Roles of CTCF and YY1 in T Cell Receptor Gene Rearrangement and T Cell Development

by

Liang Chen

Department of Immunology
Duke University

Date:_______________________
Approved:

____________________________________
Michael S. Krangel, Supervisor

____________________________________
Yuan Zhuang, Chair

____________________________________
Richard Lee Reinhardt

____________________________________
Qijing Li

____________________________________
Kenneth Kreuzer

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

2016
ABSTRACT

Roles of CTCF and YY1 in T Cell Receptor Gene Rearrangement and T Cell Development
by
Liang Chen

Department of Immunology
Duke University

Date:________________________________________

Approved:

Michael S. Krangel, Supervisor

Yuan Zhuang, Chair

Richard Lee Reinhardt

Qijing Li

Kenneth Kreuzer

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

2016
Abstract

Diversity of T cell receptors (TCR) and immunoglobulins (Ig) is generated by V(D)J recombination of antigen receptor (AgR) loci. The Tcra-Tcrd locus is of particular interest because it displays a nested organization of Tcrd and Tcra gene segments and V(D)J recombination follows an intricate developmental program to assemble both TCRδ and TCRα repertoires. However, the mechanisms that dictate the developmental regulation of V(D)J recombination of the Tcra-Tcrd locus remain unclear.

We have previously shown that CCCTC-binding factor (CTCF) regulates Tcra gene transcription and rearrangement through organizing chromatin looping between CTCF-binding elements (CBEs). This study is one of many showing that CTCF functions as a chromatin organizer and transcriptional regulator genome-wide. However, detailed understanding of the impact of specific CBEs is needed to fully comprehend the biological function of CTCF and how CTCF influences the generation of the TCR repertoire during thymocyte development. Thus, we generated several mouse models with genetically modified CBEs to gain insight into the CTCF-dependent regulation of the Tcra-Tcrd locus. We revealed a CTCF-dependent chromatin interaction network at the Tcra-Tcrd locus in double-negative thymocytes. Disruption of a discrete chromatin loop encompassing Dδ, Jδ and Cδ gene segments allowed a single Vδ segment to frequently contact and rearrange to diversity and joining gene segments and dominate the adult TCRδ repertoire. Disruption of this loop also narrowed the TCRα repertoire, which, we believe, followed as a consequence of the restricted TCRδ repertoire. Hence, a single
CTCF-mediated chromatin loop directly regulates TCRδ diversity and indirectly regulates TCRα diversity. In addition, we showed that insertion of an ectopic CBE can modify chromatin interactions and disrupt the rearrangement of particular Vδ gene segments. Finally, we investigated the role of YY1 in early T cell development by conditionally deleting YY1 in developing thymocytes. We found that early ablation of YY1 caused severe developmental defects in the DN compartment due to a dramatic increase in DN thymocyte apoptosis. Furthermore, late ablation of YY1 resulted in increased apoptosis of DP thymocytes and a restricted TCRα repertoire. Mechanistically, we showed that p53 was upregulated in both DN and DP YY1-deficient thymocytes. Eliminating p53 in YY1-deficient thymocytes rescued the survival and developmental defects, indicating that these YY1-dependent defects were p53-mediated. We conclude that YY1 is required to maintain cell viability during thymocyte development by thwarting the accumulation of p53.

Overall, this thesis work has shown that CTCF-dependent looping provides a central framework for lineage- and developmental stage-specific regulation of Tcra-Tcrd gene expression and rearrangements. In addition, we identified YY1 as a novel regulator of thymocyte viability.
Contents

Abstract ........................................................................................................................................ iv

List of Tables ................................................................................................................................ ix

List of Figures ................................................................................................................................ x

Acknowledgements .................................................................................................................. xii

1. Introduction .................................................................................................................................. 1

  1.1 V(D)J recombination .......................................................................................... 1

  1.2 Lymphocyte development ................................................................................. 4

  1.3 Structure of AgR loci .......................................................................................... 7

  1.4 Regulation of V(D)J recombination ................................................................... 11

    1.4.1 Expression of RAG ................................................................................... 11

    1.4.2 Binding and activity of RAG .................................................................. 12

    1.4.3 Locus accessibility ................................................................................... 14

  1.5 Ordered rearrangement and allelic exclusion ..................................................... 22

  1.6 Subnuclear localization ....................................................................................... 26

  1.7 Locus conformation ............................................................................................... 28

    1.7.1 CTCF ......................................................................................................... 31

    1.7.2 V(D)J recombination is regulated by the conformation of AgR loci .......... 33

2. Specific Aims ............................................................................................................................ 45

3. Material and Methods .............................................................................................................. 48

  3.1 Mice ................................................................................................................................. 48

    3.1.1 Generation of INT1-2KO, INT2M and CBE KI mice ......................... 48

    3.1.2 Other mouse strains ..................................................................................... 49
5 An ectopic intragenic CBE disturbs Vδ usage at the Tcra-Tcrd locus .........................92
   5.1 Introduction ........................................................................................................92
   5.2 Tcrd, but not Tcra rearrangement is defective in CBE KI mice..............................93
   5.3 The CBE KI does not perturb chromatin accessibility .........................................99
   5.4 CBE KI mice have an altered chromatin loop organization ...............................100
   5.5 Conclusion .......................................................................................................103
5 The role of YY1 in thymocyte development ..............................................................104
   6.1 Introduction .......................................................................................................104
   6.2 Early ablation of YY1 severely blocks DN thymocyte development ..............105
   6.3 YY1 is required for the normal life span of DP thymocytes ...............................114
   6.4 YY1 regulates the p53-dependent apoptosis pathway ......................................121
   6.5 Conclusion .......................................................................................................126
7 Discussion and future directions ..............................................................................127
Bibliography ............................................................................................................139
Biography ................................................................................................................165
List of Tables

Table 1: Primers used for 3C........................................................................................................53
Table 2: Primers used for Tcrd germline transcription and mRNA expression.............54
Table 3: Primers used for Tcr rearrangement........................................................................56
Table 4: Primers used for Chromosomal DNA retention.......................................................58
Table 5: Primers used for ChIP.................................................................................................60
List of Figures

Figure 1: Overview of V(D)J recombination ................................................................. 4
Figure 2: Structure and organization of Ig and Tcr loci .................................................. 9
Figure 3: Long-range interaction network within the Igh locus ...................................... 41
Figure 4: Long-range interaction network within the Tcra-Tcrd locus .......................... 70
Figure 5: Generation of INT1-2KO mice .................................................................... 72
Figure 6: Thymocyte development in INT1-2KO mice .................................................. 73
Figure 7: Vδ4 and Vδ6.3 usage in γδ T cells from INT1-2KO mice ............................... 74
Figure 8: Restricted TCRδ repertoire in INT1-2-deficient mice .................................... 80
Figure 9: INT1-2-deletion alters chromatin conformation but not accessibility .......... 84
Figure 10: Generation of INT2 mutant (INT2M) mice .................................................. 86
Figure 11: Partial redundancy between INT1 and INT2 .............................................. 87
Figure 12: Restricted TCRα repertoire in INT1-2KO mice ......................................... 90
Figure 13: Generation of CBE KI mice ...................................................................... 96
Figure 14: Tcrd rearrangement is disturbed in CBE KI mice ........................................ 98
Figure 15: Tcra rearrangement is normal in CBE KI mice .......................................... 99
Figure 16: CBE KI alters chromatin conformation but not accessibility .................... 103
Figure 17: Early ablation of Yy1 severely blocks T cell development ......................... 108
Figure 18: β-selection in YY1-deficient mice ............................................................... 111
Figure 19: Increased cell death in YY1-deficient DN thymocytes ............................... 113
Figure 20: YY1 regulates the survival of DP thymocytes ............................................ 116
Figure 21: YY1 regulates the TCRα repertoire ............................................................ 118
Figure 22: YY1 regulates DP thymocyte survival independent of V(D)J recombination or TCR expression ................................................................. 120
Figure 23. YY1 negatively regulates p53. .................................................................122

Figure 24. Absence of p53 rescues the defect in YY1CD2 mice. ..............................126
Acknowledgements

W would like to thank C. Bock for the production of mice with gene targeting, L. Martinek, N. Martin, B. Li and M. Cook of the Duke Cancer Institute Flow Cytometry Shared Resource for help with cell sorting and analysis, S. Langdon of the DNA Analysis Facility for help with DNA sequencing, Z. Huang and A. Byrd for technical support, P-Y Yang, J. Liang and H.-Y. Shih for technical advice, Y.-W. He, G. Hammer, Y. Zhuang, Q.-J. Li, W. Zhang for reagents, and Y.-W. He, A. Feeney, and J. Rathmell for mouse lines.
1. Introduction

1.1 V(D)J recombination

The fundamental role of the adaptive immune system is to recognize and respond to a universe of foreign antigens. This function is highly dependent on the expression of diverse repertoires of antigen receptors (AgRs) by T and B lymphocytes. Diversity of AgRs is achieved by somatic assembly of Variable (V), Diversity (D) and Joining (J) gene segments, a process termed V(D)J recombination (Fig. 1a) (Schatz and Ji, 2011). The initiation of V(D)J recombination is catalyzed by a recombinase protein complex comprised of recombination activating gene 1 (RAG1) and RAG2 (collectively referred to as RAG) and high-mobility group protein B1 (HMGB1) (Schatz and Swanson, 2011). RAG can recognize recombination signal sequences (RSSs) that flank each V, D and J gene segment. An RSS is comprised of a conserved heptamer (consensus 5’-CACAGTG-3’) and a nonamer (consensus 5’-ACAAAAACC-3’) separated by a spacer of either 12 or 23 bp (termed 12-RSS and 23-RSS, respectively) (Fig. 1b and c). Efficient V(D)J recombination requires pairing of one 12-RSS and one 23-RSS, a constraint known as the 12/23 rule (Hiom and Gellert, 1998). Two RSSs captured by RAG and HMGB1 form a synaptic complex. Within this synaptic complex, RAG creates double-strand breaks (DSBs) between RSSs and coding segments, resulting in pairs of coding ends containing gene segments, and signal ends containing RSSs. Thereafter, these DSBs are processed and repaired by the classical non-homologous end joining (NHEJ) pathway. Signal ends are normally precisely repaired to form signal joints. The hairpin-sealed
coding ends are opened by DSB repair factors Artemis and DNA-PKcs. This process is often asymmetric, leading to the addition of palindromic nucleotides (P nucleotides) at the joint. In addition, joining of coding ends is imprecise, due to exonuclease-mediated deletion of nucleotides, and addition of non-templated nucleotides (N-nucleotides) to coding joints. The latter process is mediated by Terminal deoxynucleotidyl Transferase (TdT). TdT is a template-independent DNA polymerase that adds nucleotides to 3’ ends of DNA (Komori et al., 1993). This error-prone DNA repair process further augments the diversity of Ig and Tcr repertoires (Komori et al., 1996). V(D)J recombination results in deleting or inverting a fragment of chromosomal DNA that ranges from 500 bp to >1 Mb. V(D)J recombination, if mistargeted, can also cause genomic instability and promote hematopoietic malignancy (Teng et al., 2015). Therefore, it is critical for V(D)J recombination to be properly regulated during lymphoid development.
a)

12RSS

23RSS

RAG1
RAG2

HMGB1

Bind a single RSS

Capture pairing RSS Synapsis

Cleave RSSs Hairpin formation

NHEJ DSB repair

Coding joint

Signal joint

b)

Igh

V D J

Igk

V J

Igl, Tcra, Tcrg

V J

Tcrb, Tcrd

V D J

c)

Consensus 12-RSS

NNNNN CACAGTGNNNNNNNNNNNNNNNACA AAAACC

NNNNN GTGTCA CNNNNNNNNNNNNNNNTCTTTTTGG

Coding Heptamer Spacer Nonamer
1.2 Lymphocyte development

B cells develop from bone marrow progenitors and progress through pro-B cell, pre-B cell, and immature B cell stages of development (Hardy and Hayakawa, 2001). B cell development is coupled with ordered V(D)J recombination of the \(Igh\), \(Igk\) and \(Igl\) loci: \(D_H\)-to-\(J_H\) rearrangements occur in pre-pro-B cells, followed by \(V_H\)-to-\(D_HJ_H\) rearrangements in pro-B cells (Hardy and Hayakawa, 2001). Productive \(Igh\) rearrangement generates a \(\mu\) heavy chain, which then associates with the surrogate light chain (VpreB, \(\lambda5\)) to assemble the pre-BCR complex (Geier and Schlissel, 2006). Pre-BCR signaling promotes pro-B-to-pre-B cell differentiation, expansion of pre-B cells, and inhibition of further \(Igh\) gene rearrangement (Geier and Schlissel, 2006). Intact Pre-BCR signaling is critical for immune tolerance. For example, surrogate light-chain-deficient mice produce an elevated amount of serum autoantibodies and pre-B cells from these mice express prototypic autoreactive heavy chains (Keenan et al., 2008). At the following pre-B cell stage, two Ig light chain loci, \(Igk\) and \(Igl\), are activated and rearranged (Geier and Schlissel, 2006). Productive \(Igk\) or \(Igl\) rearrangement produces an Ig light chain,
which can associate with the \( \mu \) heavy chain to form a B cell receptor (BCR). BCR-bearing immature B cells are tested for reactivity against self-antigens. Such reactivity can induce receptor editing, clonal deletion or clonal energy (Melchers, 2015). Cells that can pass this checkpoint migrate out of the bone marrow as transitional B cells before further differentiating into mature naive B cells that are capable of antigen-specific activation.

Both \( \alpha\beta \) and \( \gamma\delta \) T lymphocytes develop from CD4\(^{-}\)CD8\(^{-}\) double negative (DN) thymocytes. DN thymocytes can be subdivided into four stages: DN1 (CD25\(^{-}\)CD44\(^{+}\), which includes pluripotent early thymic precursors), DN2 (CD25\(^{+}\)CD44\(^{+}\)), DN3 (CD25\(^{-}\)CD44\(^{+}\)), and DN4 (CD25\(^{-}\)CD44\(^{-}\)). DN1 and DN2 thymocytes proliferate extensively in a TCR-independent, Notch-dependent manner before progression to the DN3 stage (Yui and Rothenberg, 2014). DN3 cells can be further divided into DN3a and DN3b based on cell size and CD27 expression (Taghon et al., 2006). At the DN3a stage, most cells become quiescent and undergo V(D)J recombination at the \( \text{Tcrg, Tcrd, and Tcrb} \) loci (Capone et al., 1998; Livak et al., 1999). Thymocytes with in-frame rearrangements of \( \text{Tcrg} \) and \( \text{Tcrd} \) express the \( \gamma\delta \) TCR. Cells that successfully rearrange \( \text{Tcrb} \) produce a functional TCR\(\beta\) protein, which can assemble with pT\(\alpha\) and CD3 proteins to form pre-TCRs (Michie and Zuniga-Pflucker, 2002; Yui and Rothenberg, 2014). It has been proposed that \( \gamma\delta \) TCRs and pre-TCRs instruct the precursor T cells to commit to either \( \gamma\delta \) or \( \alpha\beta \) lineages, respectively, due to a difference in signaling strength. The \( \gamma\delta \) TCR generates a strong signal to enforce a \( \gamma\delta \) T cell fate, whereas the pre-TCR induces a weak signal and enforces an \( \alpha\beta \) T cell fate (Hayday and Pennington, 2007; Kreslavsky et al.,
Pre-TCR drives a burst of proliferation and allows cells to progress from DN3a to DN3b in a process called β-selection (Michie and Zuniga-Pflucker, 2002). Only cells that pass the β-selection checkpoint develop through the DN4 and immature single-positive stages into CD4⁺CD8⁺ double-positive (DP) thymocytes, which represents the hallmark of αβ T cell lineage commitment. Failure to assemble a functional pre-TCR complex, as occurs in mice deficient for RAG1, RAG2, pre-Tα or CD3γ, leads to a severe block of αβ T cell development at the DN3 stage (Haks et al., 1999; Jiang et al., 1996; Michie and Zuniga-Pflucker, 2002; Mombaerts et al., 1992). Signals that rescue thymocytes from death and promote their proliferation are critical for β-selection. Known trophic signals for thymocytes at the β-selection checkpoint include those generated by the pre-TCR, Notch and the IL-7 receptor (Boudil et al., 2015; Ciofani and Zuniga-Pflucker, 2005). Notch promotes thymocyte survival by regulating glucose metabolism (Ciofani and Zuniga-Pflucker, 2005). The pro-apoptotic factor p53 has been suggested to eliminate thymocytes that fail to pass β-selection, because the concurrent loss of p53 can rescue developmental defects in pre-TCR-deficient mice (Guidos et al., 1996; Haks et al., 1999; Jiang et al., 1996). Cell proliferation during this transition downregulates RAG expression through cell cycle-dependent protein degradation (Li et al., 1996; Taghon et al., 2006). RAG is then re-expressed to mediate Tcra rearrangement in quiescent DP thymocytes. Throughout their 3-4 days of lifespan, DP thymocytes undergo multiple rounds of Va-to-Jα rearrangement, with Jα gene segments used sequentially from the 5’ end to the 3’ end of the Jα array (Carico and Krangel, 2015). Because the lifespan of DP
thymocytes impacts the progression of Vα-to-Jα rearrangements and positive selection of T cells, factors that regulate the survival of DP thymocytes (e.g., RORγ an orphan nuclear receptor, and Bcl-xL, an anti-apoptotic Bcl-2 family protein) are essential regulators of TCR repertoire diversity (Sun et al., 2000). Depleting RORγ shortens the lifespan of DP thymocytes and limits Vα-to-Jα rearrangements to the most 5’ Jα gene segments, whereas extending the lifespan of DP thymocytes with a Bcl-xL transgene skews Vα-to-Jα rearrangement towards 3’ Jα gene segments (Guo et al., 2002).

Thymocytes that have productive Tcra rearrangements assemble TCRs. These cells further undergo positive selection, which selects for weak reactivity with self peptide:MHC complexes, and negative selection, which selects against strong reactivity with self peptide:MHC complex.

1.3 Structure of AgR loci

V, D and J gene segments are arrayed along the length of three T cell receptor (Tcr) loci (Tcra-Tcrd, Tcrb, Tcrg) and three immunoglobulin (Ig) loci (Igh, Igk, and Igλ).

The mouse Igh locus spans about 3 Mb on chromosome 12. Approximately 150 V<sub>H</sub> gene segments are distributed over 2.5 Mb at the 5’ portion of the locus (Fig. 2). V<sub>H</sub> gene segments are divided into distal (V<sub>H</sub>358 and V<sub>H</sub>3609), intermediate (V<sub>H</sub>S107) and proximal (V<sub>H</sub>Q52 and V<sub>H</sub>7183) gene families. The most 3’ V<sub>H</sub> gene segment, V<sub>H</sub>81X, recombines at high frequency in fetal liver-derived pre-B cells, although it is designated as a pseudogene. The D<sub>H</sub> region consists of 10-15 D<sub>H</sub> gene segments, including DFL16.1, intermediate DSPs and DQ52. The first J<sub>H</sub> gene segment is located less than 1 kb
downstream of DQ52, followed by three other J<sub>H</sub> gene segments and the exons of C<sub>μ</sub> and C<sub>δ</sub>. Other C<sub>H</sub> exons extending 140 kb downstream of C<sub>δ</sub> include C<sub>μ</sub>, C<sub>δ</sub>, C<sub>γ3</sub>, C<sub>γ1</sub>, C<sub>γ2b</sub>, C<sub>γ2a</sub>, C<sub>ε</sub> and C<sub>α</sub> (Perlot et al., 2005).

The mouse Igk locus spreads over 3 Mb on chromosome 6. It consists of approximately 140 functional V<sub>κ</sub> gene segments within a 3.2-Mb region, followed by a 30-kb J<sub>κ</sub>-C<sub>κ</sub> region that contains 4 functional J<sub>κ</sub> gene segments and 1 C<sub>κ</sub> gene (Ribeiro de Almeida et al., 2015). A unique feature of the Igk locus is its ability to undergo secondary rearrangement by joining a germline V<sub>κ</sub> gene segment with a J<sub>κ</sub> gene segment located downstream of an existing V<sub>κ</sub>J<sub>κ</sub> exon (Feddersen and Van Ness, 1985). This process, termed receptor editing, replaces the existing Igk chain of a pre-B cell in an attempt to produce a functional and non-autoreactive BCR (Nemazee, 2006).

The Tcrb locus spans approximately 700 kb. The V<sub>β</sub> array spans approximately 670 Mb and comprises of 20 functional V<sub>β</sub> gene segments (Fig. 2). A 150-kb region that separate the Trbv1 from the main V<sub>β</sub> cluster contains several non-functional trypsinogen genes. Another 13 trypsinogen genes are located within the interval region between the V<sub>β</sub> and D<sub>β</sub>J<sub>β</sub> clusters. Downstream of these trypsinogen genes are two D<sub>β</sub>-J<sub>β</sub>-C<sub>β</sub> clusters. A single V<sub>β</sub> gene segment (Trbv31) lies downstream of the D<sub>β</sub>-J<sub>β</sub> segments. The orientation of Trbv31 is opposite to other Tcrb segments, forcing it to rearrange by inversion (Cobb et al., 2006).
The mouse Tcrg locus spans 200 kb on chromosome 13. This locus contains three functional clusters of Vγ, Jγ, and Cγ gene segments (Cγ1, Cγ2 and Cγ4) (Fig. 2). The Cγ2 and Cγ4 clusters contain only single Vγ and Jγ gene segments, whereas the Cγ1 cluster contains four Vγ gene segments (Vγ5, Vγ2, Vγ4, Vγ3) (Xiong and Raulert, 2007).

Usage of Vγ gene segments from the Cγ1 cluster is developmentally programmed and tissue-specific. For example, the Jγ1-proximal Vγ3 and Vγ4 gene segments are exclusively rearranged in fetal thymocytes. Both Vγ3- and Vγ4-to-Jγ1Cγ1 rearrangements are preferentially coupled with Trdv4(Vδ1)-to-Dδ2Jδ2 rearrangements to
generate invariant $\gamma\delta$ TCRs. $V_\gamma 3$ is exclusively used by the TCRs of dendritic epidermal T cells (DETCs) (Havran and Allison, 1990), whereas $V_\gamma 4$-containing $\gamma\delta$ T cells are normally found in the tongue, lung, and vaginal epithelium (Itohara et al., 1990). The other $V_\gamma$ gene segments within the $C_\gamma 1$ cluster, $V_\gamma 2$ and $V_\gamma 5$, along with $V_\gamma 1.1$ in the $C_\gamma 4$ cluster and $V_\gamma 1.2$ in the $C_\gamma 2$ cluster, rearrange in adult thymus. The $\gamma\delta$ T cells using these $V_\gamma$ gene segments colonize the secondary lymphoid organs (Krangel, 2009).

The $Tcra-Tcrd$ locus of the 129/SvJ mouse strain spans approximately 1.5 Mb on chromosome 14. This locus displays a complex organization of gene segments and an intricate developmental program of V(D)J recombination that leads to the development of both $\gamma\delta$ and $\alpha\beta$ T lymphocytes. Approximately 104 V gene segments are distributed across 1.4 Mb at the 5' portion of the locus, with 16 of them designated as pseudogenes that either cannot rearrange or generate an out-of-frame exon (Fig. 2). To add to its complexity, the V-array contains a 400 kb duplication (denoted as “d”) of V gene segments. The 1.4 Mb V-array is followed by $D_\delta$, $J_\delta$, $C_\delta$, $J_\alpha$ and $C_\alpha$ gene segments clustered within the final 0.1 Mb of the locus. A subset of the V gene segments, including $Trdv1$, $Trdv2-2$ ($V_\delta 4$), $Trdv4$ ($V_\delta 1$) and $Trdv5$ ($V_\delta 5$), which is in inverted orientation, are proximal to the DJC$\delta$ cluster and rearrange to $D_\delta J_\delta$ gene segments. Notably, $Trdv4$ is exclusively used in fetal thymocytes. There are also more distal $V_\delta$ gene segments, which are used as both $V_\delta$ and $V_\alpha$ gene segments. These V gene segments include $Trav21-dv12$, $Trav13-4-dv7$, $Trav6-7-dv9$, $Trav4-4-dv10$, $Trav14d-3-dv8$, $Trav16d-dv11$ and the $Trav15-dv6$ ($V_\delta 6$) family (Carico and Krangel, 2015).
1.4 Regulation of V(D)J recombination

The initiation of V(D)J recombination is regulated by three distinct mechanisms. First, expression of RAG1 and RAG2 is tightly regulated to coordinate developmental stage-specific V(D)J recombination (Kuo and Schlissel, 2009). Second, binding and activity of RAG at AgR loci is tightly regulated during lymphocyte development. Third, the three dimensional (3D) conformation of AgR loci plays a critical role in juxtaposing widely separated gene segments during synapsis (Schatz and Ji, 2011).

1.4.1 Expression of RAG

Developing T and B cells both experience two waves of RAG expression during their differentiation. For thymocytes, RAG genes are initially expressed in DN thymocytes to mediate rearrangement of the \( Tcrb \), \( Tcrg \), and \( Tcrd \) genes. Successful rearrangement of \( Tcrb \) leads to cellular proliferation and downregulation of RAG. RAG is expressed at the DP stage to initiate \( Tcra \) rearrangement (Yannoutsos et al., 2001). In developing B cells, RAG is expressed in pre-pro- and pro-B cells to mediate \( Igh \) rearrangement, and then expressed in pre-B cells to mediate rearrangement of the \( Igk \) and \( Igl \) loci. Distinct cis-regulatory elements of the \( Rag \) locus have been shown to regulate RAG expression in T and B lineage cells at different developmental stages. Expression of \( Rag \) genes in DN thymocytes is dependent on a 10-kb regulatory region 5' of \( Rag2 \) (Monroe et al., 1999a; Yu et al., 1999), while \( Rag \) expression in DP thymocytes is dependent on an anti-silencer (ASE), an element that is ~90 kb upstream of the \( Rag2 \) promoter (Yannoutsos et al., 2004). ASE\textsuperscript{−/−} mice have a minimal defect in Rag expression.
in DN thymocytes, while DP thymocytes from these mice express remarkably low levels of RAG1 and RAG2 and they are incapable of Vα-to-Jα rearrangement (Hao et al., 2015). Mechanistically, the ASE functions as an enhancer to directly activate the Rag promoters through chromatin interaction (Hao et al., 2015). In developing B cells, Rag transcription is dependent on Erag, a 1.7 kb regulatory element ~23 kb upstream of the Rag2 promoter (Hsu et al., 2003). Elimination of Erag specifically impairs RAG expression in pro- and pre-B cells but not in thymocytes.

RAG expression is also tightly controlled at the post-transcriptional level. RAG protein is subject to ubiquitination-dependent proteasomal degradation once cells enter S phase of cell cycle (Li et al., 1996). At this stage, ubiquitination-dependent degradation of RAG2 is triggered by phosphorylation of RAG2 residue T490 (Jiang et al., 2005). Since NHEJ is the predominant DSB repair pathway in G1 phase (Sonoda et al., 2006), restriction of RAG2 to the G1 phase ensures that V(D)J recombination DSBs are resolved by the classical NHEJ pathway (Helmink and Sleckman, 2012). Defects in cell-cycle restriction of RAG expression pose a major threat to genome stability and increase the probability of tumorgenesis. For example, the RAG2T490A mutation, which prohibits cell-cycle-coupled degradation of RAG2, promoted chromosomal translocations and increased incidence of lymphomas in p53+/− mice (Zhang et al., 2011).

1.4.2 Binding and activity of RAG

As one of the major components in the recombinase complex, RAG1 has DNA binding capacity and catalytic activity. At AgR loci, RAG1 directly binds to RSSs
through its nonamer binding domain (NBD, residues 389-464) (Teng and Schatz, 2015). However, outside of the AgR loci, binding of RAG1 to DNA seems not to be restricted to RSSs (Teng et al., 2015). RAG2 binding sites are also distributed throughout the genome, with a major fraction (~60%) of RAG2 binding sites aligning with active transcriptional start sites that are enriched for histone H3K4 trimethylation (H3K4me3) (Ji et al., 2010b; Teng et al., 2015). Indeed, RAG2 can directly bind to H3K4me3 via its plant homeodomain (PHD) finger (Liu et al., 2007b; Matthews et al., 2007). Binding of RAG2 to H3K4me3 releases an auto-inhibitory effect conferred by the C-terminus of RAG2 (residues 387-520), whose presence weakens pre-cleavage RAG-DNA interactions and reduces the efficiency of RAG activity (Grundy et al., 2010). Therefore, the interaction between RAG2 and H3K4me3 enables the RAG recombinase to “interpret” the histone modification code.

Recently, a single-molecule assay showed that RAG bound to the 12-RSSs of a naked DNA substrate with higher affinity than with the 23-RSSs, and that the dwell time of RAG on 12-RSSs was longer (Lovely et al., 2015). However, in vivo ChIP assay showed that RAG is enriched at the 23-RSS bearing Jκ segments, but not at 12-RSS bearing Vκ segments at the Igκ locus (Ji et al., 2010b). This suggests that the “12-RSS-first” model might be restricted to naked DNA substrates, whereas binding of RAG on chromatin is dictated by chromatin accessibility. A 3.4 Å cryo-electron microscopy study of RAG-DNA complexes revealed that the 12/23 rule is dictated by an induced asymmetry mechanism: binding to either the 12-RSS or 23-RSS induces a conformational change of
the RAG1 NBD, which facilitates the RAG complex binding to the reciprocal RSS (Ru et al., 2015). Another protein in the recombinase complex, HMGB1, is required to bend the 23-RSS prior to RAG binding (Ciubotaru et al., 2013; Little et al., 2013; Lovely et al., 2015).

### 1.4.3 Locus accessibility

Since the seminal discovery of non-coding germline transcription from unrearranged V\_H genes and across the J\_H-C\_H regions in pro-B cells over 30 years ago (Yancopoulos and Alt, 1985), many groups have reported that the frequency of V(D)J recombination coincides with the abundance of germline transcripts running across unrearranged gene segments (Krangel, 2003). In addition, cis-regulatory elements, such as promoters and enhancers, have been shown to dictate V(D)J recombination (Krangel, 2003). Supported by these data, the accessibility hypothesis became the prevailing model to explain how V(D)J recombination could be developmentally regulated. This hypothesis is based on the notion that highly compact chromatin is an inherent barrier for RAG and other DNA-binding proteins, and that transcriptional elongation either reflected or actually promoted an open chromatin structure that was accessible to RAG. A causal relationship was formally tested through insertion of a transcription terminator downstream of T early alpha (TEA), which is the promoter for 5' J\_alpha gene segments of the Tcra locus. Transcriptional blockade imposed by this terminator caused a severe accessibility and rearrangement defect in immediately downstream J\_alpha gene segments, indicating that transcriptional elongation has a direct role in remodeling chromatin.
structure and rendering chromatin accessibility (Abarrategui and Krangel, 2006). Other than the transcription machinery, histone modifying enzymes and ATP-dependent chromatin remodeling complexes work in concert to regulate accessibility of AgR loci. In vitro biochemical assays showed that histone acetylation can stimulate RSS cleavage by RAG (Kwon et al., 2000). Indeed, histone acetylation marks active and accessible gene segments of the Tcra-Tcrd locus during thymocyte development (McMurry and Krangel, 2000). This observation has been generalized to other AgR loci. For example, acetylation of V<sub>H</sub> gene segments is high in pro-B cells to coincide with a high usage of V<sub>H</sub> gene segments for V(D)J recombination. Accordingly, V<sub>H</sub> gene segments are deacetylated in pre-B cells to prohibit any further V<sub>H</sub>-to-D<sub>H</sub>J<sub>H</sub> rearrangement (Chowdhury and Sen, 2003). On the contrary, histone H3K9 methylation has been shown to inhibit recombinase accessibility. Experimentally tethering histone methyltransferases to accessible Dβ and Jβ gene segments suppressed accessibility and inhibited Tcrb rearrangement (Osipovich et al., 2004).

Since an RSS embedded in a nucleosome is generally refractory to RAG cleavage, chromatin must undergo a change of configuration to increase its accessibility. SWI/SNF is an ATP-dependent chromatin remodeling complex that can restructure the nucleosome and overcome the chromatin barrier that inhibits RSS accessibility and RAG-cleavage (Kwon et al., 2000). Prior to V(D)J recombination, SWI/SNF is required to be recruited to gene segments to increase their accessibility to the RAG recombinase (Morshead et al., 2003; Osipovich et al., 2007). The active component of SWI/SNF, BRG1,
is known to target Tcrb promoters and promotes accessibility of Δβ RSSs (Osipovich et al., 2007). Moreover, BRG1 can promote accessibility at the Igh locus in pro-B cells and it is critical for normal V(D)J recombination of the Igh locus (Osipovich et al., 2009).

Recently, RAG was found to focally bind to a discrete chromatin region of the Igh, Igk, Tcrb and Tcra-Tcdr loci containing D or J gene segments in a developmental stage- and cell lineage-specific manner (Ji et al., 2010a; Ji et al., 2010b; Teng et al., 2015). These focal regions, designated recombination centers (RCs), are characterized by an elevated level of germline transcription and histone modifications associated with active chromatin, such as H3K4me3 and H3Ac. This discrete pattern of RAG binding is consistent with the RSS capture model, in which RAG1 and RAG2 initially engage one highly accessible RSS, followed by capture of a distal RSS to form the synaptic complex. Since the distal V gene segments do not display substantial RAG binding, they need to access RAG bound to the recombination center. Thus, recruitment of the distal RSSs to the RC requires conformational changes of the AgR loci.

Chromatin accessibility at Ig and Tcr loci is regulated by cis-regulatory elements that drive sense and antisense transcription across gene segments. For example, several cis-regulatory elements have been identified throughout the Igh locus. The intronic enhancer, Eµ, is located in the interval between the JH4 gene segment and the CH exon (Banerji et al., 1983; Gillies et al., 1983) (Fig. 2). Elimination of Eµ results in reduced ΔH-to-JH rearrangements and a severe defect in the VH-to-ΔHJH rearrangement (Afshar et al., 2006; Chen et al., 1993; Perlot et al., 2005; Sakai et al., 1999; Serwe and Sablitzky, 1993).
The residual D<sub>H</sub>-to-J<sub>H</sub> rearrangement in E<sub>µ</sub><sup>−/−</sup> mice was attributed to PDQ52, a promoter that is upstream of the most 3′ D<sub>H</sub> gene segment (Alessandrini and Desiderio, 1991). PDQ52 is thought to have a redundant role with E<sub>µ</sub> in promoting accessibility for the DQ52-J<sub>H</sub> region by initiating a sense transcript, termed µ0, that runs through the J<sub>H</sub> segments and C<sub>µ</sub> exons and is spliced from the J<sub>H</sub>1 splice donor site to C<sub>µ</sub> (Schlissel et al., 1991). Notably, PDQ52 also generates E<sub>µ</sub>-dependent antisense transcripts, which play a role in silencing the central D<sub>H</sub> segments, through a mechanism termed RNA interference-mediated transcriptional gene silencing (Chakraborty et al., 2007). Apart from these regulatory elements, each V<sub>H</sub> gene segment is actively transcribed by its own promoter, which generates germline V<sub>H</sub> transcripts in pro-B-cells. These V<sub>H</sub> promoters remain active in mature B cells to maintain expression of the fully rearranged IgH chain.

The Igk locus contains two germline promoters referred to as κ<sup>0</sup> promoters (Amin et al., 2009) (Fig. 2). The proximal κ<sup>0</sup> promoter, located upstream of the Jκ1 gene segment, promotes a bias to Jκ1 usage in primary Vκ-to-Jκ rearrangements, which spares downstream Jκ gene segments for secondary rearrangements during receptor editing (Yamagami et al., 1999). The proximal κ<sup>0</sup> promoter generates spliced 0.8 kb germline transcripts (κ<sup>0</sup> 0.8) at the pre-B cell stage, while the distal κ<sup>0</sup> promoter, located ~3.5 kb upstream of the Jκ1 gene segment, generates a spliced 1.1 kb GLT (κ<sup>0</sup> 1.1) specifically in pre-B cells. These promoters are important for the developmental regulation of Igk recombination, as deletion of both the proximal and distal κ<sup>0</sup> promoters severely impairs Vκ-to-Jκ rearrangements (Meyer and Neuberger, 1989). Three Igk
enhancers have been identified: The intronic enhancer (iEκ), the 3’ enhancer (3’Eκ) and the distal κ enhancer (dEκ) (Fig. 2). iEκ, a counterpart of Eµ, is situated between Jκ gene segments and the Cκ gene (Inlay et al., 2002). The 3’ enhancer (3’Eκ) is located approximately 9 kb downstream of the Cκ gene (Meyer and Neuberger, 1989). Individually deleting either iEκ or 3’Eκ did not completely block Igk rearrangements, whereas eliminating both iEκ and 3’Eκ completely abolished Igk rearrangements (Gorman et al., 1996; Inlay et al., 2002; van der Stoep et al., 1998; Xu et al., 1996). 3’Eκ and dEκ, but not iEκ, drive and maintain high level Igk mRNA expression of a functionally rearranged Igk gene (Liu et al., 2002; Xiang and Garrard, 2008). Mice lacking dEκ have a defect in Igk mRNA expression, leading to a reduced κ:λ ratio in mature splenic B cells (Xiang and Garrard, 2008). 3’Eκ has been shown to regulate expression of Igk in mature B cells, to direct nuclear position of the Igk loci and to enforce allelic exclusion (Xiang and Garrard, 2008).

The Tcrb locus includes several distinct regulatory elements. Located at the 3’ end of the locus (Fig 2), the Tcrb enhancer (Eβ) is specifically activated in DN thymocytes and is responsible for the transcription and accessibility of both the Dβ1–Jβ1 and Dβ2–Jβ2 clusters, which form recombination centers due to high enrichment of RAG (Osipovich and Oltz, 2010). Eβ+/− mice exhibited impaired Tcrb recombination (Bories et al., 1996; Bouvier et al., 1996). This defect is attributed to impaired accessibility across the two Dβ-Jβ clusters (Bories et al., 1996; Bouvier et al., 1996; Capone et al., 1993; Ji et al., 2010a). Similar to VH gene segments, each Vβ gene segment is associated with its own promoter
Indeed, transcription of germline Vβ gene segment is largely retained in Eβ−/− mice, indicating that Vβ accessibility is regulated in an Eβ-independent manner (Mathieu et al., 2000). In contrast to the global effect of Eβ on the Tcrb locus, PDβ1 and PDβ2, two promoters associated with Dβ1 and Dβ2, respective, can only confer accessibility to their downstream Dβ-Jβ cluster. PDβ1-deleted mice lose chromatin accessibility and rearrangement of the Dβ1–Jβ1 cluster (Whitehurst et al., 2000). Nucleosome mapping showed that PDβ1-generated transcripts are essential for repositioning nucleosomes and increasing accessibility across the Dβ1-Jβ1.3 region (Kondilis-Mangum et al., 2010). Accessibility of the Dβ2-Jβ2 cluster is activated by PDβ2 (McMillan and Sikes, 2008, 2009).

The Tcra-Tcrd locus features three well-characterized cis-regulatory elements: the Tcrd enhancer (Eδ), the Tcra enhancer (Eα), and the T early a (TEA) promoter (Fig. 2). Eδ is activated in DN thymocytes and inactivated in DP thymocytes, while Eα and TEA are activated in DP thymocytes (Hernandez-Munain et al., 1999). Developmental-stage-specific activation of these elements directs the developmental accessibility of Tcrd and Tcra gene segments.

Eδ, located within the Jδ-Cδ intron, appears to stimulate local accessibility within a 68-kb region encompassing Trdv4, Trdd1, Trdd2, Trdj1 and Trdj2 and Trdv5 gene segments. In adult thymocytes, elimination of Eδ reduced germline transcription at Trdd2, Trdj1, Trdj2 and Trdv5 gene segments and reduced histone acetylation, H3K3me2 and H3k4me3 at Trdd2, Trdj1 and Trdj2 gene segments (Hao and Krangel, 2011; Huang
and Sleckman, 2007). In fetal thymocytes, Eδ can provide accessibility to Trdv4 at a distance of 68 kb (Hao and Krangel, 2011). The Dδ-Jδ region is enriched for RAG and thus constitutes the Tcrd RC (Teng et al., 2015).

Unlike Eα, which can activate gene segments nearly 500 kb away (will be discussed below), Eδ has a limited ability to activate transcription over a long distance. Indeed, accessibility of most Vδ gene segments, except for Trdv4 (exclusively used in fetal thymus) and Trdv5, is intact in Eδ-deficient (EδKO) mice (Hao and Krangel, 2011; Huang and Sleckman, 2007). EδKO mice exhibit only a partial defect in Tcrd rearrangement with decreased Dδ-to-Jδ rearrangement and increased Vδ-Dδ rearrangement (Monroe et al., 1999b). This distance limitation seems to be an intrinsic feature of Eδ, as placing Eδ at the exact location of Eα prevented Eδ from activating its natural Dδ and Jδ targets when they were 100 kb away (Bassing et al., 2003).

Only a handful of V gene segments are found in Tcrd rearrangements. Among them, Trdv2-2, Trdv5 and the Trav15-dv6 family are the Vδ gene segments that dominate adult Tcrd repertoire (Hawwari and Krangel, 2005). These Vδ gene segments are also among the highest accessible V gene segments in DN thymocytes (Hawwari and Krangel, 2005). Theoretically, the developmental regulation of Vδ accessibility would be imparted by the promoters of Vδ gene segments. To formally test this, the promoter of a Vα gene segment, Trav12-2, was replaced with the promoter of a Vδ gene segment, Trav15-1-dv6-1, through gene targeting (Naik et al., 2015). Equipped with the Vδ promoter, Trav12-2 displayed abnormally elevated accessibility in DN thymocytes,
enabling *Trav*12-2 to rearrange to D$\delta$J$\delta$ gene segments (Naik et al., 2015). It is not clear whether this promoter-driven model is generalizable to all the V$\delta$ gene segments, since *Trav*2 and *Trav*14 family V gene segments are highly accessible in DN thymocytes, but are rarely found in Tcrd rearrangements (Hawwari and Krangel, 2005). It is possible that these V gene segments are regulated by mechanisms other than chromatin accessibility.

E$\alpha$ is responsible for the accessibility of the C$\alpha$-proximal V$\alpha$ segments and the entire set of J$\alpha$ gene segments even though it is situated at the most 3’ end of the Tcra-Tcrd locus. E$\alpha$-deletion completely blocks V$\alpha$-to-J$\alpha$ rearrangement (McMurry and Krangel, 2000; Sleckman et al., 1997). The core E$\alpha$ enhanceosome spans approximately 700 bp and consists of binding sites for transcriptional factors Sp1, Ets-1, TCF-1, CREB, RUNX1, GATA-3, E2A and HEB (Carico and Krangel, 2015). In addition, CCCTC-binding factor (CTCF), a transcription factor involved in chromatin looping, binds to a region downstream of the core E$\alpha$ enhanceosome. CTCF is believed to be critical for the long-distant enhancer activity of E$\alpha$ (reviewed in a following section).

T-early alpha (TEA) is a promoter that resides at the 5' end of the J$\alpha$ array and drives transcription across a 12 kb region housing the 5’ J$\alpha$ gene segments and promotes chromatin accessibility of this region. Elimination of TEA leads to impaired usage of 5’ J$\alpha$ gene segments (J$\alpha$61-J$\alpha$52) (Villey et al., 1996). This defect is attributed to loss of accessibility and active histone modifications in this region, including H3ac, H4ac, H3K4me2, and H3K4me3 (Abarrategui and Krangel, 2007; Hawwari et al., 2005; Mauvieux et al., 2001). Elimination of TEA also abrogates RAG binding to the 5’ J$\alpha$
region, thus disrupting the Tcra RC (Ji et al., 2010a; Ji et al., 2010b). All of these effects on chromatin structure and RAG binding are directly attributable to TEA-dependent transcription through the 5' Jα gene segments (Abarrategui and Krangel, 2006, 2007; Ji et al., 2010a). The Jα49 promoter (Jα49p) is 15 kb downstream of TEA and dictates the accessibility of more 3' Jα gene segments (Jα49-Jα45) (Hawwari et al., 2005). Other "cryptic" promoters further downstream are thought to be suppressed by transcriptional interference from TEA and Jα49p, as TEA and Jα49p double-deleted alleles exhibit upregulated promoter activity downstream of the Jα49 gene segment in RAG-deficient mice (Abarrategui and Krangel, 2007; Hawwari et al., 2005). Jα usage occurs in a sequential, coordinated manner with primary recombination favoring the 5' Jα and most 3' Vα gene segments (Trav17, Trav19 and Trav21), followed by secondary recombination using more 3' Jα gene segments and more 5' Vα gene segments (Carico and Krangel, 2015). Mechanistically, secondary Vα-to-Jα rearrangements are thought to be targeted to more downstream Jα gene segments by the promoters of Vα at 5'end of a Vα-Jα rearrangement (Hawwari and Krangel, 2007). However, whether the centrally positioned Vα gene segments are rearranged in a manner that is coordinated with Jα gene segments is an open question.

1.5 Ordered rearrangement and allelic exclusion

Tcrb and Igh rearrangement is developmentally ordered in that D-To-J rearrangements occur prior to V-to-DJ rearrangements (Krangel, 2009). Direct V-to-J rearrangement is prohibited by properties of RSSs and flanking sequences that instruct
recombination “beyond 12/23” (Bassing et al., 2000; Tillman et al., 2004). One cell-free assay using naked DNA substrates with RAG and HMGB1 has shown that the "beyond 12/23" restriction is enforced at the initial nicking and pairing steps of V(D)J recombination (Drejer-Teel et al., 2007). Ordered Tcrb rearrangement is also actively promoted by RSSs. The 23-RSS at the 3’ end of Dβ1 contains a conserved binding site for AP-1 component c-Fos, which can recruit RAG to this RSS (Wang et al., 2008). RAG recruitment to this RSS presumably prevents its recruitment to the 5’Dβ1 RSS, thus preventing Vβ-to-Dβ1 recombination. Either eliminating the Dβ1 23-RSS, or knockout of c-Fos, promotes Vβ-to-Dβ1 rearrangement (Wang et al., 2008). Physiologically, Dβ-to-Jβ rearrangement will naturally eliminate the 3’ 23-RSS of Dβ1, thus enabling Vβ-to-DJβ rearrangement after Dβ-to-Jβ has occurred.

One prominent feature of the adaptive immune system is the monospecificity of αβ T and B lymphocytes. Monospecificity of TCR and BCR allows for efficient clonal selection of foreign antigen-reactive cells, as well as elimination of self-reactive cell clones. The genetic basis underlying monospecificity is allelic exclusion, which ensures that Tcrb, Igh, Igk and Igl genes are functionally rearranged on only one allele (Brady et al., 2010; Vettermann and Schlissel, 2010).

For Tcrb and Igh loci, allelic exclusion is enforced at the V-to-DJ step in two phases: an initiation phase and a maintenance phase (Brady et al., 2010). In the initiation phase, it is thought the V-to-DJ rearrangement is not attempted simultaneously on both alleles (Vettermann and Schlissel, 2010). Two types of mechanisms are proposed to
underlie asynchronous recombination: (1) a deterministic mechanism, in which one allele is the preferred substrate for RAG. This model is supported by one study showing asynchronous replication of Tcrb loci. As was shown for the Igk locus, this might promote epigenetic changes that render the earlier replicating allele more accessible (Mostoslavsky et al., 2001); (2) a stochastic mechanism, in which low efficiency of V(D)J recombination deters simultaneous V-to-DJ recombination.

In the maintenance phase, an in-frame rearrangement inhibits V-to-(D)J recombination on the other allele through a negative feedback mechanism. Feedback inhibition of Tcrb recombination in DP thymocytes and Igk recombination in pre-B cells is associated with reduced accessibility, and locus conformational changes (discussed later) (Brady et al., 2010; Jackson and Krangel, 2006). In addition, pre-TCR and pre-BCR signaling can induce proliferation of thymocytes and pre-B cells, which downregulates RAG expression though the aforementioned cell-cycle-dependent protein degradation (Yannoutsos et al., 2001). Feedback signals induced by pre-TCR also downregulate E47 expression, which confers accessibility to Vβ gene segments through direct DNA binding (Agata et al., 2007). Indeed, ectopic expression of E47 overrode the feedback inhibition mechanism and disrupted allelic exclusion of Vβ gene segments. Notably, there are cases in which Vβ segments are accessible but still unable to rearrange in DP thymocytes. Jackson et al. generated a Tcrb allele in which Eα is knocked into the Vβ region downstream of Trbv15 (Jackson et al., 2005). Although this ectopic Eα increased the germline transcription and accessibility of Vβ gene segments in DP thymocytes,
feedback inhibition of Vβ-to-Dβ-Jβ recombination in DP thymocytes remained intact. A follow-up study showed that highly accessible Vβ segments situated upstream of a functionally rearranged Vβ-to-Dβ-Jβ segment still cannot rearrange DP thymocytes (Jackson and Krangel, 2005). Moreover, Trbv31, situated downstream of the Tcrb RC, is highly accessible in DP thymocytes but refractory to recombination (Mathieu et al., 2003). Moreover, rearrangement, but not accessibility of Trbv31 can be inhibited in mice containing a pre-assembled Vβ-Dβ-Jβ rearrangement (Yang-IoN et al., 2010). These results suggest that there are additional mechanisms to enforce allelic exclusion in DP thymocytes that may be intrinsic to Vβ segments.

Allelic exclusion of the Igk locus is also enforced at the initial DNA cleavage step. ATM activated by RAG-mediated DSBs downregulates Gadd45α, which is required for transcription of Rag1 and Rag2. Thus, the ATM-Gadd45α pathway ensures that DSBs at one Igk allele inhibit further rearrangements of the other allele even before alert of feedback based on production of a functional protein (Steinel et al., 2013). It is unclear whether this pathway regulates other AgR gene rearrangements.

Unlike Tcrb and Igh loci, rearrangement of the Tcra-Tcrd locus is not ordered and is not subject to allelic exclusion. First, Vδ-to-Dδ, Dδ-to-Dδ, and Dδ-to-Jδ rearrangements can all be detected in DN thymocytes (Chien et al., 1987). Second, productive Tcrd rearrangements can be detected from both alleles in mouse γδ hybridomas, indicating that both alleles can simultaneously undergo Tcrd rearrangements or that there is no feedback inhibition from the productive allele (Sleckman et al., 1998). Tcra
rearrangement is similarly biallelic and not allelically excluded (Malissen et al., 1992; Malissen et al., 1988).

1.6 Subnuclear localization

Subnuclear localization of genes is highly correlated with gene expression. Using unbiased whole-genome analysis, several groups have reported that lamina-associated genes are largely transcriptionally silent. Accordingly, genes are generally activated when they are distant from the nuclear lamina (Kind and van Steensel, 2010; Peric-Hupkes et al., 2010; Pickersgill et al., 2006). Chromatin regions that interact with the nuclear lamina are termed nuclear lamina-associated chromatin domains (LADs) (Guelen et al., 2008). LADs contain DNA binding motifs that can recruit transcriptional repressors, including cKrox and HDAC3, which are critical for gene silencing and for the interaction between LADs and lamina (Zullo et al., 2012). Indeed, tethering of a gene to the nuclear lamina can inhibit gene expression (Reddy et al., 2008; Zullo et al., 2012).

Moreover, tethering of a specific subset of human chromosomes to the nuclear lamina by using a lac operator/lac repressor (lacO/lacI) system can repress genes that are proximal to LADs (Finlan et al., 2008). These data indicate that subnuclear localization regulates gene expression.

Several lines of evidence indicate that subnuclear localization of AgR loci contributes to the temporal and cell-lineage specificity of V(D)J recombination. Peripheral localization of Igh and Igk loci was observed in hematopoietic progenitors, whereas these loci migrate to the center of the nucleus at the developmental stages that
support V(D)J recombination (Kosak et al., 2002; Reddy et al., 2008). The Tcrb locus is associated with the nuclear lamina in DN thymocytes, coinciding with Tcrb rearrangement (Schlimgen et al., 2008). Further examination of recently rearranged Tcrb alleles marked by the DNA-repair protein 53BP1 showed that nuclear lamina-associated Tcrb alleles rearranged at a lower frequency than free alleles, suggesting that subnuclear localization may help to enforce Tcrb allelic exclusion (Chan et al., 2013). Suppression of V(D)J recombination at the nuclear lamina was attributed to the selective exclusion of RAG2 from this compartment (Chan et al., 2013).

Whether Igh and Igk loci are regulated through subnuclear localization is controversial. Earlier studies showed that the non-productively rearranged or unrearranged allele is associated with repressive pericentromeric heterochromatin (PCH) in pre-B and activated B cells, while the productive allele is centrally localized (Goldmit et al., 2005; Kosak and Groudine, 2004; Roldan et al., 2005; Skok et al., 2001; Su et al., 2005). Monoallelic association of the Igh locus with PCH is mediated by pre-BCR and IL-7/IL-R signaling, and was proposed to be the feedback inhibition mechanism for Igh allelic exclusion (Roldan et al., 2005). However, monoallelic association of Igh with PCH is contradictory to multiple studies showing that Igh alleles are comparably transcribed and carry similar active chromatin marks in mature B cells (Daly et al., 2007; Tinguely et al., 2012). In addition, mice lacking the silencer in the intervening sequence (Sis), a regulatory element in the Igk locus that promotes monoallelic association between PCH and both Igk and Igh loci, still displayed normal Igk allelic exclusion.
Moreover, Ig loci do not exhibit differences in their subnuclear localization during B cell development, although Ig rearrangement is subjected to allelic exclusion (Kosak et al., 2002). Together, these data indicate that repositioning of the nonproductive allele to repressive nuclear environment is not responsible for maintenance of allelic exclusion.

## 1.7 Locus conformation

Eukaryotic genomes must be properly packaged to fit the capacity of cell nuclei. It has been increasingly appreciated that the 3D organization of the genome is critical for gene regulation. Fluorescent in situ hybridization (FISH) allows for in situ, but low resolution visualization of nuclear organization and 3D-structure of chromatin. Over the last decade, the development of chromosome conformation capture (3C) and other 3C-based techniques has provided the means to examine spatial proximity between DNA elements in a greater detail (Gibcus and Dekker, 2013; Sexton and Cavalli, 2015). 3C-based assays use formaldehyde to cross-link proximal (presumably interacting) DNA fragments, followed by restriction digestion and ligation of the interacting fragments. This ligated 3C library can be used for 3C analysis, which quantifies the interaction frequency between two-specific fragments by realtime-PCR using one primer for each interacting partner (Dekker et al., 2002). The 3C library can alternatively be subjected to secondary digestion and ligation, resulting in a library for 4C. 4C allows interrogation of the entire interactome of a chosen “viewpoint” fragment through inverse PCR and next-generation sequencing (Zhao et al., 2006). Hi-C allows for even more complete analysis
of the interactome without prior knowledge of fragments of interest. These technical advances have significant improved our understanding of the connection between chromatin conformation and gene expression (Sexton and Cavalli, 2015).

The earliest evidence that linked chromatin interactions and gene expression came from the studies of the haemoglobin subunit beta (Hbb) locus (Palstra et al., 2003). At this locus, distal regulatory elements interact with their target genes via chromatin loops, which create sub-megabase chromatin domains. These domains are enriched for RNA polymerase II (Palstra et al., 2003). Additional studies investigating enhancer-promoter interactions supported the idea that chromatin loops connecting promoters and enhancers are an important aspect of gene regulation (Kieffer-Kwon et al., 2013; Li et al., 2012; Sanyal et al., 2012).

Recently, several groups have reported that the mammalian genome is organized into megabase-sized topologically associating domains (TADs), characterized by high frequency intra-domain interactions (Dixon et al., 2012; Nora et al., 2012). TADs represent discrete regulatory units, as genes within the same TAD display uniform chromatin activity, including transcription, histone modifications, lamina association and DNA replication timing (Dixon et al., 2012; Le Dily et al., 2014; Nora et al., 2012; Pope et al., 2014). Genome-wide comparisons of TADs in various cells and tissues from mouse and human revealed that megabase-sized TADs are relatively invariant among cell types (Dixon et al., 2012). However, “sub-TADs” defined by sub-megabase-sized interactions are highly dynamic during stem cell differentiation, based on high-
resolution Hi-C analysis of 7 genomic loci in embryonic stem cells and neural progenitor cells (Phillips-Cremins and Corces, 2013). In addition, intra-TAD interactions appear to be highly dynamic in accord with transcriptional activation (Dixon et al., 2015; Giorgetti et al., 2014).

TADs play a critical role in transcriptional control through confining enhancer activities within TAD boundaries (Ghavi-Helm et al., 2014; Shen et al., 2012; Symmons et al., 2014). Disrupting TAD structure generated new TADs which promoted ectopic interactions between genes and enhancers and aberrant transcription (Nora et al., 2012; Giorgetti et al., 2014). Altered TAD structure has been linked to some human genetic diseases (Giorgetti et al., 2014).

How TAD boundaries are established and maintained remains unclear. In the mammalian genome, several architectural proteins, including CTCF, mediator and cohesin, are known to be enriched at the anchor sites of chromatin interactions (Sexton and Cavalli, 2015). CTCF/cohesin complexes normally anchor long-range constitutive interactions (Dixon et al., 2012; Phillips-Cremins and Corces, 2013), whereas Mediator/cohesin complexes mediate short-range dynamic interactions (Phillips-Cremins and Corces, 2013). Consistently, CTCF-deficient cells displayed increased inter-TAD interactions (Zuin et al., 2014). However, although CTCF- and cohesin-binding sites are enriched at TAD borders, the majority (~75%) of these sites are found within TADs (Dixon et al., 2012; Nora et al., 2012; Phillips-Cremins and Corces, 2013). How these architectural proteins regulate TAD structure needs further investigation.
1.7.1 CTCF

CTCF is a highly conserved, ubiquitously expressed transcription factor that contains a DNA-binding domain with 11 zinc-fingers (ZFs) (Ohlsson et al., 2001). As measured by chromatin immunoprecipitation sequencing (ChIP-seq), CTCF binds to over $5 \times 10^5$ sites in mammalian genomes, with 40-70% of these CTCF-binding sites highly conserved among various cell types across mammalian species (Chen et al., 2008; Cuddapah et al., 2009; Kim et al., 2007; Schmidt et al., 2012). Computational analyses of CTCF-binding sites in the human genome uncovered a 15-20 bp consensus motif (hereafter referred to as the core motif) (Kim et al., 2007). Recently, CTCF binding was examined across the whole genome using a novel ChIP method at single-nucleotide resolution (ChIP-exo) (Rhee and Pugh, 2011). This study identified 35,161 CTCF-binding sites in HeLa cells, 17,000 more than using the conventional ChIP method in the same cell type (Cuddapah et al., 2009). Strikingly, it uncovered a ~41-bp CTCF-binding motif that contains four combinatorial modules (Rhee and Pugh, 2011). Approximately 50% of CTCF-binding sites contain modules 2 and 3, constituting the core motif (Kim et al., 2007). Other CTCF-binding sites contain either three or four modules (~49%), or only module 3 (~1%) (Rhee and Pugh, 2011).

How CTCF binds to unique DNA sequences remains unclear. In vitro biochemical assays using CTCF mutants showed that four of its eleven ZFs are critical for its binding to the core motif (modules 2 and 3), whereas other zinc fingers are used in a combinatorial manner to recognize additional flanking sequences (Ohlsson et al.,
This model, termed the “CTCF code”, has been recently refined by Casellas and colleagues, who investigated the in vivo binding pattern of CTCF ZF mutants at 50,000 CTCF-binding sites in primary lymphocytes. They determined that the ZFs of CTCF are divided into distinct binding subdomains with ZFs 4–7 recognizing the conserved core motif (modules 2 and 3) and peripheral ZFs (ZFs 1-3 and 8-11) binding to flanking DNA sequences. The peripheral ZFs are critical for the binding stability of CTCF. This study showed that CTCF associates with a diverse array of sequences via ZF clustering (Nakahashi et al., 2013).

Initially, CTCF was identified as an insulator protein that can block enhancer-promoter interaction at the chicken β-globin locus and the Igf2/H19 locus of mouse and human (Burgess-Beusse et al., 2002). CTCF was later found at barrier elements demarcating the genome into distinct chromatin domains, which exhibit unique epigenetic states and transcriptional activities (Gaszner and Felsenfeld, 2006). For example, a genome-wide study in human cells revealed that CBEs are enriched at a fraction of the boundaries between gene-silencing H3K27me3 domains and the active H2AK5ac domains (Cuddapah et al., 2009). Moreover, CBEs were found at the borders of LADs, which are repressive chromatin domains (Guelen et al. 2008).

Recently, CTCF has been described as an organizer of chromatin structure based on its role in mediating long-distance chromatin looping between CBEs (Ong and Corces, 2014). Because it can promote chromatin looping in different chromatin contexts, CTCF can function as a transcriptional activator or repressor. For example, the CTCF-
mediated interactome was mapped in mouse ES cells using chromatin interaction analysis with paired-end tag sequencing (ChIA–PET) (Handoko et al., 2011). Combining this analysis with mapping of histone modification for 1,480 CBE-mediated loops showed that these loops can (1) facilitate promoter-enhancer interactions, thus activating gene expression; (2) confine active histone modifications; and (3) demarcate active and repressive chromatin domains. Thus, CTCF exerts its diverse influences on gene expression due to the nature of the chromatin loops.

CTCF is particularly enriched at AgR loci and it has been implicated in regulating V(D)J recombination through long-range chromatin looping (Shih and Krangel 2013) (discussed later).

1.7.2 V(D)J recombination is regulated by the conformation of AgR loci

AgR diversification must overcome daunting topological constraints to recruit gene segments for recombination that may be distributed across several Mb of DNA. Multiple studies have shown that AgR loci undergo large-scale conformational changes during lymphocyte development, bringing distant gene segments into proximity. For example, The \(Igh\) locus undergoes a large-scale conformational change during B cell development (Fuxa et al., 2004; Jhunjhunwala et al., 2008; Roldan et al., 2005). In pre-pro-B cells, the \(Igh\) locus adopts an extended conformation and \(V_H\) gene segments and \(D_HJ_H\) gene segments are segregated into distinct chromatin domains. Accordingly, only \(D_H\)-to-\(J_H\) recombination can occur at this stage. In pro-B cells, the \(Igh\) locus contracts to promote
merging and juxtaposition of $V_H$ gene segments and the $D_H$-$J_H$ recombination center, presumably allowing for high frequency interactions and rearrangements. In pre-B cells, the unrearranged $Igh$ alleles adopt a decontraction configuration to suppress further $V_H$-$D_H$-$J_H$ recombination and maintain allelic exclusion (Roldan et al., 2005).

Similarly, the $Tcrb$ locus undergoes contraction in DN thymocytes and decontraction in DP thymocytes (Skok et al., 2007). Conformations of the $Tcra$-$Tcrd$ locus in B cells, DN and DP thymocytes have been resolved by 3D-FISH using four DNA probes (Shih and Krangel, 2010; Skok et al., 2007). As compared to B cells, the locus adopts a highly contracted overall organization in DN thymocytes (Shih and Krangel, 2010). In DP thymocytes, the locus adopted a 3’ contracted and 5’ decontracted conformation (Shih and Krangel, 2010). Thus, the 3’ portion of the locus (including proximal V gene segments and the $J\alpha$ array) maintains a contraction configuration throughout thymocyte development. These unique conformations were thought to support distinct programs of $Tcrd$ and $Tcra$ rearrangement. The locus may be contracted in DN thymocytes so that distributed $V\delta$ gene segments may non-selectively interact with the $D\delta$-$J\delta$ RC, promoting $V\delta$ diversity in the single round of $Tcrd$ rearrangement. In DP thymocytes, the $V\alpha$ array adopts an extended conformation, presumably to facilitate 3’-to-5’ sequential usage of $V\alpha$ gene segments during multiple rounds of $V\alpha$-to-$J\alpha$ rearrangements.

An initial study reported that the $Igk$ locus undergoes contraction in pre-B cells to facilitate $V\kappa$-$J\kappa$ rearrangements (Roldan et al., 2005). However, Hi-C revealed an
extensive, locus-wide network of physical interactions connecting Vκ gene segments and iEκ in pro-B cells (Lin et al., 2012). This interaction network is present well before the initiation of Igk rearrangement in pre-B cells, presumably pre-conditioning the locus for V(D)J recombination (Lin et al., 2012).

Over the last decade, increasing data from 3C-based assays have shown that AgR loci are demarcated by chromatin loops that bring together distant segments of DNA. These chromatin loops appear to be highly dynamic and cell lineage specific, and are thought to ensure fine-tuning of gene segment utilization during V(D)J recombination (Shih and Krangel, 2013). Known transcription factors that play roles in Igh locus contraction and chromatin looping include Pax5 (Fuxa et al., 2004), Ikaros (Reynaud et al., 2008), YY1 (Guo et al., 2011a; Liu et al., 2007a; Verma-Gaur et al., 2012) and CTCF (Degner et al., 2011).

Pax5 is a B lineage cell-specific transcription factor that plays important roles in B cell development (Fuxa and Busslinger, 2007; Fuxa et al., 2004; Nutt et al., 1998; Nutt et al., 1997) and B cell lineage commitment (Heavey et al., 2003; Hoflinger et al., 2004; Nutt et al., 1999; Rolink et al., 1999; Schaniel et al., 2002). Pax5−/− pro-B cells have a severe defect in rearrangement of the distal V_{H}J_{558} gene family, whereas D_{H}-to-J_{H} recombination is normal (Nutt et al., 1997; Hesslein et al., 2003). In addition, rearrangement of the proximal V_{H}7183 and V_{H}Q52 gene segments is minimally affected in these cells (Nutt et al., 1997; Hesslein et al., 2003). Pax5 is required to promote Igh locus contraction (Fuxa et al., 2004). 3D-FISH studies showed that Pax5-deficient pro-B
cells have a decontracted Igh locus, which was proposed to prevent juxtaposition and rearrangement between accessible distal V\textsubscript{H} gene segments and D\textsubscript{H}J\textsubscript{H} gene segments (Fuxa et al., 2004). Moreover, Pax5 binds to multiple conserved repeat elements in the distal V\textsubscript{H} region in pro-B cells. These elements are termed Pax5-activated intergenic repeats (PAIRs) (Ebert et al., 2011). PAIRs are characterized by the presence of Pax5-dependent antisense transcription. In pre-B cells, PAIRs lose Pax5 binding and Pax5-dependent antisense transcription, indicating their involvement in the developmental regulation of distal V\textsubscript{H} usage (Ebert et al., 2011).

Ikaros is a transcription factor that plays a critical role in development of T cells (Geimer Le Lay et al., 2014; Ng et al., 2009; Urban and Winandy, 2004) and B cells (Alkhatib et al., 2012; Ferreiros-Vidal et al., 2013; Heizmann et al., 2013; Kirstetter et al., 2002; Ma et al., 2008; Thompson et al., 2007). Ikaros has been shown to regulate both accessibility and contraction of the Igh locus. Ikaros-deficient pro-B cells are defective in rearrangement of both proximal and distal V\textsubscript{H} gene segments due to locus decontraction and loss of V\textsubscript{H} accessibility (Reynaud et al., 2008).

YY1 is a ubiquitously expressed, multi-functional transcription factor that can activate or repress transcription through interactions with other transcriptional regulators (Affar el et al., 2006; Shi et al., 1997). YY1 has diverse and complex roles in multiple physiological and pathological processes. YY1 is essential for embryogenesis because germline knockout of Yy1 is embryonic lethal (Donohoe et al., 1999). YY1 regulates maintenance and differentiation of stem cells (Perekatt et al., 2014; Vella et al.,
YY1 is involved in differentiation of skeletal and cardiac muscle cells (Gregoire et al., 2013; Lu et al., 2013; Sucharov et al., 2008). YY1 is implicated in cancer progression, because its expression is elevated in various types of cancer (Castellano et al., 2009; Gordon et al., 2006). In this regard, YY1 can negatively regulate p53 in a transcription-independent manner (Sui et al., 2004). Numerous studies have demonstrated roles for YY1 in B cell development and V(D)J recombination of the *Igh* and *Igk* loci (Atchison, 2014; Liu et al., 2007a; Medvedovic et al., 2013; Pan et al., 2013; Verma-Gaur et al., 2012). In addition, YY1 is essential for germinal center B cell development (Green and Kroemer, 2009). YY1 binding motifs are associated with targeting of AID to non-Ig genes (Duke et al., 2013). Moreover, YY1 is known to regulate Th2 cytokine production (Hwang et al., 2013). Th2 cytokine genes *IL4*, *IL13*, and *IL5* are clustered on the Th2 locus and are coordinately regulated during Th2 cell differentiation. YY1 has been shown to bind to the promoters of *IL4* and *IL13* and to the LCR and to regulate chromatin conformation and intrachromosomal interactions at the Th2 locus. Accordingly, YY1 overexpression induced the expression of Th2 cytokine genes, while YY1 knockdown in Th2 cells impaired the production of Th2 cytokines (Hwang et al., 2013).

Both CTCF and YY1 are known to regulate chromatin interactions at the *Igh* locus (Fig. 3). The *Igh* locus contains over 85 CBEs (Degner et al., 2011). A majority of the *Igh* CBEs are located within the *V_{H}* region. In pro-B cells, these CBEs mediate multiple chromatin loops in both the 3’ and 5’ portion of the *V_{H}* region (Guo et al., 2011a). For example, the *V_{H}10* and *V_{H}3* gene segments interact with multiple sites spanning several
hundred kb, potentially segregating the $V_H$ region into several chromatin domains (Guo et al., 2011a).

Similarly, multiple chromatin loops have been identified in the 3’ portion of $IgH$ locus (Fig. 3), which extends from intergenic control region 1 (IGCR1) to the 3’ regulatory region (3’RR). IGCR1, located about 4 kb upstream of 5’ DFL16.1, contains two CBEs that confer insulator activity in an in vitro reporter assays (Guo et al., 2011b). At the very 3’ end of the $IgH$ locus, a ~40 kb regulatory element, termed the 3’ regulatory region (3’RR), is defined by four DNase hypersensitivity sites (DHSs) (DHS1-4) (Pinaud et al., 2011). The 3’ RR contains several B cell-specific enhancers, which are essential for class switch recombination (CSR) and $IgH$ expression in plasma cells (Birshtein, 2014). The 3’RR is followed by three DHSs (DHS5-7) that contains a cluster of CBEs (hereafter referred to as the 3’CBE). 3C assays have shown that IGCR1 interacts with Eµ to generate a 70 kb chromatin loop (Guo et al., 2011a; Verma-Gaur et al., 2012). Eµ is enriched for YY1 rather than CTCF, and its interaction with IGCR1 in pro-B cells is dependent on YY1 (Verma-Gaur et al., 2012). Another chromatin loop formed between IGCR1 and the 3’CBE is highly dependent on CTCF (Degner et al., 2011; Guo et al., 2011b). IGCR1$^{-/-}$ mice displayed increased accessibility and rearrangement of proximal $V_H$ gene segments, coupled with defects in the rearrangement of distal $V_H$ gene segments (Guo et al., 2011b). Therefore, the IGCR1-3’CBE loop defines a chromatin domain encompassing $D_H$, $J_H$ and $C_H$ gene segments, which functions as an insulator to prevent Eµ from activating the proximal $V_H$ gene segments. However, eliminating the 3’
CBE, which should also destroy the IGCR1-3’CBE loop, did not recapitulate the defects of IGCR1-deficient mice (Volpi et al., 2012). In 3’CBE-deficient mice, usage of proximal $V_H7183$ genes was minimally increased, while distal $V_HJ558$ usage was unaffected (Volpi et al., 2012). Hence, the role of the IGCR1-3’CBE interaction requires further investigation.

Notably, IGCR1 is required to enforce ordered and B cell lineage-specific $V_H$-to-$D_H$-$J_H$ rearrangement, particularly involving the proximal $V_H$ segments (Guo et al., 2011b). $V_H81X$, the most proximal $V_H$ segment, was found to directly join to DQ52 at high frequency in IGCR1-deficient but not in wild-type pro-B cells (Guo et al., 2011b). In addition, $V_H$-to-$D_H$-to-$J_H$ rearrangements using proximal $V_H7183$ and $V_HQ52$ segments, but not distal $V_HJ558$ segments, were detected in IGCR1-deficient DP thymocytes (Guo et al., 2011b). Moreover, IGCR1 is required for $Igh$ allelic exclusion of proximal $V_H$ segments (Guo et al., 2011b). These data indicate that allelic exclusion of proximal $V_H$ segments is mechanistically connected to their ordered rearrangement.

YY1 also plays a critical role in $Igh$ locus conformation. YY1-deficient pro-B cells lose $Igh$ locus contraction and have a defect in distal $V_H$ rearrangement (Verma-Gaur et al., 2012). ChIP sequencing identified YY1-binding sites at Eμ and in the distal and middle $V_H$ region (Medvedovic et al., 2013; Verma-Gaur et al., 2012). Notably, YY1-binding sites in the distal $V_H$ region colocalized with Pax5-binding sites and several PAIR elements. It has been shown that YY1 binds to PAIR4 and PAIR6 and contributes to antisense transcription, in cooperation with Pax5 (Medvedovic et al., 2013; Verma-Gaur
et al., 2012). It was proposed that YY1 and Pax5-mediated non-coding transcription promotes chromatin loops that connect PAIRs and Eµ, and drives Igh locus contraction (Verma-Gaur et al., 2012). 4C-seq analysis showed that conditional deletion of YY1 in pro-B cells disrupts long-range interactions between Eµ and distal VH region (Medvedovic et al., 2013). In addition, YY1 also mediates the interaction between Eµ and 3’ RR in pro-B cells (Medvedovic et al., 2013).

Recently, a hierarchy model has been proposed to explain how several levels of locus conformation can fold the 2.8 Mb Igh locus into a contracted conformation that facilitates its rearrangement program (Gerasimova et al., 2015) (Fig. 3). First, CTCF instigates three sets of chromatin loops that have an average of a few hundred kilobases in size: loops within the VH region, VH-to-DH loops, and DH-to-3’ RR loops. Second, Pax5 further contracts the VH region into chromatin domains of megabase size via direct interaction between Pax5 and CTCF at PAIR elements (Medvedovic et al., 2013). Finally, The megabase-sized VH chromatin domain merges with the 300 kb 3’ DH-Cµ domain. This contraction was proposed to be organized by YY1 binding at Eµ and the VH region. This model is consistent with the fact YY1-deficiency recapitulates the contraction defects seen in Eµ-deleted Igh alleles (Guo et al., 2011a).
The role of CTCF in $Igk$ rearrangement was initially evaluated in CTCF-deficient pre-B cells (Ribeiro de Almeida et al., 2011). $\kappa$ usage in these cells was strongly biased towards the proximal $\kappa$ gene segments. Biased proximal $\kappa$ usage was coupled with upregulated germline transcription of proximal $\kappa$ gene segments. CTCF-deficiency also led to increased interactions between $iE\kappa$, $3E\kappa$ and the proximal $\kappa$ segments, indicating that CTCF mainly functions as an insulator at the $Igk$ locus. The $Igk$ locus contains approximately 60 sites to which CTCF binds in pre-B cells. Two CBEs are particularly well characterized: Sis and Contracting element for recombination (Cer). In Sis-deleted $Igk$ alleles, usage of the proximal $\kappa$ gene segments was markedly increased at the expense of distal $\kappa$ gene segments (Xiang et al., 2011). This recapitulates the phenotype of CTCF-deficient pre-B cells. Thus, Sis represents an insulator element that prevents the $Igk$ enhancers from activating the proximal $\kappa$ gene segments. Similarly biased $\kappa$ rearrangement is also observed in Cer$^{-/-}$ mice (Xiang et al., 2013, 2014). Moreover, the Sis/Cer double deleted allele has even more strongly biased usage of proximal $\kappa$ gene segments.

Figure 3: Long-range interaction network within the $Igh$ locus
segments and reduced usage of middle and distal \( \kappa \) gene segments, indicating that Sis and Cer function together in insulation (Xiang et al., 2014).

The REPO domain of YY1 (201–226) was named after its ability to recruit Polycomb (Wilkinson et al., 2006). REPO domain mutant YY1 (YY1\( \Delta \)REPO) had intact DNA binding and transcriptional regulation but cannot recruit PcG proteins to DNA (Wilkinson et al., 2006). This mutation disrupted B cell development and impaired \( \kappa \)-to-J\( \kappa \) rearrangements (Pan et al., 2013). Presumably, YY1 regulates Igk rearrangement by directly binding to the Igk locus and recruiting PcG protein EZH2 and condensin. However, whether YY1 plays a role in Igk locus conformation is unclear.

Similar to the Igh and Igk loci, the Tcrb locus undergoes contraction in DN thymocytes (Skok et al., 2007). Further resolution of chromatin conformation by placing 3C anchors at the V\( \beta \)5, V\( \beta \)23, V\( \beta \)24 gene segments showed that these gene segments universally interact with the D\( \beta \)-J\( \beta \)-C\( \beta \) region at a much higher frequency in DN than in DP thymocytes (Skok et al., 2007). Therefore, these long-range interactions were proposed to reflect the contracted conformation of the Tcrb locus in DN thymocytes (Skok et al., 2007). Accordingly, loss of these interactions were thought to deconstruct the Tcrb locus in DP thymocytes (Skok et al., 2007). However, a recent 3C assay showed that spatial proximity of D\( \beta \)J\( \beta \) to most V\( \beta \) gene segments is unaltered when comparing DN and DP thymocytes. Only the more distal V\( \beta \) gene segments (Trbv1-11) are spatially segregated from the D\( \beta \)J\( \beta \) compartment in DP thymocytes. Hence, Tcrb decontraction
segregates only some Vβ gene segments from the Dβ-Jβ RC in DP thymocytes (Majumder et al., 2015).

Notably, the conformation of Tcrb in DN thymocytes is regulated by two CBEs upstream of PDβ1 (hereby referred to as the 5’ CBE and the PDβ1-CBE). The PDβ1-CBE functions as a barrier element to block the spread of active chromatin from the downstream domain, while the 5’ CBE tethers the Dβ-Jβ RC to the distal Vβ gene segments. These interactions are significantly reduced in PDβ1 deleted alleles. In those alleles, disruption of the 3’ barrier by PDβ1-deletion spreads the active chromatin to the 5’ CBE, and forces the 5’ CBE to be a new barrier, thereby impairing its tethering function. As a consequence of this disrupted conformation, the Tcrb repertoire is strongly skewed towards the more proximal Vβ gene segments (Majumder et al., 2015).

At the Tcra-Tcrd locus, a majority of CBEs are distributed at cis-regulatory elements, such as Ea, TEA and promoters of V gene segments (Shih., 2012). In DP thymocytes, Ea interacts with TEA and the promoters of proximal Vα gene segments through long-distance chromatin loops. These long-distance interactions are confined within a 500 kb locus-contraction unit at the 3’ end of the Tcra-Tcrd locus, presumably promoting activation of TEA and proximal Vα promoters by Ea and synapsis between Vα and Jα segments during primary Vα-to-Jα rearrangements. Importantly, this entire interaction network is dependent on both CTCF and Ea. Conditional deletion of CTCF in DP thymocytes, or elimination of Ea, disrupted interactions among Ea, TEA and proximal Vα promoters and severely impaired Vα-to-Jα recombination (Shih., 2012).
Thus, CTCF contributes to the long-distant activity of Eα documented by analysis of chromatin accessibility (Hawwari & Krangel, 2005).
2. Specific Aims

Adaptive immunity depends on a highly diverse repertoire of Tcr and Ig on T and B cells, respectively. This diversity is generated by V(D)J recombination of antigen receptor loci. Although the basic mechanisms of recombination are shared by AgR loci, they are differentially regulated in order to accommodate distinct locus structures and developmental programs. The gene segments of TCRδ and TCRα are arrayed in the TCRα/δ locus. Our main goal is to understand the mechanisms that dictate the developmental regulation of V(D)J recombination of the Tcra-Tcrd locus.

Our lab has shown that during T cell development, the Tcra-Tcrd locus undergoes large scale contraction to form long distance chromatin loops (Shih and Krangel, 2010). Moreover, CTCF, a highly conserved zinc-finger protein, mediates chromatin loop formation, bringing together distant Tcra-Tcrd gene segments and regulatory elements into proximity during V(D)J recombination (Shih et al., 2012). These conformational changes support Tcra-Tcrd gene rearrangement. We expect that long distance chromatin loops, directly mediated by CBEs, provides a central mechanism that regulates Tcra-Tcrd gene expression and rearrangement during thymocyte development. Thus we generated several mouse models with genetically modified CBEs to better understand the role of the CTCF-dependent interactome at the Tcra-Tcrd locus. In addition, another transcription factor, YY1 has been shown to regulate Ig locus conformation through a similar mechanism as CTCF (Atchison, 2014). Accordingly, we hypothesize that YY1 cooperates with CTCF to shape the conformation of Tcra-Tcrd locus.
role of CTCF and YY1 in the \textit{Tcra-Tcrd} gene rearrangement, we have the following objectives.

\textbf{AIM1: To determine the function of intragenic region 1-2 (INT1-2), a CTCF-dependent regulatory element, in V(D)J recombination at the \textit{Tcra-Tcrd} locus.}

INT1-2 is 5.8-kb DNA element in the \textit{Tcra-Tcrd} locus containing 2 CBEs, termed INT1 and INT2 respectively. 4C and 3C analysis showed that INT1-2 is a critical node for the interactome at the 3' end of the \textit{Tcra-Tcrd} locus. Thus, we generated the INT1-2 knockout (INT1-2KO) mice in order to dissect its physiological function in \textit{Tcra-Tcrd} gene rearrangement. To specifically test a role for INT2 in \textit{Tcra-Tcrd} gene rearrangement, we additionally generated INT2 mutant (INT2M) mice.

\textbf{AIM2: To test whether an ectopic CBE can promote CTCF-dependent gene interactions and increase diversity of the TCR\(\delta\) repertoire.}

To further examine how an ectopic CBE may influence conformation and gene rearrangement of the \textit{Tcra-Tcrd} locus, we generated a mouse model where an additional CBE is knocked into the J\(\delta\)-C\(\delta\) interval. We expected this could disrupt normal locus conformation and V(D)J recombination.

\textbf{AIM3: To determine the role of YY1 in \textit{Tcra-Tcrd} gene rearrangement and T cell development.}

YY1 initially intrigued us because it has been shown to regulate the conformation of \textit{Ig} loci through a similar mechanism as CTCF. However, we found that YY1-deficiency led to thymocyte apoptosis, which has not been documented. Thus, we extensively
investigated the role of YY1 in T cell development by conditionally deleting YY1 at two developmental stages.
3. Material and Methods

3.1 Mice

3.1.1 Generation of INT1-2KO, INT2M and CBE KI mice

Homology arms were generated by PCR using Phusion High Fidelity DNA Polymerase (Thermo Scientific) and were sequenced to confirm PCR fidelity. To generate INT1-2−/− mice, the long homology arm extended from nucleotide 1,497,612 to 1,503,426 and the short homology arm extended from nucleotide 1,509,115 to 1,510,716 of NCBI Reference Sequence NT_039614.1.

To generate INT2M mice, the long homology arm extended from the nucleotide 1,503,427 to 1,509,114, with nucleotides 1,509,043 to 1,509,062 (GAACACTAGGG-GGCAATGC) replaced with a scrambled sequence (CGACGAGAAGCTAGCAGTG) (Guo et al., 2011b). The short-arm extended from the nucleotide 1,509,115 to 1,510,716.

To generate CBE KI mice, the long homology arm extended from the nucleotide 1,570,947 to 1,665,419 of NCBI Reference Sequence NT_039614.1 and the short homology arm extended from nucleotide 1,576,382 to 1,577,981, with 70 bp nucleartides (CCCACTCTGTGTTGCTTAAAAGGTACACGGGGGCGAGCAGTACTCCACC AAAAGGCTTTCTCCCT) inserted to the 3’ end.

All the homology arms were cloned into the pGKneoF2L2DTA targeting vector containing a phosphoglycerate kinase promoter-driven neomycin resistance (neo′) cassette and diphtheria toxin A (DTA) selection marker (a gift from Dr. You-Wen He, Duke University). EcoRV-linearized targeting constructs were used to electroporate the
TC1 129S6/SvEvTAc embryonic stem (ES) cell line. Neomycin resistant ES cell clones were first screened by PCR and then verified by Southern blot. Verified ES cells were microinjected into C57BL/6J blastocysts, which were then implanted into pseudopregnant C57BL/6J female mice. Chimeric male founder mice were crossed with CMV-Cre transgenic female mice (Jackson Laboratory) to delete the loxP-flanked neo cassette and obtain germline transmission. Gene-targeted mice were bred to eliminate the CMV-Cre transgene and were of mixed C57BL/6 and 129 genetic background. Breeding schemes of Rag-sufficient mice ensured that littermate controls always segregated wild-type strain 129 Tcra-Tcrd alleles. Experiments analyzing mutant alleles on a Rag2−/− background used Rag2−/− mice on a 129 genetic background as controls. All mice were used in accordance with protocols approved by the Duke University Institutional Animal Care and Use Committee. Mice were sacrifice at 3-4 weeks of age for all experiments.

3.1.2 Other mouse strains

Yy1f/f mice (B6;129S4-Yy1tm2Yshi/J) (Affar el et al., 2006) obtained from The Jackson Laboratory were bred with Lck-Cre transgenic mice (B6.Cg-Tg(Lck-cre)548Jxm/J), a gift from J. Rathmell (Duke University), to generate Yy1f/f Lck-Cre mice and were further bred with with Rag2−/− mice to generate Yy1f/f Lck-Cre Rag2−/− mice. Yy1f/f CD2-Cre mice were a gift from A. Feeney (The Scripps Research Institute). Trp53−/− mice (B6.129S2-Trp53tm1Tyj/J) were a gift from D. L. Silver (Duke University) and were bred to Yy1f/f CD2-Cre mice to generate Yy1f/f CD2-Cre Trp53−/− mice. Mice used for experiments carried a mixed 129 and C57BL/6 background and were analyzed at 4-5 weeks of age.
Mice used as wild-type carried floxed YY1 alleles but lacked Cre expression. All mice were used in accordance with protocols approved by the Duke University Animal Care and Use Committee.

### 3.2 Flow cytometry analysis and Fluorescence Activated Cell Sorting (FACS)

All reagents were purchased from Biolegend unless otherwise indicated. FITC-anti-mouse TCR\(\gamma/\delta\) antibody (GL3) and PE-anti-mouse TCR V\(\delta 4\) antibody (GL2) were used for \(\gamma\delta\) TCR and V\(\delta 4\) staining. To analyze or sort DN3 thymocytes, thymocytes were stained with anti-CD4 (GK1.5) and anti-CD8 (53-6.7), followed by negative selection with Sheep Anti-Rat IgG Dynabeads (Life Technologies). Beads-depleted DN cells were stained with antibodies against 7AAD, APC-CD44 (IM7), FITC-CD25 (PC61), and PE-cy5-conjugated lineage makers, including Gr-1 (RB6-8C5), CD3\(\varepsilon\) (145-2C11), Ter-119/Erythroid Cells (TER-119), CD11b (M1/70), B220 (RA3-6B2), 7AAD-CD25-CD44-Lin- cells were sorted by FACS. To separate DN3a from DN3b thymocytes, the DN thymocytes were also stained with anti-CD28 (37.51). For intracellular staining with anti-TCR (H57-597) or anti-YY1 (H-414), cells were first stained with antibodies against surface markers before fixation and permeabilization (BD Cytofix/Cytoperm™ Kit).

### 3.3 Real-time PCR

Taqman-based real-time qPCR assays were performed using the LightCycler 480 Probes Master (Roche) with the following PCR program: 95 °C for 10 min, followed by 48 cycles of 95 °C for 10 s and 65 °C for 30 s. SYBR green real-time PCR was performed
using the QuantiFast SYBR Green PCR Kit (QIAGEN) with the following program: 95°C for 5 min, followed by 45 cycles of 95°C for 10 s and 62°C for 30 s. All PCR reactions were run in duplicate using a LightCycler 480 Real-Time PCR system (Roche).

3.4 3C

3C assays were performed as described previously with slight modifications (Hagege et al., 2007). In brief, eight to ten million thymocytes were cross-linked in 8 ml RPMI containing 10% fetal bovine serum and 2% paraformaldehyde for 10 min at 25°C. Crosslinking was stopped by adding glycine to a final concentration of 0.125 M. Cross-linked thymocytes were washed in PBS, pelleted and lysed in 1 ml of 10mM Tris pH 8.0, 10mM NaCl and 0.2% NP-40 for 10 min on ice, after which they were disrupted by Dounce homogenization using 15 strokes with pestle “A”. Nuclei were washed in PBS, pelleted and suspended in 0.43 ml of 1.27x NEB Buffer 2 containing 0.38% (wt/vol) SDS. After 1 hour of incubation at 37°C, 10% Triton X-100 was added to a final concentration of 2% for 1 hour incubation at 37°C. Chromatin was then digested by addition of 200 units of HindIII (NEB) for overnight at 37°C, followed by a second addition of 200 units of HindIII for 4 hours. HindIII was inactivated by adding SDS to a final concentration of 1.6% and heating at 68°C for 10 min. Digested chromatin was diluted to 7 mls in 50mM Tris, 10mM MgCl₂ and 1% Triton X-100. After a 1 hour incubation at 37°C, the solution was supplemented with DTT to 10mM and ATP to 1mM, and 200 units T4 ligase (NEB) added for 6 hour incubation at 16°C. Ligated chromatin was incubated with 30 µg proteinase K at 65°C overnight to reverse crosslinks. DNA was then purified by
phenol:chloroform extraction and isopropanol precipitation. 3C products were quantified by Taqman-based real-time PCR with Taqman probes and PCR primers shown in Table 1. To generate control PCR templates, bacterial artificial chromosomes bMQ-440L6 and bMQ-464f17 (Source BioScience) were mixed in equimolar amounts, and were digested and religated. bMQ-440L6 spans proximal Vα/δ gene segments from Trav19 to downstream of Trdv2-2, whereas bMQ-464f17 spans from INT1-2 to the central Jα gene segments. This control template was used to generate standard curves for all 3C-qPCR assays.

3.5 Analysis of Tcrd germline transcription and mRNA expression

For Tcrd germline transcription, RNA was extracted from Rag2-deficient thymocytes. RNA was extracted by using TRIzol reagent and was converted to cDNA using Superscript III and random hexamers according to the manufacturer’s instructions (Life Technologies). To analyze mRNA expression in DP thymocytes, RNA was extracted from sorted DP thymocytes by using TRIzol reagent and was converted to cDNA using Superscript III and Oligo (DT)20 according to the manufacturer’s instructions (Life Technologies). SYBR Green PCR (Qiagen) was performed to quantify cDNA. Amplification of Hprt or Gapdh was used for normalization. PCR primers are listed in Table 2.
Table 1: Primers used for 3C

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>5’-TGCAGTTCCCTGAGTAGAAG-3’</td>
</tr>
<tr>
<td>ii</td>
<td>5’-GGGAAGGTGAAAGCAGATGTAG-3’</td>
</tr>
<tr>
<td>iii</td>
<td>5’-ATCTCTGAGGGTCCAAGGC-3’</td>
</tr>
<tr>
<td>iv</td>
<td>5’-GGGCCAACCAACACTATAC-3’</td>
</tr>
<tr>
<td>v</td>
<td>5’-CCGATAGCAAGGCGAGTTC-3’</td>
</tr>
<tr>
<td>vi</td>
<td>5’-CGAATTGGAAGGACATTTGC-3’</td>
</tr>
<tr>
<td>vii</td>
<td>5’-ACCACAAATCTCTCCTCAG-3’</td>
</tr>
<tr>
<td>viii</td>
<td>5’-TCACACTCTTTCTCTCACTC-3’</td>
</tr>
<tr>
<td>ix</td>
<td>5’-GCAACCAAGAAGCTGACTC-3’</td>
</tr>
<tr>
<td>x</td>
<td>5’-GGACACCAAGATGATCAGA-3’</td>
</tr>
<tr>
<td>xi</td>
<td>5’-GATGTCAGAGCAAGATGAAGC-3’</td>
</tr>
<tr>
<td>xii</td>
<td>5’-GTGCTCTCTCTCTCCACTTC-3’</td>
</tr>
<tr>
<td>xiii</td>
<td>5’-ATGACTGAGTCCTCCCTTCCAG-3’</td>
</tr>
<tr>
<td>xiv</td>
<td>5’-GTTACACTCTGCCATTTCCC-3’</td>
</tr>
<tr>
<td>xv</td>
<td>5’-ATGAGGTCCTCAAGGCCATC-3’</td>
</tr>
<tr>
<td>Dδ2-Jδ1 Viewpoint</td>
<td>5’-CCAAACTTCTCGGTGATCTGT-3’</td>
</tr>
<tr>
<td>Jδ2-Eδ Viewpoint</td>
<td>5’-GAAGATTTGTGGAATTCGACC-3’</td>
</tr>
<tr>
<td>TEA Viewpoint</td>
<td>5’-CACCAACGAAAGCAAGAC-3’</td>
</tr>
<tr>
<td>INT1</td>
<td>5’-ATGACTGAGTCCTCTCCAG-3’</td>
</tr>
<tr>
<td>TEA probe</td>
<td>5’-TCCGAAGGTCTTTCTTCAAGCAGGA-3’</td>
</tr>
<tr>
<td>TRDV2-2 probe</td>
<td>5’-AGAACTCAGTGCTTTGGTCCGTC-3’</td>
</tr>
<tr>
<td>Eδ probe</td>
<td>5’-CCCCAGCGGAAGGTACAGTGCTG-3’</td>
</tr>
</tbody>
</table>
Table 2: Primers used for Tcrd germline transcription and mRNA expression

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRAV15/DV6 F</td>
<td>5’-AGCCTGCTCTGTGTGCTGG-3’</td>
</tr>
<tr>
<td>TRAV15/DV6 R</td>
<td>5’-CAATGCTTCTGTGTGTGATATCCTTAGTTA-3’</td>
</tr>
<tr>
<td>TRDV2-2 F</td>
<td>5’-TGAAAGGTGAGACAGTGCTAGGG-3’</td>
</tr>
<tr>
<td>TRDV2-2 R</td>
<td>5’-AGCAAATGGGTGGGTTTGCCTTCTTA-3’</td>
</tr>
<tr>
<td>Dδ1 F</td>
<td>5’-TACGGGCTGTGTCTTACTGTG-3’</td>
</tr>
<tr>
<td>Dδ1 R</td>
<td>5’-GCTCAATGGACTCTTTGCAGTG-3’</td>
</tr>
<tr>
<td>Dδ2 F</td>
<td>5’-ACAAAGCCCAGGGAAGGTTT-3’</td>
</tr>
<tr>
<td>Dδ2 R</td>
<td>5’-CTGGGAGACGTTCTTCAAC-3’</td>
</tr>
<tr>
<td>Jδ1 F</td>
<td>5’-AGCTGCTGAGGTTTGGGAATG-3’</td>
</tr>
<tr>
<td>Jδ1 R</td>
<td>5’-ATCCCTCAGACCCCTAACCAGA-3’</td>
</tr>
<tr>
<td>Jδ2 F</td>
<td>5’-AACCCTACGGGGCTCCCAAAAG-3’</td>
</tr>
<tr>
<td>Jδ2 R</td>
<td>5’-AAGCTCTGAGGTTTGGGAATG-3’</td>
</tr>
<tr>
<td>Cδ F</td>
<td>5’-GCTTGGTCAGTATGGAGATTCG-3’</td>
</tr>
<tr>
<td>Cδ R</td>
<td>5’-CAGTGTGAACTGTGATCTTGG-3’</td>
</tr>
<tr>
<td>TRDV5 F</td>
<td>5’-AAGCTTCCGAAGGTCAAGAAGA-3’</td>
</tr>
<tr>
<td>TRDV5 R</td>
<td>5’-ACCGACTGGAAGGATGATGTTG-3’</td>
</tr>
<tr>
<td>Yy1 F</td>
<td>5’-GCAGAGTGTGCGAAACGTT-3’</td>
</tr>
<tr>
<td>Yy1 R</td>
<td>5’-CCGAGCAAATCTTCTATTACAACCAG-3’</td>
</tr>
<tr>
<td>Mdm2 F</td>
<td>5’-CCAGGCCAATGTGCAATACCAACA-3’</td>
</tr>
<tr>
<td>Mdm2 R</td>
<td>5’-TGCGCTCAGACGACTTAAACAAC-3’</td>
</tr>
<tr>
<td>Gapdh F</td>
<td>5’-CAGTACATGGCTCATGGTTC-3’</td>
</tr>
<tr>
<td>Gapdh R</td>
<td>5’-CACCAGTAGACTCCACGAC-3’</td>
</tr>
<tr>
<td>Trp53 F</td>
<td>5’-ACGCTTCTCCGAAGACTG-3’</td>
</tr>
<tr>
<td>Trp53 R</td>
<td>5’-AGGGAGCTCGAGGCTGATA-3’</td>
</tr>
<tr>
<td>Bcl2 F</td>
<td>5’-TTATAAGCTGTCACAGAGGGGCTAC-3’</td>
</tr>
</tbody>
</table>
3.6 PCR analysis of Tcr rearrangement

For Tcrd or Tcra rearrangements, genomic DNA was isolated from sorted DN3 or DP thymocytes, respectively. Tcra rearrangements were analyzed in genomic DNA isolated from sorted intracellular (ic)TCRβ⁺ thymocytes. Tcra and Tcrd rearrangements
were quantified by Taqman-qPCR. Tcra rearrangements were quantified by SYBR Green PCR (Qiagen). Data were normalized to unrearranged genomic DNA region at Cd14, and further normalized to plasmids if accurate quantification was required. To generate plasmids containing Trdv2-2 rearrangements, Trdv2-2-Trdd1-Trdd2-Trdj1 rearrangements were amplified from INT1-2<sup>−</sup> DN3 thymocyte using Trdv2-2 and Trdj1 primers and cloned with a TOPO TA Cloning Kit for Sequencing (Life Technologies). To generate plasmids containing Trdv3 rearrangements, double-stranded DNA fragments of Trdv3-Trdd1-Trdd2-Trdj1 were synthesis (gBlock, Integrated DNA Technologies) and cloned with a TOPO TA Cloning Kit for Sequencing (Life Technologies).

Jα usage was analyzed by PCR of cDNA prepared from total thymocytes using Trav12 and Cα primers using following program: 94°C for 2 min, 33–35 cycles of 92°C for 30 s, 55°C for 30 s, and followed by 72°C for 30 s, and 72°C for 4 minutes. PCR products were gel-purified, cloned with a TOPO TA Cloning Kit for Sequencing (Life Technologies), and were sequenced using an internal Cα primer.

**Table 3: Primers used for Tcr rearrangement**

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trbv1</td>
<td>5’-GCCACACGGGTCACTGATAC-3’</td>
</tr>
<tr>
<td>Trbv5</td>
<td>5’-GCCCAGACAGCTCCAAGCTAC-3’</td>
</tr>
<tr>
<td>Trbv12-1</td>
<td>5’-TAGCAATGTGGTCTGGTACCAG-3’</td>
</tr>
<tr>
<td>Trbv12-2</td>
<td>5’-TCTCTCTGTGGCCTGGTATCAA-3’</td>
</tr>
<tr>
<td>Trbv13-1</td>
<td>5’-GGTACAAGGCCACCAGAACA-3’</td>
</tr>
<tr>
<td>Trbv13-2</td>
<td>5’-GCTGGCAGCACTGAGAAAGGA-3’</td>
</tr>
<tr>
<td>Description</td>
<td>Sequence</td>
</tr>
<tr>
<td>---------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Trbv13-3</td>
<td>5’-TTCCCTTTTCTCAGACAGCTGTA-3’</td>
</tr>
<tr>
<td>Trbv16</td>
<td>5’-CTCAGCTCAGATGCCCAAT-3’</td>
</tr>
<tr>
<td>Trbv29</td>
<td>5’-CTGATTCTGGATTCTGCTA-3’</td>
</tr>
<tr>
<td>Trbv31</td>
<td>5’-AAATCAAGCCCTAACCTCTAC-3’</td>
</tr>
<tr>
<td>Jβ1.1 probe</td>
<td>5’-TGTGAGTCTGTTCTTTACCA-3’</td>
</tr>
<tr>
<td>Jβ1.1R</td>
<td>5’-CTCGAATATGGACACGGAGGACATGC-3’</td>
</tr>
<tr>
<td>Trav12</td>
<td>5’-GCAGCAGCTCCTCCATC-3’</td>
</tr>
<tr>
<td>Trav13</td>
<td>5’-AAGAACGTCGCAGCTCTTTG-3’</td>
</tr>
<tr>
<td>Trav14</td>
<td>5’-TGGAGACTCAGCCACCTACT-3’</td>
</tr>
<tr>
<td>Trav17</td>
<td>5’-TGGAGCGACTCAGCCAAGTA-3’</td>
</tr>
<tr>
<td>Trav19</td>
<td>5’-CCAGCCTGAAGACACAGCAG-3’</td>
</tr>
<tr>
<td>Trav21</td>
<td>5’-GTATGGCTTTCTGGGTATTCG-3’</td>
</tr>
<tr>
<td>Jα61</td>
<td>5’-ATGAGTCTTCCAGTCATG-3’</td>
</tr>
<tr>
<td>Jα58</td>
<td>5’-GACTCACTGTGAGCTTTGCC-3’</td>
</tr>
<tr>
<td>Jα56</td>
<td>5’-ACTCAGAACGGTCTTTGACC-3’</td>
</tr>
<tr>
<td>Jα49</td>
<td>5’-GGAATGACAGTCAACTTGTTTC-3’</td>
</tr>
<tr>
<td>Jα40</td>
<td>5’-TGGTACCCTGCCAAACG-3’</td>
</tr>
<tr>
<td>Jα33</td>
<td>5’-ATTAGCTTGGTCACAGCGCC-3’</td>
</tr>
<tr>
<td>Jα31</td>
<td>5’-GGTCCCCATCACAAGGAG-3’</td>
</tr>
<tr>
<td>Jα18</td>
<td>5’-TCCCTAAGGCTGAACCCTCTATC-3’</td>
</tr>
<tr>
<td>Jα17</td>
<td>5’-TGATGGCTAGGCTCCTTTTC-3’</td>
</tr>
<tr>
<td>Jα2</td>
<td>5’-TTACCGTCACCTGGGTCCCTTTCC-3’</td>
</tr>
<tr>
<td>Trav12 (For Jα usage)</td>
<td>5’-CAGACAGAAGGCTGTCAC-3’</td>
</tr>
<tr>
<td>Cα</td>
<td>5’-TGCGGCCGGGTTGCTTTGAAG-3’</td>
</tr>
<tr>
<td>Internal Cα</td>
<td>5’-CGGCACATTGGATTTGGGAGTC-3’</td>
</tr>
</tbody>
</table>
3.7 Chromosomal DNA retention

Genomic DNA isolated from total thymocytes of wild-type and INT1-2-deficient littermates was quantified by SYBR Green real-time PCR (Qiagen). Amplicon abundance in thymocyte DNA was compared with that in kidney DNA using a kidney DNA standard curve. Samples were initially normalized to each other on the basis of the abundance of Ea; retention of the TEA amplicon in wild-type samples was then set as 100%, and amplicon abundance in both genotypes were presented relative to that value. PCR amplicons were located upstream of the identified gene segments. Primers are listed in Table 4. Although Vα-Jα rearrangement will excise amplicons onto extrachromosomal circles, this material should be retained in genomic DNA preparations of DP thymocytes, with the assumption that no thymocyte proliferation occurs after Vα-Jα rearrangement. In practice, and consistent with published work (Livak and Schatz, 1996), we observed that retention of TEA was 50% that of Ea, which indicated that some proliferation occurs after Vα-Jα rearrangement. Loss of signal due to Vα-Jα rearrangement was controlled for by setting the TEA amplicon abundance to 100% in wild-type samples.

Table 4: Primers used for Chromosomal DNA retention

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRDV1 F</td>
<td>5’-CAGTAGGAGGCAATAAGATC-3’</td>
</tr>
<tr>
<td>TRDV1 R</td>
<td>5’-TTGAAGGAACATCCCCATC-3’</td>
</tr>
<tr>
<td>TRDV2-2 F</td>
<td>5’-GGATGGCCATTGAACTCAGAG-3’</td>
</tr>
<tr>
<td>TRDV2-2 R</td>
<td>5’-AAGCCACGTGTGTTGTTATG-3’</td>
</tr>
<tr>
<td>D81 F</td>
<td>5’-GCACTGGCTGTTATGGC-3’</td>
</tr>
<tr>
<td>Primer</td>
<td>Sequence</td>
</tr>
<tr>
<td>---------------</td>
<td>----------------------------------------------------</td>
</tr>
<tr>
<td>Universal Primer</td>
<td>5’-CCGGGAGATCTGAATTCCAC-3’</td>
</tr>
<tr>
<td>Jδ1 SE</td>
<td>5’-CAGTAGGAGGCCAATAAGATC-3’</td>
</tr>
<tr>
<td>Dδ1 SE</td>
<td>5’-TTGAAGGACATCCCCCATC-3’</td>
</tr>
<tr>
<td>LM-PCR-L</td>
<td>5’-GCAGGGGACCCGGGAGATCTGAATTCC-3’</td>
</tr>
<tr>
<td>LM-PCR-S</td>
<td>5’-GAATTCAGATC-3’</td>
</tr>
<tr>
<td>Trδf1 SE Probe</td>
<td>5’-TCCTCTTCTTGGCTGTGATGTCATGCT-3’</td>
</tr>
<tr>
<td>Trδd1 SE Probe</td>
<td>5’-TCACCACCCCTCTGCCATACCCACT-3’</td>
</tr>
<tr>
<td>Trδv5 SE</td>
<td>5’-AACCTGGCATGACTTTCT-3’</td>
</tr>
<tr>
<td>Trδv2-2 SE</td>
<td>5’-CAGCCACCTCTGTCATTCT-3’</td>
</tr>
<tr>
<td>Trδv5 SE probe</td>
<td>5’-CCCTGGGACACAGTTTAGGAGTGC-3’</td>
</tr>
<tr>
<td>Trδv2-2 SE Probe</td>
<td>5’-TGACACAGGCAACAGAGGGGTCTAAC-3’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dδ1 R</td>
<td>5’-GCACAAACACTTCACACTTG-3’</td>
</tr>
<tr>
<td>Dδ2 F</td>
<td>5’-AGGTGACAGCTTCTGCTGCTC-3’</td>
</tr>
<tr>
<td>Dδ2 R</td>
<td>5’-CCCAACTGAGCTGACCAATAG-3’</td>
</tr>
<tr>
<td>Jδ1 F</td>
<td>5’-AGCTGACAGGTGGTTTGGGGAATG-3’</td>
</tr>
<tr>
<td>Jδ1 R</td>
<td>5’-ATCCCTCAGCCACTCAACCCAG-3’</td>
</tr>
<tr>
<td>TEA F</td>
<td>5’-TCCTTTCCAGTTCTTGTGAG-3’</td>
</tr>
<tr>
<td>TEA R</td>
<td>5’-CTCCAGTATGACCTGTTATTGAG-3’</td>
</tr>
<tr>
<td>Ea F</td>
<td>5’-AGGAAGTCGACAGACCTGAA-3’</td>
</tr>
<tr>
<td>Ea R</td>
<td>5’-GAGGGGAGAAGCCCTTTTG-3’</td>
</tr>
</tbody>
</table>
3.8 Chromatin immunoprecipitation (ChIP)

Total thymocytes from Rag2<sup>−/−</sup> mice were cross-linked with 1% paraformaldehyde for 10 minutes at room temperature. Cross-linked cells were lysed in cell lysis buffer (5mM PIPES, pH 8.0, 85mM KCl, 0.5% NP-40). Nuclei were precipitated, washed and lysed in nuclei lysis buffer (50mM Tris-HCl pH8.0, 10mM EDTA, 1% SDS). Chromatin was sonicated by Sonicator 3000 (Qsonica) for 4’33’’, (15s on, 30s off for 6 times). Sonicated chromatin was diluted by 10-fold and precipitated with antibodies against either CTCF (07-729; Millipore), H3Ac (06-599; Millipore) or control rabbit IgG (ab-105-c; R&D Systems). Chromatin-Ab complexes were isolated with Protein A agarose/salmon sperm DNA (Millipore). Next, chromatin was eluted and incubated at 65°C for 4 hours to reverse cross-links. DNA was purified by phenol/chloroform and precipitated by ethanol. Enrichment of chromatin was measured by quantitative PCR as previously described (Shih et al., 2012). Primers are listed in Table 5. Data from CTCF-ChIP and H3Ac-ChIP were expressed as bound/input and then normalized to values for c-Myc and b2-microglobulin (β2M) respectively.

Table 5: Primers used for ChIP

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRDV1 F</td>
<td>5’-CAGTAGGAGGCCAATAAGATC-3’</td>
</tr>
<tr>
<td>TRDV1 R</td>
<td>5’-TTGAAGGAAATCCCATCCCATC-3’</td>
</tr>
<tr>
<td>TRDV4 F</td>
<td>5’-TCCCCAATACTATCTGGCCTG-3’</td>
</tr>
<tr>
<td>TRDV4 R</td>
<td>5’-TCTGCTGAAAGGTGGTGGTTG-3’</td>
</tr>
<tr>
<td>TRDV2-2 F</td>
<td>5’-TCCTGTTTTGAGGTGAGACAG-3’</td>
</tr>
</tbody>
</table>
3.9 Ligation mediated PCR (LM-PCR)

Genomic DNA was prepared from sorted DN3 thymocytes, and linker ligation was performed as described (McMurry et al., 1997; Schlissel et al., 1993). Broken-ended RSS DNA ligated to the linker was detected by Taqman PCR with a signal-end specific primer/Taqman probe and a primer binding to the linker DNA.

3.10 Anti-CD3ε treatment

Mice were injected i.p. with 150 µl of 1 mg/ml anti-CD3 (145-2C11) or with an equal volume of PBS as previously described (Shinkai and Alt, 1994). Mice were euthanized 9 days after injection to isolate DP thymocytes by cell sorting.
3.11 BrdU assay

Mice were injected i.p. with 1 mg BrdU at 2 or 4 hours prior to analysis. BrdU incorporation was detected by intracellular staining (FITC BrdU Flow kit; BD Pharmingen).

3.12 OP9-DL1 culture

OP9-DL1 co-cultures were carried out as previously described (25). DN3a (Lin–CD25–CD44loCD28loFSChi) or DN3b (Lin–CD25–CD44loCD28hiFSClo) thymocytes were sorted and stained with Celltrace Violet (Life Technologies) and were then placed on OP9-DL1 monolayers with 5 ng/ml IL-7. Cells were harvested on days 2, 3 and 4 to measure dilution of Celltrace Violet and expression of CD4, CD8 and CD25. Apoptotic cells were assayed by staining with Annexin V (Biolegend) on day 4.

3.13 Western blot

Antibodies specific for YY1 (H-414, Santa Cruz), p53 (1C12, Cell Signaling Technology), Bcl-xL (54H6, Cell Signaling Technology), Bim (559685, BD Biosciences), caspase-3 (8G10, Cell Signaling Technology), cleaved caspase-3 (5A1E, Cell Signaling Technology) and actin (I-19, Santa Cruz) were used according to the manufacturer’s instructions.

3.14 Statistics

Statistical analyses were performed using Graphpad Prism 6.0 software.
4. A discrete chromatin loop in the mouse Tcra-Tcrd locus shapes the TCRδ and TCRα repertoires

This chapter is slightly modified from the following publication:


The 4C experiments described in 4.2 were performed by Zachary Carico and Han-Yu Shih.

4.1 Introduction

AgR diversification must overcome daunting topological constraints to recruit gene segments for recombination that may be distributed across several Mbs of DNA. Multiple studies have shown that AgR loci undergo large-scale conformational changes during lymphocyte development, bringing distant gene segments into proximity. For example, 3D-FISH has shown that the Igh, Igk, Tcrb and Tcra-Tcrd loci undergo contraction coinciding with the developmental stages during which V(D)J recombination occurs (Roldan et al., 2005; Skok et al., 2007; Jhunjhunwala et al., 2008; Shih et al., 2010). Conversely, loci can be extended to terminate V(D)J recombination, as has been documented for Igh and Tcrb (Roldan et al., 2005; Skok et al., 2007). Dynamic regulation of locus conformation ensures that V(D)J recombination occurs in a developmental-stage specific manner and provides the opportunity for distal V
segments to compete with proximal V segments to ensure the assembly of diverse AgR repertoires.

3C-based assays have shown that AgR loci are demarcated by chromatin loops that juxtapose distant segments of DNA. Although studies have implicated roles for Pax5 and YY1 in Igh loop organization, the primary mediator of chromatin looping at Igh, Igk, Tcra and Tcrb is CTCF (Atchison, 2014; Degner et al., 2011; Gerasimova et al., 2015; Medvedovic et al., 2013; Ribeiro de Almeida et al., 2011; Shih and Krangel, 2013). CTCF is a highly conserved, ubiquitously expressed, zinc-finger-containing transcription factor that binds throughout the genome and mediates long-distance looping between CBEs (Ong and Corces, 2014). CTCF can block, or insulate, enhancer activity by creating DNA loops that separate enhancers from promoters, or can facilitate gene expression by creating DNA loops that juxtapose enhancers and promoters (Ong and Corces, 2014). These two mechanisms account for the known roles of CTCF in V(D)J recombination at AgR loci (Shih and Krangel, 2013). At the Igh locus, IGCR1, an intergenic CBE between the VH and DH arrays, insulates DH-proximal VH gene segments from the influence of the Igh enhancer (Eµ) (Guo et al., 2011b). With IGCR1 deleted, rearrangements are biased towards the hyperactive DH-proximal VH segments and become disordered and lineage-nonspecific. Intergenic CBEs at the Igk locus similarly insulate proximal VK gene segments from Igk enhancers (Xiang et al., 2013, 2014). At the Tcra-Tcrd locus, CTCF marks many important cis-regulatory elements and as a result helps to target the Ea to TEA, and to the promoters of Ja-proximal Va gene segments. These interactions
promote transcription, accessibility and recombination of these Vα and Jα gene segments (Shih et al., 2012). Emerging genome-wide studies also indicate that CTCF-mediated looping may serve a structural or organizing role rather than a direct gene regulatory role (Sanyal et al., 2012; Shen et al., 2012; Phillips-Cremins et al., 2013).

The 1.6 Mb Tcra-Tcrd locus displays a complex organization of gene segments and an intricate program of V(D)J recombination that leads to the development of both γδ and αβ T lymphocytes (Carico and Krangel, 2015). Approximately 100 V gene segments are distributed across 1.5 Mb, with Tcrd D, J, and constant (C) gene segments, and Tcra J and C gene segments clustered within the final 0.1 Mb of the locus (hereafter referred to as the 3′ end of the locus). The majority of V gene segments rearrange to Jα segments in CD4+CD8+ double-positive (DP) thymocytes and contribute to the TCRα repertoire. However, only a few V gene segments rearrange to Dδ and Jδ gene segments in CD4−CD8− double-negative (DN) thymocytes and contribute to the TCRδ repertoire. Several Vδ gene segments (Trdv1, Trdv2-2, Trdv4, Trdv5) are positioned proximal to the DδJδ cluster and are thought to be used exclusively for Tcrd rearrangement. Others (Trav21-dv12, Trav13-4-dv7, Trav6-7-dv9, Trav4-4-dv10, Trav14D-3-dv8, Trav16d-dv11 and the Trav15-dv6 family) are more distal, are interspersed among Vα gene segments, and are used as both Vδ and Vα gene segments (Weber-Arden et al., 2000). How the locus produces a balanced and diverse TCRδ repertoire with representation of proximal and distal Vδ gene segments is unclear.
Here we defined a CTCF-dependent chromatin interaction network that extends across 0.5 Mb of the Tcra-Tcrd locus in DN thymocytes. We identified two intergenic CBEs, INT1 and INT2, that play central roles in this interaction network. INT1 interacts broadly and dynamically across this region of chromatin. However, INT2 specifically interacts with the CBE associated with the TEA promoter, forming a high frequency chromatin loop that segregates Tcrd D, J and C gene segments from most Vδ gene segments. Mice deleted for INT1 and INT2 on both alleles (hereafter referred to as INT1-2KO mice) had a highly restricted TCRδ repertoire, which was strongly biased towards Trdv2-2. This Vδ gene segment is normally segregated from Dδ gene segments by the INT2-TEA loop, but was newly included within the Dδ-containing loop with INT1-2 deleted. Biased Vδ usage resulted not from increased accessibility, but from increased interactions between Trdv2-2 and Dδ gene segments. Of note, the TCRα repertoire was also altered in INT1-2KO mice, implicating heterogeneity of Tcrd rearrangement as a diversifier of Tcra rearrangement. Our results argue that a CTCF-dependent chromatin interaction network creates TCRδ and TCRα repertoire diversity during T cell development.

4.2 Mapping of Long-Range Interactions at the Tcra-Tcrd Locus

We previously found that most CBEs at the Tcra-Tcrd locus are constitutively occupied by CTCF in B cells and DN and DP thymocytes (Shih et al., 2012). A majority of these CBEs are associated with cis-regulatory elements, such as V gene segment promoters, the TEA promoter and Eα. However, we also noted two prominent
intergenic CBEs, INT1 and INT2. We ask whether these CBEs are weaved into a chromatin interaction network that sets the stage for Tcrd rearrangement in DN thymocytes. To map long-range interactions, we performed 4C-sequencing. For this, we compared RAG2-deficient DN thymocytes to control splenic B cells; in both cell populations the Tcra-Tcrd locus is maintained in germline configuration. 4C libraries were prepared using HindIII for the initial chromatin digestion and DpnII for secondary digestion, with the results mapped to individual HindIII fragments. Data from four viewpoints, including TEA, INT1, INT2 and Eδ, are shown (Fig. 4a and b). In DN thymocytes, we found the TEA viewpoint to interact at high frequency with INT2; reciprocally, the INT2 viewpoint interacted frequently with TEA, forming a distinct 80-kb chromatin loop (Fig. 4a). This loop appears to confine almost all additional contacts made by TEA and INT2, since both viewpoints interacted frequently with other sites within the loop but rarely with regions outside of the loop. This loop segregates a subset of TCRδ gene segments, Trdv4, Dδ, Jδ, Cδ and Trdv5, from other gene segments in the locus. Remarkably, although located only 4.7 kb upstream of INT2, INT1 participated in numerous low-frequency, long-range interactions that extend across a 0.5 Mb region from Trav6-7 to Eα (Fig. 4a); this suggests a dynamic loop organization. Eδ almost exclusively interacted with neighboring fragments within the TEA-INT2 loop (Fig. 4a), consistent with studies showing that it can only regulate transcription in the Trdv4-Trdv5 interval (Hao et al. 2011). The interaction profiles of all four viewpoints are lineage-specific, since they were detected in DN thymocytes but not in B cells (Fig. 3a and b);
nevertheless, CTCF binding to the TEA, INT1 and INT2 CBEs is comparable in the two cell populations (Fig. 4a and b). The INT1-2 and TEA CBEs appear to be key nodes in the Tcra-Tcrd locus interactome in DN thymocytes.
Figure 4: Long-range interaction network within the Tcra-Tcrd locus

(a) (Top) CTCF binding to the 3’ portion of the C57BL/6 Tcra-Tcrd locus in DN thymocytes. Several CBEs are labeled above the CTCF track. Black vertical lines below the CTCF track mark gene segments, a subset of which are labeled. (Bottom) Interactomes of TEA, INT2, INT1 and Eδ viewpoints determined by 4C-seq analysis of Rag2−/− thymocytes (C57BL/6 background). (b) (Top) CTCF binding and (Bottom) 4C-seq analyses of splenic B cells. Sequence reads were averaged from two independent experiments for each cell type and were mapped to HindIII fragments. The viewpoint HindIII fragment is marked in red.
4.3 INT1-2KO mice have an altered TCRδ repertoire

To test for a role of the INT1-2 CBEs in Tcrδ rearrangement, we generated an INT1-2-deleted 129 allele in which the 5.8 kb DNA fragment containing both INT1 and INT2 was eliminated in the mouse germ line (Fig. 5a, b and d). Although the absolute number of total thymocytes was mildly reduced in INT1-2KO mice compared to wild-type (WT) littermates, the development of αβ TCR+ thymocytes was largely normal based on staining with anti-CD4 and anti-CD8 antibodies (Fig. 6a and b). DN thymocytes can be subdivided into four successive developmental stages based on staining with anti-CD44 and anti-CD25 antibodies: DN1 (CD25–CD44+), DN2 (CD25+CD44+), DN3 (CD25+CD44–), and DN4 (CD25–CD44–). Percentages of DN1-DN4 thymocytes were comparable between WT and INT1-2KO mice (Fig. 6a). However, as analyzed by using a γδ TCR-specific antibody, the percentage of γδ T cells in INT1-2KO mice was about half that of control littermates (Fig. 6a and b). Moreover, the percentage of Vδ4 usage among γδ TCR+ thymocytes increased by 3-fold in INT1-2KO mice (Fig. 6a and b); Vδ6.3 usage was, however, unchanged (Fig. 7). Therefore, INT1-2KO mice display defective γδ T cell development and a biased TCRδ repertoire that is heavily skewed towards Vδ4.
Figure 5: Generation of INT1-2KO mice

(a) Positions of gene segments (black rectangles), enhancers (black ovals) and CBEs (white ovals) in the 3’ 300 kb of the Tcra-Tcrd locus. (b) The wild-type 129 Tcra-Tcrd allele (top), showing HindIII sites (H) and StuI sites (S), as well as INT1 and INT2 and the Trdv4 segment (above), and probes used for Southern blot (below); the INT1-2-targeting construct (below), with a diphtheria toxin cassette (DT) and a neomycin-resistance cassette (neo'); the INT1-2KO allele with an inserted neo cassette (INT1-2KO neo'); and the INT1-2-deficient allele with deletion of the neo' cassette (INT1-2KO). (c) Southern blot analysis of genomic DNA from wild-type embryonic stem cells (WT) and INT1-2KO embryonic stem cells with targeting of the neo' cassette (INT1-2KO neo'); and the INT1-2-deficient allele with deletion of the neo' cassette (INT1-2KO). (d) Genotyping PCR analysis of wild-type mice and their INT1-2KO homozygous or heterozygous littermates.
Figure 6: Thymocyte development in INT1-2KO mice

(a) Flow cytometry analysis of thymocytes from WT and INT1-2KO littermates. CD4/CD8 and γδ TCR staining are shown for total thymocytes; CD44/CD25 staining is shown for DN thymocytes depleted of CD4⁺ and CD8⁺ cells and pre-gated as follows: 7AAD⁻CD4⁻CD8⁻CD11b⁻Ter119⁻B220⁻Gr-1⁻CD3ε⁻; Vδ4 staining is shown for pre-gated γδ TCR⁺ thymocytes. (b) Number of total thymocytes (left), abundance of γδ TCR⁺ thymocytes as a percentage of total thymocytes (middle) and percentage of Vδ4⁺ thymocytes among pre-gated γδ TCR⁺ thymocytes (right) in WT and INT1-2KO mice. For (b), each data point represents an individual mouse and the horizontal line indicates the mean. Statistical significance was evaluated by Mann-Whitney U-test.
Figure 7: Vδ4 and Vδ6.3 usage in γδ T cells from INT1-2KO mice.

Vδ4 and Vδ6.3 staining is shown for pregated γδ TCR+ thymocytes from WT and INT1-2KO littermates. Percentages of Vδ4+ and Vδ6.3+ γδ cells are indicated. Data are representative of two independent experiments.
Vδ4 is encoded by Trdv2-2, the first functional Vδ gene segment upstream of the INT1-2 CBEs (Fig. 5a). It is one of the Vδ gene segments that is commonly used in the adult TCRδ repertoire (Weber-Arden et al., 2000). We asked whether the skewed TCRδ repertoire in INT1-2KO mice could be attributed to dysregulated Tcrd rearrangements by quantifying Vδ-(Trdd1)Trdd2-Trdj1 coding joints in genomic DNA samples prepared from DN3 thymocytes. Rearrangement of Trdv2-2 was found to be markedly increased in INT1-2KO as compared to WT DN3 thymocytes (Fig. 8a and b). Trdv1, located 67-kb upstream of Trdv2-2, was equally rearranged in WT and INT1-2KO thymocytes (Fig. 8a and b). Rearrangements of Trdv2-1-(Trdd1)Trdd2-Trdj1 were not detected from both WT and INT1-2KO thymocytes (Fig. 8a and b), which was confirmed by PCR-sequencing with a common Trdv2 primer (data not shown). However, all other functional Vδ gene segments that we tested, including Trav15-1-dv6, Trav16d-dv11, Trav6-7-dv9, Trav13-4-dv7, and Trdv5, were substantially less rearranged in INT1-2KO thymocytes (Fig. 8a and b). Notably, Trdv3, a non-coding pseudogene that is rarely rearranged in WT DN3 thymocytes, is substantially rearranged to (Trdd1)Trdd2-Trdj1 in INT1-2KO DN3 thymocytes (Fig. 8c). Nevertheless, usage of Trdv2-2 in Vδ-(Trdd1)Trdd2-Trdj1 rearrangements is over 2 times higher than Trdv3 in INT1-2KO DN3 thymocytes (Fig. 8c). Therefore, Trdv2-2 is predominantly rearranged at the expense of other Vδ gene segments. This bias extended to incomplete Vδ-Dδ rearrangements as well. The Tcrd locus is unusual in that Tcrd rearrangement is unordered and Vδ-Dδ, Dδ-Dδ, and Dδ-Jδ rearrangements can all occur. Strikingly, rearrangement of Trdv2-2-to-(Trdd1)Trdd2
increased, whereas Trav15-1- and Trav5-to-(Trdd1)Trdd2 rearrangements decreased in INT1-2KO thymocytes (Fig. 8d). In fact, Trdv2-2-to-(Trdd1)Trdd2 rearrangements in INT1-2KO mice were as high as in Eδ KO mice, in which partial rearrangements are known to predominate (Monroe et al., 1999b).

To more precisely quantify dysregulation of Tcrd rearrangement in adult thymocytes, we analyzed the retention of Tcrd genomic sequences in preparations of total thymocyte DNA. Deletional rearrangement of Tcrd gene segments in DN thymocytes places intervening signal joint DNAs onto extrachromosomal circles, which will be diluted out and lost during the cell proliferation driven by pre-TCR signaling. In contrast, deletional Tcra gene rearrangement in DP thymocytes should theoretically occur without loss of genomic sequences, because signal joint-containing extrachromosomal circles are generated after the pre-TCR driven proliferative burst. Thus, genomic DNA retention in total thymocytes can faithfully and quantitatively report the spectrum of Tcrd rearrangement events. To ensure accurate quantification of DNA loss due to Tcrd rearrangement, we compared the retention of Tcrd sequences to that of TEA in WT thymocytes, because TEA will be not be excised by Tcrd rearrangements. By measuring the abundance of PCR amplicons situated immediately upstream of the indicated gene segments, we found that wild-type thymocytes had rearranged Trdd1-Trdd2 and Trdd2-Trdj1 intervals on 96% and 90% of alleles, respectively (Fig. 8e). In addition, approximately 28% of wild-type alleles had rearranged Vδ gene segments upstream of Trdv1, approximately 42% had rearranged
Trdv3 or Trdv2-2, and another 25% either had not undergone V-to-D rearrangement or had rearranged Trdv5 by inversion (which would not delete the region upstream of Trdd1) (Fig. 8e). In contrast, INT1-2-deleted alleles displayed impaired Trdd2-to-Trdj1 rearrangement, but increased Trdv2-2(plus Trdv3)-to-Trdd1-to-Trdd2 rearrangements (Fig. 8e). Precocious Vδ-to-Trdd1-to-Trdd2 rearrangements may inhibit Trdd2-to-Trdj1 recombination events on INT1-2-deleted alleles. Because the amplicon upstream of Trdv2-2 was retained on 93% of alleles whereas the amplicon upstream of Dδ1 was retained on only 6% of alleles, the data suggest that Trdv2-2 and Trdv3, together, are involved in partial V-(D)-D or complete V-(D)-D-J rearrangements on the vast majority of INT1-2 deleted alleles. Notably, since Trdv3 rearrangements cannot generate a functional TCRδ chain, cells that use Trdv3 for Tcrd rearrangements will not develop into γδ T cells. Thus, increased Trdv3 rearrangements in INT1-2-deleted alleles may underlie reduced production of γδ T cells in INT1-2KO mice (Fig. 6a and b). INT1-2 deletion also caused increased rearrangement of Trdv2-2 to the most 5’ Jα gene segments in DP thymocytes (Fig. 8f), on alleles that had not undergone Trdv2-2 to Dδ gene rearrangement in DN thymocytes. This may explain slightly reduced retention of the TEA amplicon in INT1-2KO as compared to wild-type thymocytes (Fig. 8e). However, we did not detect premature rearrangement of Trdv2-2 or proximal Vα gene segments to Jα gene segments in INT1-2KO DN3 thymocytes (Fig. 8g). Taken together, these data show that the INT1-2 genomic region is essential to generate a diverse Vδ repertoire.
In the fetal thymus, γδ T cells emerge as early as embryonic day 14. Early fetal Vδ usage is distinct from that of adult since it is strongly biased towards Trdv4 (Vδ1), which is located proximal to Dδ gene segments and within the INT2-TEA loop (Weber-Arden et al., 2000). We asked whether the regulation of Tcrd rearrangement in the fetal thymus was disrupted in INT1-2KO mice. Trdv4 rearrangement was unchanged in INT1-2KO E15.5 fetal thymus (Fig. 8h). In addition, Trdv4 rearrangements were not detected in adult thymocytes from WT and INT1-2 thymocytes (Fig. 8g), indicating that its developmental stage specificity is intact. However, we detected substantially increased rearrangement of Trdv2-2 in INT1-2KO E15.5 fetal thymus (Fig. 8h), indicating that the INT1-2 region helps to limit rearrangements of adult Vδ gene segments in the fetal thymus.
Figure 8: Restricted TCRδ repertoire in INT1-2-deficient mice.

(a) Locus map identifying Vδ gene segments analyzed. Taqman-based qPCR analysis of \(V\delta\rightarrow(Trdd1)Trdd2-Trdj1\) rearrangements (b, c) and \(V\delta\rightarrow(Trdd1)Trdd2-Trdd2\) rearrangements (d) in genomic DNA extracted from DN3 thymocytes from wild-type, INT1-2KO and Eδ KO mice (b, c, d) or from wild-type and INT1-2KO mice. Values were normalized to \(Cd14\) (b, d) or normalized first to \(Cd14\) and then to known quantities of plasmids containing cloned Trdv2-2-\(Trdd1-Trdd2-Trdj1\) or Trdv3-\(Trdd1-Trdd2-Trdj1\) rearrangements (c). (e) Quantitative PCR analysis of rearrangements, assessed by measurement of the retention of chromosomal DNA in total thymocytes (with PCR amplicons upstream of gene segments identified); results were normalized to each other on the basis of the abundance of Eα and are presented relative to retention of the TEA amplicon in wild-type cells, set as 100%. (f) Genomic DNA extracted from DP thymocytes from wild-type and INT1-2KO mice was analyzed for rearrangement of Trdv2-2 to different Jα gene segments by SYBR Green-qPCR, with normalization to \(Cd14\). (g) Genomic DNA extracted from adult DN3 thymocytes from WT and INT1-2KO mice was analyzed for rearrangement of \(V\delta\rightarrow(Trdd1)Trdd2-Trdj1\) or \(V\delta\rightarrow-Traj61\) or -\(Traj58\) by SYBR Green-qPCR, with normalization to \(Cd14\). (h) Quantitative PCR analysis of \(V\delta\rightarrow(Trdd1)Trdd2-Trdj1\) rearrangements in genomic DNA extracted from wild-type and INT1-2KO thymi at embryonic day 15.5 (two-way analysis of variance (ANOVA) with Tukey’s (b, d) or Sidak’s (c, e, g, h) multiple-comparison test). *P < 0.05, **P < 0.01, ***P < 0.001 and ****P < 0.0001. Data are pooled from three independent experiments (b, c, d; mean ± SEM of \(n = 3\) samples (wild-type or INT1-2KO) or 2–3 samples (Eδ KO), each pooled from two to three mice), two independent experiments (e; mean ± SEM of \(n = 3\) samples per genotype, each from one mouse) or two independent experiments (h; mean ± SEM of 3 samples (wild-type) or 4 samples (INT1-2KO), each pooled from two to three mice) (g) Data represent the mean ± SEM of 3 samples for each genotype, with each sample representing an individual mouse.

4.4 No change in chromatin accessibility on INT1-2-deleted alleles

The regulation of RSS accessibility to the recombinase represents a critical control point for V(D)J recombination (Schatz and Ji, 2011). Germline transcription creates accessibility by disrupting nucleosome structure and organization and depositing histone modifications that facilitate RAG binding and RAG enzymatic activity.
We asked whether abnormal \( Tcrd \) recombination on INT1-2-deleted alleles reflected altered germline transcription of \( Tcrd \) gene segments. Germline transcripts were quantified in \( Rag^{2--} \) DN thymocytes carrying wild-type or INT1-2-deleted alleles maintained in unrearranged configuration. We found no differences in germline transcription besides a modest increase at \( Trdj1 \) on INT1-2-deleted alleles (Fig. 9a). Thus, promoter activities were largely unaffected by INT1-2 deletion. We also analyzed histone H3 acetylation (H3ac) by chromatin immunoprecipitation (ChIP). Enrichment of histone H3ac was comparable at all sites examined on wild-type and mutant alleles (Fig. 9b). Hence, INT1-2-deletion did not substantially impact chromatin accessibility of \( Tcrd \) gene segments in DN thymocytes, and in particular, had no effect on V\( \delta \) gene segments.

4.5 INT1-2 deletion changes the chromatin loop organization of the \( Tcra-Tcrd \) locus

We used 3C-quantitative PCR (qPCR) to ask whether INT1-2-deletion generates an altered landscape of long-distance chromatin interactions. 3C libraries were prepared by \( HindIII \) digestion and \( HindIII \) fragments were assayed for interactions with the TEA viewpoint (Fig. 9c and d). Note that although the INT1-2-deleted allele lacks the INT1 \( HindIII \) fragment, it retains the portion of the INT2 \( HindIII \) fragment that includes the primer-binding site. Consistent with the 4C-seq data, the TEA viewpoint strongly interacted with INT2 (fragment xiv) on wild-type alleles (Fig. 9d). However, TEA interacted minimally with the residual INT2 fragment on INT1-2-deleted alleles, instead
interacting frequently with another intergenic CBE, INT3 (fragment vi), located 49 kb upstream of $T_{rdv2-2}$ (Fig. 9d). As a consequence of INT3-TEA interaction, $T_{rdv2-2}$ was confined within a new 250 kb loop that included $T_{crd}\ D$, $J$, and $C$ gene segments, $E\delta$, $T_{rdv5}$, several $V\delta$ pseudogenes, and two $V\delta$ gene segments ($T_{rdv4}$ and $T_{rdv2-1}$) that rearrange minimally in adult DN thymocytes (Data not shown) (Weber-Arden et al., 2000). Moreover, within this loop, TEA interacted more frequently with the region encompassing $T_{rdv2-2}$ and a neighboring CBE (fragments ix to xii) (Fig. 9d). As expected, interaction between TEA and $T_{rdd1}$ (fragment xv) was unaffected by INT1-2 deletion (Fig. 9d).

To test whether this new loop organization facilitates contacts between $T_{rdv2-2}$ (fragment xii) and $D\delta$ and $J\delta$ gene segments, we used fragments D2J1 (containing $T_{rdd2}$ and $T_{rdj1}$) and $E\delta$ (containing $T_{rdj2}$ and $E\delta$) as viewpoints (Fig. 9e). Interactions of $T_{rdv2-2}$ with these viewpoints were substantially more frequent on INT1-2-deleted alleles as compared to wild-type alleles (Fig. 9e). However, as expected, interactions of D2J1 with $T_{rdv5}$ (fragment xvi) and of $E\delta$ with $T_{rdd1}$ (fragment xv) were comparable on wild-type and INT1-2-deleted alleles (Fig. 9e). Therefore, INT1-2-deletion redefines the chromatin interaction landscape in a manner that facilitates contacts between the $T_{rdv2-2}$ and $D\delta$ and $J\delta$ RSSs (Fig 9f).
Figure 9: INT1-2-deletion alters chromatin conformation but not accessibility.

(a) Germline transcription was analyzed in WT and INT1-2KO DN thymocytes (both on a Rag2−/− background). Data represent the mean ± SEM of 2-4 WT and 2-3 INT1-2KO cDNA preparations, each representing a pool of 2-3 mice, with all values normalized to Hprt. Statistical analysis was by 2-way ANOVA with Sidak’s multiple comparison test. (b) Histone H3 acetylation (H3ac) was analyzed in WT and INT1-2KO DN thymocytes (both on a Rag2−/− background). Data represent the mean ± SEM of 2-3 WT and 3 INT1-2KO chromatin preparations, each representing a pool of 8-10 mice, with values of bound/input normalized to values for B2m. Statistical analysis was as in (a). (c) Long-distance interactions analyzed by 3C. CBEs (gray ovals) are indicated on the map. Viewpoint (gray rectangles) and target (numbered black rectangles) HindIII fragments are shown below the map. V gene segments shaded gray are pseudogenes. (d) WT and INT1-2KO DN thymocytes (both on a Rag2−/− background) were analyzed by 3C from the TEA viewpoint. Data represent the mean ± SEM of 3-5 WT and 3-6 INT1-2KO preparations, with normalization to a TEA nearest neighbor fragment. Statistical analysis was as in (a). (e) Similar 3C analyses from D2J1 and Eδ viewpoints. Data for D2J1 represent the mean ± SEM of 3 WT and 3 INT1-2KO preparations, with normalization to interaction between TEA and its neighbor fragment. Data for Eδ represent the mean ± SEM of 4 WT and 3-4 INT1-2KO preparations, with interactions normalized to an Eδ nearest neighbor fragment. Statistical analysis was by unpaired Student’s t-test with Holm-Sidak correction for multiple comparisons. *, P<0.05; **, P<0.001; ***, P<0.0001. (f) Chromatin looping in INT1-2KO mice facilitates Trdv2-2 contacts and rearrangement to Dδ and Jδ gene segments. CBEs (red ovals) are depicted along with their orientations (embedded white arrowheads). In wild-type, stable looping between the TEA and INT2 CBEs is depicted by the thick red line, whereas heterogeneous looping involving the INT1 CBE is depicted by the multiple thin red arrows. In INT1-2KO, the TEA CBE loops to the INT3 and Trdv2-2 CBEs (thick and thin red lines, respectively). Within these loops, Trdv2-2 more frequently contacts Dδ and Jδ gene segments (thin gray lines).

4.6 Partial redundancy between INT1 and INT2

Because the INT1-2 deletion spans 5.8 kb, we could not evaluate the specific contributions of the INT1 and INT2 CBEs to the observed dysregulation of rearrangement and chromatin looping on the mutant allele. To specifically test the INT2 CBE and the INT2-TEA chromatin loop, we generated an allele in which the INT2 CBE
was replaced with a scrambled DNA sequence (hereafter referred to as the INT2M allele; **Fig. 10a, b and c**). CTCF chromatin immunoprecipitation (ChIP) confirmed that CTCF does not bind to the mutant INT2 site (**Fig. 10d**). In contrast to INT1-2KO mice, we observed no change in the number of total thymocytes or the percentage of γδ T cells (**Fig. 11a**). However, INT2M mice had twice as many Vδ4+ γδ T cells as wild-type mice (**Fig. 11a**). Consistent with this result, *Trdv2-2* rearrangement on INT2M alleles increased by 50% relative to wild-type, whereas rearrangement of *Trdv5* and *Trav15-dv6* were each reduced by 50% (**Fig. 11b**). Rearrangement of several other Vδ gene segments was unchanged (**Fig. 11b**). Therefore, INT2M mice partially recapitulate the phenotypic defects observed in INT1-2KO mice. We also measured chromatin interactions on the INT2M allele using TEA, Eδ and Dδ2Jδ1 as viewpoints. Perhaps surprisingly, interaction between TEA and INT2 only decreased modestly on the INT2M allele, whereas interaction between TEA and INT1 doubled (**Fig. 11c**). Elevated interaction with INT1 may explain the relatively modest reduction in TEA-INT2 interaction, given the resolution of 3C. INT2M alleles also displayed moderately increased interactions between TEA and sites upstream of INT1, including INT3 and *Trdv2-2*; similarly, *Trdv2-2* interacted more frequently with D2J1 and Eδ (**Fig. 11c**). However, none of these increases were as substantial as those on INT1-2-deleted alleles. These data suggest that with the INT2 CBE eliminated, the INT1 CBE partially subsumes its function by looping to TEA. However, additional looping to upstream sites allows communication between *Trdv2-2* and Dδ and Jδ segments, leading to increased *Trdv2-2* rearrangement. Together,
these data implicate the INT2 CBE in the dysregulation on INT1-2-deleted alleles, and reveal that INT1 can partially compensate for INT2 when the latter is inactivated.

Figure 10: Generation of INT2 mutant (INT2M) mice.

(a) WT 129/SvJ allele (129), targeting construct INT2M, neomycin-resistant allele INT2M neo', and neo'-deleted allele INT2M are shown. DT, diphtheria toxin cassette; H, HindIII; S, Stul; Southern blot probe is indicated. (b) Southern blot analysis of genomic DNA from WT and INT2M neo'-targeted ES cells. Results are representative of 2 experiments. (c) Genotyping PCR of INT2M heterozygous, WT and INT2M homozygous littermates. Results are representative of >3 experiments. (d) ChIP analysis of CTCF
binding to WT and INT2M alleles in $\text{Rag2}^{+/−}$ and $\text{Rag2}^{+/−}\text{INT2M}$ thymocytes, respectively. $\text{Trdv4}$ served as a negative control. Data represent the mean ± SEM of 3 WT and 2 INT2M samples, with each sample representing a pool of 2-3 mice. Statistical significance was determined by unpaired Student’s $t$-test with Holm-Sidak correction for multiple comparisons. *, $P<0.01$.

**Figure 11: Partial redundancy between INT1 and INT2.**

(a) Number of total thymocytes (left), abundance of $\gamma\delta$ TCR$^+$ thymocytes as a percentage of total thymocytes (middle), and percentage of Vδ4$^+$ thymocytes among pre-gated $\gamma\delta$ TCR$^+$ thymocytes (right) in WT and INT2M mice. Each data point represents an individual mouse and the horizontal line indicates the mean. Statistical significance was evaluated by unpaired Student’s t-test (left) or Mann-Whitney U-test (middle, right). (b) Genomic DNA extracted from DN3 thymocytes from 3-4 week old WT and INT2M mice was analyzed for VDD-$\text{Trdj1}$ rearrangement by Taqman-qPCR, with normalization to $\text{Cd14}$. Data represent the mean ± SEM of 3 WT and 3 INT2M preparations, with each preparation representing a pool of 2-3 mice. Statistical significance was evaluated by 2-way ANOVA with Sidak’s multiple comparison test. (c) Long-distance interactions analyzed by 3C. WT, INT1-2KO and INT2M DN thymocytes (all on a $\text{Rag2}^{+/−}$ background) were analyzed by 3C from the TEA, D2J1 and Eδ viewpoints, with
normalization as in Fig. 9d,e. Data represent the mean ± SEM of 3-5 WT, 3-6 INT1-2KO and 3-4 INT2M preparations for the TEA viewpoint, 4 WT, 4 INT1-2KO and 3 INT2M preparations for the D2J1 viewpoint, and 4 preparations of each genotype for the Eδ viewpoint. All 3C preparations represent pools of 8-10 mice. TEA viewpoint data for WT and INT1-2KO sites vi, x and xiv are identical to Fig. 9d. Statistical significance was evaluated by unpaired Student’s t-test with Holm-Sidak correction for multiple comparisons (TEA viewpoint) or by unpaired Student’s t-test (D2J1 and Eδ viewpoints). *, P<0.05; **, P<0.01; ***, P<0.001; ****, P<0.0001. ND, not determined.

4.7 The TCRα repertoire is altered in INT1-2KO mice

Large Vα and Jα arrays allow for multiple rounds of Vα-Jα rearrangement. Numerous studies support a model of sequential Jα usage in DP thymocytes, with primary rearrangements targeted to the most 5′ (Cα-distal) Jα segments made accessible by TEA promoter activity, and subsequent rearrangements targeted to progressively more 3′ Jα gene segments made accessible by Vα promoters introduced in prior rounds of recombination (Hawwari and Krangel, 2007; Krangel et al., 2004). Accordingly, Vα usage must progress from Jα-proximal to Jα-distal on individual alleles. Numerous studies show that Jα-proximal Vα gene segments (Trav19 and Trav21-dv12) rearrange almost exclusively to 5′ Jα gene segments (Genolet et al., 2012; Jouvin-Marche et al., 2009; Pasqual et al., 2002). This usage is consistent with 3C data indicating that these Vα and Jα gene segments are brought into contact by Eα on unrearranged alleles in DP thymocytes (Shih et al., 2012). However, if primary rearrangement were always to initiate with the most proximal Vα gene segments, combinatorial diversity of the TCRα repertoire would be constrained. Although the most distal Vα gene segments rarely rearrange to 5′ Jα gene segments, members of centrally positioned Vα families often do
We envisage that Vα-Jα combinatorial diversity can be facilitated by heterogeneous Vδ rearrangement in DN thymocytes that variably truncates the Vα array, placing a range of more distal Vα segments in a Jα-proximal position prior to the onset of Vα-Jα recombination. This hypothesis predicts that if Vδ usage were limited to the most proximal Vδ gene segments, as on INT1-2-deleted alleles, combinatorial Vα-Jα diversity would be reduced. To understand the impact of INT1-2 deletion on the TCRα repertoire, we used qPCR to analyze Vα-Jα recombination in genomic DNA isolated from DP thymocytes of wild-type and INT1-2KO mice (Fig. 12). As expected, in wild-type DP thymocytes we found that the most proximal Vα gene segments (Trav21-dv12, Trav19, Trav17) rearranged almost exclusively to the most 5′ Jα gene segments (Traj61, Traj58, Traj56) (Fig. 12a and b). In contrast, central Vα families (Trav12, Trav13, Trav14) rearranged to broadly distributed Jα segments (Fig. 12c). Yet in INT1-2KO DP thymocytes, rearrangement of proximal Vα gene segments to 5′ Jα gene segments was markedly increased, whereas rearrangement of central Vα segments to 5′ Jα gene segments was strongly suppressed (Fig. 12b and c). We conclude that the rearrangement of broadly distributed Vδ gene segments in DN thymocytes provides an important mechanism to diversify the TCRα repertoire.
Figure 12: Restricted TCRα repertoire in INT1-2KO mice.

(a) Partial locus map, with Tcra and Tcrd gene segments denoted above and below the horizontal line, respectively. D and J segments are in black, selected Trav families are color coded, and Trdv segments are in gray. Aligned Trav and Trdv designations indicate V segments designated as Trav-Trdv. Genomic DNA extracted from DP thymocytes from WT and INT1-2KO mice was analyzed for rearrangement of (b) Jα-proximal Vα segments and (c) Jα-distal Vα-segments to different Jα segments by SYBR Green-qPCR with normalization to Cd14. Data represent the mean ± SEM of 3 preparations for each genotype, each preparation representing a different mouse. Statistical significance was evaluated by 2-way ANOVA with Sidak’s multiple comparison test. *, P<0.05; **, P<0.01; ***, P<0.001; ****, P<0.0001. ND, not detected.
4.8 Conclusion

We have defined a CTCF-dependent chromatin interaction network that organizes the 3’ portion of the \textit{Tcra-Tcrd} locus in DN thymocytes. We identified two CBEs, INT1 and INT2, as the key players in this interactome that play critical roles in diversifying the TCR\(\delta\) and TCR\(\alpha\) repertoires. Eliminating INT1 and INT2 from the \textit{Tcra-Tcrd} locus redefined the chromatin interaction network, generating a new loop organization that facilitated rearrangement of \textit{Trdv2-2}, while discouraging the rearrangement of other V\(\delta\) gene segments. Abnormally homogeneous V\(\delta\) usage at INT1-2KO allele subsequently restricted V\(\alpha\) usage during primary V\(\alpha\)-J\(\alpha\) rearrangement in DP thymocytes. Therefore, these data revealed an important and previously unappreciated link between TCR\(\delta\) and TCR\(\alpha\) repertoire diversification.
5 An ectopic intragenic CBE disturbs Vδ usage at the Tcra-Tcrd locus

5.1 Introduction

Eα interacts with the CBE-associated TEA promoter and with promoters of the Cα-proximal Vα gene segments over a distance of 500 kb (Shih et al., 2012). CTCF binding at two CBEs downstream of the core Eα enhanceosome is thought to mediate these interactions (Magdinier et al., 2004). Unlike Eα, which can activate gene segments as far as 500 kb away, Eδ, located within the Jδ-Cδ intron, appears to stimulate local accessibility within a 57-kb region encompassing Trdv4, Dδ1, Dδ2, Jδ1, Jδ2 and Trdv5 gene segments. In fetal thymocytes, Eδ can render accessibility to Trdv4 at a distance of 18 kb, while in adult thymocytes, Trdv5 is the only Vδ segments whose transcription is dependent on Eδ (Hao and Krangel, 2011; Huang and Sleckman, 2007). Whether this limitation is because Eδ lacks a CBE is not clear. To test this, we generated a mouse model in which the Eα-CBE is knocked into the region downstream of Eδ (referred to as CBE KI hereafter). We expected that the Eδ CBE KI interacts with CTCF-bound Vα promoters, which would allow Eδ to extend its influence to more distal V gene segments. However, CBE KI mice exhibited a very specific defect in the rearrangement of Trdv5, a gene segment that is positioned 10 kb downstream of Eδ. 3C assays showed that the CBE KI disrupted interaction between Trdv5 and DδJδ gene segments.
5.2 **Tcrd**, but not *Tcra* rearrangement is defective in CBE KI mice.

To test whether an ectopic CBE downstream of Eδ would alter *Tcrd* rearrangement, we generated the CBE KI allele in which the exact sequence of the Eα-CBE is knocked into the region downstream of Eδ (Fig. 13a, b, c and d). CTCF ChIP confirmed that CTCF does bind to the CBE KI site (Fig. 13e).

To analyze *Tcrd* rearrangement, we quantified Vδ-Trdd1(Trdd2)-Trdj1 coding joints in genomic DNA samples prepared from DN3 thymocytes. The frequency of *Trdv5* rearrangement was markedly decreased in CBE KI mice (Fig. 14a and b). In contrast, all other Vδ gene segments tested, including Trav16d-dv11, Trav6-7-dv9, Trav13-4-dv7, *Trdv2*-2 and two Trav15-dv6 family members, were rearranged normally in CBE KI DN3 thymocytes as compared to wild-type cells (Fig. 14a and b).

It was notable that the rearrangement defect was specific to *Trdv5*, which rearranges by inversion. We wondered whether the rearrangement was impaired at the stage of RAG-mediated cleavage, or at the joining stage of coding joint formation.

To evaluate a potential defect in RAG-mediated DNA cleavage in CBE KI mice, we performed LM-qPCR assay to detect RAG-mediated DSBs at the RSSs of *Tcrd* gene segments (Schlissel et al., 1993). We found comparable DSBs at the 5' *Trdd1*, 5' *Trdj1* and 3' *Trdv2*-2 RSSs in wild-type and CBE KI DN3 thymocytes (Fig. 14c). However, DSBs at the 5' *Trdv5* RSS were markedly decreased in CBE KI DN3 thymocytes (Fig. 14c). This result indication that the CBE KI impaired RAG-mediated DNA-cleavage at the *Trdv5* RSS.
Since the CBE KI is situated in an intergenic region that will not be deleted by 
\textit{Tcrd} rearrangements, it is possible that primary V\textalpha-J\textalpha rearrangements could be affected 
by this ectopic CBE. However, we did not detect any defect in rearrangement of 
proximal and central V\textalpha segments (\textit{Trav12}, \textit{Trav13}, \textit{Trav17}, \textit{Trav19} and \textit{Trav21}) to 5' J\textalpha 
segments (\textit{Traj61}, \textit{Traj58}, \textit{Traj56}, \textit{Traj49}) in CBE KI mice (\textbf{Fig. 15}).
Figure 1

(a) Diagram showing genomic locations and genetic markers.

(b) Detailed diagram of CBE KI targeting construct.

(c) Sequence diagrams of RUNX1, c-Myb, GATA, and CTCF binding sites.

(d) Gel electrophoresis image showing bands at 8.7 kb and 5.8 kb.

(e) Bar graph showing relative CTCF binding levels with statistical significance.
Figure 13. Generation of CBE KI mice.

(a) Positions of gene segments (black rectangles), enhancers (black ovals) and CBEs (while ovals) in the 3’ 300 kb of the Tcra-Tcrd locus. (b) WT 129/SvJ allele (129), targeting construct CBE KI, neomycin-resistant allele CBE KI neo’, and neo-deleted allele CBE KI are shown. DT, diphtheria toxin cassette; B, BglIII; Southern blot probes are indicated. (c) Sequences and relative location of 3’ Eδ and CBE KI are shown, with indication of binding motifs for CBF, c-Myb, GATA and CTCF. (d) Southern blot analyses of genomic DNA from WT and CBE KI neo’-targeted ES cells digested by BglIII. Data are representative of two experiments. (e) Chromatin-immunoprecipitation analysis of the binding of CTCF to Eδ and Eα in WT and CBE KI Tcra-Tcrd alleles of Rag2+/− thymocytes and CBE KI Rag2+/− thymocytes, respectively; Trdv4 serves as a negative control. Data represent the mean ± SEM of 3 WT and 3 CBE KI samples, with each sample representing a pool of 6-8 mice. Statistical significance was determined by unpaired Student’s t-test with Holm-Sidak correction for multiple comparisons. *, P<0.01.
Figure 14. Tcrd rearrangement is disturbed in CBE KI mice.

(a) Locus map identifying Vδ gene segments that were analyzed in (b). (b) Quantitative PCR analysis of Vδ-(Trdd1)Trdd2-Trdj1 rearrangements, with normalization to Cd14. Data represent the mean ± SEM of 3 WT and 3 CBE KI samples, with each sample representing a pool of 3 mice. Statistical significance was determined by unpaired Student’s t-test with Holm-Sidak correction for multiple comparisons. (c) LM-qPCR analysis of DSBs in genomic DNA of DN3 thymocytes sorted from WT and CBE KI mice. Specific primers were used to detect DSBs at the RSS 3’ of Trdv2-2, or 5’ of Trdd1, Trdj1 and Trdv5. DSB frequency was normalized to the level in kidney genomic DNA. Embryonic stem cells (ES cells) served as a negative control. Data represent the mean ± SEM, with each symbol representing an individual mouse. Statistical significance was determined by unpaired Student’s t-test with Holm-Sidak correction for multiple comparisons.
Figure 15. Tcra rearrangement is normal in CBE KI mice.

Genomic DNA extracted from DP thymocytes from WT and INT1-2KO mice was analyzed for rearrangement of Trav12, Trav13, Trav17, Trav19 and Trav21 to different Jα segments by SYBR Green-qPCR with normalization to Cd14. Data represent the mean ± SEM of 3 preparations for each genotype, each preparation representing a different mouse. Statistical significance was evaluated by 2-way ANOVA with Sidak’s multiple comparison test.

5.3 The CBE KI does not perturb chromatin accessibility

We envisaged two possible mechanisms underlying the defective DNA cleavage at Trdv5 in CBE KI mice. One possibility is that the CBE KI insulates Trdv5 from Eδ, resulting in reduced transcription and chromatin accessibility. Another possibility is that
CBE KI interferes with the interaction between Trdv5 and the Dδ-Jδ gene segments, and thus excludes Trdv5 from the Rag-enriched Dδ-Jδ recombination center (Teng et al., 2015).

First, we asked whether impaired Trdv5 recombination on CBE KI alleles reflected altered chromatin accessibility of Tcrd gene segments. Germline transcription was analyzed for wild-type and CBE KI alleles with both maintained in germline configuration on a Rag2−/− background. We found no significant differences in germline transcription at Trdv5 and other Vδ, Dδ and Jδ gene segments, indicating that promoter activities were unaffected by the CBE KI (Fig. 16a). This indicates that the Trdv5 promoter is not suppressed by the CBE KI between Eδ and Trdv5. In addition, enrichment of histone H3Ac was analyzed by chromatin immunoprecipitation (ChIP) in Rag2−/− DN thymocytes carrying wild-type or CBE KI alleles. We found comparable H3Ac enrichment at Trdv1, Trdv2-2, Trdd1, Trdd2, Trdj2 and particularly, Trdv5, on wild-type and CBE KI alleles. Trdj1 displayed less H3Ac on CBE KI alleles (Fig. 16b). However, apparently decreased accessibility at Trdj1 is not biologically significant, since most Vδ-to-Trdd1(Trdd2)-Trdj1 rearrangements are not affected by the CBE KI. We conclude that the rearrangement defect at Trdv5 is not attributed to altered chromatin accessibility in CBE KI DN thymocytes.

5.4 CBE KI mice have an altered chromatin loop organization

We used 3C-qPCR to examine whether the CBE KI alters the long-distance chromatin loop organization of the Tcra-Tcrd locus. 3C libraries were prepared by HindIII
digestion and *Hind*III fragments were assayed for interactions with the Eδ, Dδ2Jδ1 and TEA viewpoints (Fig. 16c and d). Note that the CBE KI is included in the Eδ *Hind*III fragment on the CBE KI allele, but not on the wild-type allele.

To test whether the CBE KI allows Eδ to better interact with other CBEs, we measured the interaction frequency between Eδ and other intergenic CBEs or CBE-containing gene segments, including *Trav19, Trav21, Trdv1, INT3, Trdv2, INT1, INT2, TEA* (Fig. 16d). We found that the Eδ-INT2 interaction was dramatically increased on CBE KI allele (Fig. 16d). However, interactions between Eδ and other CBEs were not affected, and the chromatin loop formed by TEA-INT2 interaction was not affected as well (Fig. 16d). As a consequence of strong Eδ-INT2 interaction on CBE KI alleles, DδJδ gene segments are confined within a 100 kb loop that could exclude *Trdv5*. To test whether this new loop organization prohibits contacts between *Trdv5* (fragment ix) and DδJδ gene segments, we used the fragment Dδ2Jδ1 (containing *Trdd2* and *Trdj1*) as a viewpoint (Fig. 16d). The Dδ2Jδ1 viewpoint had comparable interactions with *Trdv1* on wild-type and CBE KI alleles. However, interaction between the Dδ2Jδ1 viewpoint and *Trdv5* was substantially less frequent on CBE KI alleles as compared to wild-type alleles (Fig. 16d). Therefore, the CBE KI redefines chromatin loops at the *Tcra-Tcrd* locus in a manner that prohibits contacts between the *Trdv5* and Dδ and Jδ RSSs.
Figure 16. CBE KI alters chromatin conformation but not accessibility.

(a) Germline transcription of gene segments in WT and CBE KI thymocytes (all on a \textit{Rag2}^{−/−} background). Data represent the mean ± SEM of two experiments with 3 WT and 3 CBE KI chromatin preparations, each representing a pool of 2-3 mice. Values were normalized to \textit{Gapdh}. (b) Enrichment of H3Ac was analyzed in WT and CBE KI DN thymocytes (all on a \textit{Rag2}^{−/−} background). Data represent the mean ± SEM of 3 WT and 3 CBE KI chromatin preparations, each representing a pool of 6-8 mice, with values of bound/input normalized to values for \textit{B2m}. (c) Map of fragments analyzed by 3C: gray ovals, CBEs; below, viewpoint (gray rectangles) and target (numbered black rectangles) HindIII fragments; the gray V segments are pseudogenes. (d) (left) WT and CBE KI DN thymocytes (all on a \textit{Rag2}^{−/−} background) were analyzed by 3C from the Eδ viewpoint; (right) interactions of DδJδ-\textit{Trdv}1, DδJδ-\textit{Trdv}5 and TEA-INT2 were analyzed by 3C. All 3C data represent the mean ± SEM of three WT and four CBE KI preparations, each pooled from eight to ten mice. Data were normalized to either an Eδ nearest neighbor fragment (left) or a TEA neighbor fragment (right). (a, b, d) Statistical significance was analyzed by 2-way ANOVA with Sidak’s multiple comparison test. (e) Chromatin looping on the CBE KI allele inhibits \textit{Trdv}5 contacts and rearrangement to Dδ and Jδ segments. CBEs (ovals) are depicted along with their orientations (embedded white arrowheads).

5.5 Conclusion

Here, we showed that an ectopic CBE KI at the \textit{Tcra-Tcrd} locus specifically impairs RAG-mediated DNA-cleavage at the \textit{Trdv}5 RSS, resulting in defective rearrangement of \textit{Trdv}5 and a narrowed TCRδ repertoire. This rearrangement defect at \textit{Trdv}5 was not attributed to altered chromatin accessibility. Rather, CBE KI redefines chromatin loops at the \textit{Tcra-Tcrd} locus, which significantly reduced interaction frequency between \textit{Trdv}5 and Dδ and Jδ RSSs. Thus, CTCF regulates V(D)J recombination at the \textit{Tcra-Tcrd} locus through its role as a structural organizer, rather than as a transcriptional insulator. Our data argues that a well-organized CTCF-mediated looping network is critical for diversifying Vδ usage and the TCRδ repertoire.
6 The role of YY1 in thymocyte development.


6.1 Introduction

Although numerous studies have been devoted to understanding the roles of YY1 in B cell development and V(D)J recombination of the *Igh* and *Igk* loci, studies of YY1 in T-lineage cells have been limited to its role in regulating Th2 cytokine production (Hwang et al., 2013). In particular, the role of YY1 in TCR gene rearrangement is unknown. We envisaged that YY1 might cooperate with CTCF to shape the conformation of the *Tcra-Tcrd* locus. In this regard, YY1 is a strong candidate, as it has been documented to regulate *Igh* locus conformation (Atchison, 2014). In fact, *Yy1<sup>f/f</sup>*Lck-cre<sup>+</sup> DP thymocytes did have a skewed *Tcra* repertoire due to defective secondary Vα-Jα rearrangements. However, further analysis attributed this defect to increased apoptosis in DP thymocytes. Following this lead, we found that YY1 played an important role in maintaining the viability of thymocytes through T cell development. We found that early ablation of YY1 caused severe developmental defects in the DN compartment due to a dramatic increase in DN thymocyte apoptosis. Furthermore, YY1 emerged as a novel regulator of the lifespan of DP thymocytes, because late ablation of YY1 resulted in increased apoptosis of DP thymocytes and a restricted TCRα repertoire. Mechanistically,
we showed that p53 was upregulated in both DN and DP YY1-deficient thymocytes. Eliminating p53 in YY1-deficient thymocytes rescued the survival and developmental defects, indicating that these YY1-dependent defects were p53-mediated. We conclude that YY1 is required to maintain cell viability during thymocyte development by thwarting the accumulation of p53.

6.2 Early ablation of YY1 severely blocks DN thymocyte development

To elucidate the role of YY1 in early T cell development, we analyzed Yy1^{f/f} mice expressing a human CD2-Cre transgene. The human CD2-Cre transgene is active in common lymphoid progenitors (de Boer et al., 2003) and should promote Yy1 deletion in all T-lineage cells. These mice are hereafter referred to as Yy1^{CD2} mice. Yy1^{CD2} mice displayed a profound loss in thymocyte number, with thymus cellularity reduced to 1% of that of their wild-type (Yy1^{f/f}) littermates (Fig. 17a). Although all thymocyte subsets were reduced in number (Fig. 17b), there was a dramatic increase in the proportion of DN thymocytes relative to DP thymocytes (Fig. 17b and c), suggesting that the progression from the DN to the DP stage was compromised. The proportion of CD4^{+} thymocytes was increased in Yy1^{CD2} mice (Fig. 17b). However, these thymocytes did not express a surface TCRβ chain, indicating that they were immature (Fig. 17d). Because a similar population of CD4^{+} thymocytes was not detected in Rag2^{+/-} Yy1^{CD2} mice (Fig. 17e), the population detected in Yy1^{CD2} mice must represent bona-fide, post-β-selection
immature single positive thymocytes, rather than pre-β-selection thymocytes with
dysregulated CD4 expression.

Consistent with impaired development of αβ-lineage precursors to the more
mature DP and single positive stages, TCRβ-expressing thymocytes were significantly
reduced in Yy1CD2 mice (Fig. 17f). In contrast, the percentage of γδ T cells was
substantially increased in Yy1CD2 mice compared with wild-type littermates, indicating
that the developmental defect was restricted to the αβ-lineage (Fig. 17f). Further
delineation of DN thymocyte populations showed that the percentage of DN3
(CD25⁺CD44⁻) thymocytes was increased in Yy1CD2 mice, whereas the percentage of DN4
(CD25⁻CD44⁻) thymocytes was markedly reduced (Fig. 17g). To exclude the possibility
that the residual presence of DN4 thymocytes in Yy1CD2 mice was due to incomplete
deletion of Yy1, YY1 expression was measured by ic staining in DN3 and DN4
thymocytes from wild-type and Yy1CD2 mice. YY1 protein was substantially reduced in
both DN3 and DN4 thymocytes from Yy1CD2 mice (Fig. 17h). Consistent with this, only
4% of Yy1 alleles were intact in icTCRβ⁺ thymocytes (Fig. 17i). Moreover, intact Yy1
alleles were essentially undetectable in TCRγδ⁺ thymocytes (Fig. 17i). Hence, efficient
γδ-lineage development and partial αβ-lineage development can occur in the absence of
YY1. Taken together, our results demonstrated that Yy1CD2 thymocytes have a severe,
αβ-lineage–specific developmental defect, which impairs the DN3-to-DN4-to-DP
progression of thymocytes.
Figure 17: Early ablation of Yy1 severely blocks T cell development.

Number of total thymocytes (a) or percentages (left panel) and absolute numbers (right panel) of thymocyte subsets (b) in Yy1<sup>F</sup> (WT) and Yy1<sup>F</sup> CD2-Cre (Yy1<sup>CD2</sup>) mice. Each data point represents an individual mouse and the horizontal line indicates the mean. Statistical significance was evaluated by unpaired Student’s t-test. (c-h) Flow cytometry analysis of thymocytes from WT and Yy1<sup>CD2</sup> littermates. (c) CD4 and CD8 staining is shown for total thymocytes. (d) TCRβ staining is shown for pre-gated CD4<sup>+</sup>CD8<sup>−</sup> thymocytes. (e) CD4 and CD8 staining of total thymocytes from Rag2<sup>+/−</sup> Yy1<sup>F</sup> (Rag2<sup>+/−</sup>) and Rag2<sup>−/−</sup> Yy1<sup>F</sup> CD2-Cre (Rag2<sup>−/−</sup>Yy1<sup>CD2</sup>) mice. Data are representative of two independent experiments. (f) TCRβ and TCRδ staining is shown for total thymocytes. (g) CD44 and CD25 staining is shown for pre-gated CD4<sup>+</sup>CD8<sup>−</sup>Lin<sup>−</sup> thymocytes. (h) Intracellular staining of YY1 in pre-gated DN (CD4<sup>−</sup>CD8<sup>−</sup>Lin<sup>−</sup>), DN3 (CD4<sup>−</sup>CD8<sup>−</sup>Lin<sup>−</sup>CD25<sup>−</sup>CD44<sup>+</sup>) and DN4 (CD4<sup>−</sup>CD8<sup>−</sup>Lin<sup>−</sup>CD25<sup>−</sup>CD44<sup>−</sup>) thymocytes. The control consists of WT thymocytes incubated with anti-YY1 without fluorescent secondary antibody. Data are representative of three (c, d, f, g) or two (e, h) independent experiments. (i) Genomic DNA was extracted from sorted iCTCRβ<sup>+</sup> or TCRγδ<sup>+</sup> thymocytes and deletion of Yy1 exon1 was measured by real-time PCR with normalization to Cd14. Data represent the mean ± SEM of 3 samples for each genotype. Statistical significance was evaluated by unpaired Student’s t-test with Holm-Sidak correction for multiple comparisons. ***P < 0.001, ****P < 0.0001.

A developmental defect at the β-selection checkpoint could reflect impaired Tcrb rearrangement in Yy1<sup>CD2</sup> mice. However, a substantial proportion of DN3 thymocytes expressed iCTCRβ protein in Yy1<sup>CD2</sup> mice (Fig. 18a) and the TCRβ repertoire was minimally altered in these mice (Fig. 18b). We then asked whether DN4 thymocytes in Yy1<sup>CD2</sup> mice had undergone a normal process of β-selection. We tested CD27 expression in DN3 and DN4 thymocytes, because upregulation of CD27 during the DN3-to-DN4 transition marks cells that pass the β-selection checkpoint (Taghon et al., 2006). Although the mean fluorescence intensity of CD27 in Yy1<sup>CD2</sup> DN4 thymocytes was lower than in wild-type thymocytes (Fig. 18c), a substantial portion of Yy1<sup>CD2</sup> DN4 thymocytes had
appropriately upregulated CD27, indicating that they represented bona-fide post–β-
selection thymocytes. To rule out a possible defect in pre-TCR–driven proliferation,
Yy1CD2 mice and their wild-type littermates were pulsed with BrdU for 2 hours, and
incorporation of BrdU into proliferating cells was measured by flow cytometry.
Proliferating icTCRβ+ DN thymocytes were slightly more abundant in Yy1CD2 mice than
in wild-type mice (Fig. 18d), perhaps reflecting a compensatory mechanism in the face of
reduced cellularity. Nevertheless, this result indicated that the proliferative capacity of
pre-TCR competent thymocytes was not impaired by loss of YY1. To further examine the
dynamics of cell proliferation, purified DN3 thymocytes were stained with Celltrace
Violet and co-cultured with OP9-Delta-like 1 (OP9-DL1) stromal cells, which provided
the Notch signaling required for thymocyte development. Yy1CD2 DN3 thymocytes
proliferated with slower dynamics as compared to wild-type DN3 thymocytes (Fig. 18e).
This could reflect a primary defect in proliferation, or alternatively, higher cell death
during proliferation, resulting in lower cell numbers in successive generations.
Figure 18. β-selection in YY1-deficient mice.

(a) icTCRβ staining is shown for pregated CD4−CD8−Lin−CD44− (DN3 and DN4) thymocytes of Yy1f/f (WT) and Yy1f/f CD2-Cre (Yy1CD2) mice (left panels). Mean ± SEM of three independent experiments (right panel). Statistical significance was evaluated by two-way ANOVA with Sidak’s multiple-comparison test. (b) Tcrb rearrangement. Genomic DNA from sorted icTCRβ+ thymocytes of WT and Yy1CD2 mice was analyzed for rearrangement of Vβ-to-Dβ1Jβ1.1 by SYBR Green-qPCR, with normalization to Cd14. Data represent the mean ± SEM of 3 samples for each genotype, with each sample representing an individual mouse. Statistical significance was evaluated by unpaired student’s t-test with Holm-Sidak correction for multiple comparisons. * P<0.05, ** P<0.01, *** P<0.001. (c) Cell surface expression of CD27 analyzed in pregated DN3 and DN4 thymocytes. Data are representative of two independent experiments. (d) WT and Yy1CD2 mice were pulsed with BrdU for 2 h and the percentage of BrdU+ cells was measured in icTCRβ+ DN thymocytes. Data are presented as the mean ± SEM of two independent experiments. (e) sorted DN3a thymocytes were stained with Celltrace Violet and cultured on OP9-DL1 stromal cells (Day 0). The dilution of Celltrace Violet was measured at the indicated time points. Data are representative of three independent experiments.

To test whether increased cell death was responsible for the developmental defect in Yy1CD2 mice, we analyzed the viability of YY1-deficient thymocytes in vitro, because apoptotic thymocytes are often undetected in vivo due to clearance by phagocytosis (Schlegel et al., 2000). Pre−β-selection DN3a thymocytes (7AAD− CD25−CD44−Lin−CD28+) were sorted as previously described (Teague et al., 2010; Yui et al., 2010), stained with Celltrace Violet, and co-cultured with OP9-DL1 cells. After 4 days in culture, DN3a cells from wild-type mice developed into DP cells (Fig. 19a). In contrast, DN3a cells from Yy1CD2 mice failed to generate DP cells (Fig. 19a), in accord with the developmental defects observed in vivo (Fig. 17c). During the same time-frame, wild-type DN3a thymocytes proliferated vigorously, as determined by dilution of
Celltrace Violet, and gradually downregulated CD25 expression as they proliferated (Fig. 19b), consistent with previous analysis (Yui et al., 2010). By day 4, 75% of the cells had downregulated CD25 expression, indicating that these cells had adopted a DN4-DP phenotype (Fig. 19b). In contrast, although DN3a thymocytes from Yy1<sup>CD2</sup> mice proliferated, fewer CD25<sub>low/-</sub> cells were generated (Fig. 19b). To further analyze population dynamics, thymocytes were harvested from culture on day 4, and apoptosis was assayed in gated proliferating cells. The percentages of early apoptotic cells (Annexin V<sup>+</sup>7AAD<sup>-</sup>) and late apoptotic cells (Annexin V<sup>+</sup>7AAD<sup>+</sup>) were determined in both the CD25<sup>hi</sup> and CD25<sub>low/-</sub> populations. Apoptosis was higher in Yy1<sup>CD2</sup> than in wild-type thymocytes, particularly in the CD25<sub>low/-</sub> population (Fig. 19c). Together, these data indicate that YY1 is required for normal thymocyte development because it protects proliferating DN4 thymocytes from apoptosis.
Figure 19. Increased cell death in YY1-deficient DN thymocytes.

Sorted DN3a thymocytes from \( Yy1^{f/f} \) (WT) and \( Yy1^{f/f} CD2-Cre (Yy1^{CD2}) \) mice were labeled with Celltrace Violet and placed in OP9-DL1 co-cultures. CD4 and CD8 expression (a) and CD25 expression and dilution of Celltrace Violet (b) were analyzed at the indicated time points. The results are representative of three independent experiments. (c) Annexin V and 7AAD staining of CD25\(^{hi}\) and CD25\(^{low/-}\) proliferating cells (a combination of the left upper and lower quadrants in (b)) at day 4 of culture.
(left). Mean ± SEM of two independent experiments analyzing the results of three WT and two \(Yy1^{CD2}\) cultures (right). Statistical significance was evaluated by two-way ANOVA with Sidak’s multiple-comparison test.

6.3 YY1 is required for the normal life span of DP thymocytes

We next investigated whether YY1 regulates cell death in DP thymocytes. \(Yy1^{If}\) mice were crossed with \(Lck\)-Cre transgenic mice to generate \(Yy1^{Lck}\) mice. Although \(Lck\)-Cre has been reported to be active in DN2 and DN3 thymocytes (Lee et al., 2001), YY1 protein was not substantially depleted until the DP stage in \(Yy1^{Lck}\) mice (Fig. 20a). We found that \(Yy1^{Lck}\) mice possessed 30% of the total thymocytes detected in their wild-type littermates (Fig. 20b), with normal numbers of DN thymocytes and a normal DN1-to-DN4 progression (Fig. 20b and c). However, the percentage of DN thymocytes was higher in \(Yy1^{Lck}\) mice than in wild-type mice (Fig. 20d), and the absolute number of DP thymocytes was significantly reduced (Fig. 20b). To address whether the loss of DP thymocytes was due to a lower replenishment from proliferating precursors, we tracked DP thymocytes with a recent history of proliferation by labeling with BrdU \textit{in vivo}. Four hours after BrdU injection, the percentage of BrdU\(^+\) DP thymocytes in \(Yy1^{Lck}\) mice was comparable to that in wild-type mice (Fig. 20e). Thus, YY1 deficiency significantly decreased the number of DP thymocytes but did not affect the generation of DP thymocytes from proliferating precursors. To elucidate whether the DP thymocytes in \(Yy1^{Lck}\) mice were reduced because of a survival defect, we cultured wild-type and \(Yy1^{Lck}\) DP thymocytes in vitro to assess cell-autonomous apoptosis as previously described
(Sun et al., 2000). After 24 and 48 hours of culture, there were far fewer viable cells (Annexin V–7AAD–) in cultures from Yy1Lck mice than from wild-type mice (Fig. 20f). Reduced viability of Yy1Lck DP thymocytes was due to increased apoptosis, because these cells exhibited higher levels of caspase 3 cleavage after 6 hours of in vitro culture (Fig. 20g).
Figure 20: YY1 regulates the survival of DP thymocytes.

(a) Western blot of YY1 and actin in purified DN (CD4<sup>−</sup>CD8<sup>−</sup>Lin<sup>−</sup>) and DP thymocytes of Yy1<sup>f/f</sup> (WT) and Yy1<sup>f/f Lck-Cre</sup> (Yy1<sup>Lck</sup>) mice. Results are representative of two independent experiments. (b) Numbers of total, DN and DP thymocytes in WT and Yy1<sup>Lck</sup> mice. Each data point represents an individual mouse and the horizontal line indicates the mean. Statistical significance was evaluated by unpaired Student’s <i>t</i>-test with Holm-Sidak correction for multiple comparisons. (c) CD44 and CD25 staining of WT and Yy1<sup>Lck</sup> thymocytes pre-gated as CD4<sup>−</sup>CD8<sup>−</sup>Lin<sup>−</sup>. Results are representative of three independent experiments. (d) CD4 and CD8 staining of total thymocytes of WT and Yy1<sup>Lck</sup> mice. (e) Frequency of BrdU<sup>+</sup> WT and Yy1<sup>Lck</sup> DP thymocytes following a 4 h pulse with BrdU. The mean ± SEM of three independent experiments is shown. (f) Sorted DP thymocytes were cultured <i>in vitro</i> for 24 or 48 h and stained with Annexin V and 7AAD (left). Mean ± SEM survival is presented for three WT and four Yy1<sup>Lck</sup>
cultures (right). Statistical significance was evaluated by two-way ANOVA with Sidak’s multiple-comparison test. (g) Sorted DP thymocytes were analyzed for caspase 3 and cleaved caspase 3 by western blot either immediately ex vivo or after 6 h of in vitro culture. The wedges indicate 2-fold dilutions of cell extract. Data are representative of two independent experiments. ***P < 0.001, ****P < 0.0001.

Reduced survival of DP thymocytes in vivo should be apparent as a bias in the TCRα repertoire, since Ja usage follows a temporal progression during DP thymocyte development (Guo et al., 2002). Indeed, although rearrangements of Va segments to 5’Ja segments were comparable in wild-type and Yy1Lck DP thymocytes (Fig. 21a), rearrangements of Va segments to more 3’Ja segments were underrepresented (Fig. 21b). Impaired Tcra rearrangement was not due to defective RAG expression (Yannoutsos et al., 2001), because Rag1 and Rag2 gene expression was normal in Yy1Lck DP thymocytes (Fig. 21c). Therefore, we concluded that YY1 is required to protect DP thymocytes from apoptosis in order to generate a normal TCRα repertoire.
Figure 21. YY1 regulates the TCRα repertoire.

Genomic DNA extracted from DP thymocytes from $Yy1^{f/f}$ (WT) and $Yy1^{f/f} Lck$-Cre ($Yy1^{Lck}$) mice was analyzed for rearrangement of V$\alpha$ segments to 5$'$J$\alpha$ segments (a) and to 3$'$J$\alpha$ segments (b) by real-time PCR with normalization to Cd14. Data represent the mean ± SEM of three DNA preparations for each genotype, each preparation representing a different mouse. (c) $Rag1$ and $Rag2$ transcription in DP thymocytes was analyzed by real-time PCR with normalization to $Hprt$. Data represent the mean ± SEM of three cDNA preparations for each genotype, each preparation representing a different mouse. Statistical significance was evaluated by unpaired Student’s t-test with Holm-Sidak correction for multiple comparisons. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$. 

118
To assess whether increased death of DP thymocytes depended on the process of Tcra gene rearrangement or αβ TCR-dependent selection events, we crossed Yy1Lck mice onto a Rag2−/+ background to prevent the generation of αβ TCRs or RAG-dependent DNA breaks. Rag2−/+ background mice were treated with anti-CD3ε to mimic pre-TCR signaling and drive DN-to-DP development in the absence of TCRβ (Shinkai and Alt, 1994). Nine days after injection of anti-CD3ε, YY1-sufficient Rag2−/+ thymocytes expanded by over 100-fold, whereas YY1-deficient Rag2−/+ thymocytes expanded only 30-40-fold (Fig. 22a), with fewer DP thymocytes (Fig. 22b). Moreover, YY1-deficient Rag2−/+ thymocytes were more prone to apoptosis when cultured in vitro (Fig. 22c). Therefore, increased apoptosis of Yy1Lck DP thymocytes was not caused by an altered response to V(D)J recombination-induced double-strand breaks or TCR-dependent selection signals.
Figure 22. YY1 regulates DP thymocyte survival independent of V(D)J recombination or TCR expression.

\( \text{Rag2}^{−/−} \text{Yy1}^{f/f} \) (\( \text{Rag2}^{−/−} \)) or \( \text{Rag2}^{−/−} \text{Yy1}^{f/f} \text{Lck-Cre} \) (\( \text{Rag2}^{−/−} \text{Yy1}^{Lck} \)) mice were injected with anti-CD3ε or PBS and thymocytes were analyzed 9 d later. (a) Total number of thymocytes. Each data point represents an individual mouse and the horizontal line indicates the mean. Statistical significance was evaluated by unpaired Student’s t-test with Holm-Sidak correction for multiple comparisons. (b) CD4 and CD8 staining. Results are representative of three independent experiments. (c) Annexin V and 7AAD staining of DP thymocytes cultured in vitro for the indicated time points (top). Mean ± SEM survival (Annexin V−7AAD+) is shown for three WT and three \( \text{Yy1}^{Lck} \) cultures, each from a different mouse (bottom). Statistical significance was evaluated by two-way ANOVA with Sidak’s multiple-comparison test. *\( P < 0.05 \), **\( P < 0.001 \).
6.4 YY1 regulates the p53-dependent apoptosis pathway

Because YY1-deficient thymocytes underwent cell-autonomous apoptosis in the absence of death receptor stimulation, we assessed involvement of the mitochondrial intrinsic apoptosis pathway. This pathway is primarily regulated by three groups of Bcl-2 family proteins: Bcl-2 homology 3-only apoptosis initiator proteins including PUMA and Bid; pro-survival cell guardians including Bcl-2, Bcl-xL and Mcl-1; and pro-apoptotic effector proteins including Bax and Bak (Czabotar et al., 2014). Additionally, p53, which is negatively regulated by YY1 (Gronroos et al., 2004; Sui et al., 2004), has been connected to the intrinsic apoptosis pathway. p53 not only transactivates several genes encoding Bcl-2 family proteins but also antagonizes Bcl-2 and Bcl-xL and activates Bax and Bak through protein-protein interactions (Chipuk et al., 2004; Hagn et al., 2010; Hemann and Lowe, 2006; Leu et al., 2004; Yu and Zhang, 2005). We found that the transcription of \( Bcl2l11 \) (encoding Bim), \( Bmf \), \( Bid \), \( Bbc3 \) (encoding PUMA), \( Bcl2 \), \( Mcl1 \), \( Bcl2l1 \) (encoding Bcl-xL) and \( Bak1 \) (encoding Bak) was comparable in wild-type and \( Yy1^{Lck} \) DP thymocytes (Fig. 23a). Transcription of \( Birc5 \) (which encodes Survivin, another pro-survival factor in thymocytes), was also unchanged in \( Yy1^{Lck} \) DP thymocytes (Fig. 23b), even though a previous report showed that YY1 represses \( Birc5 \) transcription (Galloway et al., 2014). However, the transcription of \( Trp53 \) (encoding p53), \( Cdkn1a \) (encoding p21) and \( Mdm2 \) was significantly upregulated in \( Yy1^{Lck} \) DP thymocytes (Fig. 23b). These data were consistent with observations in other mammalian cells showing
that Cdkn1a and Mdm2 are direct targets of YY1 (Gronroos et al., 2004; Santiago et al., 2007). Furthermore, we found that the abundance of p53 protein was significantly increased in DN3 thymocytes from Yy1CD2 mice (Fig. 23c), and in DP thymocytes from Yy1Lck mice (Fig. 23d), as compared to wild-type controls. However, expression of Bim and Bcl-xL proteins was normal in Yy1Lck DP thymocytes (Fig. 23d). These results indicated that YY1 is a negative regulator of p53 abundance in thymocytes.

Figure 23. YY1 negatively regulates p53.

The abundance of transcripts encoding pro- and anti-apoptotic proteins, YY1, and p21 were analyzed in DP thymocytes of Yy1WT (WT) and Yy1Lck-Cre (Yy1Lck) mice, with normalization to Hprt (a) or Gapdh (b). Data represent the mean ± SEM of three WT and three Yy1Lck preparations, each from a different mouse. Statistical significance was
We originally hypothesized that YY1 was a candidate regulator of TCR locus conformation and repertoire formation. However, our study revealed major effects of YY1 on T cell development that are unrelated to any direct effects at TCR loci. For example, we only found minor defects in the TCRβ repertoire in YY1-deficient DN thymocytes. Moreover, although we did not measure Vα usage, at least Jα usage was comparable in YY1-sufficient and -deficient thymocytes in which the lifespan defect was rescued by p53 deletion. Although our analyses cannot eliminate the possibility that YY1 may regulate TCR repertoire independent of its role in cell death, our data show no indication of a major role for YY1 as a direct regulator of TCR loci by YY1. Given the absence of clear effects on TCR loci, we explored further in the role of YY1 in cellular development of thymocytes.

To evaluate whether YY1 regulation of p53 expression could account for impaired thymocyte development in the YY1-deficient mice, we crossed Yy1CD2 mice with Trp53+/− mice to generate Yy1CD2Trp53+/− double-knockout mice. Absence of p53 complemented the defect in thymus cellularity in Yy1CD2 mice (Fig. 24a), as well as the losses of DP thymocytes (Fig. 24b; left column) and DN4 thymocytes (Fig. 24b; middle and right columns). Absence of p53 also complemented the defect in Jα usage in YY1-
deficient thymocytes (Fig. 24c). Together, these data demonstrated that overexpression of p53 is responsible for impaired development of YY1-deficient thymocytes.
Figure 24. Absence of p53 rescues the defect in YY1<sup>CD2</sup> mice.

(a) Number of total thymocytes in Yy1<sup>WT</sup>, Yy1<sup>CD2</sup>, Trp53<sup>−/−</sup> and Yy1<sup>CD2</sup> Trp53<sup>−/−</sup> mice. Each data point represents an individual mouse and horizontal lines indicate the mean. Statistical significance was evaluated by one-way ANOVA with Tukey’s multiple-comparison test. **P < 0.01, ***P < 0.001. (b) Staining of CD4 and CD8 in total thymocytes (left column), of CD44 and CD25 in DN thymocytes (CD4<sup>−</sup>CD8<sup>−</sup>Lin<sup>−</sup>) (middle column), and of CD25 and icTCRβ in DN thymocytes (CD4<sup>−</sup>CD8<sup>−</sup>Lin<sup>−</sup>) (right column) are shown. (a, b) Results are representative of two independent experiments. (c) J<sub>α</sub> usage in total thymocytes. Trav12–Cα RT-PCR products obtained from thymocytes of 4- to 5-week-old Trp53<sup>−/−</sup> and Yy1<sup>CD2</sup> mice were cloned and sequenced to evaluate J<sub>α</sub> usage. Data represent the proportions of 70 Trp53<sup>−/−</sup> and 69 Yy1<sup>CD2</sup>Trp53<sup>−/−</sup> clones with the indicated J<sub>α</sub> segments.

6.5 Conclusion

By disrupting Yy1 gene expression at two distinct stages of thymocyte development, we identified a novel function for YY1 in the generation of αβ T lymphocytes. Yy1<sup>CD2</sup> mice displayed increased apoptosis of post-β-selection DN4 thymocytes, leading to severe developmental arrest at the DN4 stage. Yy1<sup>Lck</sup> mice had fewer DP thymocytes due to reduced DP thymocyte lifespan in vivo; consistent with this, YY1-deficient DP thymocytes were prone to cell-autonomous apoptosis when cultured

in vitro. Further analysis revealed elevated levels of p53 protein in both Yy1<sup>CD2</sup> DN and Yy1<sup>Lck</sup> DP thymocytes, and Trp53 gene deletion corrected the blockade of T cell development in Yy1<sup>CD2</sup> mice. Taken together, these data suggested that YY1 plays a critical cell-intrinsic role in thymocyte development by suppressing the level of p53. This function of YY1 is specific to the αβ T cell lineage, because γδ T cell development was normal in Yy1<sup>CD2</sup> mice.
7 Discussion and future directions

The central task of AgR loci is to generate diverse TCR and Ig repertoires. Among various AgR loci, the Tcra-Tcrd locus undergoes a unique and complicated recombination program, with Tcrd, primary Tcra, and secondary Tcra gene rearrangement occurring in a sequential and coordinated manner (Carico and Krangel, 2015). Our lab previously examined the chromatin architecture of the Tcra-Tcrd locus by 3D-FISH. As compared to B cells, the locus adopts a highly contracted conformation in DN thymocytes (Shih and Krangel, 2010). We envisaged that this overall contracted conformation would instigate a CTCF-dependent interactome, as CTCF was previously identified as a critical regulator of long distance interactions and Tcra rearrangement in DP thymocytes (Shih et al. 2012). Indeed, 4C and 3C analyses uncovered an extensive chromatin interactome in DN thymocytes involving a high-frequency (and thus relatively stable) chromatin loop between the TEA and INT2-CBEs, as well as multiple low frequency (and presumably more dynamic) chromatin loops between the INT1-CBE and other sites in the 3’ portion of the locus (Fig. 4).

Following this line of evidence, my thesis work took a genetic approach to study the cis-regulatory function of the CTCF-dependent interactome in Tcrd rearrangement. First, we knocked out INT1 and INT2, two CBEs that serve as important nodes for the interactome at the 3’ portion of the locus. We showed that INT1-2 deficiency redefined the chromatin interaction network, generating a new looping organization. This new loop incorporated Trdv2-2 and Trdv3 into the same chromatin domain as the Dδ-Jδ RC.
As a consequence, these two Vδ segments are strongly biased for Vδ-DδJδ rearrangements at the expense of other Vδ gene segments. To further examine the mechanistic connection between chromatin looping and Vδ usage, an ectopic CBE was knocked into the region downstream of Eδ. As the CBE KI is located within the DδJδ-Trdv5 interval, it created a loop that segregated Trdv5 from the Dδ-Jδ RC, which inhibited synopsis formation between the Trdv5 and DδJδ gene segments, as indicated by the reduced interaction frequency between these gene segments (Fig. 16d). Thus, we conclude that CTCF-dependent chromatin interactions function as a critical rheostat for synopsis frequencies between Vδ and DδJδ gene segments.

CTCF can promote, or inhibit RSS synapsis, depending on the position of CBEs relative to gene segments. In DP thymocytes, CTCF promotes primary Vα-to-Jα rearrangements by mediating interactions between Vα and Jα gene segments (Shih et al., 2012). This is because CBEs are closely associated with proximal Vα segments and the promoter of 5’ Jα segments, TEA. In contrast, INT1-2 inhibits rearrangement of Trdv2-2-to-DδJδ gene segments, because those CBEs are located in-between these gene segments. Notably, looping at the 3’ portion of the Tcra-Tcrd locus in DN thymocytes is very different from that in DP thymocytes. In DP thymocytes, interactions among Eα, TEA and 3’ Vα segments are associated with transcription (Shih et al., 2012), whereas in DN thymocytes, the TEA and INT2 CBEs are all located in transcriptionally silent regions to mediate relatively stable and transcription-independent loops (Fig. 4). Thus, we view the INT2-TEA loop to be primarily structural in nature, preparing the locus for V(D)J
recombination. Whereas locus contraction may rather non-selectively bring all Vδ gene segments into proximity of the Dδ-Jδ cluster, the detailed chromatin loop organization ensures fine-tuning of Vδ rearrangements in DN thymocytes.

CBEs have been shown to organize chromatin interactions at almost all AgR loci, although their distribution is variable among loci. At Igh, Igk and Tcrb loci, a majority of CBEs are located at intergenic regions, whereas CBEs at Tcra-Tcrd mark many critical cis-elements, including gene segment promoters and enhancers (Shih and Krangel, 2013). INT1-2-deleted Tcra-Tcrd alleles behave similarly to IGCR1/CBE-deficient Igh alleles and Cis/Cer-deficient Igk locus, in the sense that they all display dominant contributions of the immediately upstream V gene segments to the respective repertoires (Guo et al., 2011b; Xiang et al., 2013). However, the mechanism underlying the phenotypes of these CBE-deficient alleles is distinct. In contrast to the hyper-activated proximal V gene segments at IGCR1/CBE-deficient Igh alleles and Cis/Cer-deficient Igk alleles, the increased rearrangement of Trdv2-2 on INT1-2-deleted alleles was not attributed to altered accessibility. For the Igh and Igk alleles, CBEs function as insulators to restrict influence of recombination center-associated Eμ and iEκ (Guo et al., 2011b; Xiang et al., 2013). In contrast, INT1-2 did not seem to restrict the influence of Eδ (Fig 9). The distinct role of the INT1 and INT2 CBEs may be attributed to the distinct properties of Eδ as compared to Eα and Eμ. Eα and Eμ are able to directly interact with and activate distal gene segments (Guo et al., 2011a; Shih et al., 2012). Unlike those enhancers, which are capable of long-distance interactions, Eδ seems to have a limited ability to activate
transcription over long distances, since Trdv4 and Trdv5, 57 and 11 kb from Eδ, respectively, are known to be the only two Eδ-dependent Vδ segments (Hao and Krangel, 2011; Huang and Sleckman, 2007). Moreover, Eδ could not function over long distance to activate Dδ and Jδ promoters when placed in the position of Eα (Bassing et al., 2003). This "shortcoming" of Eδ can not be rescued by equipping it with the Eα-associated CBE, which is believed to interact with distal CBEs associated with TEA and numerous Vα gene segments, as the CBE KI allele did not exhibit additional interactions between Eδ and distal CBE-associated V gene segments (Fig 16d). Thus, the CBE KI appears to influence recombination only by disturbing the interaction between Trdv5 and DδJδ, but not by changing the accessibility of any gene segments (Fig 16).

A similar study examined the role of ectopic CBEs in V(D)J recombination of the Tcrb locus (Shrimali et al., 2012). In that study, the Igf2/h19 imprinting control region (H19-ICR) containing four CBEs was knocked into the Tcrb locus in the interval between the PDβ1-Dβ1Jβ1Cβ1 cluster and the PDβ2-Dβ2Jβ2Cβ2 cluster (Shrimali et al., 2012). This allele (termed TCR-Ins) exhibited reduced transcription across the Dβ1-Jβ1 region and reduced Dβ1-to-Jβ1 rearrangement but normal transcription and rearrangement of Dβ2-to-Jβ2 (Shrimali et al., 2012). These defects were attributed to disrupted interactions between PDβ1 and Eβ due to the insulator function of CBEs within the ectopic H19-ICR, which then impaired the accessibility of the Dβ1Jβ1Cβ1 cluster (Varma et al., 2015). In addition, Vβ usage at the TCR-Ins allele was strongly biased towards Trbv31, and there was a drastic rearrangement defect in Trbv13, a Vβ segment upstream of the inserted
CBEs (Shrimali et al., 2012). It was proposed that the ectopic H19-LCR disrupted interactions and synapsis between Trbv13 and D[β2]β2 gene segments. However, 3C analysis of the interaction frequency between RSSs was not performed (Shrimali et al., 2012; Varma et al., 2015). Therefore, our analyses of the INT1-2KO and CBE KI alleles have provided the first evidence that CBE can alter rearrangement, independent of transcriptional insulation, by preventing interaction and synapsis between V and (D)J gene segments.

CTCF binding to the H19-LCR imparts insulator function, whereas CTCF binding to the CBE KI does not. One possibility is that the H19-LCR may recruit other proteins, which cooperate with CTCF to mediate insulation. Another hypothesis to explain this difference is that CBE KI-recruited CTCF may be subject to distinct posttranslational modifications that might decouple its transcriptional insulation and chromatin looping functions. It has been shown that CTCF is regulated by several posttranslational modifications, such as sumoylation and poly(ADP-ribosyl)ation (Ong and Corces, 2014). Poly(ADP-ribosyl)ation is a covalent modification of proteins catalyzed by the poly(ADP-ribose) polymerases (PARPs) (Schreiber et al., 2006). Poly(ADP-ribosyl)ation of CTCF is critical to maintain its insulator function (Witcher and Emerson, 2009; Yu et al., 2004). Inhibition of CTCF Poly(ADP-ribosyl)ation impaired the insulator function of the H19-LCR with concomitant loss of Igf2 gene imprinting (Yu et al., 2004). Whether CBE KI-enriched CTCF is subject to Poly(ADP-ribosyl)ation is
unclear. Further experiments should focus on the regulation of post-translational modifications of CTCF during thymocyte development.

Remarkably, although only 4.7 kb apart, the INT1 and INT2 CBEs have distinct functions. The INT2-TEA interaction forms a high frequency, stable structural loop, whereas INT1 is involved in more dynamic tethering loops. The INT2 and TEA CBEs are convergently oriented. However, INT1, normally excluded from looping to TEA, displayed a diverse array of low frequency interactions with similarly oriented CBEs and other elements. A similar pattern of low frequency interaction was observed in the Igh locus V_11 region, which was proposed to reflect the heterogeneous nature of loops at the single cell level (Medvedovic et al., 2013). It is reasonable to believe that INT1-mediated loops help DßJß segments to tether a heterogeneous spectrum of Vß gene segments for synapsis. To date, the looping organization of the Vα/δ region of the Tcra-Tcrd locus is not well studied. Considering that looping structure of the Vα/δ region might be highly dynamic, future experiments should take advantage of the recently developed single cell Hi-C technique to tease out structural fluctuations within this region (Nagano et al., 2013; Nagano et al., 2015).

In our model, the distinction between stable and dynamic looping seems to be dictated by CBE orientation (hereafter indicated as “<“ or “>”). Genome-wide analysis recently revealed a strong preference for looping between convergently (“>“ and “<“) oriented CBEs (de Wit et al., 2015). Indeed, our data have shown that TEA (\(<\)) strongly interacts with its neighboring and convergently orientated INT2 (\(>\)), and that INT1 (\(>\)) is
normally outcompeted by INT2 (>) for the TEA-CBE (<). The INT1-TEA interaction only
becomes apparent by mutation of INT2. However, solely attributing looping to the
orientation of paired CBEs is overly simplistic. At the Tcra-Tcrd locus, INT1 and INT2
shared an orientation with 87% of CBEs, whereas the TEA-CBE (<) is the only CBE in the
reverse orientation within the 3’ 100 kb region. Strikingly, although the forward
orientation Eα (>) has a divergent orientation relative to TEA (<) and shares the same
orientation with INT1 (>) and INT2 (>, it strong interacts with TEA (<), INT1 (>) and
INT2 (>) in DP thymocytes (Shih et al., 2012 and data not shown). Therefore, a
convergent CBE orientation is not the only determinant for looping. A formal test of the
CBE-orientation hypothesis involves flipping the orientation of CBEs. For example, an
engineered INT1-CBE with reversed orientation could be useful to test whether the
tethering function of INT1 is dependent on its orientation. We noted ~14 CBEs with the
(>) orientation located within the distal and central portions of the V array. The function
of these CBEs is intriguing, as they may function like TEA (<) to serve as local
“interactome nodes”. The interactome of these CBEs could be examined by 4C analysis
using these CBEs (<) as viewpoints. Functions of these CBE can be determined using
CBE-deleted alleles.

Notably, Although INT1 and INT2 are similarly oriented, INT1 appears unable
to fully assume the stable looping function of INT2, because INT2M alleles displayed
elevated looping between TEA and upstream sites (eg., INT3). Forcing INT1 to at least
partially assume the structure role of INT2 on INT2M alleles may compromise the
dynamic tethering function of the INT1 CBE, because INT2M mice partially recapitulated the defect in Tcrd rearrangement in INT1-2KO mice.

What is the physiological function of structural loops at the Tcra-Tcrd locus? We argue that chromatin loops primarily play a role in buffering against the disparity in accessibility, distance and other intrinsic properties of various V gene segments. The INT1-2 allele provided one example: the inherently hyper-accessible Trdv2-2 and the pseudogene Trdv3 are physically segregated from Dδ and Jδ segments by the INT2-TEA loop on wild-type alleles, whereas Trdv5, an intrinsically less accessible but functional Vδ, appeared to require an additional boost for equal rearrangement. Incorporating Trdv2-2 and Trdv3 into the same loop as Trdv5 provided them with a prominent advantage over Trdv5 for synapsis with Dδ RSSs. Consistent with this, the ectopic CBE on the CBE KI allele segregated Trdv5 from the DδJδ chromatin domain, leading to defective rearrangement of Trdv5. For those more distal Vδ gene segments, tethering loops provided by INT1 may be required to compensate for the obvious distance disadvantage. The central goal of the looping network seems to provide an equal opportunity for all Vδ to be able to access the DδJδ recombination center in order to generate a balanced Tcrd repertoire. Therefore, both INT1-mediated tethering and INT2-mediated stable loops are required for a normal TCRδ repertoire.

The defect in γδ T cell production in INT1-2KO mice reflects, at least in part, an overall reduction in productive Vδ-Dδ-Jδ rearrangements, because we observed a drastic increase in rearrangement of a non-coding pseudogene, Trdv3 (Fig. 8c). In addition,
reduced γδ T cell numbers may also reflect a constraint on selection of the narrowed TCRδ repertoire. In fact, although INT1-2KO DN3 thymocytes displayed dramatically reduced Trav15-dv6 rearrangements, Vδ6.3+ cells still represented 15% of total γδ T cells in these mice (Fig. 7). In this sense, the percentage of Vδ4+ cells in the γδ T cell repertoire of INT1-2KO mice (55%, Fig 6) may not accurately reflect the extent to which Trdv2-2 rearrangements dominate the Vδ repertoire of DN3 thymocytes. A thorough analysis of TCRδ repertoire should be performed to understand how defective Tcrd rearrangement in INT1-2KO mice affects the TCRδ repertoire.

The Vα-to-Jα rearrangements follow a coordinated, reciprocal progression. 3’ Vα gene segments are preferentially rearranged to 5’ Jα gene segments (Ja61-Ja48), saving the 5’ Vα segments for 3’ Jα gene segments (Ja38-Ja16) (Carico and Krangel, 2015). However, while this sequential manner of rearrangement insures efficient sampling of the 61 Jα and >100 Vα gene segments, it actually restricts the combinatorial diversity of Vα-to-Jα rearrangements. Tcra repertoire analysis from INT1-2KO mice revealed a novel mechanism underlying Tcra repertoire diversity. We observed an altered Tcra repertoire, which reduced usage of central Vα segments with 5’Ja segments. As INT1-2 is located in a region that will be deleted by the majority of Tcrd rearrangements, it is unlikely to directly regulate Tcra rearrangements in DP thymocytes (Shih et al., 2012). Therefore, the altered TCRα repertoire in INT1-2KO mice is thought to be due to the outcome of biased Tcrd rearrangement. Most likely, in wild-type alleles, primary Vα-to-Jα rearrangements use the most 3’ Vα and most 5’ Jα gene segments (Carico and Krangel, 2015). However,
the availability of the most 3' Va gene segments is influenced by heterogeneous rearrangement of Vδ gene segments that are interspersed among Va gene segments. This should variably truncate the Va/Vδ array prior to Tcra rearrangement. In INT1-2KO mice, the majority of alleles rearrange the most proximal Vδ segments (Trdv2-2 and Trdv3) and homogeneously retain the most 3' Va gene segments (Trav17, Trav19, Trav21). These Va gene segments are preferentially used for subsequent primary Tcra rearrangements, leading to a more restricted TCRα repertoire. Thus, our data argue that Tcrod rearrangement is an inherent diversifier of the TCRα repertoire. These data not only revealed a previously unappreciated link between Tcrod and Tcra rearrangement, but also provided a rationale for having Vδ and Va gene segments in a single array in a nested configuration.

Remarkably, the chromatin interaction network of DN thymocytes is completely absent in the decontracted Tcra-Tcrod locus in B cells, even though the relevant CBES are occupied by CTCF in these cells. This suggests that CTCF is not the only factor that instigates conformational changes. We envisage that chromatin interactions at the Tcra-Tcrod locus are cooperatively mediated by several trans-factors in a fashion similar to the Igk locus (Fig. 3). In this context, we sought to identify additional factors that may share the responsibility with CTCF to organize chromatin structure.

Because it had been shown that YY1 plays critical roles in V(D)J recombination of the Igk and Igk loci (Gerasimova et al., 2015; Liu et al., 2007a; Medvedovic et al., 2013; Pan et al., 2012; Verma-Gaur et al., 2012), we initially considered YY1 as a candidate...
regulator of Tcr loci. However, we detected no substantial changes in Tcrb rearrangement in Yy1CD2 mice (Fig. 18b). Moreover, defects in Tcra rearrangement in Yy1Lck mice were attributed to impaired viability of DP thymocytes, which was rescued when YY1-deficiency was analyzed on a Trp53−/− background (Fig. 21 and Fig. 24c). Furthermore, we detected no obvious effect of YY1 deficiency on γδ T cell development (Fig. 17f).

Although we cannot completely rule out a potential role of YY1 for Tcr gene rearrangement, we believe that any direct effects of YY1 on Tcr loci are relatively subtle as compared to those at Ig loci.

Our data showed that YY1 suppresses Trp53 transcription (Fig. 23b). However, we think that YY1 may also regulate p53 expression at the post-transcriptional level, since other groups used cell lines to show that YY1 negatively regulates p53 post-transcriptionally by stabilizing the interaction of p53 with Mdm2, an E3 ubiquitin ligase (Gronroos et al., 2004; Sui et al., 2004). This interaction is known to be required for ubiquitination and proteasomal degradation of cytoplasmic p53 (Kubbutat et al., 1997; Marchenko et al., 2010). One can test whether YY1 regulates p53 through a similar mechanism in DP thymocytes by co-immunoprecipitation of p53 and Mdm2 in thymocytes that are treated with proteasome inhibitors.

In summary, this thesis work extended the understanding how chromatin conformation regulates V(D)J recombination at the Tcra-Tcrd locus and provided fundamental insights into basic mechanisms of how CTCF-dependent chromatin architecture dictates genome function at AgR loci. In addition, we investigated the role
of YY1 in early T cell development and identified YY1 as a gatekeeper for the viability of
DN and DP thymocytes. We further demonstrated that YY1 regulates thymocyte
development by setting the threshold of p53 expression and p53-dependent apoptosis.
This work is the first to link the YY1-p53 pathway to T cell development.
Bibliography


Hayday, A.C., and Pennington, D.J. (2007). Key factors in the organized chaos of early T cell development. Nature Immunology 8, 137-144.


implication for a limited T cell receptor alpha chain repertoire. The Journal of Experimental Medicine 196, 1163-1173.


Biography

Liang Chen

Education:

2003-2008, Bachelor of Medicine  The Fourth Military Medical University, Xi’an, China
2010-2016, Ph.D. of Immunology  Duke University, Durham, North Carolina, USA

Publication:


