Regulation of the Alpha and Delta T Cell Receptor Repertoires

by

Danielle Jean Dauphars

Department of Immunology
Duke University

Date:_____

Approved:

_____
Michael S. Krangel, Supervisor

_____
You-Wen He

_____
Francis Ka-Ming Chan

_____
Qi-Jing Li

_____
Micah A. Luftig

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

2021

ABSTRACT

Regulation of the Alpha and Delta T Cell Receptor Repertoires

by

Danielle Jean Dauphars

Department of Immunology
Duke University

Date:_____
Approved:

_____
Michael S. Krangel, Supervisor

_____
You-Wen He

_____
Francis Ka-Ming Chan

_____
Qi-Jing Li

_____
Micah A. Luftig

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

2021

Abstract

The adaptive immune system, comprised of B and T lymphocytes, responds with unique specificity to antigens. Specificity is imparted by the process of V(D)J recombination. During V(D)J recombination, gene segments spread across genomic space are assembled to generate unique antigen receptors by a cut-and-paste reaction in somatic cells. The mechanisms that regulate V(D)J recombination remain unclear.

In T cells, two receptor classes are possible. These TCRs, the αβTCR and γδTCR, are assembled in part from genes sharing a single genetic locus. This single locus for both the T cell receptor α and T cell receptor δ (*Tcra-Tcrd*) chains has a unique nested structure that makes its regulation and recombination particularly interesting. In the *Tcra-Tcrd* locus, V_δ segments are interspersed among the V_α segments. Rearrangement of *Tcrd* precedes *Tcra* recombination, and since these events are on the same locus, *Tcrd* recombination was hypothesized to diversify the starting point for primary *Tcra* recombinations, which are the first of a series of sequential recombination events possible for functional TCRα chain generation. We generated a novel mouse model genetically modified to be incapable of *Tcrd* recombination. Using this model, we have determined that *Tcrd* recombination indeed diversifies the *Tcra* repertoire. Particularly, primary rearrangements are skewed toward more proximal V_α segments, while later rounds of rearrangement proceed as expected from the restricted primary V_α repertoire.

We also investigated the specific impact of rearrangement to two relatively distal V_δ segments of particular interest, *Trav15d-1-dv6d-1* and *Trav15-1-dv6-1*, on the *Tcra* repertoire. To understand the significance of rearrangements to these V_δ segments, we used two lines of mice lacking each of these segments. We found that primary *Tcra* recombinations utilizing V_α segments immediately upstream of each deleted gene segment were reduced. In the case of deletion of *Trav15d-1-dv6d-1*, the most distal V_δ segment, we observed a reduction in primary recombinations using all upstream V_α segments. In the case of *Trav15-1-dv6-1* deletion, primary recombinations upstream of this segment were reduced, but utilization of V_α segments 5′ of the most proximal upstream V_δ segment was not disrupted. Unexpectedly, we also observed a reduction in secondary recombinations to V_α segments immediately upstream of either *Trav15d-1-dv6-1* or *Trav15-1-dv6-1* when either segment was deleted. These gene segments or their surrounding regions appear to be important in promoting diversity of the *Tcra* repertoire not only by their use as *Tcrd* recombination substrates, but also by a mechanism that remains to be explored.

The structure of the *Tcra-Tcrd* locus also plays a role in recombination. Early in thymocyte development, the *Tcra-Tcrd* locus is contracted. During this period, *Tcrd* recombinations are capable of assembling segments distal in linear space. Later in development, when *Tcra* recombines, the locus is decontracted, and rearrangements generally proceed using the most proximal available segments. We and others have previously demonstrated that chromatin topology has a major impact on transcription and V(D)J recombination. Control of locus conformation is heavily influenced by contacts

between CCCTC binding factor (CTCF)-bound elements. How CTCF regulates recombination at the *Tcra-Tcrd* locus is only partially understood. Herein, we used a novel mouse model lacking an intergenic CTCF binding element (CBE) in the *Tcra-Tcrd* locus ('INT1') to determine the impact of this specific CBE on *Tcrd* recombination. We find that loss of INT1 skews *Tcrd* rearrangements toward utilization of proximal V_δ segments. These results suggest that INT1 serves as an important structural feature facilitating varied *Tcrd* recombination.

Dedication

For my parents, Daniel and Jill Dauphars, who have always pushed me to succeed, been my greatest supporters, and given me the freedom to learn and grow.

Contents

List of Tables

List of Figures

Acknowledgements

Many colleagues in the Duke University Department of Immunology contributed to the success of this project. Foremost, Dr. Krangel has been the best mentor one could wish for in every conceivable way; it is my good fortune to have been his trainee. Members of the Krangel lab past and present have supported me technically, scientifically, and personally. My thesis committee provided exceptional feedback and encouragement throughout my time at Duke. Innumerable friends and family members sustained me during my studies.

Many thanks to the Transgenic and Knockout Mouse Shared Resource at Duke, particularly Cheryl Bock and Gary Kucera, who aided in generation of novel mouse models used throughout this work. Yuan Zhuang and Ariana Mihai provided mice and critical experimental design advice for Chapter 4. Liuyang Wang provided biostatistics assistance for the experiments in Chapter 4. Zachary Carico was instrumental in designing the 5'RACE methods in Chapter 4, and his findings were foundational to the hypotheses tested here. Liang Chen conducted research underlying the experiments in Chapter 5. The Duke Cancer Institute Flow Cytometry Core and DNA Analysis Facility provided technical support, and I could not be more grateful to Nancy Martin and the late Dr. Mike Cook, without whom none of this work would be possible.

Finally, my jaan, Nandan Gokhale, has been beside me on this journey, my partner at every step. He has assisted me with science, dog parenting, terrible puns, and by bringing joy and curiosity to every facet of my life.

1. Introduction

1.1 Lymphocyte development

1.1.1 Innate and adaptive immunity

Multicellular organisms have developed immune systems to protect themselves from infectious agents. To recognize and defend against pathogens, a multifactorial approach is applied. The immune system is broadly divided into two components. The innate arm is quick to respond to non-specific evolutionarily-conserved pathogen-associated features (Janeway, 1989; Medzhitov, 2013). The hallmarks of the adaptive arm, however, are its specificity and memory formation capacity. Although adaptive immunity is more specific, it acts slowly and thus serves as a second arm of defense after the initial innate immune response.

The innate immune system is comprised of barrier, secreted, and cellular components. Innate immune cells differentiate self from non-self using pattern recognition receptors (PRRs). These germline-encoded PRRs sense conserved molecular patterns found on broad classes of pathogens. Structural motifs shared among members of different groups of pathogens can be recognized by PRRs (Brubaker et al., 2015). Recognition allows clearance by phagocytes and sequestration of pathogens to prevent their spread across tissues. PRRs sense non-self but are not specific to individual pathogens. This is in direct contrast to the way adaptive immune cells recognize pathogens. In addition to its function as a barrier and in directly clearing pathogens, the innate immune system serves as an important support to the adaptive immune system by

presenting antigens derived from infectious agents to adaptive immune cells (Iwasaki and Medzhitov, 2015).

Two lymphocyte cell types, B cells and T cells, are responsible for adaptive immunity. Each B cell or T cell expresses a unique antigen receptor (AgR), called a B cell receptor (BCR), or T cell receptor (TCR), respectively. Adaptive AgRs differ from PRRs in that they are not germline encoded; their recognition is specific and a strong response is sustained by memory cells after an initial infection (Chaplin, 2010). Indeed, to achieve the diversity of AgRs found on B and T lymphocytes, a unique mechanism is required. A process of somatic gene recombination, termed V(D)J recombination, which recombines variable (V), diversity (D), and joining (J) gene segments, allows the generation of trillions of BCRs and TCRs that are not germline encoded (Nikolich-Žugich et al., 2004; Roth, 2014). Due to the inherent danger of somatic V(D)J recombination, complex regulatory mechanisms are essential to both facilitate the generation of diverse, functional AgRs and to prevent aberrant rearrangements.

1.1.2 B cell development

The two major classes of lymphocytes, B and T cells, are both derived from common lymphoid progenitors, a subset of hematopoietic stem cells of the bone marrow. B cells develop in the bone marrow, while T cells develop in the thymus. In the bone marrow, common lymphoid progenitors have the potential to adopt the B, T, or NK cell lineages. B lineage commitment is signaled by expression of B220. These pre-pro-B cells have committed to the B cell fate but have very limited recombination of

immunoglobulin (*Ig*) genes (Hardy and Hayakawa, 2001). Biallelic D-to-J_H recombination of the *Igh* locus occurs as RAG as upregulated in early pro-B cells. V-to-DJ_H recombination then takes place on one allele at a time after D-to-J_H recombination (Alt et al., 1984; Jung et al., 2006), begins in the late pro-B stage. *Igh* rearrangement success leads to the expression of the pre-BCR in large pre-B cells; μ heavy chain pairs with λ5 and V-pre-B, the surrogate light chain (Karasuyama et al., 1993, 1990). The pre-BCR serves as a checkpoint in B cell development; signaling through the pre-BCR is essential for B cell development and the proliferative burst as cells transition to the small pre-B cell stage (Karasuyama et al., 1996; Hardy and Hayakawa, 2001). At this stage light chain, *Igk* and *Igl*, recombinations take place. *Igk* recombination typically occurs first, with *Igl* recombination occurring primarily in cells with unproductive or deleted *Igk* recombinations (Wang et al., 2020; Gorman and Alt, 1998; de Almeida et al., 2015).

Successful light chain rearrangement and coupling of the new light chain with the heavy chain leads to IgM receptor, a type of BCR, expression on the surface of new entrants into the immature B cell stage. Selection then occurs to delete B cells with BCRs that are highly responsive to self-antigen. Alternatively, B cells are capable of upregulating RAG for receptor editing to generate a BCR that is not self-reactive. *Igk* recombinations can proceed until the locus is depleted of segments. The *Igl* chain can also undergo recombination to generate a BCR that is not self-reactive. B cells with only mild recognition of self-antigen can become anergic (Wang et al., 2020). If the cell passes this selection phase with a BCR unresponsive to self-antigen, these cells, now in the transitional B cell stage, leave the bone marrow to further differentiate into mature B

cells. B cells are also capable of alternative splicing of *Igh* transcripts to remove the transmembrane domain, allowing secretion of a non-membrane-bound BCR, which is referred to as an antibody.

1.1.3 T cell development

The two classes of T cells develop in the thymus; they are differentiated by surface TCR expression as γδT cells or αβT cells. While many progenitors are capable of generating T cells, those that colonize the thymus are the primary physiological progenitors of T cells (Carpenter and Bosselut, 2010). The stem cells that develop into T cells do not arise in the thymus (Donskoy and Goldschneider, 1992). In fetal life, they arise from fetal liver, but in adult life they arise from bone marrow common lymphoid progenitors (Itoi et al., 2001; Bhandoola and Sambandam, 2006; Lind et al., 2001). When multipotent stem cells arrive in the thymus, they are referred to as early thymic progenitors (ETPs). ETPs maintain the potential to give rise to many immune cell types, but primarily retain T and myeloid cell potentials (Bell and Bhandoola, 2008; Wada et al., 2008). A basic helix-loop-helix transcription factor, E2A, is expressed highly in cells selecting the T cell fate, leading to commitment (Engel et al., 2001). Developmental staging of thymocytes is typically determined by cell surface marker expression patterns which change over the course of T cell ontogeny (**Figure 1**). ETPs make up a portion of the earliest of four $CD4^{-}CD8^{-}$ (DN) thymocyte stages, referred to as DN1. DN staging is based upon expression of CD44 (or c-kit, also known as CD177) and CD25, and DN1 cells are $CD44^{+}CD25^{-}$. As thymocytes mature they acquire CD25 expression, moving

into the $CD44^{+}CD25^{+}$ (DN2) stage. At this point, thymocytes maintain limited capacity to differentiate into non-T cells. During the DN2 stage, *Tcrb*, *Tcrd*, and *Tcrg* recombination begin (Livák et al., 1999). At the *Tcrb* locus, D-to-J_{β} recombination precedes V-to-DJ_{β} rearrangement (Capone et al., 1998). These cells proliferate then downregulate CD44 expression and enter the $CD44^{-}CD25^{+}$ (DN3) stage. At DN3, rearrangement of *Tcrb*, *Tcrd*, and *Tcrg* continues, and DN3 cells retain limited multi-lineage potential (Masuda et al., 2007; Rothenberg, 2011; Livák et al., 1999). At this point, αβT and γδT cell lineages diverge (Kreslavsky et al., 2008; Parker and Ciofani, 2020; Ciofani and Zúñiga-Pflücker, 2010). Cells will join the γδT cell lineage if they receive a signal through the γδTCR. If cells instead receive a signal through the assembled pre-TCR, composed of a TCR_{β} chain and a surrogate pre-T_{α} chain, they progress toward the αβT cell fate (Fehling et al., 1995; Carpenter and Bosselut, 2010). In a process known as TCR_{β}-selection, if thymocytes receive relatively weak signals through the pre-TCR, they will develop to the $CD44^{-}CD25^{-}$ (DN4) stage. DN4 thymocytes begin proliferating and massively expanding, trafficking through the intermediate single positive stage to become $CD4^{+}CD8^{+}$ (DP) thymocytes. This proliferation and differentiation is dependent on downregulation of E2A expression by the inhibitor Id3, which is activated downstream of pre-TCR signaling (Engel et al., 2001).

At the DP stage, *Tcra* recombination occurs simultaneously on both alleles (Malissen et al., 1988, 1992). Multiple rounds of *Tcra* rearrangement are possible as V-

to-J_α recombinations leave upstream V_α and downstream J_α segments available (Petrie et al., 1993; Wang et al., 1998). The first recombination on a given allele is termed the "primary" rearrangement, and subsequent recombinations are referred to as "secondary." The sequential rounds of rearrangement provide multiple opportunities to generate in-frame recombinations that allow pairing of a TCR_α chain with a TCR_β chain to assemble an αβTCR. The αβTCR that can recognize self-antigen presented on self-major histocompatibility complex (MHC) on the surface of cortical thymic epithelial cells (cTECs) will receive a weak positive selection signal that saves it from death by neglect (Surh and Sprent, 1994). Signaling through an assembled αβTCR that recognizes self-antigen presented on self-MHC, like pre-TCR selection, causes upregulation of Id3, which inhibits E2A, allowing differentiation (Engel et al., 2001). After receiving a positive selection signal, DP thymocytes downregulate RAG, upregulate the CD3 complex, and traffic to the thymic medulla, where negative selection takes place. Medullary thymic epithelial cells (mTECs) and B cells are capable of expressing antigens typically restricted to specific tissues (Derbinski et al., 2001; Yamano et al., 2015). mTECs, B cells, and other antigen presenting cells (APCs) display these antigens to DP thymocytes during a process known as negative selection. If an αβTCR is highly reactive to self-peptide, it will receive too strong a selection signal through the TCR and die, become quiescent, or become a regulatory T cell to escape autoimmunity (Swat et al., 1991; Jordan et al., 2001; Apostolou et al., 2002). If DP thymocytes receive a positive selection signal and do not undergo negative selection, they will differentiate further to either the mature naïve $CD4^+$ or $CD8^+$ single positive (SP) subsets (Taniuchi, 2018).

Commitment to either the CD4SP or CD8SP subset depends on recognition of MHC class, along with a complex signaling program. CD4SP αβT cells are selected by MHC class II and comprise the helper T cell subset. CD8DP αβT cells are selected by MHC class I, and these cells comprise the cytotoxic T cell subset (Taniuchi, 2018).

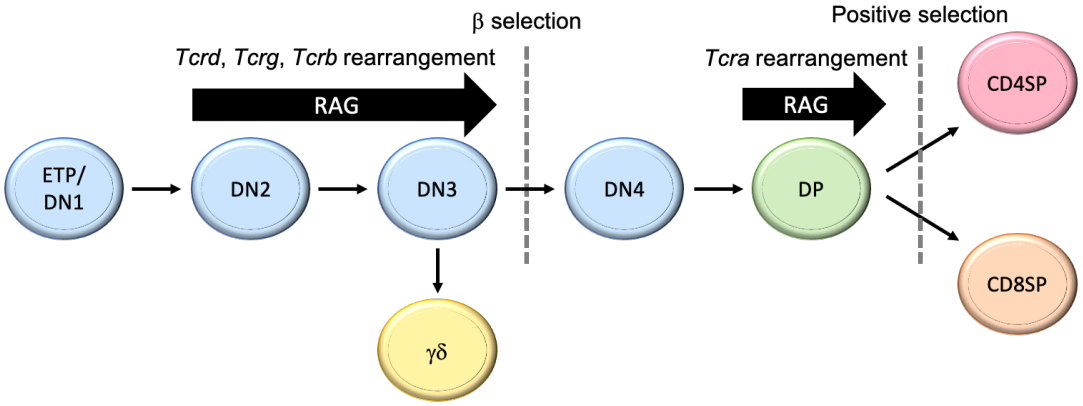

Figure 1: Thymocyte development.

1.2 Antigen receptor loci

AgR loci contain V (sometimes D) and J segments. They also have one or more constant (C) gene segments. In addition to the gene segments that serve as the building blocks of AgR chains, AgR loci also contain multiple regulatory features. BCRs are comprised of the protein products of 3 Ig loci- *Igh*, *Igk*, and *Igl*. There are also 3 TCR-encoding loci- *Tcrb*, *Tcrg*, and *Tcra-Tcrd*. At the *Igh*, *Tcrb*, and *Tcrd* loci, a V, D, and J are joined, with D-to-J recombination usually occurring prior to V-to-DJ recombination. The *Igk*, *Igl*, *Tcrg*, and *Tcra* loci are devoid of D segments, and recombinations are V-to-

J (Jung and Alt, 2004). AgR locus organization plays a role in regulating V(D)J recombination, and each locus has a unique structure (**Figure 2**).

In mice, the *Igh* locus, which is located on chromosome 12, spans nearly 3Mb. The 5′ 2.7Mb of the *Igh* locus contains up to 150 V_H segments, with the precise number varying by laboratory strain. D_H, J_H, C_H gene segments lie downstream of the V_H array. The *Igh* locus is unique among AgR loci, in that after VDJ_H recombination, mature B cells are capable of switching C_H class to generate different antibody isotypes in a process known as class-switch recombination (Jung et al., 2006; Lefranc et al., 2015).

The *Igh* locus also features several regulatory elements. Each V_H and D_H segment has an upstream promoter (Yancopoulos and Alt, 1986). An intergenic region between V_H and D_H segments, known as intergenic control region 1 (IGCR1), contains several CCCTC binding factor (CTCF)- binding elements (CBEs), which are known to regulate three-dimensional (3D) structure of genomic regions (Guo et al., 2011b). An intronic enhancer element, Eμ, is located between the J_H and C_H segments (Banerji et al., 1983; Gillies et al., 1983). Downstream of the C_H exons lies an accessible region known as the 3′ regulatory region (3′RR) comprised of several enhancers, which is followed by a cluster of downstream CBEs, known as the 3′CBEs (Lin et al., 2018; Guo et al., 2011a).

The mouse *Igk* locus is located on chromosome 6. Over 100 functional V_κ segments are localized at the 5′ end of the locus. Unusual among AgR loci, many V_κ gene segments are in inverted transcriptional orientation with respect to the rest of the locus; these segments are utilized in inversional, rather than deletional, recombination events. The 3′ end of the locus contains four functional (and one pseudogene) J_κ segments and a single

C_κ (de Almeida et al., 2015; Lefranc et al., 2015). Receptor editing of *Igk* chains is possible with secondary rearrangements on already recombined alleles utilizing upstream V_κ and downstream J_κ segments of the rearranged V_κ-J_κ joint (Feddersen and Van Ness, 1985).

The mouse *Igk* also locus features four known enhancers. The intronic kappa enhancer (iEκ) is located between the J_κ segments and C_κ (Queen and Baltimore, 1983). Another enhancer is located 9 kb downstream of C_κ; it is termed '3′Eκ' (Meyer and Neuberger, 1989). Both iEκ and 3′Eκ regulate germline transcription (GLT) and accessibility of the *Igk* locus and therefore promote *Igk* recombination. 3′Eκ is a stronger enhancer than iEκ, and it has additional roles in regulating *Igk* transcription in mature B cells and promoting somatic hypermutation (SHM) in germinal center B cells. Beyond 3′Eκ, a further downstream enhancer (dEκ) is responsible for transcriptional regulation of *Igk* in more mature B cells (Xiang and Garrard, 2008). dEκ is not required for normal *Igk* rearrangement, but it does positively regulate SHM. Recently, a fourth enhancer element, termed 'E88,' has been uncovered in the middle of the V_κ array (Barajas-Mora et al., 2019). E88 is a major anchor of the long-range chromatin structures of the *Igk* locus. It is in three-dimensional (3D) contact with the enhancers at the 3′ end of the *Igk* locus, and it regulates V_κ usage. It appears to establish a specific *Igk* chromatin architecture in pro-B cells, in advance of *Igk* rearrangement at the pre-B stage (Barajas-Mora et al., 2019).

Insulators in the *Igk* locus also play an important role in rearrangement. The two characterized insulators are located in the intergenic region between the V_κ and J_κ arrays.

The first discovered was a silencer in the intervening sequence (Sis), which reduces V-to-J_κ recombinations (Liu et al., 2002, 2006); its two CBEs are in the opposing orientations, facing away from each other (Ribeiro de Almeida et al., 2011). Upstream of Sis, another insulator localized in a DNase I hypersensitive region was shown to be important for locus contraction. It is therefore referred to as the contracting element for recombination (Cer) (Xiang et al., 2013); Cer contains two reverse-oriented CBEs. Contraction of *Igk* is necessary for appropriate V-to-J_κ recombination, and when Cer is experimentally deleted, recombinations are less likely to reach distal V_κ segments (Kleiman et al., 2018). The importance of directional CTCF binding sites suggests that both Sis and Cer are anchors of chromatin structures that organize topological domains, thereby sequestering V_κ from J_κ segments, hence their designation as insulator elements (Kleiman et al., 2018).

In mice, the *Igl* locus spans 240 kb on chromosome 16. This locus is one of the least complex AgR loci, but interestingly has several clusters of J_λ and C_λ segments and only three V_λ segments (Lefranc et al., 2015; Carson and Wu, 1989). The *Igl* locus is much more complex in humans (Kawasaki et al., 1997).

The mouse *Tcrg* locus, on chromosome 13, is only 200 kb in length (Kranz et al., 1985; Woolf et al., 1988). Like the *Igl* locus, the *Tcrg* locus, while not as large and complex as many AgR loci, contains several clusters of J_γ and C_γ segments, with few V_γ segments (Woolf et al., 1988). The promoters and enhancers near the $C_\gamma1$ cluster are the most notable additional features of the mouse *Tcrg* locus. Two promoters of interest are those for $V_\gamma2$ and $V_\gamma3$. The $V_\gamma3$ promoter is active only in fetal thymocytes, and therefore

this segment is not used in rearrangements of the *Tcrg* locus in adults. The promoter associated with $V_\gamma 2$, contrastingly, is active in both fetal and adult thymocytes (Agata et al., 2001; Baker et al., 1998). Two enhancers regulate *Tcrg* recombination, one is between the first two V_γ segments of the $C_\gamma 1$ cluster, $V_\gamma 5$ and $V_\gamma 2$. This enhancer is in a DNase hypersensitive site and is referred to as HsA. The second *Tcrg* enhancer is downstream of the $C_\gamma 1$ cluster and is thus referred to as 3′ECγ1 (Kappes et al., 1991). These enhancers are not required for *Tcrg* recombination, but they have overlapping roles in regulating *Tcrg* transcription (Xiong et al., 2002).

The mouse *Tcrb* locus stretches ~700 kb on chromosome 6. The 5′ end of the locus contains a single V_β segment, $V_\beta 1$, then a gap of 150 kb containing an array of *trypsinogen* genes. 33 more V_β segments lie downstream of $V_\beta 1$, though not all are functional. Another *trypsinogen* region follows this array of V_β segments. On the 3′ end of the locus, two $D_\beta J_\beta C_\beta$ clusters are located just upstream of a final V_β segment ($V_\beta 31$); $V_\beta 31$ lies in the opposite transcriptional orientation to the rest of the locus (Lefranc et al., 2015). Each $D_\beta J_\beta C_\beta$ cluster has an associated upstream promoter; the promoters appear to be essential for recombination of only the associated $D_\beta J_\beta C_\beta$ cluster (Whitehurst et al., 1999). The enhancer of *Tcrb* (E_β) is downstream of $C_\beta 2$. E_β is essential for all *Tcrb* recombination, including both D-to-J_β and V-to-DJ_β rearrangement. Experimental deletion of E_β causes a block in thymocyte development at the DN stage by ablating *Tcrb* recombination (Bouvier et al., 1996; Bories et al., 1996).

The mouse *Tcra-Tcrd* locus is complex due to a large array of gene segments and a unique structure among AgR loci, wherein the *Tcrd* and *Tcra* genes share one locus; D_δ, J_δ, and C_δ gene segments are nested between the V_α and J_α segments (Carico and Krangel, 2015). The *Tcra-Tcrd* locus in mice is on chromosome 14. The entire locus is 1.8 Mb in C57BL/6 mice and 1.6 Mb in all other laboratory strains, including strain 129 mice. The *Tcra-Tcrd* locus features V_δ gene segments interspersed among the V_α segments at the 5′ end, which comprise most of the length of the locus. The locus features a duplication event in the V segment array, which in C57BL/6 mice is repeated once again, leading to the increased length of the locus in this strain. All laboratory mouse strains have a distal duplication (denoted with "D" in the name of the gene segment) of the V_α-V_δ array beginning with the ninth-most 5′ gene segment. C57BL/6 mice have an additional region called the central duplication (denoted with "N" in the name of the gene segment) immediately downstream of the distal duplication. Most laboratory strains feature 104 V_α-V_δ segments in the *Tcra-Tcrd* locus, and the C57BL/6 allele contains an additional 34 V_α-V_δ segments in the central duplication. Not all of these gene segments are functional (Lefranc et al., 2015; Carico and Krangel, 2015). Humans have a similarly structured *Tcra-Tcrd* locus, but it does not contain a similar duplication, so there are fewer V_α-V_δ segments (Glusman et al., 2001). Downstream of the V_α-V_δ segments, the mouse *Tcra-Tcrd* locus contains two D_δ and two J_δ segments followed by a single C_δ segment. The lone V_δ segment (*Trdv5*) in the 3′ region is inverted with respect to the other gene segments. The J_α array contains 60 gene segments, but several are not

functional (Lefranc et al., 2015). These are followed by a lone C_α gene segment, the most 3′ protein-coding feature of the *Tcra-Tcrd* locus.

In addition to gene segments, the *Tcra-Tcrd* locus includes several regulatory features. Each V_α-V_δ gene segment has a coordinate upstream promoter. The enhancer of *Tcrd* (E_δ) is located between the J_δ segments and C_δ (Redondo et al., 1990; Gill et al., 1991). The enhancer of *Tcra*, E_α, is located just downstream of C_α at the 3′ end of the locus (Ho et al., 1989; Winoto and Baltimore, 1989). These enhancers temporally regulate rearrangement to generate *Tcrd* and *Tcra* genes, respectively. E_δ specifically promotes V-D-J_δ recombination in DN3 thymocytes, while E_α promotes V-to-J_α recombination at the DP stage (Lauzurica and Krangel, 1994; Hernández-Munain et al., 1999; Sleckman et al., 1997; Monroe et al., 1999c). A third major regulatory element, termed T early alpha (TEA), serves as a promoter for the V_α-proximal J_α segments (Villartay et al., 1987; Shimizu et al., 1993; Chasseval and Villartay, 1993). TEA is located just upstream of the most 5′ J_α segment *Traj61*. One more promoter lies proximal to *Traj49* and is therefore known as the promoter Jα49. TEA and Jα49 are activated by E_α and are responsible for promoting non-coding transcription through the 5′ portion of the J_α array. Together, TEA and the Jα49 promoter stimulate accessibility of proximal J_α segments and their use in primary recombination events (Mauvieux et al., 2003; Hawwari et al., 2005; Villey et al., 1996). Interestingly, *Traj61*, which is involved in many primary *Tcra* recombinations, has a defective splice site (Villey et al., 1997).

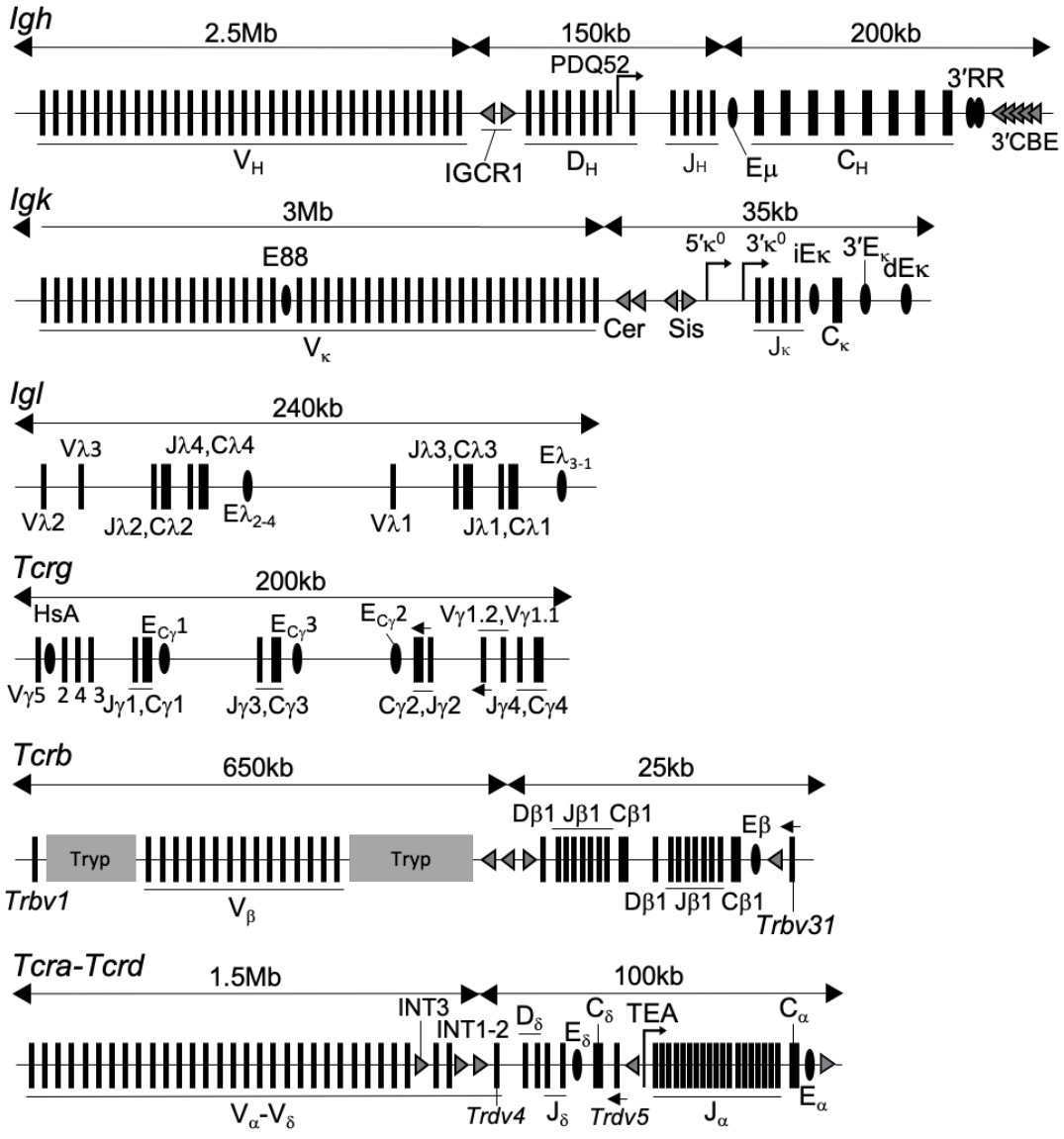

Figure 2: Linear structure of mouse antigen receptor loci.

1.3 V(D)J Recombination and Regulation

Adaptive immunity, which is a feature of all jawed vertebrates (Flajnik, 2018), relies on the *de novo* generation of receptors in somatic cells to recognize antigens. These AgRs are unique to the T and B cells of the adaptive immune system. V(D)J recombination, wherein V, D, and J gene segments are ligated in a cut-and-paste manner, is responsible for the immense combinatorial diversity of AgRs. Double-strand breaks (DSBs) in the DNA at TCR and Ig loci in developing T and B cells, respectively, are made by the protein products of the recombination activating genes (*Rag1* and *Rag2*). Collectively, the 'RAG' proteins complex with the high-mobility group protein B1 (HMGB1) to catalyze V(D)J recombination (Schatz and Swanson, 2011) (**Figure 3**).

The RAG complex recognizes and cleaves sites called recombination signal sequences (RSSs) immediately flanking V, D, and J gene segments. The RSS is made of a conserved heptamer (consensus 5′-CACAGTG-3′) and nonamer (consensus 5′-ACAAAAACC-3′) separated by a 12 or 23 bp spacer. These are respectively called 12-RSS and 23-RSS. Recombinations always take place between a 12-RSS and a 23-RSS; this conservation of recombinations is referred to as the 12/23 rule. HMGB1 promotes adherence to the 12/23 rule (Hiom and Gellert, 1998; Van Gent et al., 1996, 1997). HMGB1 enhances coupled RAG cleavage both *in vitro* and *in vivo*, particularly by promoting cleavage at 23-RSSs, and HMGB1 may function to bend DNA to induce cleavage (Van Gent et al., 1997; Thwaites et al., 2019). The structure of the RAG heterotetramer is asymmetrical, which further enforces the 12/23 rule (Ru et al., 2015; Kim et al., 2018). RAG-mediated cleavage is induced between the RSS and gene segment

when two RSSs are brought together in a structure known as the synaptic complex. RAG induces cleavage in two steps. First, one strand is nicked at the 5′ end of the RSS, leaving a 3′-OH to attack the other strand, forming a blunt signal end and a hairpin at the coding end (McBlane et al., 1995). The structure of the RAG heterotetramer closes around nicked RSS pairs, holding strands in place for hairpin formation (Ru et al., 2018; Kim et al., 2018).

After RAG-mediated DSBs take place, the synaptic complex holds the ends in proximity to facilitate ligation of the RSSs into a signal joint and the coding segments into a coding joint. Non-homologous end joining (NHEJ) is responsible for repair of RAG-mediated DSBs. In a successful rearrangement event, the signal ends and intervening sequence are cleaved from the genome and repaired to form an excision circle. The coding joint features two gene segments stitched together on the AgR allele. When coding joints are formed, palindromic (P) nucleotides are added between the two coding ends during the opening of sealed hairpins. Terminal deoxynucleotidyl transferase (TdT) is responsible for increasing receptor diversity further by adding non-templated (N) nucleotides to coding junctions during the NHEJ reaction (Gilfillan et al., 1993; Komori et al., 1993; Alt and Baltimore, 1982; Landau et al., 1987; Elliott et al., 1988). These TdT-inserted N nucleotides provide an additional layer of diversity, as they are not germline encoded, and they are unique to each recombination event (Schatz and Ji, 2011). After recombination, V(D)J transcripts are spliced to a C gene segment to generate a receptor chain.

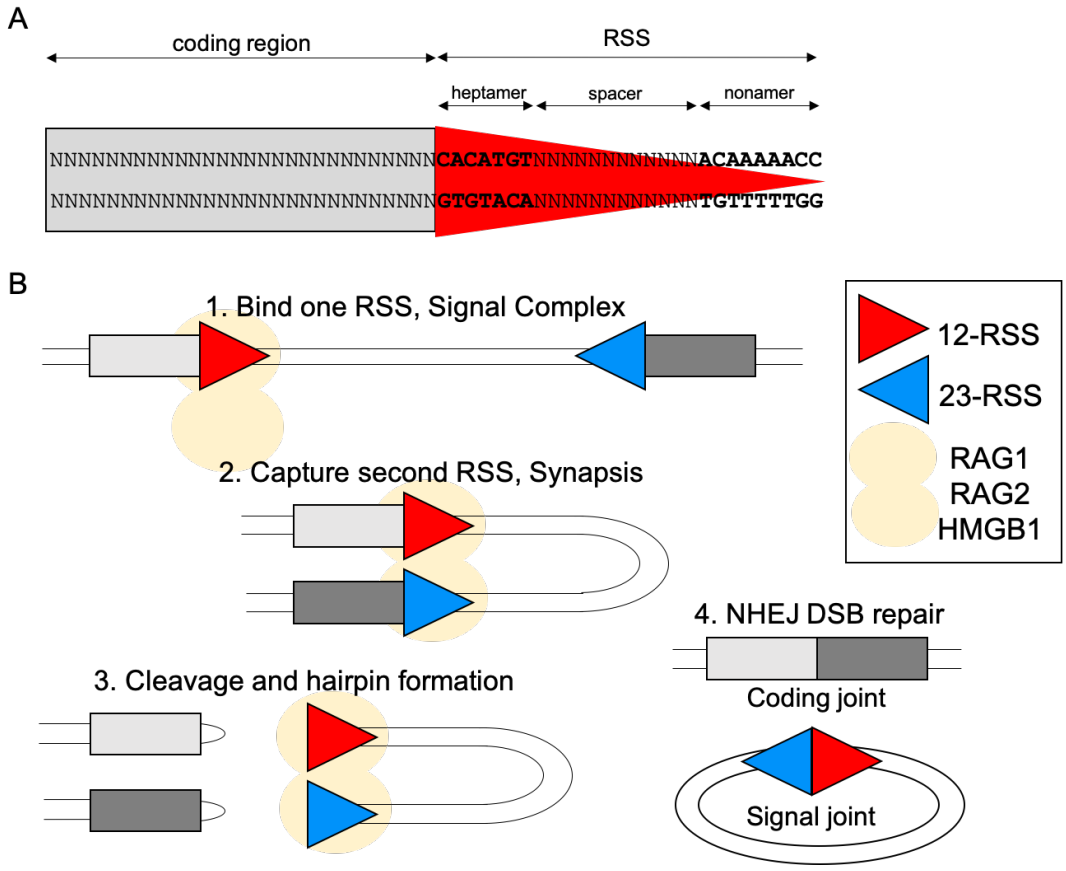

Figure 3: V(D)J Recombination.

(A) Depicts the structure of the gene-encoding segment and RSS. (B) Recombination begins when the RAG complex binds one RSS. A second RSS is captured to form a synaptic complex. Between the RSS and gene segment, RAG induces DSBs. These DSBs are processed and repaired via NHEJ, forming a signal joint and coding joint.

Complex regulation of V(D)J recombination is required to avoid the disastrous consequences of uncontrolled DSBs at other genomic locations or in cells other than B and T lymphocytes (Krangel, 2009; Teng et al., 2015). Regulation of RAG-induced breaks is absolutely essential to spur rearrangements in a cell-type and developmental

stage-specific fashion. Genomic DSBs such as those induced by RAG can lead to oncogenic translocations when not properly protected and repaired. The RSS regions RAG recognizes and cleaves are restricted to AgR loci, but many other genomic sites have sequences with enough similarity, known as cryptic RSSs, to allow off-target recombination events leading to translocations (Teng et al., 2015; Marculescu et al., 2002). As such, directing when Rag is transcribed and where RAG localizes to induce DSBs must be tightly coordinated.

1.3.1 Regulation of Rag expression

The RAG complex, composed of a heterotetramer of the protein products of *Rag1* and *Rag2* genes (Ru et al., 2015; Kim et al., 2015), is responsible for the endonuclease activity requisite for V(D)J recombination (Oettinger et al., 1990; Schatz et al., 1989). Two RAG1 proteins, which have the catalytic activity, and two cofactor RAG2 proteins comprise this heterotetramer. Expression of the two RAG proteins is tightly controlled; there are two waves of *Rag* gene expression in both developing B and developing T cells (Kuo and Schlissel, 2009; Miyazaki and Miyazaki, 2021). In developing B cells, *Rag1* and *Rag2* are first expressed at the pro-B cell stage (Yu et al., 1999). At this stage, *Igh* recombination begins. Similarly, the DN2 stage, when *Tcrd*, *Tcrg*, and *Tcrb* recombinations begin, is the first stage when *Rag* genes are expressed in thymocytes (Monroe et al., 1999b; Wilson et al., 1994). *Rag* is then downregulated in cycling pre-B and DN4 cells. Signaling through the pre-BCR, pre-TCR, or γδTCR represses *Rag* expression. During these stages, both pre-B and pre-T cells proliferate. *Rag* expression is

anti-correlated with cell division; control of *Rag* expression prompts recombination to take place during the resting and growth 1 phases of the cell cycle (Lin and Desiderio, 1994; Grawunder et al., 1995; Schlissel et al., 1993). *Rag* expression then recurs at the resting small pre-B stage for *Igk* and *Igl* recombination and the non-cycling DP stage for *Tcra* recombination. *Rag* expression continues through the pre-B and DP stages until a viable BCR or TCR has been generated or cells die from lack of AgR signaling (Kuo and Schlissel, 2009; Takahama and Singer, 1992).

The expression of *Rag* is temporally controlled in a lineage-specific fashion by enhancers at the *Rag* locus, which contains the genes encoding RAG1 and RAG2 proteins. In developing B cells, the *Rag2*-proximal Erag enhancer is stimulated by the transcription factor FOXO1 to induce *Rag* transcription. This *Rag* transcription is repressed by BCR and IL-7R signaling with downstream Akt kinase antagonism of FOXO1 activity at Erag (Amin and Schlissel, 2008). Deletion of Erag reduces *Rag1* and *Rag2* transcription, which reduces recombination and causes a partial blockade in progression of pro-B cells to the pre-B compartment (Hsu et al., 2003).

In T lineage cells, distinct regulatory features are responsible for suppressing and upregulating *Rag* expression at the DN and DP stages. A silencer element located between *Rag1* and *Rag2* has been determined to repress expression of these genes (Yannoutsos et al., 2004). An anti-silencing element 5′ of *Rag2* counteracts the silencer to direct upregulation of *Rag* in DN3 and DP thymocytes by interacting with the *Rag* promoters to facilitate activation (Monroe et al., 1999a; Yu et al., 1999; Yannoutsos et al., 2004; Hao et al., 2015). Runx1, Gata3, and E2A are all transcription factors that have

been determined to play a role in upregulating *Rag* expression in DP thymocytes, where transcript levels are much higher than in DN cells (Naik et al., 2019).

RAG is also regulated post-transcriptionally. In resting and growth 1 phase cells, RAG accumulates, allowing DSBs in quiescent and resting cells. However, in S phase, RAG is degraded. RAG2 features a phosphorylation site marking it for ubiquitination and degradation to protect dividing cells (Li et al., 1996; Lin and Desiderio, 1994). RAG expression in dividing cells puts the genome at risk, posing the possibility of DSBs being repaired by processes other than NHEJ and potentiating oncogenic translocations or genotoxicity (Teng et al., 2015; Helmink and Sleckman, 2012; Küppers and Dalla-Favera, 2001).

1.3.2 Regulation of RAG accessibility

In order to efficiently participate in V(D)J recombination, RSS regions proximal to V, D, and J gene segments must be accessible to RAG binding. Moreover, accessibility to RAG must be repressed to prevent ectopic cleavage. Multiple regulatory features control RAG accessibility in a lineage-dependent and developmental stage-specific fashion. Accessibility relies on transcriptional activation in nucleosome-depleted regions, histone modifications, and chromatin structure (Schatz and Ji, 2011; Carico and Krangel, 2015; Yancopoulos and Alt, 1985; Cobb et al., 2006). RAG proteins initially bind to AgR loci at foci of J (and sometimes D) segments referred to as the recombination center (RC). RCs are enriched for RNA polymerase II and activating histone marks. High levels

of RAG binding to RSS regions within the RC facilitate capture of a second RSS for synapsis and cleavage (Ji et al., 2010; Schatz and Ji, 2011).

GLT of AgR loci was shown to be correlated with V(D)J recombination for many years (Fondell and Marcu, 1992; Lennon and Perry, 1985; Yancopoulos and Alt, 1986). The necessity of GLT for recombination was first demonstrated at the *Tcra-Tcrd* locus (Abarrategui and Krangel, 2006). Transcription across the J_α array not only promotes recombination but suppresses activation of cryptic J_α promoters in the central portion of the J_α array. This was determined by inserting a transcription terminator upstream of *Traj53*, which reduced recombinations using immediately downstream J_α segments and led to increased permissive chromatin marks at cryptic promoters in the central portion of the J_α array (Abarrategui and Krangel, 2006). The TEA promoter is responsible for stimulating noncoding transcription through the J_α array. By experimentally introducing a transcription terminator immediately downstream of TEA, it was demonstrated that TEA-mediated transcription activated proximal J_α promoters while suppressing central J_α promoters. This pattern of transcriptional regulation enforces ordered J_α utilization during *Tcra* recombination (Abarrategui and Krangel, 2007).

Enhancers increase transcription, and thereby support RAG accessibility and cleavage (McMurry et al., 1997; Cobb et al., 2006). The high accessibility required for RC formation is facilitated by developmentally-regulated promoters and enhancers in these regions (Schatz and Ji, 2011). Lineage- and stage-specific activation of enhancers is mediated by cooperative transcription factor binding (Spitz and Furlong, 2012). In pro-B

cells, RC formation spanning the J_H array and a portion of the D_H array depends on iEμ (Afshar et al., 2006). At the *Tcra-Tcrd* locus, E_δ is active in DN thymocytes and inactivated in DP thymocytes. E_α and TEA, in turn, are active in DP thymocytes (Hernández-Munain et al., 1999). This switch between enhancers facilitates stage-specific recombination of first *Tcrd*, and later *Tcra*. E_α, which is essential for GLT of J_α segments, increases accessibility of RC-proximal V_α segments and all J_α segments. Deletion of E_α has been demonstrated to block *Tcra* recombination (McMurry and Krangel, 2000; Sleckman et al., 1997). Activation of E_α is established by a shift in transcription factor binding downstream of pre-TCR signaling (Hernández-Munain et al., 1999; Blanco et al., 2012). Similarly, *Igk* is recombinationally silent in pre-B cells, but pre-BCR signaling mediates transcription factor binding and activation of iEκ and 3′Eκ (Beck et al., 2009; Mandal et al., 2015). Histone acetylation, a marker of accessibility, is also capable of stimulating RAG-mediated cleavage (Kwon et al., 2000). At the *Tcra-Tcrd* locus histone 3 (H3) acetylation was demonstrated to promote accessibility downstream of enhancer activity (McMurry and Krangel, 2000). Further, repressive histone deacetylation suppresses regional use of gene segments in recombination, even in activated AgR loci undergoing recombination where proximal gene segments with acetylation were active recombination substrates (Chakraborty et al., 2007).

Additional histone modifications more directly impact RAG binding and activity. While RAG1 binds to RSSs based on sequence, particularly the CAC of the heptamer (Teng et al., 2015) and AT-rich nonamer (Yin et al., 2009), RAG binding is enriched at

sites with H3 lysine 4 trimethylation (H3K4me3), a marker of active transcription. The plant homeodomain (PHD) finger of RAG2 binds H3K4me3, which enhances RAG-mediated cleavage and recombination (Liu et al., 2007; Ramón-Maiques et al., 2007; Matthews et al., 2007; Ji et al., 2010). H3K4me3 serves as an allosteric activator of the RAG complex. Binding of the RAG2 PHD to H3K4me3 changes the conformation of RAG, displacing an autoinhibitory domain to increase binding affinity and stimulate cleavage (Ward et al., 2018; Bettridge et al., 2017). Therefore, RAG binding is enriched at transcriptionally active and H3K4me3-marked regions genome-wide (Teng et al., 2015).

1.4 Chromatin organization

In addition to enhancers, promoters, and epigenetic control, chromatin architecture is a major regulator of transcription. This is the case genome-wide and across species and cell types (Dixon et al., 2012). A variety of technologies have emerged to determine how chromatin is organized within the nucleus of eukaryotic cells. Low resolution determination of 3D genome architecture is achieved with three-dimensional fluorescent *in situ* hybridization (3D-FISH), which allows microscopic visualization of tagged genomic regions using fluorescent DNA probes. More recent advances in chromosome conformation capture (3C) technologies allow the study of 3D structure by interrogating interactions along the genome. These techniques rely on crosslinking DNA to capture contacts between specific genomic sites. Contacts between two specific regions of interest are then determined by PCR in ‘3C,’ whereas contacts between one

site and the rest of the genome are determined by ‘4C,’ and contacts among all genomic sites are determined by ‘Hi-C’ (Dekker et al., 2002; Zhao et al., 2006; Lieberman-Aiden et al., 2009; Denker and De Laat, 2016; Sexton and Cavalli, 2015). 4C and Hi-C rely on high-throughput sequencing (HTS) to detect genome-wide interactions.

At the highest level, sub-chromosomal DNA is organized into repressive heterochromatin and activated euchromatin. Activation state is mediated in part by subnuclear compartmentalization. Genes are more activated when they are distal to the nuclear lamina (Guerreiro and Kind, 2019; Pickersgill et al., 2006; Guelen et al., 2008). Genes tethered to the lamina are transcriptionally repressed. Transcriptional repressors are recruited to DNA and mediate gene silencing and tethering to the lamina (Zullo et al., 2012; Reddy et al., 2008).

Further, megabase-scale topologically associated domains (TADs) organize regions of interacting chromatin to regulate expression of genes by bringing promoters and enhancers into proximity (Dixon et al., 2012; Nora et al., 2012; Pope et al., 2014). Neighboring TADS, on the scale of megabases, can therefore be in opposing activation states. TAD structures are mostly static between cell types and differentiation stages (Dixon et al., 2012). Mapping the genome-wide interactome and regulome in mice revealed that TADs confined enhancers to co-regulate gene networks within these structures (Shen et al., 2012). Disruption of TAD boundaries induces aberrant co-regulation of transcription between TADs by changing the regulatory interactome (Nora et al., 2012).

Smaller regulatory units within TADs (both sub-TADs and loops) are more dynamic. A study of chromatin architecture at several loci revealed hierarchical interactions that are reorganized over the course of differentiation. These structures were mediated by combinations of the CTCF, Mediator, and cohesin proteins (Phillips-Cremins et al., 2013; Sexton and Cavalli, 2015). While stable long-range interactions were primarily anchored by CTCF and cohesin, short-range interactions were established by Mediator-cohesin interactions. Both 3D structure and coordinated gene expression were demonstrated to be altered when Mediator or cohesin were knocked down (Phillips-Cremins et al., 2013). Further interrogation of the regulators of genome structure revealed that while cohesin shapes chromatin domains, CTCF is responsible for enforcing boundaries between domains (Zuin et al., 2014). At the kilobase-scale, chromatin loops bring specific elements, primarily promoters and enhancers, into proximity while keeping others sequestered to drive coordinated patterns of gene expression (Palstra et al., 2003). These 3D interactions form a dynamic regulatory network. At AgR loci, the 3D structure of chromatin is essential in regulating orderly recombination.

1.4.1 3D architecture mediated by CTCF and cohesin

Loops and TADs are dynamic structures that differ by lineage and developmental stage as well as within individual cells. Hi-C evidence in recent years has made clear that chromatin domains are formed in 3D space by cohesin-mediated DNA extrusion (Hansen et al., 2018; Davidson et al., 2019). The multimeric ATPase cohesin complex has been well-characterized for its roles in holding sister chromatids together and DSB repair. The

ring-shaped cohesin structure with a hinge that can hold DNA in place has more recently been characterized as playing a role in transcriptional regulation (Nasmyth and Haering, 2009). It was determined that, in humans, the 11-zinc finger DNA-binding transcriptional regulator CTCF co-localizes with cohesin. Early studies concluded that cohesin and CTCF acted together as insulators to shape the chromatin into transcriptional units (Wendt et al., 2008). However, further work demonstrated that cohesin, Mediator, and CTCF act together to physically bring enhancers and promoters into 3D proximity to mediate transcription. These topological loops were determined to be cell-type specific and responsible for dynamic transcriptional regulation (Kagey et al., 2010). While it was initially postulated that interactions within loop domains happened by diffusion in 3D nuclear space, with CTCF and cohesin maintaining structures that formed by diffusion, recent evidence demonstrates that a different mechanism is at play in at least some cases (Kim et al., 2019).

Cohesin was then theorized to actively pull DNA through its ring-shaped structure in a bi-directional fashion until being stopped by CTCF on either end (Sanborn et al., 2015; Fudenberg et al., 2016). The anchors on either side of loop domains were observed in Hi-C maps as bright spots of contact between regions on the same chromosome. Interestingly, Hi-C analysis also detected areas of increased interaction in the genomic region flanked by the highly interacting anchors. These interactions visualized in the regions flanked by anchors are referred to as “stripes” due to their appearance on Hi-C maps. The stripes appear to indicate continuous sliding contact from one anchor to the next anchor. When cohesin was experimentally degraded, the loss of not only the bright

contact spots, but also the stripes, was observed. The stripes and spots were recovered quickly when cohesin was allowed to return (Rao et al., 2017). By Hi-C, it was also determined that loop domains and their stripes were dependent on ATPase activity of cohesin. Together, these findings suggested that cohesin was actively extruding DNA, pulling it into loops anchored by CTCF binding sites, rather than associating with cohesin at the boundaries of domains formed by diffusion in 3D space. This process has been termed "loop extrusion." Loop extrusion forms topologically interacting domains that bring promoters and enhancers into proximity (Vian et al., 2018). CTCF and cohesin colocalize to form a stable complex that blocks further extrusion and anchors loop and TAD boundaries (Hansen, 2020; Hansen et al., 2018). CTCF binding at loop or TAD boundaries is regularly found in a convergent orientation based on the DNA sequence of its binding sites; pairing between convergently-oriented CBEs appears to facilitate stability of topological structures by serving as a roadblock for ongoing cohesin-mediated DNA extrusion (Rao et al., 2014). This pattern of CTCF binding orientation lends further support to cohesin-mediated loop extrusion rather than diffusion as the mechanism forming chromatin loops in mammalian cells. DNA is bidirectionally extruded by cohesin at high speed, 0.5-1kb/s on average (Davidson et al., 2019; Kim et al., 2019). This rapid organization of the genome into cell type-specific functional units explains why technologies like 3C and 4C demonstrate strong interaction peaks between anchoring sites, with smaller peaks of low-frequency interactions along genomic regions flanked by anchors due to active loop extrusion.

1.4.2 V(D)J regulation by chromatin architecture

The organization of AgR architecture within the nucleus is vital in regulating V(D)J recombination. Repressive localization to the nuclear periphery suppresses recombination to enforce both stage-specific recombination and allelic exclusion. In hematopoietic stem cells, the *Igh* and *Igk* loci are localized at the nuclear periphery. This lamina-association of *Igh* and *Igk* represses transcription, and the loci localize to the center of the nucleus in a developmental stage-appropriate fashion to facilitate V(D)J recombination (Reddy et al., 2008; Kosak et al., 2002; Guo et al., 2011a). *Tcrb*, which recombines one allele at a time (allelic exclusion), is localized to the nuclear periphery during rearrangement in DN thymocytes (Schlimgen et al., 2008). *Tcrb* alleles more proximal to the lamina undergo less recombination, as RAG2 is excluded from the lamina-proximal domain (Chan et al., 2013). Allelic exclusion of the *Tcrb* locus is regulated in part by localization to the nuclear lamina as transcription and recombination are suppressed. Association with the lamina restricts E_β enhancer activity to the *Tcrb* RC, excluding V_β segments from synapsis with the RC (Chen et al., 2018).

AgR loci are also organized into TADs and loops in 3D space. This organization serves to both regulate transcription and bring some segments into proximity of the RC for recombination while others are sequestered, enforcing orderly recombination. All AgR loci have distinct patterns of CTCF binding (Loguercio et al., 2018), which enforce intra-locus structural domains. Loop boundaries enforced by CTCF restrict RAG into domains, preventing off-target RAG activity (Hu et al., 2015). These structures are dynamic over the course of development to regulate recombination. Typically, the

segments comprising an RC are in contact, and the spread of these contacts are contained within a single loop domain. At the appropriate developmental time, V segments are looped into proximity of the RC for V-to-(D)J synapsis and cleavage. All AgR loci contract at the appropriate developmental time to bring V segments into proximity of the RC to facilitate synapsis (Fuxa et al., 2004; Sayegh et al., 2005; Roldán et al., 2005; Jhunjhunwala et al., 2008; Skok et al., 2007). Synapsis may be regulated by diffusion in 3D space or by RAG chromatin scanning during cohesin-mediated loop extrusion (Jain et al., 2018; Peters, 2021).

Recent data demonstrates that the capture of V (and sometimes D) segments by the RC occurs by at least two distinct mechanisms, diffusion and RAG scanning. Much of the work on this subject has been performed on the *Igh* locus, but it is likely to hold for other AgR loci to a greater or lesser extent. Though inversional recombinations are possible when the 12/23 rule is followed, on most AgR loci inversional rearrangements appear rarely to segments upstream of the RC (Sollbach and Wu, 1995; Bolland et al., 2016). It has recently been demonstrated that the directional preference in recombination is due to at least in part to RAG scanning linked to cohesin-mediated loop extrusion within CTCF-bound regions (Zhang et al., 2019). This is strengthened by the finding that cryptic RSSs, which are present in both orientations, are also preferentially utilized in an orientation-specific manner (Hu et al., 2015; Zhao et al., 2016).

The entire *Igh* locus occupies a single TAD, which appears as a stripe in visualization of Hi-C analysis (Hill et al., 2020; Vian et al., 2018). *Igh* is in an extended conformation in pre-pro-B cells and non-B lineage cells. *Igh*, like *Tcrb*, is localized to the

nuclear periphery. The B cell lineage commitment factor Pax5 facilitates re-localization away from the lamina and induces contraction of the *Igh* locus (Jhunjhunwala et al., 2009; Fuxa et al., 2004; Guo et al., 2011a). Recently, it was determined that Pax5 mediates contraction of the *Igh* locus by repressing *Wapl*, the factor responsible for releasing cohesin from chromatin (Hill et al., 2020). In pro-B and pre-B cells, transcriptional repression of *Wapl* by Pax5 facilitates cohesin-mediated loop extrusion of the *Igh* locus as cohesin residence time is increased (Hill et al., 2020).

At the pro-B cell stage, *Igh* condenses into a rosette structure with several loops in the distal V_H, proximal V_H, and RC regions. The loop containing the RC stretches from IGCR1 to the 3′RR, containing D_H, J_H, and C_H segments, as well as E_μ (Degner et al., 2011; Guo et al., 2011b; Jain et al., 2018). This loop sequesters the RC and D_H and J_H segments from the V_H array to enforce D-to-J_H recombination prior to V-to-DJ_H recombination (Guo et al., 2011b). The V_H array becomes contracted and comes into proximity of the RC to facilitate long-range recombinations in pro-B cells (Jhunjhunwala et al., 2009; Degner et al., 2011; Jhunjhunwala et al., 2008). These long-range interactions are dynamic, suggesting cohesin-mediated loop extrusion as a mechanism, and 4C analysis demonstrated that developmental timing of these architectural changes depended on the transcription factor YY1 as well as CTCF and Pax5 (Medvedovic et al., 2013). As at other genomic locations, loop extrusion at *Igh* is ATP-dependent. In addition to bringing gene segments into proximity of the RC, loop extrusion links enhancers and promoters, regulating both transcription and recombination (Vian et al., 2018).

The orientation of the 23-RSS of J_H segments appears to direct RAG bound to the RC to scan upstream to locate convergent 12-RSSs of D_H segments, which mediates deletional joining. This scanning activity can be blocked by transcriptional interference, by experimental nuclease-dead Cas9, or by the 3′CBE loop anchor. Then, RAG accumulates in the RC and region proximal to the scanning impediment. However, the proximal D_H segment, DQ52, which lies within the *Igh* RC, appears to be used by diffusion rather than RAG scanning (Zhang et al., 2019). *Igh* recombination was reduced in a pro-B cell line when an auxin-inducible system was used to degrade cohesin. However, recombination to RC-proximal segments such as DQ52 were not reduced. When CTCF was degraded using this system, CBE-mediated interactions were reduced, but distal V_H segments were more frequently utilized in recombination. Distal V_H-RC interactions, which are typically observed in primary pro-B cells, were increased in a pro-B cell line when CTCF was experimentally degraded (Ba et al., 2020). This data suggests that physiological levels of CTCF serve as an impediment to cohesin-mediated loop extrusion and RAG scanning to distal V_H segments in the pro-B cell line. Intriguingly, IGCR1 does not appear to completely block RAG scanning from the RC in primary pro-B cells because *Wapl* repression allows increased cohesin residence time, facilitating loop extrusion- and scanning-mediated V_H capture (Dai et al., 2021; Hill et al., 2020).

CTCF binding proximally downstream of V_H segments increases their accessibility, prolonging interaction with the RC and increasing the likelihood of recombination. When IGCR1 was experimentally deleted, these interactions were increased for D_H-proximal V_H segments (Jain et al., 2018). Deletion of IGCR1 appears to

allow RAG targeting over a new, larger domain with different boundaries in pro-B cell lines (Hu et al., 2015). Experimentally inverting the distal V_H region, including both gene segments and CBEs, revealed dramatically reduced recombination of inverted V_H segments, while segments in their physiological orientation were still utilized (Zhang et al., 2019; Ba et al., 2020). It is possible that the inverted CBEs do not enforce cohesin pausing, though their distance from the V_H segments makes this a less plausible mechanism than the explanation that the inverted RSSs cannot be captured during unidirectional loop extrusion (Hill et al., 2020; Dai et al., 2021).

In total, data from the *Igh* locus supports a mechanism of loop extrusion mediating 3D structure and recombination. Pax5-mediated repression of *Wapl* allows the formation of CBE-bounded topological domains by long-range cohesin-mediated loop extrusion. Further, cohesin-mediated loop extrusion facilitates directional RAG scanning upstream of the 23-RSSs of J_H segments to preferentially allow deletional synapsis with convergently-oriented D_H 12-RSSs. In primary pro-B cells, RAG scanning has been determined to allow V_H capture by the RC. Experimental inversion of the orientation of bona fide RSSs and cryptic RSSs confirmed a preference for deletional recombinations between convergently oriented RSSs, which is indicative of scanning, as diffusion-mediated synapsis has no preference for RSS orientation. In physiological conditions, not all segments are captured by scanning. DQ52 appears to be captured by diffusion. Both scanning and diffusion play roles in mediating capture and synapsis at the *Igh* locus.

Recombination of the *Igk* locus is also impacted by CTCF binding. When CTCF is depleted, proximal V_κ segments are used at a higher frequency. Enhancer activity

spreads to the proximal V_κ segments, rather than being sequestered to the RC. Further, the reach of the iEκ and 3′Eκ enhancers extends 3′ of the locus without CTCF enforcing boundaries on the regulome typically contained within the RC (Ribeiro de Almeida et al., 2011). These findings may be consistent with loop extrusion continuing without CBE-mediated boundaries.

Experimental deletion of the Sis CBEs between V_κ and J_κ segments, reminiscent of findings from IGCR1 deletion on *Igh* alleles, lead to more proximal V_κ segment utilization. Sis appears to enforce a boundary between the RC and V_κ segments (Xiang et al., 2011). The utilization of proximal V_κ segments in the absence of Sis is most likely due to increased RAG scanning from the RC, as segments in the inverted orientation are used less frequently when Sis is experimentally deleted (Xiang et al., 2011). Intriguingly, physiological utilization of V_κ segments does not appear to depend on segment orientation, suggesting that V_κ capture for *Igk* recombination is mediated primarily by diffusion in 3D space. RAG scanning from the RC during cohesin-mediated loop extrusion appears to be blocked by Sis in physiological conditions. However, after a deletional recombination event, Sis would be removed from the locus and RAG scanning may occur during secondary recombinations of *Igk*.

By analyzing epigenetic marks along the *Igk* locus, an enhancer was recently discovered in the central portion of the V_κ array, proximal to the V_κ1-88 gene segment. This enhancer, E88, displays abundant binding of enhancer-marking transcription factors in addition to B cell lineage transcription factors and is only accessible in B lineage cells

(Barajas-Mora et al., 2019). In both cell lines and mice, E88 regulates chromatin structure. In an interactome bound by convergent CBEs, E88 interactions span sub-TADs along the *Igk* locus; the pattern of E88-mediated interactions changes from the pro-B to the pre-B stage. Interactions between E88 and enhancers in the 5′ portion of the V_κ array increase in pre-B cells, facilitating equal opportunity V_κ use by upregulating GLT and bringing the central portion of the V_κ array into proximity of Sis and Cer, which serve as a boundary between the RC and the V_κ array. The rate of rearrangement is reduced in the absence of E88, and recombinations to V_κ segments proximal to E88 occur less frequently (Barajas-Mora et al., 2019). E88 appears to be a central regulator of *Igk* topology and accessibility that facilitates diffusion-mediated RAG capture and synapsis.

1.4.3 *Tcra-Tcrd* locus 3D architecture

The topology of *Tcra-Tcrd* locus, like other AgR loci, is dynamic over development, which serves to regulate stage-specific recombination. Because this locus is a shared space for *Tcrd* and *Tcra* gene segments, which recombine at different stages, it serves as a fascinating and complex example of developmental regulation of V(D)J recombination by chromatin architecture. In 3D-FISH studies examining the *Tcra-Tcrd* locus, splenic B cells demonstrated an elongated structure. In DN thymocytes, the time at which *Tcrd* recombinations are ongoing, the *Tcra-Tcrd* locus was revealed to be contracted, with the entire V_α-V_δ array in proximity to the 3′ end of the locus. In DP thymocytes, the time of *Tcra* recombination, the 5′ end of the locus was determined to be

de-contracted while the 3′ end remained contracted (Shih and Krangel, 2010). These structural changes reveal a potential mechanism for the use of V_δ segments spread across the V_α-V_δ array in DN thymocytes, while V_α use in DP thymocytes is restricted to more J_α-proximal segments. The change in *Tcra-Tcrd* locus structure from DN to DP does not depend on E_δ or E_α enhancer activity (Shih and Krangel, 2010). Rather, it appears that locus topology regulates the reach of enhancer activity.

The *Tcra-Tcrd* locus features many CBEs in the V_α-V_δ array, as well as in the intergenic region upstream of the DJC_δ cluster, at TEA, at E_α, and at the 3′ end of the locus (Shih et al., 2012; Loguercio et al., 2018). Based on sequence analysis of the core binding motif, most CBEs are in the forward orientation, but several prominent CBEs along the locus are in the reverse orientation. The two forward-oriented intergenic CBEs upstream of *Trdv4* and the DJC_δ cluster (INT1 and INT2) are important regulators of locus structure and recombination in DN thymocytes (Chen et al., 2015; Zhao et al., 2016). A stripe between the CBEs at INT1-2 and TEA can be observed in Hi-C maps, suggesting that a loop between these anchors is formed by cohesin-mediated extrusion (Zhao et al., 2020). INT2 in particular appears to form the 5′ boundary of a loop with TEA, at which a reverse-oriented CBE is located. This loop contains the *Tcrd* RC. INT1-2 separates the *Tcrd* RC from most V_δ segments. INT1 is involved in infrequent interactions with sites upstream of, within, and downstream of the INT2-TEA loop (Chen et al., 2015). When both INT1 and INT2 were experimentally disrupted, the proximal upstream *Trdv2-2* and pseudogene *Trdv3* were used more frequently in recombinations.

RAG appears to be able track to off-target sites upstream of the *Tcrd* RC in the absence of INT1-2, suggesting that in DN thymocytes these CBEs may function similarly to Sis. Experimental deletion of INT2 promoted the use of INT1 as the upstream interacting partner of TEA. However, the interactome of TEA extended upstream of INT1 in the absence of INT2, suggesting that INT1 is not perfectly redundant with INT2 as a boundary element (Zhao et al., 2016; Chen et al., 2015). In summary, INT1-2 appears to prevent RAG tracking from the *Tcrd* RC to distal V_δ segments, making diffusion the likely mechanism by which V_δ segments are captured in physiological conditions.

A major structural loop at the *Tcra-Tcrd* locus in DP thymocytes is between TEA and E_α; both of these features bind both CTCF and cohesin (Seitan et al., 2011). This loop brings E_α into proximity of TEA, which serves as a promoter for J_α segments. When cohesin is experimentally deleted, this interaction is reduced. Less GLT and reduced H3K4me3 are observed at central and distal J_α segments in the absence of cohesin. Recombination to central and distal J_α segments is also reduced; further, GLT of C_α is reduced. The influence of E_α seems to instead spread 3′ of the *Tcra-Tcrd* locus to the downstream *Dad1* gene, which is transcriptionally upregulated, suggesting a loss of insulation (Seitan et al., 2011). Further studies also demonstrate that E_α serves as a boundary between the *Tcra-Tcrd* locus and downstream genes such as *Dad1*. When the CBE associated with E_α was experimentally deleted in mice, distal J_α utilization was reduced and interactions with *Dad1* increased (Zhao et al., 2020). Cohesin, therefore, appears necessary to restrict E_α activity by establishing the TEA-E_α CBE-anchored loop

under physiological conditions, wherein E_α serves to enhance J_α transcription and recombination, but does not enhance *Dad1*.

E_α has also been observed by 3C to interact with both V_α and J_α segments (Shih et al., 2012). In addition to binding in the reverse CBE orientation at E_α, CTCF binds many promoters in the *Tcra-Tcrd* locus in the forward CBE orientation. Thereby, E_α has an increased probability of interacting with these promoters. Experimental deletion of TEA shifts the interaction between TEA and E_α to focus E_α on promoters associated with more central J_α segments. Experimental depletion of CTCF reduces interactions between E_α and V_α and J_α segments but increases interactions with D_δ and J_δ segments. Reductions in recombinations between proximal V_α and J_α segments were observed in thymocytes experimentally depleted of CTCF (Shih et al., 2012). These changes in structure and rearrangement mediated by CTCF binding may be the result of changes to loops that facilitate diffusion-mediated synapsis, or they may reflect cohesin-mediated loop extrusion and RAG tracking. RAG might be slipping past the structural boundaries that typically restrict the RC to J_α segments.

To date, little data exists to explain whether loop extrusion, diffusion, or a combination of both mechanisms are responsible for recombinations of *Tcra-Tcrd*. Intriguingly, because two different receptor chains rearrange at this locus in two different stages of differentiation, the locus may switch forms of RSS synapsis over the course of development. Because the pattern of recombination events in DN cells is different from that in DP cells, there is a high degree of probability that synapsis is not conformationally

mediated in the same way for *Tcrd* as it is for *Tcra*. Recombination of the locus will be discussed below, but, briefly, *Tcrd* recombinations may involve distal V_δ segments, while *Tcra* recombinations rarely involve distal V_α segments.

1.5 Antigen receptor repertoires

1.5.1 *Igh* repertoire

Utilization of V_H gene segments is unequally distributed. Individual V_H genes have different intrinsic utilization frequencies. The V_H segments are divided into families along the locus, and despite clustered localization and sequence similarity among V_H gene segments, individual members of these families are utilized at different frequencies (Choi et al., 2013). Factors driving the use of particular V_H segments include RSS quality, chromatin modifications, and transcription factor binding (Feeney, 2009). Utilizing assays with recombination substrates to study competition between V_H segments for recombination, V_H utilization was demonstrated to depend partially upon RSS sequence. However, not all recombination preference was due to RSS strength; genes closer to the RC were utilized more frequently than segments with similar RSSs (Williams et al., 2001; Montalbano et al., 2003). V_H genes very proximal to the RC are used at high frequency, particularly the V_H81X (Yancopoulos et al., 1984; Perlmutter et al., 1985). Accessibility also depends on histone modifications, with V_H segment acetylation correlating with relative rearrangement (Espinoza and Feeney, 2005).

The D_H gene segment repertoire is regulated by several additional mechanisms. The most J_H-proximal and J_H-distal unique D_H gene segments are utilized more

frequently than the central D_H segments (Tsukada et al., 1990; Choi et al., 2013). This has been explained by a combination of RSS strength, localization, and epigenetic silencing of the highly repetitive central DH segments (Chang et al., 1992; Chakraborty et al., 2007). Due to the 3D topology of the *Igh* locus, these D_H gene segments have been demonstrated to frequently be in contact with the J_H array (Degner et al., 2011). Although the most J_H-proximal D_H segment, DQ52, is captured by RC-bound RAG by diffusion (Zhang et al., 2019), high utilization of the most distal D_H segment may be enhanced by paused RAG scanning due to blockade of loop extrusion by the nearby CBE.

Eμ is also an important regulator of *Igh* recombination. Eμ promotes GLT, accessibility, and D-to-J_H recombination within the *Igh* RC. However, residual D-to-J_H recombination is observed on alleles with experimental Eμ deletions (Afshar et al., 2006). V-to-D_H recombinations are nearly ablated in these experiments; it appears likely that this effect is downstream of reduced efficiency of D-to-J_H recombinations, as GLT of V_H segments is not impacted by loss of Eμ (Perlot et al., 2005). Experimental deletion of the core Eμ demonstrated the role of Eμ in recruiting RNA polymerase II and deposition of H3K4me3 at the D_H and J_H segments (Chakraborty et al., 2009).

The 3D structure of the *Igh* locus (as discussed above) also significantly impacts recombination. The strict enforcement of D-to-J_H recombination prior to V-to-DJ_H recombination appears to be mediated by RAG scanning, which allows deletional D-to-J_H recombination within the RC. A comparison between CBE localization and V_H segment utilization demonstrated that V_H segments proximal to a CBE were more likely to be recombined (Choi et al., 2013). This correlation between CTCF binding and V_H

utilization suggests that both topology and impediments to RAG tracking may be significant factors in V_H capture. Diffusional recombination events between distal segments may be mediated by 3D loops that bring V_H segments into proximity of the RC, but most V_H capture appears to occur by RAG scanning from the RC (Hill et al., 2020; Dai et al., 2021).

1.5.2 *Igk* and *Igl* repertoires

It appears that *Igk* recombinations, which occur prior to *Igl* recombinations, are typically successful within several possible rounds, as Igλ chains are very uncommon in the mouse BCR repertoire (Cotner and Eisen, 1978). The fetal *Igk* repertoire displays diverse utilization of V_κ segments, but it is restricted compared to recombinations in the adult mouse. Most *Igk* recombinations in fetal liver appear to occur by inversion. There is also a significant bias for particular V_κ-J_κ recombinations, and the centrally-located $V_\kappa 4$ family is frequently utilized (Medina and Teale, 1993; Ramsden et al., 1994). The $V_\kappa 4$ family has associated promoter elements that differ from other V_κ promoters in addition to putative transcription factor binding sites, which may account for their utilization in fetal recombinations (Brekke and Garrard, 2004). $V_\kappa 4$ family utilization in adult *Igk* recombination is uncommon (Aoki-Ota et al., 2012). Despite the diversity of the adult *Igk* repertoire, certain patterns of recombination are observed.

The *Igk* locus is capable of undergoing multiple rounds of recombination in a process known as receptor editing. The first J_κ segment used is often *Igkj1*, the most V_κ-

proximal J_κ segment (Yamagami et al., 1999). Initial rearrangement to *Igkj1* is proposed to preserve further J_κ segments for further recombinations. The bias toward utilization of *Igkj1* is due to the activity of a proximal promoter. In mice with an experimental deletion of the proximal promoter, more distal J_κ segments are marked by H3K4me3 and they are utilized more frequently (Vettermann et al., 2015). By repressing utilization of distal J_κ segments, the J_κ-proximal promoter may serve a role akin to the role of TEA in *Tcra* recombination in stimulating proximal J utilization in early recombinations to increase the potential for recombinatorial diversity and receptor editing.

Nearly all functional V_κ segments can recombine with any J_κ segment (Aoki-Ota et al., 2012). This may be due to the 3D structure of the *Igk* locus in pre-B cells, in which all functional V_κ segments are in close proximity. Such topology may facilitate equal probability diffusion-mediated RSS capture (Lin et al., 2012). However, not all recombinations occur with equivalent frequency. 43% of recombinations involved only seven V_κ segments, and these segments were biased to recombine with particular J_κ segments (Aoki-Ota et al., 2012). Early recombinations using *Igkj1* were biased toward either deletional recombinations with proximal V_κ segments or inversional rearrangements that would similarly retain a diverse array of V_κ segments (Aoki-Ota et al., 2012). These findings cannot be accounted for by strength of RSS or V_κ-associated promoter (Fitzsimmons et al., 1998). Intriguingly, an additional layer of recombinatorial diversity at the *Igk* locus appears to be facilitated by cryptic RSS segments in the V_κ array that facilitate non-functional intra-V_κ recombinations. These recombinations often

occur at V_κ segments with strong RSS sequences, which would delete these segments and balance V_κ utilization among segments with varying RSS strength (Shinoda et al., 2019). Primary recombinations involving proximal V_κ and J_κ segments maintains most segments for further recombination. Inversional primary recombinations would scramble the V_κ array, lending further diversity to secondary recombinations. Intra-V_κ recombinations would remove a bias toward V_κ segments with strong RSSs in the repertoire. Together, this data suggests that the program of *Igk* recombination facilitates retention of a diverse array of segments for multiple rounds of recombination.

Igl recombinations are not only rare, but relatively unvaried. *Iglv1* nearly always recombines with the proximal $JC_\lambda1$ or $JC_\lambda3$ clusters. *Iglv2* and *Iglv3* almost exclusively recombine with the proximal $JC_\lambda2$ cluster. However, rare exceptions to this pattern of recombination between V_λ segments and the most proximal JC_λ clusters have been observed (Reilly et al., 1984; Sanchez et al., 1987, 1991). One enhancer is associated with each JC_λ cluster, and they are responsible for transcriptional activation necessary for *Igl* recombination. These enhancers are active in *Igl* producing, but not *Igk* producing, cells (Bich-Thuy and Queen, 1989; Hagman et al., 1990; Eccles et al., 1990). This suggests that silencing of *Igl* enhancers during *Igk* recombination facilitates early exclusion of *Igl* recombination.

1.5.3 *Tcrb* repertoire

Tcrb recombination is regulated by mechanisms similar to many other AgR loci. However, the locus has unique features that set it apart. Interestingly, recombination between V_β and J_β segments without D_β segments is theoretically feasible following the 12/23 rule due to the orientation of RSSs at the *Tcrb* locus. However, V-to-J_β recombinations are not frequently observed due to control mechanisms beyond 12/23 restriction (Sleckman et al., 2000; Bassing et al., 2000).

The RC of the *Tcrb* locus is focused on two clusters of D_β, J_β, and C_β segments. E_β controls transcription of the *Tcrb* RC by stimulating the promoters upstream of each DJC_β cluster (McDougall et al., 1988). The DJC_β promoters are essential for activating transcription and accessibility of their respective DJC_β clusters (McMillan and Sikes, 2008; Whitehurst et al., 2000). Each V_β segment also has an associated promoter, but GLT of these segments is not E_β-dependent (Mathieu et al., 2000; Gopalakrishnan et al., 2013). Via repertoire analysis and computational modeling of factors controlling *Tcrb* repertoire, it was determined that RSS quality, transcriptional activation, and nucleosome depletion drive the use of particular V_β segments (Gopalakrishnan et al., 2013).

Although both allelic inclusion and two recombinations on a single *Tcrb* allele are theoretically possible, these outcomes are rarely observed. In addition to localization to the nuclear lamina (discussed above), allelic exclusion has been demonstrated to rely on poor RSS quality of V_β segments (Wu et al., 2020; Wu and Bassing, 2020). Previous RSS replacement experiments have demonstrated that replacing a poor V_β RSS with a stronger

RSS significantly increases utilization of the segment, altering the *Tcrb* repertoire (Wu et al., 2003). Interestingly, these experiments were performed on the downstream *Trvb31* segment, which is in the opposite transcriptional orientation of the other *Tcrb* gene segments, facilitating its utilization by inversional recombination. Because this segment can be used in inversional recombination and the locus has two unique DJC_β clusters, two recombinations on the same *Tcrb* allele are possible in addition to potential allelic inclusion (Lee and Bassing, 2020). These recombinations are rarely observed, and poor RSS quality has been demonstrated to be a mechanism limiting multiple rearrangements (Wu and Bassing, 2020; Wu et al., 2020). The poor quality of RSSs in the V_β array appear to reduce the rate of recombination to allow feedback inhibition by pre-TCR signaling before multiple *Tcrb* recombinations can occur in a single cell.

The topological regulation of *Tcrb* recombination is focused on intergenic CBEs upstream of the RC that regulate the structure and recombination of the *Tcrb* locus. The most 5' of these CBEs appears to anchor a loop bringing distal V_β segments into proximity of the RC (Majumder et al., 2015). The many CBEs in the V_β array are convergently oriented with this CBE, potentiating loop extrusion that may facilitate variable proximity to the RC of individual V_β segments. The other two CBEs proximal to the RC function together as a topological boundary upstream of the RC. When these CBEs are experimentally deleted, active chromatin marks spread upstream of the RC and distal V_β segments no longer come into proximity of the RC. Utilization of proximal V_β segments is not impacted, despite reductions in distal V_β use when these CBEs are

deleted (Majumder et al., 2015), suggesting that physiological RAG tracking from the *Tcrb* RC is blocked by CBEs upstream of the RC. Additionally, in physiological conditions, distance from the RC does not predict V_β utilization (Gopalakrishnan et al., 2013). To date, evidence on V_β capture by the T*crb* RC is incomplete. RAG tracking is possible, but a combination of inhibition by CBEs and nuclear lamina association makes this seem unlikely. Diffusion-mediated capture and synapsis of V_β segments variably brought into proximity of the RC by cohesin-mediated loop extrusion seems more likely. Topological features combined with weak RSSs may contribute to V_β diversity as segments brought into proximity of the RC are sub-optimal RAG substrates, which may diversify diffusion-mediated capture of V_β RSSs and allow utilization of V_β segments across the array.

1.5.4 *Tcrg and Tcrd* repertoires

Rearrangement of the *Tcra-Tcrd* locus is unique among AgR loci. It is the only AgR locus containing gene segments for two receptor chains. These chains rearrange at different stages of development, with *Tcrd* recombination in DN cells and *Tcra* recombination in DP cells. Further, *Tcrd* recombination involves V_δ, D_δ, and J_δ segments, while *Tcra* rearrangements do not include D segments. Regulatory mechanisms controlling recombination shift over the course of development.

Interestingly, *Tcrd* recombination does not proceed in as orderly a fashion as other loci with D segments, such as *Igh*. Rather than a strict program of D_δ-to-J_δ then V_δ-

to-DJ_δ recombination, many *Tcrd* recombination events proceed D_δ-to-D_δ, D_δ-to-DJ_δ, or V_δ-to-D_δ (Chien et al., 1987). Furthermore, *Tcrd* recombination does not seem to be hampered by allelic exclusion in mice; over 60% of splenic γδT cell hybridomas interrogated demonstrated complete rearrangement on both alleles, including a significant proportion with two in-frame rearrangements (Sleckman et al., 1998).

The γδTCR repertoire is regulated partly by ontogenic timing. In fetal life, specific V_δ and V_γ segments are used frequently that are never used in adult γδT cell development. The mature peripheral γδTCR repertoire includes a number of invariant or semi-invariant chains rearranged during fetal life. These γδT cells are distributed in a tissue-specific fashion. Fetal rearrangements do not include N nucleotide insertion, as TdT is not yet expressed (Rothenberg and Triglia, 1983).

As early as embryonic day 12 (E12) cells begin to seed the fetal thymus. These earliest thymic T cells arise from the fetal liver and give rise to the invariant *Trgv5*-*Trdv4* subset, peaking around E14. These cells emigrate from the thymus to comprise the majority of T cells in the skin and are known as dendritic epidermal T cells (DETCs) (Havran et al., 1991; Carding and Egan, 2002; Havran and Allison, 1988, 1990). Adult stem cells are incapable of generating DETCs (Ikuta et al., 1990). At E14 another subset of γδT cells arise in the thymus, carrying the invariant TCR comprised of *Trgv6*-*Trdv4*, development of which peaks around E16. These cells then home to mucosal epithelial tissues, particularly in the female reproductive tract, lungs, liver, kidney, and tongue (Itohara et al., 1990; Carding and Egan, 2002). These fetal recombinations are blocked by

E2A in adult cells. E2A is required for efficient *Tcrd* recombination throughout ontogeny, but dose effects determine which segments are available for recombination as the organism develops (Bain et al., 1999).

At the *Tcrg* locus, a combination of GLT, promoter activity, accessibility, and locus position mediate ontogenic timing of V_γ utilization. GLT of V_γ genes correlates with their pattern of rearrangement over the course of development (Goldman et al., 1993). In adult thymocytes, an experiment swapping the promoters of *Trgv4*, which is used in adult recombinations, and *Trgv5*, which is used exclusively in fetal recombinations, reversed the typical pattern of recombination and GLT of these segments. However, V_γ fetal utilization was not impacted by swapping these promoters (Baker et al., 1998). During fetal development, both *Trgv4* and *Trgv5* are accessible and transcriptionally active. In adult thymocytes, *Trgv4* GLT and histone acetylation are high, while *Trgv5* GLT and acetylation are reduced. Using an experimental inhibitor of histone deacetylation, *Trgv5* rearrangements could be induced in adult thymocytes (Agata et al., 2001). In another study, the *Trgv4* and *Trgv5* genes were swapped; this construct switched the position of the promoters, coding regions, and RSSs of these segments. Switching the position of these genes increased fetal utilization of *Trgv4*, which was positioned proximal to *Trgj1* in this construct. Switching the locations of these genes did not change accessibility or GLT, so the change in utilization appears to be directly related to the position of these genes with respect to *Trgj1* (Xiong et al., 2004). In fact, *Trgv5* and *Trgv6*, which are the most *Trgj1*-proximal V_γ genes under physiological conditions are the most frequently used segments in fetal *Tcrg* recombination. When these genes

were experimentally deleted, the more distal upstream *Trgv4* and *Trgv7* genes were used in their place, suggesting that gene location controls fetal V_γ utilization (Xiong et al., 2008). These findings that V_γ segments proximal to the $JC_\gamma1$ cluster are preferentially used in fetal thymocytes when segments are equivalently transcribed and accessible suggests that RAG tracking is a potential mechanism of recombination at the *Tcrg* locus (Xiong and Raulet, 2007). In adult thymocytes, it appears that V_γ segments commonly used in fetal recombination are not utilized due to depressed accessibility and GLT.

The adult *Tcrd* repertoire is more variable than the fetal repertoire (Elliott et al., 1988; Krangel et al., 1990). As demonstrated in γδT cell hybridomas, *Trdv2-2*, *Trdv5*, and *Trav15-dv6* family members are commonly used in the adult *Tcrd* repertoire to the exclusion of most other V_δ segments (Sleckman et al., 1998; Pereira et al., 2000; Weber-Arden et al., 2000). Using linear amplification-mediated high-throughput genome-wide sequencing (LAM-HTGTS), regions targeted in specific recombinations can be detected (Hu et al., 2016). LAM-HTGTS assays revealed that *Trdv2-2*, *Trdv5*, *Trav15-dv6* family members, and the pseudogene *Trdv3* are frequent recombination targets, while *Trdv4* is very rarely utilized in adult *Tcrd* recombinations (Zhao et al., 2016). These studies also confirmed previous findings that *Trdj1* is used much more frequently in adults than *Trdj2* (Chien et al., 1987; Zhao et al., 2016). In adult DN thymocytes, frequently utilized V_δ segments are hyperacetylated, which is a marker of transcriptional activation. Segments used with less frequency in adult *Tcrd* recombination, such as *Trdv4* and distal V_δ segments other than *Trav15-dv6* family members, are relatively deacetylated (Hawwari

and Krangel, 2005). A potential mechanism of preferential accessibility of *Trdv4* in fetal life is the activity of E_δ. In adult DN thymocytes, E_δ upregulates GLT of the *Tcrd* RC region, encompassing the D_δ and J_δ segments as well as the downstream *Trdv5*. However, in fetal life the influence of E_δ extends upstream to induce recombination using *Trdv4* (Hao and Krangel, 2011; Weber-Arden et al., 2000).

Intriguingly, utilization of a single D_δ segment, like orderly recombination steps, is not strictly enforced in *Tcrd* recombination as it is in recombinations of other loci with D segments (Chien et al., 1987). Via LAM-HTGTS, V_δ use in DN thymocytes was tracked. *Trdd2* is capable of recombining out of order. It can join to *Trdd1*, V_δ segments, or J_δ segments. However, *Trdd1* always recombines with *Trdd2* prior to any other recombination event (Zhao et al., 2016). *Trdd2* is highly accessible in DN thymocytes and has a strong promoter; thus, the first RC at the *Tcra-Tcrd* locus forms around *Trdd2* (Carabana et al., 2005; Hao and Krangel, 2011; Teng et al., 2015). *Trdd2* is used in most fetal rearrangements to the exclusion of *Trdd1*; later in ontogeny D_δ use expands to include both *Trdd1* and *Trdd2* (Chien et al., 1987; Elliott et al., 1988; Itohara et al., 1993; Parker and Ciofani, 2020).

Changing the 3D structure of the *Tcra-Tcrd* locus in DN thymocytes by experimentally removing the intergenic CBEs upstream of *Trdv4* perturbs the *Tcrd* repertoire (Chen et al., 2015). In these studies, two mouse models with perturbations of the *Tcra-Tcrd* locus were employed. One line lacked both INT1 and INT2; the other lacked only INT2. In the absence of INT1-2, the loop around the *Tcrd* RC is perturbed.

Trdv2-2, which is the most proximal V_δ segment upstream of INT1-2, is utilized at higher frequency when this boundary is experimentally removed (Chen et al., 2015). Deletion of INT1-2 did not impact fetal or adult *Trdv4* use, despite its position within the INT2-TEA loop (Chen et al., 2015). Further inquiry using LAM-HTGTS revealed that on INT1-2 mutant alleles, the pseudogene *Trdv3* and forward-oriented cryptic RSSs upstream of INT1-2 also became frequent recombination partners of *Trdd1* and *Trdd2* (Zhao et al., 2016). Rearrangements to more distal upstream V_δ segments are reduced on INT1-2 mutant alleles, as are rearrangements to the *Tcrd* RC-proximal downstream *Trdv5* segment. When INT2 is deleted, INT1 is able to partially compensate by forming an upstream boundary of the RC, interacting with TEA; however, INT1 alone does not form as strong a boundary as INT1-2. *Trdv2-2* was utilized more frequently when INT2 was mutated. Further, a reduction in rearrangements to both the *Trdv5* segment, which is downstream of the *Tcrd* RC, and *Trav15-dv6*, which is upstream and distal to the *Tcrd* RC, was observed in the absence of INT2. The INT1 and INT2 CBEs appear to have at least partially redundant function in regulating *Tcrd* recombination by establishing the structure of the *Tcra-Tcrd* locus in DN thymocytes (Chen et al., 2015; Zhao et al., 2016). These studies have not elucidated the role of INT1 when INT2 is intact.

1.5.5 *Traj* repertoire

Tcra recombination occurs in multiple rounds, replacing out of frame rearrangements and those that generate non-selectable TCRs, with no allelic exclusion (Malissen et al., 1992, 1988). The first indication of multiple rounds of subsequent *Tcra*

recombination in DP thymocytes came from studies of extrachromosomal circular DNA, representing circularized signal joints with intervening sequence containing DNA from a previous recombination. These studies demonstrated that not only had *Tcrd* recombination occurred on the same alleles as *Tcra* recombinations, but that successive rounds of deletional *Tcra* recombination were ongoing as well (Okazaki and Sakano, 1988; Toda et al., 1988; Takeshita et al., 1989). Many of the early rounds of *Tcra* recombination that were observed were out of frame (Okazaki and Sakano, 1988). Further studies revealed *in vivo* V-to-J_α rearrangements of endogenous segments when a rearrangement was knocked in, proving successive rearrangements enhance the opportunity for positive selection (Wang et al., 1998).

Recombinations of *Tcra* are non-random. Particularly, the order of J_α use is established as moving from V_α-proximal in early recombinations to more distal J_α segments in further rounds of recombination. Early in ontogeny, in fetal and neonatal life, the most proximal J_α segments were found to be utilized in recombination. If only one allele had undergone *Tcra* recombination, it always used V_α-proximal J_α segments. When two rearranged alleles occurred in the same cell, rearrangements were usually synchronized, using J_α segments in the same portion of the array; these findings also confirm the lack of allelic exclusion during *Tcra* recombination (Thompson et al., 1990). Frequently, primary *Tcra* rearrangements are to the most proximal J_α, *Traj61*, but this gene segment has defective splice sites (Villey et al., 1997), so further rearrangements

take place to make a rearrangement that can splice to C_α and be expressed as a functional TCRα chain.

This pattern of utilization of proximal J_α segments in early recombinations is retained in adult life in mice (Shih et al., 2012). TEA is required for utilization of the most proximal J_α segments in recombination; the transcriptional activation spurred by TEA focuses primary recombinations on the proximal portion of the J_α array (Villey et al., 1996). TEA-mediated upregulation of transcription and accessibility of the most proximal J_α segments also actively suppresses the use of central and distal J_α segments in primary recombinations. The mechanism of suppression of more distal promoters is transcriptional interference (Abarrategui and Krangel, 2007).

Cells with reduced capacity for ongoing recombination, such as those with defective RAG expression or reduced thymocyte lifespan have a repertoire skewed toward use of more proximal J_α segments (Yannoutsos et al., 2001; Guo et al., 2002), reinforcing the notion that these are the earliest recombinations. As rounds of recombination continue, J_α use proceeds to central and distal segments (Carico et al., 2017). Indeed, DP cells with extended lifespans that potentiate further rounds of recombination display a bias toward increased distal J_α use (Petrie et al., 1995; Guo et al., 2002).

To study *Tcra* recombination over time, a mouse model with Tamoxifen-inducible ER-Cre controlled by C_δ transcription and a fl-Stop-fl ZsGreen reporter (Zhang et al., 2015) was used. Tamoxifen injection labels cells transcribing *Tcrd*, which are

predominantly in the DN compartment; these cells can be tracked as they enter the DP compartment. Early J_α use was skewed, in DP thymocytes, to the proximal half of the J_α array, with ~30% of rearrangements twelve hours after Tamoxifen injection utilizing the three most proximal functional J_α segments, *Traj58*, *Traj57*, and *Traj56* (Carico et al., 2017). Therefore, the pattern of J_α utilization starts with V_α-proximal J_α segments internal to the locus and continues, using ever more distal J_α segments in successive rounds of recombination.

1.5.6 *Trav* repertoire

Due to the lack of allelic exclusion and multiple rounds of rearrangement possible on the *Tcra-Tcrd* locus, it has been hypothesized that V_α use, like J_α use, should occur from the interior of the locus to the exterior. Although this pattern of recombinations involving the most proximal available segments in each round of recombination would be beneficial in maintaining genes for further rearrangements, it would limit the proportion of alleles using distal V_α segments, thus reducing combinatorial diversity. PCR-based studies of T cell clones found frequent proximal V_α-to-proximal J_α early rearrangements. Ongoing rounds of recombination used more distal V_α and J_α segments, with central V_α-to-central J_α and distal V_α-to-distal J_α recombinations observed (Huang and Kanagawa, 2001; Aude-Garcia et al., 2001).

However, some research using HTS of the *Tcra* repertoire has suggested that V_α segments are used with relatively equal frequency across the V_α array (Genolet et al.,

2012). These studies relied on sequence information from post-selection TCRs, which inherently biases the repertoire away from an indicator of recombination *per se*. Further, the highly repetitive nature of the C57BL/6 locus hindered perfect alignment; not all V_α reads could be definitively assigned to an individual segment when they aligned to V_α families with high sequence similarity between member segments. As such, the authors assigned these ambiguous reads proportionally to each member of the V_α family they aligned with, providing recombination frequencies which may not be biologically valid (Genolet et al., 2012). These results were not replicated in research on the 129 locus, which, due to containing one fewer duplicated region, is more amenable to definitive assignment of reads to individual members of V_α families. Using pre-selection strain 129 DP thymocytes to avoid skewing of the repertoire by selection, the pattern of recombination conformed to the previous findings that V_α segments, like J_α segments, are predominantly used from proximal to distal, or from the interior of the locus toward the ends. Although most rearrangements with proximal J_α segments involved the proximal half of the V_α array, distal V_α segments were utilized in a proportion of recombinations with proximal J_α segments. Secondary rearrangements of the remaining V_α segments generally proceed from J_α-proximal to J_α-distal (Carico et al., 2017). Early proximal V_α utilization is likely regulated by E_α, which is responsible for increased accessibility (as marked by acetylation) and GLT of proximal V_α segments immediately after the DN to DP transition (Hawwari et al., 2005), but E_α alone cannot explain the use of central and distal V_α segments in early recombinations.

Due to the nested organization of *Tcrd* segments within the *Tcra* locus, it has been proposed that *Tcrd* recombinations in DN thymocytes may contribute to the diversity of V_α use in early *Tcra* recombinations. As V_δ segments are distributed along the V_α array, deletional *Tcrd* recombinations can truncate the V_α array, and the use of more distal V_δ segments in *Tcrd* recombination would leave only distal V_α segments for primary *Tcra* recombination events on some proportion of alleles (Carico et al., 2017; Chen et al., 2015). Two pieces of evidence support this supposition. One is that deletion of the intergenic CBEs, INT1-2, upstream of *Trdv4*, disrupts not only *Tcrd*, but *Tcra* recombination as well. There is a reduction in recombination of central V_α segments with proximal J_α segments. There is a concomitant increase in proximal V_α-to-proximal J_α recombinations (Chen et al., 2015; Carico et al., 2017). This may reflect a change in repertoire on alleles that have not recombined *Tcrd*, in which the 3D structure of the locus predisposes recombinations to occur using proximal V_α and J_α segments. More likely, however, as *Tcrd* recombinations occur on most *Tcra-Tcrd* loci (Shih et al., 2012), is that the reduction in distal V_δ recombinations on loci with disrupted 3D organization decreases the propensity for distal V_α use in early recombinations with proximal J_α segments. A second piece of evidence that *Tcrd* recombination diversifies V_α use in *Tcra* recombinations comes from a study of secondary recombinations on alleles incapable of *Tcrd* recombination due to knockin of a fixed primary *Tcra* recombination, which deletes the D_δ, J_δ, C_δ, and several proximal V_δ segments. In these mice, the *Tcra* repertoire is constrained, with minimal utilization of central and distal V_α segments in early

recombinations with proximal J_α segments (Carico et al., 2017). In summary, V_α use in secondary rearrangements appears to generally progress from J_α-proximal to J_α-distal, but primary V_α use is more varied. This variability may be due to deletional *Tcrd* recombination prior to *Tcra* rearrangement. The mechanisms diversifying V_α use warrant further exploration.

2. Specific Aims

Chromatin architecture shapes interactions throughout the genome. Antigen receptor (AgR) loci in adaptive immune cells require developmentally coordinated topological changes to systematically create functional yet diverse repertoires. A key mediator of chromatin interactions required for recombination is CCCTC-binding factor (CTCF). CTCF binding elements (CBEs) define the boundaries of topological structures such as chromatin loops. Loops can sequester variable (V), diversity (D), and joining (J) segments or bring these segments into proximity for recombination. Recombination of the T cell receptor (TCR) alpha-delta locus (*Tcra-Tcrd*) requires tight regulation of its complex nested organization. *Tcrd* recombination, which precedes *Tcra* recombination on the *Tcra-Tcrd* locus, is regulated by a CTCF-mediated loop surrounding the DJ_δ cluster. Two CBEs at the 5' end of this loop control V_δ interactions with the DJ_δ cluster, but their individual roles are undetermined.

In the αβT cell lineage, most *Tcra-Tcrd* alleles have undergone *Tcrd* recombination prior to *Tcra* rearrangement. While individual *Tcra* recombinations appear to use the most proximal V_α and J_α segments in successive rounds of rearrangement, V_α use in the *Tcra* repertoire is surprisingly diverse. Although *Tcrd* recombinations predominantly involve proximal V_δ segments, members of the more distal *Trav15-dv6* family are used more frequently than other distal segments. Because V_δ segments are interspersed within the V_α array, *Tcrd* recombination changes the structure of the *Tcra-Tcrd* locus, but the impact of *Tcrd* recombination on the *Tcra* repertoire is unknown.

We hypothesized that *Tcrd* recombination truncates the *Tcra-Tcrd* locus, diversifying *Tcra* repertoire formation, which would otherwise proceed using the most proximal segments in each successive round of recombination. We also hypothesized that CTCF-mediated topological domains regulate *Tcrd* recombination.

2.1 Specific Aim 1: To determine the impact of Tcrd recombination on the Tcra repertoire.

The impact of *Tcrd* recombination on *Tcra* recombination has been studied using a novel mouse, D_δ and J_δ deficient (DJD), that does not recombine *Tcrd*. Pre-selection thymocytes from these mice were subjected to high-throughput sequencing of *Tcra* repertoire. Early recombinations on these alleles were also analyzed using a Tamoxifen-inducible reporter system. Further, the impact on *Tcra* repertoire of utilization of central and distal V_δ segments, *Trav15-1-dv6-1* and *Trav15d-1-dv6d-1*, in *Tcrd* recombination was determined by *Tcra* repertoire analysis in mice lacking these gene segments.

2.2 Specific Aim 2: To determine the impact of the INT1 CBE on the Tcrd repertoire.

A novel mouse line with a mutation to intergenic CBE referred to as INT1 was established. CTCF binding at INT1 was determined to be reduced by chromatin immunoprecipitation. Thymocyte subpopulations and γδT cell development were analyzed by flow cytometry. The *Tcrd* repertoire was analyzed in these mice by quantitative PCR to determine the role of INT1 on V_δ utilization.

3. Materials and methods

3.1 Mice

3.1.1 Generation of DJD mice

To generate D_δ and J_δ deficient (DJD) mice, female C57BL/6J mice were mated to male strain 129 mice using a superovulation and timed mating strategy. The resulting F1 embryos were subject to pro-nuclear injection of reagents for two guide-mediated CRISPR/Cas9 (Cong et al., 2013; Singh et al., 2014) to delete a ~24.3kb region spanning *Trdd1-Trdj2* at the *Tcra-Tcrd* locus. **Table 1** contains the relevant guide sequences. Offspring were screened for deletion on the 129 *Tcra-Tcrd* allele using PCR and Sanger sequencing (Duke University DNA Analysis Facility). Appropriately targeted mice of mixed C57BL/6J and 129 genetic background were crossed once to strain 129 and DJD heterozygotes were then intercrossed to obtain homozygous WT and DJD littermates for analysis.

Table 1: CRISPR guide sequences for the generation of DJD mice

Guide	Sequence
Upstream	5'-AGTTTGATAACAGGTGAGTCT-3'
Downstream	5'-GCAGGTAATTTTAGCATTAT-3'

3.1.2 Generation of *Tcrd*^{CreER} *Trac*^{M} mice

Tcrd$^{CreER/CreER}$ *Rosa26*$^{ZsG/ZsG}$ mice, with modified strain 129 *Tcra-Tcrd* alleles on a C57BL/6 genetic background (backcrossed > 10 generations), were described previously (Zhang et al., 2015). These mice were further modified by deletion of 138 bp spanning the 5' portion of *Trac* exon 1 to generate *Tcrd*$^{CreER/CreER}$*Trac*$^{M/+}$ *Rosa26*$^{ZsG/ZsG}$

mice. Two-guide mediated CRISPR/Cas9 targeting was accomplished using electroporation of guides in **Table 2**. Offspring were screened by PCR and Sanger sequencing. The founder mouse was backcrossed once to *Tcrd*$^{CreER/CreER}$ *Rosa26*$^{ZsG/ZsG}$ and mice heterozygous for the *Trac* mutation mice were then intercrossed to obtain *Tcrd*$^{CreER/CreER}$*Trac*$^{M/M}$ *Rosa26*$^{ZsG/ZsG}$ mice. These were then crossed to DJD heterozygotes to obtain littermates containing a *Rosa26*^{ZsG} allele and *Tcrd*^{CreER} *Trac*^{M} allele paired with either a WT 129 or DJD 129 *Tcra-Tcrd* allele, on a mixed C57BL/6 and 129 genetic background.

Table 2: CRISPR guide sequences for the generation of *Tcrd*^{CreER} *Trac*^{M} mice

Guide	Sequence
Upstream	5'-GAGAGTTCCGCTCTTGCCTG-3'
Downstream	5'-GTTAAAAGATCCTCGGTCTC-3'

3.1.3 Generation of INT1M mice

To generate INT1M mice, superovulation was performed on C57BL/6J mice. CRISPR/Cas9 using one guide and a repair oligo was performed. The intention was to create a scrambled sequence incapable of CTCF binding at the site of an intergenic CTCF-binding element in the *Tcra-Tcrd* locus, 'INT1' (Shih et al., 2012). As such, the repair oligo contained an 18bp scrambled sequence insert (Guo et al., 2011b; Chen et al., 2015) flanked by 50bp homology arms to promote homology-directed repair. **Table 3** contains the relevant guide sequences. Unfortunately, the repair did not proceed as planned. Instead, mice with a small deletion in the region of CTCF binding were

obtained, and these were expanded to generate the INT1M line. These mice carry an 8bp deletion spanning a portion of the consensus CTCF-binding sequence.

Table 3: CRISPR guide sequences for the generation of INT1M mice

Guide	Sequence
INT1	5'-AGAGTCAGGGGACCACTAGA-3'
Repair Insert	5'-GACGAGAAGCTAGCAGTG-3'

3.1.4 Other mouse strains

Trav15-1-dv6-1 and *Trav15d-1-dv6d-1* deletions were generated by A. Mihai and Y. Zhuang in *Tcrd*$^{CreER/CreER}$ *Rosa26*$^{ZsG/ZsG}$ *Id3*$^{f/f}$ mice (Madisen et al., 2010; Guo et al., 2011c; Zhang et al., 2015) containing modified strain 129 *Tcra-Tcrd* alleles on a C57BL/6 genetic background (backcrossed > 10 generations). *Rag2*$^{-/-}$ mice on the C57BL/6J background were purchased from Jax. They were bred with INT1M mice for experiments in Chapter 5. Mice with an HYα knockin (HYα-KI) at the *Tcra-Tcrd* allele on chromosome 14 (Buch et al., 2002; Hawwari and Krangel, 2007), which are of a mixed 129 and C57BL/6J background, were bred with *Tcrd*$^{CreER/CreER}$*Trac*$^{M/M}$ *Rosa26*$^{ZsG/ZsG}$ mice. This breeding scheme generated *Rosa26*$^{ZsG/+}$ mice with two disparate *Tcra-Tcrd* alleles, one of which is HYα-KI and the other is *Tcrd*^{CreER}*Trac*^{M}. These mice facilitated analysis of *Tcra* recombination over time, with Tamoxifen-induced *Tcrd*^{CreER} expression driving ZsGreen expression in cells actively transcribing *Trdc*. *Trac*^{M} stopped primers from binding to the *Tcrd*^{CreER}*Trac*^{M} allele, which allowed analysis of *Tcra* repertoire on only the HYα-KI allele.

All CRISPR/Cas9-mediated deletions were performed by the Transgenic and Knockout Mouse Shared Resource at Duke University. Mice were sacrificed at 3-5 weeks of age. Both male and female mice were used; no differences were observed on the basis of sex. All mice were handled under protocols approved by the Duke University Institutional Animal Care and Use Committee and maintained in specific pathogen-free conditions.

3.2 Cell collection, flow cytometry, and fluorescence activated cell sorting

Thymi were collected from mice at 3-4 weeks of age for experiments involving DN thymocytes. Thymi were collected from mice at 4-5 weeks of age for experiments involving DP thymocytes. All thymi were processed into single cell suspensions in RPMI 1640 containing 10% (vol/vol) FBS. To label developing thymocytes, HYα-KI heterozygous *Tcrd*$^{CreER/+}$*Trac*$^{M/+}$ *Rosa26*$^{ZsG/+}$ and DJD heterozygous *Tcrd*$^{CreER/+}$*Trac*$^{M/+}$ *Rosa26*$^{ZsG/+}$, and control *Tcrd*$^{CreER/+}$*Trac*$^{M/+}$ *Rosa26*$^{ZsG/+}$ mice were injected i.p. with one dose of 1mg Tamoxifen (Sigma-Aldrich) in corn oil (Sigma-Aldrich) 8-72hrs prior to sacrifice.

All fluorophores were purchased from BioLegend unless otherwise indicated. Pre-selection DP thymocytes were sorted by staining with antibodies to CD4 (GK1.5), CD8α (53–6.7), CD3$_\varepsilon$ (145-2C11), 7AAD, and lineage (Lin) markers B220 (RA3-6B2), CD11b (M1/70), CD11c (N418), F4/80 (BM8), Gr-1 (RB6-8C5, Invitrogen), and Ter-119 (TER-119). For sorting, pre-selection DP thymocytes were defined as CD4$^+$CD8$^+$Lin$^-$7AAD$^-$

CD3ε^{lo}. Thymocytes from *Tcrd*$^{CreER/+}$*Trac*$^{M/+}$ *Rosa26*$^{ZsG/+}$ mice were also sorted for ZsGreen^{+}. Thymocyte analysis included the aforementioned antibodies as well as antibodies to CD44 (IM7), CD25 (PC61), cKit (2B8), and γδTCR (GL3). To sort DN2/3 thymocytes, thymocytes were stained with CD4 (GK1.5) and CD8α (53–6.7), followed by bead-based negative selection with Sheep Anti-Rat IgG Dynabeads (Life Technologies). Bead-depleted DN thymocytes were then stained with 7AAD, CD44 (IM7), CD25 (PC61), and Lin markers B220 (RA3-6B2), CD11b (M1/70), CD11c (N418), F4/80 (BM8), Gr-1 (RB6-8C5, Invitrogen), and Ter-119 (TER-119). For sorting, DN2/3 thymocytes were defined as CD4^{-}CD8^{-}Lin^{-}7AAD^{-}CD25^{+}CD44$^{+/-}$. Flow cytometry data were acquired on a BD FACSCanto II flow cytometer in 8-color configuration. Cell sorting was performed on either the Astrios (Beckman Coulter), MoFlo XDP (Beckman Coulter), or FACSDiVa (Becton Dickinson).

3.3 Tcra repertoire analysis

3.3.1 Library preparation and sequencing

Sorted pre-selection DP thymocytes were lysed in Trizol (ThermoFisher) per manufacturer's protocol and stored at -80°C until RNA extraction. Total RNA was extracted using Trizol per manufacturer's protocol. cDNA was prepared using 700ng of RNA per sample via template-switch 5' RACE with minor modifications to previously described methods (Pinto and Lindblad, 2010; Quigley et al., 2011; Carico et al., 2017). RNA, 1μl of 10μM 5' RACE Template-switch oligo adapter (5'-

GTCGCACGGTCCATCGCAGCAGTCACArGrGrG-3'), and 1µl of 10µM oligo$(dT)_{20}$ primer were added to nuclease-free water to a final volume of 12µl. This mixture was heated to 70°C for one minute then snap-annealed at -20°C for one minute. 4ul of 5X Superscript II First-Strand Buffer (ThermoFisher), 1ul of 0.1M DTT, 1µl of 10mM dNTPs, and 1µl of 100mM RNaseOUT (ThermoFisher), and 1µl of 200U/µl SuperScript II (ThermoFisher) were added at a final volume of 20µl. This reaction was incubated at 42°C for 2hrs and inactivated at 72°C for 7 minutes, synthesizing cDNA and adding the 5' RACE adapter by template switching.

5' RACE cDNA was amplified by PCR with minor modifications to described methods (Quigley et al., 2011; Carico et al., 2017). To amplify cDNA, each sample was split among 8 reactions of 50µl each. The 50µl reaction mixture consisted of final concentrations of 1U Kapa HiFi polymerase (Roche), 1x Kapa HiFi buffer (Roche), 0.3mM dNTPs, 0.3µM C_α reverse primer (5'-TACACAGCAGGTTCTGGGTTCTGGATGT-3'), 2nM Long RACE adapter primer (5'-ACGCTGACGCTGAGCCTACCTGACGTCGCACGGTCCATCGCAGCAGTC-3'), and 0.04µM Short RACE adapter primer (5'-ACGCTGACGCTGAGCCTACCTGAC-3') in nuclease-free water. Touchdown PCR was performed with one cycle at 98°C for three minutes, five cycles at 98°C for 20 seconds followed by 72°C for two minutes, five cycles at 98°C for 20 seconds followed by 70°C for two minutes, ten cycles at 98°C for 20 seconds followed by 30 seconds at 65°C and 72°C for two minutes, and one cycle at 72°C for 10 minutes. Products were pooled and QIAquick PCR purification (Qiagen) was

performed on PCR products, per manufacturer's specifications. PCR was then performed to add Illumina adapters and library-identifying barcodes using a slightly modified version of published protocols (Kozich et al., 2013; Carico et al., 2017), with each library split evenly among 8 reactions of 50µl total volume each. 50µl reactions consisted of final concentrations of 0.3µM each of forward and reverse adapter primers, 1U Kapa HiFi polymerase (Roche), 1x Kapa HiFi buffer (Roche), and 0.3mM dNTPs in nuclease-free water. Forward primers consisted of Illumina P5 adaptor, barcode, a pad sequence (increases melting temperature), and Short RACE adapter primer sequence (5'-AATGATACGGCGACCACCGAGATCTACAC-[8nt Nextera XT (Illumina) barcode]-TGTCGTCCTTACGCTGACGCTGAGCCTACCTGAC-3'). Reverse primer consisted of Illumina P7 adapter, Nextera X7 N701 barcode, a pad sequence, and the C_α reverse primer sequence (5'-CAAGCAGAAGACGGCATACGAGATTCGCCTTAAGTCAATCAATACACAGCAGGTTCTGGGTTCTGGATGT-3'). Products were subjected to a PCR with one cycle at 98°C for 30 seconds, 10 cycles of 98°C for 10 seconds followed by 65°C for 30 seconds and 72°C for 2 minutes, then one cycle at 72°C for 10 minutes. Products were pooled and kit-based PCR purification was performed. Library amplification was confirmed by gel electrophoresis.

Sequencing was performed as previously described (Carico et al., 2017), with minor modifications. Agilent Bioanalyzer analysis was used to determine library molarity and quality, and size selection was performed by the Duke University Sequencing and Genomic Technologies Shared Resource. Libraries with unique barcodes were pooled

and sequenced by the Duke University Sequencing and Genomic Technologies Shared Resource using 300nt paired-end reads on Illumina MiSeq (chemistry version 3). A PhiX control library and custom primers were added to the standard Illumina primer mix, as previously described (Carico et al., 2017). Custom primers included a Read 1 (5'-GTCGCACGGTCCATCGCAGCAGTCGGG-'3), Read 2 (5'-CAGAACCCAGAACCTGCTGTGTATTGATTGACT-'3), and P7 index read primer (5'- GCCTTAAGTCAATCAATACACAGCAGGTTCTGGGTTCTGGATGT-3'). The Duke University Sequencing and Genomic Technologies Shared Resource used Illumina MiSeq Reporter software to demultiplex libraries and assess their quality and yield. All primers and oligonucleotides were obtained, purified using standard desalting methods, from Integrated DNA Technologies; they were dissolved in nuclease-free water.

3.3.2 Analysis of repertoire data

Data was analyzed as previously described (Carico et al., 2017), with minor modifications. MiXCR (version 3.0.7) was used to align sequences and assemble clones (Bolotin et al., 2015). The reference library was edited to permit alignments only to 129 sequences, as determined by the international ImMunoGeneTics information system (IMGT) (Bosc and Lefranc, 2003). Sequencing reads were aligned to the library using the MiXCR 'align' command (Bolotin et al., 2015). The MiXCR 'assemble' command was used to identify unique clones with sequences spanning CDR2 through CDR3, capturing the V and J segments as well as the junctional sequence. As demonstrated in **Table 4**, sequencing was robust. All steady-state analyses included greater than 10^5 unique clones,

and all but the earliest 8hr HYα-KI$^{+/-}$*Tcrd*$^{CreER/+}$*Trac*$^{M/+}$ timepoint included analyses of at least 5 x 10^4 unique clones. All libraries analyzed contained over 10^6 reads passing MiXCR alignment QC. The MiXCR 'exportClones' command was used to generate a human-readable form of this data. Sequences aligning to pseudogenes and other very infrequently uses genes were manually removed, including *Traj61*, *Traj41*, *Traj25*, *Trav5d-2*, *Trav7d-6*, *Trav7-6*, and *Trav18*. *Trav11* and *Trav11d* are not distinguishable; both were maintained in this analysis, but the computed distribution of reads between the segments should be ignored. The VDJtools package was used to calculate the V and J use in unique clones (Shugay et al., 2015). 'PlotFancyVJUsage' was used to calculate clonal frequencies of V-to-J recombinations. Heatmaps were generated in R (version 3.3.3) (R Core Team, 2020) using the gplots (Warnes et al., 2009) and RColorBrewer (Neuwirth, 2014) packages. For difference maps, the WT repertoire was subtracted from the respective mutant repertoire.

Table 4: Quality of *Tcra* repertoire sequencing.

	MiXCR Aligned Reads		Unique *Tcra* Clones	
Sample	Replicate 1	Replicate 2	Replicate 1	Replicate 2
WT DP	2,724,476	2,196,267	541,510	396,722
DJD DP	2,137,553	2,430,268	388,593	377,173
ZsGreen$^+$ 12h WT DP	1,767,296	3,088,093	81,039	116,598
ZsGreen$^+$ 12h DJD DP	1,790,468	2,946,100	83,778	106,918
WT DP	2,461,779	3,175,598	262,146	324,511
Trav15-1-dv6-1 KO DP	2,741,475	2,453,179	223,783	264,507
WT DP	3,793,367	3,123,919	171,982	318,404
Trav15d-1-dv6d-1 KO DP	4,179,063	3,435,318	219,968	304,900
HYα-KI ZsGreen$^+$ 8h	1,437,116		46,124	
HYα-KI ZsGreen$^+$ 18h	2,125,663		80,146	

HYα-KI ZsGreen^{+} 30h	1,285,879		74,501	
HYα-KI ZsGreen^{+} 48h	1,472,467		43,396	
HYα-KI ZsGreen^{+} 72h	3,051,810		104,082	

3.4 PCR analysis of *Tcr* repertoire

For analysis of *Tcrd* rearrangement in INT1M and control mice, gDNA from DN2/3 cells was analyzed by Taqman-based real-time qPCR. Real-time semi-quantitative PCR assays were performed using the LightCycler 480 Probes Master (Roche). The PCR program ran at 95°C for 10 min, followed by 48 cycles of 95°C for 10 sec and 62°C for 30 sec, followed by a cooling step at 40°C for 30 sec. All gDNA rearrangements were analyzed to *Trdj1*, and all cDNA rearrangements were analyzed to *Trdc*. Data were normalized to the unrearranged genomic region *Cd14*. Data were further normalized to a standard curve of bulk gDNA from WT thymi to determine relative rearrangement frequency. **Table 5** contains relevant primer sequences.

Table 5: Primers used for *Tcrd* repertoire PCR analysis

Primer	Sequence
Trav15(d)-1-dv6(d)-1	5'-TCCATCAGCCTTGTCATTTCA-3'
Trav15(d)-1-dv6(d)-1 Probe	5'-CTCCCAGAGAGCACAGAAATACTTTCCCGAA-3'
Trav15(d)-2-dv6(d)-2	5'-GCAGCTTTTTAGTGGGAGAGATGG-3'
Trav15(d)-2-dv6(d)-2 Probe	5'-AGCCAGAGGATTCAGGGACGTACTTCT-3'
Trdv11	5'-TGCAACAGTGGGTCATTATTCT-3'
Trdv11 Probe	5'-CCATCGGACTCATCATCACCGCCACAC-3'
Trdv9	5'-TCTTACGATACACTGCAACTACTCA-3'
Trdv9 Probe	5'-ACTCGGCTGTGTACTACTGTGCTCTGG-3'
Trdv12	5'-GCTATTGCCTCTGACAGAAAGTC-3'
Trdv12 Probe	5'-GCACCTTGATCCTGCCTCATGTCAGCC-3'
Trdv1	5'-AAAGTAAGGACAAAATAACCGCTAAG-3'
Trdv1 Probe	5'-AGCCTCCCAGCCCAGCCACT-3'
Trdv2-2	5'-CCGCTTCTCTGTGAACTTCC-3'
Trdv2-2 Probe	5'-AAGCAGCTAAGTCCTTCAGCCTGGAGA-3'

Trdv3	5’-GCCAAAAATCTTAAAGAAAAGAACTTCACC-3’
Trdv3 Probe	5’-GCAGTGGTACATAGTGGAGTCTGATCACTGG-3’
Trdv4	5’-GGAAAGAGCAACCTCAAAGGG-3’
Trdv4 Probe	5’-CTGACCCACAGTAGTACGTACCAGCGTC-3’
Trdv5	5’-AACCTTCCATCTGGTGATCTCT-3’
Trdv5 Probe	5’-CAGTGAGCCTTGAAGACAGCGCTACTT -3’
$J_{\delta}1$ R	5’-CAGTCACTTGGGTTCCTTGTCC-3’
CD14 F	5’-GGCTCCGAATAGAATCCGACTA-3’
CD14 R	5’-GGCGGCAGATGTGGAATTG-3’
CD14 Probe	5’-CGCACCGTAAGCCGCTTTAAGGACAGA-3’
C_{δ}R	5’-CACCAGACAAGCAACATTTGTTCC-3’

3.5 Chromatin immunoprecipitation (ChIP)

Cells from INT1M$^{-/-}$*Rag2*$^{-/-}$ and *Rag2*$^{-/-}$ mice on the C57BL/6J background were crosslinked on ice for 10 min in RPMI 1640 (Gibco) containing 10% FBS (Gemini Bio-Products) and 1% paraformaldehyde (Electron Microscopy Sciences) at a concentration of 10^6 cells/ml. 0.125M final concentration of glycine was used to stop crosslinking. Cells were then lysed in a lysis buffer of 5mM PIPES (pH 8.0), 85mM KCl, 0.5% NP-40, 0.1mM PMSF, and 0.1mM Benzamidine. Chromatin was then sonicated using Sonicator 3000 (Qsonica) for 10 min and adjusted to IP buffer. IP buffer was composed of 0.01% SDS, 1.1% Triton-X, 1.2mM EDTA, 18.7mM Tris (pH 8.0), 167mM NaCl, 0.1mM PMSF (Sigma-Aldrich), and 0.1mM Benzamidine (Sigma-Aldrich). Preclearing was performed using protein A agarose/salmon sperm DNA slurry (Millipore) for 3-6 hours at 4°C. After spindown, sample was split into three aliquots. One was used as input control, one was used for control anti-Rabbit IgG antibody (ab-105-c; R&D Systems) precipitation, and one for the test antibody anti-CTCF (07-729; Millipore). 5µg of antibody was added, and samples were rocked at 4°C overnight. Chromatin-antibody

complexes were pulled down with protein A agarose/salmon sperm slurry. After centrifugation, supernatant was taken and processed for unbound sample. The bound pellets were washed and eluted. Crosslinks were reversed from the eluted pellets. DNA was extracted with phenol/chloroform purification and precipitated with ethanol. qPCR analysis was performed using the QuantiFast SYBR Green PCR kit (Qiagen). The program, performed on a Roche Lightcycler 480, was: 95°C for 5 minutes followed by 45 cycles of 95°C for 10 seconds and 62°C for 30 seconds. Binding calculations were as percent of input, and binding was normalized to binding at *Myc*. **Table 6** contains relevant primer sequences.

Table 6: Primers used for CTCF ChIP

Primer	Sequence
INT1 F	5'-TGCCTGTTTCTCCAAAGCTG-3'
INT1 R	5'-TCTTACCATCTGCTGCCATC-3'
INT2 F	5'-TAATCTGCACACATCCACCG-3'
INT2 R	5'-AACCGAAATGGAGCATCTGG-3'
TRDV4 F	5′-TCCCAAATACTATCTGGCCTG-3′
TRDV4 R	5′-TCTGCTGTAAAGGTGGTTTG-3′
TEA F	5'-TCCTTCCAGTGTCCTTTGAG-3'
TEA R	5'-CTCCAGTATGACCTGTTTATGG-3'
myc F	5'-GATGACCGGAAGCTTGTCTTAG-3'
myc R	5'-GGCTCTCGGATTTGTGAAAGTA-3'

3.6 E2A ChIP-seq analysis

These analyses were performed by Ariana Mihai. E2A ChIP-seq from sorted DP thymocytes ($CD1dTet^{-}CD4^{+}CD8^{+}$) isolated from $Id2^{f/f}$ $Id3^{f/f}$ *Lck*-Cre C57BL/6 mice was previously reported (Roy et al., 2018). Here, alignment was performed to mm10 using Bowtie2 (version 2.3.4.1), allowing for multiple read alignment (k = 3) to account for

sequence similarity within the *Tcra-Tcrd* locus; all other parameters were set to default. Peak-calling was performed using MACS2 (version 2.1.1.20160309); default options were used. Data visualization was performed using the Integrative Genomics Viewer (version 2.8.2).

3.7 Statistical methods

The analysis of global Shannon's Entropy was conducted by Dr. Liuyang Wang. Global V-J pair diversity was measured by Shannon entropy index (*H*). Shannon entropy quantifies both abundance and degree of unevenness of all distinct V-J pairs in a sample. A higher *H* value indicates more even distribution for distinct V-J pairs, while a lower value suggests a dominant V-J type occupies overall pairs. The Shannon diversity (H) was calculated using following equation:

$$H = -\sum_{i=1}^{S} p_i \log p_i$$

p_i is the fraction of *i*-th V-J pairs in a sample, and S is the cumulative combination of V-J pairs. We then normalized H to the maximum H_{max} using equation:

$$E_H = \frac{H}{H_{max}}$$

The normalized Shannon index E_H is bounded from 0 to 1, and a value of 1 means all V-J pairs have the same frequency.

To analyze regional changes in recombination, unique clones in each region were combined from two replicates per experiment. Change in the number of clones per region

was analyzed with two-tailed chi-square test with Yates's correction for continuity, Analyses were performed using Graphpad Prism 6 software.

Other statistical analyses were conducted using Graphpad Prism 6.0 software, as described.

4. Distal *Tcrd* recombinations diversify the *Tcra* repertoire by facilitating early distal V_α utilization

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

Dauphars, D.J., A. Mihai, L. Wang, Y. Zhuang, and M.S. Krangel. (In revision) *Trav15-dv6* family *Tcrd* rearrangements diversify the *Tcra* repertoire. *J. Exp. Med.* (A.M. and Y.Z. designed the *Trav15d-1-dv6d-1* KO and *Trav15-1-dv6-1* KO mice. A.M. performed the re-analysis of E2A binding at the *Tcra-Tcrd* locus. L.W. performed the Shannon's Entropy analysis and consulted on other statistics. Y.Z. and M.S.K. assisted in designing the study. M.S.K assisted in preparing the text and figures.)

4.1 Introduction

A cut-and-paste mechanism of RAG-mediated cleavage at the RSSs flanking V, D, and J gene segments and NHEJ repair facilitates immense combinatorial diversity of TCR and BCR repertoires during V(D)J recombination. The generated variable region is then spliced to a C segment to make an AgR chain (Schatz and Ji, 2011; Helmink and Sleckman, 2012). For V-to-(D)J recombination to occur, V segments must be brought into proximity of the RC (Jhunjhunwala et al., 2009; Lin et al., 2018). Ordered recombination occurs by a tightly regulated process that is integral to the generation of a diverse AgR repertoire. The use of individual V, D, and J gene segments to impart recombinatorial diversity on the AgR repertoire is therefore of great interest.

Unlike loci containing D gene segments, the *Tcra* locus, with its large number of V_α and J_α gene segments, can undergo multiple rounds of recombination to generate a function in-frame TCRa chain (Petrie et al., 1993; Wang et al., 1998). The use of J_α gene segments has been demonstrated, as described above, to occur from V_α-proximal to V_α-distal, maintaining the majority of J_α segments after the primary recombination for further rounds of secondary rearrangement (Thompson et al., 1990). Use of V_α gene segments, alternatively, has been demonstrated to be less constrained to the proximal segments. Early recombinations involve not only the proximal half of the V_α array, but widespread utilization of the central and distal V_α segments (Carico et al., 2017). However, secondary recombinations of an allele with a fixed primary rearrangement demonstrated orderly proximal-to-distal V_α utilization in secondary rearrangements commensurate with J_α use, in a pattern of recombinations moving through approximately 1.33 V_α segments per J_α segment ($4V_\alpha$:$3J_\alpha$) (Carico et al., 2017). A pattern of constrained V_α and J_α utilization beginning with primary recombinations would thus lead to a repertoire lacking the combinatorial diversity imparted by variable primary V_α utilization.

In thymocytes, *Tcrd* recombination, as well as *Tcrg* and *Tcrb* recombination, takes place at the DN stage (Capone et al., 1998). Later, in DP thymocytes, V-to-J_α recombination occurs for *Tcra* (Petrie et al., 1995). Most αβTCR expressing cells first undergo *Tcrd* recombination on at least one allele (Nakajima et al., 1995; Livak et al., 1995; Sleckman et al., 1998; Shih et al., 2012). Because *Tcrd* and *Tcra* share a single genomic locus, with D_δ and J_δ segments nested between the V_α and J_α segments and V_δ

segments interspersed along the V_α array, *Tcrd* recombinations can variably truncate the V_α array. This organization of the *Tcra-Tcrd* locus is found in both mouse and man (Glusman et al., 2001; Carico and Krangel, 2015). *Tcrd* recombination may consequently diversify V_α utilization in primary *Tcra* recombinations by increasing utilization of more distal V_α segments.

In a mouse model of altered chromatin looping at the Tcra-Tcrd locus, V-to-DJ_δ recombination was found to be skewed toward rearrangement of the proximal *Trdv2-2* and *Trdv3* gene segments, with concomitant reductions in rearrangements to distal V_δ segments (Chen et al., 2015; Zhao et al., 2016). *Tcra* rearrangements were also found to be less diverse, consistent with a role for *Tcrd* recombination in *Tcra* repertoire formation. However, because the primary effect of mutation in this model is on chromatin looping, it remains unclear whether the observed *Tcra* repertoire phenotype may be a direct consequence of this structural change as opposed to an indirect consequence of altered *Tcrd* rearrangements.

Proximal V_δ segments, the use of which does not delete any V_α segments, are the most frequently used segments for *Tcrd* recombination. However, one family of more distal V_δ segments, the *Trav15-dv6* family, accounts for most of the central and distal V_δ utilization (Chen et al., 2015; Zhao et al., 2016). Because *Tcrd* recombinations involving the *Trav15-dv6* family truncate the V_α array, they may fundamentally diversify the *Tcra* repertoire by allowing utilization of central and distal V_α segments in primary *Tcra* recombinations.

Here, we employ an HTS approach (Carico et al., 2017) to interrogate the *Tcra* repertoire. Utilizing a modified version of the tamoxifen-inducible *Tcrd*^{CreER} *Rosa26*^{ZsG} reporter (Zhang et al., 2015) with a mutation at *Trac* (*Tcrd*^{CreER}*Trac*^{M}) in combination with the HYα-KI allele (Buch et al., 2002), we observed restricted secondary recombinations over time on alleles with a knockin primary rearrangement. We also used primary thymocytes of a novel mouse, D_δ and J_α deleted (DJD), to determine that *Tcrd* recombinations indeed diversify the *Tcra* repertoire. Moreover, early recombinations on DJD alleles are of reduced complexity, with the majority of V_α utilization in the proximal portion of the array. Further, using two lines of mice lacking either *Trav15-1-dv6-1* or *Trav15d-1-dv6d-1*, we find that *Trav15-dv6* segments facilitate central and distal V_α utilization, respectively, in early *Tcra* recombinations. Additionally, we find that *Trav15-dv6* family members or their flanking elements impact *Tcra* repertoire diversification by an additional mechanism that is separate from their roles as *Tcrd* recombination substrates

4.2 A fixed primary recombination constrains Tcra repertoire over time

The pre-selection DP thymocyte repertoire is remarkably diverse, with recombinations observed between most V_α and J_α segments across the locus (**Figure 4A**) (Carico et al., 2017). The majority of recombinations are observed along two diagonals (**Figure 4B**). The 'major diagonal' involves a broad swath of recombinations. It originates

with proximal V_α-to-proximal J_α primary recombinations in the lower left corner of the heatmap, but these recombinations are not tightly restricted to extremely J_α-proximal V_α segments. Indeed, they involve segments spanning the proximal third of the V_α array. As rounds of secondary recombination progress through ever more distal V_α and J_α segments, central portion of the heatmap is filled with major diagonal central V_α-to-central J_α rearrangements. The major diagonal ends in the upper right corner of the heatmap with distal V_α-to-distal J_α recombinations. The 'minor diagonal' occupies the upper left corner of the heatmap, commencing with primary recombinations between segments in the distal third of the V_α array and proximal J_α segments. As rearrangements along the minor diagonal proceed, they utilize more distal V_α segments and J_α segments, ending with the most distal V_α-to-central J_α recombinations. Although secondary *Tcra* recombinations are constrained to recombinations between the most proximal available V_α and J_α segments, the utilization of V_α segments across the array in recombinations with proximal J_α segments suggests that diverse primary recombinations may lend immense diversity to the *Tcra* repertoire.

Previous analysis of the temporal kinetics of *Tcra* recombination in pre-selection DP thymocytes revealed diverse early V_α utilization, though use of J_α segments was more constrained (Carico et al., 2017). J_α segments spanning the V_α-proximal half of the array, from *Traj58* to *Traj31* were used very frequently in early recombinations. Over 27% of early recombinations involved the three most V_α-proximal J_α segments, *Traj58*, *Traj57*,

and *Traj56*. J_α utilization progressed to include more V_α-distal J_α segments at later times. V_α utilization in early recombinations was more heterogenous. V_α segments in the proximal, central, and distal portions of the array were used in early rearrangements. Over time, more distal V-to-J_α rearrangements were observed (Carico et al., 2017). Conversely, recombinations of alleles with a fixed primary recombination were mostly constrained to a narrowed major diagonal of recombinations involving proximal V_α-to-proximal J_α, central V_α-to-central J_α, and distal V_α-to-distal J_α recombinations (Carico et al., 2017). These findings suggested that variability in primary recombinations contributed considerably to the diversity of ongoing secondary rearrangements.

To examine secondary recombinations over time absent the variability of primary *Tcra* recombinations, we employ a mouse with a knockin primary *Tcra* rearrangement as a fixed starting point, HYα-KI (Buch et al., 2002). The other *Tcra-Tcrd* allele contains a tamoxifen-inducible $Tcrd^{CreER}$. When used along with the $Rosa26^{ZsG}$ reporter, tamoxifen injection causes ZsGreen expression in cells with transcriptionally active *Trdc*, including DN thymocytes; most DN thymocytes, failing to diverge to the γδT cell fate, progress to rearrange *Tcra* at the DP stage (Madisen et al., 2010; Zhang et al., 2015). We employed a 2 guide-mediated CRISPR/Cas9 (Cong et al., 2013; Singh et al., 2014) strategy to mutate the 5' portion of *Trac* on the $Trdc^{CreER}$ allele in order to stop annealing of *Trac* reverse primers and analyze transcripts from only the HYα-KI allele. The resulting experimental mouse (**Figure 5**), with one HYα-KI allele, one $Tcrd^{CreER}$ allele with a mutation in 5' *Trac* ($Tcrd^{CreER}Trac^{M}$), and one $Rosa26^{ZsG}$ reporter allele is referred to as 'HYα-$KI^{M/+}$

Tcrd$^{CreER/+}$ *Trac*$^{M/+}$ *Rosa26*$^{ZsG/+}$.' This model allowed time course-based analysis of *Tcra* recombination after tamoxifen injection (Carico et al., 2017).

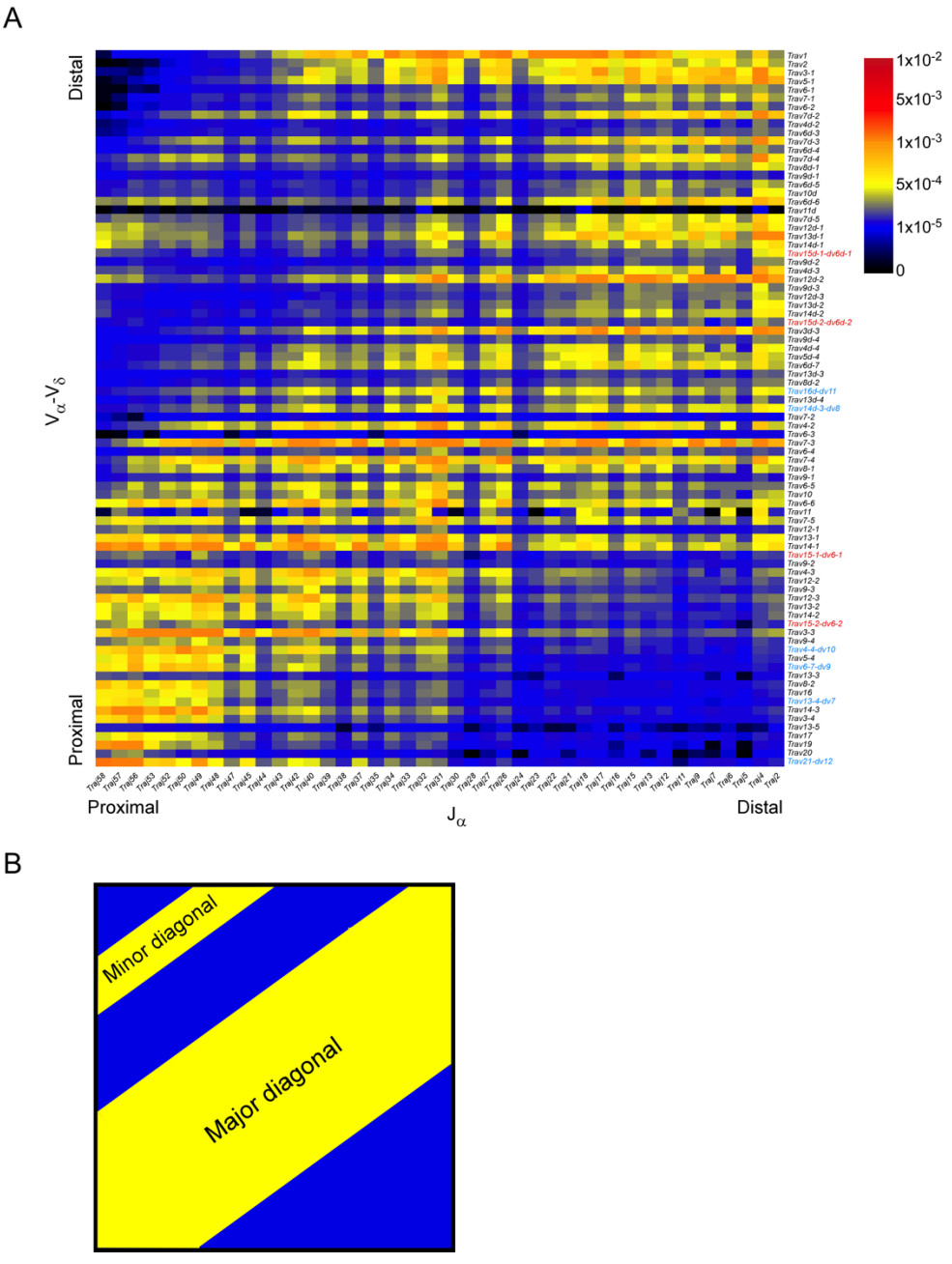

Figure 4: *Tcra* repertoire in strain 129 pre-selection thymocytes.

(A) Average (n=2) frequencies of V-J_α rearrangements in WT strain 129 $CD4^+CD8^+CD3_\varepsilon^{lo}$ thymocytes (identical to Fig 8B, left). Gene segment names are indicated on the right and lower margins. Red and blue lettering identifies *Trav15-dv6* family V_δ segments and other V_δ segments, respectively. (B) Diagram indicating locations of major and minor diagonals in yellow, corresponds to the heatmap in A.

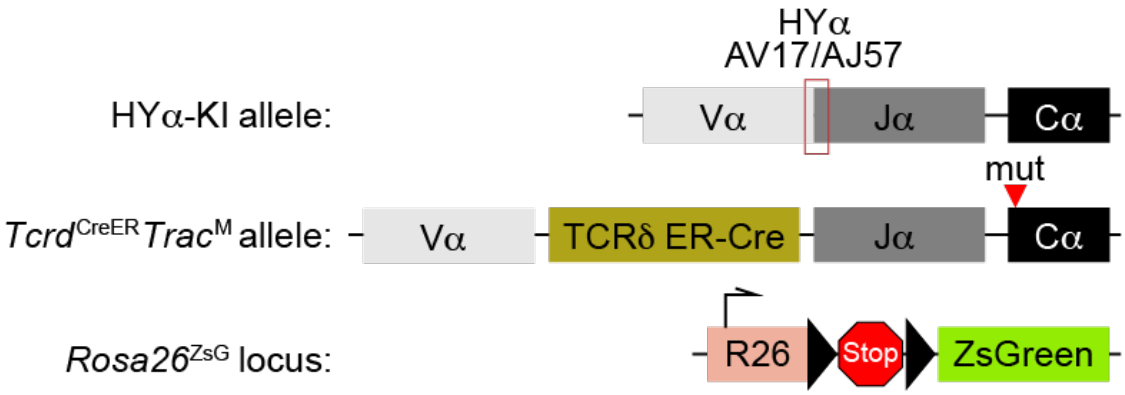

Figure 5: Diagram of HYα-KI$^{+/-}$*Tcrd*$^{CreER/+}$*Trac*$^{M/+}$ *Rosa26*$^{ZsG/+}$ mouse.

The 129 mouse strain *Tcra-Tcrd* locus, on chromosome 14, contains the HYα-KI of *Trav17-Traj57* on one allele and an ER-driven Cre within *Trdc* on the other allele. *Rosa26*, on chromosome 6, has a *loxP*-flanked Stop cassette allowing ZsGreen expression in the presence of activated Cre. CreER is induced with Tamoxifen.

Here, we have sequenced and analyzed the *Tcra* repertoire on the HYα-KI allele in pre-selection ZsGreen^{+} DP thymocytes at 8hrs, 18hrs, 30hrs, 48hrs, and 72hrs using 5'RACE and *Tcra* sequencing (**Figure 6A**). Unique clones were analyzed to determine the frequencies of V-to-J$_{\alpha}$ rearrangements. We expected that the most proximal V$_{\alpha}$ and J$_{\alpha}$ segments would be used in successive rearrangement events (Hawwari and Krangel, 2007; Carico et al., 2017). In the absence of factors that may serve to diversify V$_{\alpha}$ use, such as *Tcrd* recombination and chromatin features of the central region of the *Tcra-Tcrd* locus, we observe mostly proximal to distal use of V$_{\alpha}$ and J$_{\alpha}$ segments over time (**Figure 6B**). As expected, J$_{\alpha}$ use was biased to the proximal segments at early timepoints. The three most proximal J$_{\alpha}$ segments were involved in 33.10% of recombinations at 8hrs. By 30hrs, these three most proximal segments were involved in only 17.27% of

recombinations, and at 72hrs recombinations to the three most proximal J_α segments were represented in only 6.84% of rearrangements. V_α use was also restricted, with 53.65% of V_α utilization at 8hrs involving the most proximal 15 V_α segments, those positioned more J_α-proximally than *Trav15-2-dv6-2*. These were used in only 14.86% of recombinations at 72hrs. Central V_α use involving *Trav15-2-dv6-2* to *Trav9-2* accounted for 39.67% of rearrangements at 8hrs and peaked at 69.45% at 48hrs. Distal V_α use of segments *Trav15d-1-dv6-1* and beyond at 8hrs made up only 6.77% of recombinations. Distal V_α rearrangements reached a peak of 17.98% of recombinations at 72hrs. In the HYα-KI model of a fixed primary recombination, early recombinations were substantially less diverse than in our previously published data on early *Tcra* recombinations of WT alleles; in particular, V_α utilization is constrained in the absence of diverse primary recombinations (Carico et al., 2017). We thus conclude that the diversity of primary *Tcra* recombinations dictates the potential diversity of the fully developed *Tcra* repertoire.

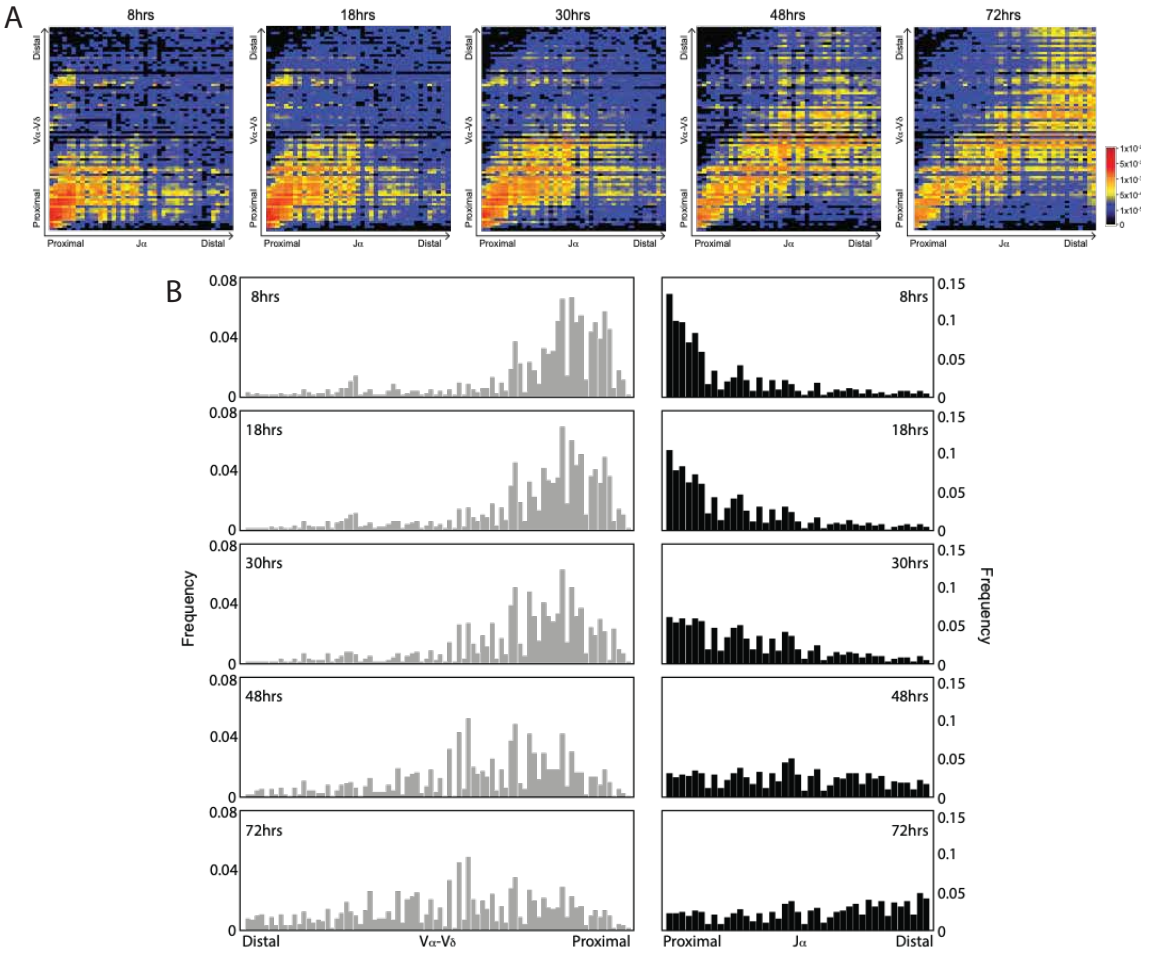

Figure 6: HYα-KI limits *Tcra* combinatorial diversity over time.

(A) V-J$_\alpha$ rearrangement frequencies on the HYα-KI allele in CD4$^+$CD8$^+$CD3$_\varepsilon^{lo}$ZsGreen$^+$ mouse thymocytes at the indicated times post-Tamoxifen injection. (B) Frequencies of V$_\alpha$-V$_\delta$ (left) or J$_\alpha$ (right) use at the indicated times post-Tamoxifen injection. One library per timepoint; libraries consist of material from pre-selection DP thymocytes from 1-3 mice.

4.3 Tcrd recombinations diversify the Tcra repertoire

One factor that may diversify primary V_α utilization is prior *Tcrd* recombination involving V_δ segments dispersed among the V_α segments. To directly assess the impact of *Tcrd* recombination on the *Tcra* repertoire, we sought a mouse strain incapable of *Tcrd* recombination. In this regard, the C_δ-knockout (KO) mouse undergoes normal *Tcrd* rearrangement (Itohara et al., 1993), and the E_δ-KO mouse has only a partial defect in *Tcrd* recombination (Monroe et al., 1999c). Therefore, we generated a novel line of mice lacking a 24 kb region containing the D_δ and J_δ segments (**Figures 7A and 8A**). The resulting DJD allele ablates V-to-DJ_δ rearrangement and γδT cells, with otherwise normal thymocyte development (**Figure 7B-D**).

Pre-selection DP thymocytes from DJD mice and their wild-type (WT) littermates were subjected to 5′ RACE and *Tcra* sequencing. Unique clones were analyzed to determine the frequencies of V-to-J_α rearrangements. As previously reported (Carico et al., 2017), in WT thymocytes we found the *Tcra* repertoire to be diverse (**Figures 4A and 8B**). Proximal J_α segments frequently rearranged with proximal, central, and distal V_α segments. Secondary rearrangements occurred mostly along two distinct diagonals (**Figure 4B**). The major diagonal started with proximal and central V-to-proximal J_α primary rearrangements, followed by secondary rearrangements through the central and distal V_α and J_α segments. The minor diagonal arose from distal V-to-proximal J_α primary rearrangements. In DJD thymocytes, absent *Tcrd* recombination, we observed restricted V_α usage; proximal J_α segments recombined almost exclusively with the most

proximal V_α segments (**Figure 8B**). This reduction in diversity of V_α utilization in early rearrangements with proximal J_α segments depressed the diversity of secondary rearrangements, with a notable loss in the use of distal V_α segments, even in recombinations with central and distal J_α segments. The major diagonal was constrained and shifted toward more proximal V_α segments along its entire length; the minor diagonal was largely depleted of unique clones.

Previous work described the progression of secondary rearrangements emanating from a single primary recombination as proceeding approximately 1.33 V_α segments per J_α (Carico et al., 2017). According to our hypothesis, DJD thymocytes should be depleted of V-J_α combinations resulting from primary rearrangements distal to the most proximal *Trav15-dv6* family member, *Trav15-2-dv6-2*, as well as any secondary rearrangements that occur as a consequence of those primary rearrangements. To assess this, we overlaid a step-function diagonal ($4V_\alpha$:$3J_\alpha$, corresponding to 1.33 V_α per J_α) over the DJD-WT difference map, with its origin at *Trav15-2-dv6-2* (**Figure 8C**). The step-function diagonal predicts almost exactly the region of depletion of the combinatorial *Tcra* repertoire in DJD thymocytes. In WT thymocytes, the region above the diagonal accounts for 43% of the repertoire; in DJD thymocytes, it accounts for 22% (**Figure 8B and 8C**; $P<0.0001$ by chi-square test with Yates's correction). However, this is a substantial underestimate of the impact on some distal V_α rearrangements. Prior work demonstrated that the standard deviation of the V_α distribution used with any J_α gradually increases from proximal to distal across the J_α array (Carico et al., 2017). Because this increase is

not accounted for by the step-function diagonal, we would expect representation of the most distal V-distal J_α combinations to persist, although at reduced frequencies, in the region above the diagonal in DJD thymocytes (compare **Figure 8B**, right panel, to **Figure 8C**). This effect should not impact quantification along the minor diagonal; rearrangements in this region account for 7.7% of the repertoire in WT, but only 2% in DJD thymocytes, a reduction of 74%. We conclude that *Tcrd* rearrangement expands the combinatorial diversity of the *Tcra* repertoire by diversifying V_α use.

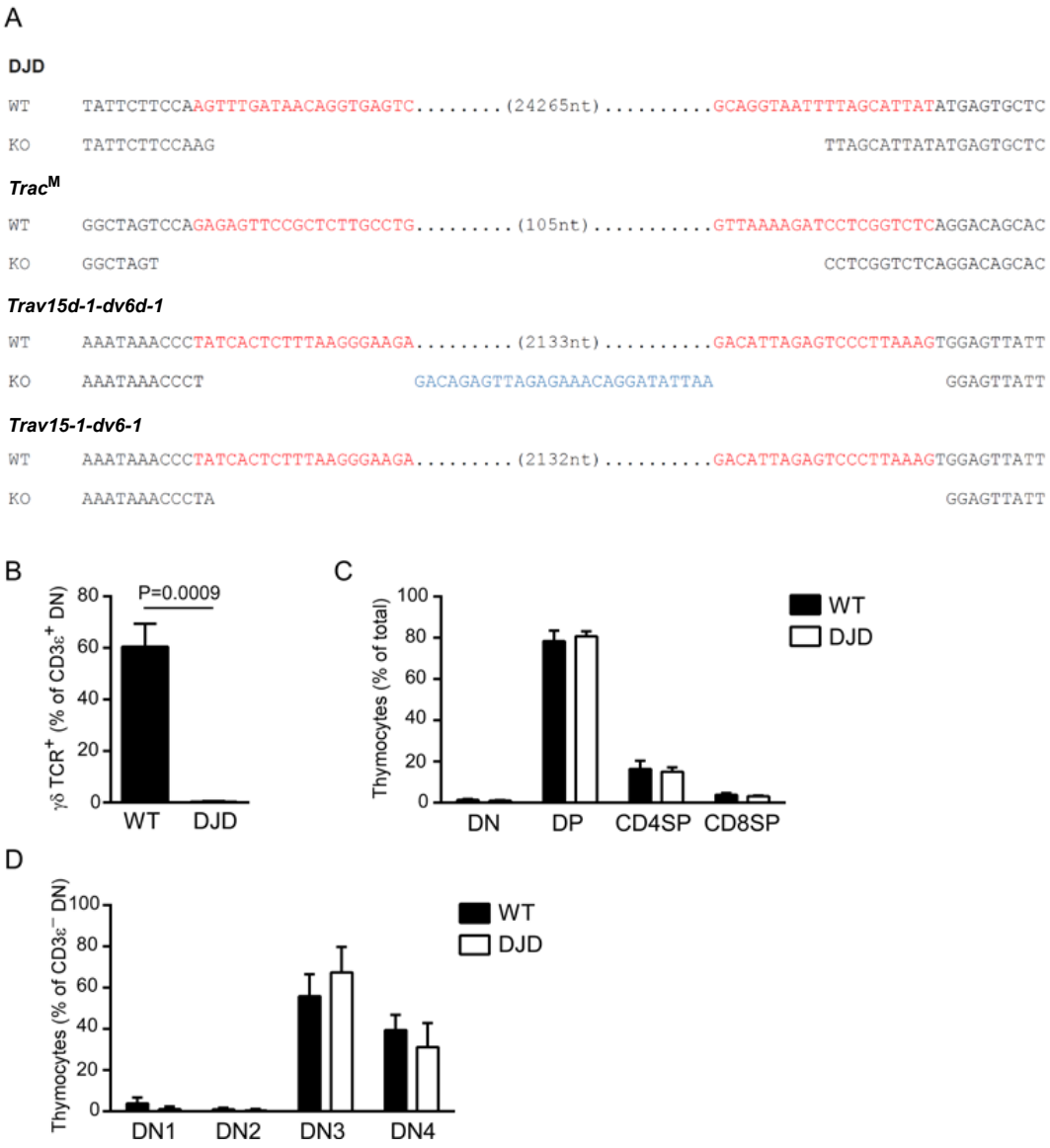

Figure 7: Characterization of mutant mouse strains.

(A) DNA sequence of mutant alleles. Red, guide sequence for CRISPR/Cas9 targeting. Blue, nucleotides inserted in KO alleles. Gaps, nucleotides deleted in KO alleles. (B-D) Flow cytometry analysis of total thymocytes from WT and DJD littermate mice. All data are presented as mean and SD of 4 WT and 5 DJD mice analyzed in two independent experiments. Statistical significance was evaluated with unpaired t- test with Welch's correction. Cells were pre-gated as 7AAD$^-$CD11b$^-$CD11c$^-$Ter119$^-$B220$^-$Gr-1$^-$F4/80$^-$.

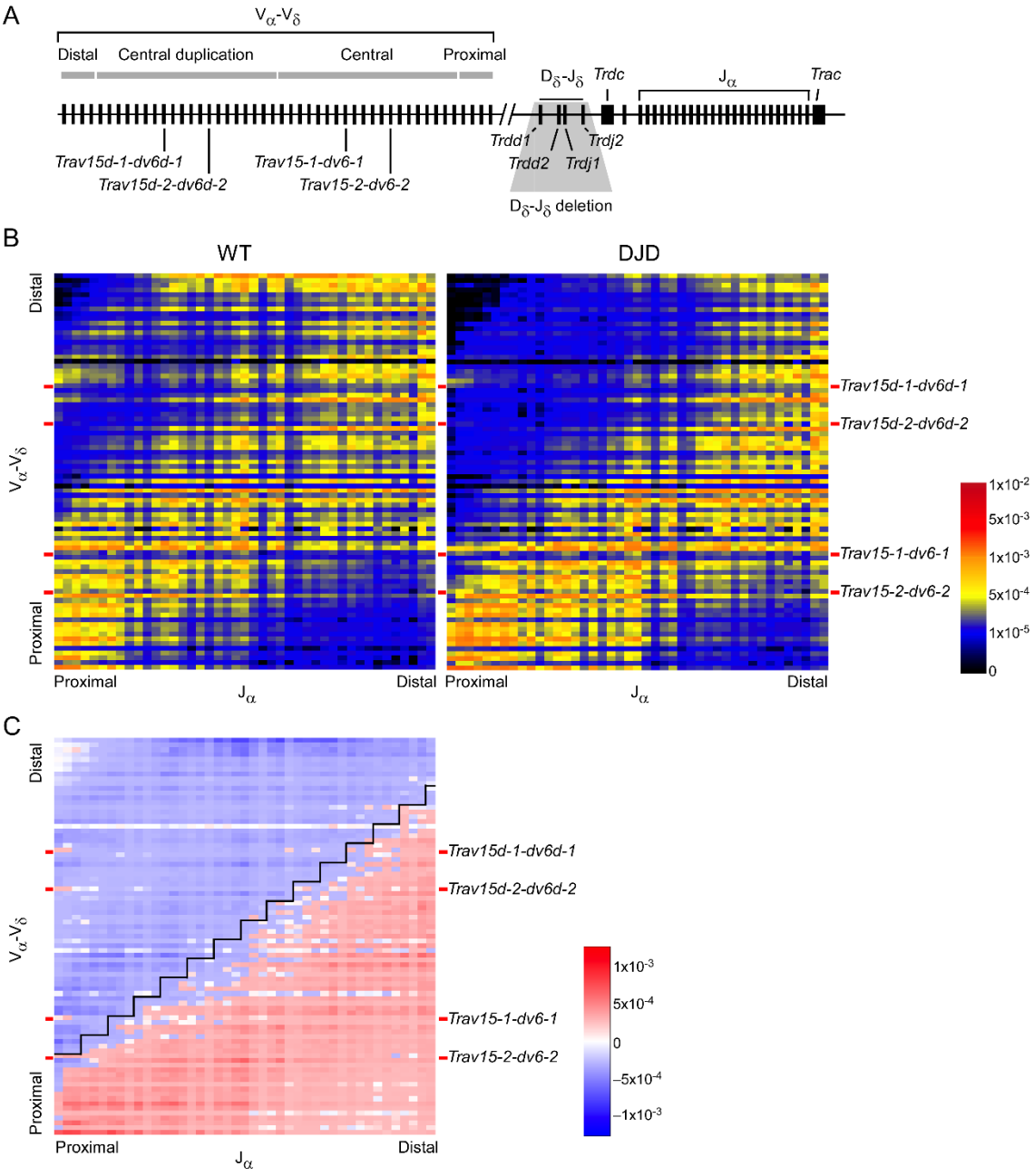

Figure 8: Tcrd recombination diversifies Tcra repertoire.

(A) Schematic of the *Tcra-Tcrd* locus with gene segments depicted. Shaded region corresponds to deletion on DJD alleles. (B) Average frequencies of V-J$_\alpha$ rearrangements in CD4$^+$CD8$^+$CD3$_\varepsilon^{lo}$ thymocytes from WT (left) and DJD (right) mice. Two DJD mice and two WT littermates (mixed 129 and C57BL/6 background) were

analyzed in two independent experiments by HTS of 5′ RACE-amplified *Tcra* transcripts. Red bars on left and right edges of heatmaps indicate locations of *Trav15-dv6* family V segments. (C) Map depicting difference between average V-J_α rearrangements in DJD and WT mice. Blue represents rearrangements under-represented in DJD; red represents rearrangements more common in DJD. Black line indicates step-function diagonal with a slope of $4V_\alpha$:$3J_\alpha$ (or 1.33 V_α per J_α, the previously established relationship for the progression of secondary rearrangements across the V and J arrays (Carico et al., 2017)). With the origin of the diagonal just distal to *Trav15-2-dv6-2*, the line separates all primary and secondary *Tcra* rearrangements hypothesized to arise as a result of *Trav15-dv6* and other distal *Tcrd* rearrangements from those expected to occur on alleles that lack such rearrangements. Normalized Shannon's entropy values were 0.953 and 0.950 for replicate WT samples and 0.936 and 0.935 for replicate DJD samples.

4.4 Tcrd recombinations diversify early Tcra rearrangements

Primary *Tcra* rearrangements are poorly represented in the steady-state *Tcra* repertoire. To more robustly investigate the impact of *Tcrd* rearrangement on primary *Tcra* rearrangement, we directly visualized *Tcra* rearrangements in the earliest DP thymocytes. We previously analyzed these cells in mice carrying a tamoxifen-inducible $Tcrd^{CreER}$ allele together with a $Rosa26^{ZsG}$ reporter allele (Carico et al., 2017). In these mice, tamoxifen injection causes ZsGreen expression in cells with transcriptionally active *Trdc*, including DN thymocytes, most of which, failing to diverge to the γδT cell fate, progress to rearrange *Tcra* at the DP stage (Madisen et al., 2010; Zhang et al., 2015). In mice containing $Rosa26^{ZsG}$ and a *Trac*-mutated $Tcrd^{CreER}$ ($Tcrd^{CreER}$ $Trac^{M}$) allele (**Figure 7A**), in combination with either a WT or DJD allele, we analyzed *Tcra* recombination in $ZsGreen^{+}$ DP thymocytes at 12 hours after tamoxifen injection. These primary and early secondary recombinations, in cells that comprise less than 20% of the DP population (Carico et al., 2017), are typically overshadowed by secondary recombinations in the

steady-state repertoire. As previously reported (Carico et al., 2017), on WT alleles we found these recombinations to be restricted to the first half of the J_α array (**Figure 9A**, left panel), with the majority of rearrangements focused on the most proximal J_α segments, *Traj58-Traj48*. V_α use involved the first half of the V_α array, along with a cluster of relatively distal V_α segments. Save for an unexplained reduction in use of *Traj58*, DJD alleles displayed minimal change in overall J_α utilization (**Figure 9A and B**). However, V_α use on DJD alleles was restricted compared to WT thymocytes, with overall utilization of segments distal to *Trav15-2-dv6-2*, the most proximal *Trav15-dv6* segment, reduced from 59.4% to 32.2% on DJD alleles ($P<0.0001$ by chi-square test with Yates's correction). We conclude that *Tcrd* recombination diversifies primary V_α use, leading to a more robust repertoire of secondary rearrangements and ultimately imparting combinatorial diversity upon the *Tcra* repertoire.

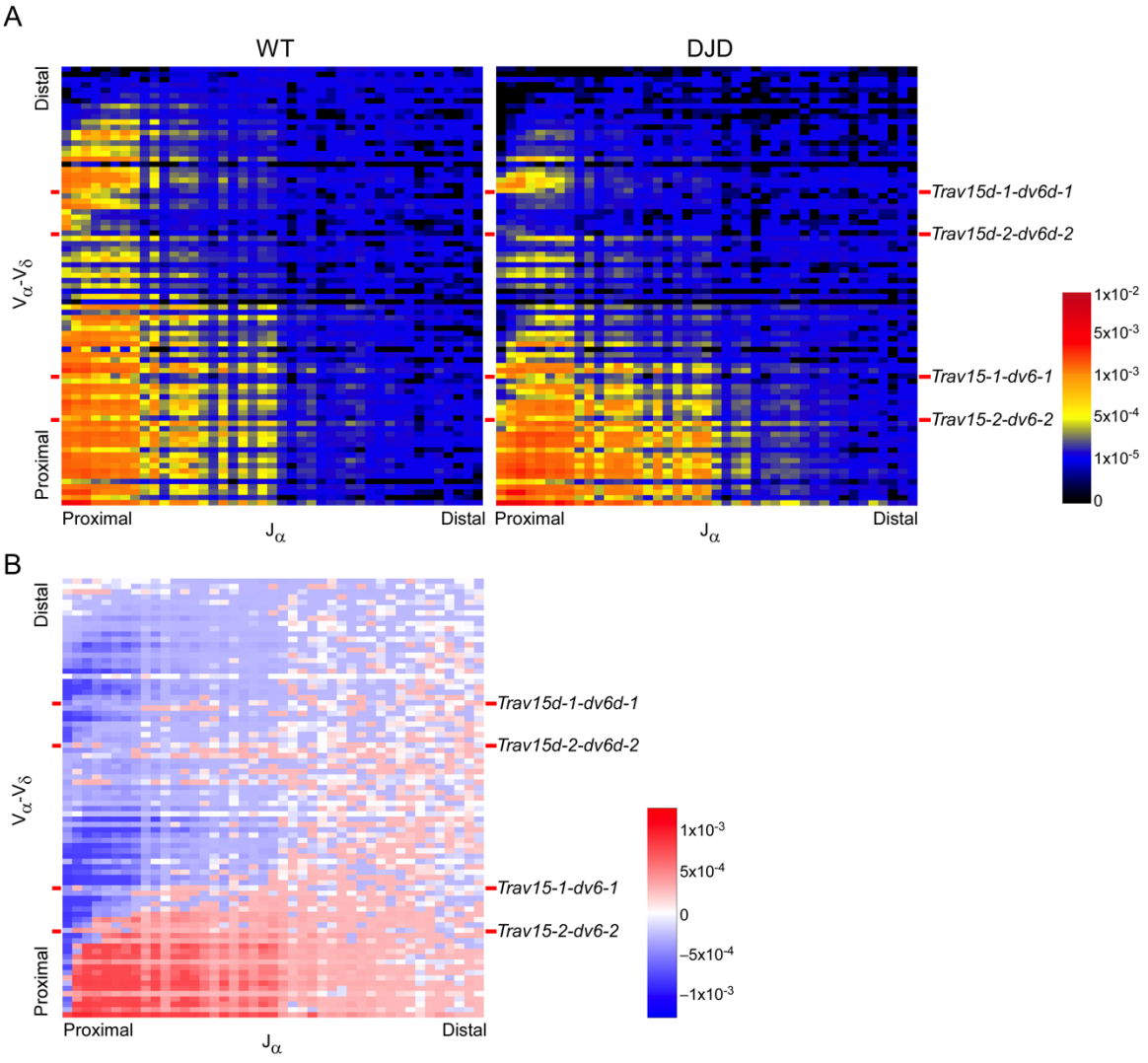

Figure 9: Early *Tcra* rearrangements are diversified by *Tcrd* recombination.

(A) Average frequencies of V-J$_\alpha$ rearrangements on WT or DJD alleles in CD4$^+$CD8$^+$CD3$_\varepsilon^{lo}$ZsGreen$^+$ thymocytes of mice containing *Rosa26*^{ZsG} and a *Tcrd*^{CreER} *Trac*^{M} allele paired with either a WT or a DJD allele at 12 hrs post-tamoxifen injection. Two DJD and two WT littermates (mixed 129 and C57BL/6 background) were analyzed in two independent experiments. (B) Map depicting difference between average V-J$_\alpha$ rearrangements in DJD and WT mice. Normalized Shannon's entropy values were 0.863 and 0.864 for replicate WT samples and 0.835 and 0.838 for replicate DJD samples.

4.5 Tcrd recombinations using Trav15d-1-dv6d-1 and Trav15-1-dv6-1 diversify Tcra repertoire

Trav15-dv6 segments comprise the most frequent of the distal and central V_δ contributions to the *Tcrd* repertoire; the only other similarly located V_δ gene segments, *Trav14d-3-dv8* and *Trav16d-dv11* (Fig S2), are rarely used (Chen et al., 2015; Zhao et al., 2016). We observed that in primary recombinations of the WT *Tcra* locus, most central and distal V_α use occurs immediately upstream of *Trav15-dv6* family members, particularly *Trav15-1-dv6-1* and *Trav15d-1-dv6d-1* (left panels of **Figures 8B, 9A, and 10A**). These early central and distal recombinations are dramatically reduced in DJD mice, (right panels of **Figures 8B, and 9A**) suggesting that *Trav15-dv6* recombinations in DN cells may contribute to diversifying the *Tcra* repertoire. To analyze the impact of *Trav15-dv6* rearrangements on *Tcra* recombination, we analyzed two lines of mice, one with a deletion of ~2kb spanning *Trav15d-1-dv6d-1* and a second with a similar deletion of *Trav15-1-dv6-1*, for *Tcra* clonal repertoire. We hypothesized that the loss of either would result in a substantial reduction in recombination events emanating in a diagonal pattern from the most proximal upstream V_α and moving distally through the remaining V_α segments.

Primary and secondary *Tcra* recombinations in the minor diagonal, predicted to occur as a consequence of *Tcrd* rearrangements involving *Trav15d-1-dv6d-1*, represented 3.1% of the *Tcra* repertoire in WT DP thymocytes, but only 0.8% of the repertoire in cells lacking *Trav15d-1-dv6d-1* (**Figure 10A**, and **10B**, region 1; $P<0.0001$ by chi-square test with Yates's correction). Major diagonal primary and secondary *Tcra*

recombinations, predicted to occur as a consequence of *Tcrd* rearrangements to *Trav15-1-dv6-1*, accounted for 21.2% of the repertoire in WT thymocytes but 16.6% of the repertoire in cells lacking *Trav15-1-dv6-1* (**Figure 11A** and **11B**, region 2; P<0.0001 by chi-square test with Yates's correction). In contrast, the representation of minor diagonal *Tcra* recombinations predicted to occur as a consequence of rearrangements involving more distal *Trav15-dv6* family members increased slightly in *Trav15-1-dv6-1*-deleted DP cells (**Figure 11A**, and **11B**, region 1). In both lines of mice, rearrangements involving V_α segments downstream of the deleted *Trav15-dv6* family member increased proportionally (**Figures 10C**, region 3, and **11C**, region 4). Because reduced minor diagonal *Tcra* rearrangements were observed in both DJD mice and *Trav15d-1-dv6d-1*-deleted mice and reduced major diagonal *Tcra* rearrangements were observed in both DJD mice and *Trav15-1-dv6-1*–deleted mice, we conclude that distal *Tcrd* rearrangements involving *Trav15-dv6* family members diversify the *Tcra* repertoire.

In thymocytes lacking either *Trav15d-1-dv6d-1* or *Trav15-1-dv6-1*, we observed changes in the *Tcra* repertoire that were not predicted to arise due to loss of *Tcrd* recombination to either segment. In the absence of *Trav15d-1-dv6d-1*, distal V_α-to-distal J_α rearrangements outside the minor diagonal were reduced to 13.1% of the repertoire from 19.1% in WT; these secondary rearrangements lie on the major diagonal and are predicted to arise from primary rearrangements involving central V_α segments (**Figure 10A** and **10B**, region 2). Similarly, in *Trav15- 1-dv6-1*-deleted mice, major diagonal secondary rearrangements involving central and distal V_α segments were unexpectedly reduced to 37.4% of the repertoire from 43.1% in WT (**Figure 11A**, and **11B**, region 3);

these secondary rearrangements are predicted to arise from primary rearrangements involving proximal V_α segments. These changes are not observed in the DJD repertoire (**Figure 8**). This implies that *Trav15-dv6* family members impact *Tcra* repertoire diversification by an additional mechanism separate from their roles as *Tcrd* recombination substrates. We propose that the *Trav15-1-dv6-1* and *Trav15d-1-dv6d-1* deletions have these additional effects on *Tcra* repertoire diversity because they impair the propagation of secondary *Tcra* rearrangements to V_α segments beyond the deleted region.

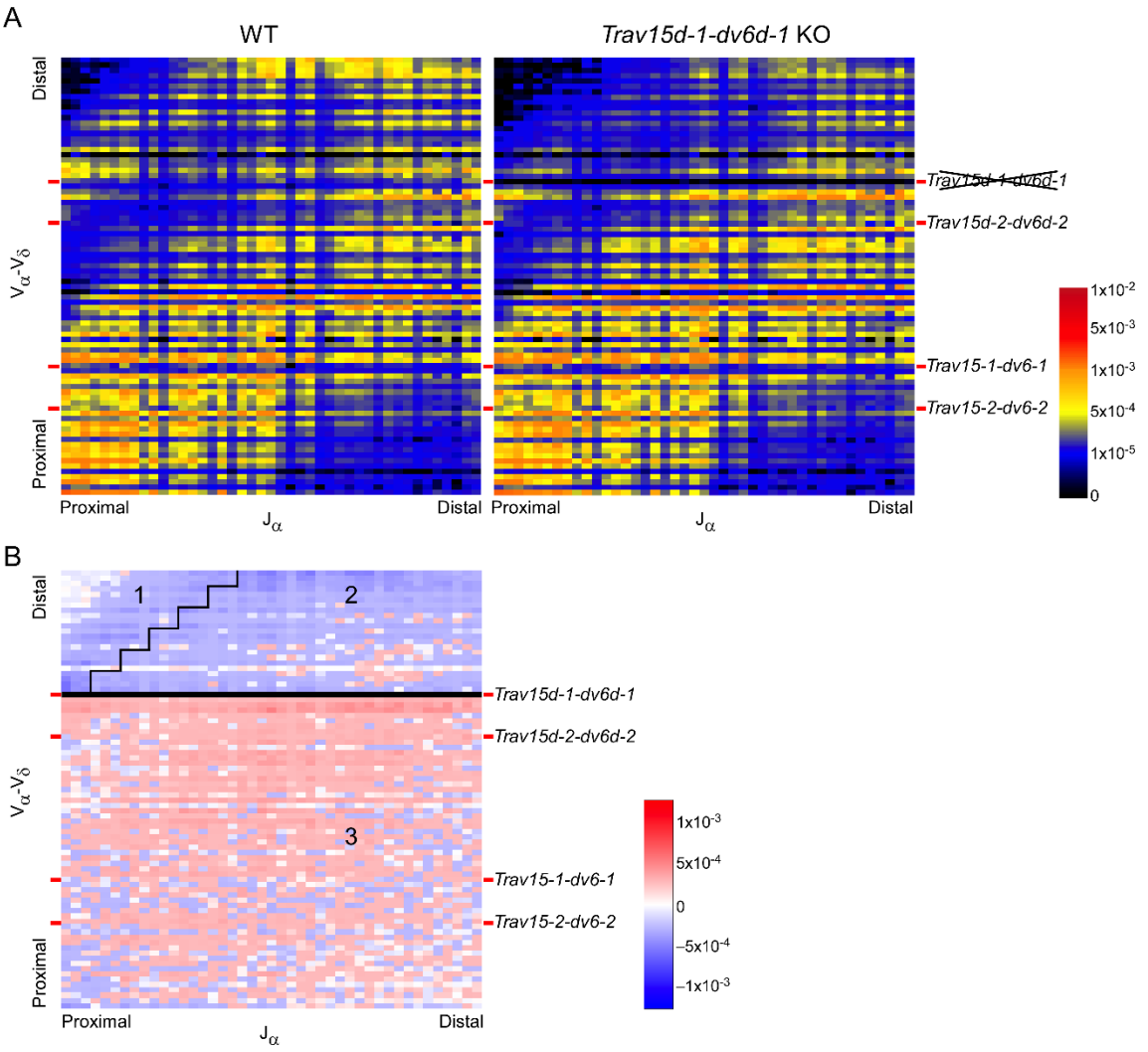

Figure 10: Reduced distal Vα use in *Trav15d-1-dv6d-1* KO mice.

(A) Average frequencies of V-Jα rearrangements in $CD4^+CD8^+CD3\varepsilon^{lo}$ thymocytes of mice carrying WT (left) or *Trav15d-1-dv6d-1* KO (right) *Tcra-Tcrd* alleles. Although irrelevant for this analysis, WT and KO *Tcra-Tcrd* alleles also carry $Tcrd^{CreER}$ (see Methods). Two littermates per genotype (C57BL/6 background) were analyzed in two independent experiments. (B) Map depicting difference between average V-J□ rearrangements on *Trav15d-1-dv6d-1* KO and WT alleles. Region 1, primary and secondary *Tcra* rearrangements predicted to depend on use of *Trav15d-1-dv6d-1* in *Tcrd* recombination. Region 2, secondary rearrangements upstream of *Trav15d-1-dv6d-1* expected to arise from primary and secondary recombinations downstream of *Trav15d-1-dv6d-1*. Region 3, primary and secondary rearrangements downstream of *Trav15d-1-*

dv6d-1. Normalized Shannon's entropy values were 0.944 and 0.947 for replicate WT samples and 0.938 and 0.942 for replicate *Trav15d-1-dv6d-1* samples.

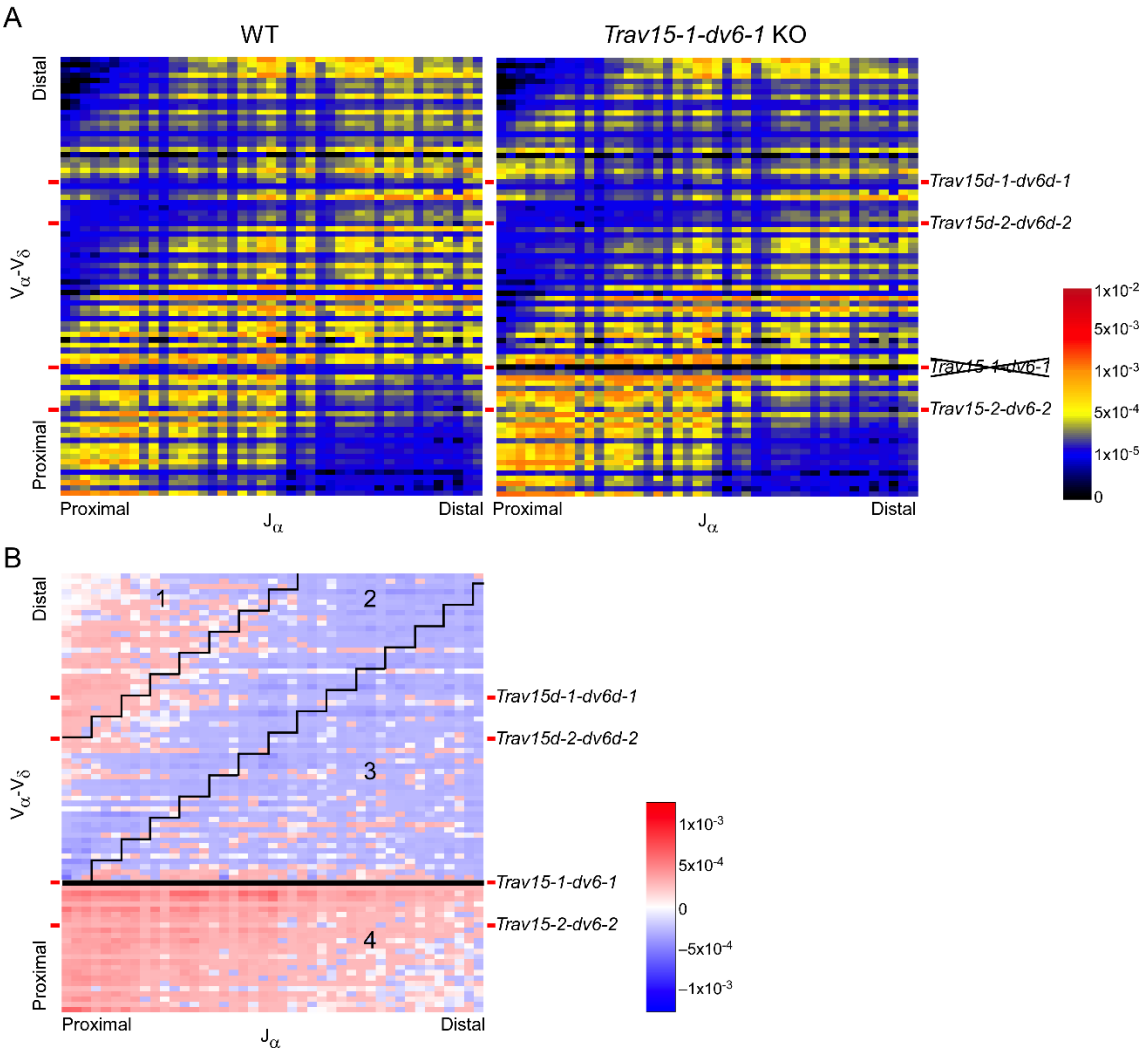

Figure 11: Reduced central Vα use in *Trav15-1-dv6-1* KO mice.

(A) Average frequencies of V-Jα rearrangements in $CD4^+CD8^+CD3\varepsilon^{lo}$ thymocytes of mice carrying WT (left) or *Trav15-1-dv6-1* KO (right) *Tcra-Tcrd* alleles. *Tcra-Tcrd* alleles. Although irrelevant for this analysis, WT and KO *Tcra-Tcrd* alleles also carry $Tcrd^{CreER}$ (see Methods). Two littermates per genotype (C57BL/6 background) were analyzed in two independent experiments. (B) Map depicting difference between average V-J□ rearrangements on *Trav15-1-dv6-1* KO and WT alleles. Region 1, primary and secondary *Tcra* rearrangements predicted to depend on *Trav15d-1-dv6d-1* and *Trav15d-*

2-dv6d-2 Tcrd rearrangements. Region 2, primary and secondary *Tcra* rearrangements predicted to depend on *Trav15-1-dv6-1 Tcrd* rearrangements. Region 3, secondary *Tcra* rearrangements upstream of *Trav15-1-dv6-1,* expected to arise from primary and secondary *Tcra* recombinations downstream of *Trav15-1-dv6-1*. Region 4, primary and secondary *Tcra* rearrangements downstream of *Trav15-1-dv6-1*. Normalized Shannon's entropy values were 0.950 and 0.947 for replicate WT samples and 0.947 and 0.939 for replicate *Trav15-1-dv6-1* samples.

The *Trav15d-1-dv6d-1* and *Trav15-1-dv6-1* deletions encompass regions spanning ~500bp upstream of the transcription start site to ~1kb downstream of the gene body. These deleted regions are highly conserved between *Trav15-1-dv6* family members and include the promoter, the RSS, and downstream E-box sites that are highly accessible and by ChIP-seq are occupied by E2A at levels surpassed only by *Tcra* enhancer E-box sites (**Figure 12**) (Heng et al., 2008; Roy et al., 2018; Yoshida et al., 2019). Loss of the *Trav15-dv6* RSS could inhibit more distal secondary rearrangements if the propagation of sequential rearrangements required closely spaced RSSs. However, *Trav15-dv6* family members were not found to be frequently recombined in the WT *Tcra* repertoire, suggesting that *Trav15-dv6* RSSs are often skipped during secondary rearrangements (**Figure 8B**). Moreover, the distance between the nearest functional V_α segments flanking *Trav15-dv6* is reduced from ~37kb to ~35kb in these deletions. We considered that *Trav15-dv6* deletion might influence secondary rearrangements by disrupting the CTCF-mediated chromatin loop organization. However, binding sites for CTCF were not disrupted on the *Trav15d-1-dv6d-1*- or *Trav15-1-dv6-1*-deleted alleles (Shih et al., 2012), and the deletions do not obviously create *de novo* CTCF-binding sites (Martin et al., 2011).

It remains possible that *Trav15-dv6* segments influence secondary *Tcra* recombinations via effects on chromatin accessibility mediated by flanking *cis*-acting elements. The *Trav15-dv6* promoter drives accessibility, transcription, and recombination of *Trav15-dv6* in DN thymocytes (Naik et al., 2015) but displays only modest accessibility and is not expected to affect neighboring V segments (Heng et al., 2008; Yoshida et al., 2019). Perhaps a better candidate is the highly accessible downstream E2A-bound element (**Figure 12**). E2A is a known regulator of *Tcrd* recombination and γδT cell fate (Bain et al., 1999) and may influence *Trav15-dv6* rearrangement in DN thymocytes. As the E2A-bound region remains highly accessible in DP thymocytes, it could have extended effects on accessibility and RAG binding at RCs that form downstream of *Trav15-dv6* family members during secondary rearrangement, or on V_{α} substrates upstream of *Trav15-dv6* family members. The *cis* region may alternatively serve as a mediator of locus structure to facilitate recombination; such features have been observed at other AgR loci (Barajas-Mora et al., 2019). Further exploration of this phenomenon is warranted to better understand these perturbations of secondary *Tcra* rearrangements on *Trav15d-1-dv6d-1-* and *Trav15-1-dv6-1*-deleted alleles.

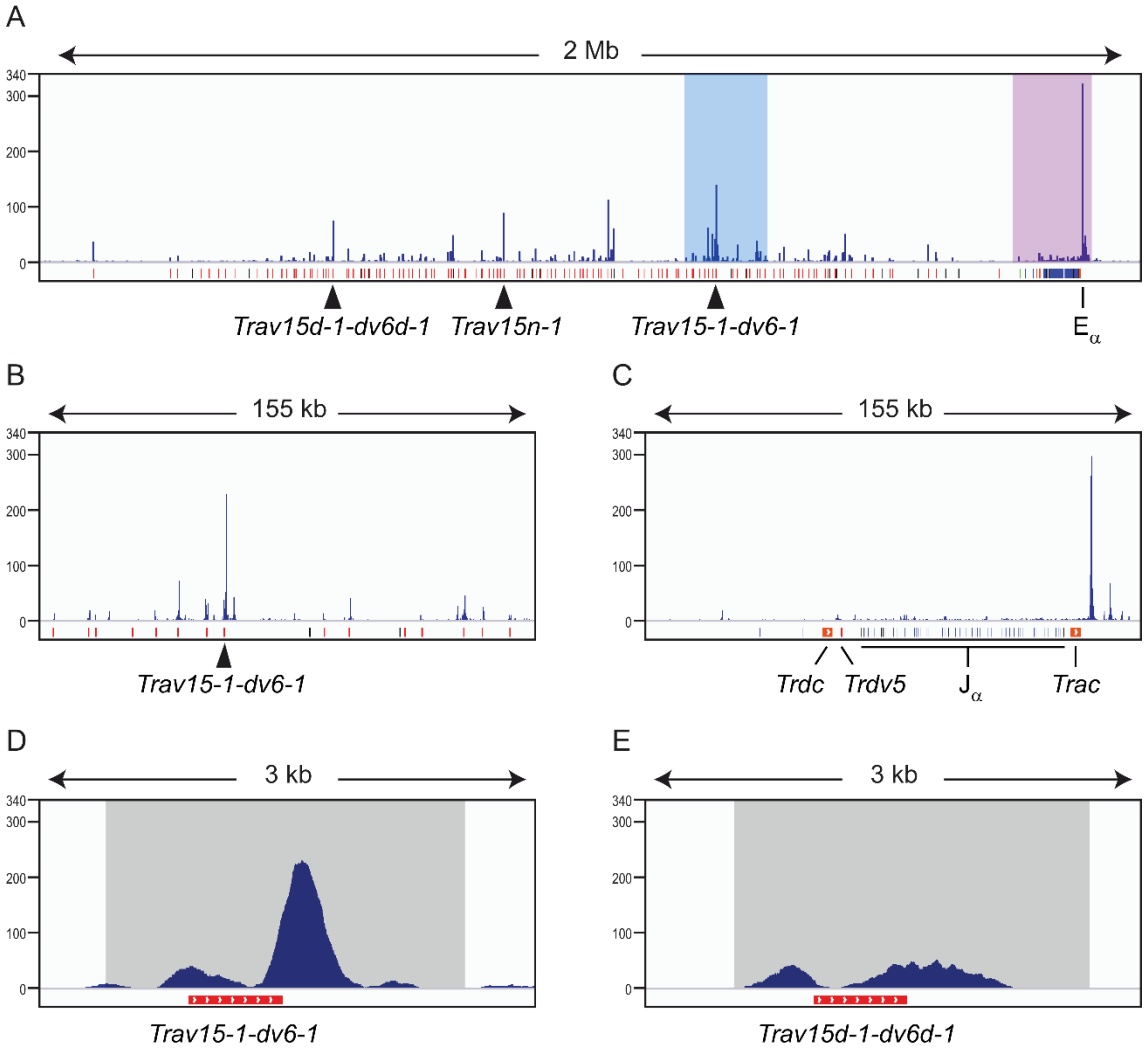

Figure 12: E2A binding at the C57BL/6 *Tcra-Tcrd* locus.

Re-analysis of published E2A ChIP-seq data obtained from *Id2*- and *Id3*-deficient DP thymocytes (Roy et al., 2018). V_α and V_δ (red bars), D_δ (green bars), J_α and J_δ (blue bars), and *Trdc* and *Trac* (orange bars) segments are indicated. Pseudogenes (black bars). Compared to strain 129, the C57BL/6 *Tcra-Tcrd* locus contains an extra copy of several V_α segments, including *Trav15n-1*. (A) E2A binding across the *Tcra-Tcrd* locus. (B) E2A binding across 155 kb surrounding *Trav15-1-dv6-1* (blue shaded region of A). (C) E2A binding across 155 kb encompassing *Trac* with major E2A peak at E_α (purple shaded region of A). (D) E2A binding at *Trav15-1-dv6-1*, with shaded region indicating the region of deletion on the $Tcrd^{CreER}$ allele (strain 129). (E) E2A binding at *Trav15d-1-dv6d-1*, with shaded region as in D.

4.6 Conclusions

We have determined that secondary *Tcra* recombinations proceed using the most proximal available segments. However, V_α utilization in primary *Tcra* rearrangements is diverse, including segments spread across the proximal, central, and distal portions of the V_α array. The diversity of V_α utilization in primary recombinations is markedly increased by *Tcrd* recombination. *Tcrd* recombination is therefore an important instigator of overall *Tcra* repertoire diversity. Further, we found that *Trav15-dv6* family members, apparently through two distinct mechanisms, expand central and distal V_α utilization. The role of *Trav15-dv6* segments as *Tcrd* recombination substrates, as expected, increased the diversity of secondary *Tcra* recombinations emanating from primary recombinations to immediately distal V_α segments. Surprisingly, though, *Trav15-dv6* segments appear to diversify the *Tcra* repertoire by a second mechanism.

Our repertoire studies emphasize that capture of V gene segments within the *Tcra-Tcrd* locus occurs via fundamentally different mechanisms during *Tcrd* and *Tcra* rearrangement. Capture of V_δ gene segments by the *Tcrd* RC in DN thymocytes can occur over long distances. By contrast, capture of V_α gene segments by *Tcra* RCs appears to occur predominantly as a result of short-range interactions, once those segments are brought into proximity of the RC by prior rounds of rearrangement. One exception is the residual rearrangement of central and distal V_α segments in early rearrangements on DJD alleles (**Figure 9**, right panel). The cluster of rearrangements between V segments upstream of *Trav15d-2-dv6d-2* and J segments *Traj58-Traj48* is not ablated, but rather

reduced from 11.41% to 3.45% of total rearrangements in early DP thymocytes. We think that these rearrangements, which include *Trav15-dv6* segments, may result from residual chromatin marks carried over from DN thymocytes in the absence of *Tcrd* recombination.

The change in the mode of V segment capture between the DN and DP stages of thymocyte development correlates with a conformational change in the *Tcra-Tcrd* locus, with the V array being contracted in DN and extended in DP (Shih and Krangel, 2010). Prior work has shown that long-range capture of V_H segments by the *Igh* RC in pro-B cells occurs via long-range RAG scanning and cohesin-mediated loop extrusion, facilitated by reduced expression of cohesin unloader *Wapl* (Hu et al., 2015; Lin et al., 2018; Ba et al., 2020; Hill et al., 2020; Dai et al., 2021). However, RAG scanning from the *Tcrd* RC appears to be effectively contained within an 80 kb chromatin loop domain formed by the INT1-2 and TEA CTCF binding elements in DN thymocytes (Chen et al., 2015; Zhao et al., 2016). This suggests that long-range capture of V_δ gene segments may occur via diffusion. In contrast, short-range V_α capture by the *Tcra* RC in DP thymocytes can be envisioned to occur either by relatively short-range RAG scanning or diffusive interactions. Regardless, our data suggest that *Tcra* repertoire diversity in mice may result from synergy between two different modes of V segment capture. We expect that our conclusions on *Tcra* repertoire diversity in mice likely extend to humans, since the human *Tcra-Tcrd* locus is similarly organized with V_δ gene segments distributed amongst central and distal V_α gene segments.

5. CTCF binding at INT1 diversifies the *Tcrd* repertoire

5.1 Introduction

Recombination of AgR loci depends entirely on capture of gene segments by the RC. This capture may be mediated by RAG tracking or by diffusion of gene segments into proximity of RAG-bound segments at the RC. In either case, the topology of an AgR locus plays a role in recruiting segments to serve as recombination substrates. CTCF binding and orientation at CBEs dictates chromatin structure (Rao et al., 2014). The 3D organization of the *Tcra-Tcrd* locus is dynamic over the course of thymocyte development, and conformation of the locus is directed, at least in part, by CTCF (Guo et al., 2011b; Rao et al., 2014; de Wit et al., 2015; Yin et al., 2017). To date, it has not been determined whether tracking, diffusion, or a combination of both mechanisms is responsible for segment capture during *Tcrd* recombination, although previous work from our lab and others holds clues to potential mechanisms of V_δ capture by the *Tcrd* RC. It is likely, however, that whatever mechanism brings V_δ segments to the RC at the DJ_δ cluster depends upon CTCF and cohesin-mediated 3D structures.

Regulation of recombination at the *Tcra-Tcrd* locus is particularly complex due to the developmental-stage specific recombination programs of both *Tcrd* and *Tcra*. While most V segments of the *Tcra-Tcrd* locus are utilized for *Tcra* recombination in DP thymocytes, a few V segments interspersed along the V_α-V_δ array and focused particularly at the 3' DJ_δ-proximal portion of the array are utilized for *Tcrd* recombination in DN thymocytes. V gene segments proximal to the DJ_δ cluster, *Trdv1*, *Trdv2-2*, the pseudogene

Trdv3, *Trdv4*, and the inverted *Trdv5*, are thought to be utilized almost exclusively for *Tcrd* recombination (Chen et al., 2015; Shih et al., 2012). A handful of segments in the central and distal portion of the V_α-V_δ array are utilized for both *Tcrd* and *Tcra* recombination. Among them, the members of the *Trav15-dv6* family are used far more frequently in *Tcrd* recombination than any other segments relatively distal to the RC at the DJ_δ cluster (Sleckman et al., 1998; Weber-Arden et al., 2000; Pereira et al., 2000).

Studies of the *Tcra-Tcrd* locus by chromatin immunoprecipitation and sequencing (ChIP-seq) found peaks of CTCF occupancy at CBEs across the *Tcra-Tcrd* locus (Shih et al., 2012; Loguercio et al., 2018). The preponderance of CBEs across the *Tcra-Tcrd* locus are in the forward orientation, including two intergenic CBEs of interest, INT1 and INT2, located between the pseudogene *Trdv3* and *Trdv4*, upstream of the *Tcrd* RC (Shih et al., 2012). Several reverse-oriented CBEs are also spread across the locus, most of which are within the V_α-V_δ array. Of note, TEA, the promoter lying just upstream of the J_α array, is flanked by a reverse-oriented CBE. Our previous work uncovered a domain bounded by INT2 on the 5' end and TEA on the 3' end in DN thymocytes (Chen et al., 2015). This domain contains the *Tcrd* RC and includes *Trdv4*, *Trdd1*, *Trdd2*, *Trdj1*, *Trdj2*, *Trdc*, and the inverted *Trdv5* segment.

In a previous collaborative study from our lab, a CBE was introduced immediately downstream of E_δ (Chen et al., 2016). The primary phenotype of this insertion was a reduction in utilization of *Trdv5*. *Trdv5* is situated between E_δ and TEA, and it is in an inverted transcriptional orientation relative to other *Tcrd* genes. The insertion of a CBE

downstream of E_δ did not impact transcription or accessibility of *Trdv5*, suggesting that the loss of *Trdv5* utilization was due to a change in the chromatin topology that allows *Trdv5* diffusional recombination (Chen et al., 2016). The knockin CBE here serves as a domain boundary sequestering *Trdv5* from the *Tcrd* RC, thereby reducing its utilization in recombination. This loss of *Trdv5* utilization when the topology of the *Tcrd* RC is changed demonstrates the important role CTCF and cohesin-mediated structures serve in regulating recombination at the *Tcra-Tcrd* locus.

By 4C, the INT2 and TEA CBEs were found to be frequent interaction partners that rarely interact with locus regions outside their bounds. The INT1 CBE, which lies directly 5' of INT2, conversely, interacts broadly with regions upstream, inside, and downstream of the INT2-TEA domain in DN thymocytes (Chen et al., 2016). We previously deleted intergenic CBEs INT1 and INT2 (INT1-2KO) in the *Tcra-Tcrd* locus and singly mutated the INT2 CBE (INT2M) in mice to deplete CTCF binding. Normal utilization of V_δ segments in mice depends on CTCF binding at the intergenic CBEs INT1 and INT2 (Chen et al., 2015). INT1-2KO mice display a disturbed *Tcrd* repertoire and diminished γδT cell numbers, with INT2M alone demonstrating a partial phenotype compared to INT1-2KO (Chen et al., 2015; Zhao et al., 2016). In two studies of V_δ utilization in INT1-2KO mice, one by qPCR and the other by LAM-HTGTS, V_δ utilization was skewed (Chen et al., 2015; Zhao et al., 2016). A large proportion of recombinations in INT1-2KO mice were skewed to *Trdv3*; this may explain the observation of reduced γδT cells, as this pseudogene cannot form a functional TCR_δ chain. Alleles rearranging to *Trdv3* are thus incapable of forming

a γδTCR. This shift in recombinations to *Trdv3* and a similar increase in *Trdv2-2* utilization were concomitant with a reduction in central and distal V_δ use (Chen et al., 2015; Zhao et al., 2016). Increased recombinations to forward-oriented cryptic RSSs in the *Trdv3*-proximal region were also observed on INT1-2KO alleles (Chen et al., 2015; Zhao et al., 2016). Further, recombinations involving *Trdv5*, which lies within the domain containing the *Tcrd* RC, were reduced. These results are compatible with RAG tracking from the *Tcrd* RC that is blocked by INT1-2 in WT mice (Zhao et al., 2016). In INT2M mice, a similar phenotype was observed. *Trdv2-2* was utilized more frequently, while distal V_δ segments and *Trdv5* were used at a reduced frequency. However, INT2M mice did not have reduced γδT cell frequency, and changes in the *Tcrd* repertoire were less pronounced than in INT1-2KO mice (Zhao et al., 2016), suggesting partial redundancy between INT1 and INT2 in insulating RAG tracking upstream from the *Tcrd* RC.

Although INT1 is capable of partially compensating as the 5' boundary of the *Tcrd* RC in INT2M mice, INT1 and INT2 appear to serve different functions based on 4C analysis of their interactomes (Chen et al., 2015). Because INT1 was observed by 4C to interact broadly across the 3' portion of the *Tcra-Tcrd* locus, we expected that CTCF binding at the INT1 CBE is responsible for interactions between distal V_δ segments and the RC at the DJ_δ cluster. The domain bounded on the 5' end by INT2 and on the 3' end by TEA does not appear to involve INT1 based on 4C analyses, and we expected the maintenance of rearrangements involving V_δ segments proximal to the DJ_δ cluster within this domain, *Trdv4* and *Trdv5*, would not depend on INT1. We therefore hypothesized that

INT1 mediates *Tcra-Tcrd* structure and *Tcrd* recombination in its capacity to interact with various regions across the 3' end of the *Tcra-Tcrd* locus. INT1 is involved in a wide array of low frequency interactions that we predict occur dynamically to bring DJ_δ-distal V_δ segments into proximity of the RC for diffusion-mediated recombination. However, the extensive contacts between INT1 and the 3' portion of the *Tcra-Tcrd* locus observed by 4C conflict with our prior finding of partial redundancy between INT1 and INT2, leaving the physiological function of INT1 uncertain.

Here, using mice with a mutation in the INT1 CBE (INT1M), we have demonstrated that CTCF binding at this CBE is essential for promoting the utilization of RC-distal V_δ segments and repressing *Tcrd* recombination to the non-coding pseudogene *Trdv3*. We found that, while DN thymocytes proceed normally along their developmental path and γδT cells are not reduced in INT1M mice, the impact on *Tcrd* recombination is substantial. Therefore, we conclude that INT1 is a critical regulator of V_δ utilization in DN thymocytes.

5.2 Reduction of CTCF binding at mutated INT1 CBE

To determine the specific role of the INT1 CBE in *Tcrd* recombination, we generated a line of mice with a deletion in the core CTCF-binding consensus sequence at INT1. Our previous mutation of the INT2 CBE was achieved by homology-directed repair (Chen et al., 2015). INT2M was made by electroporating a gene targeting vector into mouse embryonic stem cells. The oligo used to generate the mutated INT2 CBE was

constructed with homology arms surrounding a scrambled sequence previously shown to deplete CTCF binding at CBEs (Chen et al., 2015). For the present study, we attempted a similar technique, generating a break at INT1 in single-cell mouse embryos using CRISPR/Cas9 and repairing it with an oligo containing homology arms and a scrambled CBE sequence. The INT1 CBE region (**Figure 13A**, boxed and in bold font) contains the core CTCF-binding sequence recognized by zinc fingers 3-7 of CTCF in the forward orientation (Guo et al., 2011b; Chen et al., 2015). Our attempt to repair the CRISPR/Cas9 targeted break with a scrambled sequence was not successful. Instead, we generated a line of C57BL/6J mice with a small deletion in the INT1 CBE, 'INT1M,' which spanned 8bp covering a portion of the core CTCF-binding sequence (**Figure 13A and B**).

The deleted region on INT1M alleles is recognized in a sequence-specific fashion by zinc fingers 6 and 7 of CTCF (Nakahashi et al., 2013; Yin et al., 2017). The sequence generated by repair after deletion of 8bp on INT1M alleles does not recapitulate zinc finger 6 or 7 binding sequences. However, binding of CTCF zinc fingers 3-5 is not likely to be impaired by this small deletion. Previous work demonstrated that deletion of zinc fingers 6 or 7 of the CTCF protein reduced CTCF occupancy and residence time genome wide, as did deletions of zinc fingers 4 or 5 (Guo et al., 2015; Yin et al., 2017). However, a previous deletion of a small portion of the zinc finger 7 CTCF-binding sequence at one CBE did not ablate CTCF binding, while deletions of nucleotides bound by both zinc fingers 5 and 6 ablated CTCF binding (Nakahashi et al., 2013; Quitschke et al., 2000). Because we deleted the binding sequence for both zinc fingers 6 and 7 of CTCF, and the

sequence generated by the deletion does not recapitulate these binding sites, we expected CTCF binding at INT1 to be partially or completely lost on INT1M alleles.

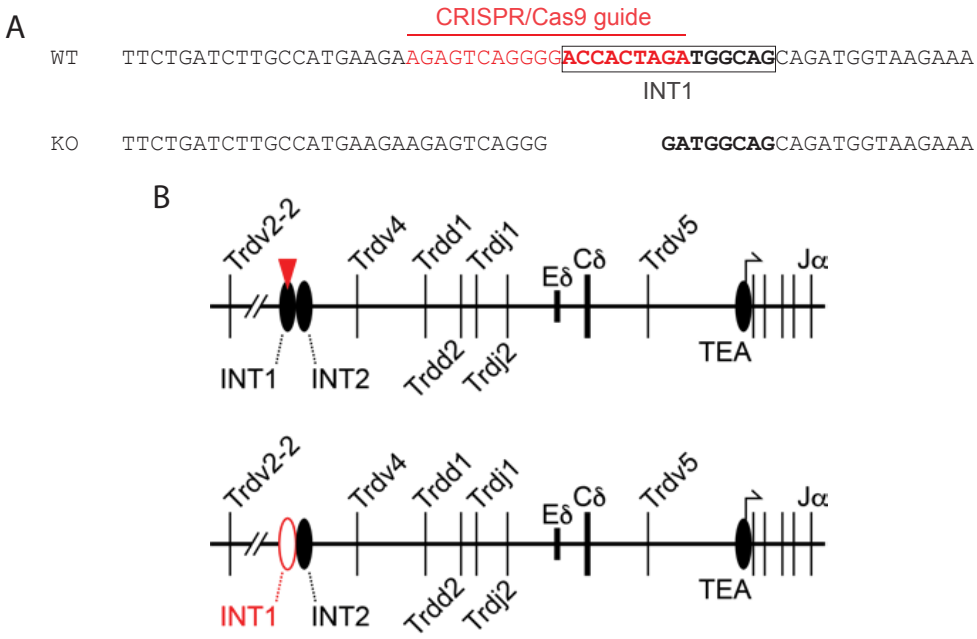

Figure 13: Schematic of INT1M allele.

(A) DNA sequence of mutant alleles. Red, guide sequence for CRISPR/Cas9 targeting. Bold and boxed, INT1 CTCF-binding consensus sequence. Gap, nucleotides deleted in INT1M allele. (B) Top: Overview of the central portion of the *Tcra-Tcrd* locus. Relevant nearby locus features are indicated. Filled black circles indicate CBEs. Red triangle represents sgRNA localization site for INT1 CBE mutation. Bottom: Depiction of *Tcra-Tcrd* locus in INT1M mice. Empty red circle represents INT1 CBE with a small deletion to stop CTCF binding.

INT1M mice were crossed to the *Rag2*$^{-/-}$ C57BL/6J line to derive thymocytes lacking the INT1 CBE that are incapable of recombination. These INT1M *Rag2*$^{-/-}$ mice were used to analyze whether the deletion in INT1 was sufficient to ablate CTCF binding to the CBE. *Rag* knockouts are a useful tool in this respect, as DN thymocytes are

maintained in a pre-recombination state. Thus, the *Tcra-Tcrd* locus is in its intact (unrearranged) state in every cell, maintaining thymocytes at the DN3 stage and facilitating analysis of features along the locus in regions that would otherwise be deleted by recombination events. Our previous analysis of INT2M alleles on the *Rag2*$^{-/-}$ background demonstrated a near complete ablation of CTCF binding at INT2 (Chen et al., 2015). Using qPCR analysis of ChIP for CTCF, we found we found a 60.4% reduction in CTCF binding at the INT1 CBE on INT1M alleles, though this reduction did not reach statistical significance (*P=0.106*) (**Figure 14**). The small deletion in the CTCF-binding consensus sequence does not appear to completely ablate CTCF binding. CTCF binding was not depleted at other nearby sites by the INT1M deletion, though a small but statistically significant increase in CTCF binding was observed at *Trdv4*, possibly due to a changed CTCF interactome on INT1M alleles facilitating increased CBE interactions with *Trdv4*, which does not bind CTCF (Chen et al., 2015).

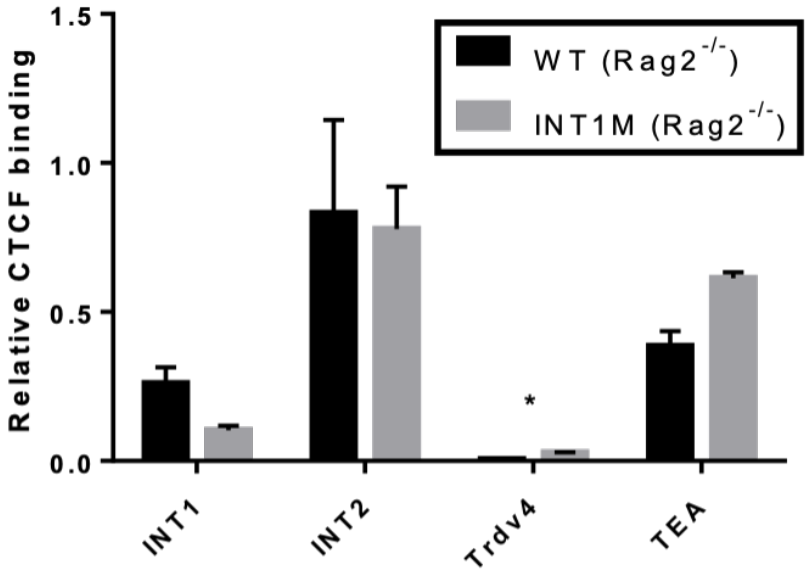

Figure 14: CTCF binding at INT1 reduced on INT1M allele.

CTCF binding was measured by ChIP-qPCR. Binding was calculated as percent of input and normalized to *Myc*. *Trdv4* serves as a negative control. (Unpaired t-test with Welch's Correction, mean with SEM plotted, 2 *Rag2*$^{-/-}$ and 2 INT1M *Rag2*$^{-/-}$ samples, each sample represents a pool of 2-3 mice, **P=0.007*)

We next interrogated whether the reduction in CTCF binding at INT1 would disturb thymocyte development. In our previous studies, the frequency of γδT cells in the INT1-2KO thymus was reduced to half that of WT controls (Chen et al., 2015). The frequency of γδT cells in the thymi of INT2M mice did not significantly differ from WT controls (Chen et al., 2015). Flow cytometric analysis of total thymi from INT1M and WT littermate control mice was performed to determine whether CTCF binding at INT1 impacts development of γδT cells or DN thymocytes.

Although INT1-2KO mice have a significant reduction in γδT cells, INT1M mice, like INT2M mice (Chen et al., 2015), have no significant change in the frequency

of γδT cells with respect to WT mice, as assayed by γδTCR-specific antibody binding to $CD3_{\varepsilon}^{+}$ DN thymocytes (**Figure 15A**). Our previous studies of INT1-2KO mice did not reveal substantial changes to DN thymocyte development. Using flow cytometry, we assayed the proportion of DN $CD3_{\varepsilon}^{-}$ cells in each of the four major DN stages. DN1($CD44^{+}CD25^{-}$), DN2($CD44^{+}CD25^{+}$), DN3($CD44^{-}CD25^{+}$), and DN4($CD44^{-}CD25^{-}$) thymocyte proportions were unchanged in INT1M relative to WT mice (**Figure 15B**). We thus concluded that thymocyte development in the DN stages and to the γδT cell subset are not dependent upon physiological levels of CTCF binding at INT1.

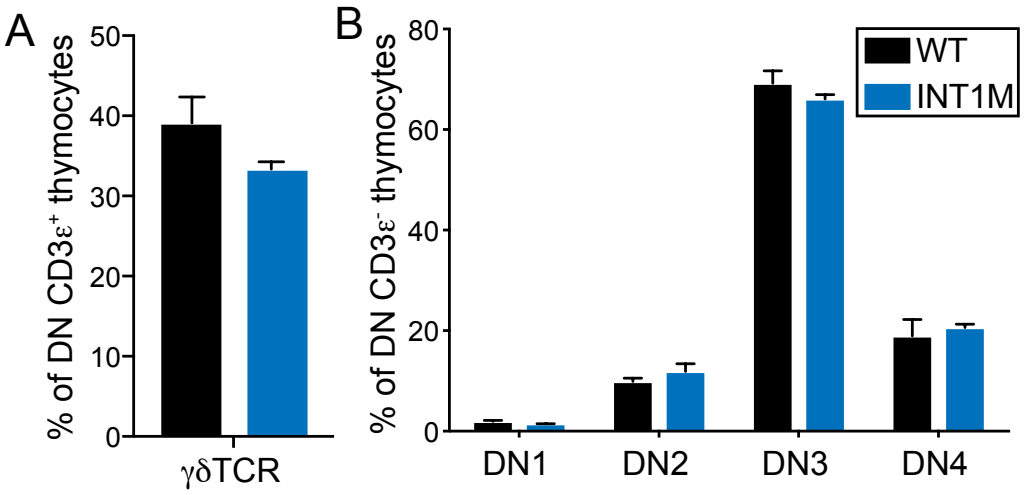

Figure 15: Normal thymic development in INT1M mice.

Flow cytometry-based comparison of thymocyte development from WT and INT1M mice stained for CD4, CD8, $CD3_{\varepsilon}$, γδTCR, CD25 and CD44. (Left) % of $Lin^{-}CD4^{-}CD8^{-}CD3_{\varepsilon}^{+}$ cells expressing γδTCR. (Right) % of $Lin^{-}CD4^{-}CD8^{-}CD3_{\varepsilon}^{-}$ cells in the DN1($CD44^{+}CD25^{-}$), DN2($CD44^{+}CD25^{+}$), DN3($CD44^{-}CD25^{+}$), and DN4($CD44^{-}CD25^{-}$) stages of thymocyte development. (Unpaired t-test with Welch's Correction, mean with SEM plotted, no significant differences, n=3-5 mice, WT are littermate controls.)

5.3 The INT1 CBE diversifies the Tcrd repertoire

Using qPCR assays, we analyzed V_δ recombinations to *Trdj1* in sorted DN2/3($CD44^{+/-}CD25^{+}$) thymocytes. As expected, we observed decreased distal V_δ usage, particularly to segments in the *Trav15-dv6* family, including those in both the *Trav15(d)-1-dv6(d)-1* and *Trdv15(d)-2-dv6(d)-2* groups. As these segments constitute the majority of central and distal V_δ utilization in adult thymocytes, these significant reductions likely represent a substantive loss of all but relatively proximal V_δ use in INT1M thymocytes (**Figure 16**). We observed concomitant increase in utilization of proximal V_δ segments, including a trend toward increased use of *Trdv2-2* and a significant increase in recombination to *Trdv3*, which lies just upstream of INT1-2. The utilization of upstream proximal V_δ segments *Trdv11*, *Trdv9*, and *Trdv1* did not change in INT1M thymocytes. However, *Trdv4* and *Trdv5*, which are located within the domain containing the *Tcrd* RC, are not utilized with increased or even equivalent frequency in INT1M mice compared to WT mice. Indeed, *Trdv4* demonstrates a trend of decreased utilization and *Trdv5* is used significantly less frequently in INT1M mice. These findings indicate decreased distal V_δ utilization and reduced recombinations involving V_δ segments within the *Tcrd* RC on INT1M alleles, as recombinations shift upstream of the RC to *Trdv3*.

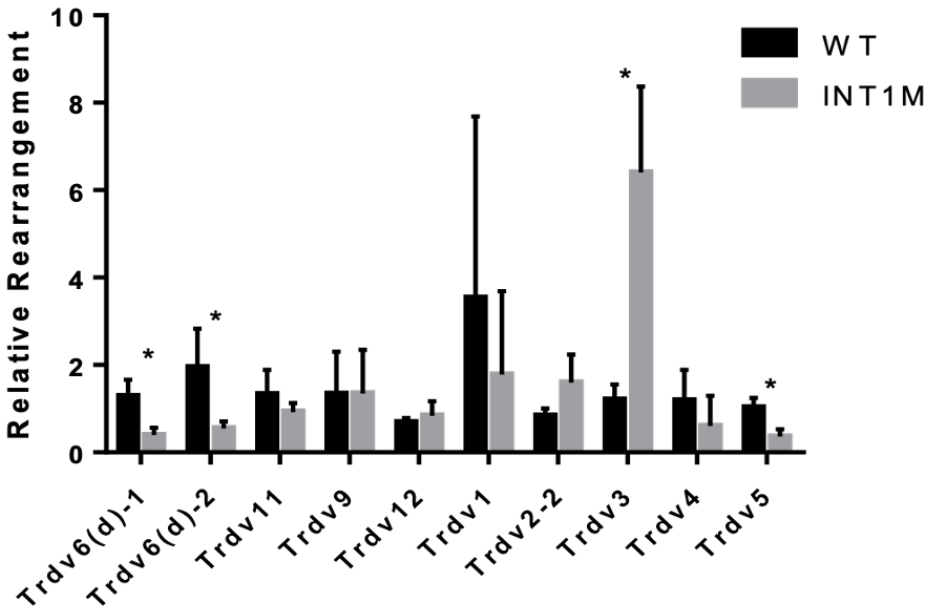

Figure 16: Disrupted *Tcrd* recombination in INT1M mice.

gDNA harvested from DN2/3 thymocytes of 6 WT and 6 INT1M mice were interrogated for VDD-to-J rearrangements to *Trdj1* with the indicated V_δ segments using Taqman-PCR. Normalized to CD14. Multiple t-tests, Holm-Sidak correction, Mean with SD plotted. WT are littermate controls derived from heterozygous breeding. All experiments performed in duplicate. *$P \leq 0.05$*

5.4 Conclusions

The INT1M mouse has a partial loss in CTCF binding to the INT1 CBE. This reduction in CTCF binding does not change thymocyte development, but it does impact *Tcrd* recombination by altering V_δ utilization. On INT1M alleles we observed a reduction in recombination to distal V_δ segments, as well as a reduction in recombinations to V_δ segments within the INT2-TEA bounded *Tcrd* RC, particularly the downstream reverse-oriented *Trdv5*. Intriguingly, as was observed in INT1-2KO mice, recombination to the

pseudogene *Trdv3*, which cannot form a functional TCR_{δ} chain, was robustly upregulated in INT1M mice. Because recombinations involving *Trdv3* cannot generate a functional γδTCR, it is surprising that INT1M mice do not have a reduction in γδT cell frequency, as was observed in INT1-2KO mice. The loss of γδT cellularity in INT1-2KO mice was attributed to the loss of potential functional δTCR chains due to more frequent recombinations involving *Trdv3* (Chen et al., 2015). γδT cell frequency may be retained in INT1M mice due to the partial phenotype of INT1M as compared to INT1-2KO. *Trdv3* utilization in INT1-2KO mice is as much as 10-fold more frequent than in WT mice (Chen et al., 2015; Zhao et al., 2016). Although *Trdv3* utilization is increased in INT1M mice, the extent of increase is not as extreme. Other *Tcrd* recombinations may be frequent enough to maintain γδT cell frequency.

The changes in the *Tcrd* repertoire suggest a role for INT1 as a cooperating partner with INT2 in serving as the upstream boundary of the *Tcrd* RC. The increased utilization of the pseudogene *Trdv3* is particularly interesting, as this segment lies immediately upstream of INT1-2. This finding may indicate that INT1, as observed in INT2M mice, serves a partial role as a support to the domain boundary activity of INT2. Perhaps INT1 serves as a secondary blockade to RAG tracking from the RC that dampens utilization of *Trdv3* and other upstream RC-proximal V_{δ} segments. In the INT1-2KO mouse, interactions with TEA spread to the upstream intergenic CBE INT3, which lies 19kb downstream of *Trdv1* (Chen et al., 2015). Tracking to INT3, which is located immediately upstream of *Trdv2-2* and *Trdv3*, would explain increased utilization of these

segments in INT1M thymocytes. A similar phenotype was observed in INT1-2KO, where not only were recombinations to *Trdv3* dramatically increased, but joining between D_δ segments and cryptic RSSs in this region occurred with increased frequency (Zhao et al., 2016). In INT2M mice, INT1 was able to partially compensate as the 5' boundary of the *Tcrd* RC, but interactions with TEA also spread upstream to INT3. It is unclear whether, like IGCR1 at the *Igh* RC, INT1-2 is a leaky boundary facilitating long-range tracking from the *Tcrd* RC (Zhao et al., 2016). This seems unlikely, because the propensity for long-range RAG tracking from the Igh RC through IGCR1 is attributed to repression of the cohesin unloading factor Wapl in B cell progenitors compared to other cell types (Hill et al., 2020; Dai et al., 2021). Reduced levels of Wapl in pre-B cells allow long-range cohesin-mediated RAG tracking to extend beyond IGCR1 due to increased cohesin occupancy time. However, in thymocytes *Wapl* is not repressed, so INT1 and INT2 likely serve as stronger barriers than IGCR1 (Hill et al., 2020). Instead, it appears that INT1-2 functions as the upstream boundary of cohesin-mediated RAG tracking from the *Tcrd* RC, more akin to the role of Sis at the *Igk* RC (Hill et al., 2020).

The other possible explanation is that INT1 serves as an anchor for cohesin-scanning that brings V_δ segments into proximity of the *Tcrd* RC for recombination by diffusion. This is supported by 4C observations that INT1 interacts broadly in the 3' portion of the *Tcra-Tcrd* locus, rather than being involved in frequent interactions with the INT2-TEA domain (Chen et al., 2015). However, this explanation for the INT1M phenotype is less plausible as it cannot explain the reduction in *Trdv5* use, as *Trdv5* lies within the INT2-TEA loop, which should not be perturbed by deletion of INT1 if INT1

serves primarily to facilitate interactions between upstream V_δ segments and the RC. It remains possible that INT1 serves both as a barrier to RAG scanning upstream of the *Tcrd* RC and as an anchor of cohesin-mediated scanning to more distal upstream V_δ segments. INT1 as an anchor of interactions between upstream V_δ segments and the RC also cannot explain the increased *Trdv3* utilization observed in INT1M thymocytes. However, our previous 4C analysis does not support the role of INT1 as an anchor for interactions with segments as distal as the *Trav15-dv6* family. The most distal upstream region of interaction with INT1 appears to terminate near *Trav6-7-dv9* (Chen et al., 2015).

The utilization of *Trdv5* suggests that at least some proportion of *Tcrd* recombinations occur by diffusion. As this segment is both downstream of the RC and inverted with respect to the rest of the locus, it cannot be utilized by RAG tracking. It is used fairly frequently in the WT *Tcrd* repertoire. However, disruption of the INT or INT2 CBEs leads to a dramatic loss of *Trdv5* utilization. This loss is likely due to increased RAG tracking upstream, which facilitates capture for RSSs oriented toward the D_δ RSSs. These increased upstream recombinations inherently reduce the proportion of *Trdv5* utilization, as only one recombination can occur per allele.

We expect that in DN thymocytes, diffusion is the primary mode of V_δ segment capture by the *Tcrd* RC. This mode of V segment capture would allow utilization of both proximal and distal V_δ segments, leading to the varied *Tcrd* repertoire. The nested structure of the *Tcra-Tcrd* locus, which is conserved from mouse to man, indicates that

this pattern of *Tcrd* recombination will impact the *Tcra* repertoire. In INT1M and INT2M mice, like INT1-2KO mice (Chen et al., 2015), the *Tcra* repertoire is likely to be impacted by proximally-skewed V_δ utilization due to increased RAG tracking upstream of the RC.

6. Discussion and future directions

6.1 Summary

The goal of this dissertation was to elucidate mechanisms underlying diversification of the *Tcra* and *Tcrd* repertoires. TCRs have the potential to recognize a broad spectrum of antigens by combinatorially assembling receptor chains from a limited set of gene segments. The process of recombination, which uses a cut-and-paste mechanism to bring these gene segments together into transcriptional units, is complex and tightly regulated. At the shared *Tcra-Tcrd* locus, *Tcrd* recombination precedes *Tcra* recombination, and distinct mechanisms dictate orderly rearrangement of each.

Herein, we have provided evidence that the 3D architecture of the *Tcra-Tcrd* locus is responsible for physiological *Tcrd* repertoire generation. Specifically, V_δ utilization was determined to be dictated in part by a CBE upstream of the *Tcrd* RC. Moreover, we found that *Tcrd* recombination redirects the starting point for V_α utilization in early V_α-to-J_α recombinations, which are vital starting points for diagonal patterns of secondary recombination. Particularly, the *Trav15-dv6* family segments or their *cis* regulatory regions were demonstrated to enhance central and distal V_α use. The role of

Tcrd recombination in imparting diversity on the *Tcra* repertoire indicates that the nested structure of the *Tcra-Tcrd* locus has a functional role in generating immensely diverse αβTCRs to better prepare the host to recognize and respond to a world of antigens. The structure of the *Tcra-Tcrd* locus is conserved from mouse to man (Glusman et al., 2001). This suggests that our findings that the unique topological and organizational structure of the locus are imperative for diverse repertoire generation may extend beyond mice.

6.2 Capture of V segments by the Tcrd *and* Tcra *RCs*

One of the most exciting avenues for further experimentation lies in uncovering the mechanisms of V segment capture by the two RCs of the *Tcra-Tcrd* locus. The central question of further research in this area is whether RAG tracking or diffusion brings segments into proximity for synapsis (Ba et al., 2020; Zhang et al., 2019; Lin et al., 2018; Hu et al., 2015; Hill et al., 2020). The pattern of *Tcrd* recombination, with its use of proximal, distal, and inverted gene segments, differs substantially from the pattern of *Tcra* recombination, which occurs mostly between proximal V_α and J_α segments. These repertoire differences suggest that V segment capture may differ in these two stages of recombination at the shared *Tcra-Tcrd* locus. Based on our findings herein and on our previous work, we hypothesize that V_δ segment capture by the *Tcrd* RC occurs primarily by diffusion. We hypothesize that V_α segment capture by the *Tcra* RC, in contrast, occurs primarily by short-range RAG tracking or short-range diffusion.

The evidence of this work and our previous research points to a diffusion-mediated mechanism of V_δ segment capture by the *Tcrd* RC. The INT1M, INT2M, and INT1-2KO mice all demonstrate reduced RC-distal V_δ use and reduced *Trdv5* use when the 5' boundary of the *Tcrd* RC is disrupted (**Figure 16**) (Chen et al., 2015; Zhao et al., 2016). Under these experimental conditions, RAG tracking appears to be the primary mode of V_δ capture, as without the INT1-2 impediment intact, cohesin-mediated extrusion can bring upstream segments into proximity of the RC until the next CBE is reached. When the 5' boundary of the domain containing the *Tcrd* RC is disrupted by the experimental deletion of INT1, INT2, or both, *Trdv3* and/or *Trdv2-2* are utilized more frequently, which appears to demonstrate RAG scanning upstream of the RC can extend at least until INT3 in these mice (**Figure 16**) (Chen et al., 2015; Zhao et al., 2016). The loss of utilization of more RC-distal V_δ segments in these mice suggests that RAG tracking does not extend to the 5' portion of the V_α-V_δ array. In physiological conditions, INT1-2 appears to serve as an insulator of RAG tracking from the RC, making diffusion the likely mode of capture of upstream V_δ segments and explaining the less frequent recombinations to *Trdv2-2* and *Trdv3* on WT alleles. Diffusion-mediated recombination would also explain the use of distal V_δ segments; if RAG tracking were the primary method of recombination, use of distal segments would occur less often due to both CTCF impediments to cohesin-mediated loop extrusion and the availability of more proximal segments for recombination. However, there is not yet conclusive evidence about the physiological mechanism of V_δ capture.

Trdv5 capture by the *Tcrd* RC appears to be straightforward. Because *Trdv5* is not only downstream of the RC, but in the inverted transcriptional orientation compared to the *Tcrd* RC, it must be captured by diffusion. Unidirectional RAG tracking from D_δ segments would not bring the 23-RSS of *Trdv5* into a synaptic complex with the 12-RSS of either D_δ segment as these heptamers are not convergent. Therefore, *Trdv5* appears to be used less often in mice with disruptions to INT1-2 (Chen et al., 2015; Zhao et al., 2016) due to increased RAG tracking upstream of the RC, which reduces the opportunity for diffusion-mediated *Trdv5* use. We previously inserted a CBE between the RC and *Trdv5*, which reduced *Trdv5* utilization without disrupting its transcription or accessibility (Chen et al., 2016). Rather, a loop between the ectopic CBE and INT1-2 forms. This loop appears to exclude *Trdv5* from the domain around the RC, as interactions between *Trdv5* and the RC are reduced (Chen et al., 2016). CTCF is not only involved in impeding cohesin-mediated RAG tracking during loop extrusion, but also in establishing the structures necessary for diffusion-mediated RSS capture by facilitating 3D structures that bring segments into proximity of the RC. Therefore, future experiments will need to examine tracking and diffusion without disrupting CTCF-mediated chromatin topology of the *Tcra-Tcrd* locus.

Recombination of *Tcra* appears to be mediated by short-range RAG tracking or diffusion. The *Tcra* RC is in the J_α array, first focused at proximal J_α segments and progressing to more distal J_α segments as rounds of recombination proceed (Ji et al., 2010; Schatz and Ji, 2011). The earliest recombinations utilize proximal J_α segments,

which are accessible and transcriptionally activated due to the TEA and Jα49 promoters (Villartay et al., 1987; Shimizu et al., 1993; Chasseval and Villartay, 1993; Villey et al., 1996; Mauvieux et al., 2003; Hawwari et al., 2005). Multiple pieces of evidence in this dissertation support short-range RAG tracking or diffusion from the *Tcra* RC. First, the observed temporal kinetics of recombination of the HYα-KI allele demonstrate restricted V_α utilization in early recombinations (**Figure 6**). Early recombinations in this model of a fixed primary rearrangement extend predominantly to the most proximal subset of V_α and J_α segments, and further rounds of recombination continue to utilize proximal V_α and J_α segments. Further, in the absence of *Tcrd* recombination, not only is the *Tcra* repertoire depleted of diversity (**Figure 8**), but early recombinations are restricted mostly to the J_α-proximal portion of the V_α array (**Figure 9**). This pattern of recombination on HYα-KI and DJD alleles suggests that V_α use, much like J_α use, occurs processively, using the most proximal available segments in any given recombination. Because the V_α-V_δ array contains a plethora of CBEs (Shih et al., 2012; Loguercio et al., 2018), cohesin-mediated extrusion would pause at various points along the array (Rao et al., 2014), allowing segments proximal to CBEs to be utilized in a given round of recombination. This is compatible with a RAG scanning mechanism from the *Tcra* RC wherein proximal V_α segments, those first encountered by scanning, are brought into synapsis with RAG-bound J_α segments at the highest frequency. It is also compatible with short-range diffusion, as cohesin extrusion anchored by the V_α-oriented CBE at TEA on the 3' end could form a loop bringing J_α-proximal segments into 3D proximity of the RC by

diffusion for primary recombination. Both of these mechanisms are compatible with the observed contracted conformation of the 3' portion and extended 5' portion of the *Tcra-Tcrd* locus in unrearranged DP thymocytes (Shih et al., 2012). Further rearrangements, after recombinatorial deletion of the TEA CBE, may involve RAG tracking or diffusion between *de novo* loops anchored by the many CBEs along the V_α array.

A central remaining question is whether *Tcrd* and *Tcra* recombination indeed occur by RAG scanning or by diffusion. This could be examined in several ways. The experimental introduction of nuclease-dead Cas9 (dCas9), as has been applied to examine RAG scanning at the *Igh* locus (Zhang et al., 2019), is one potential avenue. dCas9 was targeted to serve as a blockade to cohesin-mediated extrusion and RAG scanning at the *Igh* locus without disrupting CTCF binding. The region proximal to the experimentally introduced impediment to RAG scanning recombined more frequently, while recombination to more distal targets was reduced (Zhang et al., 2019). This demonstrated that RAG tracking within a CTCF-bounded domain was impeded by a blockade to loop extrusion. dCas9 could be targeted downstream of a V_δ segment to impede RAG scanning. If RAG scanning is the mechanism of V_δ capture, analysis of V_δ utilization by qPCR or LAM-HTGTS would demonstrate reduced utilization of segments distal to the dCas9 binding site. If diffusion is the mechanism of V_δ capture, dCas9 binding would not be expected to disrupt normal V_δ utilization. Targeting dCas9 near a J_α-proximal V_α segment would serve as an impediment to RAG scanning. Experiments to determine the method of V_α capture could take advantage of the *Tcrd*^{CreER} allele in combination with

the *Rosa26*^{ZsG} reporter to examine whether the earliest *Tcra* recombinations are restricted to V_α segments proximal to the bound dCas9. The *Tcra* repertoire could be analyzed by qPCR, LAM-HTGTS, or HTS. If RAG scanning is the mechanism of V_α capture, as we predict, we expect that more distal V_α segments would be used less frequently than those proximal to bound dCas9. If the early *Tcra* repertoire is unchanged compared to alleles without targeted dCas9, the mechanism of V_α capture could be attributed to diffusion, as diffusion would not be blocked by DNA-bound dCas9. However, *Tcrd* recombination or impacts of dCas9 binding on V_δ utilization may complicate analysis of V_α capture. Chromatin contacts between the *Tcrd* or *Tcra* RC and segments upstream and downstream of the dCas9 binding site could be analyzed by 3C or 4C to determine whether interactions with regions upstream of dCas9 are capable of interaction with the RC when cohesin-mediated loop extrusion and RAG scanning are blocked. Maintenance of WT interaction levels between segments upstream of dCas9 and the RC would demonstrate that RAG tracking is not necessary to bring V segments into proximity of the RC.

However, dCas9 targeting is not the optimal strategy to differentiate tracking from diffusion, as it is possible that impeding cohesin-mediated RAG scanning with dCas9 would impair loop extrusion and prevent the 3D structures necessary for diffusion-mediated V segment capture in addition to blocking RAG tracking. Great care would need to be taken to select a target site for dCas9 binding that is within the domain containing the RC to prevent this complication. For V_δ utilization, dCas could be targeted

downstream of *Trdv4*, which resides within the RC domain, and is predicted to be utilized by RAG tracking from the RC due to the strong preference for joints involving convergently-oriented heptamers upstream of the RC within this domain (Zhao et al., 2016; Hu et al., 2015). However, this would not allow determination of the mechanism of capture of more distal V_δ segments. For V_α utilization, alleles which have recombined *Tcrd* would delete swaths of CTCF elements between the J_α and V_α segments, and dCas9 could be targeted to an RC-proximal V_α segment to differentiate diffusion from tracking, though variable *Tcrd* recombination would likely confound these analyses. LAM-HTGTS should be employed, as it provides a robust method for detecting recombinations to cryptic RSS sites, which would provide more information about whether heptamer orientation is predictive of recombination frequency. If a strong preference for convergent heptamers is observed, RAG tracking could be determined to be the mechanism of V_α capture.

An alternative method for determining whether *Tcrd* and *Tcra* recombination are mediated by RAG scanning or diffusion, with fewer limitations, would be to invert V segments, including the RSSs. Because V segment capture by RAG scanning occurs between convergent RSS heptamers (Hu et al., 2015), inversion disrupts capture by scanning. Previous work on the *Igh* locus demonstrated that experimental inversion of V_H segments reduces their interactions with the RC (Hill et al., 2020; Dai et al., 2021). Accordingly, experimentally inverted V_H and D_H segments are utilized less frequently (Hill et al., 2020; Dai et al., 2021; Zhang et al., 2019). An optimal genetic approach that would allow assessment of V segment capture in both *Tcrd* and *Tcra* recombination

would require the generation of two novel lines of mice. One of these would carry an inversion of a frequently used V_δ segment outside the *Tcrd* RC, such as *Trdv1* or *Trdv2-2* for analysis of *Tcrd* recombination. The other would contain an inverted V_α segment that is not utilized for *Tcrd* recombination. Creating two lines of mice rather than a single line with which to analyze both *Tcrd* and *Tcra* recombination is preferable to a large inversion containing both V_δ and V_α segments. It avoids involving the inversion of CBEs, which would confound analysis due to changes in CTCF-mediated chromatin topology. Further, if *Tcrd* recombination occurs by diffusion, V_α segments between an inverted V_δ segment and the *Tcra* RC would be returned to an orientation with heptamers convergent with J_α heptamers. *Tcrd* and *Tcra* recombination on alleles with inverted V segments could be assayed by qPCR, LAM-HTGTS, or HTS and compared to their WT counterparts. If RAG scanning is the mechanism of V segment capture, the inverted segments would recombine less frequently. Alternately, if diffusion is the mechanism of V segment capture, utilization of the inverted segments would not be significantly reduced.

6.3 Putative cis *regulatory regions downstream of* Trav15-dv6 *segments*

Another intriguing finding that deserves further exploration is the putative regulatory regions we discovered proximal to *Trav15d-1-dv6d-1* and *Trav15-1-dv6-1*. Beyond confirming our hypotheses about the impact of *Tcrd* recombination on the *Tcra* repertoire, we uncovered an unexpected role for the *Trav15-dv6* family segments or their *cis* regulatory elements in driving distal *Tcra* recombinations. Secondary *Tcra*

recombinations upstream of *Trav15d-1-dv6d-1* or *Trav15-1-dv6-1* that were not expected to be impacted by the reduction in primary rearrangement diversity due to loss of specific *Tcrd* recombinations on *Trav15d-1-dv6d-1* KO or *Trav15-1-dv6-1* KO alleles were nonetheless reduced (**Figures 10** and **11**). The mechanism of this loss of recombinations has not yet been elucidated. However, we noted peaks of E2A binding in DP thymocytes immediately downstream of both segments within the deleted regions in our mouse models (**Figure 12**) (Roy et al., 2018). As E2A is known to promote *Tcrd* recombination and the use of particular V_δ segments (Bain et al., 1999), E2A may promote utilization of *Trav15-dv6* segments in DN thymocytes and *Trav15-dv6* and nearby upstream V_α segments in DP thymocytes.

These putative regulatory regions may be the reason the *Trav15-dv6* family is utilized in *Tcrd* recombination while other distal V_δ segments are used less often (Zhao et al., 2016). Further, on alleles that do not undergo *Tcrd* recombination, such as the HYα-KI and DJD alleles, these regions may drive the minor utilization of central and distal V_α segments observed in early recombinations (**Figures 6** and **9**). These recombinations are not as frequent in the steady-state HYα-KI or DJD repertoires as they are in their WT counterparts, suggesting that *Tcrd* recombinations are the major driver of central and distal V_α utilization in early rearrangements (**Figure 5**) (Carico et al., 2017). Intriguingly, however, the small region of distal V_α-to-proximal J_α recombinations observed in the minor diagonal region (diagrammed in **Figure 4B**) at early timepoints on HYα-KI and DJD alleles begins with *Trav15d-1-dv6d-1*.

What remains to be determined is whether the observed E2A binding sites immediately downstream of *Trav15d-1-dv6d-1* and *Trav15-1-dv6-1* indeed serve as regulatory regions, and how these putative regulatory elements function. One possibility is that *cis* regulatory regions deleted in *Trav15d-1-dv6d-1* KO and *Trav15-1-dv6-1* KO mice are responsible for activating transcription or accessibility of these segments for *Tcrd* recombination. These regulatory regions may additionally promote utilization of *Trav15-dv6* segments and nearby upstream V_α segments in *Tcra* recombination on alleles that do not delete these segments during *Tcrd* recombination. RAG activity depends upon accessibility and GLT, which are facilitated by histone modifications, nucleosome depletion, and chromatin structure under control of promoters and enhancers (Abarrategui and Krangel, 2006; Cobb et al., 2006; Schatz and Ji, 2011). The E88 enhancer at *Igk* is an example of a *cis* regulatory region within a V array that increases GLT and accessibility of nearby V segments (Barajas-Mora et al., 2019). We hypothesize that *cis* regulatory regions proximal to *Trav15-dv6* segments may also serve as local enhancers to promote *Trav15-dv6* utilization and utilization of nearby V_α segments.

The potential *cis* regulatory elements proximal to *Trav15-dv6* segments could be tested by deleting the E2A-binding regions downstream of a *Trav15-dv6* family segment. If a novel mouse with deletion of E2A binding sites at either *Trav15d-1-dv6d-1* or *Trav15-1-dv6-1* was generated, several tests could be conducted to determine whether E2A binding serves as a local enhancer of V segments. If the *cis* regulatory regions are serving as local enhancers, GLT and accessibility of the *Trav15-dv6* segment may be impacted by deletion of the proximal E2A binding site. GLT can be assessed by qPCR

analysis of *Trav15-dv6* on E2A binding site deleted mice. Accessibility of the *Trav15-dv6* segment can be assessed by the assay for transposase-accessible chromatin and sequencing (ATAC-seq), which detects accessibility by identifying chromatin regions vulnerable to transposase activity (Buenrostro et al., 2015). As H3K4me3 is not only known as a marker of active promoters, but also supports RAG binding and activity, it would make a good marker of whether *Trav15-dv6* segments are less likely to be utilized in recombination in the absence of the E2A binding site (Liu et al., 2007; Matthews et al., 2007; Ramón-Maiques et al., 2007; Ji et al., 2010). H3K4me3 can be assayed by ChIP and sequencing or qPCR at *Trav15-dv6* segments. All of these assays should be performed using E2A binding site deleted mice on the *Rag2*$^{-/-}$ background, in which alleles retain germline configuration. If E2A binding immediately downstream of *Trav15d-1-dv6d-1* or *Trav15-1-dv6-1* serves as a *cis* regulatory region, we expect a loss of accessibility, GLT, and H3K4me3 at the *Trav15-dv6* segment in DN thymocytes when the E2A binding site is deleted. This loss of local enhancer activity would also reduce *Tcrd* recombinations utilizing these segments, which could be assayed by qPCR. Similar analyses could be conducted on E2A binding site knockout mice to assess whether the putative *cis* regulatory regions are necessary for GLT, accessibility, H3K4me3, and recombination of *Trav15-dv6* segments and proximal upstream V_α segments in DP thymocytes.

It is also possible that these *cis* elements regulate the topology of the V_α-V_δ array. The observed loss of secondary rearrangements upstream of *Trav15-dv6* segments, even within the major diagonal, where recombinations were not expected to be disrupted in

Trav15d-1-dv6d-1 KO and *Trav15-1-dv6-1* KO mice (**Figure 10B**, region 2, and **11B**, region 3) may be due to disrupted V_α segment capture by the RC. Moreover, rearrangements using the V_α segments immediately J_α-proximal of *Trav15-dv6* segments increased in these mice. It appears as if recombinations upstream are blocked by the loss of *Trav15-dv6*, almost as if RAG tracking is impeded. However, a sequence permissive for CTCF binding is not introduced by either of these deletions, so insulator insertion is an unlikely mechanism. It is possible, however, that cohesin loading is impacted by the *Trav15d-1-dv6d-1* and *Trav15-1-dv6-1* deletions. 3D chromatin structures are established by loop extrusion; this extrusion is mediated by cohesin, and cohesin-mediated loop extrusion forms topological loops and domains (Alipour and Marko, 2012; Sanborn et al., 2015; Fudenberg et al., 2016). Cohesin is loaded on chromatin by a complex containing the protein NIPBL (Ciosk et al., 2000). Cohesin goes on to form loop structures by active ATP-driven extrusion along the chromatin (Kim et al., 2019; Davidson et al., 2019), which is stopped to form loops between convergently-oriented CBEs (Rao et al., 2014). Enhancers have been demonstrated to be associated with binding of both cohesin and its loading factor NIPBL (Kagey et al., 2010). Therefore, if the putative *cis* regulatory regions serve as enhancers, loading of cohesin at this site may contribute to loop extrusion that promotes recombination in this region. Cohesin-mediated loop extrusion can allow RAG tracking from the RC to V segments or create loops that promote diffusion-mediated synapsis in 3D space (Jain et al., 2018; Zhang et al., 2019; Hill et al., 2020; Peters, 2021). Cohesin loaded at *Trav15-dv6* segments may extrude toward the RC, facilitating recombination. It is possible that utilization of *Trav15-dv6* segments is

promoted by chromatin architecture in both DN and DP thymocytes, and this architecture may be promoted by NIPBL loading of cohesin at the putative *cis* regulatory elements.

Future work could examine whether cohesin loading is impacted by loss of either region. *Trav15d-1-dv6d-1* KO and *Trav15-1-dv6-1* KO mice or mice with specific deletions of the E2A binding regions could be used to examine whether chromatin loop formation depends upon these regions. ChIP-seq for cohesin or its loading factor NIPBL could be performed to determine whether the *cis* regulatory regions load cohesin to promote loop extrusion for V_δ or V_α capture. If *Trav15d-1-dv6d-1* or *Trav15-1-dv6-1* are important sites of cohesin loading to maintain chromatin architecture for diffusion- or tracking-mediated RAG synapsis, we would expect reduced NIPBL and cohesin around these segments and in the upstream V_α region on alleles lacking either *Trav15-dv6* segment or the associated E2A binding sites. The loss of cohesin loading and associated loop extrusion would also be expected to reduce *Tcrd* and *Tcra* recombinations using the *cis Trav15-dv6* segment and *Tcra* recombinations to upstream V_α segments in E2A binding site knockout mice, which could be assayed by qPCR, LAM-HTGTS, or HTS.

6.4 Organization of the Tcra-Tcrd locus dictates antigen-sensing potential of the αβTCR repertoire

The unique nested structure of the *Tcra-Tcrd* locus is mostly conserved between mice and humans, in addition to other animals (Glusman et al., 2001). This evolutionary conservation suggests that the nested structure of *Tcra-Tcrd* provides an evolutionary benefit for generating diverse, antigen-sensing TCR repertoires that protect hosts from

pathogens. One interesting avenue of further research would be to determine how *Tcrd* recombination's impact on *Tcra* repertoire impacts antigen recognition by αβTCRs.

Previous work from our lab studied the abundance of a particular population of αβT cells when a fixed *Tcra* recombination is introduced (HYα-KI) or when distal V_δ utilization is curtailed (INT1-2KO) in mice (Carico et al., 2017). These mucosa-associated invariant T (MAIT) cells, which recognize bacterial and fungal pathogens capable of producing riboflavin, are highly abundant in humans (Meermeier et al., 2018). In mice, the invariant αTCR chain of MAIT cells is generated by a recombination between *Trav1* and *Traj33* (Tilloy et al., 1999; Meermeier et al., 2018). Because *Trav1* is the most distal V_α segment and *Traj33* lies in the central portion of the J_α array, *Trav1*-to-*Traj33* recombinations were reduced in HYα-KI and INT1-2KO mice, where recombinations are restricted to the major diagonal (Carico et al., 2017). MAIT cells were found by flow cytometry to be depleted in these mice compared to their WT counterparts. The invariant TCR found in MAIT cells facilitates investigation of their development in these models.

We expect that, due to the reduced diversity of V_α-to-J_α recombinations observed in DJD, *Trav15d-1-dv6d-1* KO, and *Trav15-1-dv6-1* KO mice, and the expected reduction of *Tcra* diversity in INT1M mice, MAIT cells would also be depleted in these lines. This could be assayed by flow cytometry using a specialized tetramer recognized by MAIT TCRs (Corbett et al., 2014), as we did previously (Carico et al., 2017). It would also be interesting to determine whether the reduction in these cells leads to reduced

capacity to clear pathogens, such as *Candida albicans, Mycobacterium tuberculosis, and Neisseria gonorrhoeae*, which are known to be recognized by MAIT cells (Le Bourhis et al., 2010; Meierovics et al., 2013; Meermeier et al., 2018). Further, these restricted *Tcra* repertoires would likely also diminish other αβT cells responsible for robust immunity.

6.5 The role of the INT1 CBE in regulating structure and recombination of the Tcra-Tcrd locus

The study of the role of the INT1 CBE in regulating recombination and the conformation of the *Tcra-Tcrd* locus in DN thymocytes has several key limitations. The foremost of these are in the mouse model. Due to failed homology-directed repair, the small deletion on INT1M alleles does not completely ablate CTCF binding at INT1 (**Figure 14**). As such, current conclusions can only be drawn from the results of reduction in CTCF binding at INT1, rather than loss of CBE activity. Further, the mouse model was generated on the C57BL/6J background, which is not ideal for several reasons. The INT2M and INT1-2KO mice are on the 129 background (Chen et al., 2015). The disparity between the 129 *Tcra-Tcrd* locus, which features a duplication of the V_α-V_δ array, and the C57BL/6J locus, which contains a triplication of the V_α-V_δ array (Carico and Krangel, 2015), limits the feasibility of direct comparison between these lines. Moreover, the triplication of the C57BL/6J V_α-V_δ array constrains the potential for *Tcra* sequencing of INT1M mice; our previous HTS of the *Tcra* repertoire in C57BL/6J mice was unsuccessful due to the highly repetitive nature of the V_α array. In the future, V_α utilization could be assayed by real-time qPCR. Due to the similarity between the INT1M

V_δ utilization phenotype and the V_δ skewing in INT2M and INT1-2KO mice, we expect that *Tcra* recombination would also be impacted in the INT1M mouse (Carico et al., 2017). The bias toward proximal V_δ use in INT1M mice would likely translate to a reduction in distal V_α utilization in early recombinations, constraining the overall *Tcra* repertoire to recombinations along the major diagonal.

The C57BL/6J locus also limits the potential for analysis of chromatin structure in INT1M mice; alignment of 4C data to this locus would be imperfect due to its highly repetitive nature, as many reads would not be definitively assigned to a particular region. Future analyses could include 3C to determine how the reduction in CTCF binding at INT1 changes interactions between INT1, the *Tcrd* RC, and the rest of the locus. 3C may help to resolve the disparity between the 4C data suggesting that INT1 interacts broadly across the 3' portion of the *Tcra-Tcrd* locus (Chen et al., 2015; Zhao et al., 2016; Carico et al., 2017) and the data herein suggesting that INT1 cooperates with INT2 to form the 5' boundary of the *Tcrd* RC domain. The broad interactions between INT1 and the 3' end of the *Tcra-Tcrd* locus were expected to bring segments in the DJ_δ-proximal upstream region of the locus, such as *Trdv2-2* and *Trdv3*, into proximity of the RC for recombination. However, the discovery that these segments, rather than being excluded from recombination, were used more frequently in INT1M mice (**Figure 16**), suggests that the observed 4C interactome of INT1 is not predictive of its activity. We expect that interactions from the RC to INT3 would be increased in INT1M mice, concomitant with the observed increase in recombinations in the region between INT3 and INT1.

6.6 Concluding remarks

In their totality, the results of this dissertation demonstrate unique mechanisms of recombination at the *Tcra-Tcrd* locus. These findings demonstrate the importance of CTCF-mediated topology of the *Tcra-Tcrd* locus for physiological *Tcrd* recombination. *Tcrd* recombination itself is diversified by chromatin loops. *Tcrd* rearrangements not only impart diversity on the γδT cell repertoire, but also diversify the Tcra repertoire. Particularly, central and distal V_δ utilization are integral in diversifying V_α utilization in early *Tcra* recombinations. These early *Tcra* recombinations provide variable initiation points for ongoing secondary recombinations that proceed along various diagonals to generate the combinatorial pool of the *Tcra* repertoire.

References

Abarrategui, I., and M.S. Krangel. 2006. Regulation of T cell receptor-α gene recombination by transcription. *Nat. Immunol. 2006 710*. 7:1109–1115. doi:10.1038/ni1379.

Abarrategui, I., and M.S. Krangel. 2007. Noncoding transcription controls downstream promoters to regulate T-cell receptor α recombination. *EMBO J.* doi:10.1038/sj.emboj.7601866.

Afshar, R., S. Pierce, D.J. Bolland, A.E. Corcoran, and E.M. Oltz. 2006. Regulation of IgH gene assembly: role of the intronic enhancer and 5'DQ52 region in targeting DHJH recombination. *J. Immunol.* 176:2439–2447. doi:10.4049/JIMMUNOL.176.4.2439.

Agata, Y., T. Katakai, S.-K. Ye, M. Sugai, H. Gonda, T. Honjo, K. Ikuta, and A. Shimizu. 2001. Histone Acetylation Determines the Developmentally Regulated Accessibility for T Cell Receptor γ Gene Recombination. *J. Exp. Med.* 193:873–880. doi:10.1084/JEM.193.7.873.

Alipour, E., and J. Marko. 2012. Self-organization of domain structures by DNA-loop-extruding enzymes. *Nucleic Acids Res.* 40:11202–11212. doi:10.1093/NAR/GKS925.

de Almeida, C.R., R.W. Hendriks, and R. Stadhouders. 2015. Dynamic Control of Long-Range Genomic Interactions at the Immunoglobulin κ Light-Chain Locus. *In* Advances in Immunology. Academic Press Inc. 183–271.

Alt, F.W., and D. Baltimore. 1982. Joining of immunoglobulin heavy chain gene segments: implications from a chromosome with evidence of three D-JH fusions. *Proc. Natl. Acad. Sci. U. S. A.* 79:4118–22. doi:10.1073/PNAS.79.13.4118.

Alt, F.W., G.D. Yancopoulos, T.K. Blackwell, C. Wood, E. Thomas, M. Boss, R. Coffman, N. Rosenberg, S. Tonegawa, and D. Baltimore. 1984. Ordered rearrangement of immunoglobulin heavy chain variable region segments. *EMBO J.* 3:1209–19.

Amin, R.H., and M.S. Schlissel. 2008. Foxo1 directly regulates the transcription of recombination-activating genes during B cell development. *Nat. Immunol.* 9:613–622. doi:10.1038/ni.1612.

Aoki-Ota, M., A. Torkamani, T. Ota, N. Schork, and D. Nemazee. 2012. Skewed primary Igκ repertoire and V-J joining in C57BL/6 mice: implications for recombination

accessibility and receptor editing. *J. Immunol.* 188:2305–15. doi:10.4049/JIMMUNOL.1103484.

Apostolou, I., A. Sarukhan, L. Klein, and H. von Boehmer. 2002. Origin of regulatory T cells with known specificity for antigen. *Nat. Immunol.* 3:756–763. doi:10.1038/ni816.

Aude-Garcia, C., M. Gallagher, P.N. Marche, and E. Jouvin-Marche. 2001. Preferential ADV-AJ association during recombination in the mouse T-cell receptor alpha/delta locus. *Immunogenetics*. 52:224–230. doi:10.1007/S002510000266.

Ba, Z., J. Lou, A.Y. Ye, H.-Q. Dai, E.W. Dring, S.G. Lin, S. Jain, N. Kyritsis, K.-R. Kieffer-Kwon, R. Casellas, and F.W. Alt. 2020. CTCF orchestrates long-range cohesin-driven V(D)J recombinational scanning. *Nature*. 586:305–310. doi:10.1038/s41586-020-2578-0.

Bain, G., W.J. Romanow, K. Albers, W.L. Havran, and C. Murre. 1999. Positive and negative regulation of V(D)J recombination by the E2A proteins. *J. Exp. Med.* 189:289–300. doi:10.1084/jem.189.2.289.

Baker, J.E., D. Cado, and D.H. Raulet. 1998. Developmentally programmed rearrangement of T cell receptor Vgamma genes is controlled by sequences immediately upstream of the Vgamma genes. *Immunity*. 9:159–168. doi:10.1016/S1074-7613(00)80598-1.

Banerji, J., L. Olson, and W. Schaffner. 1983. A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell*. 33:729–740. doi:10.1016/0092-8674(83)90015-6.

Barajas-Mora, E.M., E. Kleiman, J. Xu, E.M. Oltz, C. Murre, and A.J. Feeney Correspondence. 2019. A B-Cell-Specific Enhancer Orchestrates Nuclear Architecture to Generate a Diverse Antigen Receptor Repertoire. *Mol. Cell*. 73:48–60. doi:10.1016/j.molcel.2018.10.013.

Bassing, C.H., F.W. Alt, M.M. Hughes, M. D'Auteuil, T.D. Wehrly, B.B. Woodman, F. Gärtner, J.M. White, L. Davidson, and B.P. Sleckman. 2000. Recombination signal sequences restrict chromosomal V(D)J recombination beyond the 12/23 rule. *Nature*. 405:583–586. doi:10.1038/35014635.

Beck, K., M.M. Peak, T. Ota, D. Nemazee, and C. Murre. 2009. Distinct roles for E12 and E47 in B cell specification and the sequential rearrangement of immunoglobulin light chain loci. *J. Exp. Med.* 206:2271–2284. doi:10.1084/JEM.20090756.

Bell, J.J., and A. Bhandoola. 2008. The earliest thymic progenitors for T cells possess myeloid lineage potential. *Nature*. 452:764–767. doi:10.1038/nature06840.

Bettridge, J., C.H. Na, A. Pandey, and S. Desiderio. 2017. H3K4me3 induces allosteric conformational changes in the DNA-binding and catalytic regions of the V(D)J recombinase. *Proc. Natl. Acad. Sci.* 114:1904–1909. doi:10.1073/PNAS.1615727114.

Bhandoola, A., and A. Sambandam. 2006. From stem cell to T cell: One route or many? *Nat. Rev. Immunol.* 6:117–126. doi:10.1038/nri1778.

Bich-Thuy, L., and C. Queen. 1989. An enhancer associated with the mouse immunoglobulin lambda 1 gene is specific for lambda light chain producing cells. *Nucleic Acids Res.* 17:5307–5322. doi:10.1093/NAR/17.13.5307.

Blanco, B. del, A. García-Mariscal, D.L. Wiest, and C. Hernández-Munain. 2012. Tcra Enhancer Activation by Inducible Transcription Factors Downstream of Pre-TCR Signaling. *J. Immunol.* 188:3278–3293. doi:10.4049/JIMMUNOL.1100271.

Bolland, D.J., H. Koohy, A.L. Wood, L.S. Matheson, F. Krueger, M.J.T. Stubbington, A. Baizan-Edge, P. Chovanec, B.A. Stubbs, K. Tabbada, S.R. Andrews, M. Spivakov, and A.E. Corcoran. 2016. Two Mutually Exclusive Local Chromatin States Drive Efficient V(D)J Recombination. *Cell Rep.* 15:2475–87. doi:10.1016/j.celrep.2016.05.020.

Bolotin, D.A., S. Poslavsky, I. Mitrophanov, M. Shugay, I.Z. Mamedov, E. V. Putintseva, and D.M. Chudakov. 2015. MiXCR: Software for comprehensive adaptive immunity profiling. *Nat. Methods*. 12:380–381. doi:10.1038/nmeth.3364.

Bories, J.C., J. Demengeot, L. Davidson, and F.W. Alt. 1996. Gene-targeted deletion and replacement mutations of the T-cell receptor beta-chain enhancer: the role of enhancer elements in controlling V(D)J recombination accessibility. *Proc. Natl. Acad. Sci. U. S. A.* 93:7871–6. doi:10.1073/PNAS.93.15.7871.

Bosc, N., and M.P. Lefranc. 2003. The mouse (Mus musculus) T cell receptor alpha (TRA) and delta (TRD) variable genes. *Dev. Comp. Immunol.* 27:465–497. doi:10.1016/S0145-305X(03)00027-2.

Le Bourhis, L., E. Martin, I. Péguillet, A. Guihot, N. Froux, M. Coré, E. Lévy, M. Dusseaux, V. Meyssonnier, V. Premel, C. Ngo, B. Riteau, L. Duban, D. Robert, S. Huang, M. Rottman, C. Soudais, and O. Lantz. 2010. Antimicrobial activity of mucosal-associated invariant T cells. *Nat. Immunol.* 11:701–708. doi:10.1038/ni.1890.

Bouvier, G., F. Watrin, M. Naspetti, C. Verthuy, P. Naquet, and P. Ferrier. 1996. Deletion of the mouse T-cell receptor beta gene enhancer blocks alphabeta T-cell development. *Proc. Natl. Acad. Sci. U. S. A.* 93:7877. doi:10.1073/PNAS.93.15.7877.

Brekke, K., and W.T. Garrard. 2004. Assembly and analysis of the mouse immunoglobulin kappa gene sequence. *Immunogenetics*. 56:490–505. doi:10.1007/S00251-004-0659-0.

Brubaker, S.W., K.S. Bonham, I. Zanoni, and J.C. Kagan. 2015. Innate Immune Pattern Recognition: A Cell Biological Perspective. *Annu. Rev. Immunol.* 33:257–290. doi:10.1146/ANNUREV-IMMUNOL-032414-112240.

Buch, T., F. Rieux-Laucat, I. Förster, and K. Rajewsky. 2002. Failure of HY-Specific Thymocytes to Escape Negative Selection by Receptor Editing. *Immunity*. 16:707–718. doi:10.1016/S1074-7613(02)00312-6.

Buenrostro, J.D., B. Wu, H.Y. Chang, and W.J. Greenleaf. 2015. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr. Protoc. Mol. Biol.* 109:21.29.1-21.29.9. doi:10.1002/0471142727.MB2129S109.

Capone, M., R.D. Hockett, and A. Zlotnik. 1998. Kinetics of T cell receptor β, γ, and δ rearrangements during adult thymic development: T cell receptor rearrangements are present in CD44+CD25+ Pro-T thymocytes. *Proc. Natl. Acad. Sci. U. S. A.* 95:12522–12527. doi:10.1073/pnas.95.21.12522.

Carabana, J., E. Ortigoza, and M.S. Krangel. 2005. Regulation of the murine Ddelta2 promoter by upstream stimulatory factor 1, Runx1, and c-Myb. *J. Immunol.* 174:4144–4152. doi:10.4049/JIMMUNOL.174.7.4144.

Carding, S.R., and P.J. Egan. 2002. γδ T cells: functional plasticity and heterogeneity. *Nat. Rev. Immunol.* 2:336–345. doi:10.1038/nri797.

Carico, Z., and M.S. Krangel. 2015. Chromatin Dynamics and the Development of the TCRα and TCRδ Repertoires. *Adv. Immunol.* 128:307–361. doi:10.1016/bs.ai.2015.07.005.

Carico, Z.M., K. Roy Choudhury, B. Zhang, Y. Zhuang, and M.S. Krangel. 2017. Tcrd Rearrangement Redirects a Processive Tcra Recombination Program to Expand the Tcra Repertoire. *Cell Rep.* 19:2157–2173. doi:10.1016/j.celrep.2017.05.045.

Carpenter, A.C., and R. Bosselut. 2010. Decision checkpoints in the thymus. *Nat.*

Immunol. 11:666–673. doi:10.1038/ni.1887.

Carson, S., and G.E. Wu. 1989. A linkage map of the mouse immunoglobulin lambda light chain locus. *Immunogenetics*. 29:173–179. doi:10.1007/BF00373642.

Chakraborty, T., D. Chowdhury, A. Keyes, A. Jani, R. Subrahmanyam, I. Ivanova, and R. Sen. 2007. Repeat Organization and Epigenetic Regulation of the DH-Cμ Domain of the Immunoglobulin Heavy-Chain Gene Locus. *Mol. Cell*. 27:842–850. doi:10.1016/J.MOLCEL.2007.07.010.

Chakraborty, T., T. Perlot, R. Subrahmanyam, A. Jani, P.H. Goff, Y. Zhang, I. Ivanova, F.W. Alt, and R. Sen. 2009. A 220-nucleotide deletion of the intronic enhancer reveals an epigenetic hierarchy in immunoglobulin heavy chain locus activation. *J. Exp. Med.* 206:1019. doi:10.1084/JEM.20081621.

Chan, E.A.W., G. Teng, E. Corbett, K.R. Choudhury, C.H. Bassing, D.G. Schatz, and M.S. Krangel. 2013. Peripheral subnuclear positioning suppresses Tcrb recombination and segregates Tcrb alleles from RAG2. *Proc. Natl. Acad. Sci. U. S. A.* 110:E4628–E4637. doi:10.1073/PNAS.1310846110.

Chang, Y., C.J. Paige, and G.E. Wu. 1992. Enumeration and characterization of DJH structures in mouse fetal liver. *EMBO J.* 11:1891–99.

Chaplin, D.D. 2010. Overview of the Immune Response. *J. Allergy Clin. Immunol.* 125:S3-23. doi:10.1016/J.JACI.2009.12.980.

Chasseval, R. De, and J.-P. De Villartay. 1993. Functional characterization of the promoter for the human germ-line T cell receptor J α (TEA) transcript. *Eur. J. Immunol.* 23:1294–1298. doi:10.1002/EJI.1830230616.

Chen, L., Z. Carico, H.-Y. Shih, and M.S. Krangel. 2015. A discrete chromatin loop in the mouse Tcra-Tcrd locus shapes the TCRδ and TCRα repertoires. *Nat. Immunol.* 16:1085–1093. doi:10.1038/ni.3232.

Chen, L., L. Zhao, F.W. Alt, and M.S. Krangel. 2016. An Ectopic CTCF Binding Element Inhibits Tcrd Rearrangement by Limiting Contact between Vδ and Dδ Gene Segments. *J. Immunol.* 197:3188–97. doi:10.4049/jimmunol.1601124.

Chen, S., T.R. Luperchio, X. Wong, E.B. Doan, A.T. Byrd, K. Roy Choudhury, K.L. Reddy, and M.S. Krangel. 2018. A Lamina-Associated Domain Border Governs Nuclear Lamina Interactions, Transcription, and Recombination of the Tcrb Locus. *Cell Rep.* 25:1729-1740.e6. doi:10.1016/J.CELREP.2018.10.052.

Chien, Y., M. Iwashima, D. Wettstein, K. Kaplan, J. Elliott, W. Born, and M. Davis. 1987. T-cell receptor delta gene rearrangements in early thymocytes. *Nature*. 330:722–727. doi:10.1038/330722A0.

Choi, N.M., S. Loguercio, J. Verma-Gaur, S.C. Degner, A. Torkamani, A.I. Su, E.M. Oltz, M. Artyomov, and A.J. Feeney. 2013. Deep Sequencing of the Murine Igh Repertoire Reveals Complex Regulation of Nonrandom V Gene Rearrangement Frequencies. *J. Immunol.* 191:2393–2402. doi:10.4049/JIMMUNOL.1301279.

Ciofani, M., and J.C. Zúñiga-Pflücker. 2010. Determining γ δ versus α β T cell development. *Nat. Rev. Immunol.* 10:657–663. doi:10.1038/nri2820.

Ciosk, R., M. Shirayama, A. Shevchenko, T. Tanaka, A. Toth, A. Shevchenko, and K. Nasmyth. 2000. Cohesin's Binding to Chromosomes Depends on a Separate Complex Consisting of Scc2 and Scc4 Proteins. *Mol. Cell*. 5:243–254. doi:10.1016/S1097-2765(00)80420-7.

Cobb, R.M., K.J. Oestreich, O.A. Osipovich, and E.M. Oltz. 2006. Accessibility Control of V(D)J Recombination. *Adv. Immunol.* 91:45–109. doi:10.1016/S0065-2776(06)91002-5.

Cong, L., F.A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P.D. Hsu, X. Wu, W. Jiang, L.A. Marraffini, F. Zhang, M.H. Porteus, D. Baltimore, J.C. Miller, J.D. Sander, A.J. Wood, M. Christian, F. Zhang, J.C. Miller, D. Reyon, J. Boch, M.J. Moscou, A.J. Bogdanove, B.L. Stoddard, M. Jinek, G. Gasiunas, R. Barrangou, P. Horvath, V. Siksnys, J.E. Garneau, H. Deveau, J.E. Garneau, S. Moineau, P. Horvath, R. Barrangou, K.S. Makarova, D. Bhaya, M. Davison, R. Barrangou, E. Deltcheva, R. Sapranauskas, A.H. Magadán, M.E. Dupuis, M. Villion, S. Moineau, H. Deveau, F.J. Mojica, C. Díez-Villaseñor, J. García-Martínez, C. Almendros, M. Jinek, J.A. Doudna, C.D. Malone, G.J. Hannon, G. Meister, T. Tuschl, M.T. Certo, P. Mali, P.A. Carr, G.M. Church, R. Barrangou, P. Horvath, S.J. Brouns, and D.Y. Guschin. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science*. 339:819–23. doi:10.1126/science.1231143.

Corbett, A., S. Eckle, R. Birkinshaw, L. Liu, O. Patel, J. Mahony, Z. Chen, R. Reantragoon, B. Meehan, H. Cao, N. Williamson, R. Strugnell, D. Van Sinderen, J. Mak, D. Fairlie, L. Kjer-Nielsen, J. Rossjohn, and J. McCluskey. 2014. T-cell activation by transitory neo-antigens derived from distinct microbial pathways. *Nature*. 509:361–365. doi:10.1038/NATURE13160.

Cotner, T., and H.N. Eisen. 1978. The natural abundance of lambda2-light chains in inbred mice. *J. Exp. Med.* 148:1388–1399. doi:10.1084/JEM.148.5.1388.

Dai, H.-Q., H. Hu, J. Lou, A.Y. Ye, Z. Ba, X. Zhang, Y. Zhang, L. Zhao, H.S. Yoon, A. Chapdelaine-Williams, N. Kyritsis, H. Chen, K. Johnson, S. Lin, A. Conte, R. Casellas, S. Lee, and F.W. Alt. 2021. Loop extrusion mediates physiological Igh locus contraction for RAG scanning. *Nature*. 590:338–343. doi:10.1038/S41586-020-03121-7.

Davidson, I.F., B. Bauer, D. Goetz, W. Tang, G. Wutz, and J.-M. Peters. 2019. DNA loop extrusion by human cohesin. *Science.* 366:1338–1345. doi:10.1126/SCIENCE.AAZ3418.

Degner, S.C., J. Verma-Gaur, T.P. Wong, C. Bossen, G.M. Iverson, A. Torkamani, C. Vettermann, Y.C. Lin, Z. Ju, D. Schulz, C.S. Murre, B.K. Birshtein, N.J. Schork, M.S. Schlissel, R. Riblet, C. Murre, and A.J. Feeney. 2011. CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the Igh locus and antisense transcription in pro-B cells. *Proc. Natl. Acad. Sci. U. S. A.* 108:9566–9571. doi:10.1073/PNAS.1019391108.

Dekker, J., K. Rippe, M. Dekker, and N. Kleckner. 2002. Capturing Chromosome Conformation. *Science.* 295:1306–1311. doi:10.1126/SCIENCE.1067799.

Denker, A., and W. De Laat. 2016. The second decade of 3C technologies: Detailed insights into nuclear organization. *Genes Dev.* 30:1357–1382. doi:10.1101/gad.281964.116.

Derbinski, J., A. Schulte, B. Kyewski, and L. Klein. 2001. Promiscuous gene expression in medullary thymic epithelial cells mirrors the peripheral self. *Nat. Immunol.* 2:1032–1039. doi:10.1038/ni723.

Dixon, J.R., S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J.S. Liu, and B. Ren. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. 485:376–380. doi:10.1038/nature11082.

Donskoy, E., and I. Goldschneider. 1992. Thymocytopoiesis is maintained by blood-borne precursors throughout postnatal life. A study in parabiotic mice. *J. Immunol.* 148:1604–12.

Eccles, S., N. Sarner, M. Vidal, A. Cox, and F. Grosveld. 1990. Enhancer sequences located 3’ of the mouse immunoglobulin lambda locus specify high-level expression of an immunoglobulin lambda gene in B cells of transgenic mice. *New Biol.* 2:801–811.

Elliott, J., E. Rock, P. Patten, M. Davis, and Y. Chien. 1988. The adult T-cell receptor delta-chain is diverse and distinct from that of fetal thymocytes. *Nature*. 331:627–

631. doi:10.1038/331627A0.

Engel, I., C. Johns, G. Bain, R.R. Rivera, and C. Murre. 2001. Early Thymocyte Development Is Regulated by Modulation of E2a Protein Activity. *J. Exp. Med.* 194:733–46. doi:10.1084/JEM.194.6.733.

Espinoza, C.R., and A.J. Feeney. 2005. The Extent of Histone Acetylation Correlates with the Differential Rearrangement Frequency of Individual VH Genes in Pro-B Cells. *J. Immunol.* 175:6668–6675. doi:10.4049/JIMMUNOL.175.10.6668.

Feddersen, R.M., and B.G. Van Ness. 1985. Double recombination of a single immunoglobulin κ-chain allele: Implications for the mechanism of rearrangement. *Proc. Natl. Acad. Sci. U. S. A.* 82:4793–4797. doi:10.1073/pnas.82.14.4793.

Feeney, A.J. 2009. Genetic and epigenetic control of V gene rearrangement frequency. *Adv. Exp. Med. Biol.* 650:73–81. doi:10.1007/978-1-4419-0296-2_6.

Fehling, H.J., A. Krotkova, C. Saint-Ruf, and H. von Boehmer. 1995. Crucial role of the pre-T-cell receptor α gene in development of αβ but not γδ T cells. *Nature*. 375:795–798. doi:10.1038/375795a0.

Fitzsimmons, S.P., B.T. Rotz, and M.A. Shapiro. 1998. Asymmetric contribution to Ig repertoire diversity by V kappa exons: differences in the utilization of V kappa 10 exons. *J. Immunol.* 1:2290–300.

Flajnik, M.F. 2018. A cold-blooded view of adaptive immunity. *Nat. Rev. Immunol.* 18:438–453. doi:10.1038/s41577-018-0003-9.

Fondell, J.D., and K.B. Marcu. 1992. Transcription of germ line V alpha segments correlates with ongoing T-cell receptor alpha-chain rearrangement. *Mol. Cell. Biol.* 12:1480–1489. doi:10.1128/mcb.12.4.1480.

Fudenberg, G., M. Imakaev, C. Lu, A. Goloborodko, N. Abdennur, and L.A. Mirny. 2016. Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep.* 15:2038–2049. doi:10.1016/J.CELREP.2016.04.085.

Fuxa, M., J. Skok, A. Souabni, G. Salvagiotto, E. Roldan, and M. Busslinger. 2004. Pax5 induces V-to-DJ rearrangements and locus contraction of the immunoglobulin heavy-chain gene. *Genes Dev.* 18:411–422. doi:10.1101/GAD.291504.

Genolet, R., B.J. Stevenson, L. Farinelli, M. Østerås, and I.F. Luescher. 2012. Highly diverse TCRα chain repertoire of pre-immune $CD8^+$ T cells reveals new insights in gene recombination. *EMBO J.* 31:1666–1678. doi:10.1038/emboj.2012.48.

Van Gent, D.C., K. Hiom, T.T. Paull, and M. Gellert. 1997. Stimulation of V(D)J cleavage by high mobility group proteins. *EMBO J.* 16:2665–2670. doi:10.1093/emboj/16.10.2665.

Van Gent, D.C., D.A. Ramsden, and M. Gellert. 1996. The RAG1 and RAG2 proteins establish the 12/23 rule in V(D)J recombination. *Cell*. 85:107–113. doi:10.1016/S0092-8674(00)81086-7.

Gilfillan, S., A. Dierich, M. Lemeur, C. Benoist, and D. Mathis. 1993. Mice lacking TdT: Mature animals with an immature lymphocyte repertoire. *Science.* 261:1175–1178. doi:10.1126/science.8356452.

Gill, L.L., K. Karjalainen, and D. Zaninetta. 1991. A transcriptional enhancer of the mouse T cell receptor δ gene locus. *Eur. J. Immunol.* 21:807–810. doi:10.1002/EJI.1830210339.

Gillies, S.D., S.L. Morrison, V.T. Oi, and S. Tonegawa. 1983. A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell*. 33:717–728. doi:10.1016/0092-8674(83)90014-4.

Glusman, G., L. Rowen, I. Lee, C. Boysen, J.C. Roach, A.F.A. Smit, K. Wang, B.F. Koop, and L. Hood. 2001. Comparative genomics of the human and mouse T cell receptor loci. *Immunity*. 15:337–349. doi:10.1016/S1074-7613(01)00200-X.

Goldman, J.P., D.M. Spencer, and D.H. Raulet. 1993. Ordered rearrangement of variable region genes of the T cell receptor gamma locus correlates with transcription of the unrearranged genes. *J. Exp. Med.* 177:729–739. doi:10.1084/JEM.177.3.729.

Gopalakrishnan, S., K. Majumder, A. Predeus, Y. Huang, O.I. Koues, J. Verma-Gaur, S. Loguercio, A.I. Su, A.J. Feeney, M.N. Artyomov, and E.M. Oltz. 2013. Unifying model for molecular determinants of the preselection Vβ repertoire. *Proc. Natl. Acad. Sci.* 110:E3206–E3215. doi:10.1073/PNAS.1304048110.

Gorman, J.R., and F.W. Alt. 1998. Regulation of immunoglobulin light chain isotype expression. *Adv. Immunol.* 69:113–181. doi:10.1016/S0065-2776(08)60607-0.

Grawunder, U., T.M.J. Leu, D.G. Schatz, A. Werner, A.G. Rolink, F. Melchers, and T.H. Winkler. 1995. Down-regulation of RAG1 and RAG2 gene expression in PreB cells after functional immunoglobulin heavy chain rearrangement. *Immunity*. 3:601–608. doi:10.1016/1074-7613(95)90131-0.

Guelen, L., L. Pagie, E. Brasset, W. Meuleman, M.B. Faza, W. Talhout, B.H. Eussen, A. de Klein, L. Wessels, W. de Laat, and B. van Steensel. 2008. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*. 453:948–951. doi:10.1038/nature06947.

Guerreiro, I., and J. Kind. 2019. Spatial chromatin organization and gene regulation at the nuclear lamina. *Curr. Opin. Genet. Dev.* 55:19–25. doi:10.1016/J.GDE.2019.04.008.

Guo, C., T. Gerasimova, H. Hao, I. Ivanova, T. Chakraborty, R. Selimyan, E.M. Oltz, and R. Sen. 2011a. Two Forms of Loops Generate the Chromatin Conformation of the Immunoglobulin Heavy-Chain Gene Locus. *Cell*. 147:332–343. doi:10.1016/J.CELL.2011.08.049.

Guo, C., H.S. Yoon, A. Franklin, S. Jain, A. Ebert, H.-L. Cheng, E. Hansen, O. Despo, C. Bossen, C. Vettermann, J.G. Bates, N. Richards, D. Myers, H. Patel, M. Gallagher, M.S. Schlissel, C. Murre, M. Busslinger, C.C. Giallourakis, and F.W. Alt. 2011b. CTCF-binding elements mediate control of V(D)J recombination. *Nature*. 477:424–30. doi:10.1038/nature10495.

Guo, J., A. Hawwari, H. Li, Z. Sun, S.K. Mahanta, D.R. Littman, M.S. Krangel, and Y.W. He. 2002. Regulation of the TCRα repertoire by the survival window of CD4+CD8+ thymocytes. *Nat. Immunol.* 3:469–76. doi:10.1038/ni791.

Guo, Y., Q. Xu, D. Canzio, J. Shou, J. Li, D.U. Gorkin, I. Jung, H. Wu, Y. Zhai, Y. Tang, Y. Lu, Y. Wu, Z. Jia, W. Li, M.Q. Zhang, B. Ren, A.R. Krainer, T. Maniatis, and Q. Wu. 2015. CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell*. 162:900–910. doi:10.1016/j.cell.2015.07.038.

Guo, Z., H. Li, M. Han, T. Xu, X. Wu, and Y. Zhuang. 2011c. Modeling Sjögren's Syndrome with Id3 Conditional Knockout Mice. *Immunol. Lett.* 135:34–42. doi:10.1016/J.IMLET.2010.09.009.

Hagman, J., C.M. Rudin, D. Haasch, D. Chaplin, and U. Storb. 1990. A novel enhancer in the immunoglobulin lambda locus is duplicated and functionally independent of NF kappa B. *Genes Dev.* 4:978–992. doi:10.1101/GAD.4.6.978.

Hansen, A.S. 2020. CTCF as a boundary factor for cohesin-mediated loop extrusion: evidence for a multi-step mechanism. *Nucleus*. 11:132–148. doi:10.1080/19491034.2020.1782024.

Hansen, A.S., C. Cattoglio, X. Darzacq, and R. Tjian. 2018. Recent evidence that TADs and chromatin loops are dynamic structures. *Nucleus*. 9:20–32. doi:10.1080/19491034.2017.1389365.

Hao, B., and M.S. Krangel. 2011. Long-Distance Regulation of Fetal Vδ Gene Segment TRDV4 by the Tcrd Enhancer. *J. Immunol.* 187:2484–2491. doi:10.4049/JIMMUNOL.1100468.

Hao, B., A.K. Naik, A. Watanabe, H. Tanaka, L. Chen, H.W. Richards, M. Kondo, I. Taniuchi, Y. Kohwi, T. Kohwi-Shigematsu, and M.S. Krangel. 2015. An anti-silencer– and SATB1-dependent chromatin hub regulates Rag1 and Rag2 gene expression during thymocyte development. *J. Exp. Med.* 212:809–824. doi:10.1084/JEM.20142207.

Hardy, R.R., and K. Hayakawa. 2001. B cell development pathways. *Annu. Rev. Immunol.* 19:595–621. doi:10.1146/annurev.immunol.19.1.595.

Havran, W., Y. Chien, and J. Allison. 1991. Recognition of self antigens by skin-derived T cells with invariant gamma delta antigen receptors. *Science.* 252:1430–1432. doi:10.1126/SCIENCE.1828619.

Havran, W.L., and J.P. Allison. 1988. Developmentally ordered appearance of thymocytes expressing different T-cell antigen receptors. *Nature*. 335:443–445. doi:10.1038/335443a0.

Havran, W.L., and J.P. Allison. 1990. Origin of Thy-1+ dendritic epidermal cells of adult mice from fetal thymic precursors. *Nature*. 344:68–70. doi:10.1038/344068a0.

Hawwari, A., C. Bock, and M.S. Krangel. 2005. Regulation of T cell receptor α gene assembly by a complex hierarchy of germline Jα promoters. *Nat. Immunol.* 6:481–489. doi:10.1038/ni1189.

Hawwari, A., and M.S. Krangel. 2005. Regulation of TCR δ and α repertoires by local and long-distance control of variable gene segment chromatin structure. *J. Exp. Med.* 202:467–472. doi:10.1084/JEM.20050680.

Hawwari, A., and M.S. Krangel. 2007. Role for rearranged variable gene segments in directing secondary T cell receptor α recombination. *Proc. Natl. Acad. Sci. U. S. A.* 104:903–907. doi:10.1073/pnas.0608248104.

Helmink, B.A., and B.P. Sleckman. 2012. The response to and repair of RAG-mediated DNA double-strand breaks. *Annu. Rev. Immunol.* 30:175–202. doi:10.1146/annurev-immunol-030409-101320.

Heng, T.S.P., M.W. Painter, K. Elpek, V. Lukacs-Kornek, N. Mauermann, S.J. Turley, D. Koller, F.S. Kim, A.J. Wagers, N. Asinovski, S. Davis, M. Fassett, M. Feuerer,

D.H.D. Gray, S. Haxhinasto, J.A. Hill, G. Hyatt, C. Laplace, K. Leatherbee, D. Mathis, C. Benoist, R. Jianu, D.H. Laidlaw, J.A. Best, J. Knell, A.W. Goldrath, J. Jarjoura, J.C. Sun, Y. Zhu, L.L. Lanier, A. Ergun, Z. Li, J.J. Collins, S.A. Shinton, R.R. Hardy, R. Friedline, K. Sylvia, and J. Kang. 2008. The immunological genome project: Networks of gene expression in immune cells. *Nat. Immunol.* 9:1091–1094. doi:10.1038/ni1008-1091.

Hernández-Munain, C., B.P. Sleckman, and M.S. Krangel. 1999. A developmental switch from TCRδ enhancer to TCRα enhancer function during thymocyte maturation. *Immunity*. 10:723–733. doi:10.1016/S1074-7613(00)80071-0.

Hill, L., A. Ebert, M. Jaritz, G. Wutz, K. Nagasaka, H. Tagoh, D. Kostanova-Poliakova, K. Schindler, Q. Sun, P. Bönelt, M. Fischer, J.M. Peters, and M. Busslinger. 2020. Wapl repression by Pax5 promotes V gene recombination by Igh loop extrusion. *Nature*. 584:142–147. doi:10.1038/s41586-020-2454-y.

Hiom, K., and M. Gellert. 1998. Assembly of a 12/23 paired signal complex: A critical control point in V(D)J recombination. *Mol. Cell*. 1:1011–1019. doi:10.1016/S1097-2765(00)80101-X.

Ho, I.C., L.H. Yang, G. Morle, and J.M. Leiden. 1989. A T-cell-specific transcriptional enhancer element 3' of C alpha in the human T-cell receptor alpha locus. *Proc. Natl. Acad. Sci. U. S. A.* 86:6714–6718. doi:10.1073/PNAS.86.17.6714.

Hsu, L.Y., J. Lauring, H.E. Liang, S. Greenbaum, D. Cado, Y. Zhuang, and M.S. Schlissel. 2003. A Conserved Transcriptional Enhancer Regulates RAG Gene Expression in Developing B Cells. *Immunity*. 19:105–117. doi:10.1016/S1074-7613(03)00181-X.

Hu, J., R.M. Meyers, J. Dong, R. Panchakshari, F.W. Alt, and R.L. Frock. 2016. Detecting DNA double-stranded breaks in mammalian genomes by linear amplification-mediated high-throughput genome-wide translocation sequencing. *Nat. Protoc.* 11:853–871. doi:10.1038/NPROT.2016.043.

Hu, J., Y. Zhang, L. Zhao, R.L. Frock, Z. Du, R.M. Meyers, F.L. Meng, D.G. Schatz, and F.W. Alt. 2015. Chromosomal Loop Domains Direct the Recombination of Antigen Receptor Genes. *Cell*. 163:947–959. doi:10.1016/j.cell.2015.10.016.

Huang, C.-Y., and O. Kanagawa. 2001. Ordered and Coordinated Rearrangement of the TCR α Locus: Role of Secondary Rearrangement in Thymic Selection. *J. Immunol.* 166:2597–2601. doi:10.4049/jimmunol.166.4.2597.

Ikuta, K., T. Kina, I. MacNeil, N. Uchida, B. Peault, Y. hsiu Chien, and I.L. Weissman.

1990. A developmental switch in thymic lymphocyte maturation potential occurs at the level of hematopoietic stem cells. *Cell*. 62:863–874. doi:10.1016/0092-8674(90)90262-D.

Itohara, S., A.G. Farr, J.J. Lafaille, M. Bonneville, Y. Takagaki, W. Haas, and S. Tonegawa. 1990. Homing of a γδ thymocyte subset with homogeneous T-cell receptors to mucosal epithelia. *Nature*. 343:754–757. doi:10.1038/343754a0.

Itohara, S., P. Mombaerts, J.J. Lafaille, J. Iacomini, A. Nelson, A.R. Clarke, M.L. Hooper, A.G. Farr, and S. Tonegawa. 1993. T cell receptor delta gene mutant mice: independent generation of alpha beta T cells and programmed rearrangements of gamma delta TCR genes. *Cell*. 72:337–348. doi:10.1016/0092-8674(93)90112-4.

Itoi, M., H. Kawamoto, Y. Katsura, and T. Amagai. 2001. Two distinct steps of immigration of hematopoietic progenitors into the early thymus anlage. *Int. Immunol.* 13:1203–1211. doi:10.1093/intimm/13.9.1203.

Iwasaki, A., and R. Medzhitov. 2015. Control of adaptive immunity by the innate immune system. *Nat. Immunol.* 16:343–353. doi:10.1038/ni.3123.

Jain, S., Z. Ba, Y. Zhang, H.Q. Dai, and F.W. Alt. 2018. CTCF-Binding Elements Mediate Accessibility of RAG Substrates During Chromatin Scanning. *Cell*. 174:102–116. doi:10.1016/J.CELL.2018.04.035.

Janeway, C.A. 1989. Approaching the asymptote? Evolution and revolution in immunology. *In* Cold Spring Harbor Symposia on Quantitative Biology. Cold Spring Harb Symp Quant Biol. 1–13.

Jhunjhunwala, S., M.C. van Zelm, M.M. Peak, S. Cutchin, R. Riblet, J.J.M. van Dongen, F.G. Grosveld, T.A. Knoch, and C. Murre. 2008. The 3D Structure of the Immunoglobulin Heavy-Chain Locus: Implications for Long-Range Genomic Interactions. *Cell*. 133:265–279. doi:10.1016/J.CELL.2008.03.024.

Jhunjhunwala, S., M.C. van Zelm, M.M. Peak, and C. Murre. 2009. Chromatin Architecture and the Generation of Antigen Receptor Diversity. *Cell*. 138:435–448. doi:10.1016/j.cell.2009.07.016.

Ji, Y., W. Resch, E. Corbett, A. Yamane, R. Casellas, and D.G. Schatz. 2010. The in vivo pattern of binding of RAG1 and RAG2 to antigen receptor loci. *Cell*. 141:419–431. doi:10.1016/j.cell.2010.03.010.

Jordan, M.S., A. Boesteanu, A.J. Reed, A.L. Petrone, A.E. Holenbeck, M.A. Lerman, A. Naji, and A.J. Caton. 2001. Thymic selection of CD4+CD25+ regulatory T cells

induced by an agonist self-peptide. *Nat. Immunol.* 2:301–306. doi:10.1038/86302.

Jung, D., and F.W. Alt. 2004. Unraveling V(D)J Recombination: Insights into Gene Regulation. *Cell*. 116:299–311. doi:10.1016/S0092-8674(04)00039-X.

Jung, D., C. Giallourakis, R. Mostoslavsky, and F.W. Alt. 2006. Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. *Annu. Rev. Immunol.* 24:541–570. doi:10.1146/annurev.immunol.23.021704.115830.

Kagey, M.H., J.J. Newman, S. Bilodeau, Y. Zhan, D.A. Orlando, N.L. van Berkum, C.C. Ebmeier, J. Goossens, P.B. Rahl, S.S. Levine, D.J. Taatjes, J. Dekker, and R.A. Young. 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature*. 467:430–435. doi:10.1038/nature09380.

Kappes, D.J., C.P. Browne, and S. Tonegawa. 1991. Identification of a T-cell-specific enhancer at the locus encoding T-cell antigen receptor gamma chain. *Proc. Natl. Acad. Sci. U. S. A.* 88:2204–2208. doi:10.1073/PNAS.88.6.2204.

Karasuyama, H., A. Kudo, and F. Melchers. 1990. The proteins encoded by the VpreB and λ5 Pre-B cell-specific genes can associate with each other and with μ heavy chain. *J. Exp. Med.* 172:969–972. doi:10.1084/jem.172.3.969.

Karasuyama, H., A. Rolink, and F. Melchers. 1993. A complex of glycoproteins is associated with VpreB/λs Surrogate Light Chain on the Surface of μ Heavy Chain-negative Early Precursor B Cell Lines. *J. Exp. Med.* 178:469–478. doi:10.1084/jem.178.2.469.

Karasuyama, H., A. Rolink, and F. Melchers. 1996. Surrogate light chain in B cell development. *Adv. Immunol.* 63:1–41. doi:10.1016/s0065-2776(08)60853-6.

Kawasaki, K., S. Minoshima, E. Nakato, K. Shibuya, A. Shintani, J.L. Schmeits, J. Wang, and N. Shimizu. 1997. One-megabase sequence analysis of the human immunoglobulin lambda gene locus. *Genome Res.* 7:250–261. doi:10.1101/GR.7.3.250.

Kim, M.S., W. Chuenchor, X. Chen, Y. Cui, X. Zhang, Z.H. Zhou, M. Gellert, and W. Yang. 2018. Cracking the DNA Code for V(D)J Recombination. *Mol. Cell*. 70:358-370.e4. doi:10.1016/J.MOLCEL.2018.03.008.

Kim, M.S., M. Lapkouski, W. Yang, and M. Gellert. 2015. Crystal structure of the V(D)J recombinase RAG1-RAG2. *Nature*. 518:507–511. doi:10.1038/nature14174.

Kim, Y., Z. Shi, H. Zhang, I.J. Finkelstein, and H. Yu. 2019. Human cohesin compacts

DNA by loop extrusion. *Science.* 366:1345–1349. doi:10.1126/SCIENCE.AAZ4475.

Kleiman, E., J. Xu, and A.J. Feeney. 2018. Cutting Edge: Proper Orientation of CTCF Sites in Cer Is Required for Normal Jκ-Distal and Jκ-Proximal Vκ Gene Usage. *J. Immunol.* 201:1633–1638. doi:10.4049/jimmunol.1800785.

Komori, T., A. Okada, V. Stewart, and F.W. Alt. 1993. Lack of N regions in antigen receptor variable region genes of TdT-deficient lymphocytes. *Science.* 261:1171–1175. doi:10.1126/science.8356451.

Kosak, S.T., J.A. Skok, K.L. Medina, R. Riblet, M.M. Le Beau, A.G. Fisher, and H. Singh. 2002. Subnuclear Compartmentalization of Immunoglobulin Loci During Lymphocyte Development. *Science.* 296:158–162. doi:10.1126/SCIENCE.1068768.

Kozich, J.J., S.L. Westcott, N.T. Baxter, S.K. Highlander, and P.D. Schloss. 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the miseq illumina sequencing platform. *Appl. Environ. Microbiol.* 79:5112–5120. doi:10.1128/AEM.01043-13.

Krangel, M.S. 2009. Mechanics of T cell receptor gene rearrangement. *Curr. Opin. Immunol.* 21:133–139. doi:10.1016/j.coi.2009.03.009.

Krangel, M.S., H. Yssel, C. Brocklehurst, and H. Spits. 1990. A distinct wave of human T cell receptor gamma/delta lymphocytes in the early fetal thymus: evidence for controlled gene rearrangement and cytokine production. *J. Exp. Med.* 172:847–859. doi:10.1084/JEM.172.3.847.

Kranz, D., H. Saito, C. Disteche, K. Swisshelm, D. Pravtcheva, F. Ruddle, H. Eisen, and S. Tonegawa. 1985. Chromosomal locations of the murine T-cell receptor alpha-chain gene and the T-cell gamma gene. *Science*. 227:941–945. doi:10.1126/SCIENCE.3918347.

Kreslavsky, T., A.I. Garbe, A. Krueger, and H. Von Boehmer. 2008. T cell receptor-instructed -αβ versus γδ lineage commitment revealed by single-cell analysis. *J. Exp. Med.* 205:1173–1186. doi:10.1084/jem.20072425.

Kuo, T.C., and M.S. Schlissel. 2009. Mechanisms controlling expression of the RAG locus during lymphocyte development. *Curr. Opin. Immunol.* 21:173–178. doi:10.1016/j.coi.2009.03.008.

Küppers, R., and R. Dalla-Favera. 2001. Mechanisms of chromosomal translocations in B cell lymphomas. *Oncogene*. 20:5580–5594. doi:10.1038/sj.onc.1204640.

Kwon, J., K. Morshead, J. Guyon, R. Kingston, and M. Oettinger. 2000. Histone acetylation and hSWI/SNF remodeling act in concert to stimulate V(D)J cleavage of nucleosomal DNA. *Mol. Cell*. 6:1037–1048. doi:10.1016/S1097-2765(00)00102-7.

Landau, N.R., D.G. Schatz, M. Rosa, and D. Baltimore. 1987. Increased frequency of N-region insertion in a murine pre-B-cell line infected with a terminal deoxynucleotidyl transferase retroviral expression vector. *Mol. Cell. Biol.* 7:3237–3243. doi:10.1128/MCB.7.9.3237-3243.1987.

Lauzurica, P., and M.S. Krangel. 1994. Temporal and lineage-specific control of t cell receptor α/δ Gene rearrangement by t cell receptor α And δ Enhancers. *J. Exp. Med.* 179:1913–1921. doi:10.1084/jem.179.6.1913.

Lee, K.D., and C.H. Bassing. 2020. Two Successive Inversional Vβ Rearrangements on a Single Tcrb Allele Can Contribute to the TCRβ Repertoire. *J. Immunol.* 204:78–86. doi:10.4049/JIMMUNOL.1901105.

Lefranc, M.P., V. Giudicelli, P. Duroux, J. Jabado-Michaloud, G. Folch, S. Aouinti, E. Carillon, H. Duvergey, A. Houles, T. Paysan-Lafosse, S. Hadi-Saljoqi, S. Sasorith, G. Lefranc, and S. Kossida. 2015. IMGT R, the international ImMunoGeneTics information system R 25 years on. *Nucleic Acids Res.* 43:D413–D422. doi:10.1093/nar/gku1056.

Lennon, G.G., and R.P. Perry. 1985. Cμ-containing transcripts initiate heterogeneously within the IgH enhancer region and contain a novel 5′-nontranslatable exon. *Nature*. 318:475–478. doi:10.1038/318475a0.

Li, Z., D.I. Dordai, J. Lee, and S. Desiderio. 1996. A Conserved Degradation Signal Regulates RAG-2 Accumulation during Cell Division and Links V(D)J Recombination to the Cell Cycle. *Immunity*. 5:575–589. doi:10.1016/S1074-7613(00)80272-1.

Lieberman-Aiden, E., N.L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B.R. Lajoie, P.J. Sabo, M.O. Dorschner, R. Sandstrom, B. Bernstein, M.A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L.A. Mirny, E.S. Lander, and J. Dekker. 2009. Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science.* 326:289–293. doi:10.1126/SCIENCE.1181369.

Lin, S.G., Z. Ba, F.W. Alt, and Y. Zhang. 2018. RAG Chromatin Scanning During V(D)J Recombination and Chromatin Loop Extrusion are Related Processes. *In* Advances in Immunology. F.W. Alt, editor. Academic Press Inc. 139:93–135.

Lin, W.C., and S. Desiderio. 1994. Cell cycle regulation of V(D)J recombination-activating protein RAG-2. *Proc. Natl. Acad. Sci. U. S. A.* 91:2733–2737. doi:10.1073/PNAS.91.7.2733.

Lin, Y.C., C. Benner, R. Mansson, S. Heinz, K. Miyazaki, M. Miyazaki, V. Chandra, C. Bossen, C.K. Glass, and C. Murre. 2012. Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. *Nat. Immunol.* 13:1196–1204. doi:10.1038/ni.2432.

Lind, E.F., S.E. Prockop, H.E. Porritt, and H.T. Petrie. 2001. Mapping precursor movement through the postnatal thymus reveals specific microenvironments supporting defined stages of early lymphoid development. *J. Exp. Med.* 194:127–134. doi:10.1084/jem.194.2.127.

Liu, Y., R. Subrahmanyam, T. Chakraborty, R. Sen, and S. Desiderio. 2007. A Plant Homeodomain in Rag-2 that Binds Hypermethylated Lysine 4 of Histone H3 Is Necessary for Efficient Antigen-Receptor-Gene Rearrangement. *Immunity*. 27:561–571. doi:10.1016/J.IMMUNI.2007.09.005.

Liu, Z., P. Widlak, Y. Zou, F. Xiao, M. Oh, S. Li, M.Y. Chang, J. Shay, and W.T. Garrard. 2006. A recombination silencer that specifies heterochromatin positioning and ikaros association in the immunoglobulin kappa locus. *Immunity*. 24:405–415. doi:10.1016/J.IMMUNI.2006.02.001.

Liu, Z.M., J.B. George-Raizen, S. Li, K.C. Meyers, M.Y. Chang, and W.T. Garrard. 2002. Chromatin Structural Analyses of the Mouse Igκ Gene Locus Reveal New Hypersensitive Sites Specifying a Transcriptional Silencer and Enhancer. *J. Biol. Chem.* 277:32640–32649. doi:10.1074/JBC.M204065200.

Livak, F., H.T. Petrie, I.N. Crisps, and D.G. Schatz. 1995. In-frame TCR δ gene rearrangements play a critical role in the αβ/γδ T cell lineage decision. *Immunity*. 2:617–627. doi:10.1016/1074-7613(95)90006-3.

Livák, F., M. Tourigny, D.G. Schatz, and H.T. Petrie. 1999. Characterization of TCR gene rearrangements during adult murine T cell development. *J. Immunol.* 162:2575–80.

Loguercio, S., E.M. Barajas-Mora, H.-Y. Shih, M.S. Krangel, and A.J. Feeney. 2018. Variable Extent of Lineage-Specificity and Developmental Stage-Specificity of Cohesin and CCCTC-Binding Factor Binding Within the Immunoglobulin and T Cell Receptor Loci. *Front. Immunol.* 9:425. doi:10.3389/FIMMU.2018.00425.

Madisen, L., T.A. Zwingman, S.M. Sunkin, S.W. Oh, H.A. Zariwala, H. Gu, L.L. Ng, R.D. Palmiter, M.J. Hawrylycz, A.R. Jones, E.S. Lein, and H. Zeng. 2010. A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* 13:133–40. doi:10.1038/nn.2467.

Majumder, K., O.I. Koues, E.A.W. Chan, K.E. Kyle, J.E. Horowitz, K. Yang-Iott, C.H. Bassing, I. Taniuchi, M.S. Krangel, and E.M. Oltz. 2015. Lineage-specific compaction of Tcrb requires a chromatin barrier to protect the function of a long-range tethering element. *J. Exp. Med.* 212:107–120. doi:10.1084/JEM.20141479.

Malissen, M., J. Trucy, E. Jouvin-Marche, P.A. Cazenave, R. Scollay, and B. Malissen. 1992. Regulation of TCR α and β gene allelic exclusion during T-cell development. *Immunol. Today*. 13:315–322. doi:10.1016/0167-5699(92)90044-8.

Malissen, M., J. Trucy, F. Letourneur, N. Rebai, D.E. Dunn, F.W. Fitch, L. Hood, and B. Malissen. 1988. A T cell clone expresses two T cell receptor α genes but uses one αβ heterodimer for allorecognition and self MHC-restricted antigen recognition. *Cell*. 55:49–59. doi:10.1016/0092-8674(88)90008-6.

Mandal, M., K.M. Hamel, M. Maienschein-Cline, A. Tanaka, G. Teng, J.H. Tuteja, J.J. Bunker, N. Bahroos, J.J. Eppig, D.G. Schatz, and M.R. Clark. 2015. Histone reader BRWD1 targets and restricts recombination to the Igk locus. *Nat. Immunol.* 16:1094–1103. doi:10.1038/ni.3249.

Marculescu, R., T. Le, P. Simon, U. Jaeger, and B. Nadel. 2002. V(D)J-mediated translocations in lymphoid neoplasms: a functional assessment of genomic instability by cryptic sites. *J. Exp. Med.* 195:85–98. doi:10.1084/JEM.20011578.

Masuda, K., K. Kakugawa, T. Nakayama, N. Minato, Y. Katsura, and H. Kawamoto. 2007. T Cell Lineage Determination Precedes the Initiation of TCR β Gene Rearrangement. *J. Immunol.* 179:3699–3706. doi:10.4049/jimmunol.179.6.3699.

Mathieu, N., W. Hempel, S. Spicuglia, C. Verthuy, and P. Ferrier. 2000. Chromatin remodeling by the T cell receptor (TCR)-beta gene enhancer during early T cell development: Implications for the control of TCR-beta locus recombination. *J. Exp. Med.* 192:625–636. doi:10.1084/JEM.192.5.625.

Matthews, A.G.W., A.J. Kuo, S. Ramón-Maiques, S. Han, K.S. Champagne, D. Ivanov, M. Gallardo, D. Carney, P. Cheung, D.N. Ciccone, K.L. Walter, P.J. Utz, Y. Shi, T.G. Kutateladze, W. Yang, O. Gozani, and M.A. Oettinger. 2007. RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. *Nature*. 450:1106–1110. doi:10.1038/nature06431.

Mauvieux, L., I. Villey, and J.P. de Villartay. 2003. TEA regulates local TCR-Jα accessibility through histone acetylation. *Eur. J. Immunol.* 33:2216–2222. doi:10.1002/eji.200323867.

McBlane, J.F., D.C. van Gent, D.A. Ramsden, C. Romeo, C.A. Cuomo, M. Gellert, and M.A. Oettinger. 1995. Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. *Cell*. 83:387–395. doi:10.1016/0092-8674(95)90116-7.

McDougall, S., C. Peterson, and K. Calame. 1988. A transcriptional enhancer 3' of C beta 2 in the T cell receptor beta locus. *Science.* 241:205–208. doi:10.1126/SCIENCE.2968651.

McMillan, R.E., and M.L. Sikes. 2008. Differential Activation of Dual Promoters Alters Dβ2 Germline Transcription during Thymocyte Development. *J. Immunol.* 180:3218–3228. doi:10.4049/JIMMUNOL.180.5.3218.

McMurry, M.T., C. Hernandez-Munain, P. Lauzurica, and M.S. Krangel. 1997. Enhancer control of local accessibility to V(D)J recombinase. *Mol. Cell. Biol.* 17:4553–4561. doi:10.1128/MCB.17.8.4553.

McMurry, M.T., and M.S. Krangel. 2000. A Role for Histone Acetylation in the Developmental Regulation of V(D)J Recombination. *Science.* 287:495–498. doi:10.1126/SCIENCE.287.5452.495.

Medina, C., and J. Teale. 1993. Restricted kappa chain expression in early ontogeny: biased utilization of V kappa exons and preferential V kappa-J kappa recombinations. *J. Exp. Med.* 177:1317–1330. doi:10.1084/JEM.177.5.1317.

Medvedovic, J., A. Ebert, H. Tagoh, I.M. Tamir, T.A. Schwickert, M. Novatchkova, Q. Sun, P.J. Huisin'tVeld, C. Guo, H.S. Yoon, Y. Denizot, S.J.B. Holwerda, W. deLaat, M. Cogné, Y. Shi, F.W. Alt, and M. Busslinger. 2013. Flexible Long-Range Loops in the VH Gene Region of the Igh Locus Facilitate the Generation of a Diverse Antibody Repertoire. *Immunity*. 39:229–244. doi:10.1016/J.IMMUNI.2013.08.011.

Medzhitov, R. 2013. Pattern Recognition Theory and the Launch of Modern Innate Immunity. *J. Immunol.* 191:4473–4474. doi:10.4049/jimmunol.1302427.

Meermeier, E.W., M.J. Harriff, E. Karamooz, and D.M. Lewinsohn. 2018. MAIT cells and microbial immunity. *Immunol. Cell Biol.* 96:607–617. doi:10.1111/IMCB.12022.

Meierovics, A., W.-J.C. Yankelevich, and S.C. Cowley. 2013. MAIT cells are critical for

optimal mucosal immune responses during in vivo pulmonary bacterial infection. *Proc. Natl. Acad. Sci. U. S. A.* 110:E3119-28. doi:10.1073/PNAS.1302799110.

Meyer, K.B., and M.S. Neuberger. 1989. The immunoglobulin kappa locus contains a second, stronger B-cell-specific enhancer which is located downstream of the constant region. *EMBO J.* 8:1959.

Miyazaki, K., and M. Miyazaki. 2021. The Interplay Between Chromatin Architecture and Lineage-Specific Transcription Factors and the Regulation of Rag Gene Expression. *Front. Immunol.* 12:813. doi:10.3389/FIMMU.2021.659761.

Monroe, R.J., F. Chen, R. Ferrini, L. Davidson, and F.W. Alt. 1999a. RAG2 is regulated differentially in B and T cells by elements 5′ of the promoter. *Proc. Natl. Acad. Sci. U. S. A.* 96:12713–12718. doi:10.1073/PNAS.96.22.12713.

Monroe, R.J., K.J. Seidl, F. Gaertner, S. Han, F. Chen, J.A. Sekiguchi, J. Wang, R. Ferrini, L. Davidson, G. Kelsoe, and F.W. Alt. 1999b. RAG2:GFP knockin mice reveal novel aspects of RAG2 expression in primary and peripheral lymphoid tissues. *Immunity*. 11:201–212. doi:10.1016/S1074-7613(00)80095-3.

Monroe, R.J., B.P. Sleckman, B.C. Monroe, B. Khor, S. Claypool, R. Ferrini, L. Davidson, and F.W. Alt. 1999c. Developmental regulation of TCRd locus accessibility and expression by the TCRd enhancer. *Immunity*. 10:503–513. doi:10.1016/S1074-7613(00)80050-3.

Montalbano, A., K. Ogwaro, A. Tang, A. Matthews, M. Larijani, M.A. Oettinger, and A.J. Feeney. 2003. V(D)J recombination frequencies can be profoundly affected by changes in the spacer sequence. *J. Immunol.* 171:5296–5304. doi:10.4049/JIMMUNOL.171.10.5296.

Naik, A.K., A.T. Byrd, A.C.K. Lucander, and M.S. Krangel. 2019. Hierarchical assembly and disassembly of a transcriptionally active RAG locus in CD4 + CD8 + thymocytes. *J. Exp. Med.* 216:231–43. doi:10.1084/jem.20181402.

Naik, A.K., A. Hawwari, and M.S. Krangel. 2015. Specification of vδ and vα usage by tcra/tcrd locus v gene segment promoters. *J. Immunol.* 194:790–4. doi:10.4049/jimmunol.1402423.

Nakahashi, H., K.R.K. Kwon, W. Resch, L. Vian, M. Dose, D. Stavreva, O. Hakim, N. Pruett, S. Nelson, A. Yamane, J. Qian, W. Dubois, S. Welsh, R.D. Phair, B.F. Pugh, V. Lobanenkov, G.L. Hager, and R. Casellas. 2013. A Genome-wide Map of CTCF Multivalency Redefines the CTCF Code. *Cell Rep.* 3:1678–1689. doi:10.1016/j.celrep.2013.04.024.

Nakajima, P.B., J.P. Menetski, D.B. Roth, M. Gellert, and M.J. Bosma. 1995. V-D-J rearrangements at the T cell receptor σ locus in mouse thymocytes of the αβ lineage. *Immunity*. 3:609–621. doi:10.1016/1074-7613(95)90132-9.

Nasmyth, K., and C.H. Haering. 2009. Cohesin: Its Roles and Mechanisms. *Annu. Rev. Genet.* 43:525–558. doi:10.1146/ANNUREV-GENET-102108-134233.

Neuwirth, E. 2014. RColorBrewer: ColorBrewer Palettes. https://cran.r-project.org/web/packages/RColorBrew.

Nikolich-Žugich, J., M.K. Slifka, and I. Messaoudi. 2004. The many important facets of T-cell repertoire diversity. *Nat. Rev. Immunol.* 4:123–132. doi:10.1038/nri1292.

Nora, E.P., B.R. Lajoie, E.G. Schulz, L. Giorgetti, I. Okamoto, N. Servant, T. Piolot, N.L. van Berkum, J. Meisig, J. Sedat, J. Gribnau, E. Barillot, N. Blüthgen, J. Dekker, and E. Heard. 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*. 485:381–385. doi:10.1038/nature11049.

Oettinger, M.A., D.G. Schatz, C. Gorka, and D. Baltimore. 1990. RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science.* 248:1517–1523. doi:10.1126/science.2360047.

Okazaki, K., and H. Sakano. 1988. Thymocyte circular DNA excised from T cell receptor alpha-delta gene complex. *EMBO J.* 7:1669–1674. doi:10.1002/J.1460-2075.1988.TB02994.X.

Palstra, R.-J., B. Tolhuis, E. Splinter, R. Nijmeijer, F. Grosveld, and W. de Laat. 2003. The β-globin nuclear compartment in development and erythroid differentiation. *Nat. Genet.* 35:190–194. doi:10.1038/ng1244.

Parker, M.E., and M. Ciofani. 2020. Regulation of γδ T Cell Effector Diversification in the Thymus. *Front. Immunol.* 11:42. doi:10.3389/FIMMU.2020.00042.

Pereira, P., V. Hermitte, M.-P. Lembezat, L. Boucontet, V. Azuara, and K. Grigoriadou. 2000. Developmentally regulated and lineage-specific rearrangement of T cell receptor Vα/δ gene segments. *Eur. J. Immunol.* 30:1988–1997. doi:10.1002/1521-4141.

Perlmutter, R., J.F. Kearney, S. Chang, and L. Hood. 1985. Developmentally controlled expression of immunoglobulin VH genes. *Science*. 227:1597–1601. doi:10.1126/SCIENCE.3975629.

Perlot, T., F.W. Alt, C.H. Bassing, H. Suh, and E. Pinaud. 2005. Elucidation of IgH intronic enhancer functions via germ-line deletion. *Proc. Natl. Acad. Sci. U. S. A.* 102:14362–14367. doi:10.1073/PNAS.0507090102.

Peters, J.M. 2021. How DNA loop extrusion mediated by cohesin enables V(D)J recombination. *Curr. Opin. Cell Biol.* 70:75–83. doi:10.1016/J.CEB.2020.11.007.

Petrie, H.T., F. Livak, D. Burtrum, and S. Mazel. 1995. T cell receptor gene recombination patterns and mechanisms: Cell death, rescue, and T cell production. *J. Exp. Med.* 182:121–127. doi:10.1084/jem.182.1.121.

Petrie, H.T., F. Livak, D.G. Schatz, A. Strasser, I.N. Crispe, and K. Shortman. 1993. Multiple rearrangements in T cell receptor α chain genes maximize the production of useful thymocytes. *J. Exp. Med.* 178:615–622. doi:10.1084/jem.178.2.615.

Phillips-Cremins, J.E., M.E.G. Sauria, A. Sanyal, T.I. Gerasimova, B.R. Lajoie, J.S.K. Bell, C.T. Ong, T.A. Hookway, C. Guo, Y. Sun, M.J. Bland, W. Wagstaff, S. Dalton, T.C. McDevitt, R. Sen, J. Dekker, J. Taylor, and V.G. Corces. 2013. Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment. *Cell*. 153:1281–1295. doi:10.1016/J.CELL.2013.04.053.

Pickersgill, H., B. Kalverda, E. de Wit, W. Talhout, M. Fornerod, and B. van Steensel. 2006. Characterization of the Drosophila melanogaster genome at the nuclear lamina. *Nat. Genet.* 38:1005–1014. doi:10.1038/NG1852.

Pinto, F.L., and P. Lindblad. 2010. A guide for in-house design of template-switch-based 5′ rapid amplification of cDNA ends systems. *Anal. Biochem.* 397:227–32. doi:10.1016/j.ab.2009.10.022.

Pope, B.D., T. Ryba, V. Dileep, F. Yue, W. Wu, O. Denas, D.L. Vera, Y. Wang, R.S. Hansen, T.K. Canfield, R.E. Thurman, Y. Cheng, G. Gülsoy, J.H. Dennis, M.P. Snyder, J.A. Stamatoyannopoulos, J. Taylor, R.C. Hardison, T. Kahveci, B. Ren, and D.M. Gilbert. 2014. Topologically associating domains are stable units of replication-timing regulation. *Nature*. 515:402–405. doi:10.1038/nature13986.

Queen, C., and D. Baltimore. 1983. Immunoglobulin gene transcription is activated by downstream sequence elements. *Cell*. 33:741–748. doi:10.1016/0092-8674(83)90016-8.

Quigley, M.F., J.R. Almeida, D.A. Price, and D.C. Douek. 2011. Unbiased molecular analysis of T cell receptor expression using template-switch anchored RT-PCR. *Curr. Protoc. Essent. Lab. Tech.* 94:10–33. doi:10.1002/0471142735.im1033s94.

Quitschke, W.W., M.J. Taheny, L.J. Fochtmann, and A.A. Vostrov. 2000. Differential effect of zinc finger deletions on the binding of CTCF to the promoter of the amyloid precursor protein gene. *Nucleic Acids Res.* 28:3370–78. doi:10.1093/NAR/28.17.3370.

Ramón-Maiques, S., A.J. Kuo, D. Carney, A.G.W. Matthews, M.A. Oettinger, O. Gozani, and W. Yang. 2007. The plant homeodomain finger of RAG2 recognizes histone H3 methylated at both lysine-4 and arginine-2. *Proc. Natl. Acad. Sci. U. S. A.* 104:18993–18998. doi:10.1073/PNAS.0709170104.

Ramsden, D.A., C.J. Paige, and G.E. Wu. 1994. Kappa light chain rearrangement in mouse fetal liver. *J. Immunol.* 153:1150–60.

Rao, S.S.P., S.C. Huang, B. Glenn St Hilaire, J.M. Engreitz, E.M. Perez, K.R. Kieffer-Kwon, A.L. Sanborn, S.E. Johnstone, G.D. Bascom, I.D. Bochkov, X. Huang, M.S. Shamim, J. Shin, D. Turner, Z. Ye, A.D. Omer, J.T. Robinson, T. Schlick, B.E. Bernstein, R. Casellas, E.S. Lander, and E.L. Aiden. 2017. Cohesin Loss Eliminates All Loop Domains. *Cell*. 171:305-320.e24. doi:10.1016/J.CELL.2017.09.026.

Rao, S.S.P., M.H. Huntley, N.C. Durand, E.K. Stamenova, I.D. Bochkov, J.T. Robinson, A.L. Sanborn, I. Machol, A.D. Omer, E.S. Lander, and E.L. Aiden. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 159:1665–1680. doi:10.1016/j.cell.2014.11.021.

Reddy, K.L., J.M. Zullo, E. Bertolino, and H. Singh. 2008. Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature*. 452:243–247. doi:10.1038/nature06727.

Redondo, J., S. Hata, C. Brocklehurst, and M. Krangel. 1990. A T cell-specific transcriptional enhancer within the human T cell receptor delta locus. *Science.* 247:1225–1229. doi:10.1126/SCIENCE.2156339.

Reilly, E.B., B. Blomberg, T. Imanishi-Kari, S. Tonegawa, and H.N. Eisen. 1984. Restricted association of V and J-C gene segments for mouse lambda chains. *Proc. Natl. Acad. Sci. U. S. A.* 81:2484–88. doi:10.1073/PNAS.81.8.2484.

Ribeiro de Almeida, C., R. Stadhouders, M.J.W. De Bruijn, I.M. Bergen, S. Thongjuea, B. Lenhard, W. Van IJcken, F. Grosveld, N. Galjart, E. Soler, and R.W. Hendriks. 2011. The DNA-Binding Protein CTCF Limits Proximal Vκ Recombination and Restricts κ Enhancer Interactions to the Immunoglobulin κ Light Chain Locus. *Immunity*. 35:501–513. doi:10.1016/J.IMMUNI.2011.07.014.

Roldán, E., M. Fuxa, W. Chong, D. Martinez, M. Novatchkova, M. Busslinger, and J.A.

Skok. 2005. Locus “decontraction” and centromeric recruitment contribute to allelic exclusion of the immunoglobulin heavy-chain gene. *Nat. Immunol.* 6:31–41. doi:10.1038/ni1150.

Roth, D.B. 2014. V(D)J Recombination: Mechanism, Errors, and Fidelity. *Microbiol. Spectr.* 2. doi:10.1128/MICROBIOLSPEC.MDNA3-0041-2014.

Rothenberg, E., and D. Triglia. 1983. Clonal proliferation unlinked to terminal deoxynucleotidyl transferase synthesis in thymocytes of young mice. *J. Immunol.* 130:1627–33.

Rothenberg, E. V. 2011. T Cell Lineage Commitment: Identity and Renunciation. *J. Immunol.* 186:6649–6655. doi:10.4049/jimmunol.1003703.

Roy, S., A.J. Moore, C. Love, A. Reddy, D. Rajagopalan, S.S. Dave, L. Li, C. Murre, and Y. Zhuang. 2018. Id Proteins Suppress E2A-Driven Invariant Natural Killer T Cell Development prior to TCR Selection. *Front. Immunol.* 9:42. doi:10.3389/fimmu.2018.00042.

Ru, H., M.G. Chambers, T.-M. Fu, A.B. Tong, M. Liao, and H. Wu. 2015. Molecular Mechanism of V(D)J Recombination from Synaptic RAG1-RAG2 Complex Structures. *Cell*. 163:1138–52. doi:10.1016/j.cell.2015.10.055.

Ru, H., W. Mi, P. Zhang, F.W. Alt, D.G. Schatz, M. Liao, and H. Wu. 2018. DNA melting initiates the RAG catalytic pathway. *Nat. Struct. Mol. Biol.* 25:732–742. doi:10.1038/s41594-018-0098-5.

Sanborn, A.L., S.S.P. Rao, S.-C. Huang, N.C. Durand, M.H. Huntley, A.I. Jewett, I.D. Bochkov, D. Chinnappan, A. Cutkosky, J. Li, K.P. Geeting, A. Gnirke, A. Melnikov, D. McKenna, E.K. Stamenova, E.S. Lander, and E.L. Aiden. 2015. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. U. S. A.* 112:E6456–E6465. doi:10.1073/PNAS.1518552112.

Sanchez, P., P.N. Marche, C. Le Guern, and P.A. Cazenave. 1987. Structure of a third murine immunoglobulin lambda light chain variable region that is expressed in laboratory mice. *Proc. Natl. Acad. Sci. U. S. A.* 84:9185–8. doi:10.1073/PNAS.84.24.9185.

Sanchez, P., B. Nadel, and P.-A. Cazenave. 1991. Vλ-Jλ rearrangements are restricted within a V-J-C recombination unit in the mouse. *Eur. J. Immunol.* 21:907–911. doi:10.1002/EJI.1830210408.

Sayegh, C., S. Jhunjhunwala, R. Riblet, and C. Murre. 2005. Visualization of looping involving the immunoglobulin heavy-chain locus in developing B cells. *Genes Dev.* 19:322–327. doi:10.1101/GAD.1254305.

Schatz, D.G., and Y. Ji. 2011. Recombination centres and the orchestration of V(D)J recombination. *Nat. Rev. Immunol.* 11:251–263. doi:10.1038/nri2941.

Schatz, D.G., M.A. Oettinger, and D. Baltimore. 1989. The V(D)J recombination activating gene, RAG-1. *Cell*. 59:1035–1048. doi:10.1016/0092-8674(89)90760-5.

Schatz, D.G., and P.C. Swanson. 2011. V(D)J Recombination: Mechanisms of Initiation. *Annu. Rev. Genet.* 45:167–202. doi:10.1146/annurev-genet-110410-132552.

Schlimgen, R.J., K.L. Reddy, H. Singh, and M.S. Krangel. 2008. Initiation of allelic exclusion by stochastic interaction of Tcrb alleles with repressive nuclear compartments. *Nat. Immunol.* 9:802–809. doi:10.1038/ni.1624.

Schlissel, M., A. Constantinescu, T. Morrow, M. Baxter, and A. Peng. 1993. Double-strand signal sequence breaks in V(D)J recombination are blunt, 5'-phosphorylated, RAG-dependent, and cell cycle regulated. *Genes Dev.* 7:2520–2532. doi:10.1101/GAD.7.12B.2520.

Seitan, V.C., B. Hao, K. Tachibana-Konwalski, T. Lavagnolli, H. Mira-Bontenbal, K.E. Brown, G. Teng, T. Carroll, A. Terry, K. Horan, H. Marks, D.J. Adams, D.G. Schatz, L. Aragon, A.G. Fisher, M.S. Krangel, K. Nasmyth, and M. Merkenschlager. 2011. A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation. *Nature*. 476:467–471. doi:10.1038/nature10312.

Sexton, T., and G. Cavalli. 2015. The Role of Chromosome Domains in Shaping the Functional Genome. *Cell*. 160:1049–1059. doi:10.1016/J.CELL.2015.02.040.

Shen, Y., F. Yue, D.F. McCleary, Z. Ye, L. Edsall, S. Kuan, U. Wagner, J. Dixon, L. Lee, V. V. Lobanenkov, and B. Ren. 2012. A map of the cis-regulatory sequences in the mouse genome. *Nature*. 488:116–120. doi:10.1038/nature11243.

Shih, H.-Y., and M.S. Krangel. 2010. Distinct contracted conformations of the Tcra/Tcrd locus during Tcra and Tcrd recombination. *J. Exp. Med.* 207:1835–41. doi:10.1084/jem.20100772.

Shih, H.-Y., J. Verma-Gaur, A. Torkamani, A.J. Feeney, N. Galjart, and M.S. Krangel. 2012. Tcra gene recombination is supported by a Tcra enhancer- and CTCF-dependent chromatin hub. *Proc. Natl. Acad. Sci. U. S. A.* 109:E3493-502. doi:10.1073/pnas.1214131109.

Shimizu, T., S. Takeshita, M. Muto, E. Kubo, T. Sado, and H. Yamagishi. 1993. Mouse germline transcript of TCR α joining region and temporal expression in ontogeny. *Int. Immunol.* 5:155–160. doi:10.1093/INTIMM/5.2.155.

Shinoda, K., Y. Maman, A. Canela, D.G. Schatz, F. Livak, and A. Nussenzweig. 2019. Intra-Vκ Cluster Recombination Shapes the Ig Kappa Locus Repertoire. *Cell Rep.* 29:4471-4481.e6. doi:10.1016/J.CELREP.2019.11.088.

Shugay, M., D. V. Bagaev, M.A. Turchaninova, D.A. Bolotin, O. V. Britanova, E. V. Putintseva, M. V. Pogorelyy, V.I. Nazarov, I. V. Zvyagin, V.I. Kirgizova, K.I. Kirgizov, E. V. Skorobogatova, and D.M. Chudakov. 2015. VDJtools: Unifying Post-analysis of T Cell Receptor Repertoires. *PLoS Comput. Biol.* 11:e1004503. doi:10.1371/journal.pcbi.1004503.

Singh, P., J.C. Schimenti, and E. Bolcun-filas. 2014. A Mouse Geneticist's practical guide to CRISPR applications. *Genetics*. 199:1–15. doi:10.1534/genetics.114.169771.

Skok, J.A., R. Gisler, M. Novatchkova, D. Farmer, W. de Laat, and M. Busslinger. 2007. Reversible contraction by looping of the Tcra and Tcrb loci in rearranging thymocytes. *Nat. Immunol.* 8:378–387. doi:10.1038/ni1448.

Sleckman, B., C. Bardon, R. Ferrini, L. Davidson, and F.W. Alt. 1997. Function of the TCR alpha enhancer in alphabeta and gammadelta T cells. *Immunity*. 7:505–515. doi:10.1016/S1074-7613(00)80372-6.

Sleckman, B.P., C.H. Bassing, M.M. Hughes, A. Okada, M. D'Auteuil, T.D. Wehrly, B.B. Woodman, L. Davidson, J. Chen, and F.W. Alt. 2000. Mechanisms that direct ordered assembly of T cell receptor β locus V, D, and J gene segments. *Proc. Natl. Acad. Sci. U. S. A.* 405:583–6. doi:10.1073/pnas.130190597.

Sleckman, B.P., B. Khor, R. Monroe, and F.W. Alt. 1998. Assembly of productive T cell receptor δ variable region genes exhibits allelic inclusion. *J. Exp. Med.* 188:1465–1471. doi:10.1084/jem.188.8.1465.

Sollbach, A.E., and G.E. Wu. 1995. Inversions produced during V(D)J rearrangement at IgH, the immunoglobulin heavy-chain locus. *Mol. Cell. Biol.* 15:671–81. doi:10.1128/mcb.15.2.671.

Spitz, F., and E.E.M. Furlong. 2012. Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* 13:613–626. doi:10.1038/nrg3207.

Surh, C.D., and J. Sprent. 1994. T-cell apoptosis detected in situ during positive and

negative selection in the thymus. *Nature*. 372:100–103. doi:10.1038/372100a0.

Swat, W., L. Ignatowicz, H. von Boehmer, and P. Kisielow. 1991. Clonal deletion of immature CD4+8+ thymocytes in suspension culture by extrathymic antigen-presenting cells. *Nature*. 351:150–153. doi:10.1038/351150a0.

Takahama, Y., and A. Singer. 1992. Post-transcriptional regulation of early T cell development by T cell receptor signals. *Science.* 258:1456–1462. doi:10.1126/SCIENCE.1439838.

Takeshita, S., M. Toda, and H. Yamagishi. 1989. Excision products of the T cell receptor gene support a progressive rearrangement model of the alpha/delta locus. *EMBO J.* 8:3261–3270. doi:10.1002/J.1460-2075.1989.TB08486.X.

Taniuchi, I. 2018. CD4 Helper and CD8 Cytotoxic T Cell Differentiation. *Annu. Rev. Immunol.* 36:579–601. doi:10.1146/annurev-immunol-042617-053411.

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/.

Teng, G., Y. Maman, W. Resch, M. Kim, A. Yamane, J. Qian, K.R. Kieffer-Kwon, M. Mandal, Y. Ji, E. Meffre, M.R. Clark, L.G. Cowell, R. Casellas, and D.G. Schatz. 2015. RAG Represents a Widespread Threat to the Lymphocyte Genome. *Cell*. 162:751–765. doi:10.1016/j.cell.2015.07.009.

Thompson, S.D., J. Pelkonen, and J.L. Hurwitz. 1990. First T cell receptor alpha gene rearrangements during T cell ontogeny skew to the 5’ region of the J alpha locus. *J. Immunol.* 145:2347–52.

Thwaites, D.T., C. Carter, D. Lawless, S. Savic, and J.M. Boyes. 2019. A novel RAG1 mutation reveals a critical in vivo role for HMGB1/2 during V(D)J recombination. *Blood*. 133:820–829. doi:10.1182/blood-2018-07-866939.

Tilloy, F., E. Treiner, S.-H. Park, C. Garcia, F. Lemonnier, H. de la Salle, A. Bendelac, M. Bonneville, and O. Lantz. 1999. An Invariant T Cell Receptor α Chain Defines a Novel TAP-independent Major Histocompatibility Complex Class Ib–restricted α/β T Cell Subpopulation in Mammals. *J. Exp. Med.* 189:1907–21. doi:10.1084/JEM.189.12.1907.

Toda, M., S. Fujimoto, T. Iwasato, S. Takeshita, K. Tezuka, T. Ohbayashi, and H. Yamagishi. 1988. Structure of extrachromosomal circular DNAs excised from T-cell antigen receptor alpha and delta-chain loci. *J. Mol. Biol.* 202:219–231. doi:10.1016/0022-2836(88)90453-6.

Tsukada, S., H. Sugiyama, Y. Oka, and S. Kishimoto. 1990. Estimation of D segment usage in initial D to JH joinings in a murine immature B cell line. Preferential usage of DFL16.1, the most 5' D segment and DQ52, the most JH-proximal D segment. *J. Immunol.* 144:4053–9.

Vettermann, C., G.A. Timblin, V. Lim, E.C. Lai, and M.S. Schlissel. 2015. The proximal J kappa germline-transcript promoter facilitates receptor editing through control of ordered recombination. *PLoS One*. 10:e0113824. doi:10.1371/journal.pone.0113824.

Vian, L., A. Pękowska, S.S.P. Rao, K.R. Kieffer-Kwon, S. Jung, L. Baranello, S.C. Huang, L. El Khattabi, M. Dose, N. Pruett, A.L. Sanborn, A. Canela, Y. Maman, A. Oksanen, W. Resch, X. Li, B. Lee, A.L. Kovalchuk, Z. Tang, S. Nelson, M. Di Pierro, R.R. Cheng, I. Machol, B.G. St Hilaire, N.C. Durand, M.S. Shamim, E.K. Stamenova, J.N. Onuchic, Y. Ruan, A. Nussenzweig, D. Levens, E.L. Aiden, and R. Casellas. 2018. The Energetics and Physiological Impact of Cohesin Extrusion. *Cell*. 173:1165-1178.e20. doi:10.1016/J.CELL.2018.03.072.

Villartay, J.P. de, D. Lewis, R. Hockett, T.A. Waldmann, S.J. Korsmeyer, and D.I. Cohen. 1987. Deletional rearrangement in the human T-cell receptor alpha-chain locus. *Proc. Natl. Acad. Sci. U. S. A.* 84:8608–12. doi:10.1073/PNAS.84.23.8608.

Villey, I., D. Caillol, F. Selz, P. Ferrier, and J.P. de Villartay. 1996. Defect in rearrangement of the most 5' TCR-J alpha following targeted deletion of T early alpha (TEA): implications for TCR alpha locus accessibility. *Immunity*. 5:331–342. doi:10.1016/S1074-7613(00)80259-9.

Villey, I., P. Quartier, F. Selz, and J.P. De Villartay. 1997. Germ-line transcription and methylation status of the TCR-Jα locus in its accessible configuration. *Eur. J. Immunol.* 27:1619–25. doi:10.1002/eji.1830270705.

Wada, H., K. Masuda, R. Satoh, K. Kakugawa, T. Ikawa, Y. Katsura, and H. Kawamoto. 2008. Adult T-cell progenitors retain myeloid potential. *Nature*. 452:768–772. doi:10.1038/nature06839.

Wang, F., C.Y. Huang, and O. Kanagawa. 1998. Rapid deletion of rearranged T cell antigen receptor (TCR) Vα-Jα segment by secondary rearrangement in the thymus: Role of continuous rearrangement of TCR a chain gene and positive selection in the T cell repertoire formation. *Proc. Natl. Acad. Sci. U. S. A.* 95:11834–11839. doi:10.1073/pnas.95.20.11834.

Wang, Y., J. Liu, P.D. Burrows, and J.Y. Wang. 2020. B Cell Development and Maturation. *In* Advances in Experimental Medicine and Biology. Springer. 1–22.

Ward, A., G. Kumari, R. Sen, and S. Desiderio. 2018. The RAG-2 Inhibitory Domain Gates Accessibility of the V(D)J Recombinase to Chromatin. *Mol. Cell. Biol.* 38:e00159-18. doi:10.1128/MCB.00159-18.

Warnes, G.R., B. Bolker, L. Bonebakker, R. Gentleman, W. Huber, A. Liaw, T. Lumley, M. Maechler, A. Magnusson, S. Moeller, M. Schwartz, B. Venables, and T. Galili. 2009. gplots: Various R programming tools for plotting data. http://cran.r-project.org/web/packages/gplots/index.html.

Weber-Arden, J., O.M. Wilbert, D. Kabelitz, and B. Arden. 2000. Vδ Repertoire During Thymic Ontogeny Suggests Three Novel Waves of γδ TCR Expression. *J. Immunol.* 164:1002–1012. doi:10.4049/JIMMUNOL.164.2.1002.

Wendt, K.S., K. Yoshida, T. Itoh, M. Bando, B. Koch, E. Schirghuber, S. Tsutsumi, G. Nagae, K. Ishihara, T. Mishiro, K. Yahata, F. Imamoto, H. Aburatani, M. Nakao, N. Imamoto, K. Maeshima, K. Shirahige, and J.-M. Peters. 2008. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature*. 451:796–801. doi:10.1038/nature06634.

Whitehurst, C.E., S. Chattopadhyay, and J. Chen. 1999. Control of V(D)J recombinational accessibility of the Dβ1 gene segment at the TCRβ locus by a germline promoter. *Immunity*. 10:313–322. doi:10.1016/S1074-7613(00)80031-X.

Whitehurst, C.E., M.S. Schlissel, and J. Chen. 2000. Deletion of germline promoter PD beta 1 from the TCR beta locus causes hypermethylation that impairs D beta 1 recombination by multiple mechanisms. *Immunity*. 13:703–714. doi:10.1016/S1074-7613(00)00069-8.

Williams, G., A. Martinez, A. Montalbano, A. Tang, A. Mauhar, K. Ogwaro, D. Merz, C. Chevillard, R. Riblet, and A.J. Feeney. 2001. Unequal VH gene rearrangement frequency within the large VH7183 gene family is not due to recombination signal sequence variation, and mapping of the genes shows a bias of rearrangement based on chromosomal location. *J. Immunol.* 167:257–263. doi:10.4049/JIMMUNOL.167.1.257.

Wilson, A., W. Held, and H.R. MacDonald. 1994. Two waves of recombinase gene expression in developing thymocytes. *J. Exp. Med.* 179:1355–1360. doi:10.1084/JEM.179.4.1355.

Winoto, A., and D. Baltimore. 1989. A novel, inducible and T cell-specific enhancer located at the 3′ end of the T cell receptor alpha locus. *EMBO J.* 8:729–733. doi:10.1002/J.1460-2075.1989.TB03432.X.

de Wit, E., E.S.M. Vos, S.J.B. Holwerda, C. Valdes-Quezada, M.J.A.M. Verstegen, H. Teunissen, E. Splinter, P.J. Wijchers, P.H.L. Krijger, and W. de Laat. 2015. CTCF Binding Polarity Determines Chromatin Looping. *Mol. Cell*. 60:676–684. doi:10.1016/j.molcel.2015.09.023.

Woolf, T., E. Lai, M. Kronenberg, and L. Hood. 1988. Mapping genomic organization by field inversion and two-dimensional gel electrophoresis: application to the murine T-cell receptor gamma gene family. *Nucleic Acids Res.* 16:3863–75. doi:10.1093/NAR/16.9.3863.

Wu, C., C.H. Bassing, D. Jung, B.B. Woodman, D. Foy, and F.W. Alt. 2003. Dramatically Increased Rearrangement and Peripheral Representation of Vβ14 Driven by the 3′Dβ1 Recombination Signal Sequence. *Immunity*. 18:75–85. doi:10.1016/S1074-7613(02)00515-0.

Wu, G.S., and C.H. Bassing. 2020. Inefficient V(D)J recombination underlies monogenic T cell receptor β expression. *Proc. Natl. Acad. Sci. U. S. A.* 117:18172–18174. doi:10.1073/PNAS.2010077117.

Wu, G.S., K.S. Yang-Iott, M.A. Klink, K.E. Hayer, K.D. Lee, and C.H. Bassing. 2020. Poor quality Vβ recombination signal sequences stochastically enforce TCRβ allelic exclusion. *J. Exp. Med.* 217:e20200412. doi:10.1084/JEM.20200412.

Xiang, Y., and W.T. Garrard. 2008. The Downstream Transcriptional Enhancer, Ed, Positively Regulates Mouse Igκ Gene Expression and Somatic Hypermutation. *J. Immunol.* 180:6725–6732. doi:10.4049/jimmunol.180.10.6725.

Xiang, Y., S.-K. Park, and W.T. Garrard. 2013. Vκ gene repertoire and locus contraction are specified by critical DNase I hypersensitive sites within the Vκ-Jκ intervening region. *J. Immunol.* 190:1819–26. doi:10.4049/JIMMUNOL.1203127.

Xiang, Y., X. Zhou, S.L. Hewitt, J.A. Skok, and W.T. Garrard. 2011. A Multifunctional Element in the Mouse Igκ Locus That Specifies Repertoire and Ig Loci Subnuclear Location. *J. Immunol.* 186:5356–5366. doi:10.4049/JIMMUNOL.1003794.

Xiong, N., J.E. Baker, C. Kang, and D.H. Raulet. 2004. The genomic arrangement of T cell receptor variable genes is a determinant of the developmental rearrangement pattern. *Proc. Natl. Acad. Sci. U. S. A.* 101:260–265. doi:10.1073/PNAS.0303738101.

Xiong, N., C. Kang, and D. Raulet. 2002. Redundant and unique roles of two enhancer elements in the TCRgamma locus in gene regulation and gammadelta T cell development. *Immunity*. 16:453–463. doi:10.1016/S1074-7613(02)00285-6.

Xiong, N., and D.H. Raulet. 2007. Development and selection of γδ T cells. *Immunol. Rev.* 215:15–31. doi:10.1111/J.1600-065X.2006.00478.X.

Xiong, N., L. Zhang, C. Kang, and D.H. Raulet. 2008. Gene placement and competition control T cell receptor gamma variable region gene rearrangement. *J. Exp. Med.* 205:929–938. doi:10.1084/JEM.20071275.

Yamagami, T., E. ten Boekel, J. Andersson, A. Rolink, and F. Melchers. 1999. Frequencies of multiple IgL chain gene rearrangements in single normal or kappaL chain-deficient B lineage cells. *Immunity*. 11:317–327. doi:10.1016/S1074-7613(00)80107-7.

Yamano, T., J. Nedjic, M. Hinterberger, M. Steinert, S. Koser, S. Pinto, N. Gerdes, E. Lutgens, N. Ishimaru, M. Busslinger, B. Brors, B. Kyewski, and L. Klein. 2015. Thymic B Cells Are Licensed to Present Self Antigens for Central T Cell Tolerance Induction. *Immunity*. 42:1048–1061. doi:10.1016/j.immuni.2015.05.013.

Yancopoulos, G., and F. Alt. 1985. Developmentally controlled and tissue-specific expression of unrearranged VH gene segments. *Cell*. 40:271–281. doi:10.1016/0092-8674(85)90141-2.

Yancopoulos, G.D., and F.W. Alt. 1986. Regulation of the Assembly and Expression of Variable-Region Genes. *Annu. Rev. Immunol.* 4:339–368. doi:10.1146/ANNUREV.IY.04.040186.002011.

Yancopoulos, G.D., S. Desiderio, M. Paskind, J.F. Kearney, D. Baltimore, and F.W. Alt. 1984. Preferential utilization of the most JH-proximal VH gene segments in pre-B-cell lines. *Nature*. 311:727–733. doi:10.1038/311727A0.

Yannoutsos, N., V. Barreto, Z. Misulovin, A. Gazumyan, W. Yu, N. Rajewsky, B.R. Peixoto, T. Eisenreich, and M.C. Nussenzweig. 2004. A cis element in the recombination activating gene locus regulates gene expression by counteracting a distant silencer. *Nat. Immunol.* 5:443–450. doi:10.1038/ni1053.

Yannoutsos, N., P. Wilson, W. Yu, H.T. Chen, A. Nussenzweig, H. Petrie, and M.C. Nussenzweig. 2001. The Role of Recombination Activating Gene (RAG) Reinduction in Thymocyte Development in Vivo. *J. Exp. Med.* 194:471–80. doi:10.1084/JEM.194.4.471.

Yin, F.F., S. Bailey, C.A. Innis, M. Ciubotaru, S. Kamtekar, T.A. Steitz, and D.G. Schatz. 2009. Structure of the RAG1 nonamer binding domain with DNA reveals a dimer that mediates DNA synapsis. *Nat. Struct. Mol. Biol.* 16:499–508.

doi:10.1038/nsmb.1593.

Yin, M., J. Wang, M. Wang, X. Li, M. Zhang, Q. Wu, and Y. Wang. 2017. Molecular mechanism of directional CTCF recognition of a diverse range of genomic sites. *Cell Res.* 27:1365–1377. doi:10.1038/cr.2017.131.

Yoshida, H., C.A. Lareau, R.N. Ramirez, J.D. Buenrostro, C. Benoist, H. Yoshida, S.A. Rose, B. Maier, A. Wroblewska, F. Desland, A. Chudnovskiy, A. Mortha, C. Dominguez, J. Tellier, E. Kim, D. Dwyer, S. Shinton, T. Nabekura, Y. Qi, B. Yu, M. Robinette, K.-W. Kim, A. Wagers, A. Rhoads, S.L. Nutt, B.D. Brown, S. Mostafavi, and T. Immunological Genome Project. 2019. The cis-Regulatory Atlas of the Mouse Immune System. *Cell*. 176:897-912.e20. doi:10.1016/j.cell.2018.12.036.

Yu, W., Z. Misulovin, H. Suh, R.R. Hardy, M. Jankovic, N. Yannoutsos, and M.C. Nussenzweig. 1999. Coordinate regulation of RAG1 and RAG2 by cell type-specific DNA elements 5' of RAG2. *Science.* 285:1080–1084. doi:10.1126/science.285.5430.1080.

Zhang, B., J. Wu, Y. Jiao, C. Bock, M. Dai, B. Chen, N. Chao, W. Zhang, and Y. Zhuang. 2015. Differential Requirements of TCR Signaling in Homeostatic Maintenance and Function of Dendritic Epidermal T Cells. *J. Immunol.* 195:4282–4291. doi:10.4049/jimmunol.1501220.

Zhang, Y., X. Zhang, Z. Ba, Z. Liang, E.W. Dring, H. Hu, J. Lou, N. Kyritsis, J. Zurita, M.S. Shamim, A. Presser Aiden, E. Lieberman Aiden, and F.W. Alt. 2019. The fundamental role of chromatin loop extrusion in physiological V(D)J recombination. *Nature*. 573:600–604. doi:10.1038/s41586-019-1547-y.

Zhao, H., Z. Li, Y. Zhu, S. Bian, Y. Zhang, L. Qin, A.K. Naik, J. He, Z. Zhang, M.S. Krangel, and B. Hao. 2020. A role of the CTCF binding site at enhancer Eα in the dynamic chromatin organization of the Tcra-Tcrd locus. *Nucleic Acids Res.* 48:9621–9636. doi:10.1093/NAR/GKAA711.

Zhao, L., R.L. Frock, Z. Du, J. Hu, L. Chen, M.S. Krangel, and F.W. Alt. 2016. Orientation-specific RAG activity in chromosomal loop domains contributes to Tcrd V(D)J recombination during T cell development. *J. Exp. Med.* 213:1921–1936. doi:10.1084/JEM.20160670.

Zhao, Z., G. Tavoosidana, M. Sjölinder, A. Göndör, P. Mariano, S. Wang, C. Kanduri, M. Lezcano, K. Singh Sandhu, U. Singh, V. Pant, V. Tiwari, S. Kurukuti, and R. Ohlsson. 2006. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat. Genet.* 38:1341–1347. doi:10.1038/ng1891.

Zuin, J., J.R. Dixon, M.I.J.A. van der Reijden, Z. Ye, P. Kolovos, R.W.W. Brouwer, M.P.C. van de Corput, H.J.G. van de Werken, T.A. Knoch, W.F.J. van IJcken, F.G. Grosveld, B. Ren, and K.S. Wendt. 2014. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl. Acad. Sci. U. S. A.* 111:996–1001. doi:10.1073/PNAS.1317788111.

Zullo, J.M., I.A. Demarco, R. Piqué-Regi, D.J. Gaffney, C.B. Epstein, C.J. Spooner, T.R. Luperchio, B.E. Bernstein, J.K. Pritchard, K.L. Reddy, and H. Singh. 2012. DNA Sequence-Dependent Compartmentalization and Silencing of Chromatin at the Nuclear Lamina. *Cell*. 149:1474–1487. doi:10.1016/J.CELL.2012.04.035.

Biography

Danielle Jean Dauphars graduated from Kent State University in Kent, OH in May 2011 with a Bachelor of Science degree in Integrated Life Sciences, with honors. Danielle went on to Duke University to obtain a Ph.D. in Immunology.

Publications:

Lu, L., H. Zhang, M. Zhan, J. Jiang, H. Yin, **D.J. Dauphars**, S.Y. Li, Y. Li, and Y.W. He. 2020. Preventing Mortality in COVID-19 Patients: Which Cytokine to Target in a Raging Storm? *Front. Cell Dev. Biol.* doi:10.3389/fcell.2020.00677.

Lu, L., H. Zhang, M. Zhan, J. Jiang, H. Yin, **D.J. Dauphars**, S.-Y. Li, Y. Li, and Y.-W. He. 2020. Antibody response and therapy in COVID-19 patients: what can be learned for vaccine development? *Sci. China Life Sci. 2020 6312*. 63:1833–1849. doi:10.1007/S11427-020-1859-Y.

Karimi, M.M., Y. Guo, X. Cui, H.A. Pallikonda, V. Horková, Y.-F. Wang, S.R. Gil, G. Rodriguez-Esteban, I. Robles-Rebollo, L. Bruno, R. Georgieva, B. Patel, J. Elliott, M.H. Dore, **D. Dauphars**, M.S. Krangel, B. Lenhard, H. Heyn, A.G. Fisher, O. Štěpánek, and M. Merkenschlager. 2021. The order and logic of CD4 versus CD8 lineage choice and differentiation in mouse thymus. *Nat. Commun.* 12:1–14. doi:10.1038/s41467-020-20306-w.

Lu, L., H. Zhang, **D.J. Dauphars**, and Y.W. He. 2021. A Potential Role of Interleukin 10 in COVID-19 Pathogenesis. *Trends Immunol.* 42:3–5. doi:10.1016/J.IT.2020.10.012.

Dauphars, D. J., G. Wu, C. H. Bassing, and M. S. Krangel. (In Press). Molecular Analysis of Mouse T Cell Receptor a and b Gene Rearrangements. *Methods Mol. Biol*.

Dauphars, D. J., A. Mihai, L. Wang, Y. Zhuang, and M. S. Krangel. (In Review). *Trav15-dv6* family *Tcrd* rearrangements diversify the *Tcra* repertoire.