacement rearrangements on the rearranged allele.

(A) Schematic showing an example of a Trbv2-Trdj2-5 replacement rearrangement. (B) Relative frequencies of the original VL3-3M2 Trbv3-Trbd1-Trbj2-4 rearrangement and Trbv2-Trbj2-5 replacement rearrangements in three wild-type and three ΔusRC cultures, both pre-sort (mostly TCRβ−, taken at day 21) and post-sort (all TCRβ+, taken at day 31). Rearrangement values were normalized to Cd14. (C) Flow cytometric analysis of Vβ4 (encoded by Trbv2) usage in the same TCRβ+-sorted populations as in (B). The wild-type sample containing primarily replacement rearrangements was excluded from analysis in Figures 33B and 33C.

The repertoire of Tcrb rearrangements in VL3-3M2 thymocytes recapitulated the transcription phenotype, with a very strong bias towards the usage of Trbv29 and Trbv30 in both ΔusRC and Δ�5′PC+ΔPDβ1, with a weaker effect observed in ΔPDβ1. Although
we did not assay recombination in the two Eβ-deleted lines, we do not expect to see any recombination as it is well-known from *in vivo* experiments that the deletion of Eβ completely abolishes *Tcrb* recombination.

**5.5. Analysis of chromatin looping by 3C**

To gain an understanding into the regulation of transcription and recombination by usRC, we analyzed the spatial organization of the *Tcrb* locus by 3C looking at the relative interaction profile of HindIII-digested fragments. We first assayed the interactions between the Eβ-containing viewpoint fragment and the rest of the *Tcrb* locus. The Eβ-containing HindIII fragment interacted broadly across the *Tcrb* locus in wild-type cells, and these interactions were diminished by an average of 60% in ΔEβ, showing that Eβ is an essential regulator of *Tcrb* locus conformation (Figure 34A and 34B). Eβ interactions were biased towards RC-proximal fragments in ΔusRC and Δ5′PC+ΔPDβ1, with significant increases at *Trbv29* and *Try5*, and was accompanied by a decrease at RC-distal Vβ gene segments that was statistically significant at *Trbv13-1* (Figure 34A and 34B). To verify the interaction profile we observed in these cell lines, we analyzed the reciprocal set of interactions by using *Trbv29* as a viewpoint fragment and obtained a similar set of results, where *Trbv29* interactions with 5′PC and the RC were elevated in ΔusRC and Δ5′PC+ΔPDβ1 and reduced by 2- to 3-fold in ΔEβ (Figure 34A and 34C).
Figure 34. 3C profiles of the Tcrb locus using a HindIII digest.

(A) Location of viewpoint and test fragments analyzed. Viewpoints are indicated below the locus schematic and test fragments are indicated above. (B-C) 3C analysis of long-distance interactions between HindIII fragments with an Eβ (B) or Trbv29 (C) viewpoint. Results for different ligation products were normalized to their abundance in a digested and religated BAC standard and were expressed relative to the frequency of ligation with a fragment adjacent to the viewpoint. The data represent the mean and SE of 5 (ΔusRC; Δ5′PC+ΔPDβ1; ΔEβ) or 6 (wild-type) independent experiments. * p < 0.05, ** p < 0.01, *** p < 0.001, **** p < 0.0001, by two-way ANOVA with Holm-Sidak’s multiple comparisons test.
One limitation with the 3C data generated using a HindIII digest is that the interactions of Trbv30 cannot be analyzed due to issues with primer design. Therefore, we repeated the 3C experiments with a BglII digest. The interactions of Eβ with RC-proximal fragments analyzed were all increased in ΔusRC (Figure 35A and 35B), and the interactions of Trbv30 with 5'PC and the RC were all significantly increased in ΔusRC (Figure 35A and 35C). These experiments illustrated that the interaction profiles of Trbv29 and Trbv30 are highly similar. Therefore, we showed that Eβ mediates looping between Vβ gene segments and the RC, and the usRC balances these interactions by reducing the frequency of looping to RC-proximal targets in favor of RC-distal targets.
Figure 35. 3C profiles of the Tcrb locus using a BglII digest.

(A) Location of viewpoint and test fragments analyzed. Viewpoints are indicated below the locus schematic and test fragments are indicated above. (B-C) 3C analysis of long-distance interactions between BglII fragments with an Eβ (B) or Trbv30 (C) viewpoint. Results for different ligation products were normalized to their abundance in a digested and religated BAC standard and were expressed relative to the frequency of ligation with a fragment adjacent to the viewpoint. The data represent the mean and SE of 4 (ΔusRC) or 5 (wild-type) independent experiments. * p < 0.05, *** p < 0.001, by two-way ANOVA with Holm-Sidak’s multiple comparisons test.

5.6. H3K27ac profile of the Tcrb locus

To delve into the mechanism by which usRC regulates NL association, transcription and recombination, we analyzed the profile of histone modifications across the Tcrb locus. We analyzed the distribution of H3K27ac because this histone mark is known to be enriched at actively-transcribed genes and their enhancers (Creyghton et al.,
We performed H3K27ac ChIP-seq on wild-type, ΔusRC and ΔEβ lines. Wild-type cells displayed a huge enrichment of H3K27ac at the RC and smaller peaks at Vβ gene segments (Figure 36). There was a large peak at Trbv3, which most likely came from the rearranged allele. Similar to the transcription profile, ΔEβ cells were completely depleted in H3K27ac across the Tcrb locus (Figure 36), demonstrating that Eβ is absolutely necessary for Tcrb locus activity. Notably, ΔusRC exhibited an elevated level of H3K27ac that extended from the RC into the RC-proximal trypsinogen region and the Vβ gene segments Trbv29 and Trbv30, terminating at a CBE located in between Trbv27 and Trbv28 (Figure 36). Therefore, the increased levels of H3K27ac in the Tcrb locus is correlated with regions of the locus that exhibited elevated transcription and Vβ-to-DJβ recombination.
Figure 36. H3K27ac ChIP-seq profile in VL3-3M2 lines.

H3K27ac ChIP-seq and DamID-seq profile of the Tcrb locus. Tcrb locus schematic is indicated above, with functional gene segments colored in black and pseudogenes colored in red. CTCF ChIP-seq data are from GSE41743 (Shih et al., 2012). H3K27ac CTCF ChIP-seq values are expressed as reads per million. DamID-seq data are identical to Figure 17. Reads were merged from two independent replicates.

Two studies reported that H3K27me3 is enriched in regions of LADs adjacent to LAD borders (Guelen et al., 2008; Harr et al., 2015), although conflicting reports exist (Bian et al., 2013; Kind et al., 2015). It was hypothesized that the enrichment of H3K27me3 near LAD borders maintains the integrity of LADs, with disruption of H3K27me3 causing a loss of LAD identity (Harr et al., 2015). Because H3K27ac modifications occur on the same histone residue as H3K27me3, the elevated levels of H3K27ac we observe in ΔusRC should serve to erase H3K27me3 modifications, causing the loss of NL association of the H3K27ac-enriched region. In order to test this
hypothesis, we performed H3K27me3 ChIP-qPCR at sites across the RC-proximal portions of the *Tcrb* locus. One of the sites chosen was 5'PC, which had been shown to be H3K27me3-enriched in mouse embryonic fibroblasts (Simon et al., 2013). However, relative to H3K27me3-enriched control loci (*Ins1* and *Olfr446*), we only managed to detect low or background levels of H3K27me3 at all tested sites across the *Tcrb* locus in wild-type and ΔusRC lines. We detected a small trend towards increased H3K27me3 in ΔEβ, although the increased levels were still below our positive control loci (Figure 37). Thus, the hypothesis that H3K27me3 is involved in the maintenance of LAD borders (Harr et al., 2015) does not hold true at the *Tcrb* locus.

**Figure 37. H3K27me3 enrichment in VL3-3M2 lines.**

H3K27me3 ChIP profile of the *Tcrb* locus and non-*Tcrb* control genes. ChIP values are expressed relative to input. The data represent the mean and SE of 3 independent experiments. * p < 0.05, by two-way ANOVA with Holm-Sidak’s multiple comparisons test. Δ, region deleted.
The propagation of H3K27ac from the RC into RC-proximal areas support our assertion that usRC functions by preventing the ingress of the E\(\beta\)-activated region into the upstream LAD. It is unclear if the propagation of H3K27ac and associated transcriptional activation is required for the initiation or the maintenance of the NL-dissociated section in the \(\Delta\)usRC cell line. However, it is clear that this mechanism does not involve H3K27me3, at least at the \(Tcrb\) locus.

5.7. **H3K9me2 profile of the Tcrb locus**

Given that LADs have been shown to be enriched in the repressive modification H3K9me2 (Kind et al., 2013; Peric-Hupkes et al., 2010; Poleshko et al., 2017; Wen et al., 2009) and that the disruption of H3K9me2 modifications causes a loss of NL association (Bian et al., 2013; Harr et al., 2015; Kind et al., 2013; Towbin et al., 2012), we analyzed the distribution of H3K9me2 modifications across the \(Tcrb\) locus by ChIP-seq in wild-type, \(\Delta\)usRC and \(\Delta\)E\(\beta\) lines. In wild-type VL3-3M2 cells, we observed a correlation of H3K9me2 with DamID (Figure 38). \(\Delta\)E\(\beta\) cells showed increased H3K9me2 across the V\(\beta\) array and the RC, consistent with increased NL association observed by DamID-qPCR (Figure 38). Surprisingly, we did not detect any loss of \(Tcrb\) locus H3K9me2 in \(\Delta\)usRC, despite the substantially-reduced NL association of the RC-proximal region (Figure 38).
Figure 38. H3K9me2 ChIP-seq profile in VL3-3M2 lines.

H3K9me2 ChIP-seq and DamID-seq profile of the Tcrb locus. Tcrb locus schematic is indicated above, with functional gene segments colored in black and pseudogenes colored in red. CTCF ChIP-seq data are from GSE41743 (Shih et al., 2012). H3K9me2 ChIP-seq values are expressed as the log$_2$ ratio of bound H3K9me2 over input. CTCF ChIP-seq values are expressed as reads per million. DamID-seq data are identical to Figure 17. Reads were merged from two independent replicates.

To determine if the pattern of histone modifications at Tcrb is a general feature of genome regulation, we performed a genome-wide analysis of our DamID-seq and ChIP-seq data from wild-type VL3-3M2 cells. We obtained the positions of LAD borders across the genome and graphed the average intensity of DamID, H3K9me2 and H3K27ac readings across a 100 kb region centered on the LAD border. This revealed that H3K9me2 was indeed enriched in LADs and depleted in non-LADs, and that there was a sharp transition from H3K9me2 to H3K27ac across the LAD border (Figure 39A). Next,
we divided the genome into 100 kb bins and plotted DamID signals against H3K9me2 signals. We observed a positive correlation between DamID and H3K9me2 with most points lying along the diagonal, but also observed that a fraction of the genome had low DamID signals despite enrichment for H3K9me2 (Figure 39B), similar to the result seen at the Tcrb locus in ΔusRC. An example of such a region is presented below (Figure 39C).

![Graphs and plots showing correlations between DamID and H3K9me2 signals.](image)

**Figure 39. Genome-wide features of LADs and LAD borders.**

(A) Feature profiles of LAD borders. LAD borders were identified genome-wide, and average signals across LAD borders were graphed across 100 kb centered on the LAD border. Non-LADs are oriented on the left and LADs are oriented on the right. (B) Correlation between H3K9me2 ChIP-seq and DamID-seq in wild-type VL3-3M2 cells. Reads were placed into 10 kb bins. Pearson correlation coefficient = 0.718. (C) Example of a region of the genome exhibiting a negative correlation between H3K9me2 and DamID. Location of genes within the displayed region are indicated above.
LADs have been shown to be enriched for H3K9me2 and this histone modification is believed to be essential for the anchoring of chromatin to the NL (Harr et al., 2015; Kind et al., 2013; Peric-Hupkes et al., 2010; Poleshko et al., 2017). Our H3K9me2 ChIP-seq data for the Tcrb locus in wild-type and ΔEβ cells agrees with this. However, we did not detect any loss of H3K9me2 in ΔusRC. Furthermore, we observed other regions of the genome with low DamID and high H3K9me2 signals. These observations indicate that H3K9me2 is necessary but not sufficient for NL association.

5.8. Histone modifications in primary thymocytes

Because it is not possible to perform DamID in vivo, we required an alternative method to assay NL association of the Tcrb locus in primary thymocytes. As the DamID and H3K9me2 profiles looked very similar in wild-type VL3-3M2 cells, we decided to perform H3K9me2 ChIP-seq in primary Rag2−/− DN and DP thymocytes as a surrogate measure of NL association. Both DN and DP thymocytes had similar H3K9me2 profiles where both trypsinogen regions had high signals while the RC and sections of the Vβ array were depleted of H3K9me2 (Figure 40). These H3K9me2 profiles in primary thymocytes are broadly similar to that in wild-type VL3-3M2 cells, although we did observe a greater depletion of H3K9me2 across the Vβ array in primary thymocytes.
Figure 40. H3K9me2 ChIP-seq profile in primary thymocytes.

H3K9me2 ChIP-seq profile of the Tcrb locus. Tcrb locus schematic is indicated above, with functional gene segments colored in black and pseudogenes colored in red. VL3-3M2 data are from wild-type VL3-3M2 cells. CTCF ChIP-seq data are from GSE41743 (Shih et al., 2012). H3K9me2 ChIP-seq values are expressed as the log2 ratio of bound H3K9me2 over input. CTCF ChIP-seq values are expressed as reads per million. DamID-seq data are identical to Figure 14. Reads were merged from two independent replicates.

To determine if the lower H3K9me2 levels in primary thymocytes were due to a greater activation state in the Vβ region, we performed H3K27ac ChIP-seq on primary Rag2−/− DN and DP thymocytes. We observed very high levels of H3K27ac deposition at the Vβ array in DN thymocytes, with levels almost similar to that of the RC (Figure 41). Although the H3K27ac DP thymocyte profile looked superficially similar to that in wild-type VL3-3M2 cells, we observed a high background visible as a low-level H3K27ac enrichment in the Tcrb locus (Figure 41) and also throughout the genome. This high
background was present in both replicates, and caused a suppression of peak intensities due to the method of normalization performed. Due to these issues, we were unable to draw any conclusions about the deposition of H3K27ac in primary DP thymocytes.

**Figure 41. H3K27ac ChIP-seq profile in primary thymocytes.**

H3K27ac ChIP-seq profile of the Tcrb locus. Tcrb locus schematic is indicated above, with functional gene segments colored in black and pseudogenes colored in red. VL3-3M2 data are from wild-type VL3-3M2 cells. CTCF ChIP-seq data are from GSE41743 (Shih et al., 2012). H3K27ac and CTCF ChIP-seq values are expressed as reads per million. DamID-seq data are identical to Figure 14. Reads were merged from two independent replicates.

Although we did not directly assay NL association of the Tcrb locus in primary thymocytes, the similarities between the ChIP-seq profiles of primary thymocytes and wild-type VL3-3M2 cells suggest that the Tcrb locus organization is likely to be similar. In all cell types, there was an enrichment of H3K9me2 and a lack of H3K27ac at the trypsinogen regions, suggesting that these regions serve as anchors that tether the Tcrb
locus to the NL. The converse is true of the RC and parts of the V_{\beta} region, which are likely to be looped away from the NL.

5.9. Discussion

Here we showed that the loss of RC-proximal NL association observed upon deletion of the usRC LAD border resulted in the activation of genes within the RC-proximal region (Figure 42). We observed a spreading of H3K27ac from the RC into the RC-proximal region, terminating at a CBE located between Trbv27 and Trbv28. The two RC-proximal V_{\beta} gene segments that became enriched in H3K27ac, Trbv29 and Trbv30, exhibited a massive increase in transcription and recombination. The RC-proximal trypsinogen genes, which are normally silent in T lineage cells, also became transcriptionally active upon deletion of the usRC LAD border. The activation of the RC-proximal region also resulted in increased looping to the RC, while the looping of the RC to RC-distal sites decreased.
Figure 42. Model of usRC LAD border function in VL3-3M2 cells

Tcrb locus organization at the NL is depicted for wild-type VL3-3M2 and UsRC. Dotted lines indicate looping interactions with the RC; line thickness represents relative frequency of looping interactions.

Using the data generated from the VL3-3M2 derivative lines, we established that the loss of NL association is a prerequisite for activation of the RC-proximal region. DamID showed that Δ5’PC+ΔPDβ1, which phenocopied ΔusRC in all assays performed, had a similar effect size to Δ5’PC+ΔPDβ1<sup>CBE</sup> (Figure 24). However, transcriptional activation of RC-proximal genes in Δ5’PC+ΔPDβ1<sup>CBE</sup> was lower compared to both
ΔusRC and Δ5′PC+ΔPDβ1 (Figure 29D), which implied that additional elements work in tandem with the CBEs in usRC to provide full LAD border functionality.

The studies performed herein were mostly done in VL3-3M2 cells because DamID cannot be performed in vivo, and also due to the relative simplicity of generating derivative cell lines with desired deletions as compared to making mouse models for each deletion. To translate our cell line studies into primary cells, we created a conditional knockout mouse model that deleted Ctcf in thymocytes, and also performed ChIP-seq in Rag2−/− thymocytes. The former would have provided an overall understanding of the function of CTCF in regulating Tcrb recombination, with the caveat that the deletion is not specific to the Tcrb locus and other aspects of CTCF-mediated cell regulation would also be affected. Although the conditional knockout of Ctcf was unsuccessful due to substantial levels of CTCF protein remaining in thymocytes, ChIP-seq showed that the H3K9me2 profile of the Tcrb locus was highly similar in the VL3-3M2 cell line and primary DN and DP thymocytes (Figure 40). Furthermore, the H3K27ac ChIP-seq profiles of VL3-3M2 and primary DN thymocytes showed a lack of H3K27ac at both RC-proximal and -distal trypsinogen regions (Figure 41). Therefore, the ChIP-seq suggest that the LAD organization of the Tcrb locus is likely applicable to both DN and DP thymocytes in vivo.

Factors that regulate antigen receptor recombination are typically studied in mouse models due to the lack of suitable cell line systems, with the exception of Abelson virus-transformed B cell lines used to study Igh and Igk recombination (Jain et al., 2018; Muljo and Schlissel, 2003). Here we utilized the VL3-3M2 cell line to study Tcrb
recombination. Although VL3-3M2 cells are DP cells with one functionally rearranged Tcrb allele, these cells can still undergo Tcrb recombination on the second, experimental allele, albeit at a low frequency. The Vβ repertoire on the experimental allele in wild-type VL3-3M2 cells was well-distributed across all Vβ segments, unlike IgH recombination in Abelson-transformed pro-B cell lines which are substantially biased towards RC-proximal rearrangements (Jain et al., 2018).

We characterized the appearance of direct Vβ-to-Jβ recombination events which are not naturally observed in vivo, in apparent violation of the beyond 12/23 rule of recombination (Bassing et al., 2000; Sleckman et al., 2000). The beyond 12/23 rule is mediated by poor RAG-mediated nicking of the Jβ RSSs, but can be stimulated by synapsis with a 3'Dβ RSS, creating a mechanism that favors Dβ-to-Jβ over Vβ-to-Jβ recombination (Banerjee and Schatz, 2014; Drejer-Teel et al., 2007). There are two possible reasons why we have managed to capture these beyond 12/23 events in our recombination system in VL3-3M2 cells. The first is that Vβ-to-Jβ recombination events are occurring naturally but at such low rates that they are swamped out by regular Tcrb recombination events. Because the recombination rate of the experimental allele is so low, these Vβ-to-Jβ recombination events can be observed at a comparatively greater frequency. The second is that a VDJβ recombination event causes a change in locus structure that allows Eβ to activate Vβ gene segments upstream of the site of recombination, allowing for increased RAG activity that eventually results in the generation of a Vβ-to-Jβ recombination event utilizing upstream Vβ and downstream Jβ.
gene segments. These two possibilities do not disprove the beyond 12/23 rule, and further study is required to understand how this restriction is mediated.

Experiments characterizing antigen receptor loci are typically performed on recombination-deficient cells, as recombination events can confound analysis by drastically altering the structure of the locus and the proximity of various genetic elements. However, this is not a great concern when studying \textit{Tcrb} in VL3-3M2 cells, as the recombination rate is so low that the vast majority of experimental alleles in a population do not undergo rearrangement. Nevertheless, we took additional precautions where possible, which included the constant replenishment of cell lines from frozen stocks to limit the amount of time that the cells are growing in culture and accumulating recombination events, and also designing of primers to avoid the detection of recombined products whenever possible.
6. Conclusions

6.1. The generation of antigen receptor diversity

Lymphocyte antigen receptors are highly diverse in order to respond to a wide variety of threats that the host organism encounters. This is achieved through the recombination of large arrays of gene segments into a functional gene. The large and complex nature of antigen receptor genes require a multitude of mechanisms in order to maximize antigen receptor diversity, like the regulation of chromatin accessibility to RAG (Schatz and Ji, 2011) and through locus conformation (Bossen et al., 2012; Jhunjhunwala et al., 2009). We have focused on characterizing the mechanism by which the subnuclear localization of the Tcrb locus at the nuclear periphery regulates Tcrb recombination.

Among antigen receptor loci, Tcrb is unique in that it is frequently associated with the NL during the developmental stage in which Tcrb undergoes recombination (Kosak et al., 2002; Schlimgen et al., 2008). Tcrb alleles positioned at the NL undergo recombination at lower rates than alleles in the nuclear interior, and one potential mechanism by which the NL suppresses recombination is through the sequestration of Tcrb alleles from RAG (Chan et al., 2013).

Here we determined the association of the Tcrb locus with the NL at high resolution using DamID. We observed an alternating pattern of NL association where the RC was free of the NL, the RC-proximal and -distal trypsinogen regions were highly associated with the NL, while the Vβ array had a mix of NL-associated and NL-free regions. We discovered a LAD border that we named usRC, located in between the RC
and the RC-proximal trypsinogen region. Deletion of the usRC LAD border caused the dissociation of 300 kb of RC-proximal chromatin from the NL, which included the RC-proximal trypsinogen region and the V_\beta gene segments Trbv29 and Trbv30. Genes and gene segments within the affected region became activated, which resulted in a massive increase in recombination of Trbv29 and to a lesser extent, Trbv30. Our results demonstrate that the NL association of the Tcrb locus, mediated by the usRC LAD border, contributes to the regulation of Tcrb repertoire diversity by reducing the activation and rearrangement of Trbv29 and Trbv30 in order to limit the overall contribution of these two V_\beta gene segments to the Tcrb repertoire.

6.2. NL association of the Tcrb locus

By performing DamID in VL3-3M2 cells, we showed that the Tcrb locus was strongly associated with the NL at both RC-proximal and -distal trypsinogen regions, while NL association was lower at the V_\beta array and the RC. NL association of the RC-proximal region was maintained by the usRC LAD border located between the RC and the RC-proximal trypsinogen region, which blocked the effects of E_\beta from propagating upstream past usRC.

The NL association of the Tcrb locus suggests a conformation where the RC-proximal and -distal trypsinogen regions act as tethers which anchor the locus to the NL. By deletion of usRC, we directly showed that the loss of NL tethering at the RC-proximal region caused the activation of Trbv29 and Trbv30, resulting in increased recombination of these V_\beta genes. In the future, we would like to recreate the usRC deletion in a mouse
model in order to show that the LAD border functionality of usRC we characterized in the VL3-3M2 cell line is similar in vivo. A somewhat similar study had been performed in vivo where 475 kb of the Tcrb locus spanning Trbv5 to Trbj1-6, inclusive, was deleted (Senoo et al., 2003). The region deleted includes the RC-proximal trypsinogen region and usRC. In this mouse model, the authors showed a massive increase in transcription and recombination of the most RC-proximal Vβ gene Trbv4. This phenotype is highly similar to the phenotype observed in our usRC-deleted VL3-3M2 cells. However, the deletion of an exceptionally large part of the Tcrb locus confounds interpretation of the results.

We hypothesize that a usRC-deleted mouse model will phenocopy the changes in Tcrb locus regulation that we observed in VL3-3M2 cells. The primary advantage of this mouse model would be the study of recombination during natural thymocyte development, as compared to studying infrequent recombination events in a DP cell line. One disadvantage of this mouse model is that we will be unable to directly assess the NL association of the Tcrb locus using DamID, as DamID has not been successfully performed in vivo due to toxicity issues resulting from Dam expression. However, as we have showed in section 5.8, we can perform H3K9me2 ChIP-seq as a surrogate measure of NL association. H3K9me2 modifications have been shown to be enriched in LADs (Guelen et al., 2008; Kind et al., 2013; Poleshko et al., 2017), and that the disruption of H3K9me2 causes LADs to dissociate from the nuclear periphery (Bian et al., 2013; Harr et al., 2015; Towbin et al., 2012). However, if we do not observe a change in the H3K9me2 profile of the Tcrb locus, similar to the result in the ΔusRC line (Figure 38), we can utilize H3K27ac ChIP-seq to determine if there is a spreading of H3K27ac marks
from the RC, which would indicate that the NL association of the RC-proximal region is disrupted.

Another potential method to assay the importance of the trypsinogen regions in anchoring the Tcrb locus to the NL would be to delete both trypsinogen regions, but leaving usRC intact. This can be performed either in VL3-3M2 cells or in a mouse model. We hypothesize that the deletion of both trypsinogen regions will cause the entire Tcrb locus to lose association with the NL, despite the presence of the usRC LAD border. This would result in a massive increase in activation across the Vβ array. As NL association was postulated to contribute to Tcrb allelic exclusion (Chan et al., 2013), we predict primary murine thymocytes with both trypsinogen regions deleted will exhibit a partial loss of allelic exclusion in thymocytes. This genetic model will show whether the silent trypsinogen regions interspersed within the Tcrb locus serve a functional regulatory role or are simply bystander genes.

### 6.3. Mechanisms of LAD border regulation

Most studies of chromatin-NL interactions have generally focused on characterizing the factors that instruct the positioning of chromatin at the NL (Meuleman et al., 2013; Poleshko et al., 2017; Reddy et al., 2008; Zullo et al., 2012). However, based on our understanding of TAD chromatin organization, domain borders are extremely important for proper compartmentalization of regulatory elements (Bonev and Cavalli, 2016; Yu and Ren, 2017). Hence, it is surprising that to date, only two studies have
performed cursory analyses of LAD borders (Guelen et al., 2008; Harr et al., 2015), and no in-depth studies of LAD borders have been performed.

The first study that described LAD borders utilized a purely bioinformatics approach to extract information from multiple high throughput datasets in order to illustrate the general features of LAD borders (Guelen et al., 2008). The authors showed that three factors showed enrichment at LAD borders: CBEs, active promoters oriented away from the LAD and CpG islands. The authors also showed that substantial H3K27me3 was detected in the outer 200 kb of LADs adjacent to LAD borders, and speculated that H3K27me3 enrichment prevents the spread of active chromatin into LADs. As the analysis presented the average of features across all LAD borders, it cannot be determined if the features described apply to all LAD borders genome-wide, or that LAD borders exhibit more heterogeneity that was not picked up by the method of analysis. An alternative method of investigation utilizing principal component analysis may have been used to subset LAD borders based on their feature profiles, which would have allowed a better understanding of whether heterogeneity exists at LAD borders.

The second study described LAD borders at *Igh* and *Ikzf1* loci, which are LADs in fibroblasts but non-LADs in pro-B cells (Harr et al., 2015). Similar to the first study (Guelen et al., 2008), the authors of the second study identified H3K27me3 peaks at the outer regions of the *Igh* and *Ikzf1* LADs adjacent to LAD borders (Harr et al., 2015). The authors showed that binding of the transcription factor YY1 could promote deposition of H3K27me3 by recruitment of the H3K27me3 methyltransferase, enhancer of zeste homolog 2 (EZH2). They further showed that the global perturbation of H3K27me3 by
knockdown of either YY1 or EZH2, or the inhibition of H3K27me3 deposition using 3-deazaneplanocin A (DZNep) caused the loss of NL association of Igh and Ikzf1 loci. The authors concluded that H3K27me3 is required to demarcate regions of chromatin destined for targeting to the NL. It is worthwhile to note that in fibroblasts, the Ikzf1 LAD borders lack active, divergent promoters oriented away from the LAD but do contain CBEs and also nearby CpG islands associated with convergent promoters.

The usRC LAD border contains multiple CBEs and a nearby active promoter oriented away from the LAD in the form of PDβ1. H3K27me3 modification (Figure 37) and CpG islands are absent from usRC and nearby regions. Upon deletion of usRC, a new LAD border was created at a CBE located in between Trbv27 and Trbv28, which blocked the propagation of active chromatin features like H3K27ac and transcription past the CBE. This new LAD border also features the presence of an active promoter oriented divergently from the LAD at Trbv29, recapitulating the two features we observed at usRC. The location of this new LAD border is interesting in that it is not formed at the very first CBE located upstream of the usRC-deleted CBEs, but instead skips two CBEs at Trbv29 and Trbv30. This is different to previous characterized CBE deletions, where the regulatory features of a deleted CBE were transferred to the CBE closest to the deleted CBE (Chen et al., 2015; Jain et al., 2018; Majumder et al., 2015b; Qiu et al., 2018). We are unsure why the CBEs at Trbv29 and Trbv30 do not assume the position of the new LAD border.

usRC and the new LAD border in ΔusRC cells, located between Trbv27 and Trbv28, are positive for two of the four features that are enriched at LAD borders
genome-wide (Guelen et al., 2008). The lack of H3K27me3 at the Tcrb locus is in contrast to the Ikzf1 LAD (Harr et al., 2015), suggesting that there are at least two subsets of LAD borders. The first subset, that includes usRC and the Trbv27/28 LAD border, are enforced by CBE insulation coupled with a divergent active promoter to create a sharp transition between active and inactive regions of chromatin. The second subset, that includes Ikzf1, are enforced by H3K27me3-mediated repression to prevent the spreading of active chromatin into the LAD.

6.4. Antigen receptor diversification by intergenic CBEs

Intergenic CBEs located downstream of V gene arrays have been documented in the Tcra/Tcrd, Igh and Igk loci. These intergenic CBEs have been documented to be critical in promoting the diversification of V gene segment usage. Of the three loci with documented intergenic CBEs, the Igh locus contains the most extensively-studied set of intergenic CBEs, IGCR1. IGCR1 forms a chromatin loop with distal CBEs located downstream of the 3’RR enhancer. Enhancer activity is constrained within this chromatin loop, limiting the activation of Igh gene segments to D<sub>H</sub> and J<sub>H</sub> gene segments found within (Guo et al., 2011b). The deletion of IGCR1 caused a new chromatin loop to form between the distal CBEs and a CBE upstream of IGCR1. This new loop domain contains the V<sub>H</sub> gene segments V<sub>H</sub>7183 and V<sub>H</sub>Q52, which became activated in IGCR1-deleted cells (Qiu et al., 2018). The activation of these V<sub>H</sub> gene segments caused a heavy bias in the V<sub>H</sub> repertoire toward the two activated V<sub>H</sub> gene segments (Guo et al., 2011b; Jain et al., 2018; Lin et al., 2015; Qiu et al., 2018). Furthermore, aberrant V<sub>H</sub>-to-D<sub>H</sub>
rearrangements could also be detected in IGCR1-deleted cells (Guo et al., 2011b; Lin et al., 2015).

The intergenic CBEs in the Tcra/Tcrd locus, INT1/2, have a very similar function to IGCR1. In wild-type DP thymocytes, INT1/2 forms a chromatin loop with the TEA promoter. Deletion of INT1/2 caused the formation of a new chromatin loop between TEA and another intergenic CBE termed INT3, causing a large increase in Tcrd recombination utilizing Trdv2-2 and Trdv3, gene segments that were included in the new loop domain (Chen et al., 2015; Zhao et al., 2016). Aberrant Vδ-to-Dδ rearrangements could also be detected (Chen et al., 2015).

Finally, the Igk locus contains the intergenic CBEs Cer and Sis. The deletion of either Cer or Sis caused a biased Vκ gene segment usage towards RC-proximal gene segments and also changes in Igk locus contraction (Xiang et al., 2013; Xiang et al., 2011).

usRC bears structural and functional similarities to all three aforementioned clusters of intergenic CBEs. The most striking phenotype in usRC-deleted VL3-3M2 cells was that the transcription and recombination of the RC-proximal Vβ gene segments Trbv29 and Trbv30 were massively increased, with rearrangements of these two Vβ gene segments comprising half of the Tcrb repertoire in usRC-deleted cells. While we did not detect the formation of a new chromatin loop that sequestered Trbv29 and Trbv30 together with the RC, we did observe increased interactions of the RC with RC-proximal regions at the expense of RC-distal regions. Therefore, all four of the complex antigen receptor loci containing large arrays of V gene segments, Tcra/Tcrd, Tcrb, Igh and Igk,
have intergenic CBEs located downstream of the V gene array that serve in promoting repertoire diversification. The mechanism to regulate repertoire diversification in Tcρb is different from the other three intergenic CBEs, given that the Tcρb locus is the only antigen receptor locus that is frequently associated with the NL during the developmental stage when recombination takes place (Kosak et al., 2002; Schlimgen et al., 2008).

One important feature of the intergenic CBEs at IGCR1 and INT1/2 is the suppression of V-to-D recombination. We were unable to assay this in VL3-3M2 cells because the experimental allele had already undergone Dβ-to-Jβ recombination. However, a mouse model bearing a usRC deletion, as suggested in section 6.2, would allow us to address this question. Similar to the V-to-D recombination events observed in IGCR1 and INT1/2-deleted developing lymphocytes, we expect to observe Vβ-to-Dβ recombination events utilizing Trbv29 and Trbv30 in usRC-deleted primary DN thymocytes.

### 6.5. Targeted deletions of regulatory elements

Due to the large, 28 kb size of the usRC LAD border, we dissected this region by deleting smaller regions containing known elements as opposed to performing targeted deletions of these elements. This allowed us to rapidly identify regulatory regions important for usRC LAD border functionality. However, one limitation of this strategy is that the regions deleted could contain important but uncharacterized regulatory elements, and that the functionality of usRC could have been attributed to the wrong element.

We believe that we have identified the correct regulatory elements within the deleted regions. At 5′PC and PDβ1CBE, CTCF binding to these sites are well-established
(Figure 27) and (Majumder et al., 2015b; Shih et al., 2012), and that the observed functions of these CBEs are consistent with the known roles of CTCF (Ong and Corces, 2014). At \( PD_\beta 1^{700} \), multiple studies have been performed characterizing regulatory regions within \( PD_\beta 1 \), and the Sp1 binding element is the only known element within \( PD_\beta 1^{700} \) (Sikes et al., 1998). Nevertheless, it is possible to directly show that these elements mediate usRC LAD border function by specifically mutating the elements described. However, such a strategy would involve making more successive CRISPR/Cas9 mutations, which increases the chances of off-target effects and aberrations caused by culturing cell lines for extended periods of time. For example, creating a line that deletes the regulatory elements in \( \Delta 5'PC+\Delta PD_\beta 1 \) would require four successive CRISPR/Cas9 targeting steps.

Another regulatory element that we have attributed usRC LAD border functionality to is the core promoter elements of \( PD_\beta 1 \) which drive transcription through the RC (Figure 20). Although we did not create a line that directly disrupted \( PD_\beta 1 \) activity, we believe that a VL3-3M2 derivative cell line bearing such a deletion will phenocopy the \( \Delta E_\beta \) line. This is because \( PD_\beta 1 \) is the only active promoter in the RC in VL3-3M2 cells, and transcription through the RC is required for recombination to take place (Abarrategui and Krangel, 2006, 2007; Yancopoulos and Alt, 1985).

### 6.6. Concluding remarks

In closing, the work described here identified and characterized a method of chromatin regulation that serves to modulate V(D)J recombination in the Tcrb locus. We
described the organization of the *Tcrb* locus with respect to the NL, and identified a LAD border termed usRC that kept RC-proximal chromatin distinct from the RC. We showed that the usRC LAD border functioned to limit the activation of RC-proximal genes from the influence of Eβ, in order to suppress the recombination of RC-proximal Vβ gene segments for the purposes of maintaining a diverse *Tcrb* repertoire. We put forth a model where the trypsinogen regions act as anchors to mediate localization of the *Tcrb* locus at the NL to suppress recombination by reducing the likelihood of RAG2 protein binding, integrating together previous research performed by the laboratory. Finally, we propose that the LAD organization and regulatory mechanisms of the *Tcrb* locus characterized in the VL3-3M2 cell line system is applicable to developing thymocytes *in vivo*. 
Appendix A. R code used in high throughput sequencing analysis

The following is a sample of the code used to perform the high throughput sequencing analysis in R, which is available at https://www.r-project.org/. It does not include the LADetector code used to generate DamID tracks, which is available at https://github.com/thereddylab/pyLAD. Only the code used to analyze wild-type samples are shown; analysis of other samples utilize similar code with the filenames altered appropriately.

# Part 1: Setting up the R environment

# All files are stored in “D:/ChIP”, which is further divided into the following directories:
# “/fastq_raw” contains the original .fastq sequence files
# “/fastq_trim” contains the trimmed .fastq files
# “/bam” contains the aligned .bam files
# “/bin_200bp” contains the intermediate 200 bp bin files
# “/bin_10kb” contains the intermediate 10 kb bin files
# “/bed” contains the .bed and .bedgraph track files
# “/misc” contains other files generated during the analysis pipeline
# Directories are accessed in R using the “setwd” command

# Installation of Bioconductor packages
source("https://bioconductor.org/biocLite.R")
biocLite()
biocLite("BSgenome.Musculus.UCSC.mm9")
biocLite("dada2")
biocLite("QuasR")
biocLite("Rsamtools")
biocLite("mosaics")
biocLite("genomation")
library(BSgenome.Musculus.UCSC.mm9)
library(dada2)
library(QuasR)
library(Rsamtools)
library(mosaics)
library(genomation)

# Setting up working directories
options(scipen = 500) # Removal of scientific notation
# Part 2: Sequence alignment

This part converts reads from raw .fastq files into aligned .bam files containing the coordinates of all the reads. Runs from multiple sequencing runs and replicates are ultimately combined into a single file.

# Assigning samples

```r
system.file(wd_fastq_raw)
WT1_input1_raw <- "WT1_input1_4625-S1_S23_L007_R1_001.fastq.gz"
WT1_input2_raw <- "WT1_input2_4625-S1_S13_L002_R1_001.fastq.gz"
WT2_input1_raw <- "WT2-input_S1_L002_R1_001.fastq.gz"
WT2_input2_raw <- "WT2-input2_4587-S10_S12_L002_R1_001.fastq.gz"
WT1_H3K9me2_raw <- "WT1_H3K9me2_4689-S1_S1_L001_R1_001.fastq.gz"
WT2_H3K9me2_raw <- "WT2-H3K9me2_S4_L002_R1_001.fastq.gz"
WT1_H3K27ac_raw <- "WT1_H3K27ac_4625-S2_S24_L007_R1_001.fastq.gz"
WT2_H3K27ac_raw <- "WT2-H3K27ac_S7_L002_R1_001.fastq.gz"
```

# Trimming low-quality portions of reads

```r
fastqFilter(fn = WT1_input1_raw, fout = paste(wd_fastq_trim, "WT1_input1.fastq.gz", sep = ""), truncLen = 0, trimLeft = 0, maxN = 0, minQ = 0, rm.phix = TRUE, n = 5e+06, compress = TRUE, verbose = FALSE)
fastqFilter(fn = WT1_input2_raw, fout = paste(wd_fastq_trim, "WT1_input2.fastq.gz", sep = ""), truncLen = 0, trimLeft = 0, maxN = 0, minQ = 0, rm.phix = TRUE, n = 5e+06, compress = TRUE, verbose = FALSE)
fastqFilter(fn = WT2_input1_raw, fout = paste(wd_fastq_trim, "WT2_input1.fastq.gz", sep = ""), truncLen = 0, trimLeft = 0, maxN = 0, minQ = 0, rm.phix = TRUE, n = 5e+06, compress = TRUE, verbose = FALSE)
fastqFilter(fn = WT2_input2_raw, fout = paste(wd_fastq_trim, "WT2_input2.fastq.gz", sep = ""), truncLen = 0, trimLeft = 0, maxN = 0, minQ = 0, rm.phix = TRUE, n = 5e+06, compress = TRUE, verbose = FALSE)
fastqFilter(fn = WT1_H3K9me2_raw, fout = paste(wd_fastq_trim, "WT1_H3K9me2.fastq.gz", sep = ""), truncLen = 0, trimLeft = 0, maxN = 0, minQ = 0, rm.phix = TRUE, n = 5e+06, compress = TRUE, verbose = FALSE)
fastqFilter(fn = WT2_H3K9me2_raw, fout = paste(wd_fastq_trim, "WT2_H3K9me2.fastq.gz", sep = ""), truncLen = 0, trimLeft = 0, maxN = 0, minQ = 0, rm.phix = TRUE, n = 5e+06, compress = TRUE, verbose = FALSE)
```
fastqFilter(fn = WT1_H3K27ac_raw, fout = paste(wd_fastq_trim, "WT1_H3K27ac.fastq.gz", sep = ""), truncLen = 0, trimLeft = 0, maxN = 0, minQ = 0, rm.phix = TRUE, n = 5e+06, compress = TRUE, verbose = FALSE)

fastqFilter(fn = WT2_H3K27ac_raw, fout = paste(wd_fastq_trim, "WT2_H3K27ac.fastq.gz", sep = ""), truncLen = 0, trimLeft = 0, maxN = 0, minQ = 0, rm.phix = TRUE, n = 5e+06, compress = TRUE, verbose = FALSE)

# Making matrices for qAlign step
setwd(wd_fastq_trim)
WT1_input1_matrix <- "WT1_input1_matrix.txt"
WT1_input2_matrix <- "WT1_input2_matrix.txt"
WT2_input1_matrix <- "WT2_input1_matrix.txt"
WT2_input2_matrix <- "WT2_input2_matrix.txt"
WT1_H3K9me2_matrix <- "WT1_H3K9me2_matrix.txt"
WT2_H3K9me2_matrix <- "WT2_H3K9me2_matrix.txt"
WT1_H3K27ac_matrix <- "WT1_H3K27ac_matrix.txt"
WT2_H3K27ac_matrix <- "WT2_H3K27ac_matrix.txt"

WT1_input1_matrixfile <- matrix(c("FileName", "WT1_input1.fastq.gz", "SampleName", "WT1 input 1"), nrow = 2, ncol = 2)
WT1_input2_matrixfile <- matrix(c("FileName", "WT1_input2.fastq.gz", "SampleName", "WT1 input 2"), nrow = 2, ncol = 2)
WT2_input1_matrixfile <- matrix(c("FileName", "WT2_input1.fastq.gz", "SampleName", "WT2 input 1"), nrow = 2, ncol = 2)
WT2_input2_matrixfile <- matrix(c("FileName", "WT2_input2.fastq.gz", "SampleName", "WT2 input 2"), nrow = 2, ncol = 2)
WT1_H3K9me2_matrixfile <- matrix(c("FileName", "WT1_H3K9me2.fastq.gz", "SampleName", "WT1 H3K9me2"), nrow = 2, ncol = 2)
WT2_H3K9me2_matrixfile <- matrix(c("FileName", "WT2_H3K9me2.fastq.gz", "SampleName", "WT2 H3K9me2"), nrow = 2, ncol = 2)
WT1_H3K27ac_matrixfile <- matrix(c("FileName", "WT1_H3K27ac.fastq.gz", "SampleName", "WT1 H3K27ac"), nrow = 2, ncol = 2)
WT2_H3K27ac_matrixfile <- matrix(c("FileName", "WT2_H3K27ac.fastq.gz", "SampleName", "WT2 H3K27ac"), nrow = 2, ncol = 2)

write.table(WT1_input1_matrixfile, "WT1_input1_matrix.txt", sep="\t", row.names = FALSE, col.names = FALSE)
write.table(WT1_input2_matrixfile, "WT1_input2_matrix.txt", sep="\t", row.names = FALSE, col.names = FALSE)
write.table(WT2_input1_matrixfile, "WT2_input1_matrix.txt", sep="\t", row.names = FALSE, col.names = FALSE)
write.table(WT2_input2_matrixfile, "WT2_input2_matrix.txt", sep="\t", row.names = FALSE, col.names = FALSE)
write.table(WT1_H3K9me2_matrixfile, "WT1_H3K9me2_matrix.txt", sep="\t", row.names = FALSE, col.names = FALSE)
write.table(WT2_H3K9me2_matrixfile, "WT2_H3K9me2_matrix.txt", sep="\t", row.names = FALSE, col.names = FALSE)
write.table(WT1_H3K27ac_matrixfile, "WT1_H3K27ac_matrix.txt", sep="\t", row.names = FALSE, col.names = FALSE)
write.table(WT2_H3K27ac_matrixfile, "WT2_H3K27ac_matrix.txt", sep="\t", row.names = FALSE, col.names = FALSE)
#Aligning reads using Bowtie, accessed by the qAlign wrapper
#Reads that map to duplicate regions are discarded

no_cores <- detectCores() - 1
cl <- makeCluster(no_cores)  # Bowtie uses all but 1 CPU core

WT1_input1_align <- qAlign(sampleFile = WT1_input1_matrix, genome = "BSgenome.Mmусculus.UCSC.mm9",
                           maxHits = 1, paired = NULL, alignmentsDir = wd_bam, clObj = cl)
WT1_input2_align <- qAlign(sampleFile = WT1_input2_matrix, genome = "BSgenome.Mmусculus.UCSC.mm9",
                           maxHits = 1, paired = NULL, alignmentsDir = wd_bam, clObj = cl)
WT2_input1_align <- qAlign(sampleFile = WT2_input1_matrix, genome = "BSgenome.Mmусculus.UCSC.mm9",
                           maxHits = 1, paired = NULL, alignmentsDir = wd_bam, clObj = cl)
WT2_input2_align <- qAlign(sampleFile = WT2_input2_matrix, genome = "BSgenome.Mmусculus.UCSC.mm9",
                           maxHits = 1, paired = NULL, alignmentsDir = wd_bam, clObj = cl)
WT1_H3K9me2_align <- qAlign(sampleFile = WT1_H3K9me2_matrix, genome = "BSgenome.Mmусculus.UCSC.mm9",
                           maxHits = 1, paired = NULL, alignmentsDir = wd_bam, clObj = cl)
WT2_H3K9me2_align <- qAlign(sampleFile = WT2_H3K9me2_matrix, genome = "BSgenome.Mmусculus.UCSC.mm9",
                           maxHits = 1, paired = NULL, alignmentsDir = wd_bam, clObj = cl)
WT1_H3K27ac_align <- qAlign(sampleFile = WT1_H3K27ac_matrix, genome = "BSgenome.Mmусculus.UCSC.mm9",
                           maxHits = 1, paired = NULL, alignmentsDir = wd_bam, clObj = cl)
WT2_H3K27ac_align <- qAlign(sampleFile = WT2_H3K27ac_matrix, genome = "BSgenome.Mmусculus.UCSC.mm9",
                           maxHits = 1, paired = NULL, alignmentsDir = wd_bam, clObj = cl)

# Workspace cleanup
stopCluster(cl)
file.remove(WT1_input1_matrix)
file.remove(WT1_input2_matrix)
file.remove(WT2_input1_matrix)
file.remove(WT2_input2_matrix)
file.remove(WT1_H3K9me2_matrix)
file.remove(WT2_H3K9me2_matrix)
file.remove(WT1_H3K27ac_matrix)
file.remove(WT2_H3K27ac_matrix)

# Generating quality check reports
setwd(wd_misc)
qQCReport(WT1_input1_align, pdfFilename = "WT1_input1_QCReport.pdf")
qQCReport(WT1_input2_align, pdfFilename = "WT1_input2_QCReport.pdf")
qQCReport(WT2_input1_align, pdfFilename = "WT2_input1_QCReport.pdf")
qQCReport(WT2_input2_align, pdfFilename = "WT2_input2_QCReport.pdf")
qQCReport(WT1_H3K9me2_align, pdfFilename = "WT1_H3K9me2_QCReport.pdf")
qQCReport(WT2_H3K9me2_align, pdfFilename = "WT2_H3K9me2_QCReport.pdf")
qQCReport(WT1_H3K27ac_align, pdfFilename = "WT1_H3K27ac_QCReport.pdf")
qQCReport(WT2_H3K27ac_align, pdfFilename = "WT2_H3K27ac_QCReport.pdf")
# Renaming output files for easier access later

```r
setwd(wd_bam)

file.rename(list.files(pattern = paste("WT1_input1", "(.*).bam\$", sep = "")), to = "WT1_input1.bam")
file.rename(list.files(pattern = paste("WT1_input1", "(.*).bam.bai\$", sep = "")), to = "WT1_input1.bam.bai")
file.rename(list.files(pattern = paste("WT1_input1", "(.*).bam.txt\$", sep = "")), to = "WT1_input1.bam.txt")

file.rename(list.files(pattern = paste("WT1_input2", "(.*).bam\$", sep = "")), to = "WT1_input2.bam")
file.rename(list.files(pattern = paste("WT1_input2", "(.*).bam.bai\$", sep = "")), to = "WT1_input2.bam.bai")
file.rename(list.files(pattern = paste("WT1_input2", "(.*).bam.txt\$", sep = "")), to = "WT1_input2.bam.txt")

file.rename(list.files(pattern = paste("WT2_input1", "(.*).bam\$", sep = "")), to = "WT2_input1.bam")
file.rename(list.files(pattern = paste("WT2_input1", "(.*).bam.bai\$", sep = "")), to = "WT2_input1.bam.bai")
file.rename(list.files(pattern = paste("WT2_input1", "(.*).bam.txt\$", sep = "")), to = "WT2_input1.bam.txt")

file.rename(list.files(pattern = paste("WT2_input2", "(.*).bam\$", sep = "")), to = "WT2_input2.bam")
file.rename(list.files(pattern = paste("WT2_input2", "(.*).bam.bai\$", sep = "")), to = "WT2_input2.bam.bai")
file.rename(list.files(pattern = paste("WT2_input2", "(.*).bam.txt\$", sep = "")), to = "WT2_input2.bam.txt")

file.rename(list.files(pattern = paste("WT1_H3K9me2", "(.*).bam\$", sep = "")), to = "WT1_H3K9me2.bam")
file.rename(list.files(pattern = paste("WT1_H3K9me2", "(.*).bam.bai\$", sep = "")), to = "WT1_H3K9me2.bam.bai")
file.rename(list.files(pattern = paste("WT1_H3K9me2", "(.*).bam.txt\$", sep = "")), to = "WT1_H3K9me2.bam.txt")

file.rename(list.files(pattern = paste("WT2_H3K9me2", "(.*).bam\$", sep = "")), to = "WT2_H3K9me2.bam")
file.rename(list.files(pattern = paste("WT2_H3K9me2", "(.*).bam.bai\$", sep = "")), to = "WT2_H3K9me2.bam.bai")
file.rename(list.files(pattern = paste("WT2_H3K9me2", "(.*).bam.txt\$", sep = "")), to = "WT2_H3K9me2.bam.txt")

file.rename(list.files(pattern = paste("WT1_H3K27ac", "(.*).bam\$", sep = "")), to = "WT1_H3K27ac.bam")
file.rename(list.files(pattern = paste("WT1_H3K27ac", "(.*).bam.bai\$", sep = "")), to = "WT1_H3K27ac.bam.bai")
file.rename(list.files(pattern = paste("WT1_H3K27ac", "(.*).bam.txt\$", sep = "")), to = "WT1_H3K27ac.bam.txt")

file.rename(list.files(pattern = paste("WT2_H3K27ac", "(.*).bam\$", sep = "")), to = "WT2_H3K27ac.bam")
file.rename(list.files(pattern = paste("WT2_H3K27ac", "(.*).bam.bai\$", sep = "")), to = "WT2_H3K27ac.bam.bai")
file.rename(list.files(pattern = paste("WT2_H3K27ac", "(.*).bam.txt\$", sep = "")), to = "WT2_H3K27ac.bam.txt")
```

154
# Merging .bam files

# DamID-seq .bam files are also processed here

# DamID-seq .bam files are unsorted, require an additional sorting step

```r
setwd(wd_bam)
mergeBam(c("WT1_input1.bam", "WT1_input2.bam",  
            "WT2_input1.bam",  
            "WT2_input2.bam"), destination = "WT_input.bam")
mergeBam(c("WT1_H3K9me2.bam", "WT2_H3K9me2.bam"), destination =  
            "WT_H3K9me2.bam")
mergeBam(c("WT1_H3K27ac.bam", "WT2_H3K27ac.bam"), destination =  
            "WT_H3K27ac.bam")
mergeBam(c("WT1_DO1.bam", "WT1_DO2.bam", "WT1_DO3.bam", "WT2_DO1.bam",  
            "WT2_DO2.bam", "WT2_DO3.bam"), destination = "WT_DO.bam")
mergeBam(c("WT1_DL1.bam", "WT1_DL2.bam", "WT1_DL3.bam", "WT2_DL1.bam",  
            "WT2_DL2.bam", "WT2_DL3.bam"), destination = "WT_DL.bam")
sortBam("WT_DO.bam", destination = "WT_DO_sorted")
sortBam("WT_DL.bam", destination = "WT_DL_sorted")
```

# Part 3: Assigning reads into bins

# An intermediate step required to process the .bam files into a format  
# more suitable for downstream analysis.

# Assigning variables

# Reads falling into unmapped parts of the chromosome or to chromosome Y  
# and M are excluded

```r
fragment_length <- 150
bin200 <- 200
bin10kb <- 10000
excludeChr <- c("chrY", "chrM", "chr1_random", "chr2_random",  
                "chr3_random", "chr4_random", "chr5_random", "chr6_random",  
                "chr7_random", "chr8_random", "chr9_random", "chr10_random",  
                "chr11_random", "chr12_random", "chr13_random", "chr14_random",  
                "chr15_random", "chr16_random", "chr17_random", "chr18_random",  
                "chr19_random", "chrX_random", "chrY_random", "chrM_random",  
                "chrUn_random")
```

# Assigning reads into 200 bp bins

```r
setwd(wd_bam)
constructBins(infile = "WT_input.bam", fileFormat = "bam", outfileLoc =  
              wd_bin200, byChr = FALSE, excludeChr = excludeChr, PET = FALSE,  
              fragLen = fragment_length, binSize = bin200, capping = 50)
constructBins(infile = "WT_H3K9me2.bam", fileFormat = "bam", outfileLoc  
              = wd_bin200, byChr = FALSE, excludeChr = excludeChr, PET = FALSE,  
              fragLen = fragment_length, binSize = bin200, capping = 50)
constructBins(infile = "WT_H3K27ac.bam", fileFormat = "bam", outfileLoc  
              = wd_bin200, byChr = FALSE, excludeChr = excludeChr, PET = FALSE,  
              fragLen = fragment_length, binSize = bin200, capping = 50)
```

# Renaming output files for easier access later

```r
setwd(wd_bin200)
```
file.rename(list.files(pattern = paste("WT_input", "(.*)\.(txt$", sep = "")), to = "WT_input_200.txt")
file.rename(list.files(pattern = paste("WT_H3K9me2", "(.*)\.(txt$", sep = "")), to = "WT_H3K9me2_200.txt")
file.rename(list.files(pattern = paste("WT_H3K27ac", "(.*)\.(txt$", sep = "")), to = "WT_H3K27ac_200.txt")

#Assigning reads into 10 kb bins
setwd(wd_bam)
constructBins(infile = "WT_input.bam", fileFormat = "bam", outfileLoc = wd_bin10kb, byChr = FALSE, excludeChr = excludeChr, PET = FALSE, fragLen = fragment_length, binSize = bin10kb, capping = 50)
constructBins(infile = "WT_H3K9me2.bam", fileFormat = "bam", outfileLoc = wd_bin10kb, byChr = FALSE, excludeChr = excludeChr, PET = FALSE, fragLen = fragment_length, binSize = bin10kb, capping = 50)
constructBins(infile = "WT_DO_sorted.bam", fileFormat = "bam", outfileLoc = wd_bin10kb, byChr = FALSE, excludeChr = excludeChr, PET = FALSE, fragLen = fragment_length, binSize = bin10kb, capping = 50)
constructBins(infile = "WT_DL_sorted.bam", fileFormat = "bam", outfileLoc = wd_bin10kb, byChr = FALSE, excludeChr = excludeChr, PET = FALSE, fragLen = fragment_length, binSize = bin10kb, capping = 50)

#Renaming output files for easier access later
setwd(wd_bin10kb)
file.rename(list.files(pattern = paste("WT_input", "(.*)\.(txt$", sep = "")), to = "WT_input_10kb.txt")
file.rename(list.files(pattern = paste("WT_H3K9me2", "(.*)\.(txt$", sep = "")), to = "WT_H3K9me2_10kb.txt")
file.rename(list.files(pattern = paste("WT_DO", "(.*)\.(txt$", sep = "")), to = "WT_DO_10kb.txt")
file.rename(list.files(pattern = paste("WT_DL", "(.*)\.(txt$", sep = "")), to = "WT_DL_10kb.txt")

#Part 4: Generating .bedgraph tracks

#.bedgraph tracks are uploaded onto the UCSC browser for viewing

#Creating H3K9me2 .bedgraph track
setwd(wd_bin200)
WT_input <- read.table("WT_input_200.txt", header = FALSE, skip = 1)
WT_H3K9me2 <- read.table("WT_H3K9me2_200.txt", header = FALSE, skip = 1)

#Setting up chromosomal location columns
chr <- WT_input[,1]
chrstart <- WT_input[,2]
chrend <- chrstart + 199
chrend[985978] <- 197195431
chrend[1894719] <- 181748086
chrend[2692718] <- 159599782
chrend[3470869] <- 155630119
chrend[4233556] <- 152537258
chrend[4981142] <- 149517036
chrend[5743765] <- 152524552
chrend[6402460] <- 131738870
chrend[7022841] <- 129993254
chrend[7672808] <- 121843855
chrend[8282028] <- 121257529
chrend[8888316] <- 120284311
chrend[949738] <- 125194863
chrend[1003188] <- 103494973
chrend[11124784] <- 98319149
chrend[11601148] <- 95272650
chrend[12055009] <- 90772030
chrend[12361722] <- 61342429
chrend[13194974] <- 166650295

#Normalizing differences between input and H3K9me2 read counts
df <- data.frame(WT_input[,3], WT_H3K9me2[,3])
norm <- sum(df[,2]) / sum(df[,1])
df[,2] <- df[,2] / norm

#Adding 1 to windows with no reads
df[,1] <- sapply(df[,1], FUN = function(x) ifelse(x == 0, 1, x))
df[,2] <- sapply(df[,2], FUN = function(x) ifelse(x == 0, 1, x))

#Creating normalized tracks
df <- log(df) / log (2)
WT <- df[,2] - df[,1]

#Removing deleted regions
WT[4440974:4441022] <- WT[4440974:4441022] * 0
us[4440832:4440971] <- us[4440832:4440971] * 0 #To delete usRC and
#Eb[4441067:4441083] <- Eb[4441067:4441083] * 0 #Eb in other samples

#Assembling .bedgraph
setwd(wd_bed)
WT <- data.frame(chr, chrstart, chrend, WT)
WT <- WT[WT[,4] != 0,]
write.table(WT, "WT_H3K9me2.bedgraph", quote = FALSE, sep="\t",
col.names = FALSE, row.names = FALSE)

#Creating H3K27ac .bedgraph track
setwd(wd_bin200)
WT_H3K27ac <- read.table("WT_H3K27ac_200.txt", header = FALSE, skip = 1)

#Normalizing to reads per million
WT_H3K27ac <- WT_H3K27ac[,3] * 10^6 / sum(WT_H3K27ac[,3])

# Assembling .bedgraph
setwd(wd_bed)
WT_H3K27ac <- data.frame(chr, chrstart, chrend, WT_H3K27ac)
WT_H3K27ac <- WT_H3K27ac[WT_H3K27ac[,4] > 0,]
write.table(WT_H3K27ac, "WT_H3K27ac.bedgraph", quote = FALSE, sep = "\t",
col.names = FALSE, row.names = FALSE)

# Adding headers to .bedgraph files - easier to perform by opening the .bedgraph files and adding the header as the first line
track type=bedGraph name="WT H3K9me2" description="WT H3K9me2"
visibility=full color=0,0,0 autoScale=off alwaysZero=on viewLimits=-3:3 yLineMark=0 yLineOnOff=On windowingFunction=mean
track type=bedGraph name="WT H3K27ac" description="WT H3K27ac"
visibility=full color=0,0,0 autoScale=off alwaysZero=on viewLimits=0:5 yLineMark=0 yLineOnOff=On windowingFunction=mean

# Part 5: H3K9me2/DamID correlation analysis

# The part performs a pairwise correlation analysis between H3K9me2 and DamID, with reads assigned into non-overlapping 10 kb bins
setwd(wd_bin10kb)
WT_DO <- read.table("WT_DO_10kb.txt", header = FALSE, skip = 1)
WT_DL <- read.table("WT_DL_10kb.txt", header = FALSE, skip = 1)
WT_input <- read.table("WT_input_10kb.txt", header = FALSE, skip = 1)
WT_H3K9me2 <- read.table("WT_H3K9me2_10kb.txt", header = FALSE, skip = 1)

# Normalizing differences between read counts
df <- data.frame(WT_DO[,3], WT_DL[,3], WT_input[,3], WT_H3K9me2[,3])
norm1 <- sum(df[,2]) / sum(df[,1])
norm2 <- sum(df[,4]) / sum(df[,3])
df[,2] <- df[,2] / norm1

# Adding 1 to windows with no reads
df[,1] <- sapply(df[,1], FUN = function(x) ifelse(x == 0, 1, x))
df[,2] <- sapply(df[,2], FUN = function(x) ifelse(x == 0, 1, x))
df[,3] <- sapply(df[,3], FUN = function(x) ifelse(x == 0, 1, x))
df[,4] <- sapply(df[,4], FUN = function(x) ifelse(x == 0, 1, x))

# Creating normalized tracks
df <- log(df) / log(2)
WT_DamID <- df[,2] - df[,1]
WT_H3K9me2 <- df[,4] - df[,3]

# Correlation graph
plot(WT_DamID, WT_H3K9me2, xlab = "DamID", ylab = "H3K9me2", pch = 20,
cex = 0.5)
abline(h = 0)
abline(v = 0)
# Correlation testing

```r
WT_cor <- data.frame(WT_DamID, WT_H3K9me2)
cor(WT_cor, use = "complete.obs", method = "pearson")
cor.test(WT_DamID, WT_H3K9me2, alternative = "two.sided", method = "pearson")
```

# Part 6: Graphing features around LAD borders

# This part identifies the 100 kb region surrounding LAD borders and graphs the enrichment of features around the LAD borders

```r
setwd(wd_bed)
LAD <- read.table("WT_DamID_LAD.bed", header = FALSE, skip = 1)
LAD[,3] <- LAD[,2] + 50000
LAD[,2] <- LAD[,2] - 50000
write.table(LAD, "WT_DamID_LAD_border_50kb.bed", quote = FALSE,
            sep="\t", col.names = FALSE, row.names = FALSE)
LAD <- readGeneric("WT_DamID_LAD_border_50kb.bed", header = FALSE, skip = 0)
DamID <- readGeneric("WT_DamID.bedgraph", header = FALSE,
                     keep.all.metadata = TRUE, skip = 1)
score_DamID <- ScoreMatrix(target = DamID, windows = LAD, weight.col = "X")
plotMeta(score_DamID, xcoords = c(-50000,50000), lwd = 2)
```
References


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


Biography

Shiwei Chen was born in Singapore on April 13th, 1987. Shiwei completed his undergraduate studies at Imperial College London (United Kingdom) and graduated with a Bachelor of Science in Biology with First Class Honours in 2011. After graduating from Imperial College London, Shiwei went to Duke University to obtain a Ph.D. in Immunology.

Papers published:
