THE ELASTIC ELASMOBRANCH: NURSE SHARKS COMMISSION B CELL COMPONENTS AND SOMATIC HYPERMUTATION MECHANISMS TO DIVERSIFY T CELLS DURING THYMIC DEVELOPMENT

A Dissertation

by

JEANNINE ANNETTE OTT

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Michael F. Criscitiello
Committee Members,	Terje Raudsepp
	David Riley
	Noushin Ghaffari
Head of Department,	Ramesh Vemulapalli

May 2020

Major Subject: Biomedical Sciences

ABSTRACT

Since the discovery of the T cell receptor (TCR) in 1983, immunologists have assigned somatic hypermutation (SHM) as a mechanism employed solely by B cells to diversify their antigen receptors. Remarkably, we found SHM acting in the thymus on the α chain locus of shark TCR for TCR repertoire generation. SHM in developing shark T cells likely is catalyzed by activation-induced cytidine deaminase (AID) and results in both point and tandem mutations that accumulate non-conservative amino acid replacements within complementarity-determining regions (CDRs). Mutation frequency at TCRα was as high as that seen at B cell receptor loci (BCR) in sharks and mammals, and the mechanism of SHM shares unique characteristics first detected at shark BCR loci. Additionally, fluorescence *in situ* hybridization showed the strongest AID expression in thymic corticomedullary junction and medulla. We suggest that TCRα utilizes SHM to broaden diversification of the primary αβ T cell repertoire in sharks, the first reported use of this process in thymic diversification in vertebrates.

In addition to canonical T and B cell receptors, cartilaginous fish assemble non-canonical TCR that employ various B cell components. For example, shark T cells associate alpha (TCR-alpha) or delta (TCR-delta) constant (C) regions with immunoglobulin (Ig) heavy chain (H) variable (V) segments or TCR-associated Ig-like V (TAIL V) segments to form chimeric IgHV-T cell receptors, and combine TCR δC with both Ig-like *and* TCR-like V segments to form the doubly-rearranging NAR-TCR. Here, we found that the use of SHM

by nurse shark TCR varies depending on the particular V segment or C region used. First, SHM significantly alters alpha/delta V (TCR αδV) segments using TCR αC but not TCR δC. Second, mutation to IgHV segments associated with TCR δC was reduced compared to mutation to TCR αδV associated with TCR αC. Mutation was present but limited in V segments of all other TCR chains, including NAR-TCR. Unexpectedly, we found preferential rearrangement of the non-canonical IgHV-TCR δC over canonical TCR αδV-TCR δC receptors. The differential use of SHM may reveal how AID targets V regions.

DEDICATION

To Sam and Lily, my favorite people on the planet.

“It is never too late to be what you might have been.”

— Mary Anne Evans (*nom de plume*, George Eliot)

ACKNOWLEDGEMENTS

First and foremost, I thank my advisor, mentor, and friend, Mike Criscitiello, for taking on a divorced mother of two in her mid-40s as a Ph.D. student and for believing in me even when I found it hard to believe in myself. I also thank the members of my committee for their thoughtful guidance and reassurance and Dr. Ellen Hsu for agreeing to be an outside reader for someone you never met before. Dr. Terje Raudsepp, your intelligence is as contagious as your laughter. Thank you for your continuous insight on such a broad range of topics. Dr. David Riley I appreciate your calm voice and reassuring demeanor, especially during my toughest prelim days. And Dr. Noushin Ghaffari your optimism is inspiring. Thank you all for mentoring me through one of the most demanding mental challenges of my life.

To all the shark folks at the University of Maryland, especially Dr. Martin Flajnik for teaching me how to get blood from a shark (and other sundry adventures) and for your continued hospitality during my visits to Baltimore; Dr. Helen Dooley for advice and encouragement against all odds; and to Yuko Ohta for your constant support of my research and dietary limitations. To the current and former members of the Comparative Immunogenetics Lab: Dr. Sara Mashoof, Dr. Breanna Breaux, Dr. Thad Deiss, Jenna Harrison, Brooke Norwood Minal Jamsandekar, Kelly Head, Christian Mitchell, Ashley Marchand, Ruth Scego, Kaitlyn Romoser, Cody Horton, and Omar Manzur. Each one of you contributed to the success of this document and provided the skills, laughter, and general craziness that kept me sane.

To the National Geographic Society for your episode on Hawai`i, where I first discovered the cliffs of Moloka`i and fell in love with biology. Two conservation biologists, dangling from the cliffs to hand pollinate endangered flowers. In those five seconds you sealed my fate as a research biologist.

To my family, especially my mom Kathy and my sister Julie, for cheering me on and reminding me why I started a Ph.D. in the first place, especially when I just wanted to lounge on the couch, knit a few rows, and watch television instead. To Brian for your constant support and forgiveness when my constant refrain ("next week will be less busy") never rang true. I promise, one day it will be less busy. And to my friends, especially Stephanie Sword, Dawn Harvey, Rebecca Morgan, Pam Criscitiello, and Tina Houser, for your comfort over cups of coffee or glasses of wine and for your faith in me as a person first and a Ph.D. student second.

And finally, to my children, Sam and Lily, for allowing me this time to explore new opportunities in my life at a time when most parents are settled. You truly are the light at the end of a long day.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by my dissertation committee consisting of my advisor Dr. Michael F. Criscitiello, and committee members Dr. Terje Raudsepp of the Department of Veterinary Integrative Biosciences, Dr. David Riley of the Department of Animal Sciences, and Dr. Noushin Ghaffari of the Computer Science Department at Prairie View A&M University. Dr. Ellen Hsu of the Department of Physiology and Pharmacology at State University of New York served as external advisor during my dissertation defense.

All data presented in Chapters 2 and 3 were conducted in collaboration with Dr. Martin F. Flajnik and Dr. Yuko Ohta of the Department of Microbiology and Immunology at the University of Maryland School of Medicine.

All other work for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by a merit fellowship from Texas A&M University College of Veterinary Medicine and a research assistantship from the National Science Foundation.

This work was also made possible in part by National Science Foundation under Grant Numbers IOS-1257829 and IOS-1656870 to Michael F. Criscitiello and by National Institutes of Health under Grant Number R01OD0549 to Martin F. Flajnik. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NSF or NIH.

NOMENCLATURE

Ab	antibody
Ag	antigen
BCR	B cell receptor
Ig	immunoglobulin
IgSF	immunoglobulin superfamily
IgH	immunoglobulin heavy chain
HC	heavy chain
IgL	immunoglobulin light chain
LC	light chain
TCR	T cell receptor
MHC	major histocompatibility complex
V	variable segment
C	constant region
FR	framework region
CDR	complementarity-determining region
TCR αδV	T cell receptor alpha/delta variable region
TCR βV	T cell receptor beta variable region
TCR γV	T cell receptor gamma variable
IgH V	immunoglobulin heavy chain (IgM/IgW) variable region

NAR	new (or nurse shark) antigen receptor
IgNAR	immunoglobulin heavy chain new antigen receptor
TAIL V	TCR-associated immunoglobulin-like variable region
VHδ	IgH-like TCRδ V
NTCR V	NAR T cell receptor membrane-distal V domain
STCR δV	NAR T cell receptor supporting (membrane-proximal) V domain
RAG	recombination-activating genes
APOBEC	apolipoprotein B RNA-editing catalytic component
CDA	cytidine deaminase
AID	activation-induced cytidine deaminase
SHM	somatic hypermutation
CSR	class switch recombination
IGC	immunoglobulin gene conversion
GC	germinal center
S/N	substitutions per nucleotide
R	replacement (non-synonymous) mutation
S	silent (synonymous) mutation

TABLE OF CONTENTS

Page

Page

Page

LIST OF FIGURES

Page

Page

LIST OF TABLES

Page

1. INTRODUCTION

"Nothing in biology makes sense except in the light of evolution"

– Theodosius Dobzhansky

Vertebrate immunity fights a recurring battle between an endless onslaught of pathogen offenders and persistent opposition by an imposing lymphocyte defense. Not surprisingly, organisms developed sophisticated immune strategies to recognize and combat pathogen invasion while sparing host tissues from harm (Beutler 2004). These strategies are twofold: an immediate, general response by a pervasive innate immune system and a long-term, highly specific response by an acquired adaptive immune system (Murphy and Weaver 2017). Vertebrate innate immunity primarily relies on myeloid cells that engulf pathogens and kill them (Beutler 2004). Often considered the "first responders" of immunity, these dendritic cells, neutrophils, macrophages, and other leukocytes of the vertebrate innate immune system defend against pathogens that breach anatomical barriers (e.g., epithelial tissues of skin or mucosa), resulting in infection (Beutler 2004; Murphy and Weaver 2017). Using germline-encoded pattern recognition receptors (PRRs), leukocytes recognize unique structural molecules shared by a broad array of microbes (called pattern-associated molecular patterns, or PAMPs). In mammals, leukocytes express many different PRRs [e.g., toll-like receptors (TLR) and nucleotide-binding oligomerization domain (NOD)-like receptors (NLR)] that recognize distinct PAMPs found in extracellular or intracellular spaces (Beutler 2004; Murphy and Weaver 2017). Recognition triggers innate immune cells to secrete cytokines that alert

other immune cells and recruit them to sites of infection, activates a complement cascade that identifies and helps clear pathogens, and (in vertebrates) stimulates lymphocytes to induce an adaptive immune response (Beutler 2004; Murphy and Weaver 2017).

While some form of this early warning system exists in all living organisms (i.e., bacteria, plants, fungi), lymphocyte-based adaptive immunity occurs only in vertebrates (Beutler 2004). When innate immunity is insufficient to control infection, the adaptive immune system facilitates both humoral and cell-mediated responses to mount a highly specialized attack against the invader (Murphy and Weaver 2017). However, while innate immune cells are scattered throughout the body, lymphocytes of the adaptive immune system await activation (by the innate immune system or other lymphocytes) within specialized secondary lymphoid tissues (e.g., spleen, gut, and lymph nodes of humans) (Kipps 2010). The hallmarks of adaptive immunity are specificity and memory, creating

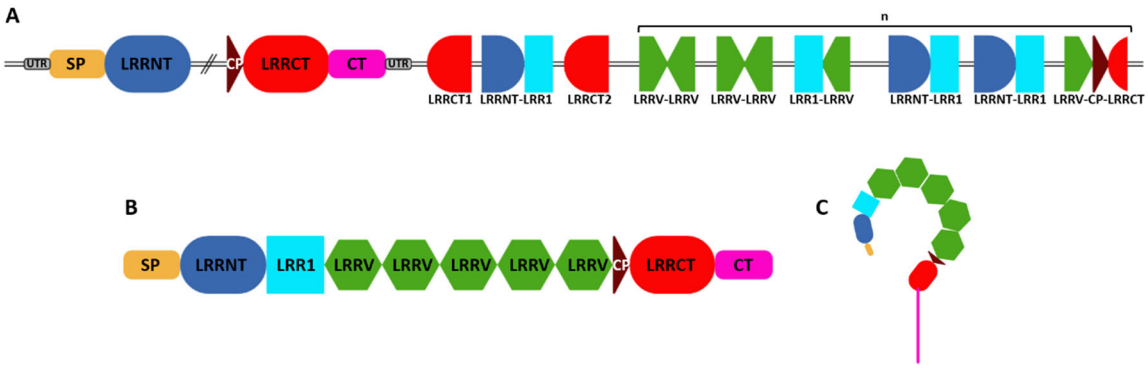

Figure 1-1. Agnathan variable lymphocyte receptors (VLR) are assembled from somatically rearranged leucine rich repeats located in variable cassettes within the germline locus. **[A]** assembled VLR transcript; **[B]** VLR germline locus; **[C]** Assembled VLR [SP: signal peptide; LRR: leucine rich repeat; V: variable; CP: connecting peptide; NT: amino-terminal; CT: carboxy-terminal] [Figure created with BioRender.com; adapted from Das et al. 2014]

both a highly specific preliminary response to a pathogen and a long-term memory response against future invasions by the same pathogen (Ahmed and Gray 1996). While many studies of immunity historically occurred in mice and humans to answer medically-relevant questions, more recent evidence suggest the adaptive immune system of mice and humans only broadly resembles the primordial system founded by the most ancient vertebrates at the dawn of adaptive immunity (Ohta et al. 2019). Vertebrates subsequently evolved two structurally different but functionally similar adaptive immune strategies, both capable of generating highly specific responses while creating immunological memory against future attacks.

Jawless (agnathan hagfish and lamprey) vertebrates evolved a lineage of antigen receptors encoded by genes that somatically rearrange leucine rich repeats (LRR) to generate three forms of mature variable lymphocyte receptors (VLR) – VLR A, B, and C –

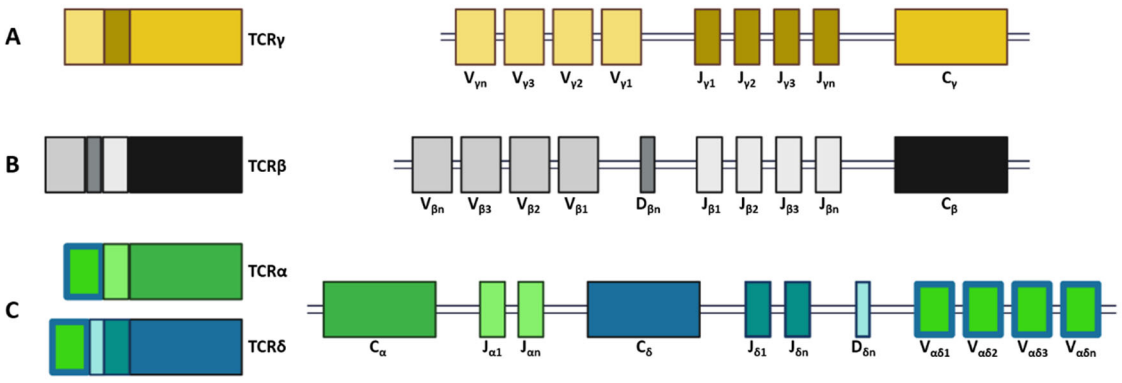

Figure 1-2.T cell receptors (TCR) of jawed vertebrates are assembled from somatically rearranged variable (V), diversity (D), and joining (J) gene segments. [A] Assembled transcripts of TCR Beta (top) and TCR Alpha/Delta (bottom) chains and an assembled αβ TCR; [B] Assembled transcripts of TCR Alpha/Delta (top) and TCR Gamma (bottom) chains and an assembled γδ TCR. [Figure created with BioRender.com; adapted from Deiss et al. 2019]

that are tethered to the cell as membrane-bound receptors or (for VLRB only) secreted as soluble protein (see Figure 1-1) (Das et al. 2014; Herrin et al. 2008; Herrin and Cooper 2010; Jung et al. 2006; Kasahara and Sutoh 2014; Kasamatsu et al. 2010; Pancer et al. 2005; Saha et al. 2010). In comparison, jawed (gnathostome fish, reptiles, amphibians, birds and mammals) vertebrates evolved a separate triad of immunoglobulin superfamily (IgSF)-based lymphocyte antigen receptors comprised of somatically rearranged variable (V), diversity (D), and joining (J) gene segments to form the B cell receptors (BCR) of B cells, and either αβ or γδ T cell receptors (TCR) of T cells (see Figure 1-2). As in agnathans, gnathostome B and T cells express membrane-bound receptors or (for B cells only) secrete antibodies into the blood (Cooper and Alder 2006; Criscitiello and Flajnik 2007; Flajnik 2002; Hsu 2018; Kasahara et al. 1992; Lee et al. 2008; Rast et al. 1997; Rast and Litman 1994; Schatz 2004). In addition to B and T cells, the IgSF-based adaptive immune system is characterized by a polymorphic and polygenic major histocompatibility complex (MHC) Class I and Class II (Figure 1-3), recombination-activating gene (RAG)-mediated somatic recombination, and activation-induced cytidine deaminase (AID)- mediated receptor diversifying events. Both VLR-based and IgSF-based lymphocyte lineages form the necessary munitions that allowed vertebrates to survive the battle against invading pathogens (Cooper and Alder 2006; Litman et al. 2005; Ohta et al. 2019; Schluter SF et al. 1999).

Recent work by us and others suggests that some vertebrate groups developed a variety of mechanisms that augment these basic immune strategies. Here we explore the

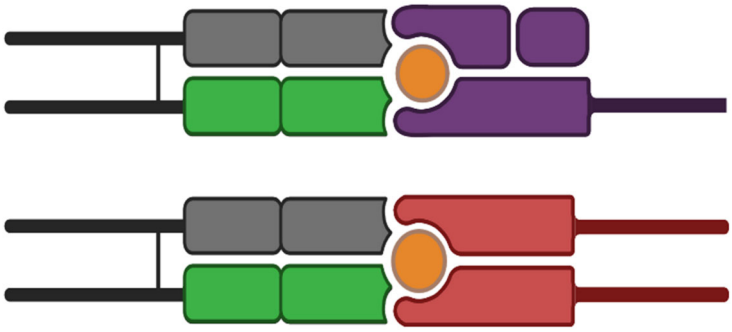

Figure 1-3. Major histocompatibility complex (MHC) Class II (top right) and MHC Class I (bottom right) in complex with peptide antigen, shown bound to an αβ TCR (left). [Figure created with BioRender.com]

conventional and the not-so-conventional antigen receptor components and diversification strategies of vertebrate adaptive immunity and the tremendous insight achieved by taking a comparative approach to studying immune mechanisms. We focus our review on the approaches employed by T cell receptors of nurse sharks, a member of the most divergent gnathostome group (Chondrichthyes) with the same fundamental, IgSF immune system components as humans.

1.1. The gnathostome immunoglobulin superfamily-based adaptive immune system

1.1.1. Receptor function is related to its structure

Most functional B cell receptors (BCR) are comprised of a heterodimer of two protein chains: a heavy chain (HC) and a light chain (LC). Each HC or LC is composed of a variable region that contains an antigen (Ag)-binding site and a constant (C) region that identifies the BCR isotype (see Figure 1-4) (Gellert 2002; Tonegawa 1983). Additionally, all

jawed vertebrates appear to have the four canonical T cell receptor (TCR) chains (α, β, γ, δ) and typically pair α chain with β chain to form αβ TCR and γ chain with δ chain to form γδ TCR (Figure 1-2). Both TCR types occur only as transmembrane proteins on the surface of T cells (Rast et al. 1997; Rast and Litman 1994). A BCR (or immunoglobulin, Ig) isotype is defined by its heavy chain and can occur as either a membrane-bound receptor or a secreted antibody (Ab) protein. In humans, there are five HC isotypes: Igμ (IgM), Igδ (IgD), Igγ (IgG), Igα (IgA), and Igε (IgE) (Murphy and Weaver 2017). Only two of the conventional isotypes discovered in gnathostomes are found in sharks, IgM and an IgD-like isotype called IgW (Ohta and Flajnik 2006; Zhu et al. 2012b).

Complete BCR and TCR variable region exons are formed by the rearrangement and assembly of variable (V), diversity (D), and joining (J) gene segments within a locus

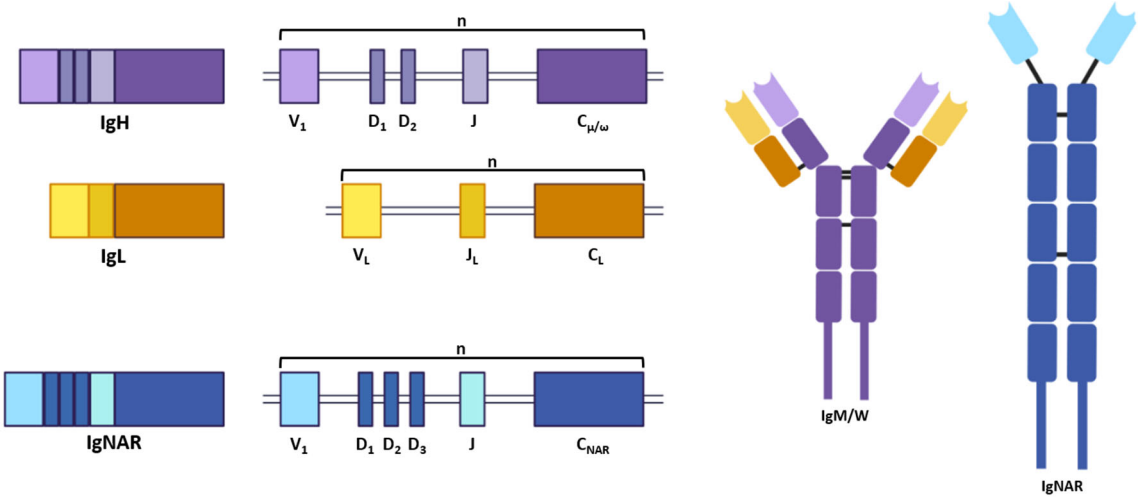

Figure 1-4. Immunoglobulin (Ig) heavy chain (top left), light chain (middle left), and new (or nurse shark) antigen receptor (NAR, bottom left) transcripts are formed by somatically rearranged variable (V), diversity (D), and joining (J) gene segments (middle) to form B cell receptors (IgM/W or IgNAR, right). [IgH: Ig heavy chain; IgL: Ig light chain] [Figure created with BioRender.com; adapted from Hsu 2016]

(see below). The V gene segment encodes three of the four framework regions (FR) and the first two complementarity-determining regions (CDR) of the assembled chain. The junction between V and J segments of LC or V, D and J segments of HC (called the V(D)J junction) encodes the third complementarity-determining region (CDR3) and the C-terminal part of the J gene segment forms the fourth framework region (see Figure 1-5A) (Gellert 2002; Lefranc 2014; Lefranc et al. 2003; Tonegawa 1983). Once assembled, each variable region chain folds to form a nine β-strand support structure (comprised of the framework regions) for the Ag binding loops (CDR) at the membrane-distal end of the receptor (Kikutani et al. 1986). In a complete TCR, Ag specificity is determined by the six CDR loops (three from TCRβ and three from TCRα) that form a single paratope (Figure 1-5B) (Jack and Du Pasquier 2019; Tonegawa 1983). These same six CDR loops form the Ag binding region in B cells, though the bivalent receptor can bind two antigen molecules simultaneously. While γδ T cells generally bind free Ag (in a manner similar to B cells), αβ T cells typically are restricted to binding peptide Ag in complex with the major histocompatibility complex (MHC, see Figure 1-3) (Jack and Du Pasquier 2019). other vertebrates illustrate contributions of TCR CDR loops to epitope binding. In a mouse αβ TCR structure (in complex with MHC Class I bound to self-peptide), peptide was bound in a diagonal direction from CDR1 to CDR3 loops of both α and β chains, and CDR1-α, CDR1-β, and CDR3-α simultaneously contact both peptide and MHC while CDR2 loops only contact α-helices of MHC (Garcia et al. 1998). However, comparison of this mouse complex to a similar model from humans (A6-HLA-A2) indicated that β chain CDR1 and

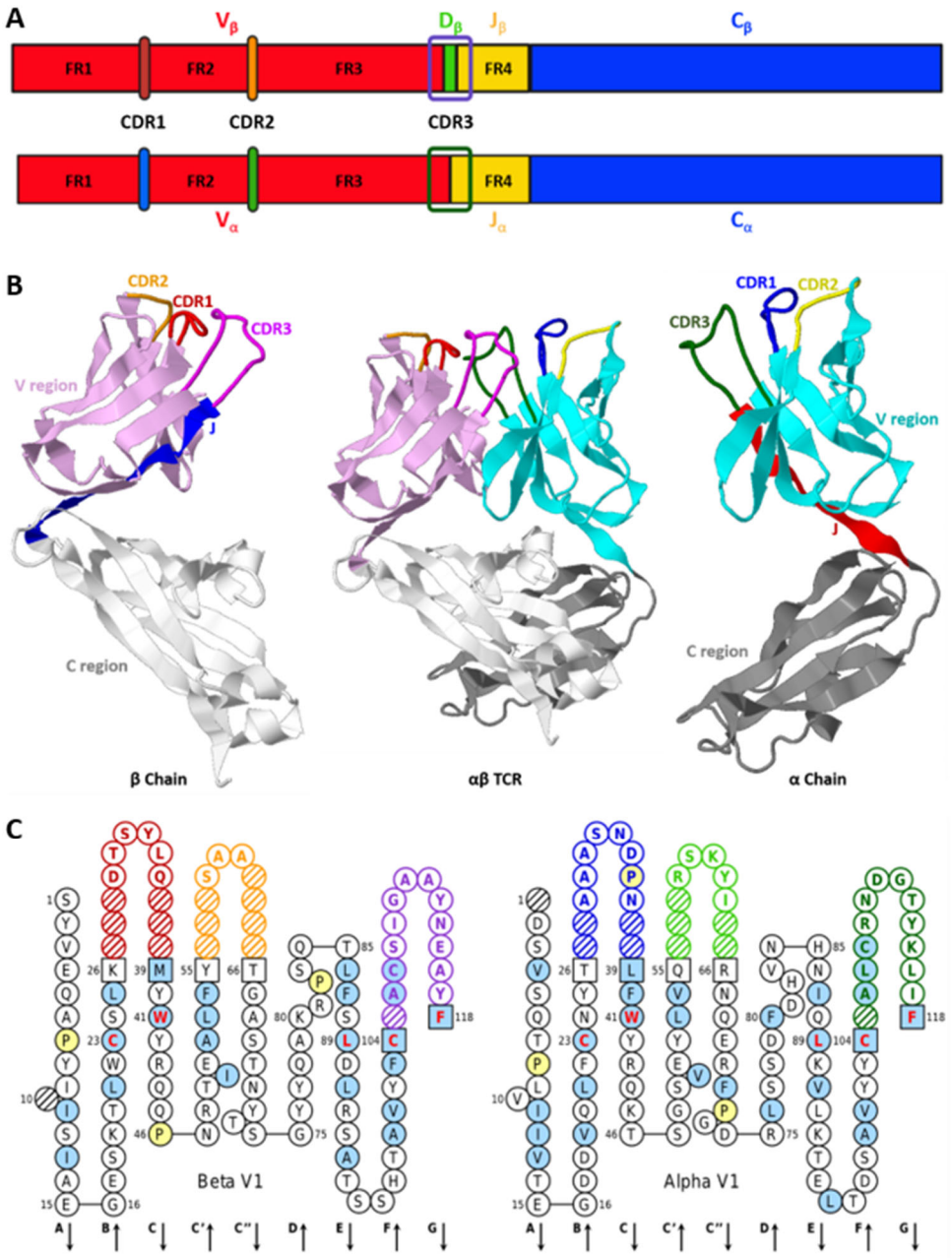

Figure 1-5. TCR variable region exons are formed by the rearrangement and assembly of variable (V), diversity (D), and joining (J) gene segments. [A] V gene segments encode three of the four framework regions (FR) and the first two complementarity-determining regions (CDR) of the assembled chain. The junction between V and J segments of alpha chain or V, D and J segments of beta chain encodes the third complementarity-determining region (CDR3). The C-terminal part of the J gene segment forms the fourth framework region. [B] 3-dimensional ribbon structure of β and α chain variable and constant regions. Nurse shark sequences modeled on a mouse αβ TCR (PDB model 4g9F). CDR binding loops colored as in the [C] 2-dimensional Collier de Perles (one layer) representation of the same amino acid sequences. [Figure 1A created with BioRender.com; Figure 1C created with IMGT Collier de Perles tool]

CDR2 make no contact with the MHC: peptide complex (Garcia et al. 1998). TCR repertoires generated to a single peptide antigen (A2/Melan-A tetramer or HIVgp160 peptide) demonstrated that TCRα was more important for antigen recognition than was TCRβ (Mantovani et al. 2002; Yokosuka et al. 2002). Orientation and recognition of MHC: peptide complexes may result from selection during T cell ultimately, CD8/CD4 lineage choice may create distinctive molecular constraints that enhance or hinder optimal binding (Buslepp et al. 2003). While Ag presentation by MHC is not essential for γδ T cells to bind Ag, crystalline structures of γδ T cells bound to non-classical MHC Class I demonstrate that MHC is recognized primarily by CDR3 of TCRδ (Adams et al. 2005; Allison and Garboczi 2002; Allison et al. 2001).

1.1.2. Lymphocytes rearrange complex loci to form highly diverse lymphocyte antigen receptors

During lymphocyte development in primary lymphoid tissues, both B and T cells employ recombination activating genes (RAG1/RAG2) to assemble variable regions from V, (D), and J gene segments. B cells develop within bone marrow or analogous primary tissue, such as epigonal or Leydig organs in sharks, while T cells develop within the thymus (Gellert 2002). Variable regions of Ig heavy chains (IgH) and TCR β and δ chains contain rearranged V, D, and J gene segments while those of Ig light chains (IgL) and TCR α and γ chains contain rearranged V and J gene segments only. Rearrangement is directed by recombination signal sequences (RSS) adjacent to each gene segment that guide RAG

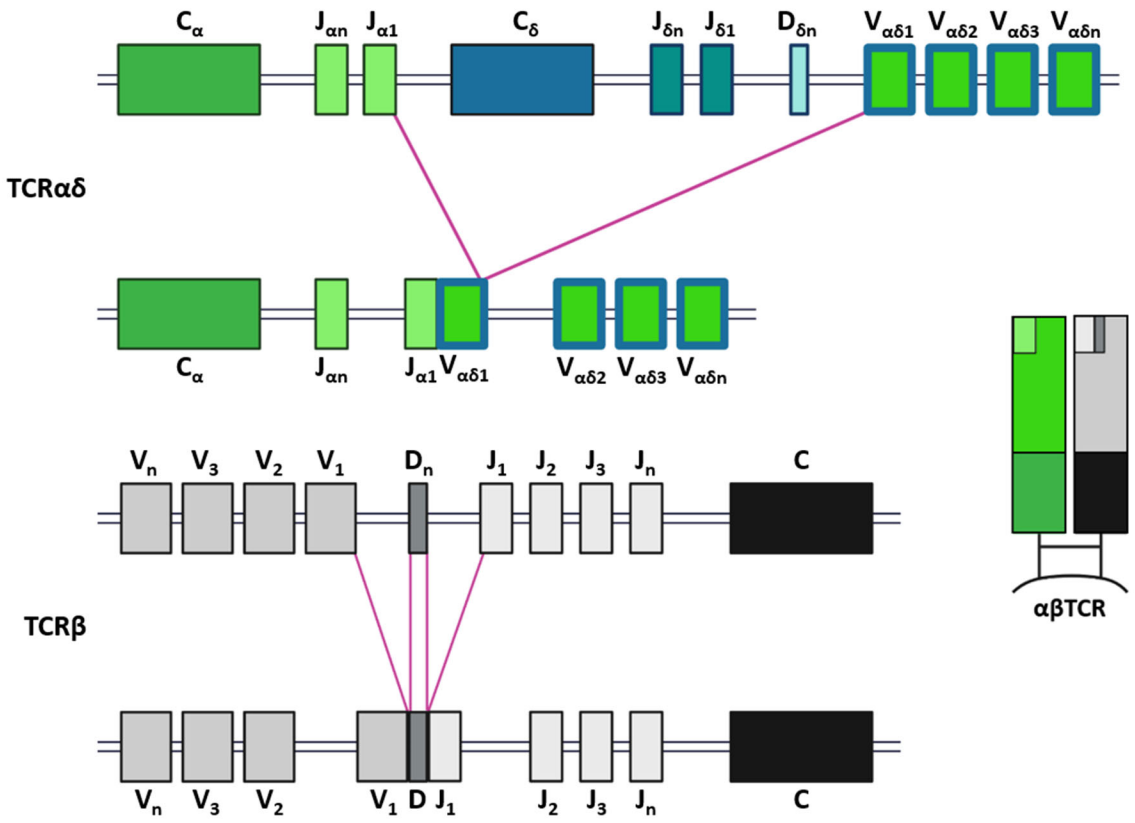

Figure 1-6. Alpha (α) and beta (β) chains are somatically rearranged to form an αβ T cell receptor (TCR). T cells recombine variable (V) and joining (J) gene segments within the αδ locus (top) or V, diversifying (D) and J gene segments within the β locus (bottom) to form complete VJ and VDJ exons. RNA splicing of beta constant region (C) forms a TCRβ chain. RNA splicing of alpha C forms a TCRα chain. Translation of both chains creates a complete TCR protein (right). [Figure created with BioRender.com]

binding to the correct location and gene segment (see Figure 1-6). An RSS is composed of a heptamer, a conserved block of seven base pairs (bp) that is contiguous with the coding sequence, followed by a 12 or 23bp non-conserved spacer, followed by a nonamer, a nine bp conserved block. Spacer lengths correspond to one turn (12bp) or two turns (23bp) of a DNA helix, bringing the necessary proteins together that catalyze recombination (Gellert 2002). Recombination is governed by the 12/23 rule, where gene segments flanked by a

12bp spacer RSS can recombine only with gene segments flanked by a 23bp spacer RSS. In this way, cells regulate the arrangement of gene segments during recombination. V segments typically contain a 23bp spacer RSS, while J segments have a 12bp spacer RSS, directing V-J recombination in IgL, TCRγ, and TCRα. Further, the coding region of D segments contain both a 12bp spacer RSS at the 5' end and a 23bp spacer RSS at the 3' end, guiding D-J and then V-DJ recombination in IgH, TCRβ, and TCRδ (Gellert 2002; Schatz et al. 1989). Recombination creates two new juxtapositions, joining the RSSs and the coding regions together. After RAG brings two gene segments together, it catalyzes double-strand breaks at each RSS between the heptamer and the coding sequence, ultimately producing two types of ends: signal and coding. The blunt signal ends are RSSs joined precisely to form signal joints. However, coding ends form sealed DNA hairpins that must be resolved before they can be joined together. Asymmetric hairpin cleavage generates single-stranded complementary (palindromic, P) nucleotides at the cleavage sites. The hanging DNA ends are either extended with random non-template (N) nucleotides by the enzyme terminal deoxynucleotidyl transferase (TdT) or removed by exonucleases, creating imprecise coding joints that contain added or deleted nucleotides. This V(D)J junction encodes the CDR3 that is typically an important site for Ag binding. The resulting variability within this join (called junctional diversity) results in a unique nucleotide signature in the resulting chain (Gellert 2002; Murphy and Weaver 2017; Swanson et al. 2009; Tonegawa 1983).

While ancestral TCR and Ig genes likely occurred within a single locus linked to

prototypic MHC genes within a pre-vertebrate primordial immune complex, these immune genes underwent duplication (via two rounds of genome-wide duplication) and translocation events to fashion the immune loci of extant vertebrates (Ohta et al. 2019). Presently, genes and gene segments of IgH and IgL chains and TCR β and TCR γ chains each are encoded by separate loci, while TCR δ is embedded within the TCR α locus. As a result, rearrangement of TCR α deletes the embedded TCR δ locus. In most jawed vertebrates, loci are organized as contiguous translocons, with numerous V, (D), and J gene segments preceding constant (C) region exons (V_n-D_n-J_n-C) that can stretch up to 3Mbp in length (Criscitiello and Flajnik 2007; Flajnik and Rumfelt 2000; Gellert 2002; Hsu 2018; Schatz 2004; Tonegawa 1983). For rearrangement to occur, DNA must undergo conformational changes that permit chromatin to fold and bring segments together (Jhunjhunwala et al. 2009). However, the loci of some organisms (e.g., shark IgH) are organized as multiple clusters of V, D, and J gene segments and C region exons $(V\text{-}D\text{-}J\text{-}C)_n$ that are closer together than in a translocon, making the sequential rearrangement of gene segments less necessary (Dooley and Flajnik 2006; Hsu 2009; Hsu 2018).

Research in mouse and human models demonstrates that both B cells and αβT cells rearrange and assemble Ag-binding receptors in similar ways during their development. Generally, IgH and TCRβ chain first combine D and J gene segments and then join V segments to the recombined DJ. The successful creation of a functional IgH or TCRβ chain halts RAG expression and gene rearrangement, and the cell undergoes a clonal expansion. Cells then express RAG again during rearrangement of V segments to J

segments in IgL and TCRα (Figure 1-6) (Bassing et al. 2002; Murphy and Weaver 2017). T cells simultaneously rearrange β, γ, and δ chains during development. While the mechanisms of gene rearrangement in γδ T cells are analogous to those of αβT cells, the timing of events differs. To simplify our discussion here, we first focus on rearrangement in BCR and αβ TCR and then consider temporal aspects of γδ TCR rearrangement later.

Rearrangement of IgH, IgL, and TCR γ, δ, and β chains is regulated, in part, by allelic exclusion, which (by definition) permits one allelic copy of a locus to be expressed at the surface of a cell, ensuring that each cell recognizes only a single ligand [reviewed in Brady et al. (2010)]. Locus rearrangement and expression of the first allele thus inhibits rearrangement of the second (Brady et al. 2010; Gascoigne and Alam 1999). The exception is TCRα, which can rearrange the loci of both alleles simultaneously. While allelic inclusion in TCRα can produce a small proportion of dual-expressing thymocytes, it also provides a cell more opportunities to find a functional α chain rearrangement. This issue often is resolved through post-translational mechanisms that prevent two TCRα chains being expressed simultaneously at the surface even if both chains are expressed in the cytoplasm (Gascoigne and Alam 1999). B and T cells that produce auto-reactive receptors or unsuccessful IgL or TCRα chains (respectively) can undergo receptor editing, rearranging both alleles of a locus multiple times until a productive arrangement is made or the cell undergoes apoptosis (Kuklina 2006; McGargill et al. 2000; Schatz 2004).

In addition to the junctional diversity created within V(D)J joins, lymphocyte receptors generate diversity through two types of combinatorial diversity. First, since

each locus typically contains multiple V, (D), and J gene segments each with distinct nucleotide sequences, cells can utilize different combinations of gene segments during rearrangement. Second, cells can pair different light chains together with the same heavy chain to form different Ag-binding regions of the receptor. Junctional and combinatorial diversity contribute to much of the variability between V regions (Gellert 2002; Tonegawa 1983). Somatic hypermutation, a process typically used by B cell receptor genes, further augments diversity within rearranged V regions by introducing mutations that can alter the binding affinity and specificity of receptors (see "affinity maturation" below). Together, these processes generate an efficient adaptive immune repertoire needed for response to infection.

1.1.3. T cell development and antigen receptor gene rearrangement occurs within specialized thymus tissue

T cell receptor gene rearrangement occurs as thymocytes develop within the thymus. In sharks, the thymus is bilaterally located dorsomedial to the gill arches and arranged as discrete lobules separated by trabeculae. Similar to the architecture of human thymus, each lobule consists of a large outer cortical region containing densely packed immature thymocytes and branched cortical epithelial cells and a smaller interior medullary region of loosely packed mature thymocytes, medullary epithelial cells, macrophages, and dendritic cells. The junction between the cortex and medulla is called the cortico-medullary junction (CMJ) and the outer region of the cortex is called the

subcapsular region (Luer et al. 1995; Murphy and Weaver 2017).

Stages of thymocyte development correlate with αβ TCR gene rearrangement and expression of key proteins on the T cell surface. As RAG mediates the V(D)J gene recombination events leading to lymphocyte receptor formation (see above), the products of these gene rearrangements in turn regulate RAG transcription. RAG expression occurs in two waves (corresponding to rearrangement of β chain and then α chain) and defines the T cell lymphopoietic stages (see Figure 1-7). In mammals, T-cell precursors migrate from bone marrow into the thymus through venules in the inner medulla and receive signals to undergo development as T cells, resulting in cells "double negative" (DN) for T cell co-receptor proteins CD8 and CD4. Immature T cells migrate through the CMJ to the outer cortex and then back towards the medulla in four sequential stages based on expression of cell surface markers CD44, a cell adhesion molecule, and CD25, the α chain of the IL-2 receptor. At the DN1 stage, cells express CD44 but not CD25, and both β and α loci occur in their germline configuration. As T cells migrate through the cortex to the subcapsular region, they begin to express CD25 and the β chain rearranges D to J segments (DN2) and then V to DJ (DN3). Rearranged β chains are expressed with a surrogate light chain (pre-Tα), and these pre-T-cell receptors are tested for functionality. Successful receptors cause ligand-independent signaling that downregulates RAG expression and halts rearrangement of the β chain (DN4). Cells undergo a proliferative burst and begin to express both CD8 and CD4, becoming double

positive (DP) thymocytes (Germain 2002; Kuo and Schlissel 2009; Murphy and Weaver 2017).

As DP thymocytes continue migrating towards the inner cortex, a second wave of RAG activity rearranges the α chain V to J gene segments. RAG expression continues to mediate rearrangement of the α locus until an MHC-compatible receptor is rearranged or the cell dies (which happens to the majority of thymocytes). DP cells that successfully recognize self-MHC Class I or Class II pass positive selection and mature to express either CD8 or CD4 (respectively), becoming CD8 or CD4 single positive (SP) thymocytes. Receptors also are tested against self-recognition (negative selection) during both DP and SP stages, eliminating cells that react to self Ag. In mice thymus, only about 2% of thymocytes survive selection mechanisms in the cortex to become mature T cells that enter the medulla and exit the thymus to form the peripheral T cell repertoire. Naïve αβ T cells are found primarily in secondary lymphoid organs and perform a vital role in the adaptive immune system (Germain 2002; Kuo and Schlissel 2009; Murphy and Weaver 2017). Thus, the cortex contains immature thymocytes actively rearranging and testing their receptor loci, and the medulla contains mature naïve CD8 or CD4 SP T cells post recombination and selection that are ready to leave the thymus.

In contrast to the MHC-restricted αβ T cells, both γ and δ chains of γδ T cells undergo receptor gene rearrangement simultaneously with β locus rearrangement during DN2/DN3 stages of T cell development. TCR signal strength during the DN3 stage instructs αβ or γδ T cell lineage fate, with strong signaling promoting the γδ T cell line while weak

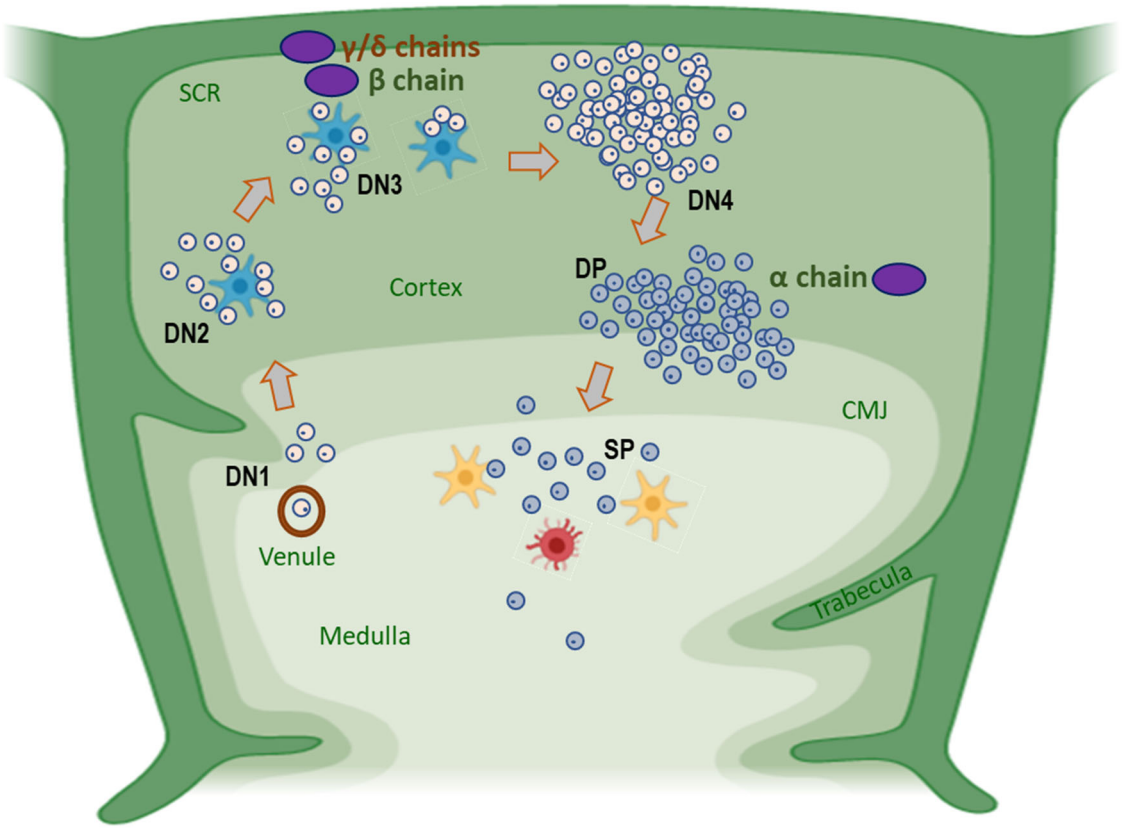

Figure 1-7. T cells develop as they migrate through the thymus. Thymocyte precursors enter through venules in the inner medulla near the cortico-medullary junction (CMJ) as double negative (DN) cells. As they differentiate through the four DN stages, they migrate to the outer cortex near the subcapsular region (SCR). Beta chain of αβ T cells and both gamma and delta chains of γδ T cells rearrange in the DN3 stage. Cells complete rearrangement at the DN4 stage and undergo cell proliferation. αβ T cells then rearrange alpha chain as they migrate back through the cortex, maturing to double positive (DP) cells that express both CD4 and CD8 co-receptor proteins. Finally, as thymocytes mature into single positive (SP) T cells, they express either CD4 or CD8 and eventually exit the thymus. Positive and negative selection of αβ T cells occurs during DP and SP stages. [Figure created with BioRender.com]

TCR signaling favoring commitment to the αβ T cell line (Fahl et al. 2014; Kreslavsky et al. 2010; Lafaille et al. 1990). In the DN3 stage, three of the four T cell loci (β, γ, and δ) have undergone rearrangement. Cells that successfully express TCRβ with pre-TCRα undergo proliferation, upregulation of CD4 and CD8 co-receptors, cessation of TCRγ

rearrangement, and rearrangement of TCRα chain, resulting in the deletion of TCRδ genetic components from the locus. Progression to the CD4/CD8 DP stage commits cells to the αβ T cell lineage (Kreslavsky et al. 2010). However, cells that rearrange successful TCRγ and TCRδ loci express γδ TCR at the surface, stimulating clonal proliferation but do not progress to the DP stage and thus emerge from the thymus committed to the γδ lineage (Kreslavsky et al. 2010). γδ T cells remain DN as mature thymocytes and do not express CD8 or CD4 co-receptors. Further, because γδ T cells are MHC-unrestricted, they likely do not undergo the same positive or negative selection processes during development as αβ T cells. Thus, while αβ T cells emerge from the thymus as mature naïve thymocytes, γδ T cells depart the thymus as fully mature cells capable of specific effector functions (Fahl et al. 2014; Lafaille et al. 1990). The "molecular switch" confirming lineage choice appears to be the ERK (extracellular signal regulated kinase)-Egr (Early growth response)-Id3 (inhibitor of differentiation 3) pathway, where strong signaling by γδ T cells stimulates ERK phosphorylation and induces Egr transcription and Id3 upregulation. Since are αβ T cell progenitors are generated only when Id3 expression is low, the upregulation caused by γδ T cell signaling disrupts development of the αβ lineage (Kreslavsky et al. 2010).

While some jawed vertebrates (ruminants, chickens, and likely sharks) have an abundance of γδ T cells, the majority of circulating T cells in mice and humans are of the αβ lineage (~95%) (Chien and Bonneville 2006; Telfer and Baldwin 2015). Murine γδ T cells appear early in embryonic development, and function as a first line of defense within

epithelial tissues lining the interior and exterior surfaces of the body (i.e., skin, lung, and intestines) (Kazen and Adams 2011; Xiong and Raulet 2007). T cells express temporal waves of invariant γδ receptors with limited junctional diversity in the γV or δV domain, with each wave programmed to appear at specific stages of development, binding ligands specifically expressed within a particular tissue (Xiong and Raulet 2007). Some γδ TCR bind lipid antigens presented by non-classical MHC molecules such as CD1d (CD1-restricted γδ T cells) (Castro et al. 2015). However, unrestricted γδ TCR typically bind free antigen in a manner similar to an antibody receptor (Chien et al. 1996). Inflammation stimulates activation of γδ T cells earlier in an immune response, releasing pro-inflammatory cytokines and killing infected macrophages. These antibody-like γδ T cells incorporate an innate response with an adaptive recognition strategy, providing both an immediate response to pathogen invasion and an ongoing (immunological memory), adaptive response to inflammation (Adams et al. 2005; Allison and Garboczi 2002; Allison et al. 2001).

1.1.4. Activation-induced cytidine deaminase optimizes lymphocyte receptors for antigen

While RAG mediates receptor gene recombination, activation-induced cytidine deaminase (AID) triggers BCR and antibody diversification through somatic hypermutation (SHM), class-switch recombination (CSR), and immunoglobulin gene conversion (IGC) events in humoral adaptive immunity (Arakawa et al. 2002; Muramatsu

et al. 2000). AID belongs to the much larger AID/APOBEC (apolipoprotein B RNA-editing catalytic component) family of zinc-dependent deaminases. While zinc-dependent deaminases are found in nearly all life forms on Earth (including bacteria, archaea, yeast, plants, and animals), the ancestral APOBEC emerged at the beginning of the vertebrate radiation, with the appearance of AID and APOBEC2 coinciding with the evolution of RAG-mediated adaptive immunity and the divergence of cartilaginous fish (Conticello et al. 2007). Of the major APOBEC family groups (AID, APOBEC1, APOBEC2, APOBEC3, and APOBEC4), only AID (found in all jawed vertebrates) and APOBEC3 (found only in placental mammals) function to mediate receptor diversification in immunity (Conticello 2008; Conticello et al. 2007; Conticello et al. 2005). AID targets the ssDNA that is opened up during transcription of Ig loci, catalyzing the deamination of cytidine to uridine within the variable regions of lymphocyte antigen receptors. The presence of uridine in DNA creates a mismatch between guanidine and uridine, which activates DNA repair mechanisms (i.e., mismatch repair, base-excision repair) to correct the mismatch. B cells are capable of manipulating these pathways so the repair is less effective at Ig loci, leading to the substitution of non-template bases at the affected site (Álvarez-Prado et al. 2018; Maul and Gearhart 2010). AID and other APOBEC deaminases catalyze the deamination of either DNA or mRNA within the nucleus. To protect other nuclear material from inadvertent targeting by AID overexpression, B cells actively transport transcribed AID to the cytoplasm where it accumulates, thereby sequestering it from both its intended and unintended targets (Ito et al. 2004). AID then is shuttled back to the nucleus

when required for SHM, CSR, or IGC events (Conticello 2008; Ito et al. 2004; Maul and Gearhart 2010; Muramatsu et al. 2007).

In T-cell dependent, antigen-driven immune responses of most jawed vertebrates, SHM is used to alter the affinity of BCR to Ag during a process called affinity maturation. After a naïve B cell is exposed to Ag, it is stimulated to proliferate within peripheral lymphoid tissues. In mammals and birds, activated B cells develop within B cell follicles of germinal centers (GC) within spleen, tonsils, and (in mammals) lymph nodes [(Good and Finstad 1966); reviewed in MacLennan (1994)]. However, reptiles, amphibians, and fish do not form GC, and B cells develop within lymphocyte-rich follicles of splenic white pulp or (in teleost fish) melanomacrophage clusters of liver and kidney (Magor 2015; Neely et al. 2018; Rios and Zimmerman 2015; Zapata et al. 1981; Zimmerman et al. 2010). During this proliferation event, AID catalyzes SHM within rearranged variable region gene exons to enhance binding affinity and specificity for the particular Ag that stimulated the cell (Conticello et al. 2005; Odegard and Schatz 2006). Affinity maturation occurs in a stepwise manner that repeatedly selects modified BCR with improved binding to the original Ag. Mutation is biased towards transitions and is targeted to particular motifs within variable region nucleotide sequences, focusing replacement mutation to particular hotspots of AID activity, particularly G and C residues within D<u>G</u>WY and WR<u>C</u>H motifs [where D is adenosine (A), guanosine (G), or thymidine (T); Y is cytosine (C) or T; W is A or T; and R is A or G]. Further, an abundance of these motifs within CDR concentrates mutation within the Ag-binding regions of the structure, thereby improving humoral

immunity (Álvarez-Prado et al. 2018; Muramatsu et al. 2000; Odegard and Schatz 2006; Saini and Hershberg 2015).

In some birds and mammals, V(D)J recombination creates limited combinatorial diversity due to a restricted availability or a preferential usage of HC and/or LC V gene segments. In rabbits, almost all BCR rearrangements (80-90%) utilize the same IgH V1 gene segment despite the presence of nearly 100 IgHV gene segments in the locus (Weinstein et al. 1994). However, many of these V segments are pseudogenes located upstream of the V1 gene segment. To offset the resulting limited diversity, young rabbits rely on immunological gene conversion (IGC) events within appendix germinal centers to diversify their primary Ig repertoire (Weinstein et al. 1994). Similarly, chickens construct their entire IgL (λ) repertoires from a single VJ rearrangement (IgH also is restricted), using IGC within the bursa of Fabricius (a primary lymphoid organ located in the bird hind gut) to augment diversity (Reynaud et al. 1987). During IGC, AID generates double strand breaks within the rearranged V(D)J sequence, and DNA repair mechanism transfer short "donor" DNA from upstream V gene segments into the rearranged sequence (Arakawa et al. 2002; Weinstein et al. 1994; Winstead et al. 1999). Rabbits further diversity their Ig repertoire by incorporating SHM within the V regions (Winstead et al. 1999). Sheep also form primary IgL repertoires from a limited pool of functional V and J gene segments but do not employ IGC to improve receptor diversity. Instead, sheep B cells diversify their primary repertoire with SHM during proliferation in ileal Peyer's patches (IPP), a primary gut-associated lymphoid tissue (GALT) composed of tightly packed B cell follicles and

small T cell zones. Patterns of SHM in IPP displayed the same bias towards replacement in CDR as affinity-matured BCR in secondary lymphoid tissues, especially at the V-J junction, suggesting selection for receptors with higher affinity towards antigen (Reynaud et al. 1991a). Thus, AID-catalyzed diversification mechanisms (e.g., IGC and SHM) are useful not only to modify BCR during secondary immune responses but, in some species, generate the primary antibody repertoire as well.

While AID-catalyzed SHM and IGC affect V region genes of both HC and LC, AID mediates CSR of C region genes for HC only. During CSR, AID induces nicks to both strands of dsDNA within donor and acceptor switch regions (long IGC-rich regions located in the introns between HC J segments and C regions). These dsDNA breaks trigger double-strand break repair mechanisms that join the two switch regions (and deleting the intervening exons), replacing the original C region with a different functional isotype (Honjo et al. 2002; Hwang et al. 2015). Thus, CSR results in a functional switch of antibody production without altering the receptor specificity achieved through affinity maturation. In nurse sharks, SHM and CSR differ in two important ways from these same events in humans or mice. First, nurse sharks do not form GC, and B cell SHM occurs within splenic white pulp, a well-defined aggregate of variably-sized lymphocytes within the spleen (Rumfelt et al. 2002; Rumfelt 2014). Second, sharks do not have classical switch regions like those found in tetrapods, though AID does mediate isotype switching between IgM and IgW in sharks (Hsu 2016). There is no evidence that IGC takes place in sharks.

1.2. APOBEC/AID and the evolution of lymphocyte receptors

1.2.1. The prototypic VLR-based adaptive immune system employs an AID-like mechanism

Agnathan vertebrates (lampreys and hagfishes) have an older, alternative type of adaptive immunity that depends on variable lymphocyte receptors (VLRs), somatically generated antigen receptors assembled from leucine-rich repeat (LRR) modules that are expressed clonally on lymphocytes (Pancer et al. 2004). Comparatively, VLRs have three lineages (A, B, and C) that resemble the B cell and T cell lineages of gnathostomes. VLR Type B (VLRB) can be membrane-bound or secreted and function in adaptive humoral responses, much like the BCR of jawed vertebrates (Alder et al. 2008). VLR Type A (VLRA) occur only as a membrane-bound receptors and function in cell-mediated immune responses analogous to αβ T cells (Kasamatsu et al. 2010). Finally, VLR Type C (VLRC), phylogenetically more similar to VLRA, are less diverse, found in limited tissues within the body, and transcriptionally are most similar to γδ T cells (Alder et al. 2008; Kasamatsu et al. 2010).

Lamprey and hagfish evolved homologs to the gnathostome AID that mediate the assembly of VLR genes into lymphocyte receptors in the absence of RAG. These cytidine deaminase (CDA) genes represent a more basal clade of mutators but share a common ancestor with the AID/APOBEC genes (Conticello 2008). In fact, lamprey CDA emerged phylogenetically as the closest sister group to the AID used by gnathostomes for SHM, GC, and CSR (Rogozin et al. 2007). Lamprey and hagfish lymphocytes express at least two

forms of CDA (CDA1/CDA2). CDA1 expression occurs selectively in VLRA (and likely VLRC) lymphocytes and likely orchestrates gene recombination of LRR cassettes into functional VLRA (and VLRC) within the thymoid (thymus like) region. In contrast, CDA2 expression occurs exclusively in VLRB lymphocytes and thus likely mediates VLRB gene assembly (Marshall et al. 1999). CDA-mediated gene rearrangement in lampreys occurs in a manner similar to AID-induced immunoglobulin (Ig) gene conversion in some birds and mammals (Deng et al. 2010a; Flajnik 2014; Rogozin et al. 2007; Zheng et al. 1994).

Based on the similarities between these two adaptive immune strategies, it is possible that the common ancestor of modern vertebrates also exploited an APOBEC- (or similar zinc-) family deaminase for lymphocyte receptor development prior to the evolution of CDA- or RAG-mediated gene rearrangement. Perhaps from this common ancestor, agnathans evolved specific APOBEC molecules for diversification of their B and T like VLRs, while gnathostomes evolved AID for T cell primary repertoire diversification and B cell affinity maturation, eventually co-opting AID for use in GC, SHM, and CSR for primary B cell repertoires.

1.3. Unconventional T cell receptors and diversifying mechanisms used by vertebrates

In addition to the canonical B and T cell receptors found in all jawed vertebrates, some vertebrates construct receptors that utilize conventional BCR components to generate unconventional TCR. In fact, T cells appear adept at creating novel, diverse receptors by capitalizing on the accessibility of IgH, IgH-like, and TCR V gene segments available to them. Whether these segments are housed within a conventional IgH locus or

in separate, distinct loci (trans-locus or trans-chromosomal rearrangements), or contained within the TCRαδ locus itself (cis-locus rearrangements), they are indistinguishable from those used by B cells to form BCR HC. T cells can further exploit traditional B cell diversifying mechanisms to expand their receptor repertoires, utilizing SHM to alter paratopes. Here we discuss three novel TCR types created by recombining distinctly Ig or Ig-like V gene segments with TCR constant regions: 1.) IgH-like TCRδ V (VHδ), 2.) IgHV and TCRδ-associated Ig-like V (TAILV) rearranged to TCRδ (or rarely, TCRα), and 3.) receptor chains with double V domains (TCRμ and NAR-TCR). We end with a discussion of SHM as a repertoire diversifying mechanism in T cells.

1.3.1. VHδ gene segments found in all jawed vertebrates except teleosts and eutherian mammals

Functional Ig-like TCRδ V (VHδ) gene segments have been found in genomes of all gnathostome groups studied except teleosts and placental mammals. The coelacanth TCRαδ locus includes a track of 25 VHδ gene segments between the TCRα and TCRδ genes (Saha et al. 2014). In the frog *Xenopus tropicalis*, the 5' end of the conventional TCRαδ locus encodes a separate cluster of VHδ gene segments that are expressed exclusively with a second distinct TCR δC (Parra et al. 2010). At least some birds express VHδ gene segments with TCRδ as well. In the passerine zebra finch, a single VHδ gene segment is present in the TCRαδ locus and is expressed with TCR δC. However, galliform birds (chicken, turkey, and likely duck) have a second non-syntenic TCRδ locus containing a

single VHδ-Dδ-Jδ-Cδ cluster that rearranges to form one TCRδ product. The conventional TCRαδ locus contains no VHδ segment (Parra et al. 2012b). The only mammal known to have functional VHδ gene segments is the monotreme platypus, which has a single VHδ gene segment located within the TCRαδ locus (Parra et al. 2012a). However, our lab located a single VHδ pseudogene in the TCRαδ locus of Florida manatee (Breaux et al. 2018).

1.3.2. Ig-TCR chimeric isoforms enhance diversity of T cell receptor repertoire in nurse sharks

Additionally, some vertebrates construct distinct Ig-TCR chimeric isoforms from an Ig or Ig-like V gene segment rearranged to a TCR C region. Criscitiello et al. (2010) first identified unusual transcripts in nurse sharks that recombine IgM or IgW V gene segments to TCRδ (or rarely, TCRα) C regions. The IgH V gene segments used by TCR are genetically indistinguishable from those used by BCR and consequently, presumed to be from the conventional Ig locus (Criscitiello et al. 2010). However, the lack of an assembled genome or complete Ig/TCR loci in nurse shark complicates our complete understanding of the genomic origin of these IgHV gene segments, and whether IgHV associated with TCR are located within the conventional TCRαδ locus (cis-chromosomal rearrangements), the conventional Ig locus (trans-locus rearrangements), or in a separate locus altogether ("*trans*" rearrangements) remains unclear (Criscitiello et al. 2010; Deiss et al. 2019). Partial assembly of the TCRδ locus uncovered unique Ig-like V gene segments nestled

within the TCRαδ translocon that show the greatest identity to IgMV, and mRNA transcripts indicate these V gene segments (termed TCRδ-associated Ig-like V, or TAILV) recombine successfully with TCR Cδ regions but not BCR C regions (Deiss et al. 2019). The presence of Ig-like TAILV within the TCRαδ locus of nurse sharks (and VHδ gene segments in a number of vertebrate lineages) suggests that T cells likely have been assimilating both Ig and TCR V gene segments into functional TCR since the genesis of the IgSF-family based adaptive immune system.

1.3.3. Complex receptors evolved convergently in sharks and mammals

Perhaps the most complex TCR isoform in sharks is the doubly-rearranging NARTCR, composed of two V domains (that undergo separate VDJ recombination events) and a TCRδ C domain (Criscitiello et al. 2006). The membrane-distal V domain (NAR V) is closely related to IgNAR (variably called “nurse shark antigen receptor” and “new antigen receptor”), a distinct IgH isotype found only in cartilaginous fish that does not associate with light chain (Criscitiello et al. 2006; Greenberg et al. 1995). The NAR V domain is supported by a membrane-proximal TCRδ V domain that is assembled from a distinct cluster of TCRδ V segments (Criscitiello et al. 2006). A draft assembly of the TCRαδ locus identified blocks of NARTCR V, D, and J gene segments located in a separate stretch of the TCRαδ translocon from the canonical TCRδ VDJ gene segments, with each NARTCR VDJ block located upstream of an apparently dedicated supporting TCRδ V gene segment (Deiss et al. 2019). NAR-TCR likely partners with TCRγ chain, likely forming an MHC-

unrestricted receptor with a protruding NAR V domain that sits atop a base formed by the γ and δ TCR chains, with only the NAR V CDRs constructing the antigen-binding site of the receptor (Criscitiello et al. 2006).

The discovery of a unique TCR locus (TCRμ) in monotreme and marsupial mammals further muddles the distinction between B and T cell receptor components (Parra et al. 2007; Wang et al. 2011). In opossum (*Monodelphis domestica*), the TCRμ locus is located on a separate chromosome from conventional TCR loci and is atypically organized as tandem clusters of Vμ, Dμ, and Jμ gene segments followed by a Cμ exon (Parra et al. 2007). In addition, an exon encoding a complete V domain, with rearranged VDJ gene segments already joined together in germline DNA ($V\mu_j$), sits between the Jμ and Cμ of each cluster. TCRμ expresses two functional transmembrane isoforms. The short form, TCRμ1.0, encodes a receptor chain composed of a single $V\mu_j$ domain and Cμ, forming an invariant binding site that is structurally more similar to conventional TCR (Parra et al. 2007). The long form (TCRμ2.0, the dominant isoform in peripheral lymphoid tissue) encodes a receptor chain containing two V domains and Cμ and is structurally analogous to the NARTCR of sharks (Parra et al. 2007). The membrane-distal V of TCRμ2.0 is formed by RAG-recombined V, D, and J gene segments that incorporates junctional diversity within the V domain, whereas the membrane-proximal V is always a (pre-joined) $V\mu_j$ exon that forms an invariant V domain (Parra et al. 2007). The two V domains are linked through a mRNA splice site in the $V\mu_j$ leader sequence that splices the recombined VDJ of the membrane-distal V to framework 1 of the membrane-proximal V (Parra et al.

2007). Phylogenetically, V gene segments of TCRμ2.0 membrane-distal V domain and the sequence corresponding to the V gene segments (FR1 through FR3) of $V\mu_j$ (in both isoforms) are closely related to Ig heavy-chain V gene segments (VH) while Cμ likely derived from a TCRδ ancestor (Parra et al. 2008; Parra et al. 2007).

Like that of opossum, the platypus (*Ornithorhyncus anatinus*) TCRμ locus also occurs in a separate location from conventional TCR genes, but platypus express a single TCRμ isoform composed of two V domains that each somatically rearrange V, D, and J gene segments (Wang et al. 2011). The V1 domains encode longer and more junctionally diverse CDR3 because they rearrange two to four Dμ gene segments and add non-template (N) nucleotides during assembly. However, while V2 domains incorporate both palindromic (P) and N nucleotide additions, they do not appear to use Dμ gene segments, likely because the locus encoding the V2 domain lacks D segments (Wang et al. 2011). As in opossum TCRμ, both V1 and V2 domains of platypus TCRμ are more similar to VH while Cμ is related to TCRδ (Wang et al. 2011).

1.3.4. Somatic hypermutation augments T cell receptor repertoire diversity in sharks and camelids

In addition to capitalizing on the availability of Ig V gene segments, T cells also can exploit traditional B cell diversifying mechanisms to expand their receptor repertoires. One such mechanism is the use of AID-catalyzed SHM to augment TCR repertoire diversity. Chen et al. (2009) reported the first evidence of targeted mutation to TCRγ V

regions in the sandbar shark (*Carcharhinus plumbeus*). The authors first sequenced the TCRγ locus and then evaluated the V region repertoire diversity using a 5' RACE library from a single animal. Typical of TCR loci in many other vertebrates, sandbar shark TCRγ is arranged as a single translocon containing at least five V gene segments, three J gene segments, and a single C region. Expressed transcripts revealed no V segment bias for four of the five known Vs but a reduction in the use of the most 5' (distal) V segment in the locus (Chen et al. 2009). However, comparison of cDNA clones to genomic sequences revealed a high frequency of mutation that could not be attributed to allelic variation or PCR error. Mutation was targeted to AID hotspot motifs within CDR of V segments (specifically CDR1), was biased towards AID-favored G and C nucleotides, resulted in more transition than transversion changes, and included both single base and consecutive (tandem) base changes that favored amino acid replacement (R), patterns that mirror SHM during affinity maturation of activated B cells (Chen et al. 2012; Chen et al. 2009). Further, mutation frequency (0.018/bp) was comparable to that observed in Ig LC of mice (0.016/bp) and sharks (0.015/bp), suggesting that base changes resulted from AID-mediated SHM within the V region (Chen et al. 2012). However, because there was no evidence of antigen selection for mutated receptors [CDR and FR showed similar ratios of replacement (R) and silent (S) changes], Chen et al. (2012) concluded that TCRγ instead utilizes SHM to enhance repertoire diversity in γδ T cells. Antigen also does not drive selection in nurse shark IgL chains except by limiting mutation to FR2, suggesting a mechanism for maintaining structural stability rather than enhanced affinity (Zhu and Hsu

2010).

Similar analyses in both γ and δ chain of dromedary camel (*Camelus dromedaries*) indicated that mutation altered γδ TCR of camelids (Antonacci et al. 2011; Ciccarese et al. 2014; Vaccarelli et al. 2012). Using RT-qPCR and a 5' RACE library, Antonacci et al. (2011) evaluated the expressed TCRδ chain repertoire of peripheral lymphoid tissues (spleen, tonsils, and blood) from a single adult camel. These transcripts were used to identify genes encoding TCRδ V gene segments in the germline. Analyses identified 13 putative germline TCRδ V gene segments belonging to three family groups. Comparing these germline sequences to cDNA clones revealed mutations to V regions at a rate (0.013/bp in spleen) similar to those reported in sandbar shark TCRγ and in mouse and shark Ig LC (see above). However, although nucleotide changes did appear to favor transitions (and included both point and tandem base changes in spleen), mutation did not target CDR over FR but instead was distributed throughout the V region (Antonacci et al. 2011). Comparison of synonymous and nonsynonymous (replacement) changes suggested (like in sandbar shark) that mutated receptors were not under antigen selection. The authors did not report specific analyses to examine whether mutation was AID mediated (e.g., bias to AID-favored G and C bases or targeted mutation to AID hotspot motifs), but concluded that mutation in TCRδ chain did contribute to γδTCR repertoire diversity (Antonacci et al. 2011).

In a follow-up study in camelids, the same group reported evidence that mutations to genes encoding TCRγ chain also may generate diversity within the γδTCR

repertoire (Vaccarelli et al. 2012). The group assembled and mapped the TCRγ locus from PCR products and chromosome walking fragments to identify two V-J-J-C cassettes within the locus. While a cluster organization is atypical for TCR loci in general, this same basic cassette (V-J-J-C) structure is found in the TCRγ locus of a number of organisms (including sheep, cattle, and buffalo) and modifications to this structure are found in mice (Antonacci et al. 2007; Vaccarelli et al. 2008; Vernooij et al. 1993). An analysis of expressed transcripts from a spleen 5' RACE library revealed targeted mutation biased towards G and C bases within AID-favored hotspot motifs. Further, although there was (again) no evidence of selection for modified receptors, the accumulation of nonconservative changes within CDR (specifically CDR2) intimated that somatic mutation contributed to the overall paratope diversity of TCRγ V regions (Vaccarelli et al. 2012). *In silico* structural models indicated that mutation to γ or δ chain enhances the structural stability of the γδ TCR, regardless of where (FR or CDR) these mutational changes occur within the V region (Ciccarese et al. 2014).

The presence of mutation within γδ TCR genes is not altogether surprising given the ability of γδ T cells to traverse the boundary between the innate and adaptive immune systems. Similar to αβ T cells, γδ T cells recombine V, (D), and J gene segments to create a highly specific adaptive repertoire with immunological memory (Kazen and Adams 2011). However, γδ T cells can assert an innate role in immunity as well, producing cytokines (e.g., TNFα and IFN-γ) in response to infection or tumor antigens (Beetz et al. 2008; Gober et al. 2003). In humans, γδ T cells can act as efficient antigen-presenting cells

to $CD8^+$ αβ T cells, synthesizing antigens through immunoproteasomes for cross-presentation via MHC class I (Brandes et al. 2009). Additionally, specific subsets of γδ T cells in humans (Vδ2 T_{regs}) express FOXP3 (forkhead/winged helix transcription factor box P3) and function as regulatory T cells, suppressing proliferation of peripheral blood mononuclear cells through the TGF-β1 signaling pathway (Casetti et al. 2009). Thus, γδ T cells combine both immediate innate-like responses to infection with on-going adaptive recognition responses [also reviewed in Kabelitz (2011)]. While some γδ TCR bind free antigen in a manner similar to BCR, some γδ TCR interact with non-classical MHC as tissue-specific receptors using restricted sets of variable and joining genes with limited junctional diversity (Adams et al. 2005; Allison and Garboczi 2002; Kazen and Adams 2011). In either case, SHM-mediated changes to paratopes could offer receptors the flexibility to recognize new pathogens or adapt to rapidly changing ligands within restricted environments.

While it is clear that T cells retain the same basic machinery that allows B cells to affinity mature receptors (Gellert 2002), somatically mutating αβ TCR may not provide the same benefits as to BCR or γδ TCR. Because αβ T cells are restricted to binding antigen in the context of self MHC, altering receptors that already have passed selection in the thymus could have profound consequences to receptor functionality. In a study striving to identify targeting elements of SHM in mice, Hackett et al. (1992) designed a rearranged TCR transgene capable of being expressed on B cells. The authors then examined cDNA transcripts of both endogenous IgH and TCR transgenes expressed on B

cells to determine if TCR are targeted by SHM. Though they did observe some mutation (0.00017/bp) in the TCR transgenes, the frequency of mutation was minimal compared to rates observed in endogenous IgH genes (0.0021/bp), suggesting that TCR genes do not contain the required transcriptional elements for SHM (Hackett et al. 1992).

T cells of B10.A transgenic mice that recognize pigeon cytochrome *c* (PCC) typically express TCR encoded by Vβ3 and Vα11 gene segments (Zheng et al. 1994). Once activated by PCC, $CD4^+$ T cells rapidly increase in number within the periarteriolar T-cell sheath (PALS) of mouse spleen until cells migrate to other sites (including GC). Analysis of T cells from PALS and GC of immunized mice revealed mutation to V regions of TCR α chain (but not β chain) that was substantially higher than expected for PCR error. Further, mutation to TCRα V mirrored that of IgH V acquired from adjacent sites within the GC, suggesting a mechanism for SHM in T cells (Zheng et al. 1994). The significance of these results was questioned, citing insufficient evidence to support the claim (Bachl and Wabl 1995). However, the mutation may suggest that AID expression within splenic GC (during affinity maturation of B cells) also can impact V regions of TCRα.

In another study using Cre-ires-hCD2 (Cre) transgenic mice with a genetic reporter knocked into the AID locus, Qin et al. (2011) assessed endogenous AID production by B and T cells within spleen, lymph nodes and Peyer's patches. The authors found that a surprisingly large number of $CD4^+$ memory T cells in these tissues express AID, likely resulting from T cell activation in peripheral lymphoid tissues. Activation of these T cell subsets produced a unique cytokine profile that increased with mouse age, suggesting a

function in cellular aging. While they made no attempt to examine cDNA transcripts for evidence of mutation, Qin et al. (2015) suggested that AID may play a role in T cell function or tumorigenesis.

Other reports of SHM in αβ T cells suggest it occurs as a result of a diseased state. For example, SHM of TCR from alloreactive T-cell hybridomas alter specificity to MHC Class II alleles (Augustin and Sim 1984). Additionally, researchers found that TCR Vα and TCR Vβ genes from splenic white pulp of HIV-1 positive patients somatically mutate in a manner similar to IgV H genes during BCR affinity maturation (Cheynier et al. 1998). Finally, overexpression of AID in transgenic mice results in T cell lymphomas, lung and liver tumors, or B cell lymphoma (Morisawa et al. 2008; Okazaki et al. 2003; Rucci et al. 2006). Thus, while AID-mediated mutation may benefit certain populations of T cells, it is clear that mutation is not always beneficial.

While we found no study that specifically assesses the presence or absence of SHM in endogenous TCR, the fact that SHM is not commonly observed in mice or humans (except in diseased states) led immunologists to assume that αβ T cells cannot utilize SHM or any other receptor-modifying mechanism (Kronenberg et al. 1986; Vitetta et al. 1991). However, anecdotal evidence from our lab indicated that α chain of αβ T cells in nurse sharks may use SHM in a manner similar to γδ T cells. Further, preliminary analyses of AID hotspot motifs within CDR and FR of human and nurse shark TCR Vα genes suggests that Vα of nurse sharks contain more AID-preferred motifs (WRCH/DGYW) per sequence than do human Vα segments, and these motifs occur 2-3x more often in shark

Vα CDR than in human CDR. This suggests that, while the costs associated with somatically mutating TCR genes may outweigh the benefits for humans and mice, the same may not be true for more evolutionarily basal organisms like sharks. Sharks may be more resistant to the dangers of aberrant mutation because of their inherently slow rates of molecular mutation (10x slower than in mammals), long lifespans (>272 years in Greenland shark), and (in many species) large body size (Marra et al. 2019; Martin 1999; Nielsen et al. 2016). Additionally, because of their considerable size and highly repetitive nature (>50%), shark genomes may exhibit more flexibility than those of mice or humans (Hara et al. 2018; Rocco et al. 2007; Rocco et al. 2002; Stingo and Rocco 2001). To realize any benefit of SHM, TCR modification would have to occur prior to or coincident with selection events in the thymus, since changes to a receptor that already passed selection could negatively affect its ability to bind self-MHC or permit binding to self-antigen.

1.3.5. Dissertation aims

To begin to understand the potential evolutionary role of AID in TCR repertoire development, we first must determine the extent to which V regions of TCR loci are impacted by mutation within both thymus and peripheral lymphoid tissues. If mutation is catalyzed by AID, it follows that nurse shark thymus must express AID within regions of the thymus where receptor rearrangement takes place. Because TCRβ rearranges first (in mammalian αβ T cells) and successful rearrangements inhibit further rearrangement of β chain genes, it is unlikely that a T cell could later salvage a β chain that receives

detrimental mutations. However, TCRα retains the ability to rearrange gene segments of both alleles until a successful TCRα chain is generated. Thus, we hypothesize that 1) V regions of TCRα should mutate at higher rates than TCRβ and 2) AID expression should be greatest within the inner cortex, corticomedullary junction, and inner medulla of thymus tissue where TCRα chain is rearranged. We also predict that mutation to other canonical T cell chains (β, γ, δ) and to non-canonical receptor chains (NAR-TCR, IgHV-TCRC, TAILV-TCRC) occurs as a byproduct of AID transcription during thymocyte development.

In chapter 2, we evaluate the hypothesis that AID is actively expressed within nurse shark thymus and mediates SHM of TCRα chain to improve repertoire diversity of αβ T cells. In chapter 3, we examine whether other canonical or non-canonical TCR chains utilize AID for receptor diversification and explore possible mechanisms that might regulate these AID-mediated diversification strategies.

2. SOMATIC HYPERMUTATION OF T CELL RECEPTOR α CHAIN CONTRIBUTES TO SELECTION IN NURSE SHARK THYMUS*

2.1. Introduction

All jawed vertebrates share fundamental components of the adaptive immune system. The cartilaginous fish (including sharks) are the most divergent jawed vertebrate group relative to mammals and use a polymorphic major histocompatibility complex (MHC) (Kasahara et al. 1992), multiple isotypes of immunoglobulin (Ig) heavy and light chains (Criscitiello and Flajnik 2007; Flajnik 2002), and the typical four T cell receptor (TCR) chains (Rast et al. 1997). Shark lymphocyte antigen receptors are diversified by RAG-mediated V(D)J somatic rearrangement (Bernstein et al. 1994). After antigen exposure, B cells also use the enzyme activation-induced cytidine deaminase (AID) for receptor modification via somatic hypermutation (SHM) (Conticello et al. 2005), allowing activated B cells to extensively alter their rearranged Ig variable region genes (Muramatsu et al. 2000). Some of the variants produced by this process bind antigen with higher affinity, enhancing humoral immunity through affinity maturation. In addition to SHM in all jawed vertebrate Ig, AID catalyzes the processes of heavy chain class switch recombination (CSR) in tetrapods (and is implicated in shark CSR (Zhu et al. 2012a) and Ig gene conversion (in birds and some mammals) (Barreto and Magor 2011). AID is a

---

* Reprinted from Ott JA, Castro CD, Deiss TC, Ohta Y, Flajnik MF, Criscitiello MF (2018). "Somatic hypermutation of T cell receptor α chain contributes to selection in nurse shark thymus." *eLife* 7: e28477. Copyright 2018 by Ott *et al*. [https://elifesciences.org/articles/28477#info]

member of the APOBEC family of nucleic acid mutators, two of which likely diversify the variable lymphocyte receptor (VLR) system in the more ancient vertebrate lineages of lamprey and hagfish (Alder et al. 2005; Guo et al. 2009).

Although in general immunologists think TCR loci do not undergo somatic hypermutation, a few reports do exist of AID-mediated SHM in T cells. However, non-productive TCRα rearrangements in hybridomas (Marshall et al. 1999), TCRβ sequences from HIV-positive individuals (Cheynier et al. 1998), and reports of SHM in TCRα murine germinal center T cells (Zheng et al. 1994) are not thought to describe any normal physiology (Bachl and Wabl 1995). More recent studies indicate that TCR δ and γ in the dromedary camel and TCR γ in the sandbar shark somatically hypermutate (see below) (Antonacci et al. 2011; Chen et al. 2012). Despite these findings, the general consensus has remained that AID does not target TCR loci (Choudhary et al. 2018; Pavri and Nussenzweig 2011). In over thirty years of studies of TCR repertoires, it has been clear that SHM is not functioning to either generate or further enhance the TCR repertoire of mouse and human.

Recent studies in the sandbar shark (*Carcharhinus plumbeus*) revived the notion of SHM at TCR loci. Sequencing of the entire TCRγ translocon in *C. plumbeus* showed definitively that SHM is occurring at that locus (Chen et al. 2009). Shark TCRγ SHM occurs in two distinct patterns: point mutations and tandem mutations characteristic of B cell SHM in cartilaginous fish (Anderson et al. 1995; Lee et al. 2002; Rumfelt et al. 2002; Zhu et al. 2012a), possibly suggesting two different cellular mechanisms for generating

mutations (Chen et al. 2012). The sandbar shark analysis found targeted nucleotide motifs of AID activity at the TCRγ locus. Chen et al. (2012) examined ratios of replacement (R) and silent (S) mutations between CDR and framework regions to determine if mutation altered affinity of receptors, a method commonly used to study B cell affinity maturation by SHM. Finding no difference between R/S ratios in CDR versus framework regions, they concluded that TCRγ uses SHM to generate a more diverse repertoire rather than for affinity maturation. SHM-induced changes to TCRδ in camels showed similar results.

Early work in our lab also suggested that SHM occurs in the less restricted γδ T cells in nurse shark (*Ginglymostoma cirratum*) and perhaps in the alpha chain of MHC-restricted αβ T cells (Criscitiello et al. 2010), encouraging us to examine this phenomenon further. Thus, we performed a systematic analysis of this process in shark primary and secondary lymphoid tissues using thymocyte clones containing the same unique third complementarity-determining region (CDR3). Our data suggest that SHM of TCRα is involved in primary T cell repertoire diversification and the enhancement of positive or negative selection in the thymic cortex. This finding is consistent with a model put forward by Niels Jerne over 40 years ago to explain antigen receptor positive selection (Jerne 1971).

2.2. Results

2.2.1. Somatic hypermutation in TCR γV and TCR δV

We assessed the presence of possible SHM within TCR V segments from γ, δ, and β chains. Using neighbor-joining consensus trees, we grouped sequences from each chain into V families based on 85% nucleotide identity and into V subfamilies based on 90% nucleotide identity. We then examined each V subfamily for the possibility of SHM. However, since none of the sequences from these chains contained CDR3 regions, we did not rigorously analyze mutations within these chains.

We first corroborated the original finding of SHM in TCRγ variable regions (V) in sandbar shark spleen (Chen et al. 2012; Chen et al. 2009) using peripheral lymphoid tissue from the spiral valve (intestine) of nurse shark. We also examined mutation in clones from the thymus. We conservatively assigned 69 sequences to nine V genes from three TCRγ families (Figure 2-1). Even with a conservative assignment of clones to predicted germline V sequences, we found nearly twice as many V genes as in the sandbar shark locus (which only contains five Vs). Only TCR γ4 did not display mutations in the nurse shark, but since we found only two γ4 sequences, it is possible that it occurs but our sample was too small to observe it.

We then investigated the possibility of mutation occurring at the TCRδ locus. We analyzed mutation in clones from nurse shark thymus, peripheral blood leukocytes, and spiral valve, conservatively grouping 111 clones into 12 V genes from seven TCRδ families (Figure 2-2). Only one of these families (TCR δ10) lacked mutation. We found that in five

of the seven δV families, the same V-δ segments used to generate δ cDNA sequences also generated α cDNA sequences. The sequence diversity at TCRγ and TCRδ was in contrast to TCRβ, where we found no such evidence for mutation in 56 sequences representing six V segments from six different V families (Figure 2-3). Limited existing data also do not support mutation at NAR-TCR, a distinct TCR containing a NAR V domain supported by a more canonical V δ domain, each resulting from independent V(D)J rearrangements (data not shown) (Criscitiello et al. 2006).

2.2.2. Identification of TCR Vα Genes in the nurse shark genome

We identified 17 germline α/δ V gene sequences corresponding to 12 unique V segments. Unfortunately, these segments matched only two groups of sequences in our TCRα dataset (TCRα/δ V4 and V9), likely due to inter-individual polymorphisms. In the absence of a complete germline sequence for this locus, we limited our database for TCRα to thymocyte clones with the same unique CDR3 signature. The CDR3 region results from the somatic recombination and assembly of variable (V) and joining (J) gene segments during lymphocyte development in the thymus (Kuklina 2006; Lantelme et al. 2008). The recombination process cleaves DNA and initiates repair mechanisms that result in the random insertion of non-template (N) nucleotides within the join, forming a unique binding sequence that contributes to the diversity and specificity of a TCR (Gellert 2002; Kuklina 2006). Once recombination ceases and the thymocyte proliferates, this

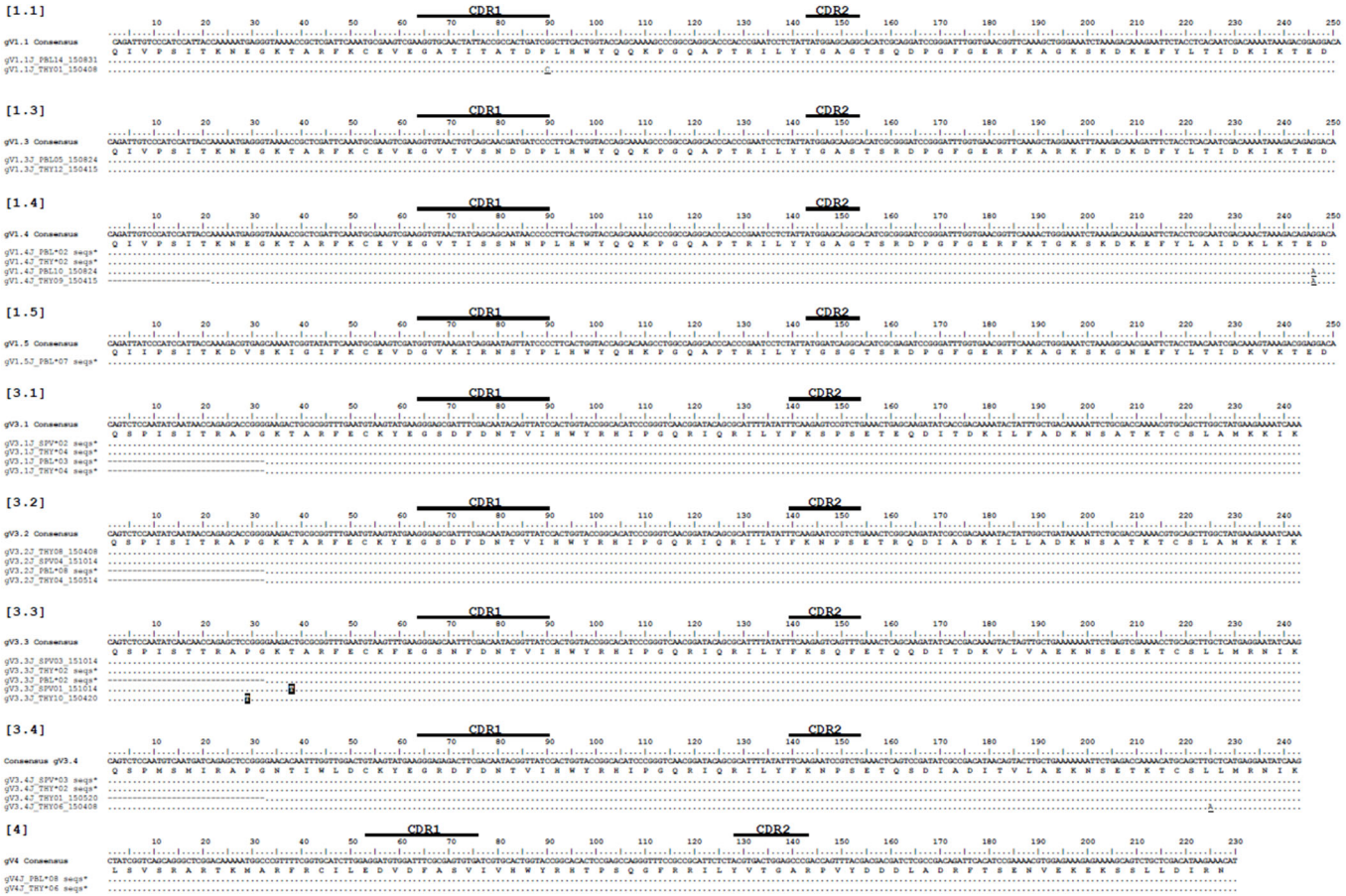

Figure 2-1. Alignment of Gamma V clones suggests minimal somatic hypermutation. Thymocyte clones for nine γV groups from three different predicted V genes. CDR regions are marked above the scale for each γV alignment. Amino acids are shown under the nucleotide consensus sequence, and dots represent identity to this sequence. We highlighted nonsynonymous changes in black; synonymous changes are underlined. Gaps are used for alignment purposes and indicate a shortened sequence (at the beginning or end). Sequences are identified by a single clone number or a group of identical clones condensed to a single line (the number of clones is indicated). Clone numbers that contain 'THY' are from thymus, 'PBL' are from peripheral blood leukocytes, and 'SPV' are from spiral valve (intestine). We deposited all 69 sequences into GenBank under accession numbers KY351639 – KY351707.

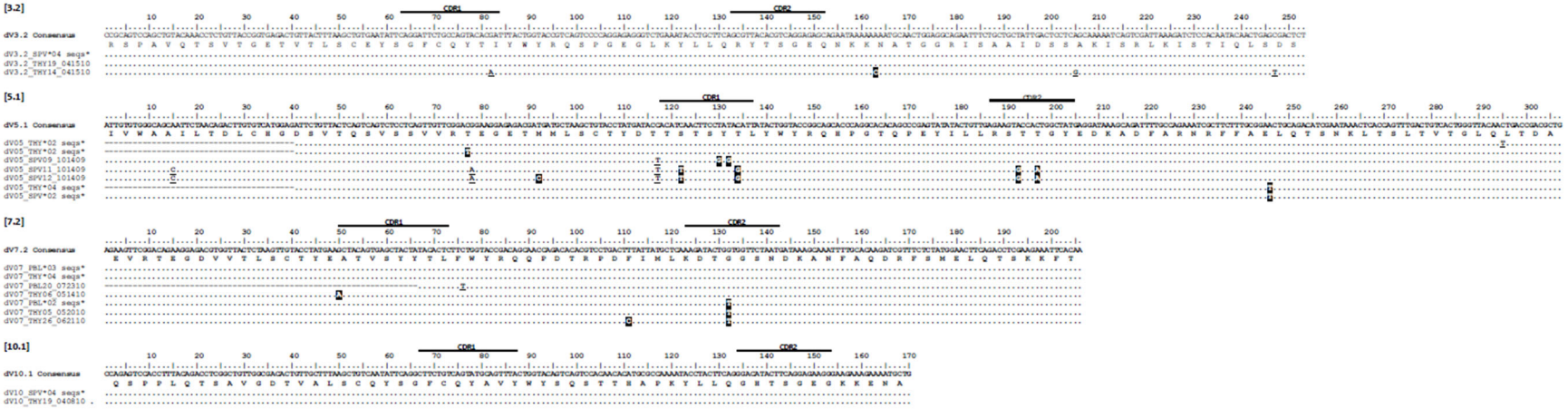

Figure 2-2. Alignment of Delta V clones suggests somatic hypermutation. Thymocyte clones for 12 δV groups from seven different predicted V genes. CDR regions are marked above the scale for each δV alignment. Amino acids are shown under the nucleotide consensus sequence, and dots represent identity to this sequence. We highlighted nonsynonymous changes in black; synonymous changes are underlined. Gaps are used for alignment purposes and indicate a shortened sequence (at the beginning or end). Sequences are identified by a single clone number or a group of identical clones condensed to a single line (the number of clones is indicated). Clone numbers that contain 'THY' are from thymus, 'PBL' are from peripheral blood leukocytes, and 'SPV' are from spiral valve (intestine). We deposited all 112 sequences into GenBank under accession numbers KY346705 –KY346816

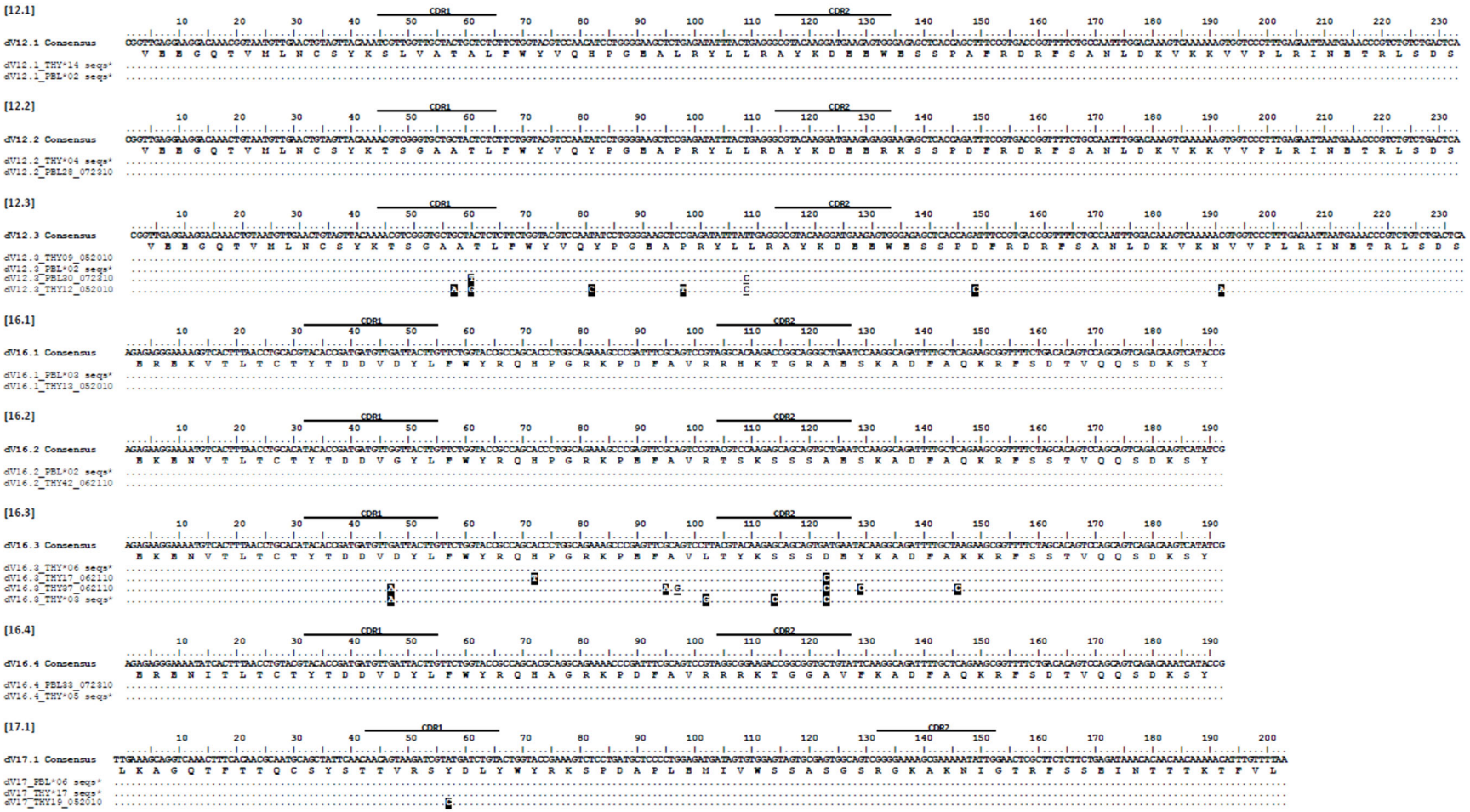

Figure 2-2. (*continued*)

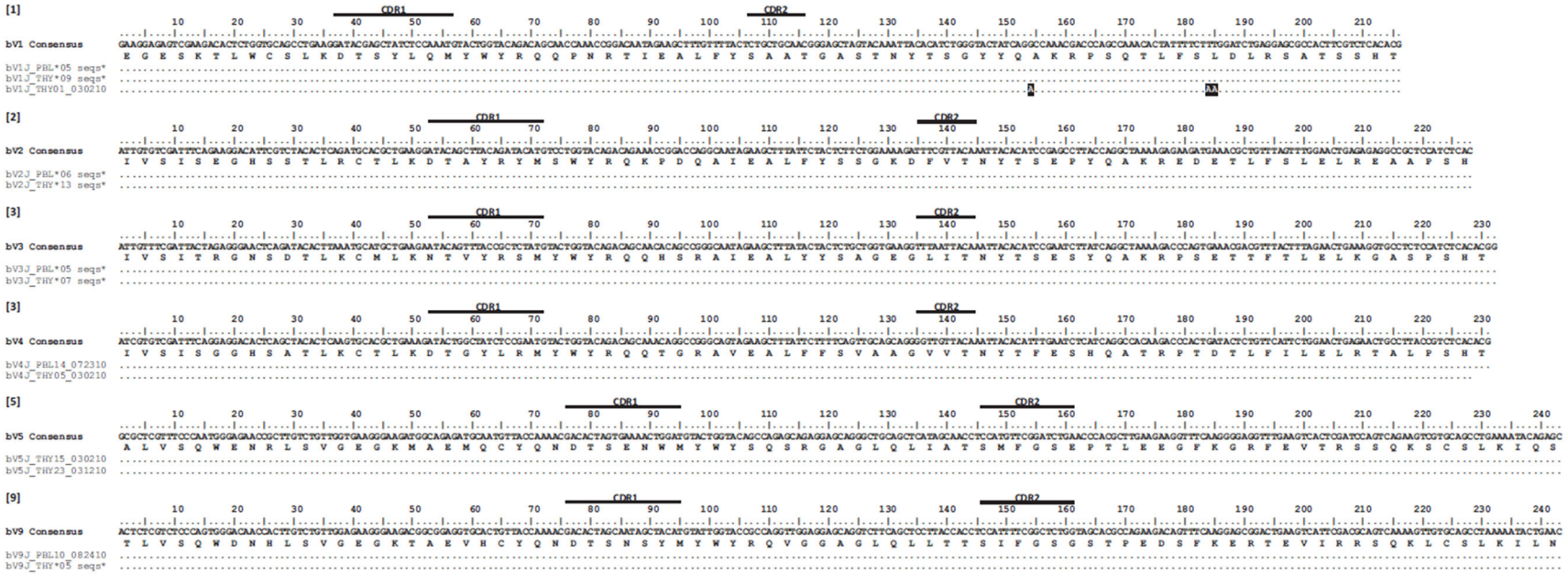

Figure 2-3. Alignment of Beta V clones suggests a lack of somatic hypermutation. Thymocyte clones for βV groups from six different predicted V genes. CDR regions are marked above the scale for each βV alignment. Amino acids are shown under the nucleotide consensus sequence, and dots represent identity to this sequence. We observed three nonsynonymous changes within a single sequence (highlighted in black). Sequences are identified by a single clone number or a group of identical clones condensed to a single line (the number of clones is indicated). Clone numbers that contain "PBL" are from peripheral blood leukocytes and "THY" are from thymus. We deposited all 57 sequences into GenBank under accession numbers KY351708 – KY351764.

distinctive CDR3 sequence is perpetuated in all daughter cells (Murphy and Weaver 2017). Since it is unlikely that two thymocytes would generate CDR3 sequences during VJ recombination, we predict that amplicons containing identical CDR3 sequences derived from the same progenitor and thus must contain the same germline V and J segments. Alpha CDR3s exhibited substantial variation within our shark sequences, despite the absence of diversity (D) segments. For example, TCRα V1 sequences using the same V and J segments had CDR3 lengths that differed by as many as 6 amino acids (18 nucleotides), and few CDR3s shared more than one amino acid within this V-J join (Figure 2-4). Further, the majority of sequences in our overall dataset do contain N and palindromic (P) nucleotides within this join. Using clones containing the same V segment from different sharks, we determined the putative end of each V segment. We first aligned sequences containing the same V segment. Then, assuming that any nucleotide present in the same position within more than one shark must be germline, we determined which nucleotides within a join belong to the V segment. We then repeated this process for each J segment. Of 290 clones, we found 197 (68%) unique sequences within this join (0 – 34 nucleotides in length), suggesting that most sequences contain N/P nucleotides (see Figure 2-4). Finally, we never observed the same CDR3 sequence in more than one shark, suggesting both exonuclease activity and addition of N and P nucleotides help diversify alpha CDR3s in nurse shark. Therefore, even in the absence of an assembled locus, we were able to evaluate mutation to germline αV segments by considering changes within only those thymocyte clones containing identical CDR3s. Our extremely conservative approach of

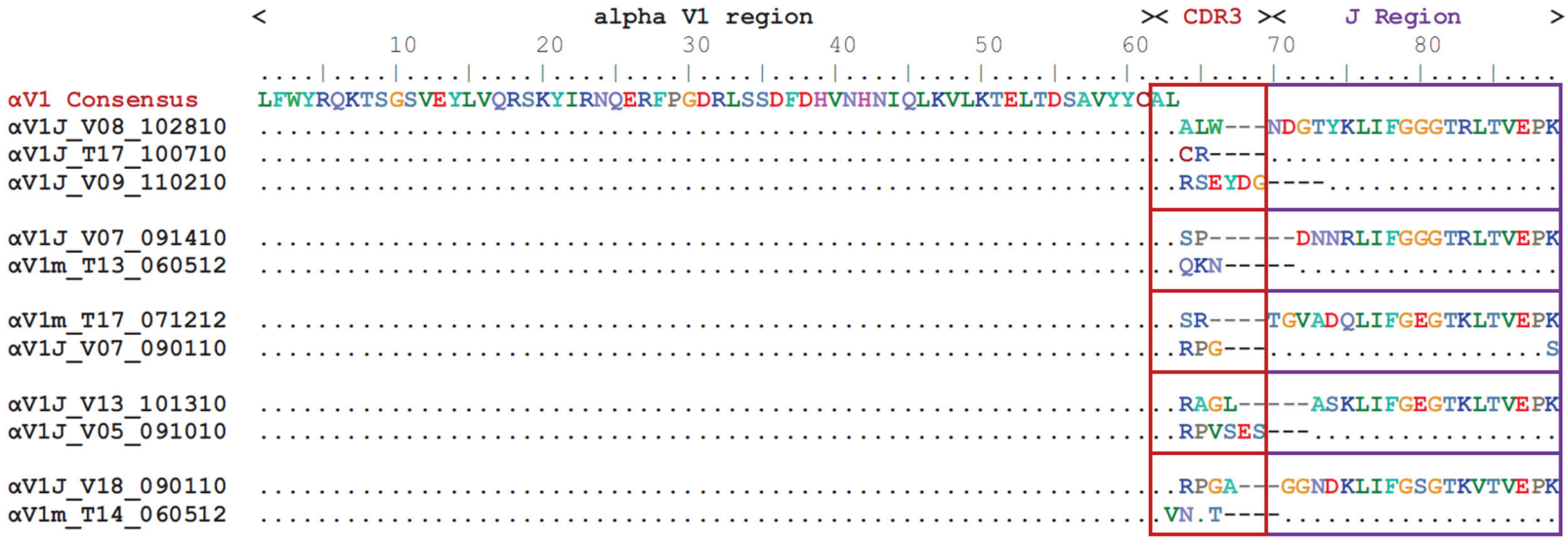

Figure 2-4. CDR3s of TCR Alpha chain are diverse. Amino acid (aa) alignment of TCR aV1 thymocyte clones illustrating diversity of the third complementarity-determining region (CDR3). All clones contain identical variable (V) region sequence (aa 1–61). We grouped clones by shared, identical joining (J) regions (purple boxes) and highlight the differences in the V-J join (CDR3 region) in red boxes. [Source data can be found in Appendix B-1]

using *only clones containing the same CDR3 encoding rearrangement and N and P nucleotide sequences* provided us the assurance that we had distinct αVs descendant from clonal T cells, since it would be extremely unlikely that two T cells created receptors that contained the exact same nucleotide sequence by chance.

2.2.3. Somatic hypermutation in nurse shark TCR αV

With SHM confirmed in γ and δ TCR chains but apparently not the TCR beta chain of nurse shark, we checked for mutation of the TCRα locus. One might expect mutation in γδ T cells since antigen binding more closely mirrors that of B cells. However, mutations to receptors of MHC-restricted αβ T cells would be surprising given that even minor modifications to these receptors could risk incompatibility with MHC.

Our preliminary V α dataset contained 539 TCRα clones (encoding 286 unique amino acid sequences representing nine V α families) from three tissues (PBL, spleen, thymus) of two sharks (*Joanie*, *Mary Junior*). Of this total, only 447 sequences contained complete CDR3-J junctions (including all bases after the last predicted nucleotide of the V segment to the last predicted nucleotide of the J segment; see Table 2-1). We observed 239 (53.5%) sequences with unique V regions (from the first predicted nucleotide of the V segment to the last predicted nucleotide of the J segment) and 179 (40%) with unique V segments (from the first to the last predicted nucleotide of the V segment only). We found 226 sequences (50%) containing unique CDR3-J regions (all bases after the last predicted nucleotide of the V to the last predicted nucleotide of the J). However, we

found 48 groups containing *identical* CDR3-J sequences across all nine V α families (suggesting they bear the V-J rearrangement from a single founder thymocyte), each V α family containing anywhere from one to ten clonal groups (Table 2-1). The majority of these groups contained no mutation within V, J or C regions. For example, one α V3 sequence group occurred 131 times, the most numerous sequences in the dataset, yet contained no mutation in any sequence. We did observe mutation in 12 of these 48 groups belonging to seven different V α families. Each family contained between one and four clonal groups, and each group contained two to nine sequences with identical CDR3-J regions (45 total sequences; see Table 2-1). We include these 45 sequences in our TCR V α dataset.

Table 2-1. Summary of sequence data used in this paper. Putative subfamilies within each TCR alpha V family share at least 85% nucleotide identity using nearest-neighbor consensus trees of V segments. Number of TCR alpha nucleotide (NUC), amino acid (AA) sequences or sequence groups within each category. Highlighted columns specifically refer to data used in this study. (See results for detailed descriptions of sequences included within each column.)

alpha V Segment	Putative # Subfamilies	All Cloned Sequences	Complete CDR3-J Junction[#]	Unique V Region[†]		Unique V Segment[††]		Unique CDR3-J[‡]		Groups with Identical CDR3-J[o]	CDR3-J Groups in Study[*]	Sequences in each dataset[**]
				NUC	AA	NUC	AA	NUC	AA			
TCRA V1	1	40	40	35	34	24	16	34	32	7	4	2, 3, 3, 5
TCRA V2	3	18	18	13	13	13	10	12	12	3	1	5
TCRA V3	3	217	194	55	52	35	34	51	50	5	1	4
TCRA V4	3	60	28	22	22	21	21	21	21	6	2	2, 2
TCRA V5	4	35	34	15	13	9	8	13	13	3	1	2
TCRA V6	2	9	9	7	7	5	6	7	7	1	0	0
TCRA V7	5	96	60	49	48	39	38	48	50	9	2	2, 2
TCRA V9	3	19	19	14	14	12	11	13	13	4	0	0
TCRA V10	2	45	45	29	26	21	19	27	26	10	2	4, 9
		539	447	239	229	179	163	226	224	48	**13**	**45**

[#]A full list of these sequences can be found in Table 1 - source data.
[†]V Region includes all bases between the 1st predicted nucleotide of the V segment to the last predicted nucleotide of the J segment (V and J).
[††]V Segment includes all bases between the 1st predicted nucleotide of the V segment to the last predicted nucleotide of the V segment (V only).
[‡]CDR3-J includes all bases after the last predicted nucleotide of the V segment to the last predicted nucleotide of the J segment.
[o]Number of groups with identical CDR3-J sequences, which we used to determine sequence relatedness (see text for details).
[*]Number of groups with identical CDR3-J sequences used in this study. (Those not used contained no mutation with V segments.)
[**]Total number of sequences for each alpha V used to assess somatic hypermutation within this study (e.g., for aV1, 4 different clonal groups contained 2, 3, 3, and 5 identical CDR3-J regions, respectively).

Using these 45 sequences from both thymus and peripheral immune tissues, we found evidence for SHM acting on the TCRα genes. We divided our 45 clones into 13 CDR3-sharing groups from seven different α V families and then analyzed sequences within groups for potential mutation (Figure 2-5). We excluded two groups (4 sequences) from analyses (αV1.3 and αV7.1) that contained no mutations within FR or CDR regions (leaving 41 clones for analysis). Two sequences (*aV7.4m_THY09_051410* and *aV5J_SPV17_102810*) contained one 3-base insertion and one sequence (*aV1.4J_SPV07_090110*) contained an 18-base insertion; although SHM can result in insertions and deletions (Diaz et al. 2002), we did not include insertion nucleotides in mutation counts. All sequences were in-frame and contained no internal stop codons, suggesting functionality of cells. Average lengths of CDRs were as follows: CDR1: 7.0 amino acids (range: 5 – 8); CDR2: 5.7 amino acids (range: 5 – 7); and CDR3: 6.0 amino acids (range: 1 - 10). Naming of families and subfamilies followed Criscitiello *et al*. (Criscitiello et al. 2010). However, with the accumulation of sequence data over our previous analysis of nurse shark TCRα (Criscitiello et al. 2010), we expanded our nomenclature considerably. Additional annotation followed the IMGT guidelines for TCRs (Lefranc et al. 2003).

Figure 2-5 shows all 12 αV CDR3 groups exhibiting mutation. The overall TCRα mutation frequency was 0.0226 substitutions per nucleotide (S/N), with 66% of all substitutions (187 of 283) resulting in amino acid replacements. The CDRs accumulated significantly more mutations than FRs (CDR: 0.0352 S/N; FR: 0.0188 S/N; df=1 p=0.0373),

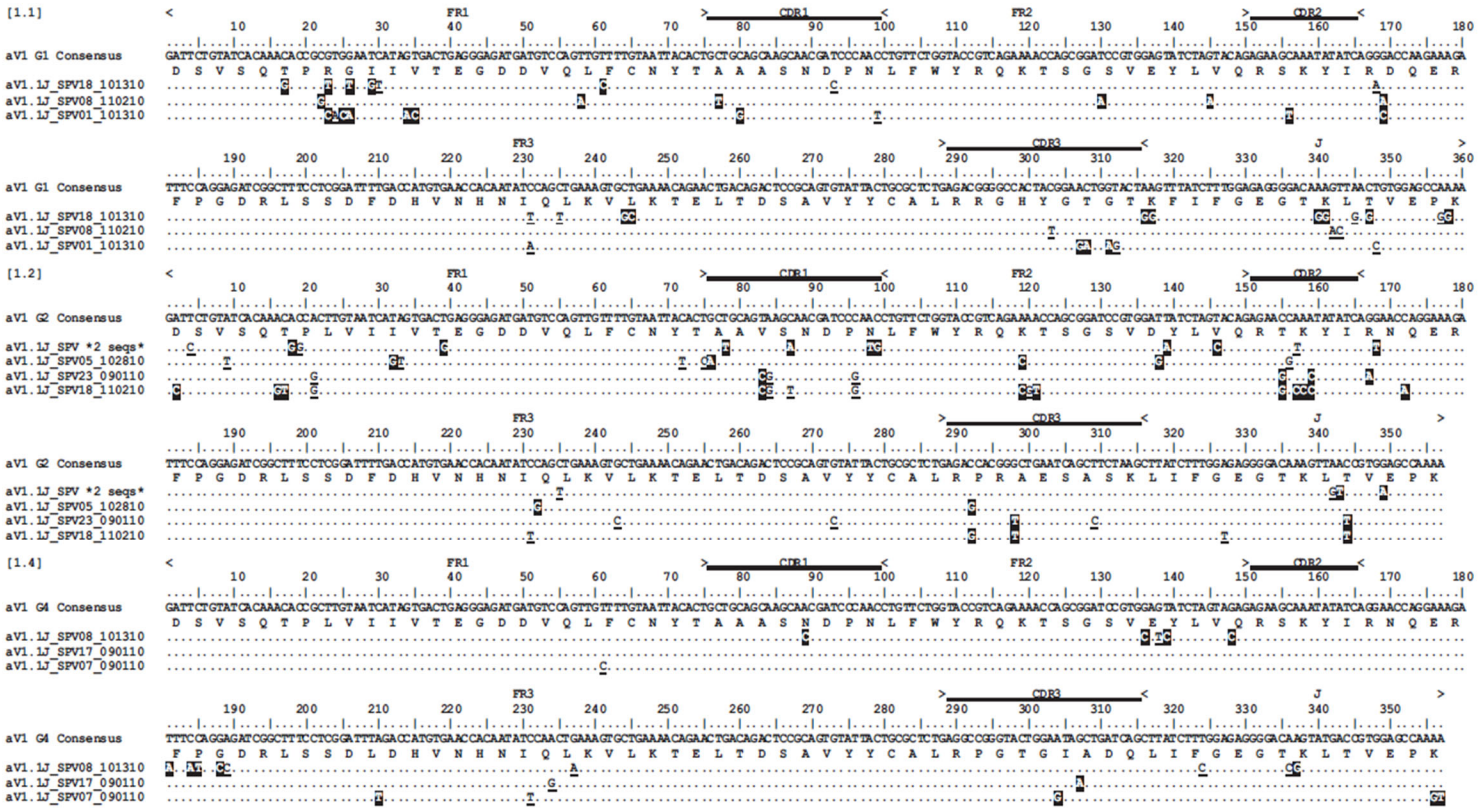

Figure 2-5. Alignment of Alpha V cDNA clones suggest somatic hypermutation at shark TCRα. Thymocyte clones for all 11 αV groups with the same CDR3 from six different predicted V genes. Locations of framework regions (FR), complementarity determining regions (CDR), joining regions (J), and constant (C) regions are marked above the scale for each αV. In absence of germline sequence information, we used a Geneious-derived nucleotide consensus sequence for analysis of nucleotide changes in thymocyte clones. Amino acids are shown under the consensus sequence, and dots represent identity to this sequence. Nonsynonymous changes are highlighted in black; synonymous changes are underlined. Gaps are used for alignment purposes and indicate either a shortened sequence (at the beginning or end) or insertions or deletions within the sequence. Sequences are identified by a single clone number or a group of identical clones condensed to a single line (the number of clones is indicated). Clone numbers that contain 'THY' are from thymus, 'PBL' are from peripheral blood leukocytes, and 'SPV' are from spiral valve (intestine). We did not use clones from aV1.3 and aV7.1 because they did not contain mutations in FR or CDR regions. We deposited all 42 sequences into GenBank under accession numbers KY189332 – KY189354 or KY366469 – KY366487.

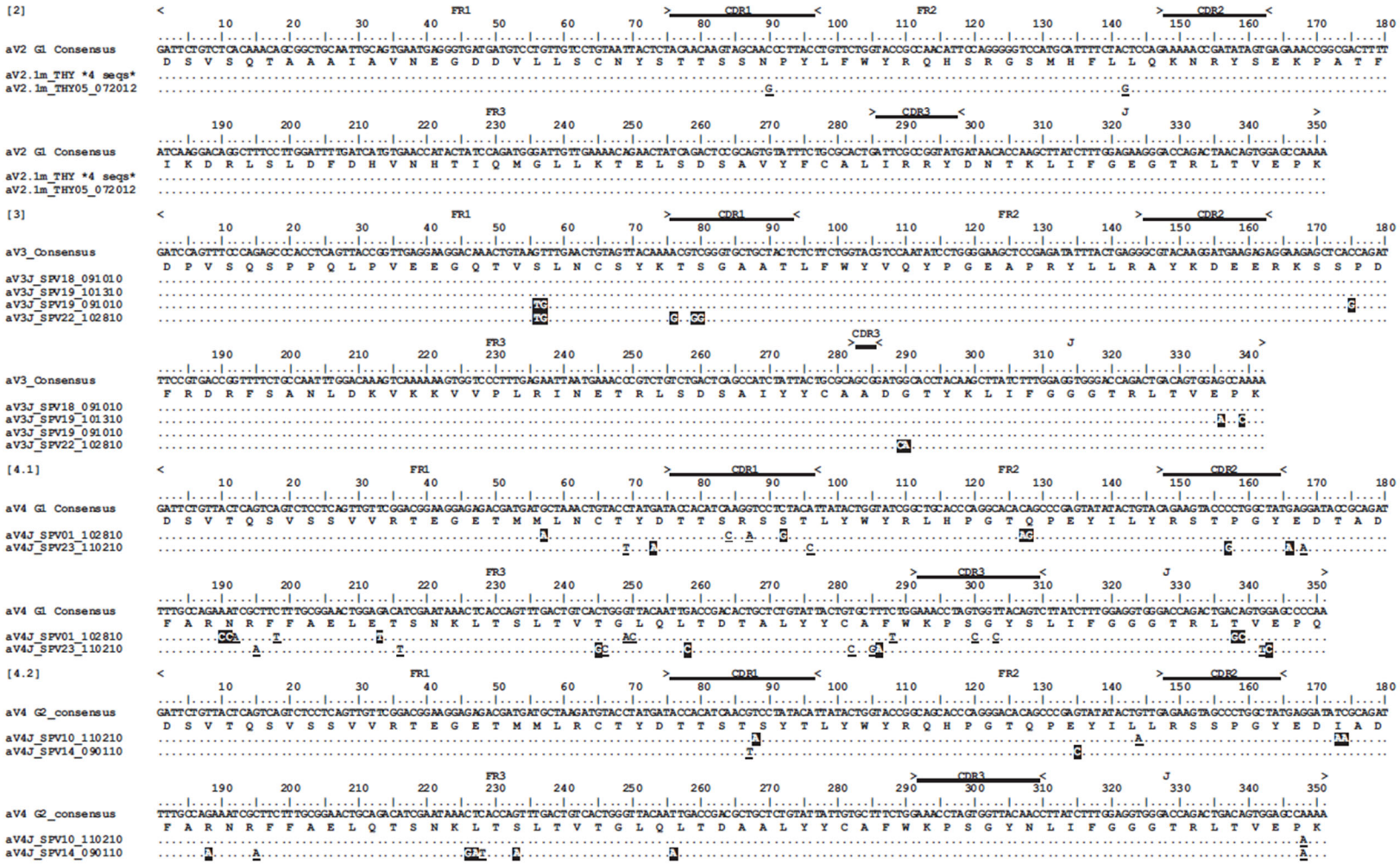

Figure 2-5. (*continued*)

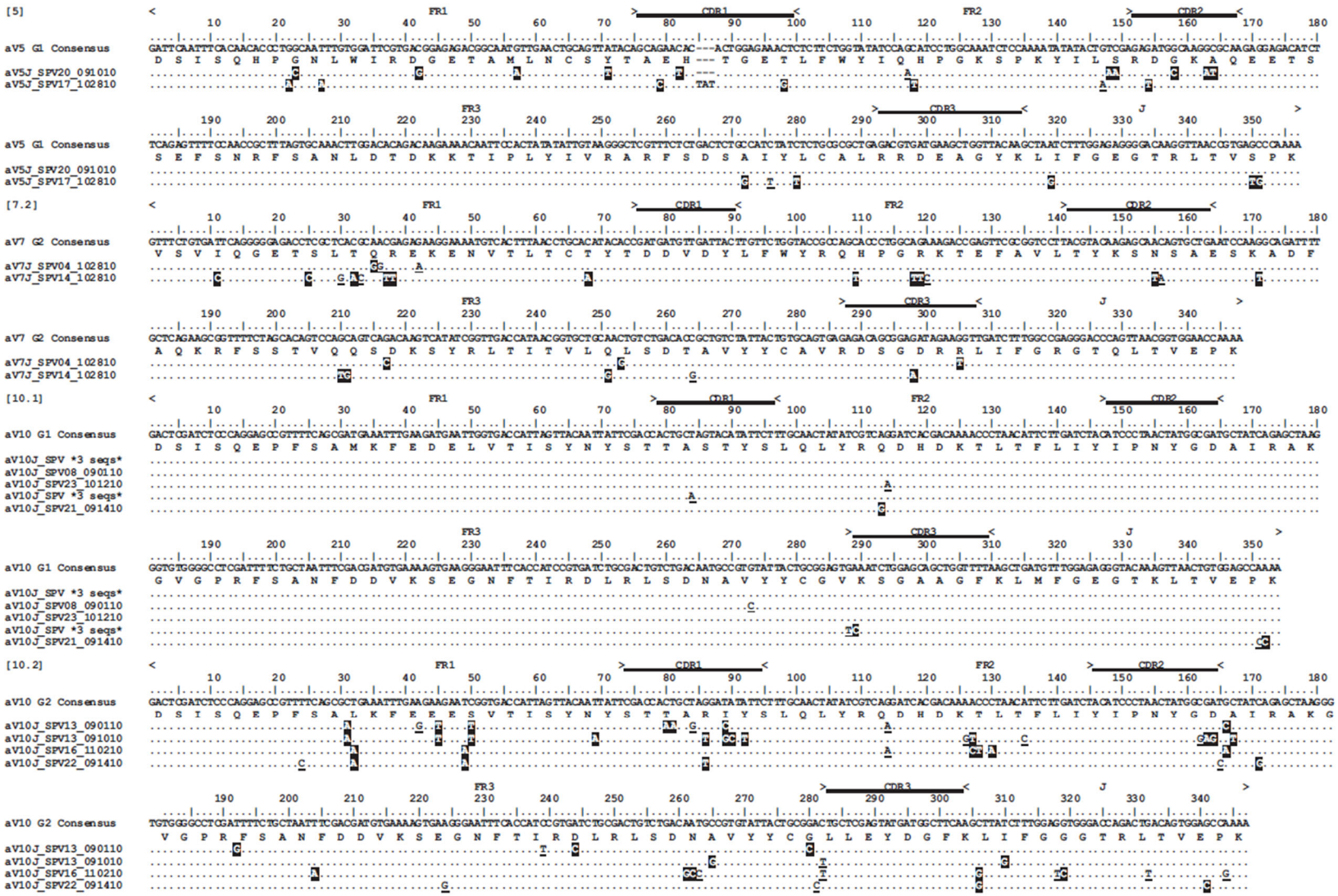

Figure 2-5. (*continued*)

and substitutions in CDRs were twice as likely to be nonsynonymous changes (NSYN) than those in FRs (CDR: 0.0235 S/N; FR: 0.0122 S/N; df=1; p= 0.0312; Table 2-2). There was no difference in frequency of synonymous (SYN) mutations between regions (CDR: 0.0117 S/N, FR: 0.0066 S/N; df=1; p=0.0705). Finally, though we found more tandemly mutated bases in CDRs (41 of 81, or 50.6% of all CDR mutations) than in FRs (73 of 192, or 38.0% of all FR mutations), this difference was not significant; df=1; p=0.721; Table 2-3). Tandem mutations ranged from two to four bases in length (mean = 2.78 bases). That this feature of SHM, specific of cartilaginous fish Ig (Diaz et al. 1999; Lee et al. 2002; Rumfelt et al. 2001), also occurs in the TCR strongly supports the validity of our analyses.

Mutation frequency also varied by region (CDR or FR). The highest mutation frequency occurred in CDR1 and accumulated significantly more mutations overall than other regions ($\bar{x}$=4.48%; see Figures 2-6 and 2-7a, Table 2-2). CDR3 displayed an unusually low mean mutation frequency (2.97%). However, these frequencies may be artificially low since our groups were based only on clones containing the same CDR3 sequence, and clones whose CDR3s deviated markedly from the consensus would have been excluded by our conservative grouping approach. We observed the lowest mean mutation frequency in FR3 (1.64%). These results are consistent with what is known about human TCR binding to MHC: Ag structures. CDR1 and CDR3 make more contacts with Ag while CDR2 interacts primarily with non-polymorphic regions of MHC (Buslepp et al. 2003; Garcia and Adams 2005), making mutation in CDR2 less favorable; further, FR regions are important for the structural stability of the domain and mutations to these regions may affect the ability of

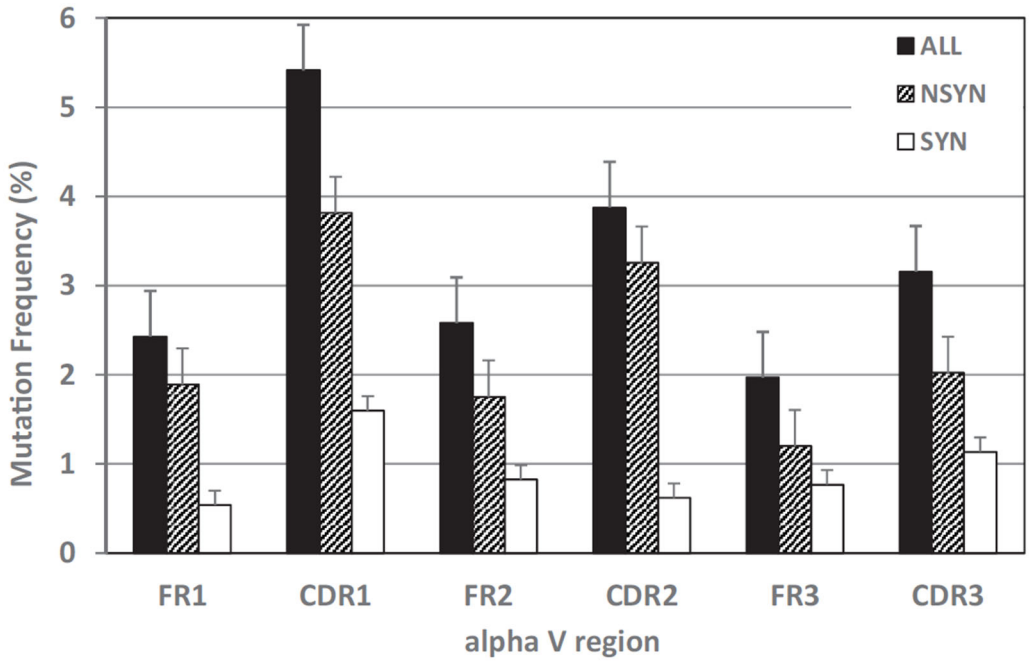

Figure 2-6. Mutation frequencies differed within TCR V regions. Mutability of complementarity determining regions (CDR), especially CDR1, exceeded that of framework regions (FR) for all mutations together (black bars) and for nonsynonymous mutations alone (NSYN, hatched bars). We found no statistical difference in synonymous mutations (SYN, white bars) between CDRs and FRs. [Source data can be found in Appendix B-2]

the Ig superfamily domain to fold properly (Mantovani et al. 2002; Reinherz et al. 1999).

2.2.4. Hotspots

We found 1,327 G/C base pairs within DGYW/WRCH hotspot motifs (4,402 G/C base pairs occurred outside these motifs) and 2,931 A/T base pairs within WA/WT hotspot motifs (3,690 A/T base pairs occurred outside these motifs; see Table 2-4). Mutations of G:C nucleotides were strongly associated with DGYW/WRCH hotspots (Figure 2-7). Overall, G:C mutations occurred 4.3x as often within hotspots as those

outside hotspots. However, G:C mutations were 1.4x more likely within CDRs than within FRs. A:T nucleotides did not appear to prefer WA/TW hotspots (Figure 2-7), though they were 2.3x as likely to mutate within hotspots than outside hotspots. Further, while A:T mutations occurred more often than expected in both FRs and CDRs, the frequency of A:T mutations were 1.9x more likely in CDRs than FRs. Chen et al (2012) observed similar results in TCR γV sequences of sandbar shark. The authors suggested that A:T mutations still are likely the result of DNA polymerase η use during mismatch repair mechanisms due to the lack of A:T point mutations in that study (the majority occurring in tandem with other mutations) (Chen et al. 2012; Rogozin et al. 2001; Wei et al. 2015). In nurse shark Ig light chain genes, the majority of mutation to A:T nucleotides (55%) also occurred in tandem (Alder et al. 2005). However, 63% (76 of 121) of A:T mutations occurred as point mutations in our study, which is inconsistent with DNA polymerase η use during mismatch repair (Chen et al. 2012; Rogozin et al. 2001; Wei et al. 2015). This result suggests that an alternate A:T motif is targeted or that shark TCRs employ a different mechanism to alter A:T nucleotides.

2.2.5. Base substitution indices

We identified 283 mutations within the eleven CDR3 groups. There was a bias towards mutations of G and C nucleotides (χ^2=9.34, df=1; p=0.0022) and G:C changes comprised 57% of all mutations (Table 2-5a). More mutations to both G and C nucleotides occurred in FRs (χ^2=8.43, df=1; p=0.0037) while only C nucleotides showed greater

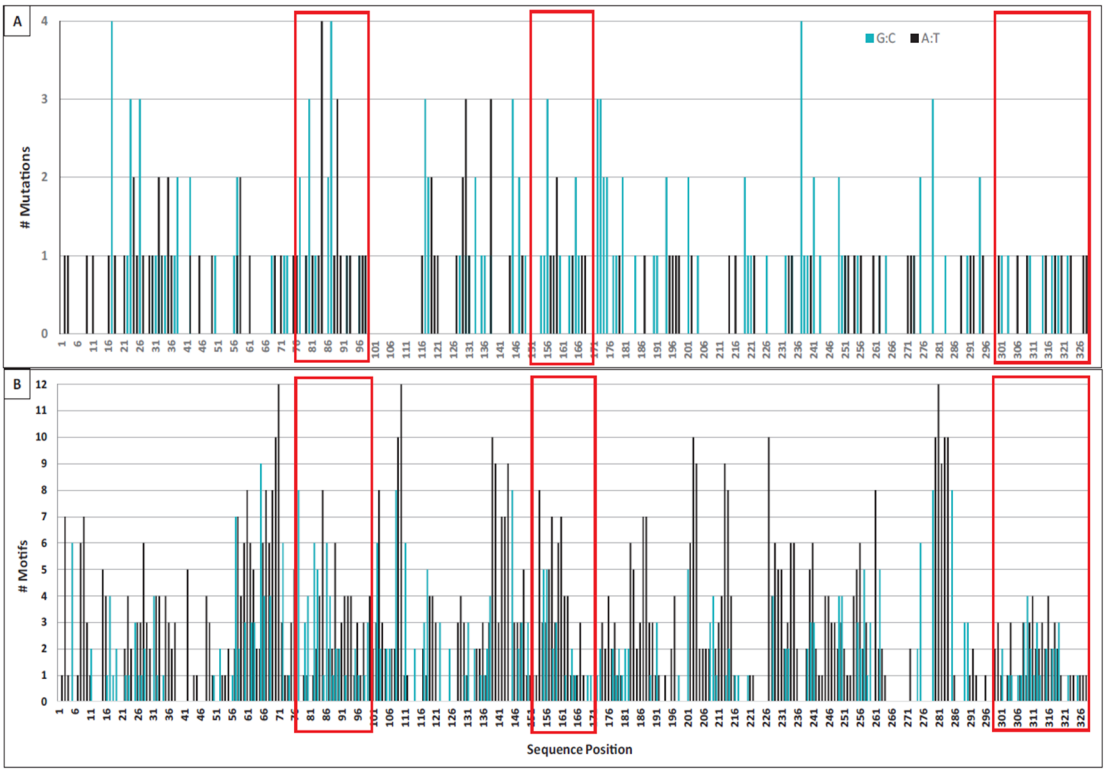

Figure 2-7. Mutation and motif locations within individual domains of TCR αV sequences. (A) Number of mutations (both single and tandem, 283 total) to A:T nucleotides (black bars, 122 mutations) or G:C nucleotides (blue bars, 161 mutations) observed at each position along the sequence length of TCR αV sequences. We counted mutations to Geneious-derived consensus sequences within framework and complementarity-determining region (CDR) domains (41 sequences within 11 αV groups). (B) Number of WA/TW motifs (black bars, indicating possible polymerase h action) or DGYW/WRCH motifs (blue bars, indicating hotspots for AID activity, and thus possible mutation) at each position along the sequence length. Position indicates the forward (3' to 5') location of the mutable base of each motif within a Geneious-derived consensus sequence for each TCR αV group (eleven groups). Red boxes indicate the location of CDR1 (positions 76–99), CDR2 (positions 151–171), and CDR3 (positions 300–328) within each panel. [WA/TW: A:T is the mutable position; DGYW/WRCH: G:C is the mutable position; W = A/T, D = A/G/T, Y = C/T, R = A/G, and H = T/C/A]

Table 2-2. Frequencies of somatic hypermutation in nurse shark alpha V groups (αV G) containing the same CDR3. Mutation frequency is the total number of nucleotide changes to a Geneious-derived consensus sequence divided by the total number of nucleotides. We counted synonymous (S) and nonsynymous (N) mutations separately for each FR and CDR for eleven different CDR3 groups in seven predicted alpha V genes. [FR: framework region; CDR: complementarity-determining region; Seqs: sequences; Nuc: nucleotides; Freq: frequency]

		FR1						FR2						FR3						FR Means		
αV	#	#	Total	Mutations		Total	Mutation	#	Total	Mutations		Total	Mutation	#	Total	Mutations		Total	Mutation			
Group	Seqs	Codons	Nuc	S	N	Mutation	Freq (%)	Codons	Nuc	S	N	Mutation	Freq (%)	Codons	Nuc	S	N	Mutation	Freq (%)	S	N	ALL
αV1.1	3	25	225	2	12	14	**6.222**	17	153	0	2	2	**1.307**	41	369	4	4	8	**2.168**	6	18	24
αV1.2	5	25	375	10	8	18	**4.800**	17	255	1	8	9	**3.529**	41	615	5	5	10	**1.626**	16	21	37
αV1.4	3	25	225	1	0	1	**0.444**	17	153	1	3	4	**2.614**	41	369	4	5	9	**2.439**	6	8	14
αV2.1	5	25	375	0	0	0	**0.000**	17	255	1	0	1	**0.392**	42	630	0	0	0	**0.000**	1	0	1
αV3.1	4	25	300	0	4	4	**1.333**	17	204	0	0	0	**0.000**	40	480	0	1	1	**0.208**	0	5	5
αV4.1	2	25	150	1	2	3	**2.000**	17	102	0	2	2	**1.961**	43	258	11	7	18	**6.977**	12	11	23
αV4.2	2	25	150	0	0	0	**0.000**	17	102	1	1	2	**1.961**	43	258	2	7	9	**3.488**	3	8	11
αV5.1	2	25	150	0	6	6	**4.000**	17	102	2	3	5	**4.902**	42	252	1	2	3	**1.190**	3	11	14
αV7.2	2	25	150	4	7	11	**7.333**	17	102	1	3	4	**3.922**	41	246	1	6	7	**2.846**	6	16	22
αV10.1	4	25	300	0	0	0	**0.000**	17	204	1	1	2	**0.980**	41	492	4	0	4	**0.813**	5	1	6
αV10.2	9	25	675	2	11	13	**1.926**	17	459	4	4	8	**1.743**	40	1080	3	11	14	**1.296**	9	26	35
Sum	**41**	**275**	**3075**	**20**	**50**	**70**		**187**	**2091**	**12**	**27**	**39**		**455**	**5049**	**35**	**48**	**83**		**67**	**125**	**192**
Mean Mutation Freq (%)				**0.65**	**1.63**	**2.28**				**0.57**	**1.29**	**1.87**				**0.69**	**0.95**	**1.64**		**0.66**	**1.22**	**1.88**
Standard Deviation				**2.994**	**4.547**	**6.531**				**1.136**	**2.252**	**2.841**				**3.125**	**3.414**	**5.429**		**4.742**	**8.201**	**11.861**

		CDR1						CDR2						CDR3						CDR Means		
AlphaV	#	#	Total	Mutations		Total	Mutation	#	Total	Mutations		Total	Mutation	#	Total	Mutations		Total	Mutation			
Group	Seqs	Codons	Nuc	S	N	Mutation	Freq (%)	Codons	Nuc	S	N	Mutation	Freq (%)	Codons	Nuc	S	N	Mutation	Freq (%)	S	N	ALL
αV1.1	3	8	72	2	2	4	**5.556**	5	45	1	0	1	**2.222**	10	90	1	6	7	**7.778**	4	8	12
αV1.2	5	8	120	5	4	9	**7.500**	5	75	1	2	3	**4.000**	8	120	1	4	5	**4.167**	7	10	17
αV1.4	3	8	72	0	1	1	**1.389**	5	45	0	0	0	**0.000**	5	45	0	2	2	**4.444**	0	3	3
αV2.1	5	7	105	0	1	1	**0.952**	5	75	0	0	0	**0.000**	3	45	0	0	0	**0.000**	0	1	1
αV3.1	4	6	72	0	3	3	**4.167**	6	72	0	0	0	**0.000**	1	12	0	0	0	**0.000**	0	3	3
αV4.1	2	7	42	3	1	4	**9.524**	6	36	0	1	1	**2.778**	6	36	2	0	2	**5.556**	5	2	7
αV4.2	2	7	42	1	1	2	**4.762**	6	36	0	0	0	**0.000**	6	36	0	0	0	**0.000**	1	1	2
αV5.1	2	7	42	0	3	3	**7.143**	6	36	0	7	7	**19.444**	7	42	0	0	0	**0.000**	0	10	10
αV7.2	2	5	30	0	0	0	**0.000**	7	42	1	1	2	**4.762**	6	36	0	2	2	**5.556**	1	3	4
αV10.1	4	7	84	3	0	3	**3.571**	6	72	0	0	0	**0.000**	7	84	0	3	3	**3.571**	3	3	6
αV10.2	9	7	189	1	8	9	**4.762**	6	162	2	2	4	**2.469**	7	189	3	0	3	**1.587**	6	10	16
Sum	**41**	**77**	**870**	**15**	**24**	**39**		**63**	**696**	**5**	**13**	**18**		**66**	**735**	**7**	**17**	**24**		**27**	**54**	**81**
Mean Mutation Freq (%)				**1.72**	**2.76**	**4.48**				**0.72**	**1.87**	**2.59**				**0.95**	**2.31**	**3.27**		**1.17**	**2.35**	**3.52**
Standard Deviation				**1.690**	**2.316**	**2.979**				**0.688**	**2.089**	**2.248**				**1.027**	**2.067**	**2.272**		**2.659**	**3.754**	**5.626**

Table 2-3. Number and frequency of DNA mutations that occur in tandem within framework regions (FR) and complementarity determining regions (CDR) in nurse shark alpha V (αV) groups. All mutation includes both tandem and point mutations within a region. [Seqs: sequences]

		FR						CDR					
		# Nucleotides Tandemly Mutated				All Mutation	Frequency of Tandem Mutation	# Nucleotides Tandemly Mutated				All Mutation	Frequency of Tandem Mutation
αV Group	# Seqs	2	3	4	Sum			2	3	4	Sum		
αV1.1	3	1	0	1	6	24	25.0	3	0	0	6	12	50.0
αV1.2	5	4	1	0	11	37	29.7	5	1	0	13	17	76.5
αV1.4	3	3	0	0	6	14	42.9	0	0	0	0	3	0.0
αV2	5	0	0	0	0	1	0.0	0	0	0	0	1	0.0
αV3	4	2	0	0	4	5	80.0	1	0	0	2	3	66.7
αV4.1	2	4	1	0	11	23	47.8	0	1	0	3	7	42.9
αV4.2	2	1	1	0	5	11	45.5	0	1	0	3	2	150.0
αV5	2	1	0	0	2	14	14.3	1	1	0	5	10	50.0
αV7.2	2	4	1	0	11	22	50.0	1	0	0	2	4	50.0
αV10.1	4	3	0	0	6	6	100.0	0	0	0	0	6	0.0
αV10.2	9	4	1	0	11	35	31.4	2	1	0	7	16	43.8
Total	41	27	5	1	73	192	38.0	13	5	0	41	81	50.6

Table 2-4. Target nucleotide mutation frequency in DGYW/WRCH or WA/TW mutation hotspots within framework regions (FR) and complementarity determining regions (CDR). [DGYW/WRCH (G:C is the mutable position; D=A/G/T, Y=C/T, W=A/T, R=A/G, and H=T/C/A); WA/TW (A:T is the mutable position; W=A/T); "ALL" refers to nucleotides within a hotspot motif; "All other" refers to nucleotides outside a hotspot motif]

Mutated Base	Hotspot Motif	Region	Total Nucleotides	Observed # Mutations	Expected # Mutations	Mutation Freq (%)	X^2 *p*	T-test *p*
G/C	DGYW/ WRCH	FR	927	56	18.16	6.04	0.0000	0.0267
		CDR	400	35	26.22	8.75		
		ALL	1327	91	37.52	6.86	0.0000	
	Outside Motif		4402	71	124.48	1.61		
A/T	WA/TW	FR	2531	54	36.07	2.13	0.0015	0.2248
		CDR	400	26	21.12	4.03		
		ALL	2931	80	55.97	2.52	0.0000	
	Outside Motif		3690	41	65.03	1.11		

[+]T-test analysis was used to compare mutations within hotspot motifs to those outside hotspot motifs. Mutations to G and C nucleotides occurred significantly more often within DGYW/ WRCH motifs than outside these motifs, while mutations to A and T nucleotides showed no preference for WA/TW motifs.

[*]χ2 analysis was used to compare observed and expected numbers of mutations between FR and CDR regions and between mutations inside and outside hotspot motifs. More mutations to all nucleotides occurred within hotspots than outside hotspots, and significantly more mutations occurred to nucleotides within CDRs than FRs.

Table 2-5. Bias in base substitution during somatic hypermutation of TCR alpha V genes within all sequence regions (ALL), framework regions (FR), or complementarity determining regions (CDR). Probability of occurrence is the proportion of that base out of the total nucleotides.
[Nuc: nucleotides; OBS: Observed; EXP: expected; MI: mutability index; ChiSq: Chi squared]

a

ALL	Base	Occurrence	Probability of Occurrence	OBS	EXP	MI[a]	ChiSq
	G	2895	0.230	77	65.05	1.18	0.0022
	C	2834	0.225	85	63.68	1.33	
	A	3498	0.278	64	78.60	0.81	0.0068
	T	3368	0.267	57	75.68	0.75	
	Total	12595	1.00	283	283		

GC Mutation: 57.0%; Transitions: 42.8%; Transversions: 25.1%

b

FR	Base	Occurrence	Probability of Occurrence	OBS	EXP	MI[a]	ChiSq
	G	2378	0.23	58	44.50	1.30	0.0037
	C	2268	0.22	56	42.44	1.32	
	A	2779	0.27	38	52.00	0.73	0.0082
	T	2835	0.28	40	53.05	0.75	
	Total	10260	1.00	192	192		

GC Mutation: 59.0%; Transitions: 41.7%; Transversions: 24.0%

c

CDR	Base	Occurrence	Probability of Occurrence	OBS	EXP	MI[a]	ChiSq
	G	517	0.22	19	20.15	0.94	0.1336
	C	566	0.24	29	22.06	1.31	
	A	719	0.31	26	28.02	0.93	0.3620
	T	533	0.23	17	20.77	0.82	
	Total	2335	1.00	91	91		

GC Mutation: 52.9%; Transitions: 45.1%; Transversions: 27.5%

[a]Mutability Index, as first defined in Chen et al 2012. χ^2 analysis was used to compare observed and expected numbers of mutations. G and C mutated significantly more often than expected, while A and T mutated significantly less often than expected. Base composition: 23.0% G, 22.5% C 27.8% A, 26.7% T.

Table 2-6. Frequencies of somatic hypermutation in nurse shark thymus and peripheral lymphoid tissue (blood and spiral valve). Mutations were analyzed only in alpha V groups containing the same third complementarity determining region (CDR). Mutation frequency was measured as the total number of nucleotide changes to a Geneious-derived consensus sequence divided by the total number of nucleotides in all sequences. Nonsynymous (N) and synonymous (S) mutations (mut) were counted separately for each framework (FR) and CDR for two predicted alpha V genes. [FR1, FR2, FR3, CDR1, CDR2, and CDR3 refer to the first, second, or third FR or CDR region, respectively.]

Tissue Type	Mut Type	FR Mutations (#)				CDR Mutations (#)				FR Mutation Frequency			CDR Mutation Frequency		
		FR1	FR2	FR3	All FR	CDR1	CDR2	CDR3	All CDR	FR1	FR2	FR3	CDR1	CDR2	CDR3
Thymus (6 sequences)	N	8	8	7	23	1	3	3	7	0.570	0.871	0.317	0.071	0.327	0.136
	S	8	2	6	16	1	0	0	1	0.570	0.218	0.272	0.071	0.000	0.000
	ALL	16	10	13	39	2	3	3	8	1.140	1.089	0.590	0.529	1.075	1.235
Total Nucleotides		1404	918	2205	4527	378	279	243	900						
Periphery (2 sequences)	N	7	4	6	17	0	2	2	4	1.496	1.307	0.833	0.000	1.852	1.587
	S	4	0	1	5	0	0	0	0	0.855	0.000	0.139	0.000	0.000	0.000
	ALL	11	4	7	22	0	2	2	4	2.350	1.307	0.972	0.000	1.852	1.587
Total Nucleotides		468	306	720	1494	126	108	126	360						

mutation than expected in CDRs (χ^2=2.25; df=1; p=0.1136). There were fewer mutations of A and T nucleotides in FRs (χ^2=6.98; df=1; p=0.0082; Table 2-5b), but mutations of A and T nucleotides in CDRs did not differ from random (χ^2=0.83; df=1; p=0.3620; Table 2-

5c). The frequency of mutated nucleotides also varied by region: In FR1, more A and T nucleotides mutated while fewer C nucleotides mutated. Mutations of G and C nucleotides were lower in FR2 and FR3, respectively (Table 2-5b,c). In both CDR1 and CDR2, there were more A mutations and fewer G mutations than expected. However, more G nucleotides mutated in CDR3. We saw a bias towards G:A and C:T transitions (42.8%) among the TCR α mutations. Transition mutations appeared only slightly more often in CDRs (45.1%) than in FRs (41.7%). Transversions of C:A or G:T mutations occurred only 25.1% of the time (Table 2-5).

2.2.6. Mutations, in situ *hybridization, and AID expression in thymus*

Since we cloned TCR αV sequences in both central (thymus) and peripheral (blood, spiral valve) lymphoid tissues, we analyzed mutation frequencies by tissue type. Though data were limited to only six sequences in two CDR3 groups (αV2, αV7.2), we found more mutations to FRs of peripheral sequences, though this was not significant (p=0.0541; Table 2-6). We found no differences between tissues within CDRs (p=0.2) from this limited data set. It is possible that positive and negative selection pressures remove more clonal sequences within the thymus, but with so few sequences to compare it is difficult to determine.

The mutated sequences in the primary T lymphoid tissue suggests the activity of activation-induced cytidine deaminase (AID) in the thymus. We confirmed the expression of AID in the thymus through real-time RT-qPCR, where thymus tissue expressed AID at

more than half (0.7x) the levels found in spleen (positive control, where B cell SHM is known to occur) and nearly 6x the levels observed in forebrain (negative control; Figure 2-8a). Further, colorimetric *in situ* hybridization (CISH) of nurse shark thymus revealed a diffuse signal of TCR β (Figure 2-8b, panels 1,2) and AID (Figure 2-8b, panels 3,4) mRNA expression throughout the thymic cortex. However, AID expression was greatest in the central cortex and cortico-medullary junction (CMJ), while TCR β expression was highest in the outer cortex. The thymic sequence data combined with the tissue localization of AID message strongly suggest that TCRα loci undergo AID-dependent SHM in the shark thymus.

We refined our CISH results with RNA fluorescence *in situ* hybridization (FISH) using probes against the TCR alpha constant region (αC) and exons 2 and 3 of *AID*. Since Stellaris RNA FISH uses a mixture of shorter probes (~20 bp in length) that hybridize along the length of the target RNA, the resulting signal is detectable only when tens of probes hybridize to the target. This makes the technology very specific -- only those transcripts bound to numerous probes are visible -- and limits the potential for false-positive or false-negative results (Orjalo et al. 2011; Raj and Tyagi 2010). Our FISH results indicated a more specific TCR αC signal within the inner cortex and medulla adjacent to the CMJ (Figure 2-9; Figure Supplement A-1, A-2), regions where, in mammals, developing cells actively rearrange their alpha chain genes and where mature αβ T cells are found. AID expression occurred in “rings” around areas of expressed TCR αC messages within the inner cortex and CMJ (where positive selection occurs in mammals) and the medulla (where negative

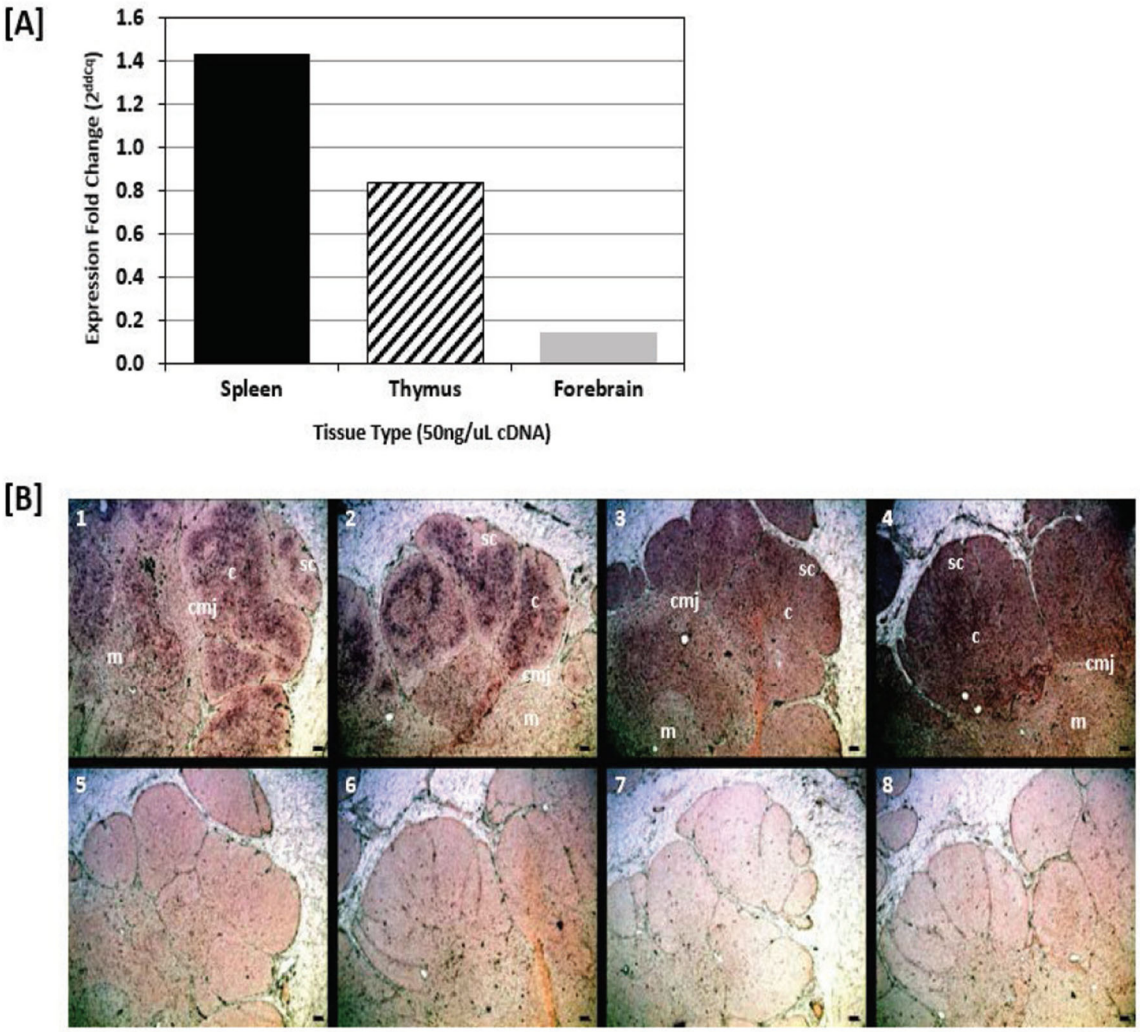

Figure 2-8. Shark thymus expresses AID. (A) Expression of AID in shark spleen (black bar), thymus (hashed bar), and forebrain (gray bar) using real-time quantitative PCR. Expression was measured using the delta delta Cq method and are normalized against shark muscle tissue using β2M as a reference gene. Data represent expression fold-change differences at four cDNA concentrations. (B) *In situ* hybridization of mRNA using ribo-probes on adult shark thymus sections. 1,2: Two different fields probing for TCRβ antisense. 3,4: AID antisense. 5,6: TCRβ sense. 7,8: AID sense. All micrographs at 10X magnification; black scale bar in lower right of each panel is 100 µM. Anatomical structures are designated on the top panels [c: cortex; m: medulla; sc: subcapsular region; cmj: corticomedullary junction].

selection occurs in mammals). Further, AID always co-localized with TCR αC (Figure 2-9). Thus, we observed a consistent pattern of expression where a "ring" of cells expressing both TCR αC and AID surround a central cell expressing only TCR αC. The more specific signal generated by FISH may suggest that, once a T cell completes RAG-mediated somatic rearrangement of its alpha chain locus, it clonally expands to form a ring of daughter cells around it. These daughter T cells then express AID (and TCR αC), promoting somatic hypermutation within their TCR alpha sequences during times when cells also undergo positive and negative selection.

2.3. Discussion

The role and diversifying mechanisms of SHM in B cells are well known (Li et al. 2004) as are the consequences of off-target AID activity (Álvarez-Prado et al. 2018). In B cells, AID mediates SHM within germinal centers of lymph nodes and spleen in mammals (Crouch et al. 2007) and we predict this is similarly occurring in the B cells zones identified in the shark splenic white pulp (Rumfelt et al. 2002). Somatic mutations occur in rearranged variable regions of B cells responding to antigen at rates of 10^{-3} mutations per base pair per cell division (Odegard and Schatz 2006). These changes are dominated by point mutations (and in shark, tandem mutations), biased towards transitions (G:A and C:T), and preferentially targeted to the AID motifs D<u>G</u>YW/WR<u>C</u>H (and less to W<u>A</u>/<u>T</u>W) (Li et al. 2004; Malecek et al. 2005; Odegard and Schatz 2006; Rogozin and Diaz 2004). The sequences of B cell V genes have evolved to maximize mutational effects, targeting the accumulation of replacement mutation within the antigen-binding CDRs and limiting

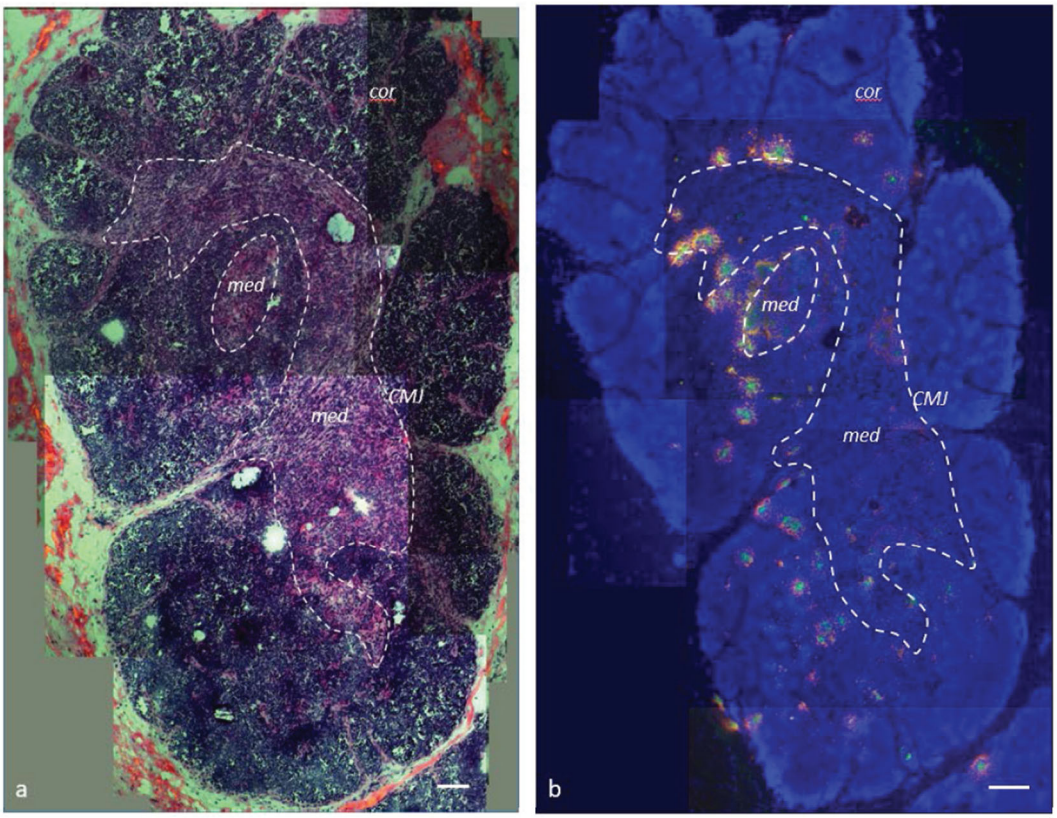

Figure 2-9. AID expression localized to inner cortex and cortico-medullary junction. (a) H and E staining of fixed shark thymus tissue illustrating thymic architecture (10x). The densely packed cells at the margins of the image comprise the cortex (cor), while the less densely packed cells in the center constitute the medulla (med). The region at the junction between cortex and medulla incorporates the corticomedullary junction (CMJ), delineated generally by a hashed white circle. [b-z] Single molecule RNA fluorescence *in situ* hybridization (FISH) probing fixed thymus sections simultaneously for AID (probes labeled with Quasar 670; pseudo colored red) and TCRα (probes labeled with CalFluor Red 610; pseudo colored green) and counterstained with DAPI (blue). (b) Composite of seven Z-stacked images (10x) depicting overall thymic architecture and the localization of AID expression to the inner cortex and cortico-medullary junction regions of shark thymus. We superimposed (and minimally adjusted) the outlined CMJ boundaries from (a) onto (b) to elucidate the junction between cortex and medulla. [c-z] We obtained images of each fluorophore using 10x, 20x, and 63x magnification and merged Z-stacked images together. Individual (10x) fluorophore images of DAPI [c-f], TCRα [g-j], and AID [k-n] and Z-stacked merged images [o-r] illustrate AID and TCRα expression in four locations of shark thymus. White boxes indicate the magnified regions of the 10x and 20x images shown in the 20x [s-v] and 63x [w-z] images, respectively. Scale bars [a,b,c,g,k,o] 150 mm, [s] 75 mm, and [w] 30 mm. [cor: cortex; med: medulla; CMJ: corticomedullary junction].

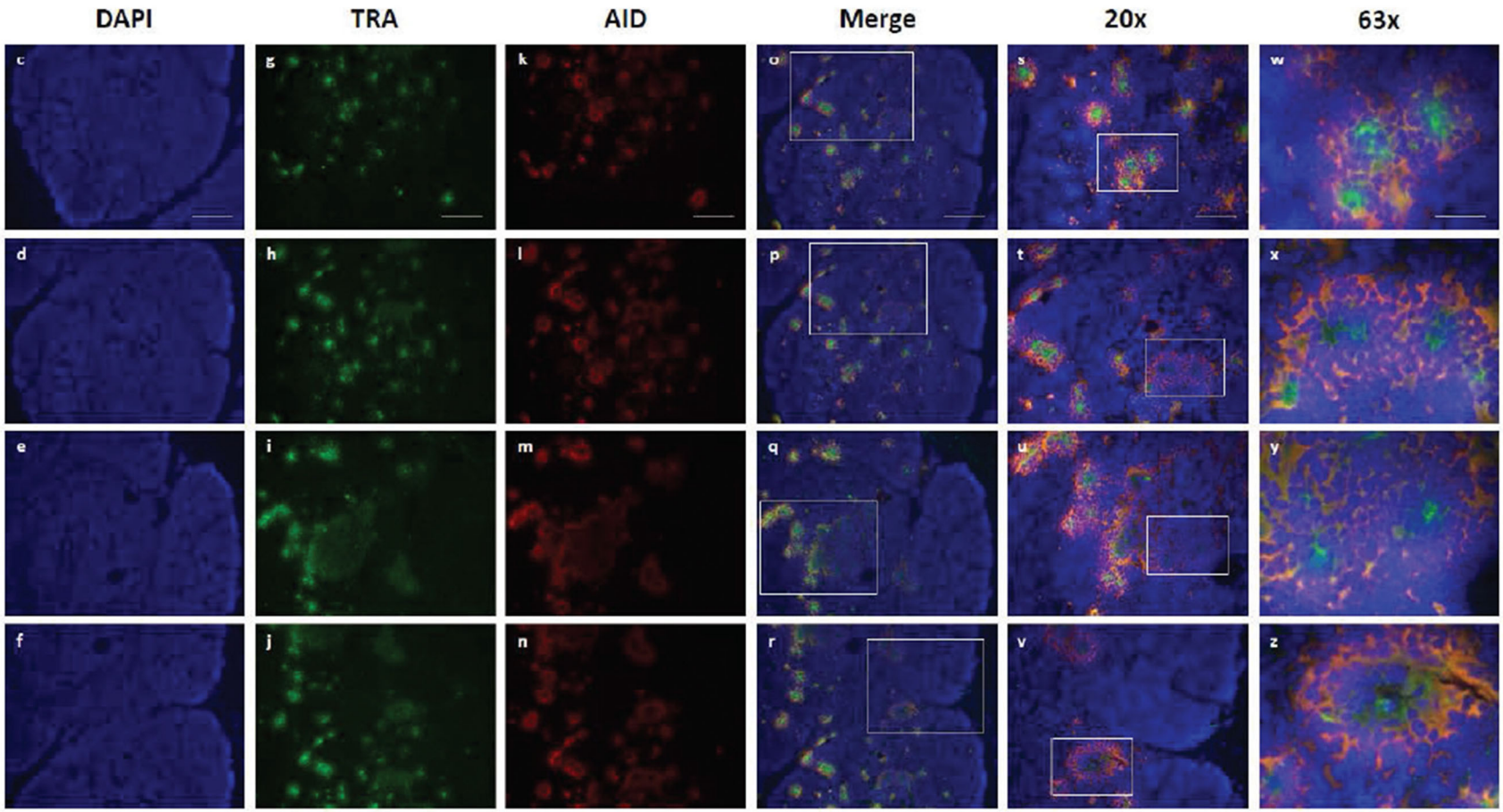

Figure 2-9. (*continued*)

mutation within the more structural FR. In humans, this focused mutation correlates with the long-term survival of B cell receptor repertoires (Saini and Hershberg 2015). The ability of B cells to use SHM for receptor diversification and improved antigen affinity is the basis of adaptive immunity (Saini and Hershberg 2015). Despite having similar developmental machinery as B cells (Gellert 2002), the assumption has long been held that αβ T cells do not undergo SHM because mutation could have deleterious effects on the binding of TCRs to MHC: Ag complexes (MHC: Ag) (Mantovani et al. 2002; Wagner et al. 1995). Despite some suggestion that SHM was occurring in T cells, studies designed to either quantify or characterize mutation in mouse or human TCRs did not gain traction and the textbook definition of SHM still defines it as an exclusively B cell mechanism (Murphy and Weaver 2017).

Recent studies reported the incidence of SHM in the γ chain of γδ T cells of sandbar shark (Chen et al. 2012; Chen et al. 2009) and in both γ and δ chains of dromedary camel (Antonacci et al. 2011; Ciccarese et al. 2014; Vaccarelli et al. 2012). In each study, SHM mirrored the mutational patterns observed in B cells during affinity maturation. However, in both sandbar shark γ and dromedary camel γ and δ chains, the authors hypothesized that T cells employ mutation as a means to generate a more diverse receptor repertoire rather than to improve receptor affinity to Ag (Antonacci et al. 2011; Chen et al. 2012; Vaccarelli et al. 2012). In contrast to αβ T cells, γδ T cells that interact with non-classical MHC often recombine tissue-specific, restricted sets of genes that have limited junctional diversity (Adams et al. 2005; Allison et al. 2001). Thus, it is reasonable

to consider that SHM could be used as a receptor-diversifying mechanism to fine-tune ligand recognition within a prospective tissue or to allow changes within the loci that allow receptors to evolve more rapidly to changing ligand environments (Adams et al. 2005; Kazen and Adams 2011). Further, many γδ T cells typically bind Ag in a manner more similar to that of Ig than to αβ T cells, recognizing and directly binding to small molecules and intact proteins without presentation by classical MHC: Ag complexes (Adams et al. 2005; Allison and Garboczi 2002; Allison et al. 2001). Inflammation stimulates activation of γδ T cells earlier in an immune response, releasing pro-inflammatory cytokines and killing infected macrophages. Thus, γδ T cells combine an innate-like response with an adaptive recognition strategy, providing both an immediate response to pathogen invasion and an ongoing, adaptive response to inflammation (Adams et al. 2005; Allison and Garboczi 2002; Allison et al. 2001). It is evident then how SHM presents a useful solution for accomplishing these tasks by creating a more diverse repertoire of these antibody-like γδ TCRs. Taken together, these studies clearly demonstrate that we can no longer regard SHM as a uniquely B cell mechanism. Considering the diversity of TCRs and TCR diversification mechanisms being found even in mammals (Hansen and Miller 2015; Miller 2010), perhaps we should prepare for more surprises in TCR antigen recognition.

In the present study, we verified SHM occurring within the γ and δ chains of γδ T cells in both thymic and peripheral immune tissue of nurse shark. Remarkably we also detected SHM occurring in the α chain of αβ T cells. We observed mutational

characteristics within α chain of nurse sharks similar to those found in B cell SHM. We observed an overall mutation frequency of 0.0226 substitutions per nucleotide (S/N) and a bias towards transition mutations. Further, we detected both single and tandem mutations, a pattern unique to sharks that also occurs in shark B cells. Changes to G and C nucleotides comprise 66.1% of all mutations. Mutation was twice as frequent in CDRs as in FRs (0.0352 versus 0.0188 S/N, respectively), and substitutions in CDRs were significantly more likely to result in amino acid changes. Further, mutations were strongly associated with AID hotspots, and substitutions to G and C nucleotides occurred nearly 1.4x as often within CDR hotspots than FR hotspots. Out of curiosity, we compared counts of AID hotspot motifs within CDR and FR regions between our eleven nurse shark TCR αV consensus sequences and six human TCR αV segments (V1.1, V1.2, V2, V3, V4 and V5). We found that shark TCR V segments exhibit far more WRCH/DGYW motifs per sequence than do human V segments (p=0.02). Further, motifs in CDRs of shark occurred 2-3x as often as in humans [human: average of 2.27 motifs per FR (range 1.97 – 2.65), 2.28 per CDR (range 1.97 – 2.59); shark: average of 3.25 motifs per FR (range 2.85 – 3.88), 5.09 per CDR (range 4.02 – 6.16); data not shown]. Importantly, the bias we found for nonsynonymous and non-conservative mutations in TCRα CDRs in the thymus are consistent with more than simple repertoire diversification; it suggests selection for changes in paratope.

We identified SHM from identical cDNA clones originating from both thymus and spiral valve tissues (see Table 2-6), suggesting that T cells with SHM-modified receptors

must have originated within the thymus and then traveled to peripheral gut-associated lymphoid tissue. Unsurprisingly, we detected the most AID expression within the inner cortex, medulla, and CMJ of shark thymus, where rearrangement and testing of TCRα takes place in mammals. Positive selection on self-MHC/ self-peptide for mature thymocytes begins with the CD4/CD8 double positive (DP) stage of development while differentiation into CD4/CD8 single positive (SP) cells requires that the TCR interact with MHC (Huesmann et al. 1991). If there is no TCR: MHC/peptide match found, T cell differentiation stalls with failure to be positively selected (Reinherz et al. 1999). However, the unusual nature of the TCRα locus, with up to 100 J segments depending on species, permits multiple successive rearrangements within a single cell, rescuing non-productive or self-selectable receptors with further gene rearrangements, a process called *receptor editing* (Bedel et al. 2012; Borgulya et al. 1992; Guo et al. 2002; Petrie et al. 1993). In mice, unlike the situation in developing B cells, receptor editing does not seem to rescue T cells from negative selection (Kreslavsky et al. 2013) and thus provides several opportunities for positive selection of DP thymocytes. Thymic nurse cells may help optimize these opportunities for selection by providing microenvironments favorable to secondary alpha chain rearrangement (Nakagawa et al. 2012).

In developing shark thymocytes, SHM in TCRα loci in conjunction with receptor editing (note that sharks, like all other gnathostomes, have a large number of TCRα J segments) could be involved in salvaging cells for positive selection or rescuing cells from death by negative selection. If AID-induced SHM occurs in conjunction with receptor

editing and positive selection, AID should be upregulated in cells undergoing RAG-mediated alpha rearrangement (and thus in cells also expressing RAG). However, if SHM occurs *after* rearrangement of TCR α and thus used for rescuing cells during negative selection, the same T cell would not express both AID and RAG. While we cannot determine conclusively without RAG expression data, the patterns of AID and TCRα expression (Figure 2-9) suggest that AID is upregulated after cells proliferate and diversify following alpha rearrangement (within the "ring" of cells). Thus, it is likely that AID is used primarily to rescue cells from negative selection, providing a 'mini-expanded self-referential repertoire' (Figure 2-10) and reducing the "profligate waste of thymocytes" (Murphy and Weaver 2017). However, based on the works above by Kreslavsky et al. and Nakagawa et al. in mice, we cannot discount the possibility that TCRs use SHM in conjunction with receptor editing for positive selection since developing shark T cells could still undergo negative selection after SHM (Figure 2-10). Further studies examining expression data from single cells could elucidate the timing of AID-catalyzed SHM in relation to T cell development. Further, we cannot completely rule out AID use in mature shark T cells, though our sequence data show no greater mutation frequency in the periphery, and abatement of AID expression in the thymic medulla are consistent with AID being a mechanism used only in T cell development.

These results are not without precedent. Qin *et al*. (2011) reported endogenous AID expression by peripheral $CD4^+$ T cells and immature B cells in mice. T cells that expressed AID also produce a distinctive cytokine profile, are associated with cell

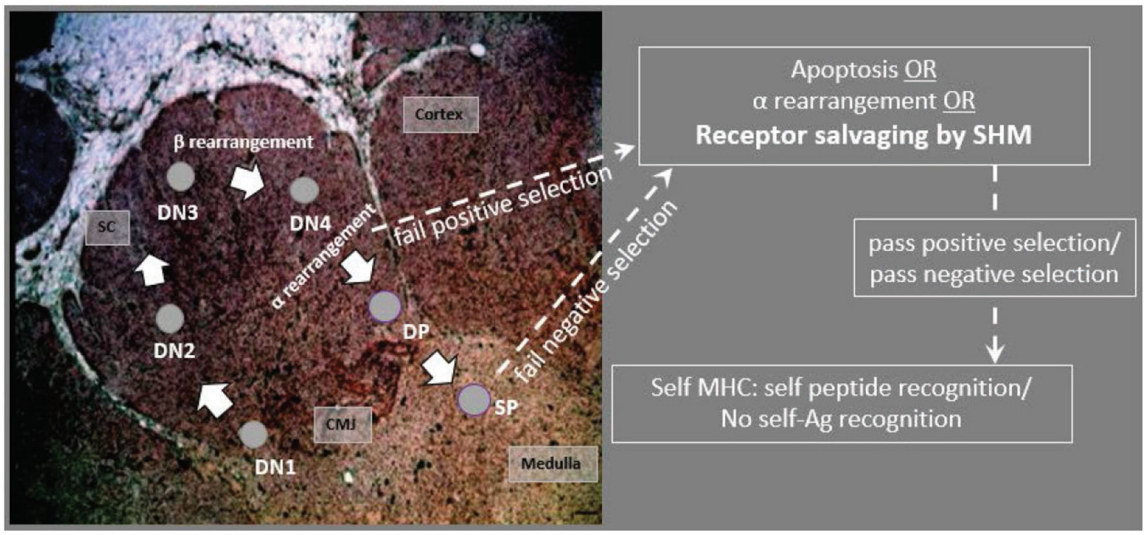

Figure 2-10. Model predicting how AID acts on T cells in the thymus. CD4/CD8 double negative (DN) thymocytes in the subcapsular region (SC) and cortex rearrange the b chain, using a surrogate pTa receptor to test for expression signaling. Cells with productive b arrangements then proliferate and express both CD4 and CD8, becoming double positive thymocytes (DPs). As DPs move toward the inner cortex and cortico-medullary junction (CMJ) where a chain rearranges, cells may begin to express AID. Non-productive rearrangements can be rescued from apoptosis by receptor editing or by receptor salvaging, in which AID catalyzes SHM to produce cells with improved affinity to MHC:Ag complexes (to pass positive selection). Salvaged thymocytes then proliferate and express either CD4 or CD8 on their surface as single-positive (SP) cells. AID-mediated receptor salvaging may also reduce recognition of self-peptide, rescuing self-reactive thymocytes from apoptosis (to pass negative selection). [sc: subcapsular region; cmj: corticomedullary junction; green shading indicates region of AID expression].

activation, and increase in abundance with age, suggesting these cells have distinctive long-term functions in aging cells. In immature B cells, AID expression may help negatively select autoreactive B cells and contribute to primary repertoire diversity. Together these results may indicate that AID expression in mice is a relic of a more extended expression in other species like sharks (Qin et al. 2011). Some mammals (e.g., rabbit, sheep, and cattle) use SHM in B cell primary repertoire diversification in the gut-associated lymphoid tissues (Alitheen et al. 2010; Archer et al. 1963; Becker and Knight 1990; Butler et al.

2011; Lanning et al. 2000; Reynaud et al. 1995; Reynaud et al. 1991b; Reynolds and Morris 1983). Even more recently, research implicated AID in central B cell tolerance in mammalian pre-B cells (Cantaert et al. 2015; Kuraoka et al. 2011), and there are examples of AID-mediated SHM being used alongside RAG-mediated V(D)J recombination in primary B cell repertoire generation in many mammals and in AID-mediated primary B cell diversification during gene conversion in the bird bursa of Fabricius (Reynaud et al. 1987; Thompson and Neiman 1987). We suggest that these B cell-specific features are much later mechanisms of AID-driven primary lymphocyte repertoire diversification and honing, perhaps mechanisms convergent with the thymic process we describe here. However, we predict that studies of this type in other vertebrate species will reveal the use of SHM as a T cell diversifying mechanism in a much broader collection of species.

Although recent studies report B cells present in the thymus of mice (Perera and Huang 2015; Perera et al. 2013), Miracle et al. examined B cell expression levels (IgM and IgX) in various tissues (including thymus) at different ages in the clearnose skate. While skate thymus expressed both IgM and IgX in early life stages, adult skates no longer express Ig in the thymus (Miracle et al. 2001). Criscitiello and Flajnik (2007) also found little evidence that Ig light chain is expressed in adult nurse shark thymus (see Fig S3 of paper) corroborating older northern blot studies with heavy chain in the species (Rumfelt et al. 2001; Rumfelt et al. 2004). A concurrent study within our lab used RT-qPCR to analyze IgM expression within shark thymus and found that, while young sharks do express IgM in the thymus, adult nurse sharks (which we used in the current study) do

not express IgM (data not shown). Further, preliminary results using FISH probes to IgM constant region also indicate that adult nurse sharks do not express IgM in the thymus. What little Ig expression we found in the thymus did not co-localize with AID expression.

Shark TCRs are capable of a wide variety of diversification mechanisms. In addition to RAG- mediated combinatorial and junctional diversity from complex loci, TCRδ (and possibly α) rearrange Ig-TCR chimera using IgM and IgW V exons with TCR D-J-C (Criscitiello et al. 2010). The TCR δ locus also encodes the doubly rearranging NAR-TCR, a δ chain with two diverse V domains (Criscitiello et al. 2006). Interestingly, our existing data do not suggest that SHM targets the IgHV or the TCR δV of the NAR-TCR δV chains for diversification. This suggests great control over the (nearly synchronous if not concomitant) expression of the potentially genotoxic AID and RAG. Thus, SHM at the γ, δ, and α loci adds to the battery of extraordinary diversification mechanisms used by shark lymphocytes in antigen recognition, although we do not yet understand the full effects of SHM on the animal's immunity to infectious disease. As for the dangers of any aberrant mutational activity genome-wide, sharks could be more resilient than other taxa due to their inherent slow rate of mutation (Martin 1999), possibly linked to the exceptional longevity of some individuals (Nielsen et al. 2016).

In the broader scope of lymphocyte evolution, we must consider whether the ancestral vertebrate lymphocyte employed APOBEC-family mediated diversification before the "big bang" of RAG (reviewed in (Hirano 2015)). Lamprey lymphocytes express at least two lymphocyte-specific cytidine deaminases (CDA1/CDA2) in the AID/ APOBEC

family. These deaminases emerged phylogenetically as the closest sister group in the AID/APOBEC family to the AID used by gnathostomes for Ig class-switch recombination, somatic hypermutation, and Ig gene conversion (Rogozin et al. 2007). CDA-mediated gene rearrangement in lampreys occurs in a manner similar to AID-induced immunoglobulin gene conversion in some birds and mammals (Rogozin et al. 2007; Zheng et al. 1994). One study suggested that VLRA (the analog to αβ TCR in jawed vertebrates) also might use CDA to affinity mature its receptors, indicating that CDA contributes both to repertoire generation and to somatic mutation after antigen exposure (Deng et al. 2010a; Flajnik 2014). If this is true, it may be possible that the ancestor to modern vertebrates also used an AID-like enzyme to assist with lymphocyte receptor development in a thymus-like organ. The expression of AID in the thymus of primitive sharks may be a remnant of this ancestral process, a mechanism lost in later vertebrates because of its potential for breaking down self-tolerance in mature lymphocytes. Perhaps agnathans evolved specific APOBEC molecules for diversification of their B and T like VLRs, while gnathostomes evolved AID for T cell primary repertoire diversification (Neils Jerne's "mutant breeding organ" (Jerne 1971)) and B cell affinity maturation, eventually co-opting AID for use in class switch recombination at IGH translocons, and later still, gene conversion and SHM for primary B cell repertoires.

From this trend of comparative TCR studies, we conclude with two hypotheses that we will test with further immunogenetic and functional studies in shark and other vertebrate models. First, the division between B and T cell repertoire diversification

components and mechanisms was not as clear-cut in ancestral lymphocytes as in modern humans and mice. The second is that different vertebrate groups have not only evolved myriad diversifications for Ig repertoires and function, but TCR biology may be just as varied. This premise is already accumulating ample supporting evidence as IgHV domains (Parra et al. 2012b; Parra et al. 2010), high allelic polymorphism (Criscitiello et al. 2004a; Criscitiello et al. 2004b), germline joined V exons (Wang and Miller 2012; Wang et al. 2011), and now mechanisms such as SHM, all once considered the immune privilege of Ig or MHC genes, are also employed for TCRs.

2.4. Materials and methods

2.4.1. Study animals

TCR sequence data used in this study came from two adult female nurse sharks ("Joanie" and "Mary Junior"; *Ginglymostoma cirratum*) delivered by caesarian section off the Florida Keys and matured in the aquatic vivarium of the University of Maryland's Center of Marine Biotechnology. We used published T cell αV sequences (Criscitiello et al. 2010) as a baseline for αV locus numbering (sharks Yellow and 1299), though we did not analyze any of these sequences for mutation.

2.4.2. Total RNA isolation and cDNA synthesis

We harvested tissues from animals after MS-222 (Argent, Redmond, WA) overdose, and immediately purified RNA with TRIzol reagent (Life Technologies, Carlsbad CA). Nurse shark thymi are located dorsomedial to the gills (Luer et al. 1995) in the

crevasse between the epaxial and brachial constrictor muscle groups. We used 5ug total RNA from spiral valve, spleen, thymus, and peripheral blood leukocytes (PBL) for oligo-dT primed cDNA generation with Superscript III First Strand Synthesis System (Thermo Fisher Scientific, Inc., Waltham, MA, USA). (Criscitiello et al. 2010)We estimated cDNA concentration using a Nanodrop 2000 Spectrophotometer (Thermo Fisher Scientific, Inc.).

2.4.3. RACE PCR, cloning, and Sanger sequencing

We generated a 5' RACE (Rapid Amplification of cDNA Ends) library using the GeneRacer Kit (Life Technologies) and reverse primers designed to the end of the shark TCRβ, TCRγ, or TCRδ variable (V) region or to the middle of the shark TCRα constant (C) region. We amplified RACE products using Phusion High-Fidelity DNA polymerase (New England Bio Labs, Inc., Ipswich, MA, USA) to minimize PCR errors under these specific PCR conditions: primary denaturing at 94^0C for 2 min; 30 cycles at 94^0C for 30s and 78^0C for 1 min; and a final extension at 72^0C for 10 min. Using this RACE library, we then amplified a specific α V region using a gene-specific primer to its leader region and the following PCR conditions: 98°C for 1 min; 25 cycles of 98°C for 5s, 49-60°C for 30s, 72°C for 150s; 72°C for 10 min. Annealing temperatures varied for each amplified αV (see Table 2-7). We visualized PCR products with agarose gel (8%) electrophoresis and then excised bands of correct size. We then isolated amplified bands from agarose gels using the PureLink Quick Gel Extraction Kit (Life Technologies) or RICO chips (TaKaRa Bio USA, Mountain View, CA, USA).

We transformed PCR amplicons into One-Shot Top10-competent cells (Thermo Fisher) using a pCR4-TOPO TA blunt end vector and cloning kit (Thermo Fisher) followed by a Zyppy plasmid miniprep kit (Zymo Research, Irvine, CA, USA) for plasmid purification of individual clones (Criscitiello et al. 2012). We checked insert size using an *Eco* RI restriction enzyme (Promega Corp, Madison, WI, USA), then amplified and purified the sequencing reaction using BigDye xTerminator Sequencing and Purification Kit (Thermo Fisher Scientific, Inc.). We submitted samples for sequencing to the DNA Technologies Core Lab on the Texas A&M University campus (College Station, TX). We deposited sequences in GenBank with the following accession numbers: *Alpha* KY189332-KY189354 and KY366469-KY355487; *Beta* KY351708-KY366487; *Gamma* KY351639-KY351707; *Delta* KY346705-KY346816.

2.4.4. Sequence alignment and tree building

We used Geneious and BioEdit (v7.2.5, Ibis BioSciences, Carlsbad, CA, USA) software to manage DNA sequence data. We aligned nucleotide and amino acid sequences using the ClustalW Multiple Alignment tool in Geneious with a gap penalty of 15, a gap extension penalty of 6.66, and free end gaps. We manually adjusted the alignments as necessary. We determined sequence relationships phylogenetically using the Geneious tree builder with default settings. We grouped sequences into unique V families based on 70% nucleotide sequence identity and 75% amino acid sequence identity (Brodeur and Riblet 1984; Rumfelt et al. 2004) using the same α V numbering

Table 2-7. List of forward (F) and reverse (R) primers used to generate T cell receptor (TCR) sequences and expression data. [AID: Activation induced cytidine deaminase; B2M: beta-2 microglobulin; α: alpha; β: beta; γ: gamma; δ: delta; V: variable region; C: constant region]

Primer	F/R	ID	Location	Nucleotide Sequence (5' to 3')	Amino Acid	Tm
TcR αV1	F	MFC370	leader region of αV1	ATG TTG CCT GAA GCT C	MLPEA	55
	R	MFC191	alpha C region	CAT TGG TGG ATA GCA AGC CCT TCG AT	SKGLLSTN	76
TcR αV4	F	MFC122	beginning of αV4	GTC TCC TCA GTT GTT CGT AC	VSSVVR	58
	R	MFC123	end of αV4	CAG TAA TAC ACA GCA GCG TC	DAAVYY	58
	F	MFC374	leader region of αV4	TGG ATT GTG TGG GCA GTA	WIVWAV	54
	R	MFC191	alpha C region	CAT TGG TGG ATA GCA AGC CCT TCG AT	SKGLLSTN	76
TcR αV5	F	MFC124	beginning of αV5	CTC AGG AAG GAG AGA TTA TCA C	QEGEII	60
	R	MFC125	end of αV5	CAA TGA TAC ACG GCG GAG TC	DSAVYH	60
	F	MFC124	beginning of αV5	CTC AGG AAG GAG AGA TTA TCA C	QEGEII	60
	R	MFC191	alpha C region	CAT TGG TGG ATA GCA AGC CCT TCG AT	SKGLLSTN	76
TcR αV7	F	MFC376	end of leader αV7	AGC GAT GGA GTT TCT GTG ATT	SDGVSVI	58
	R	MFC191	alpha C region	CAT TGG TGG ATA GCA AGC CCT TCG AT	SKGLLSTN	76
TcR αV10	F	MFC378	leader region of αV10	CTA TTT CTT CAC TAC CGC AG	YFFTTA	56
	R	MFC191	alpha C region	CAT TGG TGG ATA GCA AGC CCT TCG AT	SKGLLSTN	76
TcR α 5'	F	GeneRacer 5' Nested	homologous to RNA oligo	GGA CAC TGA CAT GGA CTG AAG GAG TA	--	78
TcR α 3'	R	MFC191	alpha C region	CAT TGG TGG ATA GCA AGC CCT TCG AT	SKGLLSTN	76
TcR βV1	F	MFC126	beginning of bV1	CTC CGT ACA TCG TCT CTA TTG	PYIVSI	60
	R	MFC127	end of βV1	CAC GCA CAG AAA TAG ACA GC	AVYFCA	58
TcR βV2	F	MFC128	beginning of βV2	CTA CGT GGA GCA GTC TCC ATC	YVEQSP	63
	R	MFC129	end of βV2	GCA CGC ACA ATA ATA GAC AGC C	AVYYCAC	62
TcR βV3	F	MFC130	beginning of βV3	CTA CGT GGA ACA GTC TCC TTC	YVEQSP	61
	R	MFC131	end of βV3	CAC GCG CAG AAA TAG ACA G	VYFCA	57
TcR βV5	F	MFCb50	beginning of βV5	GTT CGG TGC TCT TTC TCT GC	MFGALSLH	60
	R	MFCb54	end of βV5	GAC TGC AGT ATC AGT CGG CAC C	LVPTDTAV	66
TcR γV1	F	MFCg56	beginning of γV1	GTC GCT GTA TTA CTG GCT CAT TG	MSLYYWL	63
	R	MFCg59	end of γV1	GAG CGC ACA GTA ATA GGT GGC AG	TATYYCAL	67
TcR γV3	F	MFCg58	beginning of γV3	GAA GGG TCA CGT CCT TGC G	MKGHVLA	62
	R	MFCg61	end of γV3	GAT CCC AGA GTC ATC CTC	EDDSGI	56
	F	MFC170	beginning of γV3	CAA TAA CCA GAG CAC CGG G	ITRAP	56
	R	MFC171	end of γV3	AGA TCC CAG AGT CGT CCT C	EDDSGI	56
TcR δV3	F	MFCd62	beginning of δV3	GAT TCC CCG TCC CTG GTG TC	DSPSLVS	65
	R	MFCd66	end of δV3	CAG TGC ACA GTG ATA CAC AGC	AVYHCAL	61
TcR δV5	F	MFCd63	beginning of δV5	GCA GCT ACT CAG TAT CTG G	MQLLSIW	57
	R	MFCd67	end of δV5	GAA AGC ACA GTA ATA CAG AG	ALYYCAF	54
TcR δV7	F	MFC172	beginning of δV7	CTG TCA CTC AGT TAT TCT CCT C	VTQLFS	60
	R	MFC173	end of δV7	GCA GCC CAG TTA TAG TCA AAC	LTITGL	60
TcR δV12	F	MFC174	beginning of δV12	CAG AGC CCA CCT CAG TTA C	QSPPQ;	60
	R	MFC175	end of δV12	GAG CGC AGT AAT AGA TGG C	AIYYCA	57
TcR δV16	R	MFC176	end of δV16	GCA GCT CCG AGA TAG ACA AC	LSISEL	60
	F	MFC177	beginning of δV16	GAG TCC TGG CTC ACG CAA TC	ESWLTQ	63
TcR δV17	F	MFC178	beginning of δV17	CAG TCT TGG TCA GAA ATA ACC	QSWSEIT	57
	R	MFC179	end of δV17	CAA CTG AAG ATA AGT GAT CG	ITYLQL	54
AID	F	MFC342	beginning of AID exon 1	AGG CAC GAG ACC TAC ATG TTG	RHETYML	61
	R	MFC347	end of AID exon 2	TGA ACC AGG TGA GGC GGT A	YRLTWF	60
B2M	F	MFC211	first cysteine	AAC GTG TTG CTC TGT CAT GC	NVLLCHA	58
	R	MFC212	before second cysteine	GGG GTG AAC TCC ACA TAA CG	RYVEFTP	60

scheme as in Criscitiello *et al.* (Criscitiello et al. 2010). We created graphical alignments in BioEdit and imported these files into Microsoft Word to generate figures. Our preliminary dataset contained 564 TCRα clones (encoding 286 unique amino acid sequences representing nine V α families) from three tissues (PBL, spleen, thymus) of two sharks (*Joanie*, *Mary Junior*). Using this dataset, we separated sequences containing identical CDR3 rearrangements and counted mutations within each TCR "clone family" bearing the V-J rearrangement from single founder thymocytes.

2.4.5. Identification of TCR Vα genes in the nurse shark genome

We probed the filter sets for the *G. cirratum* BAC library (Arizona Genomics Institute) of shark "Yellow" and screened with variable segment and constant region probes for TCRα and TCRδ. We cultured several positive clones and isolated BAC DNA according to manufacturer's protocol with the Qiagen Large Construct Kit (Qiagen, Valencia, CA). We sent purified BAC DNA to the Duke University Center for Genomic and Computational Biology (Durham, NC) for PacBio SMRT (Menlo Park, CA) large insert (15-20kb) library preparation, sequenced on the PacBio RSII platform with P6-C4 chemistry. Read correction and contig assembly were performed with the PBcR software (Koren et al. 2012), using the BLASR error correction method and the Celera Assembler 8.2. We annotated the resulting sequencing within the Geneious software suite (v9.1.5, Biomatters Ltd., Auckland, NZ) using a custom BLAST database of all TCR and IgH sequences for *G. cirratum* in the IMGT database (Montpellier, France).

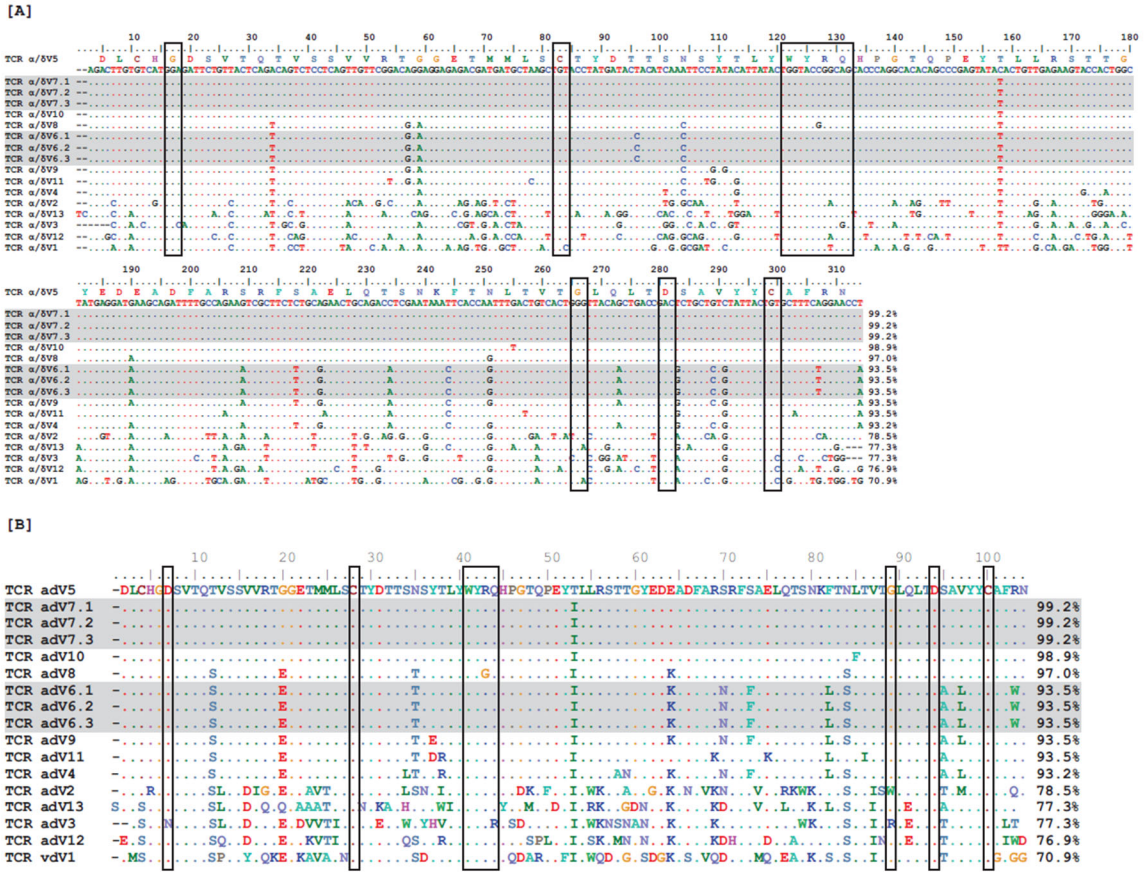

Figure 2-11. Observed TCR Alpha/ Delta germline Vs exhibit high sequence identity. Nucleotide [A] and amino acid [B] alignments of 17 germline variable (V) region gene segments. Two V groups contained three identical germline gene segments each (highlighted in gray), leaving only 13 unique V gene segments. Boxes surround conserved amino acids. Numbers at the ends of sequences indicate percent identity to the first germline sequence (α/δ V5) within the alignment.

Our search yielded 17 α/δ V germline segments, significantly fewer than expected based on TCR αV segment numbers in other species (Murphy and Weaver 2017). Of these 17 segments, only 13 contained unique nucleotide sequences, and all V segments were highly similar to each other (69-100% nucleotide and 52-100% amino acid identity). Twelve germline V segments shared >93% nucleotide identity (>85% amino acid identity), with three segments differing by only a single nucleotide (Figure 2-11). Based on the

variability we observed in our sequence data, these 17 germline α/δ V segments must represent only a small portion of the available Vs in the nurse shark genome.

We compared these 17 germline α/δ V segments to our TCR αV database containing all nine potential V families from two different sharks. All 17 germline α/δ V segments aligned to our αV4 data with >75% nucleotide identity, while 15 segments shared >93% nucleotide identity to at least one sequence in our αV4 dataset. Of the 60 sequences in our αV4 dataset, 37 sequences aligned specifically to eight germline α/δ V segments, with alignments containing one to 17 aV4 sequences per germline segment (nucleotide alignments shared >97% identity; Figure 2-12). While we did observe nucleotide differences within alignments, most differed by fewer than four nucleotides from the germline α/δ V segments. Because several germline segments differed by only a single nucleotide and we are certain that we have not found all α/δ V segments in the genome, these differences could represent variation in alleles or individuals rather than mutation. Thus, we chose not to rely on these data for mutational analysis.

2.4.6. Mutation frequency

We defined mutation frequency as the number of nucleotide changes divided by the total number of nucleotides within a particular region (e.g., FR, CDR, J, C) based on differences to a consensus sequence. We classified all nucleotide changes as either synonymous (SYN) or non-synonymous (NSYN) mutations based on whether or not the codon was unaltered or altered, respectively. For tandem base changes, we assessed the

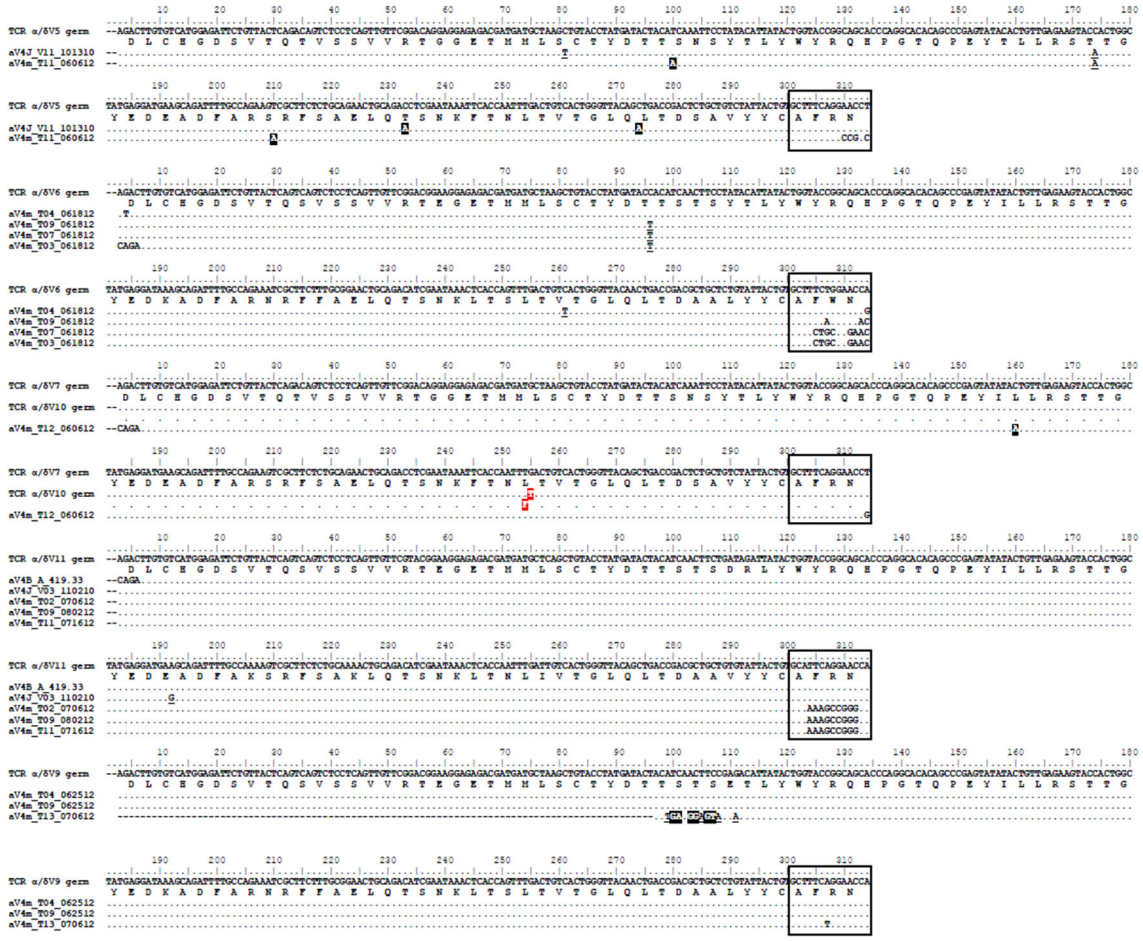

Figure 2-12. Observed germline sequences align only to TCR αV4 clones. Nucleotide alignments of TCR αV4 thymocyte clones to known germline V segments. Highlighted and underlined bases denote nonsynonymous and synonymous differences (respectively) to the germline V segment. Boxed regions represent nucleotides of the third complementarity-determining region (CDR3) according to IMGT guidelines, accounting for differences between clone and germline sequences. We highlight the single nucleotide/ amino acid change between α/δ V7 and α/δ V10 germline segments in red.

effect of each nucleotide change independent of its neighboring mutation(s). We then compared mutation frequencies between CDR and FR regions for all clone families that contained mutations using a Student's 1-tailed t-test unless otherwise noted.

2.4.7. Determination of hotspots

We searched for the ProSite motifs DGYW/WRCH (G:C mutable target) and WA/TW (A:T mutable target) using the motif search function in Geneious. These motifs serve as common "hotspots" for SHM within Ig variable regions, where AID favors the G/C bases within DGYW/WRCH motifs during the first phase of SHM while low-fidelity polymerases (i.e., polymerase η) preferentially target A/T bases within WA/TW motifs during the second phase of SHM (Chen et al. 2012; Rogozin and Diaz 2004; Wei et al. 2015). We counted only motifs present in the consensus sequence (rather than those created by the mutation) as hotspots. For each domain, we first counted the number of target nucleotides within DGYW/WRCH or WA/TW hotspots. We then examined each mutation to determine if it occurred inside or outside a hotspot. We counted changes from the consensus sequence to a target nucleotide within its respective motif as a hotspot mutation. We defined the frequency of hotspot mutation as the number of mutations (of target nucleotides) occurring within hotspots divided by the total number of mutations for each region. From these data, we compared the mutability of bases between FR and CDR regions using χ^2 Analysis.

2.4.8. Base substitution indices

We calculated a mutability index for each nucleotide using methods similar to Chen et al. (Chen et al. 2012). In our case, we derived the expected number of mutations by multiplying the frequency of a particular nucleotide within a family of sequences (e.g.,

α V2) by the total number of observed mutations within that family. We then defined the mutability index as in Chen et al (Chen et al. 2012) [the observed number of mutations of a specific nucleotide divided by the expected number of mutations of that nucleotide, with a value of 1.00 indicating random mutation]. We used χ^2 analysis to compare mutability indices between FR and CDR regions.

2.4.9. In situ *hybridization*

We used thymus tissue from an adult nurse shark for *in situ* hybridization as previously described (Criscitiello et al. 2010). We generated a probe for *G. cirratum* AID mRNA with primers NSAIDEH2 and NSAIDEH1 (Table 2-7) designed to amplify 211 base pairs of cDNA sequence of the first two AID exons (sequence and primers kindly shared by Ellen Hsu). We acquired images on an Axioscop2 microscope with AxioCam MRc5 (Zeiss, Thornwood CT) using Zeiss Axio Vision software.

Additionally, we performed fluorescence *in situ* hybridization (FISH) on adult nurse shark ("Black") thymus tissue. Slides contained two 8-µm thick sections of flash frozen thymus tissue preserved in OCT. We designed custom Stellaris® FISH Probes against the TCR alpha constant region (αC) for T cell identification and exons 1 and 2 of *AID* by utilizing the Stellaris® RNA FISH Probe Designer (Biosearch Technologies Inc., Petaluma, CA) available online at www.biosearchtech.com/stellarisdesigner (Version 2). We hybridized TCR αC with the CalFluor® Red 610 fluorophore and the AID sequence with the Quasar® 670 fluorophore for the Stellaris RNA FISH Probe set (Biosearch Technologies,

Inc.). We followed all manufacturer's instructions for frozen tissue (available online at www.biosearchtech.com/stellarisprotocols), allowing hybridization probes to incubate for 16 hours. We counterstained slides with wash buffer containing 5 ng/mL of DAPI (Sigma-Aldrich, St. Louis, MO). We obtained 10x, 20x, and 63x images using a Zeiss Stallion Digital Imaging Workstation including a 2x CoolSnap HQ Camera and Zeiss Stallion software. We merged Z-stacked images of each fluorophore together and edited and processed images using ImageJ software, version 1.47 (Schneider et al. 2012).

2.4.10. Real-Time qPCR for AID expression

We synthesized cDNA from nurse shark spleen, thymus, muscle, and forebrain RNA (see RNA purification and extraction methods above) using SuperScript III First-strand Synthesis System (ThermoFisher Scientific) and a 1:1 mixture of oligo-dT and random hexamer primers. We then amplified cDNA using touchdown PCR on an MJ mini thermal cycler (Bio-Rad, Hercules, CA) and GoTaq colorless DNA polymerase (Promega Corp) using the following conditions: primary denaturing at 94^{o}C for 2 min, five cycles at 94^{o}C for 30s and 56^{o}C for 4 min; five cycles at 94^{o}C for 30s and 54^{o}C for 4 min; 20 cycles at 94^{o}C for 30s, and 52^{o}C for 30 sec, and 72^{o}C for 4 min; with a final extension at 72^{o}C for 10 min. We visualized PCR products using agarose gel electrophoresis (as described above) to verify presence of AID in each tissue. We then cloned and sequenced the resulting PCR products to confirm the sequence was AID.

We looked for relative AID expression in shark spleen (positive control, where B cell AID-mediated SHM is known to occur), thymus, and forebrain (negative control) at four tissue concentrations (50ng, 25ng, 12.5ng, and 6.25ng) using the SYBR-green RT-PCR reagents kit (ThermoFisher Scientific) on a LightCycler 480 System (Roche Diagnostics Corp, Indianapolis, IN). We analyzed relative quantification using the LightCycler480 software and quantified relative AID expression using the ΔΔCq method. We normalized results against shark muscle tissue using beta2-microglobulin (β2M) as a reference gene. We present data as expression-fold changes of AID to β2M.

3. NURSE SHARK T CELL RECEPTORS EMPLOY SOMATIC HYPERMUTATION PREFERENTIALLY TO ALTER ALPHA/DELTA VARIABLE SEGMENTS ASSOCIATED WITH ALPHA CONSTANT REGION*

3.1. Introduction

Jawed vertebrates evolved a sophisticated immunoglobulin superfamily (IgSF)-based adaptive immune system composed of B and T cells, a polymorphic and polygenic major histocompatibility complex (MHC), recombination-activating gene (RAG)-mediated somatic recombination, and activation-induced cytidine deaminase (AID)-mediated somatic hypermutation (SHM) (Bernstein et al. 1994; Kasahara et al. 1992; Rast et al. 1997). This system relies on the rearrangement of variable (V), diversity (D), and joining (J) gene segments to generate the immunoglobulin (Ig) heavy and light chains of B cell receptors (BCR) and the four canonical T cell receptor (TCR) chains during lymphocyte development (Criscitiello and Flajnik 2007; Flajnik 2002; Rast and Litman 1994). Loci encoding each chain contain numerous V, (D), and J gene segments, and the resulting combinatorial potential results in a highly diverse immune repertoire (see Figure 3-1) (Schatz 2004). Each chain is encoded on separate loci (except TCR-delta, which is embedded within TCR-alpha) and loci are organized either as clusters of V, (D), and J segments followed by constant (C) region exons $(V\text{-}D\text{-}J\text{-}C)_n$ or as a contiguous translocon

*Reprinted with permission Ott, JA, Harrison, J, Flajnik, MF, Criscitiello, MF (2020). "Nurse shark T cell receptors employ somatic hypermutation preferentially to alter alpha/delta variable segments associated with alpha constant region" by *European Journal of Immunology* DOI:10.1002/eji.201948495. Copyright 2020 by Wiley.

containing numerous V segments, (D segments), and J segments followed by C region exons ($V_nD_nJ_nC$) (Hsu 2018; Jhunjhunwala et al. 2009). Lymphocytes further diversify antigen receptors during recombination by adding and subtracting nucleotides at gene segment joins, creating a unique third complementarity-determining region (CDR3) that is highly variable in sequence and length. Traditionally, after gene recombination, Ig heavy chains (IgH) dimerize with Ig light chains (IgL) to form BCR expressed on the B cell surface or antibodies secreted into the body humors, and TCR α and β or γ and δ chains dimerize to form canonical TCR expressed on the surface of T cells (see Figure 3-1) (Litman et al. 1999; Rast and Litman 1994). Together these mechanisms construct the efficient and effective adaptive immune repertoire necessary to respond to infection.

In B cells, receptor gene recombination occurs during lymphocyte development and cells exit bone marrow (or analogous primary lymphoid tissues such as epigonal organ in sharks) as naïve lymphocytes with functional receptors. Exposure to antigen in the follicles of peripheral lymphoid tissue activates naïve mature B cells, stimulating BCR to undergo affinity maturation. During this process, activation-induced cytidine deaminase (AID) catalyzes SHM of V regions followed by selection of the B cell, ultimately creating highly honed receptors for particular antigens (Li et al. 2004). Receptor gene recombination in T cells occurs similarly during thymic development. However, αβ TCR must undergo both positive and negative selection to ensure suitable binding to self MHC but not to self antigen; in this way, self-MHC-referential yet self-tolerant T cells emerge

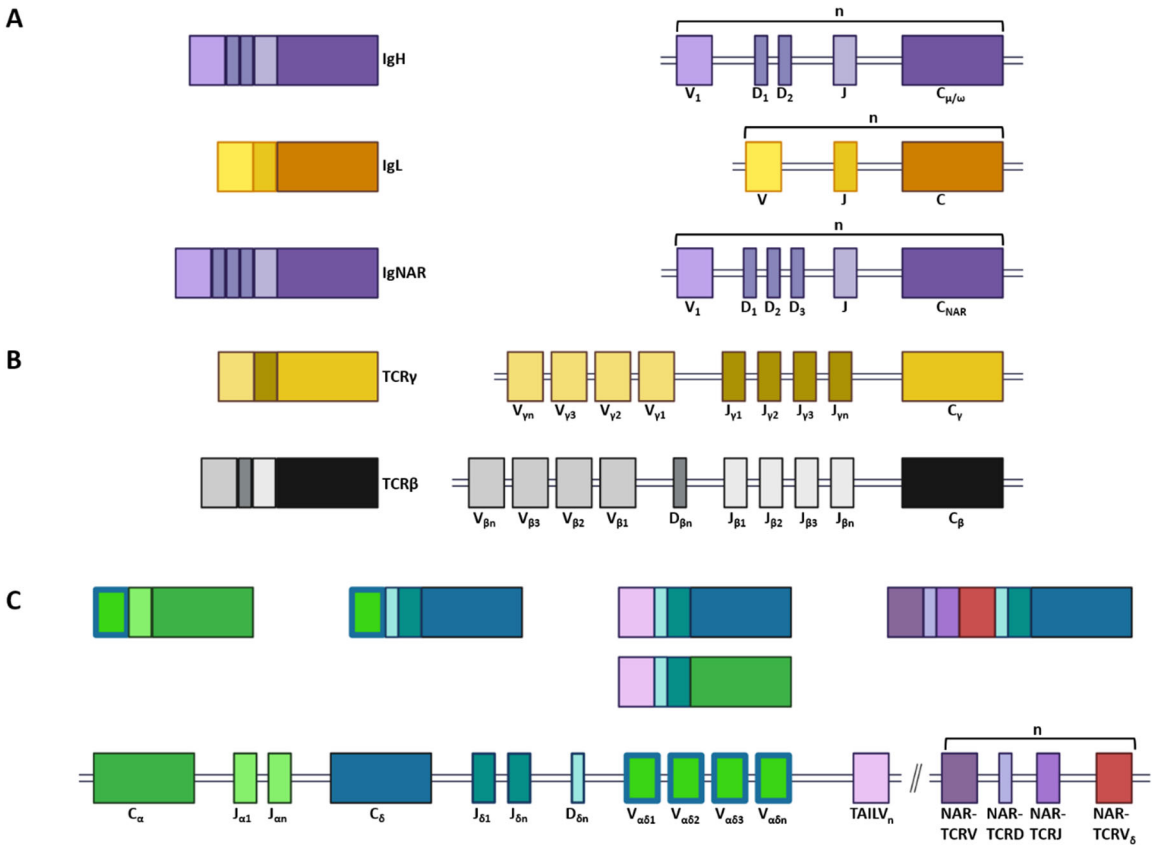

Figure 3-1. Nurse sharks generate complex B and T cell receptors. **[A]** Immunoglobulin (Ig) heavy (top) and light (middle) chains, and new (or nurse shark) antigen receptor (IgNAR, bottom) form from traditional B cell components. B cell (Ig) loci are organized in clusters of variable, (diversifying), and joining gene segments followed by a constant region (VDJ-C or VJ-C, shown as a single cluster for each chain). **[B]** T cell receptor (TCR) gamma (γ, top) and beta (β, bottom) chains form from traditional T cell components located within individual translocons. **[C]** T cell receptor delta (δ) chain is embedded within the TCR alpha (α) translocon, and share a common pool of V gene segments. Additionally, TCR-associated Ig-like V (TAILV) and both domains of NAR-TCR are found within this same locus. NAR-TCR gene segments occur in clusters of NARTCR V, NARTCR D, NARTCR J, NAR-TCR Vδ (shown as a single cluster). Rearranged gene segments create traditional TCRα and TCRδ chains as well as unique, untraditional receptors (TAILVα or TAILVδ, and both domains of NAR-TCR). Further, IgHV-TCRδ rearrangements integrate components of IgH and TCR loci. VDJ gene segments and constant regions are color coded to correspond to the receptor chain they encode. [Figure created with BioRender.com]

from the thymus as mature cells (Huesmann et al. 1991; Mantovani et al. 2002). Research in mice and humans demonstrates that unsuccessful receptors can be rescued by further locus rearrangement (receptor editing), but ultimately most cells undergo apoptosis and are removed from the potential repertoire (Bedel et al. 2012; Borgulya et al. 1992; Guo et al. 2009; Petrie et al. 1993).

Recent studies in nurse sharks (*Ginglymostoma cirratum*) and other non-model vertebrates suggest that the boundaries between B and T cell components and repertoire diversification mechanisms are blurred in comparison to mouse and human. For example, marsupials and monotremes (e.g., *Monodelphis domesticus*, *Ornithorhyncus anatinus*) contain a unique TCR locus (TCRμ) that contains V, D, and J gene segments that somatically recombine, or are pre-joined within germline DNA (Parra et al. 2007; Wang et al. 2011), to form a receptor chain with two variable domains, the membrane-distal of which resembles IgH. Further, IgHV or Ig-like TCR-delta V segments (VHδ) are found in TCR-alpha/delta loci of all gnathostome groups except teleosts and placental mammals (Breaux et al. 2018; Criscitiello et al. 2010; Deiss et al. 2019; Parra et al. 2008; Parra et al. 2012b; Parra et al. 2010; Saha et al. 2014). While many TCR-associating IgHV or VHδ genes are housed within the conventional αδ TCR locus, VHδ segments in Galliform birds are found in a second distinct TCR locus (Parra et al. 2012b). Nurse shark T cells assemble TCR using components traditionally considered BCR components, rearranging IgM or IgW (analogous to IgD) V segments to TCR alpha or delta constant (C) regions (see Figure 3-1), though it remains unclear whether sharks are using IgHV only from within the

conventional TCR-alpha/delta locus (*cis*-rearrangements) or are recombining Ig and TCR from separate loci (*trans*-rearrangements) as well (Criscitiello et al. 2010; Deiss et al. 2019). Doubly-rearranging NAR-TCR, composed of a membrane-distal Ig-like NAR V domain and a proximal, supporting TCR δV domain, also marries unique Ig and TCR components into a single receptor (see Figure 3-1) (Criscitiello et al. 2006; Venkatesh et al. 2014). Our lab recently discovered Ig-like V segments in nurse sharks that associate with TCR-alpha or TCR-delta C regions (T-cell-associated Ig-like V, TAIL V, see Figure 3-1) (Deiss et al. 2019). Additionally, T cells can exploit BCR diversification mechanisms like AID-catalyzed SHM to generate additional thymic diversity: Chen et al. (2012) presented definitive evidence that sandbar sharks utilize SHM to diversify gamma chain of γδ T cells, and camels employ SHM to diversify both TCR gamma and delta chains (Antonacci et al. 2011; Chen et al. 2012; Chen et al. 2009; Ciccarese et al. 2014; Vaccarelli et al. 2012). Additionally, nurse sharks utilize SHM for AID-catalyzed receptor salvaging to assist thymocytes through selection during thymic development (Ott et al. 2018). Thus, gnathostome adaptive immunity displays remarkable elasticity in T cell diversification mechanisms.

We examined a large dataset of TCR sequences to assess whether nurse sharks utilize SHM specifically for alpha chain receptor salvaging or if SHM affects other canonical TCR chains and non-canonical receptors (IgH-TCRC rearrangements, NAR-TCR, and TAIL V-TCR C) alike. Additionally, this dataset compelled us to revise the current

nomenclature for V gene segments within the alpha/delta (TCR-alpha/delta) locus. Finally, we examine the use of SHM in light of the immunogenetic elasticity observed within the nurse shark TCR-alpha/delta locus.

3.2. Results

3.2.1. Canonical nurse shark T cell receptor chains suggest few V segment families with many subfamilies

Our TCR data set contained 229 TCR-beta (TCR βV), 158 TCR-gamma (TCR γV), and 761 TCR-alpha/ delta (TCR αδV) newly cloned or previously published V gene sequences (1149 total clones, Table 3-1, see Table Supplement A-1 for accession numbers of new and published sequences). Using a refined approach to grouping V segments, we reduced the putative number of published TCR βV families to four, with TCR βV1 and TCR βV2 containing four and two subfamilies each, respectively (Figure 3-2A; Figure Supplements A-3, A-7). We reclassified TCR γV clones into four families (TCR γV1 – TCR γV4) with multiple subfamilies in all but TCR γV4 (Figure 3-2B; Figure Supplements A-3, A-7). We did not identify any alleles for either chain. The 761 TCR-alpha/delta clones sorted into 11 putative TCR αδV families and 24 subfamilies (Figure 3-2C; Figure Supplements A-4, A-7). Five of the 11 TCR αδV families spliced only to TCR-alpha constant (C) region (TCR αδV 1, 4, 5, 7, and 9), two utilized only TCR-delta C (TCR αδV 3 and 10), and four spliced interchangeably with TCR αC and TCR δC (TCR αδV 2, 6, 8, and 11). Five TCR αδV subfamilies included at least two alleles. For all canonical TCR chains, complete V gene

segments contained the conserved tryptophan and two cysteine residues common to the IgSF domain (see Figure 3-2) except TCR αδV11 (which includes only the conserved cysteine at position 104). This finding contrasts previous results indicating that TCR δV16 (TCR αδV8.2) lacks the first cysteine residue (Criscitiello et al. 2010).

3.2.2. Non-canonical T cell receptor variable gene segments are highly conserved

In addition to the canonical αβ and γδ TCR, we previously identified three "chimeric" nurse shark TCR containing Ig or Ig-like components: 1) IgHV can be associated with TCR-delta C [(δC) or rarely, TCR-alpha C (αC)], rearranging an IgM or IgW (analogous to IgD) V segment to a TCR-delta (or TCR-alpha) C region (Criscitiello et al. 2010); 2) doubly-rearranging NAR-TCR are composed of membrane-distal Ig-like V (NTCR V) and membrane-proximal or "supporting" TCR δV (STCR δV) (Criscitiello et al. 2006); and 3) T-cell-associated Ig-like V (TAIL V) segments recombine Ig-like V and D segments to TCR J segments and can associate with either TCR αC or TCR δC regions (see Figure 3-1) (Deiss et al. 2019). We sequenced 195 IgMV-TCR δC, 77 IgWV-TCR δC, 69 NAR-TCR (51 NTCR V and 62 STCR δV complete domains), and 9 TAIL V clones (Table 3-2). IgHV-TCR δC clones aligned with five of the six canonical IgM germline groups (IgM V1-V5) and three of the six canonical IgW groups (IgW V1-V3) (Malecek et al. 2008). We identified three IgM V2 subfamilies and two IgW V1 subfamilies in our dataset (Figure 3-2D; Figure Supplements A-5, A-7). IgMV2C is an Ig pseudogene (due to defective Ig constant region exons) but we

Table 3-1. Summary of T cell receptor (TCR) alpha (TCRα), delta (TCRδ), gamma (TCRγ), and beta (TCRβ) chain sequence data used in this paper. Alpha and delta V segments are encoded by the same locus and thus are defined by the spliced C region. Putative families share at least 70% nucleotide identity (subfamilies within each TCR V family share at least 80% nucleotide identity) using nearest-neighbor consensus trees of V segments. Alleles share at least 90% nucleotide identity but differ from each other by the same set of base changes (observed in more than one shark). Number of TCR nucleotide (NUC) or amino acid (AA) sequences or sequence groups within each category. Sums include sequences for each locus independently.

C Region	V Segment	Putative # Subfamilies: Alleles	All cloned sequences	Complete CDR3-J junction	Unique V Region[a]		Unique V segment[b]		Unique CDR3[c]		Groups with identical CDR3-J[d]	# Sequences in each CDR3-J group[e]
					NUC	AA	NUC	AA	NUC	AA		
TCRα	TCRαδV1	4:1/1/1/1	57	57	53	52	29	27	47	46	8	5,2,2,2,2,2,2,2
TCRδ				0								
TCRα	TCRαδV2	4:6/2/1/1	97	26	22	21	60	50	20	20	5	2,2,2,3,3,3
TCRδ				18	18	18			18	18	0	
TCRα	TCRαδV3	1:1	36	0	13	12	8	4	12	12	2	24,2
TCRδ				36								
TCRα	TCRαδV4	2:1/3	243	194	53	52	41	41	51	50	5	130,11,2,2,3
TCRδ				0								
TCRα	TCRαδV5	1:1	3	3	3	3	3	3	3	3	0	
TCRδ				0								
TCRα	TCRαδV6	1:1	26	24	3	3	3	3	3	3	2	21,2
TCRδ				2	2	2			2	1	0	
TCRα	TCRαδV7	1:1	8	7	7	7	6	4	7	7	0	
TCRδ				0								
TCRα	TCRαδV8	6:1/10/2/1/1/1	208	65	55	53	78	77	55	53	7	4,3,2,2,2,2,2
TCRδ				71	11	11			11	11	2	57,2
TCRα	TCRαδV9	1:1	3	3	1	1	1	1	1	1	0	
TCRδ				0								
TCRα	TCRαδV10	1:1	35	0	11	11	6	6	11	11	0	
TCRδ				11								
TCRα	TCRαδV11	2:1/1	45	43	28	27	13	9	24	23	3	2,2,2
TCRδ				2	2	2			2	2	0	
TCRαδV Sum	**11**	**24:42**	**761**	**562**	**282**	**275**	**248**	**225**	**267**	**261**	**34**	
TCRγ	TCRγ1	6:1/1/1/1/1/1	62	45	38	38	31	26	38	38	5	2,2,2,2,3
	TCRγ2	2:2/2	18	18	17	17	13	13	17	17	2	2,2
	TCRγ3	4:1/1/1/1	60	22	20	20	17	16	20	20	2	2,2
	TCRγ4	1:1	18	4	4	4	3	3	4	4	0	
TCRGV Sum	**4**	**13:15**	**158**	**89**	**79**	**79**	**64**	**58**	**79**	**79**	**9**	
TCRβ	TCRβ V1	4:1/1/1/1	92	92	88	87	40	25	82	82	3	4,2,4
	TCRβ V2	2:1/1	49	49	46	46	23	13	46	46	3	2,2,2
	TCRβ V3	1:1	52	51	43	43	25	13	43	43	12	7,3,2
	TCRβ V4	1:1	36	36	34	34	23	16	33	33	3	2,2,2
TCRβV Sum	**4**	**8:8**	**229**	**228**	**211**	**210**	**111**	**67**	**204**	**204**	**21**	

a) V Region includes all bases from the 1st predicted nucleotide of the V segment to the last predicted nucleotide of the J segment (V and J).
b) V Segment includes all bases from the 1st predicted nucleotide of the V segment to the last predicted nucleotide of the V segment (V only) and is determined from all cloned sequences irregardless of constant region.
c) CDR3 includes all bases between the C (YxC motif, position 104) of the V segment and the F (FGxG motif) of the J segment.
d) Number of groups with identical CDR3-J sequences.
e) Number of sequences within each V group containing identical CDR3-J regions.

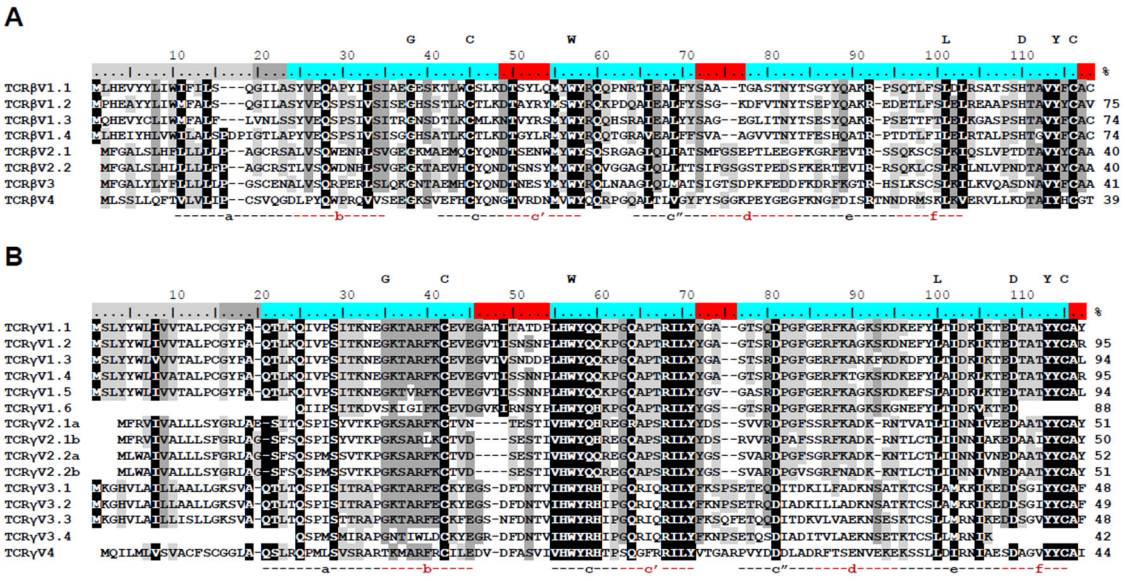

Figure 3-2. Consensus sequence alignments for T cell receptor V segments indicate substantial conservation between segments. V gene segments are grouped by identity for [A] beta (TCRβV) [B] gamma (TCRγV), [C] alpha/delta (TCRαδV), [D] immunoglobulin (Ig) (IgMV and IgWV) and TCR-associated Ig-like V (TAILV), and [E] NAR-TCR distal V domain (NTCRV) and [F] NAR-TCR proximal Vδ domain (STCRδV). V segment families share >70% nucleotide identity (e.g., TCRαδV2) and subfamilies have >80% nucleotide identity (e.g., TCRαδV2.1). Alleles share >90% nucleotide identity and common differences appear in more than one shark (e.g., TCRαδV2.1a). Letters above the scale denote conserved residues of antigen receptor domains. Regions below the alignment designate predicted beta strand location and direction. Shading within an alignment indicates amino acid conservation [Blosum62 score matrix (Threshold=1): black=100% similar; dark grey= 80-100%; light grey=60-80%]. Values to the right of the alignments show the percent nucleotide identity to the first sequence. Highlighting within the scale indicates leader peptides (gray), framework regions (blue), and complementarity-determining regions (CDR, red). Coloring within the TCRαδV consensus sequence names identify the constant region used (green=TCR αC; blue=TCR δC; orange=TCR αV or TCR δC). IgM or IgW germline sequence accession numbers are in sequence titles. IgM V2C is an Ig pseudogene due to defective Ig constant region exons but can form functional transcripts when associated with TCR-alpha or TCR-delta C. Each NAR-TCR V domain is encoded by V gene segments from a single gene family, but we employed original names to indicate the NARTCR cluster used. Gaps within a sequence are for alignment purposes only.

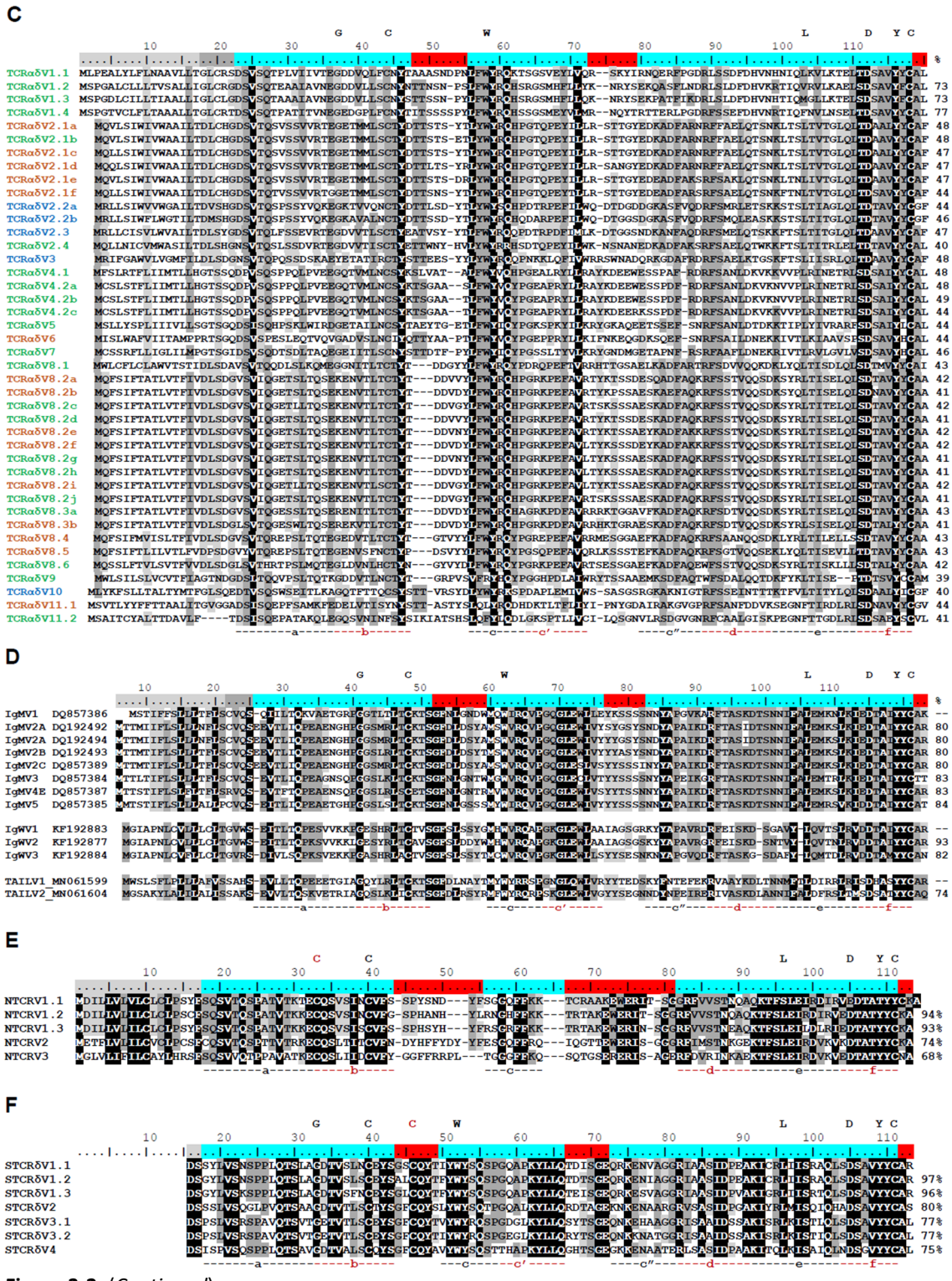

Figure 3-2. (*Continued*)

observed functional transcripts associated with TCR δC. Interestingly, our 5' RACE libraries primed with TCR δC-specific primers generated more clones associated with IgM/IgW V segments (58%) than to canonical TCR αδV segments. These libraries comprised data from two "young" sharks (Tom Thumb, a neonate and Florence Nightingale, 3' in length) and two "old" sharks (White and Grumpy, both greater than 8' in length). Libraries from younger sharks generated more canonical TCR αδV-TCR δC arrangements (69 of 115 clones, 60%) and those from older sharks generated more non-canonical IgHV-TCR δC arrangements (75 of 126 clones, 60%). However, further study characterizing IgHV-TCR δC rearrangements is required to verify these observations.

Based on our conservative naming strategy, all NTCR V gene segments belonged within a single family containing three subfamilies, and subfamily NTCR V1 included four different alleles (Figure 3-2E; Figure Supplements A-6, A-7). All "supporting" V gene segments (STCR δV) comprised a single gene family composed of four subfamilies. Both STCR δV1 and STCR δV3 contained multiple alleles. However, we retained subfamily names in V segment identities for consistency with published data. We observed multiple combinations between NTCR V and STCR δV domains, but in general NTCR V1 associated with STCR δV1 (NTCR V1.1- STCR δV1.1a, NTCR V1-STCR δV1.1b; NTCR V1.2-STCR δV1.1b; NTCR V1.3-STCR δV1.2; NTCR V1.4-STCR δV1.3), NTCR V2 associated with both STCR δV2 and STCR δV4, and NTCR V3 associated with STCR δV3 (STCR δV3.1 and STCR δV3.2). In addition to the conserved tryptophan and two cysteine residues found in other TCR, all

Table 3-2. Summary of non-canonical T cell receptors (TCR) data used in this paper. TAILV segments are TCR-associated immunoglobulin (Ig)-like V segments. Trans-rearrangements assemble Ig heavy chain variable regions (IgM or IgW) to alpha (TCRα) or delta (TCRδ) constant regions. Doubly-rearranging NAR-TCR recombine both a NAR (NTR) V segment and a supporting TRD (STRD) V segment to a TRD constant (C) region. Putative families share at least 70% nucleotide identity (subfamilies within each TCR V family share at least 80% nucleotide identity) using nearest-neighbor consensus trees of V segments. Number of TCR nucleotide (NUC), amino acid (AA) sequences or sequence groups within each category.

C Region	V Segment	Putative # Subfamilies: Alleles	All cloned sequences	Complete CDR3-J junction	Unique V Region[a]		Unique V segment[b]		Unique CDR3-J[c]		Groups with identical CDR3-J[d]	Sequences in each dataset[e]
					NUC	AA	NUC	AA	NUC	AA		
TCRα	TAILV1	1:2	9	9	5	5	1	1	5	5	2	3,3
TCRδ	TAILV1	1:2	1	1	1	1	1	1	1	1	0	
	TAILV2	1	1	1	1	1	1	1	1	1	0	
	TAILV Sum	**2**	**11**	**11**	**7**	**7**	**3**	**3**	**7**	**7**	**2**	
	IgM V1	1	69	67	65	65	21	17	65	65	2	2,2
	IgM V2	3:2,1,1	74	69	38	35	13	11	36	32	5	16,12,6,2,2
	IgM V3	1	5	1	1	1	3	2	1	1	0	
	IgM V4	1	2	2	2	2	1	1	2	2	0	
	IgM V5	1:3	45	37	26	24	9	9	26	24	6	4,4,3,2,2,2
	IgW V1	2	36	28	27	27	21	19	27	27	1	2
	IgW V2	1	34	29	29	29	25	17	29	29	3	6,3,2
	IgW V3	1	7	4	4	4	4	3	4	4	0	
	IgHV Sum	**11**:14	**272**	**237**	**192**	**187**	**97**	**79**	**190**	**184**	**17**	
	NTCR V1	1:4	27	25	24	24	24	24	24	24	0	
	NTCR V2	1	17	17	17	17	17	17	17	17	0	
	NTCR V3	1	7	5	5	5	5	5	5	5	0	
	STCRδ V1	1:3	24	23	23	23	12	12	23	23	0	
	STCRδ V2	1	26	20	20	20	11	11	20	20	0	
	STCRδ V3	1:2	6	6	6	6	4	4	6	6	0	
	STCRδ V4	1	6	1	1	1	1	1	1	1	0	
	NARTCR Sum	**7**:13	**113**	**97**	**96**	**96**	**74**	**74**	**96**	**96**	**0**	

a) V Region includes all bases from the 1st predicted nucleotide of the V segment to the last predicted nucleotide of the J segment (V and J).
b) V Segment includes all bases from the 1st predicted nucleotide of the V segment to the last predicted nucleotide of the V segment (V only).
c) CDR3-J includes all bases between the C (YxC motif, position 104) of the V segment and the F (FGxG motif) of the J segment.
d) Number of groups with identical CDR3-J sequences, which we used to determine sequence relatedness (see text for details).
e) Number of sequences within each V group containing identical CDR3-J regions.

functional NAR-TCR sequences contained the non-canonical inter-domain cysteine in FR1 of NTCR V (and CDR1 of STCR δV) required for domain stability (Flajnik et al. 2011).

3.2.3. Hotspot motifs in nurse shark T cell receptor variable segments do not necessarily predict mutation

AID preferentially alters C and G residues of WRCH/DGYW motifs of antigen receptors (Rogozin and Diaz 2004). The number of WRCH/DGYW AID hotspot motifs in CDR did not differ from FR motifs in any of the canonical TCR V segments (see Figures 3-3, 3-4). However, FR2 of TCR βV contained more WRCH/DGYW motifs than other FR. As expected, both CDR of IgHV contained more motifs than FR domains. NAR-TCR domains (NTCR V and STCR δV) contained the fewest WRCH/DGYW motifs in any region of all V segment types and within NTCR V, most WRCH motifs overlapped DGYW motifs, a pattern not observed in other V gene segments. Motif patterns did not vary by region. Thus, motif patterns alone do not predict mutation.

3.2.4. Mutation occurs in TCR αδV associated with TCR αC but not TCR δC or other canonical T cell receptor chains

SHM within TCR was first identified in TCRγ chain of sandbar sharks (Chen et al. 2012; Chen et al. 2009). However, although mutation appeared to target nucleotide motifs preferred by AID (WRCH/DGYW), mutation tended not to result in amino acid replacement within CDR, a requisite for paratope changes during affinity maturation.

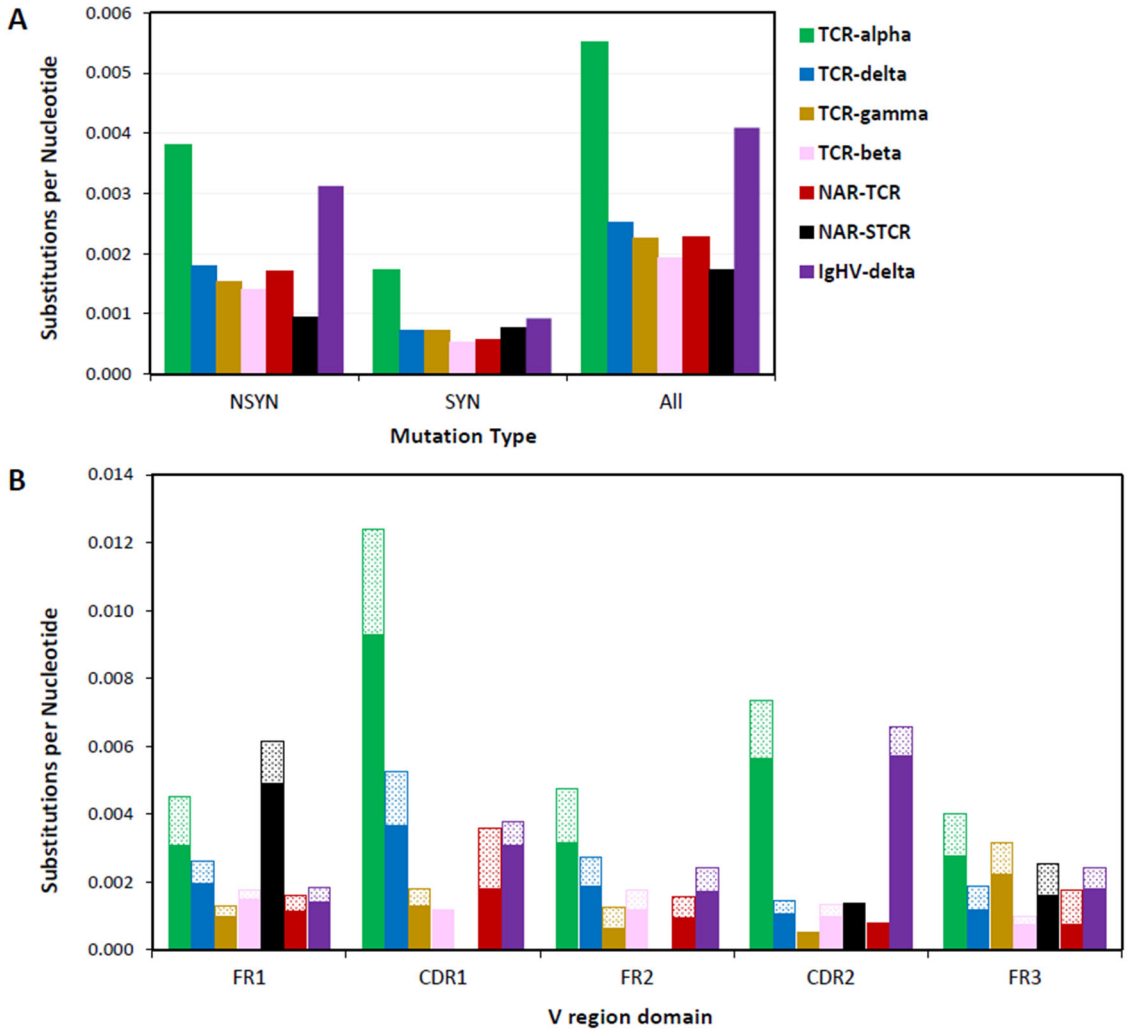

Figure 3-3. Despite the presence of AID hotspot motifs within all canonical V segment types, only TCR αδV segments associated with TCR-alpha constant (TCR αC) regions accumulate significant mutation. Heatmap coloring indicates [A] the proportion of sequences (n=number of sequences represented) within each segment type that contain a mutation or the [B] proportion of Vs within each segment type (n=number of V family groups) that contain a G/C target (within a DGYW/WRCH motif) at any single position along the sequence length (5' to 3'). White boxes indicate the locations of conserved antigen receptor domain residues. Putative locations of framework regions (FR, light blue) and complementarity-determining regions (CDR, pink) are indicated above the heatmaps. Gaps (grey space) within the sequences are for alignment purposes only. (Segments are aligned by domain and conserved residue placement only.) Numbers within each color scale represent the greatest value for that color. [TCR: T cell receptor, V: V segment, TCR αδV: TCR-alpha/delta V, TCR βV: TCR-beta V, TCR γV: TCR-gamma V, IgHV: immunoglobulin heavy chain V, NTCR V: NAR-TCR V, STCR δV: supporting TCR-delta V (NTCR V-STCR δV for doubly-rearranged NAR-TCR)] [Source data can be found in Appendix B-3]

Rather, sandbar sharks appeared to use SHM to generate a more diverse repertoire (Chen et al. 2012; Chen et al. 2009). We previously confirmed that SHM occurred within TCRγ and TCR δV segments of nurse shark, but we found that SHM altered TCR αV far more than it did gamma or delta, with replacement mutation targeting AID-preferred motifs of CDR within the thymus. This suggested that TCR-alpha likely uses SHM to salvage failing receptors during thymic selection (Ott et al. 2018). In both B and T cell receptors of sharks, SHM can occur as single point mutations or tandem mutations of two or more contiguous nucleotides, indicating at least two different cellular mechanisms generate mutations (Anderson et al. 1995; Lee et al. 2002; Rumfelt et al. 2002).

We first attempted to corroborate earlier findings of SHM in canonical TCR of nurse sharks. However, despite having unique CDR3 regions, TCRβ and TCRγ showed very little variation within V segment nucleotide sequences (see Table 3-1), with about 0.002 substitutions per nucleotide (S/N) for both chains (See Figures 3-3, 3-4, 3-5A). While we observed contiguous mutations within both chains, the majority of mutation occurred as single base changes and most base changes resulted in amino acid replacement (R) rather than silent (S) mutation (TCR βV: R/S=2.7; TCR γV: R/S=2.1; Figures 3-3, 3-5A). As previously observed, TCR γV segments contained substantially more mutation than TCR βV segments. While TCR βV sequences accumulated more mutations to FR1 (0.0018 S/N) and fewer to CDR1 (0.0012 S/N; Figures 3-3, 3-4, 3-5A), TCR γV mutation was highest in FR3 (0.0032) and lowest in CDR2 (0.0005; Figure 3-3, 3-4, 3-5A). Most mutations in TCR

βV (61%) and TCR γV (67%) were to C and G nucleotides, though only 35% of C/G mutations in TCR βV and 30% in TCR γV actually occurred within WRCH/DGYW motifs (Figures 3-3, 3-4, 3-5; Table 3-3). Both TCR βV and TCR γV mutations were biased towards transitions (TCR βV FR: 49%, CDR: 44%; TCR γV FR:56%, CDR:63%; Table 3-4). TCR βV4 sequences exhibited the most mutation, with 64% of sequences (23 of 36) containing at least one nucleotide change (Figure Supplement A-3). Again, we observed no mutation within any TCR γV4 segment (except a single base change in a leader region), suggesting this V segment may be useful as a partner chain with non-canonical receptors or receptors highly specific for particular antigen.

We analyzed mutation data for V segments within the TCR-alpha/delta locus only if we could identify with certainty the constant region associated with the V segment. Thus, we included 422 TCR αδV associated with TCR-alpha C (TCR αC) and 140 TCR αδV associated with TCR-delta C (TCR δC) in our analyses. We first confirmed previous results that CDRs of TCR αδV-TCR αC mutated significantly more than FRs (Figures 3-3, 3-4, 3-5B), with CDR1 accumulating the most mutation (Ott et al. 2018). Within the TCR-alpha/delta locus, TCR αδV-TCR αC accrued more than twice as many mutations (0.0055 S/N) as TCR αδV-TCR δC (0.0025 S/N; Figures 3-3, 3-4, 3-5B). As expected, WRCH/DGYW motifs within V segments used by TCR αC strongly correlate with those used by TCR δC (Pearson correlation, $r=0.94$; $p<0.001$). However, TCR αδV-TCR αC mutated more than other canonical TCR chains. Mutation in TCR αδV-TCR αC was biased towards C/G mutations

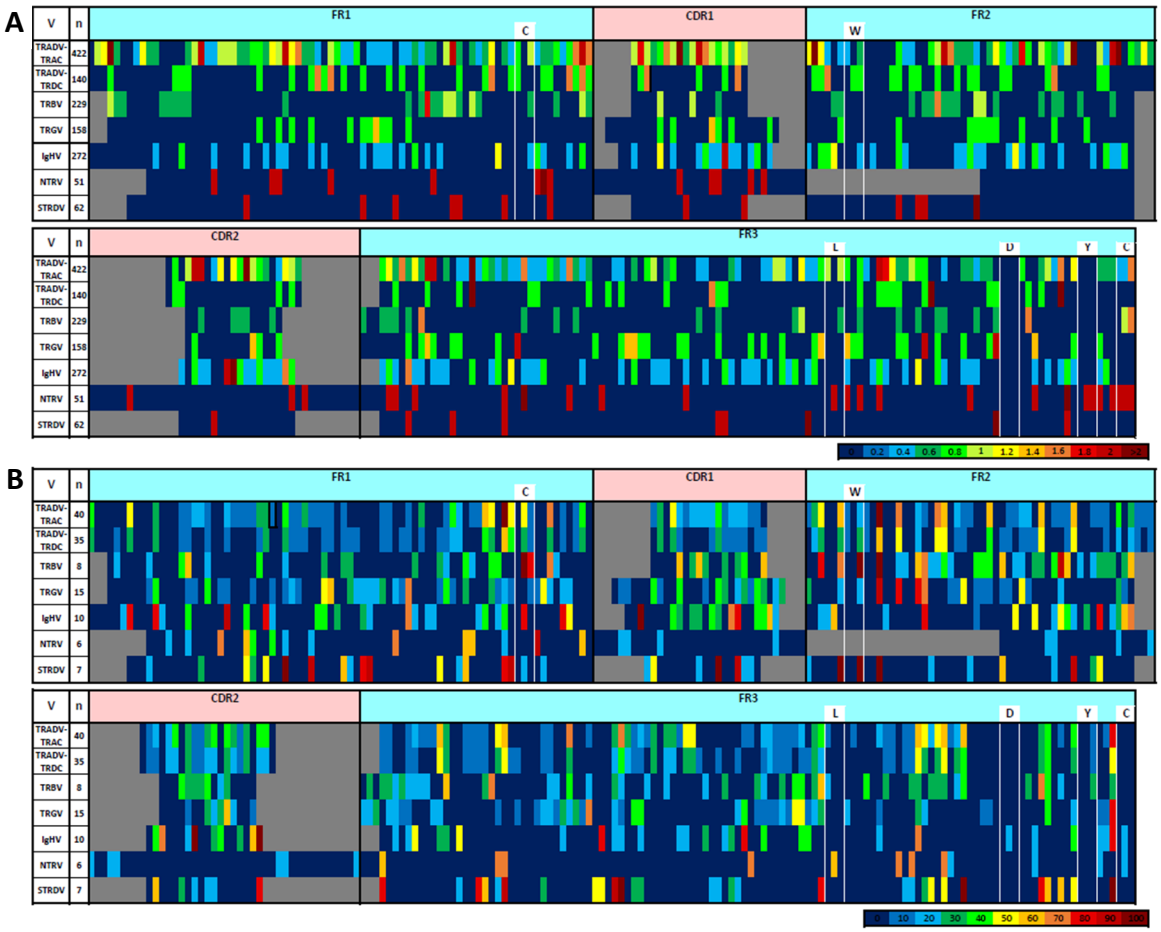

Figure 3-4. SHM targets complementarity-determining regions (CDR) of TCR-alpha associated with alpha constant regions. [A] TCR-alpha accumulates significantly more nonsynonymous (NSYN, solid) mutations than all other TCR chains except IgHV-delta sequences and significantly more synonymous (SYN, stippled) mutations than all other chains. [B] TCR-alpha/delta V gene segments associated with TCR-alpha constant (C) regions accumulate significantly more mutations in both framework (FR) and CDR than when associated with TCR-delta C, and in general, accumulate significantly more mutation than all other TCR chains except CDR2 domains of IgHV-delta [Student's one-way, unpaired T-test, $p<0.01$]. We counted the number of mutations within 422 TCR-alpha/delta V-TCR-alpha C (green), 137 TCR-alpha/delta V-TCR-delta C (blue), 158 TCR-gamma V (gold), 237 TCR-beta V (pink), 51 NARTCR V (red), 62 supporting NARTCR-delta V (NAR-STCR, black), and 275 IgHV-TCR-delta C (purple) sequences. [Student's one-way, unpaired T-test, *$p<0.05$; **$p<0.01$]

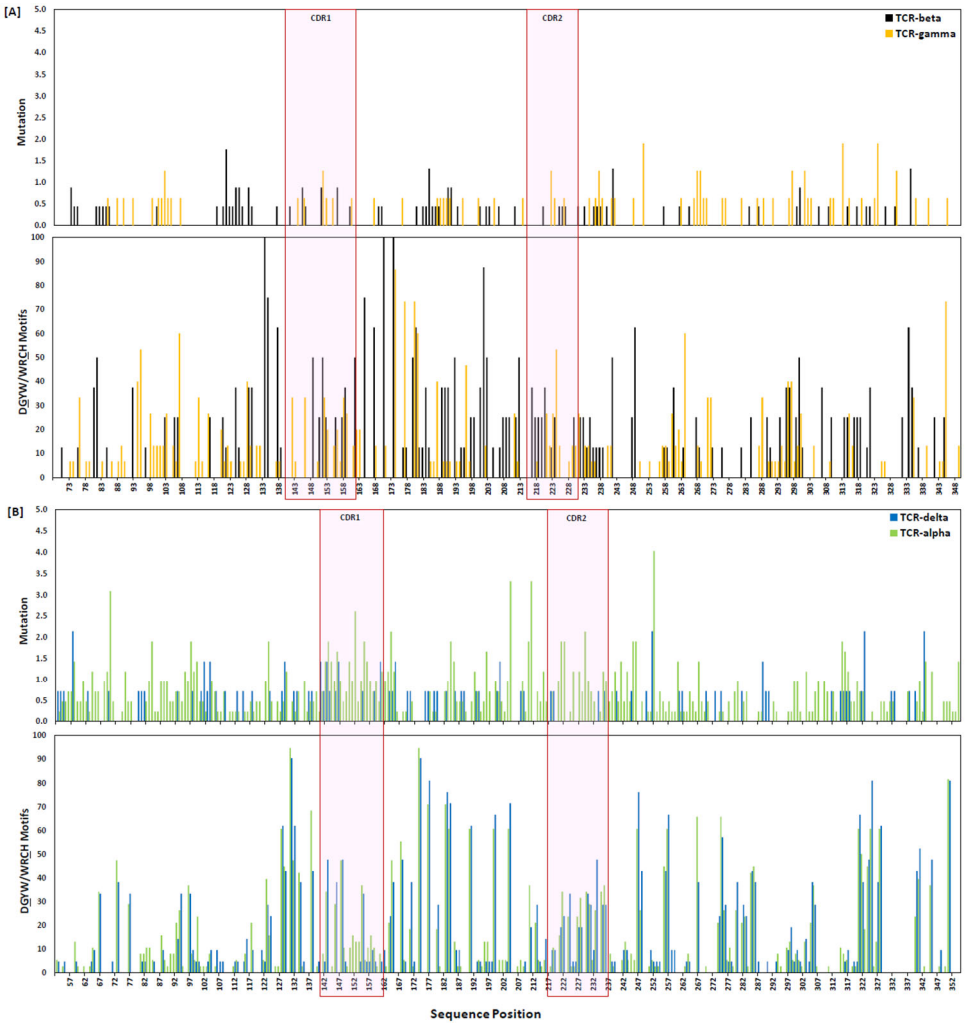

Figure 3-5. Despite the presence of AID-preferred hotspot motifs (DGYW/WRCH) in all V segments used by T cell receptors, only alpha/delta V (TCRαδ V) segments associated with alpha constant regions incorporate significant mutation. Proportion of mutations (both single and tandem) [A-D, top panels] and DGYW/WRCH motifs [A-D, bottom panels] observed at each position along the sequence length of [A] TCRβ V (black, 89 mutations; 228 sequences; 8 V gene segment groups) and TCRγ V (gold, 82 mutations; 158 sequences; 15 V gene segment groups), [B] TCRαδ V associated with TCRδ constant region (blue, 102 mutations; 140 sequences; 21 V gene segment groups) and TCRα constant region (green, 769 mutations; 422 sequences; 38 V gene segment groups), [C] NARTCR distal V domains (red, 37 mutations; 51 sequences; 6 V gene segment groups) and proximal supporting V domains (black, 25 mutations; 62 sequences; 7 V gene segment groups), and [D] IgHV associated with TCRδ constant region (purple, 100 mutations; 237 sequences; 14 V gene segment groups) and TCR-associated Ig-like V (TAILV, no mutations; 11 sequences; 2 V gene segment groups) associated with TCRδ or TCRα constant regions. Position indicates the forward (3' to 5') location of the mutable base within a Geneious-derived consensus sequence for each group. We counted mutations and motifs by sequence domain [framework region (FR) and complementarity-determining region (CDR)]. Red (and black) boxes indicate the locations of CDR [TCRβ CDR1: 145-162, CDR2: 214-231; TCRγ CDR1: 139-165, CDR2: 217-231; TCRαδ CDR1: 139-162, CDR2: 217-237; distal NARTCR V domain (red) CDR1: 130-152, CDR2: 184-226; supporting NARTCRδ V domain (black) CDR1: 133-147, CDR2: 202-219; IgHV CDR1: 139-162, CDR2: 217-237; TAILV CDR1: 133-147, CDR2: 202-219]. DGYW/WRCH: G:C is the mutable position; W = A/T, D = A/G/T, Y = C/T, R = A/G, and H = T/C/A.

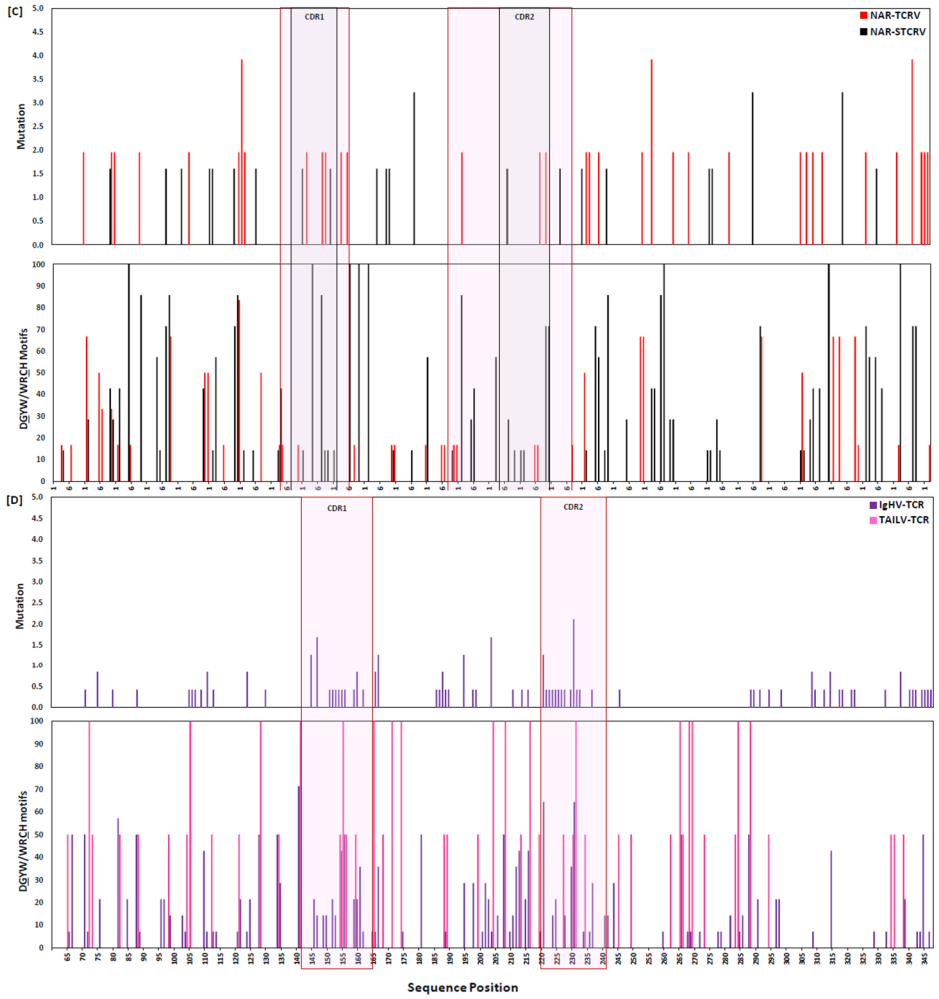

Figure 3-5. (*continued*)

Table 3-3. Target nucleotide mutation frequency in DGYW/WRCH mutation hotspots within framework regions (FR) and complementarity-determining regions (CDR) of T cell receptor (TCR) variable region (V) segments. [DGYW/WRCH (G/C is the mutable position; D=A/G/T, Y=C/T, W=A/T, R=A/G, and H=T/C/A); "ALL" refers to G and C nucleotides found within hotspot motifs along the entire V segment (FR and CDR); "Outside Motif" refers to G and C nucleotides outside a hotspot motif; S/N= substitutions per nucleotide; MI=mutability index*; %G/C= proportion of all nucleotides in that category that are G or C. TCRαδV-TCRαC=alpha; TCRαδV-TCRδC= delta; TCRβV=beta; TCRγV=gamma; IgHV-TCRδC= trans-rearrangements between immunoglobulin heavy chain (IgH) and delta TCR constant regions; NTCRV and STCRδV= components of NARTCR]

V segment	Hotspot Motif	Region	# G/C Nucleotides	% G/C	S/N	Observed G/C Mutations (#)	Expected G/C Mutations (#)	MI*	χ^2 p**
TCRαδV-TCRαC	DGYW/WRCH	FR	11008	1.9	1.27	140	5.9	23.62	0.0000
		CDR	2997	0.5	1.94	58	2.6	22.73	
		Inside	14005	2.4	1.41	198	8.3	23.96	0.0000
	Outside Motif		568931		0.03	146	335.7	0.43	
TCRαδV-TCRδC	DGYW/WRCH	FR	4012	4.2	0.65	26	2.3	11.15	0.0000
		CDR	714	0.7	0.84	6	0.5	11.15	
		Inside	4726	4.9	0.68	32	2.9	11.15	0.0000
	Outside Motif		90762		0.03	26	55.1	0.47	
TCRβV	DGYW/WRCH	FR	5476	9.8	0.29	16	5.5	2.92	0.0000
		CDR	925	1.7	0.32	3	0.7	4.28	
		Inside	6401	11.5	0.30	19	6.2	3.07	0.0000
	Outside Motif		49435		0.07	35	47.8	0.73	
TCRγV	DGYW/WRCH	FR	3099	5.6	0.42	13	3.4	3.78	0.0000
		CDR	581	1.0	0.69	4	0.4	11.27	
		Inside	3680	6.6	0.46	17	3.7	4.58	0.0000
	Outside Motif		51898		0.08	39	52.3	0.75	
IgHV-TCRδC	DGYW/WRCH	FR	4604	6.2	0.56	26	6.0	4.37	0.0000
		CDR	1995	2.7	1.10	22	3.5	6.33	
		Inside	6599	8.9	0.73	48	9.4	5.12	0.0000
	Outside Motif		67254		0.08	57	95.6	0.60	
NTCRV	DGYW/WRCH	FR	644	14.6	0.47	3	1.7	1.80	0.2240
		CDR	102	2.3	2.94	3	2.1	1.44	
		Inside	746	16.9	0.80	6	3.6	1.69	0.1542
	Outside Motif		3664		0.41	15	17.4	0.86	
STCRδV	DGYW/WRCH	FR	1019	11.3	0.20	2	1.4	1.40	0.5413
		CDR	279	3.1	0.00	0	0.1	0.00	
		Inside	1298	14.3	0.15	2	1.6	1.27	0.7161
	Outside Motif		7754		0.12	9	9.4	0.96	

*Mutability index is the observed number of mutations of a specific nucleotide divided by the expected number of mutations of that nucleotide, with a value of 1.00 indicating random mutation

**χ^2 analysis was used to compare observed and expected numbers of mutations between (a) FR and CDR regions and (b) all mutations inside and outside hotspot motifs for each V segment type.

Table 3-4. Mutation bias within T cell receptor variable (TCRV) segments differed by V segment type.
[FR: framework region; CDR: complementarity-determining region; TCRαδV-TCRαC=alpha; TCRαδV-TCRδC= delta; TCRβV=beta; TCRγV=gamma; IgHV-TCRδC= trans-rearrangements between immunoglobulin heavy chain (IgH) and delta TR constant regions; NTCRV and STCRδV= distal and proximal (supporting) components of NAR-TCR]

V segment	Region	% Transition Mutations	% Tranversion Mutations	% All mutations that are G or C nucleotides	% of G and C Mutations inside Motifs
TCRαδV-TCRαC	FR	45.58	43.54	59.41	53.44
	CDR	43.40	43.40	51.57	70.73
	All	45.00	43.50	57.33	57.56
TCRαδV-TCRδC	FR	53.33	34.67	62.67	55.32
	CDR	52.94	47.06	64.71	54.55
	All	53.26	36.96	63.04	55.17
TCRβV	FR	48.75	36.25	60.00	33.33
	CDR	44.44	44.44	66.67	50.00
	All	48.31	37.08	60.67	35.19
TCRγV	FR	56.00	34.67	65.33	26.53
	CDR	62.50	25.00	87.50	57.14
	All	56.63	33.73	67.47	30.36
IgHV-TCRδC	FR	50.72	37.68	49.28	38.24
	CDR	51.92	46.15	71.15	59.46
	All	51.05	40.00	55.26	45.71
NTCRV	FR	42.86	47.62	42.86	30.00
	CDR	62.50	25.00	37.50	27.27
	All	48.28	41.38	41.38	28.57
STCRδV	FR	40.91	54.55	45.45	20.00
	CDR	66.67	33.33	33.33	0.00
	All	44.00	52.00	44.00	18.18

(57%) and slightly biased towards transitions (45%) (Table 3-4). Most C/G mutations occurred within WRCH/DGYW motifs, especially in CDR (FR: 58%, CDR: 71%; 57% overall). TCR αδV-TCR δC mutation also was biased towards C and G nucleotides (63%) within WRCH/DGYW motifs (55%) and towards transition mutations despite having a much lower rate of mutation (53%; Figures 3-3, 3-4, 3-5B; Table 3-3). Replacement mutations occurred more often than silent mutation regardless of constant region utilized (alpha: R/S=2.2; delta: R/S=2.48). Generally, TCR αδV gene segments incurred more tandem base mutations than all other chains except perhaps IgHV.

Our previous data suggested that mutation to TCR-alpha may be higher in thymus than in peripheral lymphoid tissues (spleen, spiral valve, and blood), but the limited data set constrained our ability to find a significant difference between tissue types (Ott et al. 2018). Thus, we attempted to evaluate any difference in mutation between primary and secondary lymphoid tissues here. We compared frequency of mutation in clones originating from thymus tissue to those originating from peripheral lymphoid tissues. Unfortunately, even our larger dataset constrained analysis. We identified four groups of sequences with identical CDR3 that contained clones from both the thymus and the periphery. Unfortunately, we observed no mutation to any of these sequence groups, so we were unable to compare tissues directly. Using our entire dataset, we analyzed mutation separately for sequences derived from thymus and from peripheral lymphoid tissues. For TCR-alpha, TCR-alpha/delta, TCR-beta, and TCR-gamma, results suggested that peripheral lymphoid tissues have a higher frequency of mutation (Table Supplement

A-2). However, because we are unable to directly assess mutation in clones derived from common progenitors found in both thymus and peripheral tissues, we hesitate to definitively claim that additional mutation occurs after T cells arrive in the periphery. Further experiments are necessary to ascertain whether mutation frequencies of non-alpha Vs differ between thymus and periphery. Specifically, we need germline genomic Vs and a much deeper CDR3 family clone analysis, but unfortunately no nurse shark genome exists at this time.

3.2.5. Mutation only minimally affects NAR-TCR and IgHV-TCR δC rearrangements

We observed less mutation in both variable domains of NAR-TCR (NTCR V and STCR δV) compared to alpha chain V. This lack of mutation confirms earlier reports that NAR-TCR does not utilize SHM (Criscitiello et al. 2006). We also found that silent mutation occurred nearly as often as replacement mutation (Figures 3-3, 3-4, 3-5C). Interestingly, mutations in FR tended to be transversions (NTCR V: 47%; STCR δV: 54%) while those in CDR were biased towards transitions (NTCR V: 63%; STCR δV: 67%; Table 3-3). Rearrangements between IgHV and TCR constant regions (IgHV-TCR δC) contained more mutation in CDR2 than other regions, and mutation frequency was greater in IgHV-TCR δC than all other TCR V segment types except alpha (TCR αδV-TCR αC; Figure 3-3, 3-4, 3-5D). However, mutation to IgHV was substantially lower than would be expected during an antigen-specific response in B cells as they affinity mature in spleen (Dooley et al. 2006;

Lee et al. 2002; Zhang et al. 2013). Mutations within CDR were biased towards G and C nucleotides (71%) within WRCH/DGYW motifs (59%) and tended to be transitions (51%). However, G and C mutation within FR was much lower (49%) and only occurred in motifs 38% of the time (Table 3-3). Mutations to IgHV-TCR δC typically caused amino acid replacement at that position (R/S=3.4).

3.2.6. Summary

In our 5' RACE library, we found that SHM appears to target T cell receptors (TCR) differentially based generally on the V segment and specifically on the associated C region incorporated into a transcript. We observed limited mutation to all V segments. However, despite the presence of AID-preferred motifs in all V segments (including both conventional TCR chains and Ig or Ig-like chains), we observed significantly more mutation to TCR αδV associated αC regions than any other T cell type. Mutation occurred as both point and tandem base changes, typical of SHM seen in shark Ig and TCR before, and appeared in transcripts from both thymus and peripheral lymphoid tissues. The limited mutation observed in V segments associated with βC, γC, or δC regions or in TAIL V associated with αC or δC regions did not suggest selection for paratope changes (see Figure Supplement A-8). However, mutation patterns in TCR αδV-TCR αC were characteristic of AID-catalyzed changes to binding regions of receptors, lending further evidence that AID expression in nurse shark thymus alters paratopes as thymocytes

migrate through during development. The increased mutation in TCR αδV associated with TCR αC suggests that the associated C region is important for AID targeting (see Figure Supplement A-8).

3.3. Discussion

We previously published the novel use of somatic hypermutation (SHM) in TCR-alpha chain of αβ T cells within the thymus of nurse sharks (Ott et al. 2018). The frequency of mutation at α was similar to that seen in B cell receptor (BCR) loci in sharks and mammals. As in B cells, SHM in shark T cells appeared catalyzed by activation-induced cytidine deaminase (AID) and resulted in both point and tandem mutations that accumulated non-conservative amino acid replacements within antigen-binding regions (complementarity-determining regions, CDR) of receptors. However, unlike B cells that use SHM for affinity maturation after exposure to antigen, shark T cells instead use SHM for repertoire diversification during T cell development within the thymus, implying that SHM contributes to receptor modifications that enhance selection. Here we extend these findings with an analysis of SHM in other canonical TCR chains, including TCR-beta chain of αβ T cells and TCR-gamma and TCR-delta chains of γδ T cells, as well as in non-canonical TCR that associate immunoglobulin (Ig) or Ig-like variable (V) segments with TCR constant (C) regions.

We our previous report, we described an overall mutation frequency of 0.0225 substitutions per nucleotide (S/N) in TCR-alpha chain, with 66% of all mutations to G and

C nucleotides (Ott et al. 2018). Here we observed an overall mutation rate of 0.0055 S/N in TCR-alpha/delta V segments (TCR αδV) associated with TCR-alpha constant regions (TCR αC). While these frequencies are considerably lower than those we found before, they reflect the much larger data set used in the current study and likely are more representative of actual substitution rates within TCR-alpha chain. Again, we observed that mutation within TCR αδV-TCR αC affected C and G nucleotides 57% of the time, and 70% of these mutations occurred within complementarity-determining regions (CDR) and targeted WRCH/DGYW motifs. Mutation was 2.3 times more likely within CDR than framework regions (FR), similar to the 1.87 times we reported before, and mutation was biased towards transitions. Although some V segments accumulated mutation at higher rates than others (i.e., TCR αδV2.4 contained no mutation while TCR αδV5 mutated at a rate of 0.026 S/N), it is clear that AID-catalyzed SHM alters T cell receptor alpha chain in nurse shark.

We then assessed whether SHM is used by the other canonical T cell chains (beta, gamma, and delta) or by the non-canonical receptor chains that associate IgV segments with TCR C regions (e.g., IgHV-TCR δC rearrangements, doubly-rearranging NAR-TCR, and Ig-like TAIL V; see Figure 3-1). To fairly assess mutation patterns between TCR-alpha and TCR-delta (since they share a locus), we reclassified and renumbered all new and previously published alpha and delta V segments and identified the constant region associated with each V segment for all sequences within our dataset (see Figure 3-2C and

Figure Supplement A-4). We then analyzed mutation data separately for sequences associated with alpha constant region (TCR αδV-TCR αC) and those associated with TCR-delta constant region (TCR αδV-TCR δC). We found that SHM in nurse sharks significantly alters TCR αδV segments bound to TCR αC but not to TCR δC ($p<0.001$). Mutation was found twice as often in TCR-alpha (TCR αδV-TCR αC, 0.0051 S/N) than TCR-delta (TCR αδV-TCR δC, 0.0021 S/N) overall, and CDR of TCR-alpha (0.0098 S/N) accumulate more than three times as many mutations as TCR-delta CDR (0.0032 S/N). Further, 71% of all G/C mutations to TCR αδV-TCR αC were located within AID-preferred WRCH/DGYW hotspot motifs, compared to 52% in TCR αδV-TCR δC. Mutation in TCR-beta V (TCR βV) and TCR-gamma V (TCR γV) reflected similar patterns as TCR-delta, with relatively low mutation (TCR βV: 0.0014 S/N; TCR γV: 0.0021 S/N) that is not directed at hotspot motifs (TCR βV: 35%; TCR γV: 30%). The differential use of SHM by TCR-alpha compared to other TCR chains suggests a regulatory control mechanism modulates AID access to the locus during transcription. In mice and humans, V(D)J recombination in IgH V regions is controlled by regulatory elements upstream of the IgH constant (C) region exons (between the V and D clusters) (Guo et al. 2011; Rouaud et al. 2013). However, AID accessibility to the IgHC locus is controlled by transcriptional enhancers located in the 3' regulatory region (RR) downstream of the IgHC exons (Dunnick et al. 2009; Komori et al. 2006; Rouaud et al. 2013), and SHM requires both the transcription of IgH V regions and the upregulation of AID expression (Komori et al. 2006). It is likely that these same

transcriptional enhancers regulate SHM of nurse shark TCR, allowing AID access to TCRα chain but limiting access in other TCR chains, evidenced by the greater use of SHM by TCR αδV segments associated with TCR-alpha than with TCR-delta constant regions. Since SHM appears to occur in thymus tissue, this differential use of SHM also suggests that AID may be targeting hotspot motifs of TCR-alpha chain to salvage receptors during selection. Roughly half of all beta, gamma, and delta transcripts in our sequence library originated from thymus tissue, and we observed mutation within transcripts from thymus in all TCR chains. Thus, is seems likely that mutation to TCR-delta, TCR-beta, and TCR-gamma chain occurs inadvertently as T cells migrate through the thymic cortex during development.

Mutation was low in both V domains – NAR-TCR V and supporting TCR δV – of doubly-rearranging NAR-TCR (NTCR V: 0.0023 S/N and STCR δV: 0.0017 S/N). We observed more replacement mutation in NTCR V FR, but replacements did not occur more than silent mutation in either domain. Mutation also did not differ between FR and CDR and did not appear to target G and C nucleotides (NTCR V: 41%; STCR δV: 44%) or WRCH/DGYW motifs (NTCR V: 29%; STCR δV: 18%). Mutation in NTCR V and STCR δV mirrors patterns observed in IgNAR transmembrane (Tm) transcripts (0.007 S/N), where mutation has been shown to not target CDR or display bias towards replacements. Though secreted IgNAR forms use SHM for antigen-driven immune responses following antigen stimulation, Tm IgNAR is not used to generate the primary IgNAR repertoire (Diaz et al. 1998; Dooley et al. 2006). Diaz et al. suggested the reduced mutation within Tm

IgNAR may result from down-regulating mutation mechanisms to avoid the risk of creating non-functional receptors after four gene rearrangement events (Diaz et al. 1998). Unusually, mutation in NTCR V and STCR δV favored transversions within FR (NTCR V: 48%; STCR δV: 55%) but transitions within CDR (NTCR V: 63%; STCR δV: 67%). While it is possible that NAR-TCR mutations result unintentionally during thymocyte migration as above, the disparate bias towards transitions and transversions in both domains suggests an alternate or additional process besides AID-catalyzed SHM (Krijger et al. 2013; Thientosapol et al. 2018). Our observations of more limited mutation in TCR γV4 also may indicate that NAR-TCR preferentially pairs with TCR γV4, though no data currently exists to confirm this prediction.

Mutation to V segments of IgHV-TCR δC rearrangements significantly altered nucleotides of CDR2 ($p<0.0001$) but not of other regions. Similar to IgHV in B cells, mutation in CDR appeared targeted to G and C nucleotides (71%) within DGWY/WRCH hotspot motifs (59%) and were biased towards transitions (52%). IgHV segments used by T cell receptors are identical to germline IgHV segments used by B cells, so the similarity in mutation patterns makes sense. Although mutation to IgHV is substantially reduced when associated with a T cell receptor constant region versus a B cell one (Lee et al. 2002), the significantly greater mutation in CDR2 that is targeted to AID-favored G and C nucleotides suggests that SHM is altering IgHV segments in T cells as well as B cells. The antigen ligands that IgHV-TCR δC receptors bind is not known, nor is the developmental

pathway of these cells within the thymus. Nurse shark T cells may simply be using these additional V segments to improve thymic diversity by increasing the pool of Vs available during recombination (Ott et al. 2018). Successful rearrangements between an IgHV pseudogene and T cell receptor constant regions support this idea. The motif-rich V domains may be inadvertently mutated as the cells travel through thymic cortical areas where AID is being expressed. However, it also is plausible that these IgHV-T cells are actively modified for simple diversification for free antigen or to salvage receptors during selection that are unable to receive adequate survival signals, though it is unclear whether or not selection is required by this cell type to exit the thymus. Additional studies that assess the anatomical location of developing double-negative thymocytes in addition to functional studies of receptors are necessary to discern further why these V domains are altered.

Recent studies in non-mouse/human organisms confirm the versatility of the TCR-alpha/delta locus in the vertebrate immune system. Besides IgHV-TCR δC, TAIL V and NAR-TCR rearrangements in shark, TCRμ of monotremes and marsupials is a hybrid receptor derived from Ig and TCR components (Parra et al. 2008). Isoform TCRμ2.0 requires the rearrangement (and in platypus, recombination) of two V domains and is structurally analogous to the doubly-rearranging NAR-TCR in nurse sharks (Criscitiello et al. 2006; Parra et al. 2007). While the C regions of TCRμ resemble traditional TCR δC, the V regions are more similar to IgH V genes, and the TCRμ locus itself reflects the tandem

cluster arrangement of Ig rather than the translocon arrangement of TCR (Wang et al. 2011). In fact, antibody-like TCR δV segments (VHδ) occur in nearly all gnathostome groups: elasmobranchs (nurse sharks), bony fish (coelacanth), amphibians (*Xenopus*), birds (chickens), and mammals (platypus, Florida manatee) (Breaux et al. 2018; Criscitiello et al. 2010; Deiss et al. 2019; Parra et al. 2008; Parra et al. 2012b; Parra et al. 2010; Saha et al. 2014). The passerine zebra finch contains VHδ within a conventional TCR-alpha/delta locus, while Galliform birds house VHδ gene segments within a second non-syntenic TCR-delta locus (Parra et al. 2012b). Thus, the sharing of antigen receptor gene segments seems to be a primordial condition, where the basic immune system started as a "big soup" of receptor parts that, in some lineages, became more and more divergent (and arguably, constrained) as vertebrates moved to land and evolved warm-blooded systems.

Other examples exist to suggest that not only are Ig and TCR gene components interchangeable, but mechanisms of receptor diversification also are shared between B and T cells. In camelids, both gamma and delta V gene segments employ SHM to improve structural stability of receptors rather than antigen selection (Ciccarese et al. 2014). Analyses included only sequences from peripheral lymphoid tissues, but mutation patterns parallel those of TCR γV and TCR αδV-TCR δC in our study. While overall mutation was fairly low (and did not favor G/C nucleotides), replacement mutation was biased towards AID hotspot motifs and resulted in greater length and diversity of CDR3

(Antonacci et al. 2011; Ciccarese et al. 2014). Crystalline structures of γδ TCR bound to non-classical MHC demonstrate that MHC recognition occurs through direct contact by CDR3 of TCR-delta (Adams et al. 2005), suggesting that cells with longer or more diverse CDR3 may be selected for survival. We have evidence that IgHV-TCR δC rearrangements also create longer CDR3 by incorporating one or two diversifying (D) segments from Ig in addition to the single D and J segments from TCR-delta (IgV-IgD1-(IgD2)-TCR δD-TCR δJ) during recombination (data not shown). While MHC presentation is not obligatory for γδ T cells, receptors may require stimulation by cell to cell contact (Allison and Garboczi 2002). Thus, SHM may provide a tool to fine-tune receptors to recognize particular antigens, and since both gamma and delta chain are preloaded with numerous hotspot motifs, opportunistic mutation along the V segment is likely.

As we learn more about the immune systems of non-model vertebrates, it becomes clearer that ancient lymphocytes likely did not follow the unambiguous rules of B cell and T cell biology found in modern textbooks, and we only just are beginning to understand the myriad schema different vertebrate groups evolved to diversify Ig and TCR repertoires. For example, lymphocyte rearrangement and diversification mechanisms predate the primordial “big bang” of IgSF-based adaptive immunity, as AID-like APOBEC mutators (i.e., CDA1 and CDA2) exist in agnathan vertebrates (hagfish and lamprey) to diversify the variable lymphocyte receptors (VLR) of this more ancient lymphocyte adaptive antigen receptor system (Alder et al. 2008; Guo et al. 2009). The data presented

here offer additional clues to the possible evolutionary relationship between the immune systems of agnathan and jawed vertebrates, suggesting a sustained bipartite use of APOBEC family enzymes to diversify humoral and cellular antigen receptor repertoires, with CDA acting upon VLR and AID upon Ig and TCR loci.

3.4. Materials and methods

3.4.1. Study animals

TCR sequence data used in this study came from six nurse sharks (*Ginglymostoma cirratum*) collected off the Florida Keys. We harvested peripheral blood and immune tissues (thymus, spleen, spiral valve) after MS-222 (Argent, Redmond, WA) overdose from five sharks ("Joannie", "Mary Junior", "White", "Grumpy", "Tom Thumb") at the University of Maryland's Center of Marine Biotechnology and one shark ("Florence Nightingale") at Texas A&M University's College of Veterinary Medicine. We immediately purified RNA with TRIzol reagent (Life Technologies, Carlsbad CA). We conducted all research under the Florida Fish and Wildlife Commission Special Activity License SAL-18-2013-SR.

Additionally, we incorporated T cell sequences from published datasets (Criscitiello et al. 2006; Criscitiello et al. 2010; Deiss et al. 2019; Lee et al. 2008; Malecek et al. 2005; Malecek et al. 2008; Rumfelt et al. 2004) and transcripts from an unpublished multi-tissue Illumina transcriptome to improve analysis of V segment alignments. We reference published sequences by their GenBank accession number in all relevant figures.

3.4.2. 5' RACE library generation, cloning, and Sanger sequencing

We used 5ug total RNA to generate a 5' RACE (Rapid Amplification of cDNA Ends) library using the GeneRacer Kit (Life Technologies) and a 50:50 mix of oligo-dT and random hexamer primers for cDNA synthesis (Superscript III First Strand Synthesis System, Thermo Fisher Scientific, Inc., Waltham, MA, USA). We estimated cDNA concentration using a Nanodrop 2000 Spectrophotometer (Thermo Fisher Scientific, Inc.).

We used the GeneRacer 5' forward primer (Life Technologies) and reverse primers designed to T cell receptor beta (TCRβ), gamma (TCRγ), or delta (TCR-delta) constant (C) regions to amplify RACE products for Sanger sequencing (Table Supplement A-3). We followed protocols outlined in (Ott et al. 2018) for PCR conditions, product visualization, and cloning. We submitted either plasmids or purified PCR products for Sanger sequencing to the DNA Technologies Core Lab on the Texas A&M University campus (College Station, TX, USA) or to GENEWIZ (South Plainfield, NJ, USA). We deposited annotated sequences into the National Center for Biotechnology Information's (NCBI) GenBank sequence database with the following accession numbers: MN748005 - MN748891 and MN788155 - MN788287.

3.4.3. Sequence alignment and analysis

We used Geneious (v9.1.8, Biomatters Inc., Auckland, NZ) bioinformatics software to manage DNA sequence data following the same methods as Ott et al. (2018). Sequence alignments and region annotations followed IMGT guidelines (Lefranc et al. 2003). We predicted the location of signal peptide cleavage sites between leader and V segment sequences using SignalP (v5.0) (Almagro Armenteros et al. 2019). A V segment included all bases from the 1st predicted nucleotide of the V segment to the conserved cysteine (C) residue (YxC motif) at position 104 ("V only"). A V Region included all bases between the 1st predicted nucleotide of the V segment to the last predicted nucleotide of the J segment (V and J). The CDR3 included all bases after the conserved cysteine of the V segment and before the conserved phenylalanine (F) residue (FGxG motif) of the J segment.

We grouped sequences into unique V families based on 70% nucleotide sequence identity and further refined groups into subfamilies based on 80% nucleotide identity (Brodeur and Riblet 1984; Rumfelt et al. 2004). For beta and gamma Vs, we revised the V-segment numbering scheme used in Criscitiello et al. (2010) to reflect these parameters. However, we integrated alpha and delta V segments into a single group of "αδ Vs" (TCR αδV) to more clearly characterize the locus and avoid confusing name replication within the data. We then numbered the segments according to their position on a phylogenetic tree, from the most divergent branch (TCR αδV1) to the most recent branch (TCR αδV11).

For NAR-TCR V segments, we reassigned numbers using this same strategy and renamed the supporting TCR δV segments "STCR δV" to distinguish them from the canonical TCR αδV segments in alignments. We used the same locus-informed numbering system for IgH V identity (Lee et al. 2008) for all IgHV-TCR δC and IgHV-TCR αC rearrangements. We retained constant region identity (TCR αC/TCR δC) for use in mutation analyses. This system allowed us to clarify V segment usage and analyze differences in mutation patterns for the entire locus. Sequence names followed IMGT unique numbering standards for T cell receptors (Lefranc 2014; Lefranc et al. 2003; Ohlin et al. 2019). Sequence clone names signify the individual shark [letter following V segment (e.g., TCR γV1X) J=Joannie, M=Mary Junior, G=Grumpy, F=Florence Nightingale, T=Tom Thumb, or W=White) and the tissue type [letter preceding clone number (e.g., T19): T=thymus, S=spleen, B=peripheral blood leukocytes, and V=spiral valve (intestine)] from which the clone came. For previously published sequences, we use the accession numbers as clone names.

Our preliminary dataset contained 761 TCR αδV, 229 TCR βV, 158 TCR γV, 195 IgMV, 77 IgWV, 11 TAIL V, and 113 NAR-TCR V sequences. For individual Vs within a group (subfamily or allele), we generated a consensus sequence by evaluating sequences found in multiple tissue types and/or sharks, assuming if the exact V nucleotide sequence appeared in more than one tissue or individual it did not result from the same clone. We equated this consensus sequence to the germline sequence. For all "trans"-

rearrangements, we compared mutation in IgHV segments to published germline IgHV sequences (Lee et al. 2008). We considered only mutations to IgHV segments associated with TCR δC (not resulting from affinity maturation of B cells). Finally, we compared the incidence of AID hotspot motifs (WRCH/DGYW) between V segments that exhibited mutation and those that lacked mutation to determine the likelihood that AID might catalyze these changes.

3.4.4. Mutation analysis

We analyzed mutation data in beta, gamma, and alpha/delta variable (V) gene segments using methods described in Ott et al. (2018) for alpha chain. Mutation frequency was the number of nucleotide changes divided by the total number of nucleotides within a region (e.g., FR, CDR, J, and C) based on differences to a consensus sequence and recorded disagreements to the consensus as synonymous (S, amino acid unaltered) or non-synonymous (N, amino acid altered). Although insertions and deletions (indels) likely result from similar DNA repair mechanisms as single and tandem mutations, we did not include indels in our mutation data and removed them from nucleotide alignments. However, we indicated the location of each indel in nucleotide alignments by highlighting the nucleotides to either side of the indel (Suppl. Figs. 1-4). Further, individual sequences that shared less than 70% identity to any group were excluded from mutational analyses. We indicate these sequences using an asterisk next to the clone

name within the alignment. We used one-tailed, unpaired Student's t-test to compare mutation rates between alpha V (TCR αδV-TCR αC) and other V segment types.

To assess whether mutation was AID mediated, we examined V segment consensus sequences for the AID-favored ProSite motifs WRCH/DGYW (G:C mutable target) (International Union of Pure and Applied Chemistry, Zürich, Switzerland; see Sigrist et al. (2010)). These motifs serve as common "hotspots" for SHM, where AID favors the G/C bases within WRCH/DGYW motifs (Chen et al. 2012; Rogozin and Diaz 2004; Wei et al. 2015). We counted motifs present in consensus sequences (rather than those created by mutation) and compared mutations within clone sequences to target nucleotides within the motif. We defined the frequency of hotspot mutation as the number of mutations to target nucleotides within hotspots for a region (FR or CDR) divided by the total number of mutations in that region and compared FR and CDR regions using χ^2 analysis.

We calculated a mutability index for each nucleotide as the observed number of mutations of a specific nucleotide divided by the expected number of mutations of that nucleotide, with a value of 1.00 indicating random mutation (Chen et al. 2012). We derived the expected number of mutations by multiplying the frequency of a particular nucleotide within a family of sequences by the total number of observed mutations within that family. We used χ^2 analysis to compare mutability indices between FR and CDR regions.

4. CONCLUSIONS

To be successful, the vertebrate adaptive immune system must protect its host from a multitude of potential pathogens over a lifetime – defending against existing onslaughts as well as anticipating future encounters with infectious agents. Pivotal to this function is the creation of a diverse repertoire of lymphocyte receptors. This vast repertoire cannot be encoded directly by the genome. Rather, lymphocyte receptor loci are comprised of numerous gene segments that are somatically assembled into complete genes (Tonegawa 1983). Two strategies evolved in vertebrates: the variable lymphocyte receptor-based system of jawless vertebrates and the immunoglobulin superfamily-based system of jawed vertebrates. Both systems evolved a tripartite adaptive defense strategy with three distinct somatically assembled receptor types. The three lineages of agnathan variable lymphocyte receptors (VLR A, B, and C) are analogous to the B cell and T cell lineages of gnathostomes (Das et al. 2015). Like B cells, VLR Type B (VLRB) can be membrane-bound or secreted and functions in adaptive humoral responses (Alder et al. 2005; Pancer et al. 2005). Both VLR Type A (VLRA), transcriptionally more similar to αβ T cells, and VLR Type C (VLRC) transcriptionally similar to γδ T cells, occur only as a membrane-bound receptors and are predicted to function (as do T cells) in cell-mediated immune responses (Alder et al. 2005; Kasamatsu et al. 2010). However, unlike αβ T cells, neither VLRA nor VLRC seem to require antigen presentation for recognition (Deng et al. 2010b). The similarities in immune defense strategies between agnathan and gnathostome vertebrates suggests there were three lymphocyte lineages present in the

vertebrate common ancestor, with discernable components of the immune system labor partitioned among them (Flajnik 2014). Further insight into these similarities and differences in defense strategies could help elucidate the origins of lymphocyte receptors, making the study of “lower” fish immune systems ideal for comparative studies of immune evolution.

In most jawed vertebrates (including sharks), primary diversification of lymphocyte receptors occurs during RAG-mediated V(D)J recombination in bone marrow or analogous primary tissue (i.e., epigonal or Leydig organ in sharks, B cells) or thymus (T cells) tissues. Activation of B cells by antigen stimulates BCR genes to further diversify through AID-mediated SHM and CSR events in secondary lymphoid tissues. Together, these mechanisms generate a diverse and highly effective adaptive immune system. However, given the variety of intrinsic (e.g., physiological, genetic) and extrinsic (e.g., environment, pathogen type) constraints imposed on animals, it is not surprising that some groups evolved unique strategies to enhance or replace these fundamental mechanisms.

The ability to generate a diverse gnathostome antigen receptor repertoire depends, in part, on the number of separate V, (D), and J gene segments available during assembly (generated through combinatorial diversity). In chickens, rabbits, sheep, and cattle, combinatorial diversity is limited by a restricted availability or a preferential use of V gene segments during V(D)J recombination, and these animals must improve upon this

limited diversity through additional post-rearrangement mechanisms. These processes diversify the primary repertoire within secondary gut-associated lymphoid tissues (GALT) rather than in conventional bone marrow. One such mechanism is AID-catalyzed IGC, borrowing sequence from upstream (typically pseudogenized) V gene segments to augment the static combinatorial diversity (Arakawa et al. 2002; Becker and Knight 1990; McCormack and Thompson 1990; Parng et al. 1996; Reynaud et al. 1987; Reynaud et al. 1989; Thompson and Neiman 1987). Embryonic chickens employ IGC to diversify rearranged genes within the bursa of Fabricius, a sac-like lymphoid organ in the hindgut of birds (Glick et al. 1956; Huang and Dreyer 1978). Following IGC events in the appendix, rabbits then incorporate SHM-induced changes to V regions to further diversify the primary repertoire (Weinstein et al. 1994; Winstead et al. 1999). Sheep also generate highly restricted primary repertoires within ileal Peyer's patches (dense lymphoid follicles within the ileum of the small intestine), but use SHM rather than IGC to increase diversity of rearranged BCR (Reynaud et al. 1991a; Reynolds and Morris 1983). Thus, at least some vertebrates combine RAG-mediated V(D)J recombination and AID-catalyzed secondary diversification processes (IGC and SHM) to generate their primary antibody repertoire.

Based on research presented here, it is clear that nurse sharks also incorporate secondary diversification mechanisms during primary TCR repertoire development. In Chapter 2, we demonstrate that nurse sharks express AID in inner cortex, cortico-medullary junction, and inner medullary regions of thymus tissue, specifically coinciding

with rearrangement and assembly of TCR alpha chain and thymic selection. Analysis of TCR clones containing the same CDR3 [V(D)J junction] revealed patterns of mutation that mirrored those of affinity-matured BCR in germinal center follicles of mice and humans. Evidence of SHM in TCR alpha clones from thymus tissue suggests that T cells employ AID during thymic selection to alter the TCR alpha chain variable domain. While chickens, rabbits, and sheep evolved strategies to expand upon restricted primary RAG-generated B cell repertoires, it does not appear that nurse sharks generate T cell repertoires from a similar dearth of choices. In fact, the TCR alpha chain is built from a substantial pool of V gene segments. In Chapter 3, we report sequences encoding apparently functional receptors utilizing at least 37 different V gene segments from four putative V families. AID does not appear to edit sequences encoding other canonical or non-canonical receptor chain variable domains in the same way that it targets alpha.

We propose that AID alters the TCR alpha chain as receptors audition for the primary repertoire, salvaging receptors that would otherwise fail selection because they bind self antigen or are incompatible with self-MHC. Further, mutation to other chains occurs inadvertently as T cells migrate through the thymus during development.

4.1. Nurse sharks diversify primary T cell repertoires in unconventional ways

In addition to the canonical αβ and γδ TCR, nurse shark T cells recombine immunoglobulin (Ig) heavy chain (H) and Ig-like V gene segments into non-canonical receptors that improve overall repertoire diversity. IgH-TCR hybrids marry the V (and

often D) gene segments of conventional BCR with TCR delta (or rarely alpha) constant regions (C) to create unique, chimeric receptors (Criscitiello et al. 2010). Using our 5' RACE cDNA libraries, we determined that TCR utilize at least five IgM and three IgW V segment groups, and expression of chimeric IgH-TCRδ chains is comparable to or even exceeds expression of canonical TCRδ chains [see also Deiss et al. (2019)]. Though phylogenetically more like IgH V, TCR-associated Ig-like V (TAILV) gene segments are located within a translocon stretch of TCRαδ V and are expressed exclusively with TCR delta C (primarily) or alpha C (Deiss et al. 2019). Both IgHV and TAILV gene segments recombine with TCRδ (or TCRα) to produce unique TCR chains, expanding the combinatory potential of developing receptors. Nurse sharks also generate a unique TCR chain, the doubly-rearranging NAR-TCR, that adjoins both Ig and TCR V domains atop a TCRδ constant region (Criscitiello et al. 2006). The TCRαδ locus contains at least three NAR-TCR blocks, each consisting of Ig-like VDJ gene segments upstream of TCRδ VDJ gene segments, both of which rearrange to form the two V domains of NARTCR (Criscitiello et al. 2006; Deiss et al. 2019). The membrane-proximal TCRδ V domain supports a membrane-distal Ig-like V domain [phylogenetically similar to new or nurse-shark antigen receptor (NAR) V domains of IgNAR] that is predicted to bind antigen (Criscitiello et al. 2006). Only cartilaginous fish appear to produce these three unique receptor types (IgH-TCR, TAILV-TCR, and NAR-TCR), though convergent or orthologous types are found in other vertebrate classes (e.g., VHδ, TCRμ).

In Chapter 2, we provide evidence that nurse sharks also enrich their TCR repertoire by exploiting SHM during repertoire generation in the thymus. Using real-time RT-qPCR and *in situ* hybridization expression data from nurse shark thymus, we confirmed AID expression in thymus at levels roughly half those observed in spleen (where B cell SHM occurs). Using probes specific to either TCR αC or AID on thymus tissue, we observed a consistent "ring" pattern, where cells expressing both TCR αC and AID message surrounded a central cell expressing only TCR αC. Further, we determined that AID expression is localized to the inner cortex and medulla adjacent to the cortico-medullary junction, coincident with the location of TCRα rearrangement and thymic selection in mice (Huesmann et al. 1991). Thus, T cells appear to actively express AID during RAG-mediated somatic recombination of the alpha locus, permitting mutation to TCRα chain sequences while cells are being selected in the thymus.

We then assessed TCR transcripts for evidence of mutation and analyzed mutation patterns for similarities to AID-catalyzed mutation in affinity-matured BCR. Variable region sequences of BCR evolved to maximize the impacts of mutation, targeting replacement mutation to antigen-binding CDR and limiting mutation to structurally important framework regions (FR) (Saini and Hershberg 2015). Somatic mutation in mouse, human, and shark BCR is biased towards G:A and C:T transitions and targeted to AID-preferred nucleotide motifs (DGYW/WRCH) (Anderson et al. 1995; Lee et al. 2002; Li et al. 2004; Odegard and Schatz 2006; Rumfelt et al. 2002; Zhu and Hsu 2010). In addition

to SHM-induced point mutations observed in other vertebrates, nurse shark IgH, IgL, and IgNAR sequences generate tandem substitutions of 2-5 adjacent nucleotides (Dooley et al. 2006; Greenberg et al. 1995; Lee et al. 2002; Malecek et al. 2005). Though tandem mutations demonstrate a bias towards AID hotspot motifs, they do not typically favor transitions, suggesting that an additional mechanism may contribute to V region changes in nurse sharks (Zhu and Hsu 2010). Despite decades of assertions that SHM does not shape MHC-restricted αβ TCR repertoires, we identified mutation to nurse shark TCRα transcripts characteristic of AID-catalyzed SHM in shark BCR – point and tandem mutations focused on CDR, biased towards transitions, and targeted to AID motifs. Further, we detected mutation in transcripts from both thymus and peripheral lymphoid tissues, suggesting mutated receptors originated in the thymus prior to contact with foreign antigen. Together with corresponding evidence that AID expression overlaps TCRα chain rearrangement and selection in thymus, these data indicate that AID catalyzes SHM of TCRα for repertoire diversification during T cell development, implying that SHM contributes to receptor modifications that enhance selection.

Our discovery of AID-mediated somatic mutation in TCRα during primary lymphocyte development in thymus compelled us to examine the extent to which SHM alters the primary repertoire of other canonical (β, γ, and δ) and non-canonical (Ig or Ig-like) TCR chains. Studies in sandbar shark and dromedary camel reported mutation to TCR γ and δ chains that mirrored mutation in B cells, and study authors hypothesized that T

cells employ SHM as a means to generate a more diverse receptor repertoire (Chen et al. 2012; Chen et al. 2009; Ciccarese et al. 2014; Vaccarelli et al. 2012). However, these studies limited analyses to transcripts from peripheral lymphoid tissues, hampering thorough evaluation of the possible evolutionary mechanisms driving mutation. Based on our results from Chapter 2, we predicted that mutation in thymus would be confined to TCRα, since testing and selection of TCRβ chain is complete by the time TCRα rearrangement begins, and γδ T cells likely do not undergo thymic selection at all. Limited mutation to other canonical TCR chains could result inadvertently as cells migrate through the thymus during development. However, it is plausible that Ig or Ig-like V regions could respond to expressed AID message in thymus, since IgH, IgL, and IgNAR sequences evolved to maximize mutation in response to AID.

In Chapter 3, we examined transcripts from 5' RACE cDNA libraries from nurse shark thymus to analyze mutation patterns in other TCR chains. Our results indicate that SHM targets TCR sequences preferentially based (generally) on the V segment used and (specifically) the C region associated with it. Despite the varying presence of AID hotspot motifs within V gene segments of all canonical and non-canonical TCR chains, only TCRα V accumulated significant mutation. Though TCR β, γ, and δ chains exhibited limited mutation, patterns paralleled those observed in BCR and TCRα of nurse sharks, with point (and tandem) mutation biased towards transitions and focused on AID hotspot motifs within CDR. In TAILV and both V domains of NAR-TCR V, the infrequent mutation we

observed likely reflected the limited number of AID hotspot motifs present in sequences from these chains. Thus, AID-catalyzed mutation does not affect V segments of all chains equally. Comparing mutation between genomic V gene segments used with both alpha and delta C regions, when an alpha/delta V segment is associated with TCR αC it acquired more than twice as many mutations as when it was associated with TCR δC regions, suggesting that, in thymus, AID displays a proclivity for mutating V regions of the TCRα chain. Even IgHV gene segments, laden with abundant AID-preferred motifs, accrued substantially lower rates of mutation than TCRα V regions associated with TCRα C regions. Further, mutation was considerably lower in IgHV associated with TCR in thymus than one would expect of the same IgHV associated with a BCR undergoing affinity maturation in spleen. The increased mutation in V regions associated with TCR αC in thymus suggests that the C region may be particularly important for AID targeting.

4.2. Importance of studying immune mechanisms in non-traditional animal models

The basic components of adaptive immunity (RAG-mediated recombination of V, D, and J gene segments, B and T cell receptors, MHC class I and II, and AID-mediated somatic diversification mechanisms) are similar among extant gnathostome groups, owing to the fairly recent divergence of gnathostomes from their jawless ancestors roughly 500 Mya (Brazeau and Friedman 2015; Hsu 2009; Janvier 2011). Since this divergence, host immune systems evolved rapidly as a consequence of rapidly evolving pathogens, and the selective pressures caused by the ensuing genetic “arms race”

permitted new, innovative features to supplant existing ones (Bailey et al. 2013). Thus, gnathostomes evolved various accessory immune components as solutions to specific selective pressures of their environments (e.g., heavy-chain only antibodies of camels and sharks), and these accessory features can provide alternate views of the adaptive immune system through the window of evolution.

Much of our current understanding of vertebrate adaptive immunity comes from research in mice, despite a plethora of studies suggesting the ancestral adaptive immune system differed considerably from the "refined" features of the mouse and human. On its surface, this approach makes sense. On an evolutionary scale, rodents only recently diverged from primates (87 Mya), and because they evolved in and adapted to similar environments, mice and humans share broadly similar genetic and physiological traits (Bailey et al. 2013). Importantly, mice breed quickly in laboratory conditions, a result of their shorter generation time, and mouse genetics are more easily manipulated, evidenced by the variety of transgenic, knock-out, and knock-in mice available, making studies in mice easier than in their human counterparts (Morse and Fox 2006; Perlman 2016; Spencer 2002). Consequently, the mouse model became the experimental tool of choice in studies of human immunology, and mouse models of human disease have generated many advances in human medicine (Morse and Fox 2006). However, despite their phylogenetic similarities, differing selective pressures on humans and mice over the past 87 million years drove the evolution of distinct genetic solutions to immunogenetic

problems in each group (Bailey et al. 2013). Most notably, there are drastic differences in body size and lifespan, exposing mice and humans to different immunogens (a mouse nose is only an inch from the ground) for different periods of time [reviewed in Mestas and Hughes (2004)]. Thus, genes (or phenotypes) linked to disease may differ between the groups, leading to conflicting treatment outcomes, suggesting mice may not be the best model for studying human disease.

Phylogenetically, sharks and other chondrichthyans are the earliest members of the group that evolved a RAG-mediated Ig superfamily (IgSF)-based adaptive immune system (Flajnik and Rumfelt 2000). They also are positioned near the nexus between two vertebrate adaptive immune systems (VLRs of agnathans and IgSF of gnathostomes), and thus, can teach us a great deal about the origin of adaptive immunity. While studying a modern shark cannot tell us directly what adaptive immunity looked like at its genesis, it can shed light on its origins and help distinguish the fundamental aspects of immunity from the more accessory (or derived) elements within particular groups. Understanding these accessory components can give us potential opportunities to derive new and novel therapeutics for human disease. New human therapeutics already exploit single domain antibodies of camels and llamas (Cortez-Retamozo et al. 2002; Hamers-Casterman et al. 1993; Harmsen and De Haard 2007; Pant et al. 2006), and diagnostic therapies against nervous necrosis virus (NNV) or avian influenza virus (H9N2) utilize antigen-specific VLRB antibodies from hagfish or lamprey (Im et al. 2018; Jung et al. 2020). Thus, examining

adaptive immunity as a common evolutionary thread within vertebrates as a whole can illuminate these alternate strategies, and this understanding can aid in the development of new therapeutic tools for future immunoapplications.

Compared to mice and humans, nurse sharks retain impressive TCR repertoire diversification strategies. Nurse sharks assemble TCR from IgM or IgW V gene segments or Ig-like TAILV, expanding the combinatorial potential of developing receptors (Criscitiello et al. 2010; Deiss et al. 2019). Additionally, doubly-rearranging NARTCR combines both an Ig-like (NAR) V domain with a supporting TCR V domain to create a novel receptor type (Criscitiello et al. 2006). These diversifying strategies are not limited to nurse sharks. Coelacanth, Xenopus frogs, passeriform birds, and platypus all harbor Ig-like V gene segments (VHδ) in their conventional alpha/delta loci, while galliform birds house these Ig-like VHδ segments in a separate locus (Deiss et al. 2019; Parra et al. 2012a; Parra et al. 2012b; Parra et al. 2010; Saha et al. 2014). Marsupial and monotreme mammals acquired an additional T cell locus (TCRμ) that somatically recombines V, D, and J gene segments (or uses pre-joined segments) into a unique TCR chain with two variable domains, the most distal of which resembles IgH (Parra et al. 2007; Wang et al. 2011). Not only do nurse shark T cells borrow Ig components when recombining and assembling receptors, they derive additional diversity by pirating mechanisms traditionally used by B cells (i.e., AID-catalyzed SHM) to alter antigen binding sites. However, unlike B cells that employ SHM to affinity mature antigen receptors in secondary lymphoid tissues, nurse

sharks incorporate AID-catalyzed SHM in the thymus, likely to salvage TCR in danger of failing thymic selection. Sandbar shark and dromedary camelids also use SHM to alter V region sequences of γδ TCR (Chen et al. 2012; Ciccarese et al. 2014), and most recently, reports indicate the teleost fish Ballan wrasse (Labrus bergulta) somatically mutate both V and C regions of TCRα (Bilal et al. 2018). While these studies in sandbar sharks, Ballan wrasse, and camelids were limited to peripheral lymphoid tissues, it is possible that SHM-induced changes to T cells originated in the thymus of these groups as well.

Agnathan vertebrates (hagfish and lamprey) assemble VLR genes into lymphocytes using two AID homologs (CDA1 and CDA2), and CDA-mediated gene rearrangement in lampreys occurs through a serial gene conversion mechanism, similar to AID-catalyzed Ig gene conversion in some birds and mammals (Guo et al. 2009; Rogozin et al. 2007). CDA1 expression occurs selectively in VLRA (and likely VLRC) lymphocytes within a thymoid (thymus like) region and orchestrates VLRA (and VLRC) gene recombination, while CDA2 expression occurs exclusively in VLRB lymphocytes and mediates VLRB gene assembly (Guo et al. 2009; Rogozin et al. 2007). Thus, there is a precedent of AID (or its homologs) being used in thymus (or thymoid organ) during primary lymphocyte diversification in vertebrates, and the use of AID during primary T cell development in nurse sharks may suggest that AID (or likely another APOBEC-family mutator) mediated lymphocyte diversification in the earliest vertebrate ancestor. The potential consequences of indiscriminate AID transcription (e.g., autoimmune disease, cancer) within the highly

regulated, progressively compartmentalized nuclei of warm-blooded animals likely contributed to the loss of this ancestral mechanism in later vertebrates (like mice and humans).

4.3. Future directions

4.3.1. Assembly of TCRαδ is crucial to understanding mechanisms and regulation of mutation to T cells

Definitive quantification of SHM in nurse shark TCR necessitates complete assembly and annotation of each TCR locus. While we present sound evidence that AID-mediated SHM alters the variable domain of TCRα chain, it remains a possibility that the nurse shark TCRαδ locus includes more V gene segments than expected [based on assembled loci in mouse (~75 Vα) and human (~70 Vα)] (Arden et al. 1995a; Arden et al. 1995b; Murphy and Weaver 2017). Our conservative strategy for grouping V gene segments (into putative family, subfamily, and allelic sequences) identified 42 likely germline sequences. However, none of the V gene segments expressed by the six sharks in this study correspond to those identified in the partial draft assembly of the locus (Deiss et al. 2019). Thus, it still is possible (though unlikely) that the disparities in our dataset are the result of germline-encoded differences between (a great many) V gene segments rather than mutation. A complete germline assembly of T cell loci will provide a comprehensive list of V segments for analysis of SHM within RACE amplicons. With a completely assembled locus, we could directly compare germline and clone sequences for

evidence of base substitutions and definitively establish that these variations are the result of mutation to transcribed sequences.

Additionally, a complete TCR locus assembly could facilitate karyotype construction, which could elucidate the cytogenetic mechanisms responsible for mutation in nurse shark T cells. Currently, no cytogenetic or karyological data exist for the nurse shark or any shark in the same order (Order Orectolobiformes), primarily because it can be difficult to obtain clear bands from traditional chromosome staining techniques (Rocco et al. 2002). Because bands reflect differences in chromatin structure and nucleotide composition between regions of a chromosome, the lack of heterochromatin organization in sharks (and other elasmobranchs) suggests a lack of nuclear compartmentalization as well (Bickmore 2001; Gavrilov and Razin 2015; Jhunjhunwala et al. 2009). Thus, as chromosomes fold to bring V, (D), and J gene segments together during recombination, these loci may align with genes nearby in the nuclear neighborhood. If genes encoding AID are in close proximity to rearranging TCRα chain gene segments, ensuing transcription of rearranged TCRα genes may induce transcription of AID (or other genes), leading to somatic mutation of TCRα chain. However, identifying the nuclear position of transcriptionally active genes in an interphase nucleus (using fluorescent *in situ* hybridization or similar techniques) requires long translocon stretches of nucleotide sequence (>1 kb). Therefore, not only would a complete assembly of the TCRαδ locus assist with mutation analysis of transcribed clones, it could inform the mechanisms

driving this mutation as well. It is important to note that, although the mutational benefits to αβ TCR (assisting failing receptors pass selection) may outweigh the potential dangers of inadvertent AID transcription in nurse sharks, the same benefits may not exist for other vertebrate groups. In fact, nuclear compartmentalization may have evolved in response to such inadvertent transcription of genes. Thus, the lack of compartmentalization (and resulting chromosomal territories) in sharks provides a unique opportunity to examine the means by which sharks regulate of gene expression in the nucleus.

4.3.2. Thymoproteasome may participate in thymic positive selection in nurse shark

Recent studies in mice identified the importance of the thymoproteasome during thymic selection of developing αβ T cells (Murata et al. 2008; Nakagawa et al. 2012; Nitta et al. 2010; Sasaki et al. 2015; Takahama et al. 2012). Standard proteasomes, expressed by all somatic cells, are multi-unit proteolytic structures that degrade peptides for cell recycling. Immune cells and infected somatic cells also express immunoproteasomes, which replace the constitutive catalytic subunits (β1, β2, and β5) of the proteasome by subunits (β1i, β2i, and β5i) specialized to generate peptides suitable for MHC Class I display, thus triggering $CD8^+$ T cells into action (Takahama et al. 2012). In contrast, thymoproteasomes (incorporating a novel β5t subunit in place of β5 or β5i), are expressed exclusively by thymic cortical epithelial cells (cTECs) and produce unique self

peptide motifs that help train $CD8^+$ T cells during positive selection (Kincaid et al. 2016; Nitta et al. 2010; Ohigashi et al. 2017; Sasaki et al. 2015; Takada et al. 2015). A specialized fraction of these cTECs form a distinct subpopulation called nurse cells, which trap viable thymocytes within a cellular "cage" formed by their plasma membrane (Nakagawa et al. 2012). While not essential for T cell differentiation or positive selection, these nurse cell cages provide an MHC-rich microenvironment that encourages secondary TCRα rearrangement (receptor editing), thus optimizing positive selection of TCR (Nakagawa et al. 2012).

Importantly, fluorescent images of flow-sorted mouse thymic cells stained for β5t (thymoproteasome marker) and CD45 (T cell marker) show a ring of $β5t^+$ nurse cells enclosing $CD45^+$ thymocytes (Nakagawa et al. 2012). These nurse cell rings mirror the "AID circles" we observed in nurse shark thymus (see Figure 2-9), where a circle of AID activity surrounded a central TCRα-expressing cell or small cluster of cells. While there are no reports of nurse cells or genes encoding the β5t thymoproteasome component in nurse sharks, a search of the elephant shark genome did reveal a β5t orthologue, suggesting that nurse sharks likely create thymoproteasomes as well (Venkatesh et al. 2014). Identifying this gene in nurse sharks would enable us to develop FISH probes for β5t which, when used with current TCRα and AID probes, could help identify the cells comprising our AID circles. If nurse cells encircle this outer ring of thymocytes, it is possible that, as TCR-alpha chain undergoes receptor editing and testing, nurse cells

simultaneously present self peptide to TCR via MHC, training TCR for self-MHC (and possibly against self-antigen). If this is true, it would lend support to our hypothesis that AID is being used to salvage receptors that fail selection. In this case, during secondary TCRα chain rearrangement (via receptor editing), AID catalyzes mutation to CDR binding regions to alter affinity to self-MHC or self-peptide, salvaging failing receptors. The fact that we observed AID circles in the inner cortex, cortico-medullary junction, and inner medulla of nurse shark thymus suggests that nurse cells may assist with both positive and negative selection (rather than just positive selection, as was observed in mice).

4.3.3. Maintenance of T cells capable of binding free antigen

The conserved tripartite division of lymphocytes in both jawless and jawed vertebrates implies that there were three lymphocyte lineages present in the vertebrate common ancestor, with discernable components of immune system labor partitioned among them (Flajnik 2014). The humoral arm of adaptive immunity, mediated by B (and likely B-like VLRB) cells, responds to extracellular pathogens through neutralization, opsonization, or complement activation (Matsushita 2018; Murphy and Weaver 2017). Cell mediated responses, arbitrated by αβ T (and perhaps VLRA) cells, either aid lymphocyte maturation (i.e., B cell activation by CD4+ helper T cells) or kill infected cells (i.e., CD8+ cytotoxic T cells) (Murphy and Weaver 2017). Less is known about the third division, γδ TCR (and VLRC), likely because γδ TCR represent only 2-10% of circulating T

cells in humans and mice (Murphy and Weaver 2017; Shin et al. 2005; Willcox and Willcox 2019). Yet the conservation of a third cell lineage through nearly 600 million years of evolution strongly suggests the protective and regulatory function(s) of γδ TCR are equally important.

Unlike mice and humans, many vertebrates (e.g., ruminants, chickens, and likely sharks) have circulating T cell populations biased towards γδ T cells rather than αβ T cells (Arstila and Lassila 1993; Mackay and Hein 1989; Telfer and Baldwin 2015). Some γδ TCR recognize and bind intact peptide or non-peptide antigens without requiring presentation by MHC in a manner similar to B cells, and activated γδ T cells mount cytotoxic responses reminiscent of cytotoxic αβ T cells (Allison et al. 2001; Zeng et al. 2012). Still others bind antigen presented by non-classical MHC and generate tissue-specific, restricted repertoires and bind invariable ligands (Adams et al. 2005; Allison and Garboczi 2002; Castro et al. 2015). Like αβ T cells, γδ T cells mature in thymus tissue. However, while thymic selection schools αβ T cells on which peripheral antigens to recognize, γδ T cells do not require exposure to ligands during development and mature cells leave the thymus acquiring only an effector function (Jensen et al. 2008). Despite this lack of training, (ligand-naïve) peripheral γδ T cells mount immediate responses during onset of infectious disease or following immunization through expression of IL-17, a cytokine that recruits neutrophils and monocytes to sites of inflammation (Henry et al. 2010; Jensen et al. 2008). Though many of the ligands that γδ T cells bind remain unclear, the rapid

response to infection implies a critical role in regulation of immune responses. In mice, γδ T cells appear early in embryonic development and are expressed in temporal waves, with each wave colonizing and surveilling specific epithelial tissues with ligand-specific, invariant receptors (Kazen and Adams 2011; Xiong and Raulet 2007). Thus, γδ T cells provide an innate-like immediate response to pathogen invasion and an ongoing (immunological memory) adaptive response to inflammation.

In sharks, the majority of tactics used to supplement combinatorial and/or receptor diversity (e.g., TAILV, IgHV-TCR, NARTCR) primarily refashion TCRδ of γδ T cells (assuming these TCRδ chains pair with TCRγ). Likewise, the Ig-like (VHδ) V gene segments of fish, frogs, birds, and platypus associate with TCRδ constant regions (Parra et al. 2012a; Parra et al. 2012b; Parra et al. 2010; Saha et al. 2014), and TCRμ constant regions of monotremes and mammals are phylogenetically most similar to TCRδ (Parra et al. 2007; Parra et al. 2012a). Thus, the presence of these peculiar receptor components may simply be a strategy for improving upon diversity within the less restrictive γδ T cells. Perhaps the most important purpose of this third lineage of cells is in maintaining a T cell population capable of binding free antigen, combining functions of both fast-acting but general responses of innate immunity and long-term, acquired responses of adaptive immunity.

The limited use of SHM by sharks and camelids in TCRγ and δ chains may be a mechanism to fine-tunes γδ TCR paratopes, to provide a means to alter invariable

receptors as ligands evolve, or (as hypothesized in chapter 3) result inadvertently as cells migrate through the thymus (Adams et al. 2005; Kazen and Adams 2011). The presence of these enigmatic γδ T cell receptors lends credence to the idea that the T cell (and specifically the γδ T cell) evolved first within the adaptive immune system (Pancer et al. 2004). Further insight into these similarities and differences in defense strategies could help elucidate the origins of lymphocyte receptors, making the study of lower fish immune systems ideal for comparative studies of immune evolution.

REFERENCES

Adams EJ, Chien YH, Garcia KC (2005) Structure of a gammadelta t cell receptor in complex with the nonclassical mhc t22. Science 308*:*227-31. doi: 10.1126/science.1106885

Ahmed R, Gray D (1996) Immunological memory and protective immunity: Understanding their relation. Science 272*:*54-60. doi: 10.1126/science.272.5258.54

Alder MN, Herrin BR, Sadlonova A, Stockard CR, Grizzle WE, Gartland LA, Gartland GL, Boydston JA, Turnbough CL, Cooper MD (2008) Antibody responses of variable lymphocyte receptors in the lamprey. Nat Immunol 9*:*319-327. doi: 10.1038/ni1562

Alder MN, Rogozin IB, Iyer LM, Glazko GV, Cooper MD, Pancer Z (2005) Diversity and function of adaptive immune receptors in a jawless vertebrate. Science 310*:*1970-3. doi: 10.1126/science.1119420

Alitheen NB, McClure S, McCullagh P (2010) B-cell development: One problem, multiple solutions. Immunol Cell Biol 88*:*445-50. doi: 10.1038/icb.2009.119

Allison TJ, Garboczi DN (2002) Structure of gammadelta t cell receptors and their recognition of non-peptide antigens. Mol Immunol 38*:*1051-61. doi: 10.1016/S0161-5890(02)00034-2

Allison TJ, Winter CC, Fournie JJ, Bonneville M, Garboczi DN (2001) Structure of a human gammadelta t-cell antigen receptor. Nature 411*:*820-824. doi: 10.1038/35081115

Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H (2019) Signalp 5.0 improves signal peptide predictions using deep neural networks. Nature Biotechnology 37*:*420-423. doi: 10.1038/s41587-019-0036-z

Álvarez-Prado ÁF, Pérez-Durán P, Pérez-García A, Benguria A, Torroja C, de Yébenes VG, Ramiro AR (2018) A broad atlas of somatic hypermutation allows prediction of activation-induced deaminase targets. The Journal of Experimental Medicine 215*:*761-771. doi: 10.1084/jem.20171738

Anderson MK, Shamblott MJ, Litman RT, Litman GW (1995) Generation of immunoglobulin light chain gene diversity in raja erinacea is not associated with somatic rearrangement, an exception to a central paradigm of b cell immunity. J.Exp.Med. 182*:*109-119. doi: 10.1084/jem.182.1.109

Antonacci R, Mineccia M, Lefranc MP, Ashmaoui HM, Lanave C, Piccinni B, Pesole G, Hassanane MS, Massari S, Ciccarese S (2011) Expression and genomic analyses of camelus dromedarius t cell receptor delta (trd) genes reveal a variable domain repertoire enlargement due to cdr3 diversification and somatic mutation. Mol Immunol 48*:*1384-96. doi: 10.1016/j.molimm.2011.03.011

Antonacci R, Vaccarelli G, Di Meo GP, Piccinni B, Miccoli MC, Cribiu EP, Perucatti A, Iannuzzi L, Ciccarese S (2007) Molecular in situ hybridization analysis of sheep and goat bac clones identifies the transcriptional orientation of t cell receptor gamma genes on chromosome 4 in bovids. Veterinary Research Communications 31*:*977-983. doi: 10.1007/s11259-006-0202-x

Arakawa H, Hauschild J, Buerstedde J-M (2002) Requirement of the activation-induced deaminase (aid) gene for immunoglobulin gene conversion. Science 295*:*1301-1306. doi: 10.1126/science.1067308

Archer OK, Sutherland DE, Good RA (1963) Appendix of the rabbit: A homologue of the bursa in the chicken? Nature 200*:*337-9. doi: 10.1038/200337a0

Arden B, Clark SP, Kabelitz D, Mak TW (1995a) Human t-cell receptor variable gene segment families. Immunogenetics 42*:*455-500. doi: 10.1007/BF00172176

Arden B, Clark SP, Kabelitz D, Mak TW (1995b) Mouse t-cell receptor variable gene segment families. Immunogenetics 42*:*501-530. doi: 10.1007/BF00172177

Augustin A, Sim G (1984) T-cell receptors generated via mutations are specific for various major histocompatibility antigens. Cell 39*:*5-12. doi: 10.1016/0092-8674(84)90186-7

Bachl J, Wabl M (1995) Hypermutation in t cells questioned. Nature 375*:*285-6. doi: 10.1038/375285c0

Barreto VM, Magor BG (2011) Activation-induced cytidine deaminase structure and functions: A species comparative view. Dev Comp Immunol 35*:*991-1007. doi: 10.1016/j.dci.2011.02.005

Bassing CH, Swat W, Alt FW (2002) The mechanism and regulation of chromosomal v(d)j recombination. Cell 109*:*S45-S55. doi: 10.1016/S0092-8674(02)00675-X

Becker RS, Knight KL (1990) Somatic diversification of immunoglobulin heavy chain vdj genes: Evidence for somatic gene conversion in rabbits. Cell 63*:*987-997. doi: 10.1016/0092-8674(90)90502-6

Bedel R, Matsuda JL, Brigl M, White J, Kappler J, Marrack P, Gapin L (2012) Lower tcr repertoire diversity in traj18-deficient mice. Nat Immunol 13*:*705-6. doi: 10.1038/ni.2347

Beetz S, Wesch D, Marischen L, Welte S, Oberg H-H, Kabelitz D (2008) Innate immune functions of human γδ t cells. Immunobiology 213*:*173-182. doi: 10.1016/j.imbio.2007.10.006

Bernstein RM, Schluter SF, Lake DF, Marchalonis JJ (1994) Evolutionary conservation and molecular cloning of the recombinase activating gene 1. Biochem. Biophys. Res. Commun. 205*:*687-692. doi: 10.1006/bbrc.1994.2720

Beutler B (2004) Innate immunity: An overview. Molecular Immunology 40*:*845-859. doi: 10.1016/j.molimm.2003.10.005

Bickmore WA (2001) Karyotype analysis and chromosome banding. Encyclopedia of life sciences. John Wiley & Sons, Ltd doi: 10.1038/npg.els.0001160

Borgulya P, Kishi H, Uematsu Y, von Boehmer H (1992) Exclusion and inclusion of alpha and beta t cell receptor alleles. Cell 69*:*529-37. doi: 10.1016/0092-8674(92)90453-J

Brady BL, Steinel NC, Bassing CH (2010) Antigen receptor allelic exclusion: An update and reappraisal. Journal of Immunology 185*:*3801-3808. doi: 10.4049/jimmunol.1001158

Brandes M, Willimann K, Bioley G, Lévy N, Eberl M, Luo M, Tampé R, Lévy F, Romero P, Moser B (2009) Cross-presenting human γδ t cells induce robust cd8⁺ αβ t cell responses. Proceedings of the National Academy of Sciences 106*:*2307-2312. doi: 10.1073/pnas.0810059106

Brazeau MD, Friedman M (2015) The origin and early phylogenetic history of jawed vertebrates. Nature 520*:*490-497. doi: 10.1038/nature14438

Breaux B, Hunter ME, Cruz-Schneider MP, Sena L, Bonde RK, Criscitiello MF (2018) The florida manatee (trichechus manatus latirostris) t cell receptor loci exhibit v subgroup synteny and chain-specific evolution. Developmental & Comparative Immunology 85*:*71-85. doi: 10.1016/j.dci.2018.04.007

Brodeur P, Riblet R (1984) The immunoglobulin heavy chain variable region (igh-v) locus in the mouse. I. One hundred igh-v genes comprise seven families of homologous genes. European Journal of Immunology 14*:*922-930. doi: 10.1002/eji.1830141012

Buslepp J, Wang H, Biddison WE, Appella E, Collins EJ (2003) A correlation between tcr vα docking on mhc and cd8 dependence: Implications for t cell selection. Immunity 19*:*595-606. doi: 10.1016/S1074-7613(03)00269-3

Butler JE, Santiago-Mateo K, Sun XZ, Wertz N, Sinkora M, Francis DH (2011) Antibody repertoire development in fetal and neonatal piglets. Xx. B cell lymphogenesis is absent in the ileal peyer's patches, their repertoire development is antigen dependent, and they are not required for b cell maintenance. J Immunol 187*:*5141-9. doi: 10.4049/jimmunol.1101871

Cantaert T, Schickel JN, Bannock JM, Ng YS, Massad C, Oe T, Wu R, Lavoie A, Walter JE, Notarangelo LD, Al-Herz W, Kilic SS, Ochs HD, Nonoyama S, Durandy A, Meffre E (2015) Activation-induced cytidine deaminase expression in human b cell precursors is essential for central b cell tolerance. Immunity 43*:*884-95. doi: 10.1016/j.immuni.2015.10.002

Casetti R, Agrati C, Wallace M, Sacchi A, Martini F, Martino A, Rinaldi A, Malkovsky M (2009) Cutting edge: Tgf-β1 and il-15 induce foxp3⁺ γδ regulatory t cells in the presence of antigen stimulation. The Journal of Immunology 183*:*3574-3577. doi: 10.4049/jimmunol.0901334

Castro CD, Luoma AM, Adams EJ (2015) Coevolution of t-cell receptors with mhc and non-mhc ligands. Immunological Reviews 267*:*30-55. doi: 10.1111/imr.12327

Chen H, Bernstein H, Ranganathan P, Schluter S (2012) Somatic hypermutation of tcr γ v genes in the sandbar shark. Developmental & Comparative Immunology 37*:*176-183. doi: 10.1016/j.dci.2011.08.018

Chen H, Kshirsagar S, Jensen I, Lau K, Covarrubias R, Schluter SF, Marchalonis JJ (2009) Characterization of arrangement and expression of the t cell receptor gamma locus in the sandbar shark. Proc Natl Acad Sci U S A 106*:*8591-6. doi: 10.1073/pnas.0811283106

Cheynier R, Henrichwark S, Wain Hobson S (1998) Somatic hypermutation of the t cell receptor v beta gene in microdissected splenic white pulps from hiv-1-positive patients. European Journal of Immunology 28*:*1604-1610. doi: 10.1002/(SICI)1521-4141(199805)28:05<1604::AID-IMMU1604>3.0.CO;2-R

Chien Yh, Bonneville M (2006) Gamma delta t cell receptors. Cellular and Molecular Life Sciences CMLS 63*:*2089-2094. doi: 10.1007/s00018-006-6020-z

Chien YH, Jores R, Crowley MP (1996) Recognition by γ/δ t cells. Annual Review of Immunology 14*:*511-532. doi: 10.1146/annurev.immunol.14.1.511

Choudhary M, Tamrakar A, Singh AK, Jain M, Jaiswal A, Kodgire P (2018) Aid biology: A pathological and clinical perspective. International Reviews of Immunology 37*:*37-56. doi: 10.1080/08830185.2017.1369980

Ciccarese S, Vaccarelli G, Lefranc MP, Tasco G, Consiglio A, Casadio R, Linguiti G, Antonacci R (2014) Characteristics of the somatic hypermutation in the camelus dromedarius t cell receptor gamma (trg) and delta (trd) variable domains. Dev Comp Immunol 46*:*300-13. doi: 10.1016/j.dci.2014.05.001

Conticello S (2008) The aid/apobec family of nucleic acid mutators. GenomeBiology.com 9*:*229. doi: 10.1186/gb-2008-9-6-229

Conticello SG, Langlois MA, Yang Z, Neuberger MS (2007) DNA deamination in immunity: Aid in the context of its apobec relatives. Advances in immunology. Academic Press

Conticello SG, Thomas CJF, Petersen-Mahrt SK, Neuberger MS (2005) Evolution of the aid/apobec family of polynucleotide (deoxy)cytidine deaminases. Molecular Biology and Evolution 22*:*367-377. doi: 10.1093/molbev/msi026

Cooper MD, Alder MN (2006) The evolution of adaptive immune systems. Cell 124*:*815-22. doi: 10.1016/j.cell.2006.02.001

Criscitiello M, Flajnik M (2007) Four primordial immunoglobulin light chain isotypes, including lambda and kappa, identified in the most primitive living jawed vertebrates. European Journal of Immunology 37*:*2683-2694. doi: 10.1002/eji.200737263

Criscitiello M, Saltis M, Flajnik M (2006) An evolutionarily mobile antigen receptor variable region gene: Doubly rearranging nar-tcr genes in sharks. Proceedings of the National Academy of Sciences of the United States of America 103*:*5036-5041. doi: 10.1073/pnas.0507074103

Criscitiello MF, Kamper SM, McKinney EC (2004a) Allelic polymorphism of tcralpha chain constant domain genes in the bicolor damselfish. Dev Comp Immunol 28*:*781-92. doi: 10.1016/j.dci.2003.12.004

Criscitiello MF, Ohta Y, Graham MD, Eubanks JO, Chen PL, Flajnik MF (2012) Shark class ii invariant chain reveals ancient conserved relationships with cathepsins and mhc class ii. Dev Comp Immunol 36*:*521-33. doi: 10.1016/j.dci.2011.09.008

Criscitiello MF, Ohta Y, Saltis M, McKinney EC, Flajnik MF (2010) Evolutionarily conserved tcr binding sites, identification of t cells in primary lymphoid tissues, and surprising trans-rearrangements in nurse shark. J Immunol 184*:*6950-60. doi: 10.4049/jimmunol.0902774

Criscitiello MF, Wermenstam NE, Pilstrom L, McKinney EC (2004b) Allelic polymorphism of t-cell receptor constant domains is widespread in fishes. Immunogenetics 55*:*818-24. doi: 10.1007/s00251-004-0652-7

Crouch EE, Li Z, Takizawa M, Fichtner-Feigl S, Gourzi P, Montano C, Feigenbaum L, Wilson P, Janz S, Papavasiliou FN, Casellas R (2007) Regulation of aid expression in the immune response. J Exp Med 204*:*1145-56. doi: 10.1084/jem.20061952

Das S, Li J, Hirano M, Sutoh Y, Herrin BR, Cooper MD (2015) Evolution of two prototypic t cell lineages. Cellular Immunology 296*:*87-94. doi: 10.1016/j.cellimm.2015.04.007

Das S, Li J, Holland SJ, Iyer LM, Hirano M, Schorpp M, Aravind L, Cooper MD, Boehm T (2014) Genomic donor cassette sharing during vlra and vlrc assembly in jawless vertebrates. Proceedings of the National Academy of Sciences 111*:*14828-14833. doi: 10.1073/pnas.1415580111

Deiss TC, Breaux B, Ott JA, Daniel RA, Chen PL, Castro CD, Ohta Y, Flajnik MF, Criscitiello MF (2019) Ancient use of ig variable domains contributes significantly to the tcrδ repertoire. The Journal of Immunology 203*:*1265-1275. doi: 10.4049/jimmunol.1900369

Deng L, Velikovsky CA, Xu G, Iyer L, Tasumi S, Kerzic M, Flajnik M, Aravind L, Pancer Z, Mariuzza R (2010a) A structural basis for antigen recognition by the t cell-like lymphocytes of sea lamprey. Proceedings of the National Academy of Sciences of the United States of America 107*:*13408-13413. doi: 10.1073/pnas.1005475107

Deng L, Velikovsky CA, Xu G, Iyer LM, Tasumi S, Kerzic MC, Flajnik MF, Aravind L, Pancer Z, Mariuzza RA (2010b) A structural basis for antigen recognition by the t cell-like lymphocytes of sea lamprey. Proceedings of the National Academy of Sciences of the United States of America 107*:*13408-13413. doi: 10.1073/pnas.1005475107

Diaz M, Greenberg A, Flajnik M (1998) Somatic hypermutation of the new antigen receptor gene (nar) in the nurse shark does not generate the repertoire: Possible role in antigen-driven reactions in the absence of germinal centers. Proceedings of the National Academy of Sciences of the United States of America 95*:*14343-8. doi: 10.1073/pnas.95.24.14343

Diaz M, Stanfield RL, Greenberg AS, Flajnik MF (2002) Structural analysis, selection, and ontogeny of the shark new antigen receptor (ignar): Identification of a new locus preferentially expressed in early development. Immunogenetics 54*:*501-512. doi: 10.1007/s00251-002-0479-z

Diaz M, Velez J, Singh M, Cerny J, Flajnik MF (1999) Mutational pattern of the nurse shark antigen receptor gene (nar) is similar to that of mammalian ig genes and to spontaneous mutations in evolution: The translesion synthesis model of somatic hypermutation. Int.Immunol. 11*:*825-833. doi: 10.1093/intimm/11.5.825

Dooley H, Flajnik MF (2006) Antibody repertoire development in cartilaginous fish. Dev Comp Immunol 30doi: 10.1016/j.dci.2005.06.022

Dooley H, Stanfield RL, Brady RA, Flajnik MF (2006) First molecular and biochemical analysis of *in vivo* affinity maturation in an ectothermic vertebrate. Proceedings of the National Academy of Sciences of the United States of America 103*:*1846-1851. doi: 10.1073/pnas.0508341103

Dunnick WA, Collins JT, Shi J, Westfield G, Fontaine C, Hakimpour P, Papavasiliou FN (2009) Switch recombination and somatic hypermutation are controlled by the heavy chain 3' enhancer region. Journal of Experimental Medicine 206*:*2613-2623. doi: 10.1084/jem.20091280

Fahl SP, Coffey F, Wiest DL (2014) Origins of γδ t cell effector subsets: A riddle wrapped in an enigma. The Journal of Immunology 193*:*4289-4294. doi: 10.4049/jimmunol.1401813

Flajnik MF (2002) Comparative analyses of immunoglobulin genes: Surprises and portents. Nat.Rev.Immunol. 2*:*688-698. doi: 10.1038/nri889

Flajnik MF (2014) Re-evaluation of the immunological big bang. Current Biology 24*:*R1060-R1065. doi: 10.1016/j.cub.2014.09.070

Flajnik MF, Deschacht N, Muyldermans S (2011) A case of convergence: Why did a simple alternative to canonical antibodies arise in sharks and camels? PLoS Biol 9*:*e1001120. doi: 10.1371/journal.pbio.1001120

Flajnik MF, Rumfelt LL (2000) The immune system of cartilaginous fish. Curr Top Microbiol Immunol 248doi: 10.1007/978-3-642-59674-2_11

Garcia KC, Adams EJ (2005) How the t cell receptor sees antigen—a structural view. Cell 122*:*333-336. doi: 10.1016/j.cell.2005.07.015

Garcia KC, Degano M, Pease LR, Huang M, Peterson PA, Teyton L, Wilson IA (1998) Structural basis of plasticity in t cell receptor recognition of a self peptide-mhc antigen. Science 279*:*1166-72. doi: 10.1126/science.279.5354.1166

Gascoigne NRJ, Alam SM (1999) Allelic exclusion of the t cell receptor α-chain: Developmental regulation of a post-translational event. Seminars in Immunology 11*:*337-347. doi: 10.1006/smim.1999.0190

Gavrilov AA, Razin SV (2015) Compartmentalization of the cell nucleus and spatial organization of the genome. Molecular Biology 49*:*21-39. doi: 10.1134/s0026893315010033

Gellert M (2002) V(d)j recombination: Rag proteins, repair factors, and regulation. Annu Rev Biochem 71*:*101-32. doi: 10.1146/annurev.biochem.71.090501.150203

Germain RN (2002) T-cell development and the cd4–cd8 lineage decision. Nature Reviews Immunology 2*:*309-322. doi: 10.1038/nri798

Glick B, Chang TS, Jaap RG (1956) The bursa of fabricius and antibody production. Poultry Science 35*:*224-225. doi: 10.3382/ps.0350224

Gober H-J, Kistowska M, Angman L, Jenö P, Mori L, De Libero G (2003) Human t cell receptor gammadelta cells recognize endogenous mevalonate metabolites in tumor cells. The Journal of Experimental Medicine 197*:*163-168. doi: 10.1084/jem.20021500

Good RA, Finstad J (1966) The phylogenetic development of immune responses and the germinal center system. *In* Cottier H, Odartchenko N, Schindler R, Congdon CC (eds.) Germinal Centers in Immune Responses. Springer-Verlag, New York Inc, University of Bern, Switzerland.

Greenberg AS, Avila D, Hughes M, Hughes A, McKinney EC, Flajnik MF (1995) A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. Nature 374*:*168-173. doi: 10.1038/374168a0

Guo C, Yoon HS, Franklin A, Jain S, Ebert A, Cheng H-L, Hansen E, Despo O, Bossen C, Vettermann C, Bates JG, Richards N, Myers D, Patel H, Gallagher M, Schlissel MS, Murre C, Busslinger M, Giallourakis CC, Alt FW (2011) Ctcf-binding elements mediate control of v(d)j recombination. Nature 477*:*424-430. doi: 10.1038/nature10495

Guo J, Hawwari A, Li H, Sun Z, Mahanta SK, Littman DR, Krangel MS, He Y-W (2002) Regulation of the tcrα repertoire by the survival window of cd4+cd8+ thymocytes. Nature Immunology 3*:*469. doi: 10.1038/ni791

Guo P, Hirano M, Herrin BR, Li J, Yu C, Sadlonova A, Cooper MD (2009) Dual nature of the adaptive immune system in lampreys. Nature 459*:*796-801. doi: 10.1038/nature08068

Hackett J, Jr, Stebbins C, Rogerson B, Davis MM, Storb U (1992) Analysis of a t cell receptor gene as a target of the somatic hypermutation mechanism. The Journal of Experimental Medicine 176*:*225-231. doi: 10.1084/jem.176.1.225

Hansen VL, Miller RD (2015) The evolution and structure of atypical t cell receptors. *In* E. H, L. DP (eds.) Pathogen-host interactions: Antigenic variation v. Somatic adaptations. Results and problems in cell differentiation. Springer, Cham

Hara Y, Yamaguchi K, Onimaru K, Kadota M, Koyanagi M, Keeley SD, Tatsumi K, Tanaka K, Motone F, Kageyama Y, Nozu R, Adachi N, Nishimura O, Nakagawa R, Tanegashima C, Kiyatake I, Matsumoto R, Murakumo K, Nishida K, Terakita A, Kuratani S, Sato K, Hyodo S, Kuraku S (2018) Shark genomes provide insights into elasmobranch evolution and the origin of vertebrates. Nature Ecology & Evolution 2*:*1761-1771. doi: 10.1038/s41559-018-0673-5

Herrin BR, Alder MN, Roux KH, Sina C, Ehrhardt GRA, Boydston JA, Turnbough CL, Cooper MD (2008) Structure and specificity of lamprey monoclonal antibodies. Proceedings of the National Academy of Sciences 105*:*2040-2045. doi: 10.1073/pnas.0711619105

Herrin BR, Cooper MD (2010) Alternative adaptive immunity in jawless vertebrates. The Journal of Immunology 185*:*1367-1374. doi: 10.4049/jimmunol.0903128

Hirano M (2015) Evolution of vertebrate adaptive immunity: Immune cells and tissues, and aid/apobec cytidine deaminases. Bioessays 37*:*877-87. doi: 10.1002/bies.201400178

Honjo T, Kinoshita K, Muramatsu M (2002) Molecular mechnism of class switch recombination: Linkage with somatic hypermutation. Annual Review of Immunology 20*:*165-196. doi: 10.1146/annurev.immunol.20.090501.112049

Hsu E (2009) V(d)j recombination: Of mice and sharks. *In* Ferrier P (ed.) V(d)j recombination. Springer New York, New York, NY.

Hsu E (2016) Assembly and expression of shark ig genes. The Journal of Immunology 196*:*3517-3523. doi: 10.4049/jimmunol.1600164

Hsu E (2018) Immune system receptors in vertebrates: Immunoglobulins. Reference Module in Life Sciences doi: 10.1016/B978-0-12-809633-8.20721-8

Huang HV, Dreyer WJ (1978) Bursectomy *in ovo* blocks the generation of immunoglobulin diversity. The Journal of Immunology 121*:*1738-1747. doi: 121/5/1738

Huesmann M, Scott B, Kisielow P, von Boehmer H (1991) Kinetics and efficacy of positive selection in the thymus of normal and t cell receptor transgenic mice. Cell 66*:*533-540. doi: 10.1016/0092-8674(81)90016-7

Hwang JK, Alt FW, Yeap L-S (2015) Related mechanisms of antibody somatic hypermutation and class switch recombination. Microbiology Spectrum 3doi: 10.1128/microbiolspec.MDNA3-0037-2014

Ito S, Nagaoka H, Shinkura R, Begum N, Muramatsu M, Nakata M, Honjo T (2004) Activation-induced cytidine deaminase shuttles between nucleus and cytoplasm like apolipoprotein b mrna editing catalytic polypeptide 1. Proceedings of the National Academy of Sciences of the United States of America 101*:*1975-1980. doi: 10.1073/pnas.0307335101

Jack R, Du Pasquier L (2019) The triumph of individualism: Evolution of somatically generated adaptive immune systems. Evolutionary concepts in immunology. Springer International Publishing, Cham.

Janvier P (2011) Comparative anatomy: All vertebrates do have vertebrae. Current Biology 21*:*R661-R663. doi: 10.1016/j.cub.2011.07.014

Jerne NK (1971) The somatic generation of immune recognition. Eur J Immunol 1*:*1-9. doi: 10.1002/eji.1830010102

Jhunjhunwala S, van Zelm MC, Peak MM, Murre C (2009) Chromatin architecture and the generation of antigen receptor diversity. Cell 138*:*435-448. doi: 10.1016/j.cell.2009.07.016

Jung D, Giallourakis C, Mostoslavsky R, Alt FW (2006) Mechanism and control of v(d)j recombination at the immunoglobulin heavy chain locus. Annu Rev Immunol 24doi: 10.1146/annurev.immunol.23.021704.115830

Kabelitz D (2011) Γδ t-cells: Cross-talk between innate and adaptive immunity. Cellular and Molecular Life Sciences 68*:*2331. doi: 10.1007/s00018-011-0696-4

Kasahara M, Sutoh Y (2014) Chapter two - two forms of adaptive immunity in vertebrates: Similarities and differences. *In* Frederick WA (ed.) Advances in immunology. Academic Press

Kasahara M, Vazquez M, Sato K, McKinney EC, Flajnik MF (1992) Evolution of the major histocompatibility complex: Isolation of class ii a cdna clones from the cartilaginous fish. Proceedings of the National Academy of Sciences of the United States of America 89*:*6688-6692. doi: 10.1073/pnas.89.15.6688

Kasamatsu J, Sutoh Y, Fugo K, Otsuka N, Iwabuchi K, Kasahara M (2010) Identification of a third variable lymphocyte receptor in the lamprey. Proceedings of the National Academy of Sciences 107*:*14304-14308. doi: 10.1073/pnas.1001910107

Kazen AR, Adams EJ (2011) Evolution of the v, d, and j gene segments used in the primate γδ t-cell receptor reveals a dichotomy of conservation and diversity. Proceedings of the National Academy of Sciences of the United States of America 108*:*E332-E340. doi: 10.1073/pnas.1105105108

Kikutani H, Inui S, Sato R, Barsumian EL, Owaki H, Yamasaki K, Kaisho T, Uchibayashi N, Hardy RR, Hirano T, Tsunasawa S, Sakiyama F, Suemura M, Kishimoto T (1986) Molecular structure of human lymphocyte receptor for immunoglobulin e. Cell 47*:*657-665. doi: 10.1016/0092-8674(86)90508-8

Kincaid EZ, Murata S, Tanaka K, Rock KL (2016) Specialized proteasome subunits play an essential role in thymic selection of cd8(+) t cells. Nature Immunology 17*:*938-945. doi: 10.1038/ni.3480

Kipps TJ (2010) Chapter 5. The organization and structure of lymphoid tissues. *In* Lichtman MA, Kipps TJ, Seligsohn U, Kaushansky K, Prchal JT (eds.) Williams hematology, 8e. The McGraw-Hill Companies, New York, NY.

Komori A, Xu Z, Wu X, Zan H, Casali P (2006) Biased da/dt somatic hypermutation as regulated by the heavy chain intronic iemu enhancer and 3'ealpha enhancers in human lymphoblastoid b cells. Molecular Immunology 43*:*1817-1826. doi: 10.1016/j.molimm.2005.10.018

Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, Ganapathy G, Wang Z, Rasko DA, McCombie WR, Jarvis ED, Phillippy AM (2012) Hybrid error correction and de novo assembly of single-molecule sequencing reads. Nat Biotech 30*:*693-700. doi: 10.1038/nbt.2280

Kreslavsky T, Gleimer M, Garbe AI, von Boehmer H (2010) Aβ versus γδ fate choice: Counting the t-cell lineages at the branch point. Immunological Reviews 238*:*169-181. doi: 10.1111/j.1600-065X.2010.00947.x

Kreslavsky T, Kim HJ, Koralov SB, Ghitza D, Buch T, Cantor H, Rajewsky K, von Boehmer H (2013) Negative selection, not receptor editing, is a physiological response of autoreactive thymocytes. J Exp Med 210*:*1911-8. doi: 10.1084/jem.20130876

Krijger PH, Tsaalbi-Shtylik A, Wit N, van den Berk PCM, de Wind N, Jacobs H (2013) Rev1 is essential in generating g to c transversions downstream of the ung2 pathway but not the msh2+ung2 hybrid pathway. European Journal of Immunology 43*:*2765-2770. doi: 10.1002/eji.201243191

Kronenberg M, Siu G, Hood LE, Shastri N (1986) The molecular genetics of the t cell antigen receptor and t cell antigen recognition. Ann Rev Immunol 4doi: 10.1146/annurev.iy.04.040186.002525

Kuklina EM (2006) Revision of the antigen receptor of t-lymphocytes. Biochemistry 71:827-837. doi: 10.1134/S0006297906080025

Kuo TC, Schlissel MS (2009) Mechanisms controlling expression of the rag locus during lymphocyte development. Current Opinion in Immunology 21:173-178. doi: 10.1016/j.coi.2009.03.008

Kuraoka M, Holl TM, Liao D, Womble M, Cain DW, Reynolds AE, Kelsoe G (2011) Activation-induced cytidine deaminase mediates central tolerance in b cells. Proc Natl Acad Sci U S A 108:11560-5. doi: 10.1073/pnas.1102571108

Lafaille JJ, Haas W, Coutinho A, Tonegawa S (1990) Positive selection of γδ t cells. Immunology Today 11:75-78. doi: 10.1016/0167-5699(90)90030-D

Lanning D, Sethupathi P, Rhee KJ, Zhai SK, Knight KL (2000) Intestinal microflora and diversification of the rabbit antibody repertoire. J Immunol 165:2012-9. doi: 10.4049/jimmunol.165.4.2012

Lantelme E, Orlando L, Porcedda P, Turinetto V, De Marchi M, Amoroso A, Mantovani S, Giachino C (2008) An in vitro model of t cell receptor revision in mature human cd8+ t cells. Mol Immunol 45:328-37. doi: 10.1016/j.molimm.2007.06.153

Lee SS, Tranchina D, Ohta Y, Flajnik MF, Hsu E (2002) Hypermutation in shark immunoglobulin light chain genes results in contiguous substitutions. Immunity 16:571-82. doi: 10.1016/s1074-7613(02)00300-x

Lee V, Huang JL, Lui MF, Malecek K, Ohta Y, Mooers A, Hsu E (2008) The evolution of multiple isotypic igm heavy chain genes in the shark. J Immunol 180:7461-70. doi: 10.4049/jimmunol.180.11.7461

Lefranc M-P (2014) Immunoglobulin and t cell receptor genes: Imgt and the birth and rise of immunoinformatics. Frontiers in Immunology 5*:*22-22. doi: 10.3389/fimmu.2014.00022

Lefranc MP, Pommie C, Ruiz M, Giudicelli V, Foulquier E, Truong L, Thouvenin-Contet V, Lefranc G (2003) Imgt unique numbering for immunoglobulin and t cell receptor variable domains and ig superfamily v-like domains. Dev Comp Immunol 27*:*55-77. doi: 10.1016/S0145-305X(02)00039-3

Li Z, Woo CJ, Iglesias-Ussel MD, Ronai D, Scharff MD (2004) The generation of antibody diversity through somatic hypermutation and class switch recombination. Genes Dev 18*:*1-11. doi: 10.1101/gad.1161904

Litman GW, Anderson MK, Rast JP (1999) Evolution of antigen binding receptors. Annu Rev Immunol 17doi: 10.1146/annurev.immunol.17.1.109

Litman GW, Cannon JP, Dishaw LJ (2005) Reconstructing immune phylogeny: New perspectives. Nat Rev Immunol 5*:*866-879. doi: 10.1038/nri1712

Luer C, Walsh CJ, Bodine AB, Wyffels JT, Scott TR (1995) The elasmobranch thymus: Anatomical, histological, and preleminary functional characterization. Journal of Experimental Zoology 273*:*342-354. doi: 10.1002/jez.1402730408

MacLennan ICM (1994) Germinal centers. Annual Review of Immunology 12*:*117-139. doi: 10.1146/annurev.iy.12.040194.001001

Magor BG (2015) Antibody affinity maturation in fishes-our current understanding. Biology 4*:*512-524. doi: 10.3390/biology4030512

Malecek K, Brandman J, Brodsky JE, Ohta Y, Flajnik MF, Hsu E (2005) Somatic hypermutation and junctional diversification at ig heavy chain loci in the nurse shark. J Immunol 175*:*8105-15. doi: 10.4049/jimmunol.175.12.8105

Malecek K, Lee V, Feng W, Huang JL, Flajnik MF, Ohta Y, Hsu E (2008) Immunoglobulin heavy chain exclusion in the shark. PLoS Biol 6*:*e157. doi: 10.1371/journal.pbio.0060157

Mantovani S, Palermo B, Garbelli S, Campanelli R, Robustelli Della Cuna G, Gennari R, Benvenuto F, Lantelme E, Giachino C (2002) Dominant tcr-alpha requirements for a self antigen recognition in humans. The journal of Immunology 169*:*6253-6260. doi: 10.4049/jimmunol.169.11.6253

Marra NJ, Stanhope MJ, Jue NK, Wang M, Sun Q, Pavinski Bitar P, Richards VP, Komissarov A, Rayko M, Kliver S, Stanhope BJ, Winkler C, O'Brien SJ, Antunes A, Jorgensen S, Shivji MS (2019) White shark genome reveals ancient elasmobranch adaptations associated with wound healing and the maintenance of genome stability. Proceedings of the National Academy of Sciences 116*:*4446-4455. doi: 10.1073/pnas.1819778116

Marshall B, Schulz R, Zhou M, Mellor A (1999) Alternative splicing and hypermutation of a nonproductively rearranged tcr alpha-chain in a t cell hybridoma. J Immunol 162*:*871-7. doi: 162/2/871

Martin AP (1999) Substitution rates of organelle and nuclear genes in sharks: Implicating metabolic rate (again). Mol Biol Evol 16*:*996-1002. doi: 10.1093/oxfordjournals.molbev.a026189

Maul RW, Gearhart PJ (2010) Chapter six - aid and somatic hypermutation. *In* Alt FW (ed.) Advances in immunology. Academic Press

McCormack WT, Thompson CB (1990) Chicken igl variable region gene conversions display pseudogene donor preference and 5' to 3' polarity. Genes Dev 4*:*548-558. doi: 10.1101/gad.4.4.548

McGargill MA, Derbinski JM, Hogquist KA (2000) Receptor editing in developing t cells. Nature Immunology 1*:*336-341. doi: 10.1038/79790

Miller RD (2010) Those other mammals: The immunoglobulins and t cell receptors of marsupials and monotremes. Seminars in Immunology 22*:*3-9. doi: 10.1016/j.smim.2009.11.005

Miracle A, Anderson M, Litman R, Walsh C, Luer C, Rothenberg E, Litman G (2001) Complex expression patterns of lymphocyte-specific genes during the development of cartilaginous fish implicate unique lymphoid tissues in generating an immune repertoire. International Immunology 13*:*567-580. doi: 10.1093/intimm/13.4.567

Morisawa T, Marusawa H, Ueda Y, Iwai A, Okazaki I-m, Honjo T, Chiba T (2008) Organ-specific profiles of genetic changes in cancers caused by activation-induced cytidine deaminase expression. International Journal of Cancer 123*:*2735-2740. doi: 10.1002/ijc.23853

Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T (2000) Class switch recombination and hypermutation require activation-induced cytidine deaminase (aid), a potential rna editing enzyme. Cell 102*:*553-563. doi: 10.1016/S0092-8674(00)00078-7

Muramatsu M, Nagaoka H, Shinkura R, Begum NA, Honjo T (2007) Discovery of activation-induced cytidine deaminase, the engraver of antibody memory. Advances in immunology. Academic Press

Murata S, Takahama Y, Tanaka K (2008) Thymoproteasome: Probable role in generating positively selecting peptides. Current Opinion in Immunology 20*:*192-196. doi: 10.1016/j.coi.2008.03.002

Murphy K, Weaver C (2017) Janeway's immunobiology. Garland Science, New York.

Nakagawa Y, Ohigashi I, Nitta T, Sakata M, Tanaka K, Murata S, Kanagawa O, Takahama Y (2012) Thymic nurse cells provide microenvironment for secondary t cell receptor α rearrangement in cortical thymocytes. Proceedings of the National Academy of Sciences 109*:*20572-20577. doi: 10.1073/pnas.1213069109

Neely HR, Guo J, Flowers EM, Criscitiello MF, Flajnik MF (2018) "Double-duty" conventional dendritic cells in the amphibian xenopus as the prototype for antigen presentation to b cells. European Journal of Immunology 48*:*430-440. doi: 10.1002/eji.201747260

Nielsen J, Hedeholm RB, Heinemeier J, Bushnell PG, Christiansen JS, Olsen J, Ramsey CB, Brill RW, Simon M, Steffensen KF, Steffensen JF (2016) Eye lens radiocarbon reveals centuries of longevity in the greenland shark (somniosus microcephalus). Science 353*:*702-4. doi: 10.1126/science.aaf1703

Nitta T, Murata S, Sasaki K, Fujii H, Ripen AM, Ishimaru N, Koyasu S, Tanaka K, Takahama Y (2010) Thymoproteasome shapes immunocompetent repertoire of cd8+ t cells. Immunity 32*:*29-40. doi: 10.1016/j.immuni.2009.10.009

Odegard VH, Schatz DG (2006) Targeting of somatic hypermutation. Nat Rev Immunol 6*:*573-83. doi: 10.1038/nri1896

Ohigashi I, Ohte Y, Setoh K, Nakase H, Maekawa A, Kiyonari H, Hamazaki Y, Sekai M, Sudo T, Tabara Y, Sawai H, Omae Y, Yuliwulandari R, Tanaka Y, Mizokami M, Inoue H, Kasahara M, Minato N, Tokunaga K, Tanaka K, Matsuda F, Murata S, Takahama Y (2017) A human psmb11 variant affects thymoproteasome processing and cd8(+) t cell production. JCI Insight 2*:*e93664. doi: 10.1172/jci.insight.93664

Ohlin M, Scheepers C, Corcoran M, Lees WD, Busse CE, Bagnara D, Thörnqvist L, Bürckert J-P, Jackson KJL, Ralph D, Schramm CA, Marthandan N, Breden F, Scott J, Matsen Iv FA, Greiff V, Yaari G, Kleinstein SH, Christley S, Sherkow JS, Kossida S, Lefranc M-P, van Zelm MC, Watson CT, Collins AM (2019) Inferred allelic variants of immunoglobulin receptor genes: A system for their evaluation, documentation, and naming. Frontiers in Immunology 10*:*435-435. doi: 10.3389/fimmu.2019.00435

Ohta Y, Flajnik M (2006) Igd, like igm, is a primordial immunoglobulin class perpetuated in most jawed vertebrates. Proceedings of the National Academy of Sciences of the United States of America 103*:*10723-10728. doi: 10.1073/pnas.0601407103

Ohta Y, Kasahara M, O'Connor TD, Flajnik MF (2019) Inferring the "primordial immune complex": Origins of mhc class i and antigen receptors revealed by comparative genomics. The Journal of Immunology 203*:*1882-1896. doi: 10.4049/jimmunol.1900597

Okazaki I-m, Hiai H, Kakazu N, Yamada S, Muramatsu M, Kinoshita K, Honjo T (2003) Constitutive expression of aid leads to tumorigenesis. The Journal of Experimental Medicine 197*:*1173-1181. doi: 10.1084/jem.20030275

Orjalo A, Johansson HE, Ruth JL (2011) Stellaris[trade] fluorescence in situ hybridization (fish) probes: A powerful tool for mrna detection. Nat Meth 8*:*i-ii. doi: 10.1038/nmeth.f.349

Ott JA, Castro CD, Deiss TC, Ohta Y, Flajnik MF, Criscitiello MF (2018) Somatic hypermutation of t cell receptor α chain contributes to selection in nurse shark thymus. eLife 7*:*e28477. doi: 10.7554/eLife.28477

Pancer Z, Amemiya CT, Ehrhardt GRA, Ceitlin J, Larry Gartland G, Cooper MD (2004) Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. Nature 430*:*174-180. doi: 10.1038/nature02740

Pancer Z, Saha NR, Kasamatsu J, Suzuki T, Amemiya CT, Kasahara M, Cooper MD (2005) Variable lymphocyte receptors in hagfish. Proc Natl Acad Sci U S A 102*:*9224-9. doi: 10.1073/pnas.0503792102

Parng CL, Hansal S, Goldsby RA, Osborne BA (1996) Gene conversion contributes to ig light chain diversity in cattle. J Immunol 157*:*5478-86. doi: 157/12/5478

Parra ZE, Baker ML, Hathaway J, Lopez AM, Trujillo J, Sharp A, Miller RD (2008) Comparative genomic analysis and evolution of the t cell receptor loci in the opossum monodelphis domestica. BMC Genomics 9*:*111. doi: 10.1186/1471-2164-9-111

Parra ZE, Baker ML, Schwarz R, Deakin J, Lindblad-Toh K, Miller RD (2007) A unique t cell receptor discovered in marsupials. Proc Natl Acad Sci USA 104doi: 10.1073/pnas.0609106104

Parra ZE, Lillie M, Miller RD (2012a) A model for the evolution of the mammalian t-cell receptor α/δ and μ loci based on evidence from the duckbill platypus. Molecular Biology and Evolution 29*:*3205-3214. doi: 10.1093/molbev/mss128

Parra ZE, Mitchell K, Dalloul RA, Miller RD (2012b) A second tcrdelta locus in galliformes uses antibody-like v domains: Insight into the evolution of tcrdelta and tcrmu genes in tetrapods. J Immunol 188*:*3912-9. doi: 10.4049/jimmunol.1103521

Parra ZE, Ohta Y, Criscitiello MF, Flajnik MF, Miller RD (2010) The dynamic tcrdelta: Tcrdelta chains in the amphibian xenopus tropicalis utilize antibody-like v genes. Eur J Immunol 40*:*2319-29. doi: 10.1002/eji.201040515

Pavri R, Nussenzweig MC (2011) Chapter 1 - aid targeting in antibody diversity. *In* Alt FW, Austen KF, Honj T, Melchers F, Uhr JW, Unanue ER (eds.) Advances in immunology. Academic Press

Perera J, Huang H (2015) The development and function of thymic b cells. Cellular and Molecular Life Sciences 72*:*2657-2663. doi: 10.1007/s00018-015-1895-1

Perera J, Meng L, Meng F, Huang H (2013) Autoreactive thymic b cells are efficient antigen-presenting cells of cognate self-antigens for t cell negative selection. Proceedings of the National Academy of Sciences 110*:*17011-17016. doi: 10.1073/pnas.1313001110

Petrie HT, Livak F, Schatz DG, Strasser A, Crispe IN, Shortman K (1993) Multiple rearrangements in t cell receptor alpha chain genes maximize the production of useful thymocytes. J Exp Med 178*:*615-22. doi: 10.1084/jem.178.2.615

Qin H, Suzuki K, Nakata M, Chikuma S, Izumi N, Thi Huong L, Maruya M, Fagarasan S, Busslinger M, Honjo T, Nagaoka H (2011) Activation-induced cytidine deaminase expression in cd4+ t cells is associated with a unique il-10-producing subset that increases with age. PLOS ONE 6*:*e29141. doi: 10.1371/journal.pone.0029141

Qin T, Zhao H, Zhu H, Wang D, Du W, Hao H (2015) Immunoglobulin genomics in the prairie vole (microtus ochrogaster). Immunol Lett 166doi: 10.1016/j.imlet.2015.06.001

Raj A, Tyagi S (2010) Chapter 17 - detection of individual endogenous rna transcripts in situ using multiple singly labeled probes. *In* Walter NG (ed.) Methods in enzymology. Academic Press

Rast JP, Anderson MK, Strong SJ, Luer C, Litman RT, Litman GW (1997) A, β, γ and δt cell antigen receptor genes arose early in vertebrate phylogeny. Immunity 6doi: 10.1016/s1074-7613(00)80237-x

Rast JP, Litman GW (1994) T-cell receptor gene homologs are present in the most primitive jawed vertebrates. Proceedings of the National Academy of Sciences of the United States of America 91*:*9248-9252. doi: 10.1073/pnas.91.20.9248

Reinherz EL, Tan K, Tang L, Kern P, Liu J, Xiong Y, Hussey RE, Smolyar A, Hare B, Zhang R, Joachimiak A, Chang HC, Wagner G, Wang J (1999) The crystal structure of a t cell receptor in complex with peptide and mhc class ii. Science 286*:*1913-21. doi: 10.1126/science.286.5446.1913

Reynaud C-A, Anquez V, Grimal H, Weill J-C (1987) A hyperconversion mechanism generates the chicken light chain preimmune repertoire. Cell 48*:*379-388. doi: 10.1016/0092-8674(87)90189-9

Reynaud C-A, Mackay CR, Müller RG, Weill J-C (1991a) Somatic generation of diversity in a mammalian primary lymphoid organ: The sheep ileal peyer's patches. Cell 64*:*995-1005. doi: 10.1016/0092-8674(91)90323-Q

Reynaud CA, Dahan A, Anquez V, Weill JC (1989) Somatic hyperconversion diversifies the single vh gene of the chicken with a high incidence in the d region. Cell 59doi: 10.1016/0092-8674(89)90879-9

Reynaud CA, Garcia C, Hein WR, Weill JC (1995) Hypermutation generating the sheep immunoglobulin repertoire is an antigen-independent process. Cell 80*:*115-125. doi: 10.1016/0092-8674(95)90456-5

Reynaud CA, Mackay CR, Muller RG, Weill JC (1991b) Somatic generation of diversity in a mammalian primary lymphoid organ: The sheep ileal peyer's patches. Cell 64*:*995-1005. doi: 10.1016/0092-8674(91)90323-Q

Reynolds JD, Morris B (1983) The evolution and involution of peyer's patches in fetal and postnatal sheep. Eur J Immunol 13*:*627-35. doi: 10.1002/eji.1830130805

Rios FM, Zimmerman LM (2015) Immunology of reptiles. Els. John Wiley & Sons, Ltd

Rocco L, Liguori I, Costagliola D, Morescalchi MA, Tinti F, Stingo V (2007) Molecular and karyological aspects of batoidea (chondrichthyes, elasmobranchi) phylogeny. Gene 389*:*80-6. doi: 10.1016/j.gene.2006.09.024

Rocco L, Morescalchi MA, Costagliola D, Stingo V (2002) Karyotype and genome characterization in four cartilaginous fishes. Gene 295*:*289-98. doi: 10.1016/s0378-1119(02)00730-8

Rogozin IB, Diaz M (2004) Cutting edge: Dgyw/wrch is a better predictor of mutability at g:C bases in ig hypermutation than the widely accepted rgyw/wrcy motif and probably reflects a two-step activation-induced cytidine deaminase-triggered process. The journal of immunology 172*:*3382-3384. doi: 10.4049/jimmunol.172.6.3382

Rogozin IB, Iyer LM, Liang L, Glazko GV, Liston VG, Pavlov YI, Aravind L, Pancer Z (2007) Evolution and diversification of lamprey antigen receptors: Evidence for involvement of an aid-apobec family cytosine deaminase. Nat Immunol 8*:*647-656. doi: 10.1038/ni1463

Rogozin IB, Pavlov YI, Bebenek K, Matsuda T, Kunkel TA (2001) Somatic mutation hotspots correlate with DNA polymerase η error spectrum. Nature Immunology 2*:*530. doi: 10.1038/88732

Rouaud P, Vincent-Fabert C, Saintamand A, Fiancette R, Marquet M, Robert I, Reina-San-Martin B, Pinaud E, Cogné M, Denizot Y (2013) The igh 3' regulatory region controls somatic hypermutation in germinal center b cells. Journal of Experimental Medicine 210*:*1501-1507. doi: 10.1084/jem.20130072

Rucci F, Cattaneo L, Marrella V, Sacco MG, Sobacchi C, Lucchini F, Nicola S, Bella SD, Villa ML, Imberti L, Gentili F, Montagna C, Tiveron C, Tatangelo L, Facchetti F, Vezzoni P, Villa A (2006) Tissue-specific sensitivity to aid expression in transgenic mouse models. Gene 377*:*150-158. doi: 10.1016/j.gene.2006.03.024

Rumfelt L, McKinney E, Taylor E, Flajnik M (2002) The development of primary and secondary lymphoid tissues in the nurse shark ginglymostoma cirratum: B-cell zones precede dendritic cell immigration and t-cell zone formation during ontogeny of the spleen. Scand J Immunol 56*:*130-148. doi: 10.1046/j.1365-3083.2002.01116.x

Rumfelt LL (2014) Shark reproduction, immune system development and maturation, a review. *In* Smith SL, Sim RB, Flajnik MF (eds.) Immunobiology of the shark. CRC Press LLC, Baton Rouge, UNITED STATES.

Rumfelt LL, Avila D, Diaz M, Bartl S, McKinney EC, Flajnik MF (2001) A shark antibody heavy chain encoded by a nonsomatically rearranged vdj is preferentially expressed in early development and is convergent with mammalian igg. Proceedings of the National Academy of Sciences 98*:*1775-1780. doi: 10.1073/pnas.98.4.1775

Rumfelt LL, Lohr RL, Dooley H, Flajnik MF (2004) Diversity and repertoire of igw and igm vh families in the newborn nurse shark. BMC Immunology 5*:*1-15. doi: 10.1186/1471-2172-5-8

Saha NR, Ota T, Litman GW, Hansen J, Parra Z, Hsu E, Buonocore F, Canapa A, Cheng J-F, Amemiya CT (2014) Genome complexity in the coelacanth is reflected in its adaptive immune system. Journal of Experimental Zoology. Part B, Molecular and Developmental Evolution 322*:*438-463. doi: 10.1002/jez.b.22558

Saha NR, Smith J, Amemiya CT (2010) Evolution of adaptive immune recognition in jawless vertebrates. Seminars in Immunology 22*:*25-33. doi: 10.1016/j.smim.2009.12.002

Saini J, Hershberg U (2015) B cell variable genes have evolved their codon usage to focus the targeted patterns of somatic mutation on the complementarity determining regions. Mol Immunol 65*:*157-67. doi: 10.1016/j.molimm.2015.01.001

Sasaki K, Takada K, Ohte Y, Kondo H, Sorimachi H, Tanaka K, Takahama Y, Murata S (2015) Thymoproteasomes produce unique peptide motifs for positive selection of cd8(+) t cells. Nature Communications 6*:*7484. doi: 10.1038/ncomms8484

Schatz DG (2004) V(d)j recombination. Immunological Reviews 200*:*5-11. doi: 10.1111/j.0105-2896.2004.00173.x

Schatz DG, Oettinger MA, Baltimore D (1989) The v(d)j recombination activating gene, rag-1. Cell 59*:*1035-1048. doi: 10.1016/0092-8674(89)90760-5

Schluter SF, Bernstein RM, Bernstein H, JJ M (1999) 'Big bang' emergence of the combinatorial immune system. Developmental & Comparative Immunology 23*:*107-111. doi: 10.1016/S0145-305X(99)00002-6

Schneider CA, Rasband WS, Eliceiri KW (2012) Nih image to imagej: 25 years of image analysis. Nat Methods 9*:*671-5. doi: 10.1038/nmeth.2089

Sigrist CJA, Cerutti L, de Castro E, Langendijk-Genevaux PS, Bulliard V, Bairoch A, Hulo N (2010) Prosite, a protein domain database for functional characterization and annotation. Nucleic Acids Research 38*:*D161-D166. doi: 10.1093/nar/gkp885

Stingo V, Rocco L (2001) Selachian cytogenetics: A review. Genetica 111*:*329-47. doi: 10.1023/a:1013747215866

Swanson PC, Kumar S, Raval P (2009) Early steps of v(d)j rearrangement: Insights from biochemical studies of rag-rss complexes. *In* Ferrier P (ed.) V(D)J recombination. Springer New York, New York, NY.

Takada K, Van Laethem F, Xing Y, Akane K, Suzuki H, Murata S, Tanaka K, Jameson SC, Singer A, Takahama Y (2015) Tcr affinity for thymoproteasome-dependent positively selecting peptides conditions antigen responsiveness in cd8+ t cells. Nature Immunology 16*:*1069. doi: 10.1038/ni.3237

Takahama Y, Takada K, Murata S, Tanaka K (2012) B5t-containing thymoproteasome: Specific expression in thymic cortical epithelial cells and role in positive selection of cd8+ t cells. Current Opinion in Immunology 24*:*92-98. doi: 10.1016/j.coi.2012.01.006

Telfer JC, Baldwin CL (2015) Bovine gamma delta t cells and the function of gamma delta t cell specific wc1 co-receptors. Cellular Immunology 296*:*76-86. doi: 10.1016/j.cellimm.2015.05.003

Thientosapol ES, Bosnjak D, Durack T, Stevanovski I, van Geldermalsen M, Holst J, Jahan Z, Shepard C, Weninger W, Kim B, Brink R, Jolly CJ (2018) Samhd1 enhances immunoglobulin hypermutation by promoting transversion mutation. Proceedings of the National Academy of Sciences 115*:*4921-4926. doi: 10.1073/pnas.1719771115

Thompson CB, Neiman PE (1987) Somatic diversification of the chicken immunoglobulin light chain gene is limited to the rearranged variable gene segment. Cell 48*:*369-78. doi: 10.1016/0092-8674(87)90188-7

Tonegawa S (1983) Somatic generation of antibody diversity. Nature 302*:*575-581. doi: 10.1038/302575a0

Vaccarelli G, Antonacci R, Tasco G, Yang F, Giordano L, El Ashmaoui HM, Hassanane MS, Massari S, Casadio R, Ciccarese S (2012) Generation of diversity by somatic mutation in the camelus dromedarius t-cell receptor gamma variable domains. Eur J Immunol 42*:*3416-28. doi: 10.1002/eji.201142176

Vaccarelli G, Miccoli MC, Antonacci R, Pesole G, Ciccarese S (2008) Genomic organization and recombinational unit duplication-driven evolution of ovine and bovine t cell receptor gamma loci. BMC Genomics 9*:*81. doi: 10.1186/1471-2164-9-81

Venkatesh B, Lee AP, Ravi V, Maurya AK, Lian MM, Swann JB, Ohta Y, Flajnik MF, Sutoh Y, Kasahara M, Hoon S, Gangu V, Roy SW, Irimia M, Korzh V, Kondrychyn I, Lim ZW, Tay B-H, Tohari S, Kong KW, Ho S, Lorente-Galdos B, Quilez J, Marques-Bonet T, Raney BJ, Ingham PW, Tay A, Hillier LW, Minx P, Boehm T, Wilson RK, Brenner S, Warren WC (2014) Elephant shark genome provides unique insights into gnathostome evolution. Nature 505*:*174-179. doi: 10.1038/nature12826

Vernooij BTM, Lenstra JA, Wang K, Hood L (1993) Organization of the murine t-cell receptor γ locus. Genomics 17*:*566-574. doi: 10.1006/geno.1993.1373

Vitetta E, Berton M, Burger C, Kepron M, Lee Wa, Yin X (1991) Memory b and t cells. Annual Review of Immunology 9*:*193-217. doi: 10.1146/annurev.iy.09.040191.001205

Wagner SD, Milstein C, Neuberger MS (1995) Codon bias targets mutation. Nature 376*:*732-732. doi: 10.1038/376732a0

Wang X, Miller RD (2012) Recombination, transcription, and diversity of a partially germline-joined vh in a mammal. Immunogenetics 64*:*713-7. doi: 10.1007/s00251-012-0627-z

Wang X, Parra ZE, Miller RD (2011) Platypus tcrmu provides insight into the origins and evolution of a uniquely mammalian tcr locus. J Immunol 187*:*5246-54. doi: 10.4049/jimmunol.1101113

Wei L, Chahwan R, Wang S, Wang X, Pham PT, Goodman MF, Bergman A, Scharff MD, MacCarthy T (2015) Overlapping hotspots in cdrs are critical sites for v region diversification. Proc Natl Acad Sci 112*:*E728-37. doi: 10.1073/pnas.1500788112

Weinstein PD, Anderson AO, Maget RG (1994) Rabbit igh sequences in appendix germinal centers: Vh diversification by gene conversion-like and hypermutation mechanisms. Immunity 1*:*647-659. doi: 10.1016/1074-7613(94)90036-1

Winstead CR, Zhai S-K, Sethupathi P, Knight KL (1999) Antigen-induced somatic diversification of rabbit igh genes: Gene conversion and point mutation. The Journal of Immunology 162*:*6602-6612. doi: 162/11/6602

Xiong N, Raulet DH (2007) Development and selection of γδ t cells. Immunological Reviews 215*:*15-31. doi: 10.1111/j.1600-065X.2006.00478.x

Yokosuka T, Takase K, Suzuki M, Nakagawa Y, Taki S, Takahashi H, Fujisawa T, Arase H, Saito T (2002) Predominant role of t cell receptor (tcr)-α chain in forming preimmune tcr repertoire revealed by clonal tcr reconstitution system. The Journal of Experimental Medicine 195*:*991-1001. doi: 10.1084/jem.20010809

Zapata A, Leceta J, Barrutia MG (1981) Ultrastructure of splenic white pulp of the turtle, mauremys caspica. Cell and Tissue Research 220*:*845-855. doi: 10.1007/bf00210466

Zhang C, Du Pasquier L, Hsu E (2013) Shark igw c region diversification through rna processing and isotype switching. Journal of Immunology (Baltimore, Md. : 1950) 191*:*10.4049/jimmunol.1301257. doi: 10.4049/jimmunol.1301257

Zheng B, Xue W, Kelsoe G (1994) Locus-specific somatic hypermutation in germinal centre t cells. Nature 372*:*556-9. doi: 10.1038/372556a0

Zhu C, Hsu E (2010) Error-prone DNA repair activity during somatic hypermutation in shark b lymphocytes. Journal of Immunology (Baltimore, Md. : 1950) 185*:*5336-5347. doi: 10.4049/jimmunol.1000779

Zhu C, Lee V, Finn A, Senger K, Zarrin AA, Du Pasquier L, Hsu E (2012a) Origin of immunoglobulin isotype switching. Curr Biol doi: 10.1016/j.cub.2012.03.060

Zhu C, Lee V, Finn A, Senger K, Zarrin Ali A, Du Pasquier L, Hsu E (2012b) Origin of immunoglobulin isotype switching. Current Biology 22:872-880. doi: 10.1016/j.cub.2012.03.060

Zimmerman LM, Vogel LA, Bowden RM (2010) Understanding the vertebrate immune system: Insights from the reptilian perspective. The Journal of Experimental Biology 213:661-671. doi: 10.1242/jeb.038315

APPENDIX A

SUPPLEMENTAL FIGURES AND TABLES

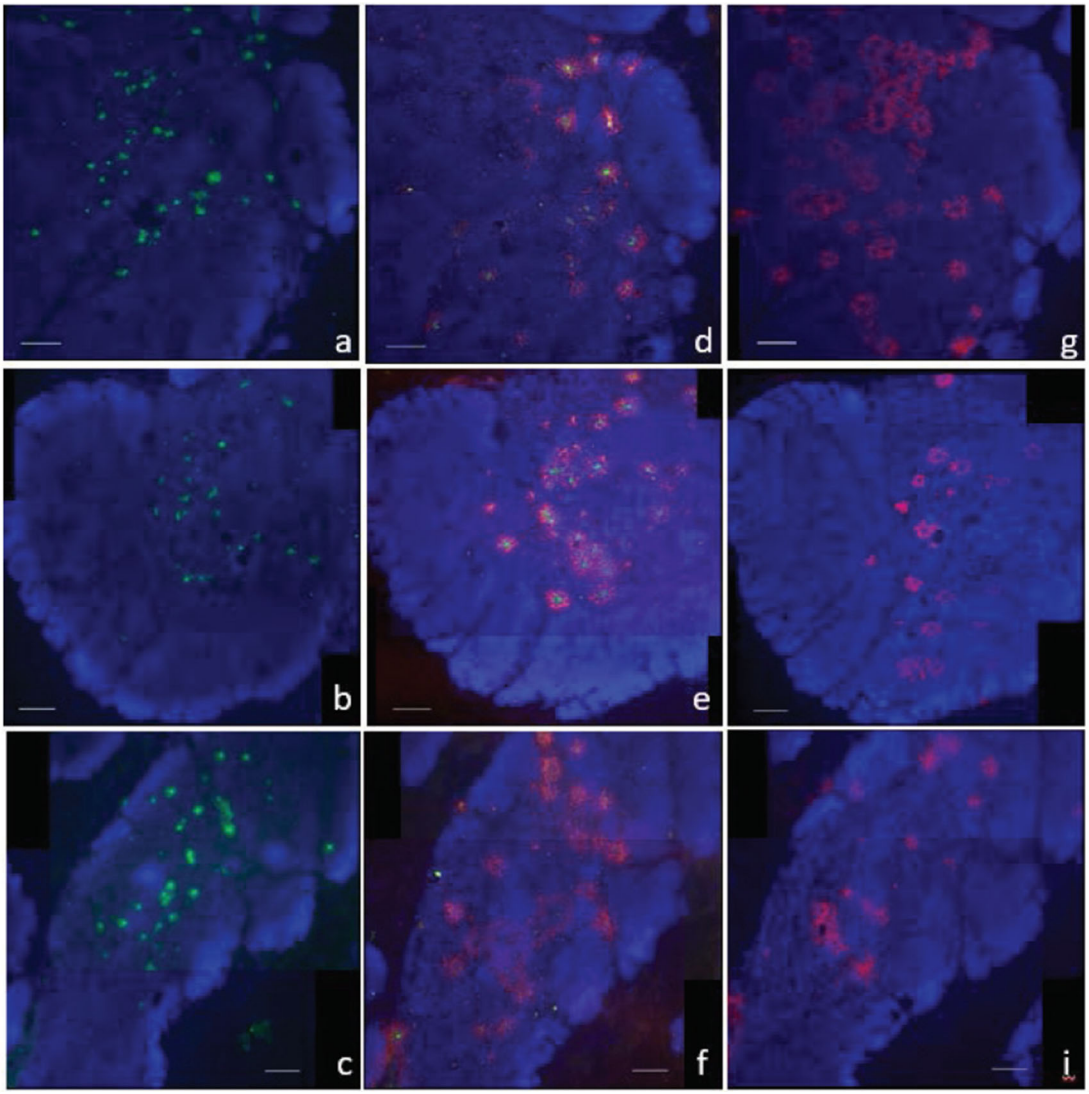

Figure supplement A-1. Localization of AID and TCRα probes is independent. Single molecule RNA fluorescence *in situ* hybridization (FISH) showing three separate regions of fixed thymus sections demonstrating the independent localization of our probes. We probed individually for TCRα (a-c) (probes labeled with CalFluor Red 610; pseudo colored green), simultaneously for both TCRα and AID (d-f), or individually for AID (probes labeled with Quasar 670; pseudo colored red) and counterstained with DAPI (blue). There were two tissue sections in between imaged regions of TCRα section (a-c) and combined AID/TCRα section (d-f). The combined AID/TCRα section (d-f) was consecutive with the AID only section (g-h). We obtained images of each fluorophore using 10x magnification and merged Z-stacked images together. Scale bar 100 mm.

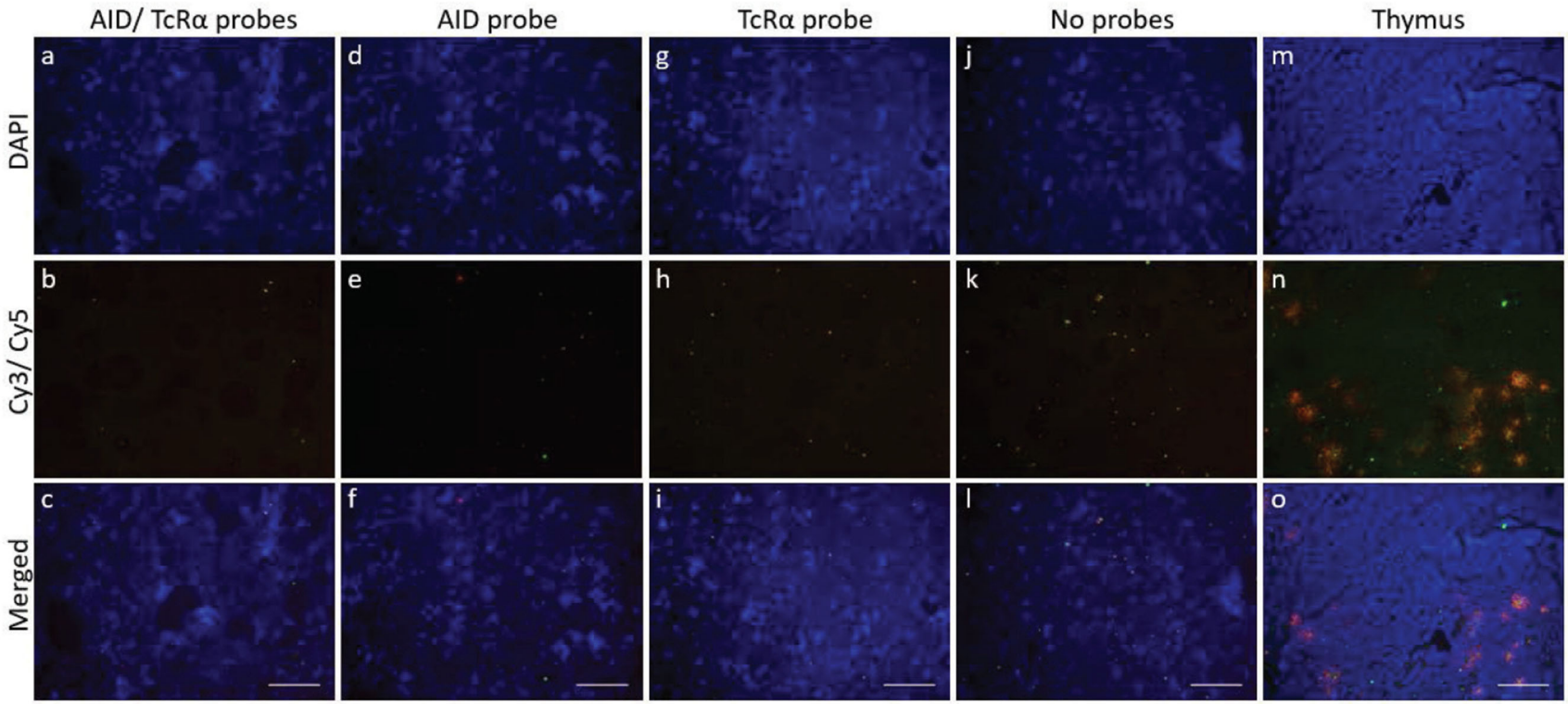

Figure Supplement A-2. Lack of AID and TCRα probe hybridization in shark brain. Single molecule RNA fluorescence *in situ* hybridization (FISH) showing fixed brain sections on four separate slides demonstrating the lack of probe localization in non-immune tissue. We probed simultaneously for both TCRα and AID (a-c), individually for AID (probes labeled with Quasar 670 and imaged with Cy5 filter; pseudo colored red) (d-f), individually for TCRα (probes labeled with CalFluor Red 610 and imaged with Cy3 filter; pseudo colored green) (g-i), or for neither probe (negative control) (j-l). We show probe staining in shark thymus tissue as a positive control [m-o]. We counterstained all slides with DAPI (blue). We imaged fluorescence detected using each filter (DAPI, Cy3/Cy5) regardless of the probe used to illustrate that background fluorescence is not dependent on probe application. We obtained images of each fluorophore using 10x magnification and merged Z-stacked images together (DAPI fluorescence [a,d,g,j,m]; Cy3 (green) and Cy5 (red) fluorescence [b,e,h,k,n]; merged [c,f,I,l,o]). Scale bar 150 mm.

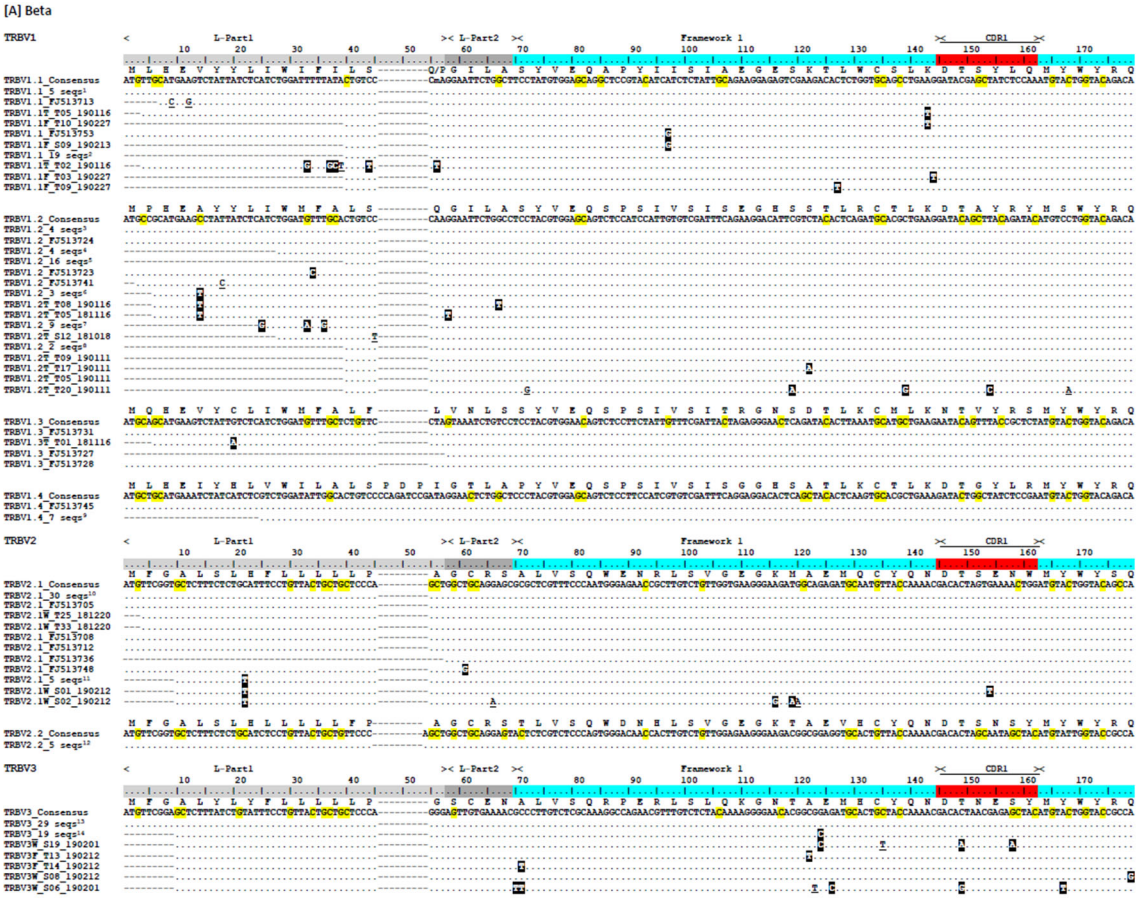

Figure Supplement A-3. Alignments of variable segments from canonical (A) beta (TCRβ V) and (B) gamma (TCRγ V) T cell chains suggest minimal somatic hypermutation. We include amino acids above the nucleotide consensus sequence; dots signify identity to this sequence. Nucleotides highlighted in black indicate non-synonymous changes to the consensus sequence and underlined nucleotides indicate synonymous changes. We denote leader [L-Part1 (light gray) is exon 1; L-Part2 (dark gray) is exon 2], framework (blue), and complementarity-determining regions (CDRs, red) below the scale. Each alignment contains only unique nucleotide sequences, with corroborating clones for lines containing identical bases referenced in Supplemental Figure 5. We used degenerate symbols (lower case) to indicate nucleotide differences likely due to allelic polymorphisms rather than mutation (see text for explanation). Highlighted nucleotides within the consensus sequence represent G or C target nucleotides within AID hotspot motifs (DGYW/WRCH). Highlighted nucleotides (pink) within a transcript mark the location of introns or indels present within a sequence but removed from the alignment (see text for details). Dashes at the 5' or 3' end indicate shortened sequences. Gaps within a sequence are for alignment purposes only. Clone numbers that contain "T" are from thymus, "S" from spleen, "B" from peripheral blood leukocytes, and "V" from spiral valve (intestine). We deposited all sequences into GenBank under accession numbers MN748625-MN748891.

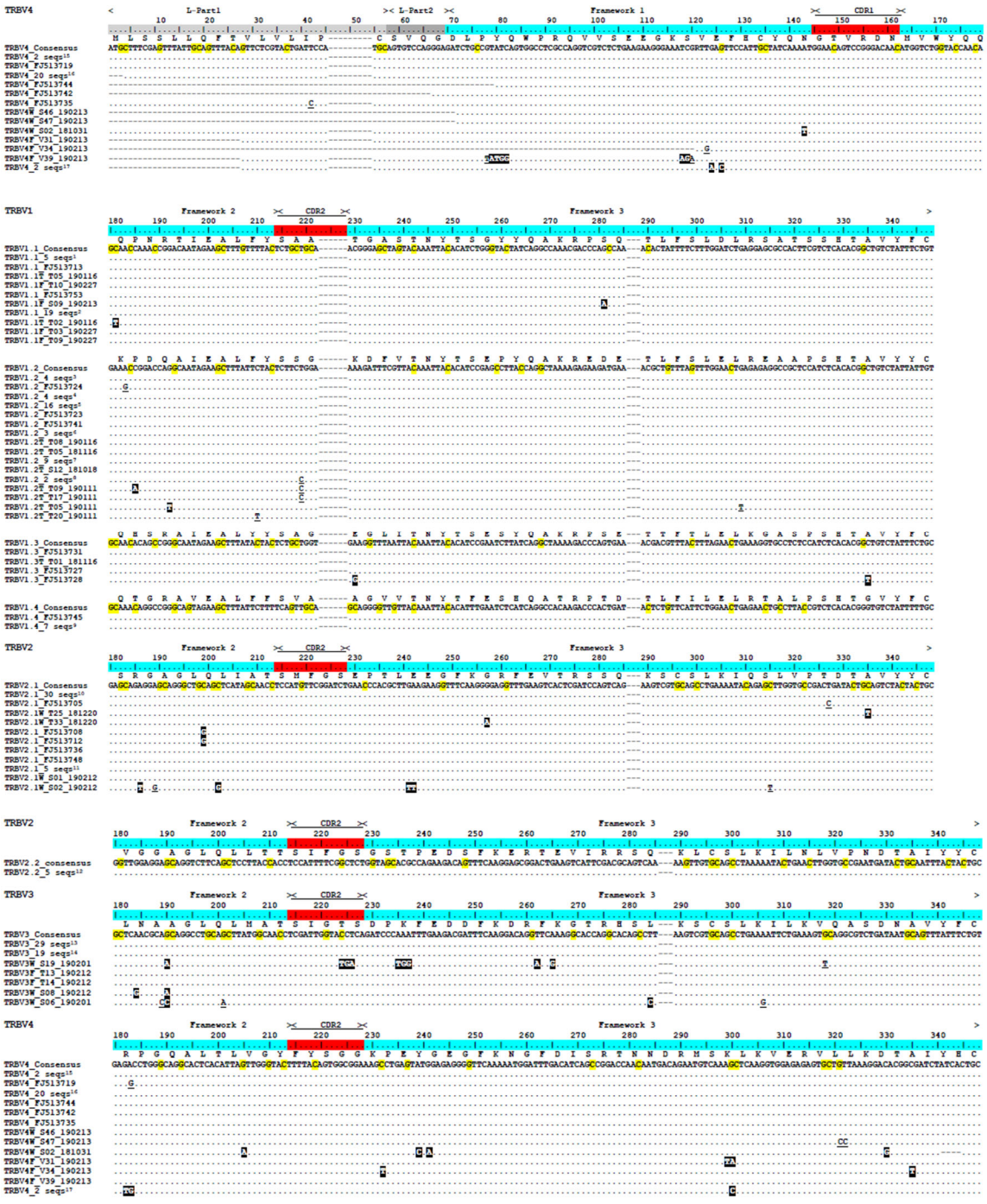

Figure Supplement A-3. (*continued*)

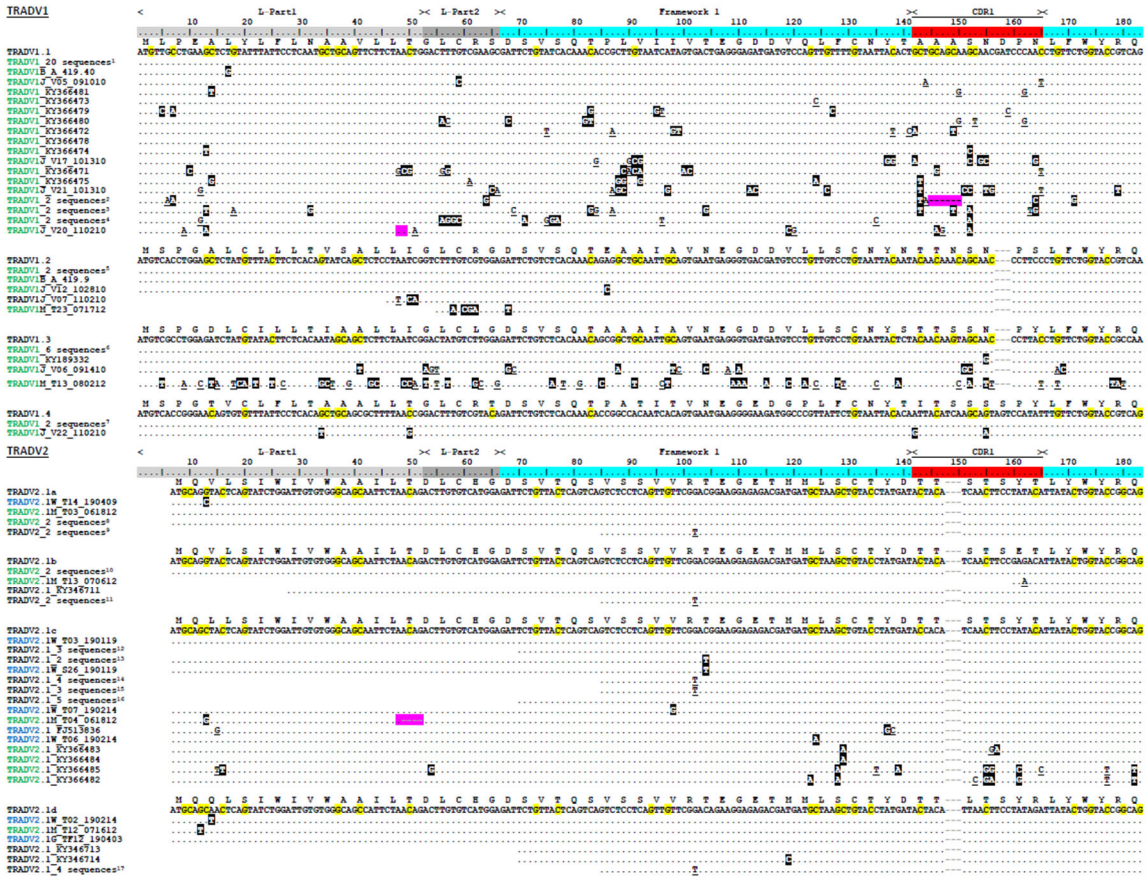

Figure Supplement A-4. Alignment of variable (V) segments from canonical alpha/delta (TCRαδ V) T cell receptor chains suggest somatic hypermutation preferentially impacts Vs associated with alpha constant regions. We include amino acids above the nucleotide consensus sequence; dots signify identity to this sequence. Nucleotides highlighted in black indicate non-synonymous changes to the consensus sequence and underlined nucleotides indicate synonymous changes. We denote leader [L-Part1 (light grey) is exon 1; L-Part2 (dark grey) is exon 2], framework (blue), and complementarity-determining regions (CDRs, red) below the scale. Each alignment contains only unique nucleotide sequences, with corroborating clones for lines containing identical bases referenced in Supplemental Figure 5. To assess mutation rather than changes due to allelic polymorphism, we divided sequences representing each putative allele (see text for details) into separate groups. Highlighted nucleotides within the consensus sequence represent G or C target nucleotides within AID hotspot motifs (DGYW/WRCH). Highlighted nucleotides (pink) mark the location of introns or indels present within a transcript but removed from the alignment (see text for details). Dashes at the 5' or 3' end indicate shortened sequences. Gaps within a sequence are for alignment purposes only. Clone numbers that contain "T" are from thymus, "S" from spleen, "B" from peripheral blood leukocytes, and "V" from spiral valve (intestine). Clones with an asterisk in front of the name are sequences that share less than 70% identity to their group and were not analyzed for mutation. We deposited all sequences into GenBank under accession numbers MN748048-MN748624.

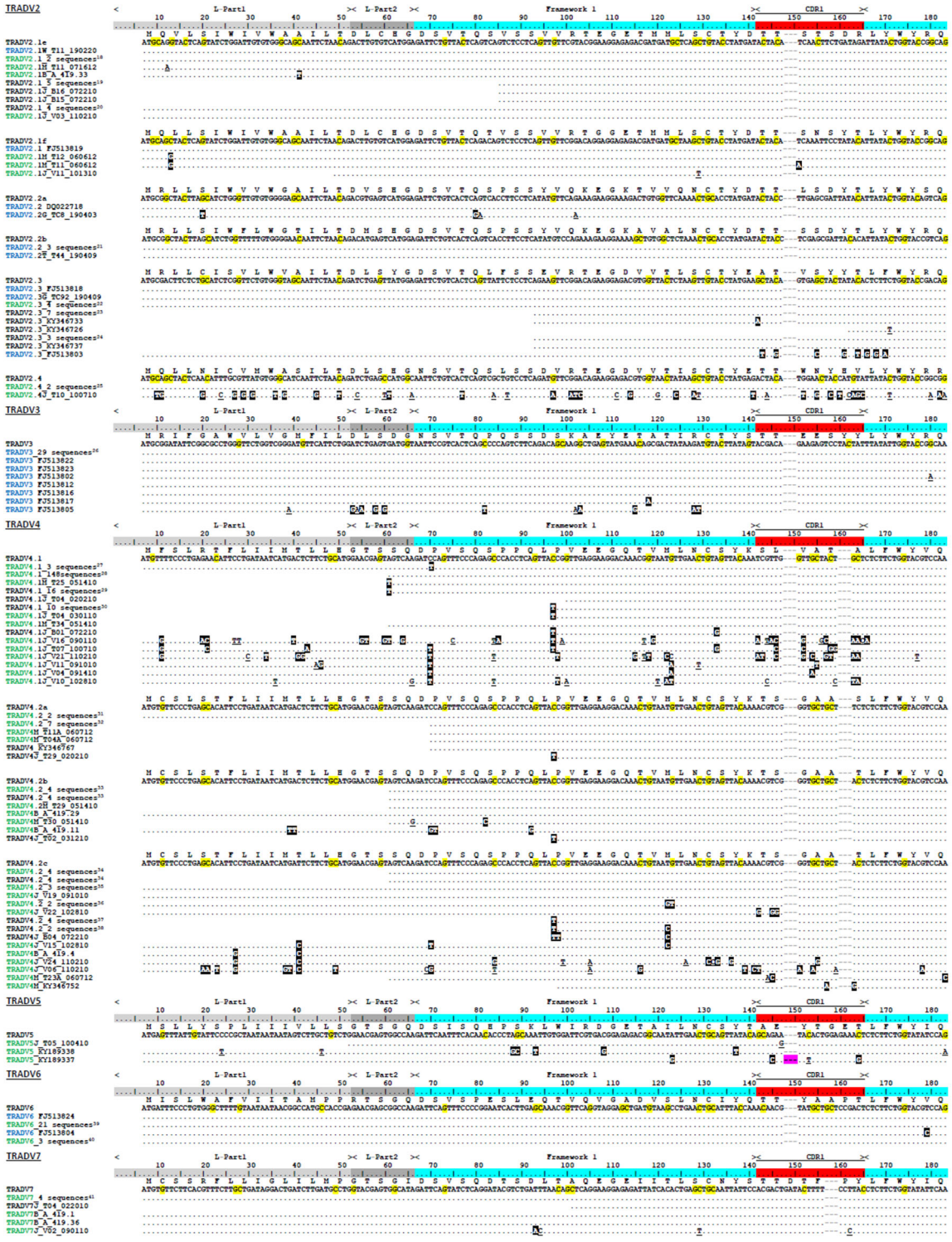

Figure Supplement A-4. (*continued*)

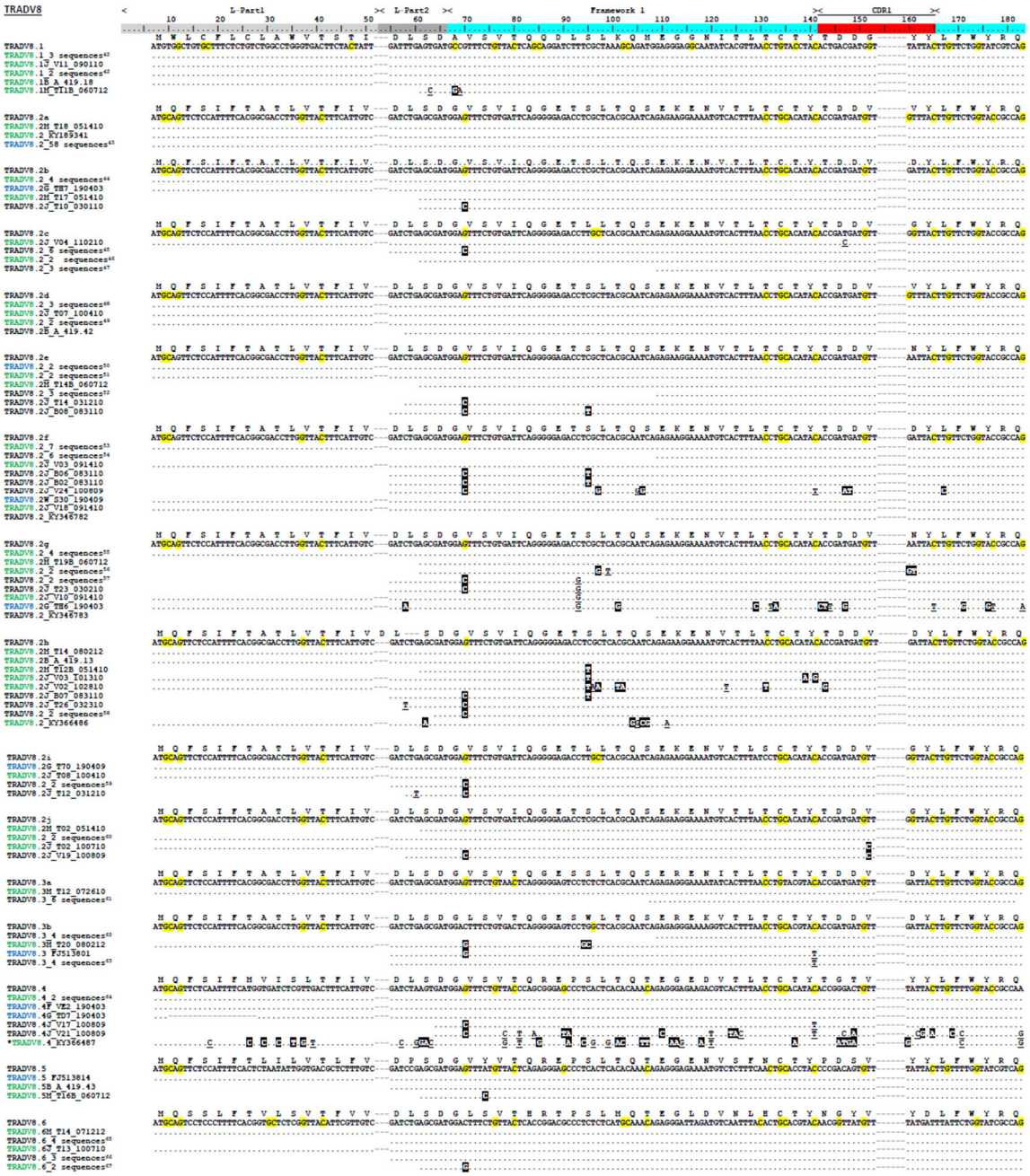

Figure Supplement A-4. (*continued*)

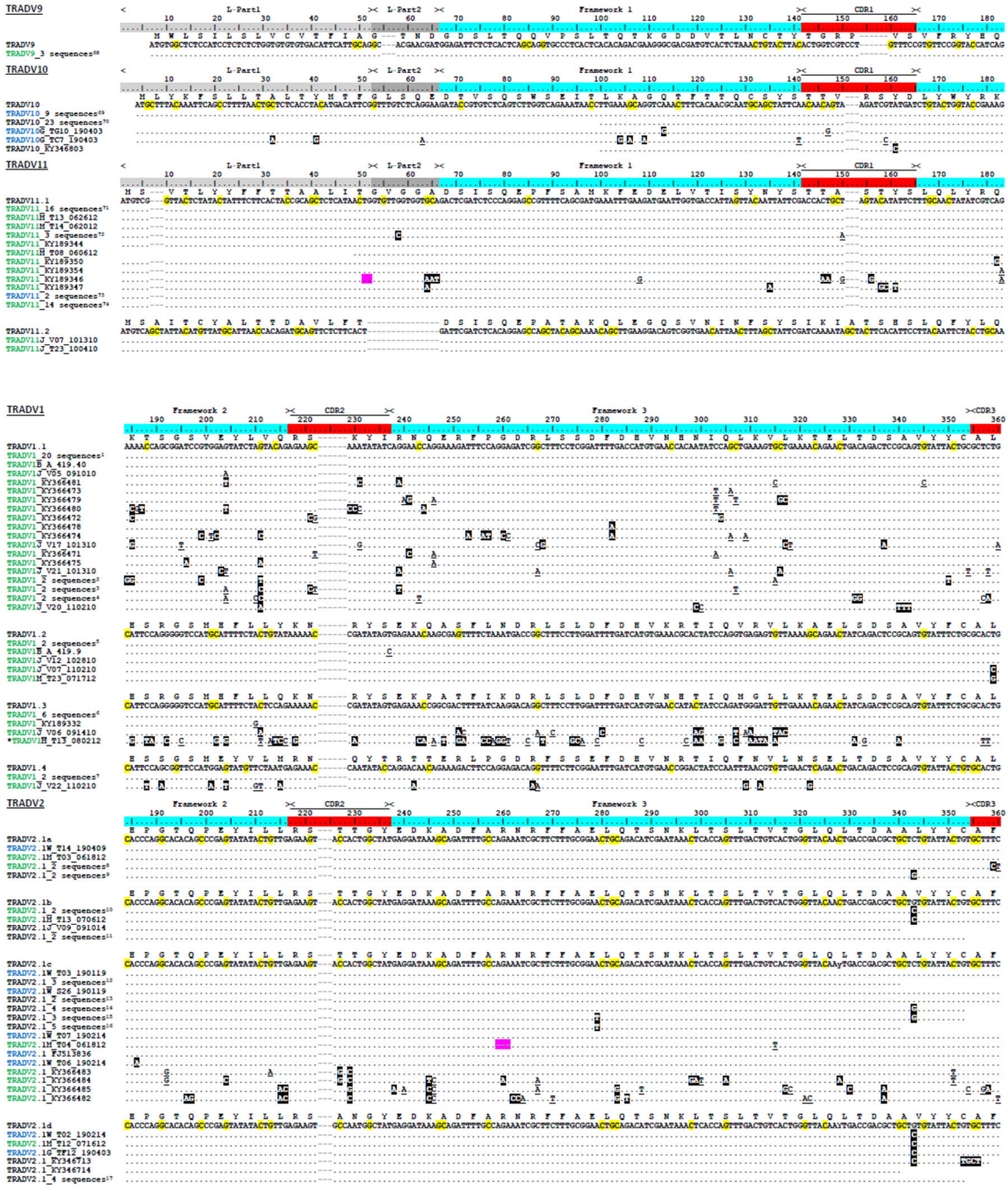

Figure Supplement A-4. (*continued*)

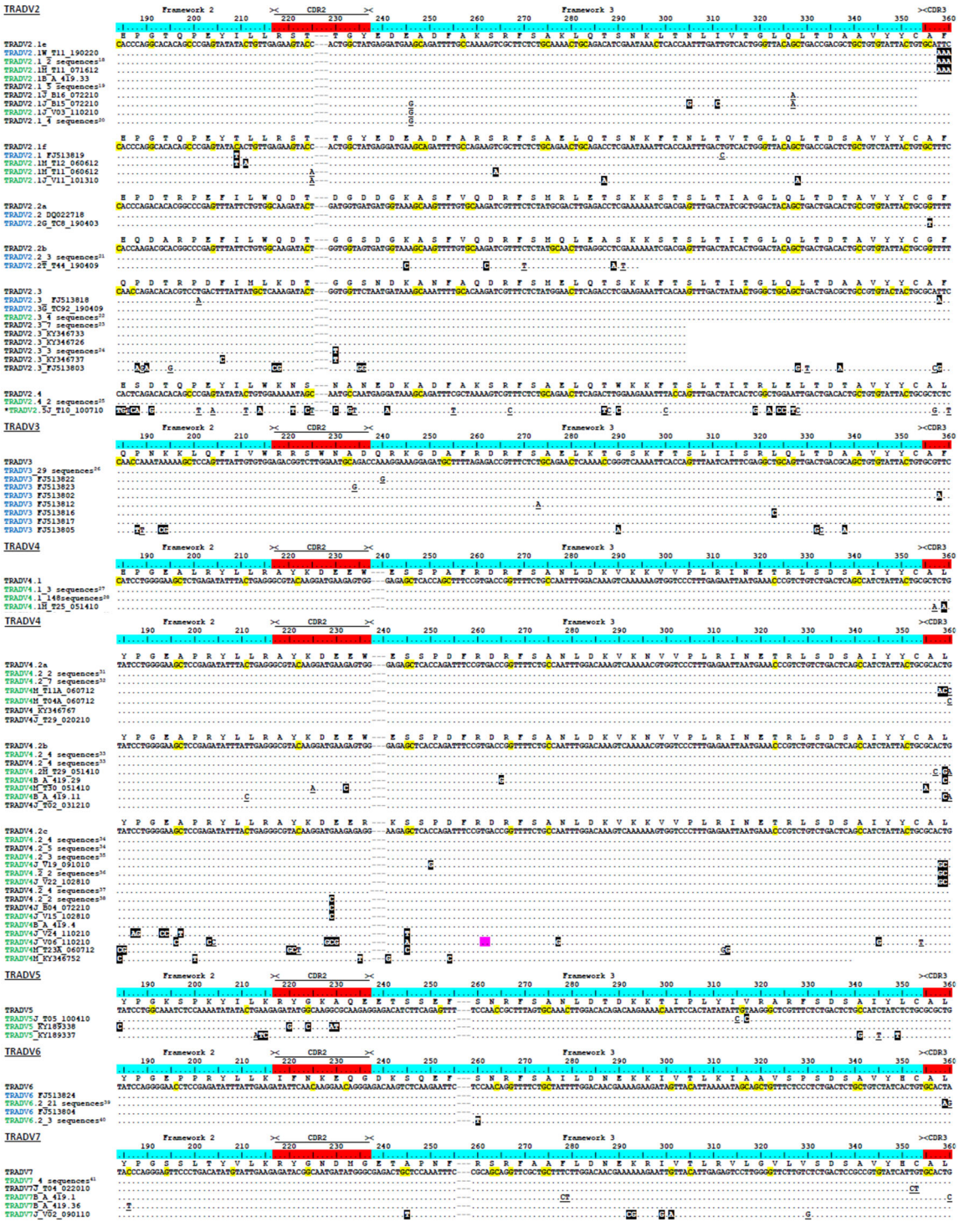

Figure Supplement A-4. (*continued*)

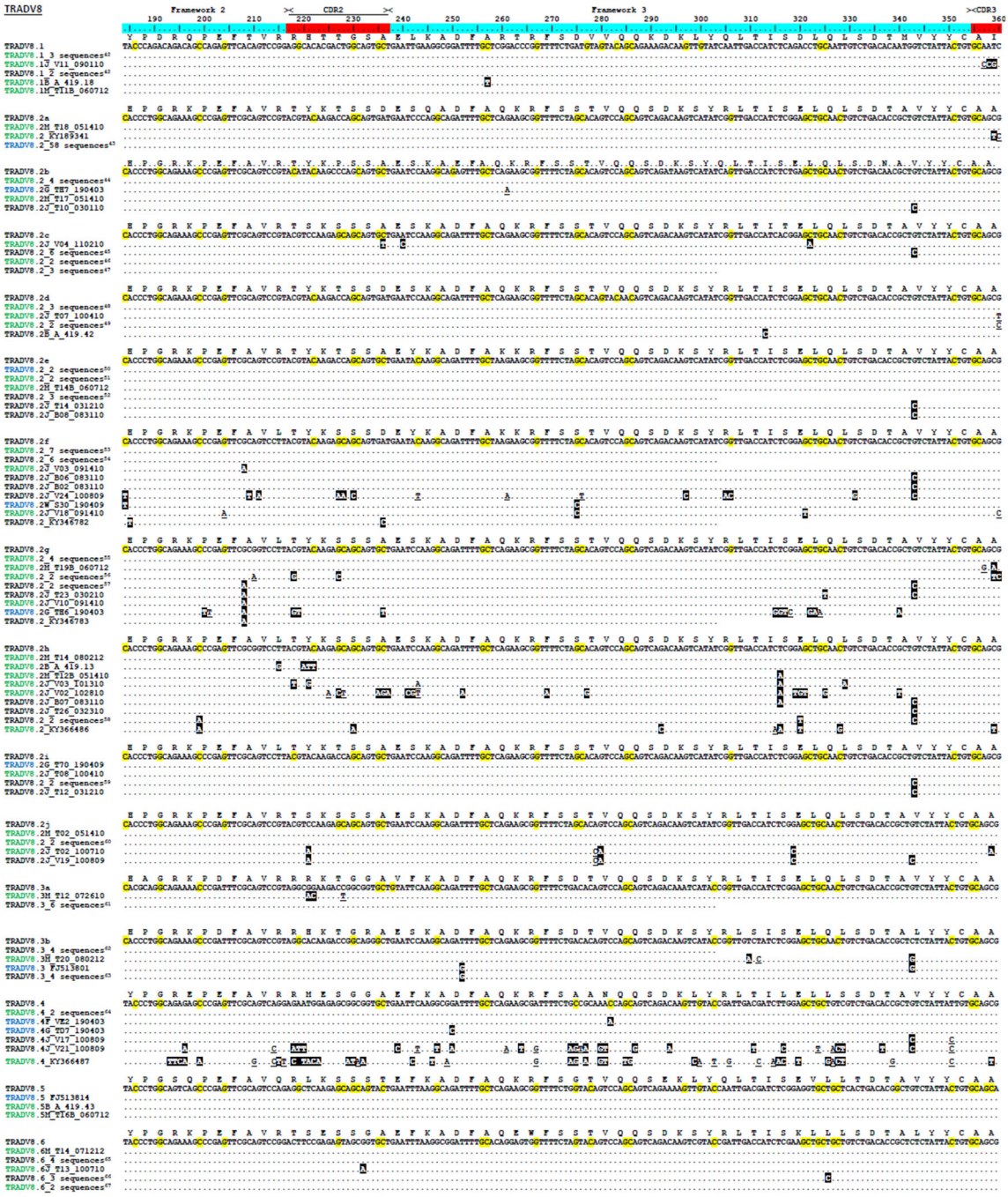

Figure Supplement A-4. (*continued*)

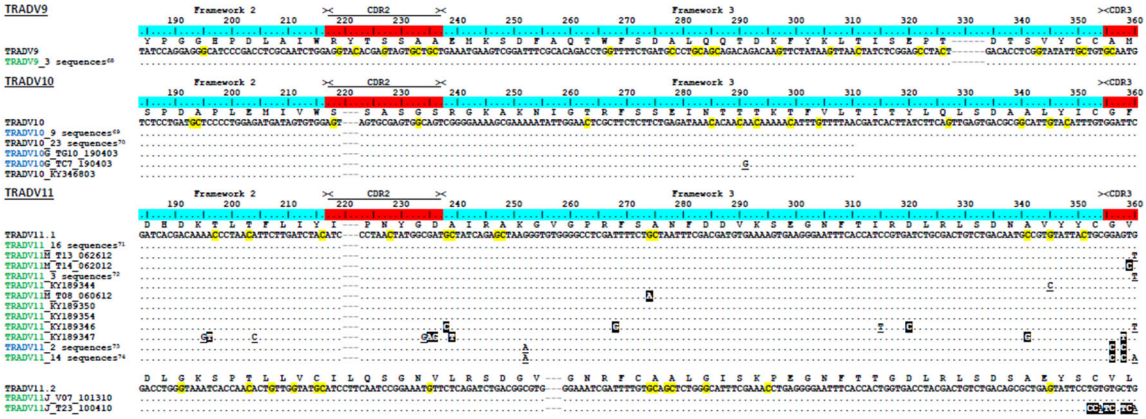

Figure Supplement A-4. (*continued*)

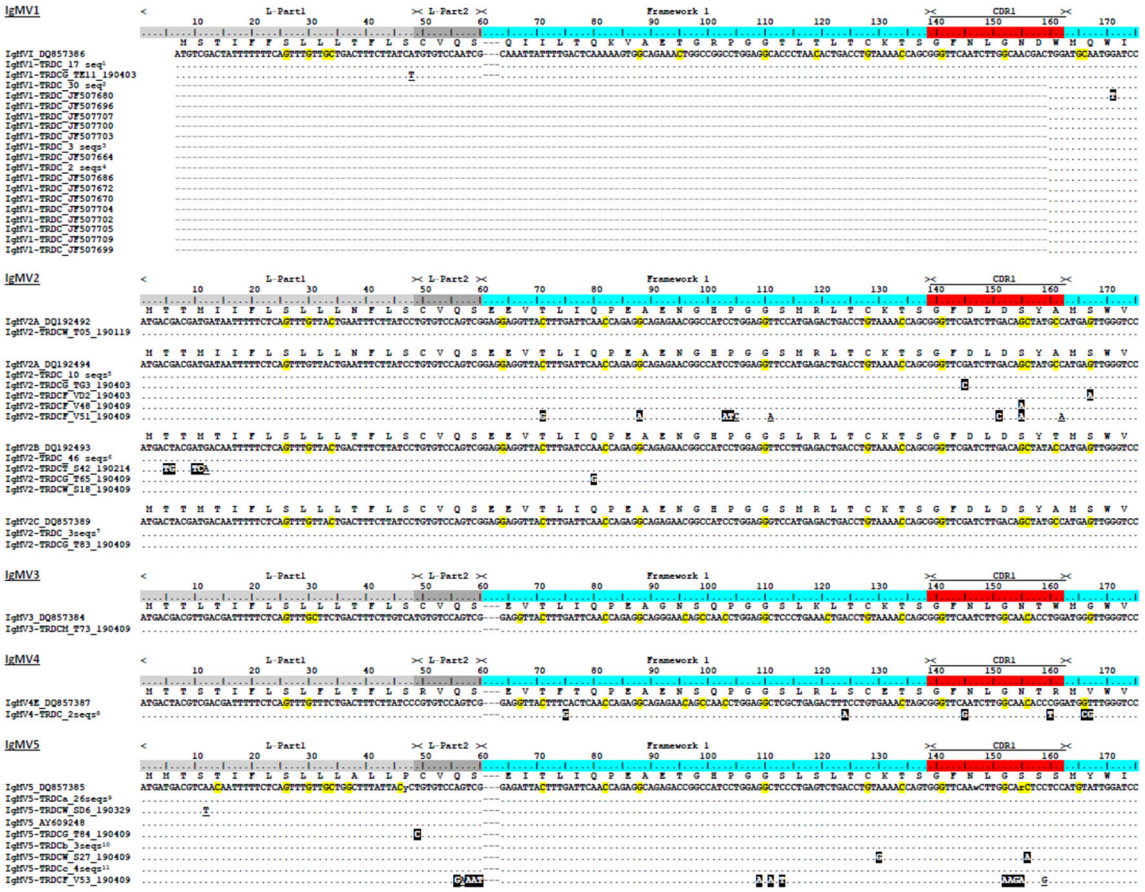

Figure supplement A-5. Alignments of variable segments from "trans" rearrangements between (A) IgM V or (B) IgW V and T cell receptor delta constant (TCRδ C) regions suggest minimal somatic hypermutation. We include amino acids above the germline nucleotide sequence; dots signify identity to this sequence. Nucleotides highlighted in black indicate non-synonymous changes to the germline sequence and underlined nucleotides indicate synonymous changes. We denote leader [L-Part1 (light gray) is exon 1; L-Part2 (dark gray) is exon 2], framework (blue), and complementarity-determining regions (CDRs, red) below the scale. Each alignment contains only unique nucleotide sequences, with corroborating clones for lines containing identical bases referenced in Supplemental Figure 5. We used degenerate symbols (lower case) to indicate nucleotide differences likely due to allelic polymorphisms rather than mutation (see text for explanation). Highlighted nucleotides within the germline sequence represent G or C target nucleotides within AID hotspot motifs (DGYW/WRCH). Highlighted nucleotides (pink) within a transcript mark the location of introns or indels present within a sequence but removed from the alignment (see text for details). Dashes at the 5' or 3' end indicate shortened sequences. Gaps within a sequence are for alignment purposes only. Clone numbers that contain "T" are from thymus, "S" from spleen, "B" from peripheral blood leukocytes, and "V" from spiral valve (intestine). We deposited all sequences into GenBank under accession numbers MN748038-MN748047 (TAILV) and MN788155 - MN788287 (IgHV).

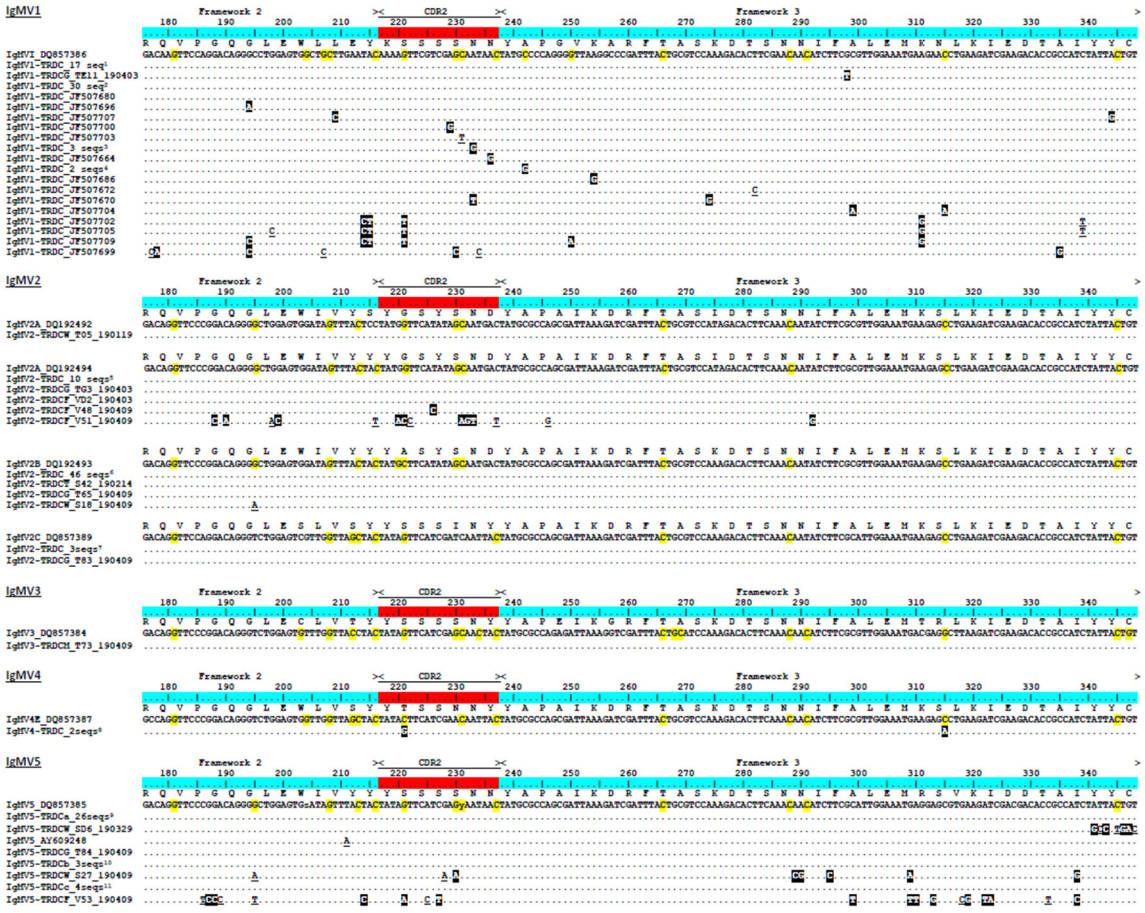

Figure supplement A-5. (*continued*)

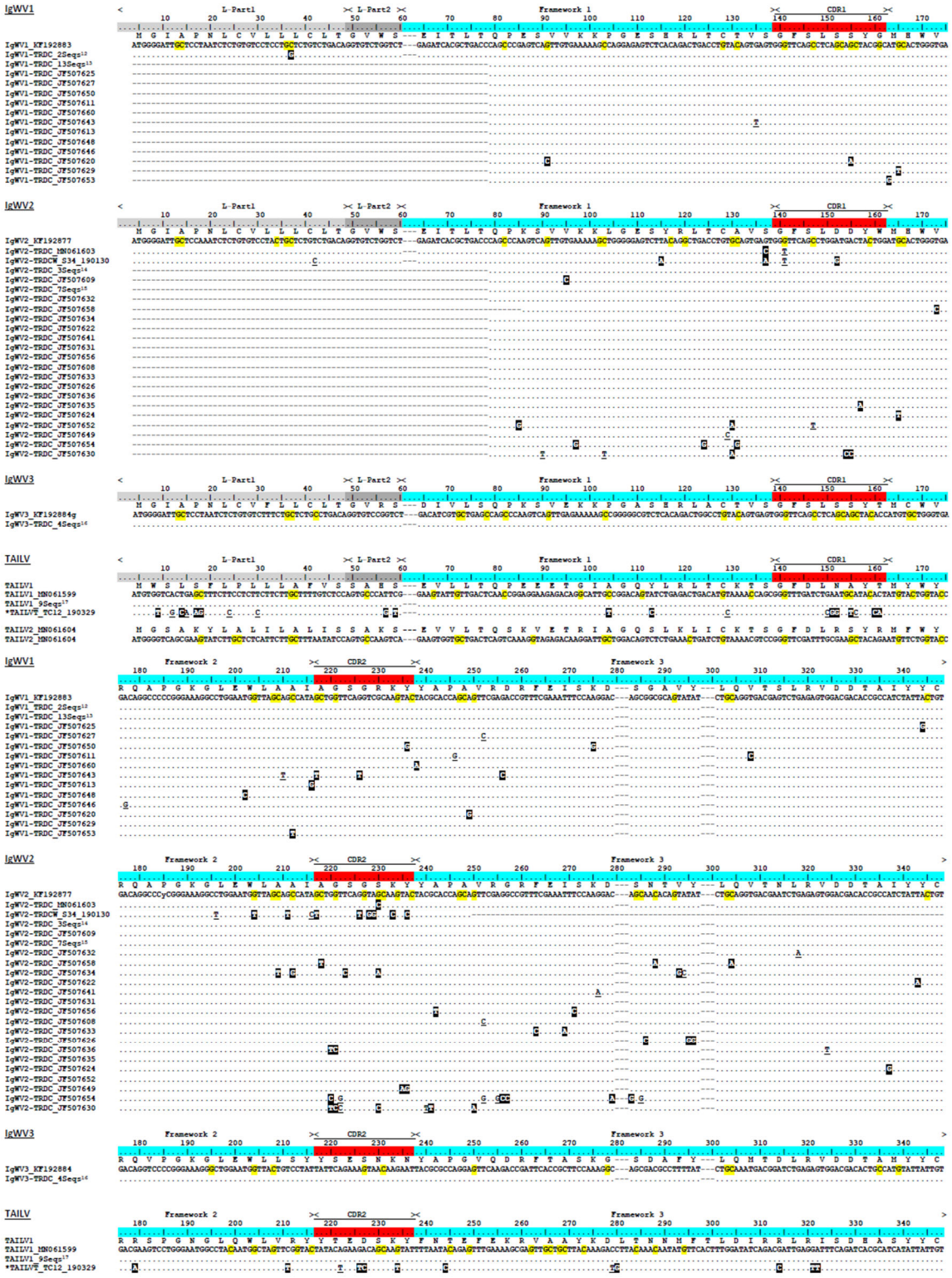

Figure supplement A-5. (*continued*)

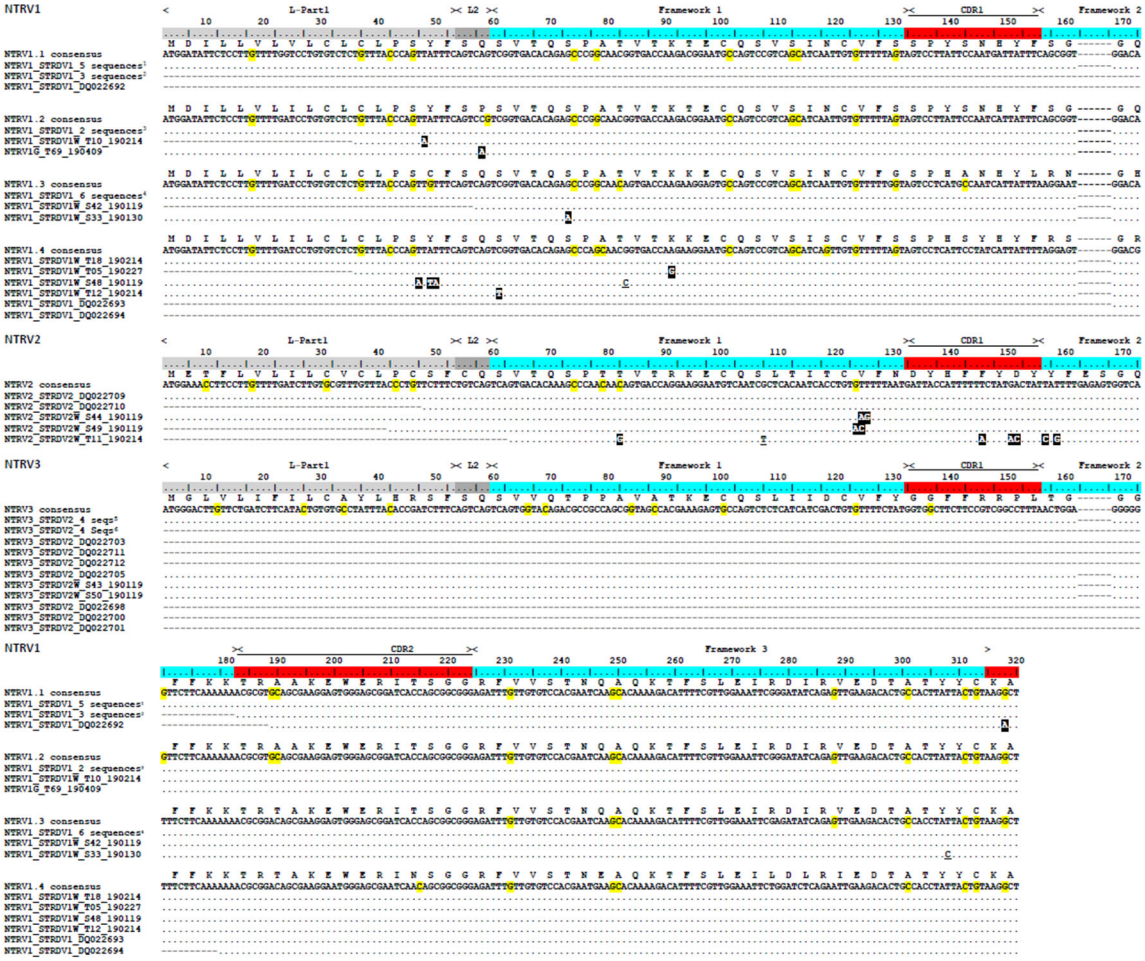

Figure Supplement A-6. Alignments of variable segments from NAR-TCR V (NTCRV) and supporting T cell receptor delta V (STCRδV) suggest minimal mutation. We include amino acids above the nucleotide consensus sequence; dots signify identity to this sequence. Nucleotides highlighted in black indicate non-synonymous changes to the consensus sequence and underlined nucleotides indicate synonymous changes. We denote leader [L-Part1 (light gray) is exon 1; L-Part2 (dark gray) is exon 2], framework (blue), and complementarity-determining regions (CDRs, red) below the scale. Each alignment contains only unique nucleotide sequences, with corroborating clones for lines containing identical bases referenced in Supplemental Figure 5. We used degenerate symbols (lower case) to indicate nucleotide differences likely due to allelic polymorphisms rather than mutation (see text for explanation). Highlighted nucleotides within the consensus sequence represent G or C target nucleotides within AID hotspot motifs (DGYW/WRCH). Dashes at the 5' or 3' end indicate shortened sequences. Gaps within a sequence are for alignment purposes only. Clone numbers that contain "T" are from thymus, "S" from spleen, "B" from peripheral blood leukocytes, and "V" from spiral valve (intestine). We deposited all sequences into GenBank under accession numbers MN748005-MN748037.

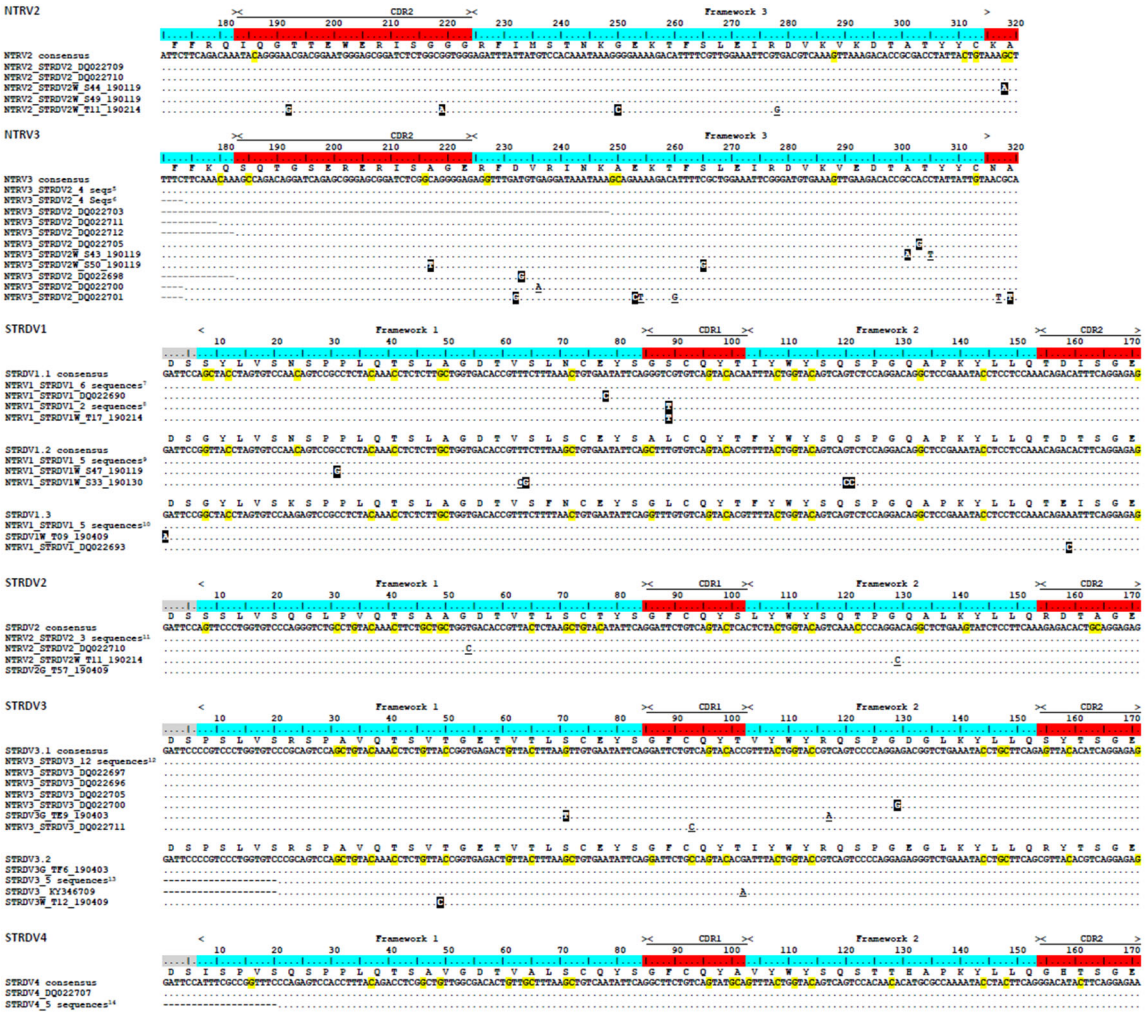

Figure Supplement A-6. (*continued*)

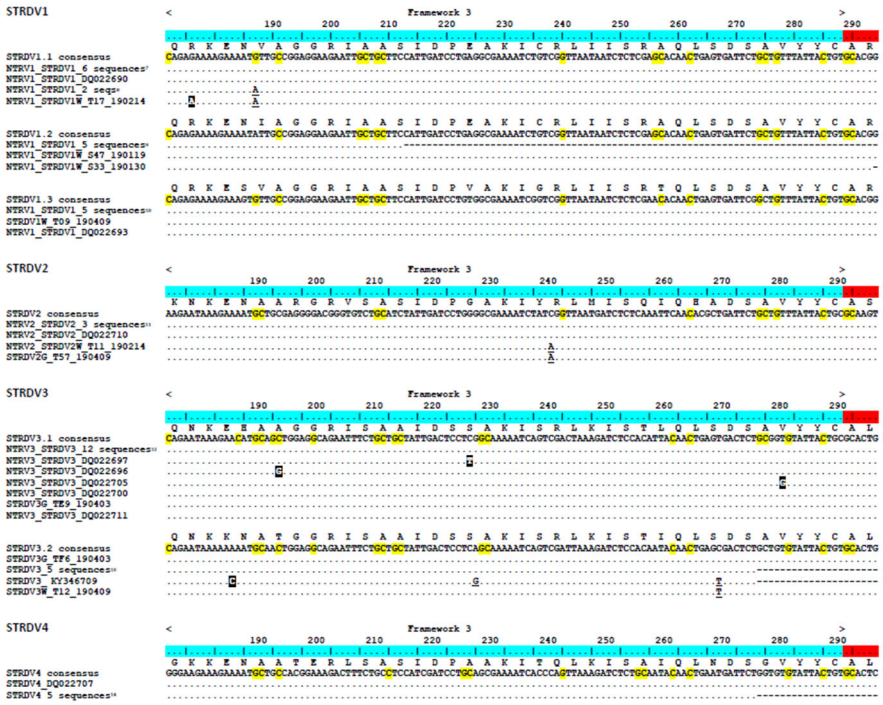

Figure Supplement A-6. (*continued*)

A

1.	TRBV1.1T_FJ513739	TRBV1.1T_T02_181116	TRBV1.1T_T09_190116	TRBV1.1T_T04_181116	TRBV1.1T_T10_190116
2.	TRBV1.1F_SPL02_190213	TRBV1.1T_T15_190111	TRBV1.1F_T17_190213	TRBV1.1T_T06_190111	TRBV1.1T_T07_190111
	TRBV1.1T_T03_190111	TRBV1.1F_T18_190213	TRBV1.1T_T22_190111	TRBV1.1F_T02_190227	TRBV1.1F_T13_190213
	TRBV1.1T_T24_190111	TRBV1.1T_T14_190111	TRBV1.1T_T08_190111	TRBV1.1F_T14_190213	TRBV1.1T_T16_190111
	TRBV1.1T_T21_190111	TRBV1.1T_T19_190111	TRBV1.1T_T10_190111	TRBV1.1F_T19_190213	
3.	TRBV1.1T_FJ513725	TRBV1.1T_FJ513766	TRBV1.1T_FJ513768	TRBV1.1T_FJ513722	
4.	TRBV1.2T_SPL10_181018	TRBV1.2F_SPL29_190320	TRBV1.2F_SPL24_190320	TRBV1.2F_SPL28_190320	
5.	TRBV1.2F_T11_190213	TRBV1.2T_T06_181116	TRBV1.2F_T20_190213	TRBV1.2F_T08_190227	TRBV1.2F_T12_190213
	TRBV1.2F_SPL08_190213	TRBV1.2F_T01_190227	TRBV1.2F_T05_190227	TRBV1.2F_T15_190213	TRBV1.2T_T18_190111
	TRBV1.2T_T11_190111	TRBV1.2T_T12_190111	TRBV1.2F_SPL10_190213	TRBV1.2F_T04_190227	TRBV1.2F_SPL03_190213
	TRBV1.2F_T07_190227				
6.	TRBV1.2T_T04_190116	TRBV1.2T_SPL02_181018	TRBV1.2T_T06_190116		
7.	TRBV1.2W_T21_190213	TRBV1.2W_T12_190227	TRBV1.2W_T17_190227	TRBV1.2W_T28_190213	TRBV1.2W_T25_190213
	TRBV1.2W_T14_190227	TRBV1.2W_T16_190227	TRBV1.2F_V33_190213	TRBV1.2W_T19_19022	TRBV1.2T_T07_190723
	TRBV1.2T_T10_190723	TRBV1.2T_T18_190723	TRBV1.2T_T21_190723		
8.	TRBV1.2T_T02_190111	TRBV1.2T_SPL24_181018			
9.	TRBV1.4W_T15_190227	TRBV1.4W_T27_190213	TRBV1.4W_T18_190227	TRBV1.4W_T20_190227	TRBV1.4W_T23_190213
	TRBV1.4W_T22_190213	TRBV1.4F_T11_190227	TRB4T_T05_190723	TRBV1.4T_T23_190723	
10.	TRBV5W_T01_181220	TRBV2.1W_T04_181220	TRBV2.1W_T15_181220	TRBV2.1W_T28_181220	TRBV2.1W_T31_181220
	TRBV2.1W_T20_181220	TRBV2.1W_T14_181220	TRBV2.1W_T11_181220	TRBV2.1W_T18_181220	TRBV2.1W_T17_181220
	TRBV2.1W_T08_181220	TRBV2.1W_T36_181220	TRBV2.1W_T34_181220	TRBV2.1W_T29_181220	TRBV2.1W_T21_181220
	TRBV2.1W_T19_181220	TRBV2.1W_T30_181220	TRBV2.1W_T16_181220	TRBV2.1W_T13_181220	TRBV2.1W_T26_181220
	TRBV2.1W_T35_181220	TRBV2.1W_T23_181220	TRBV2.1W_T27_181220	TRBV2.1W_T06_181220	TRBV2.1W_T10_181220
	TRBV2.1W_T05_181220	TRBV2.1W_T03_181220	TRBV2.1_FJ513752	TRBV2.1_FJ513711	TRBV2.1_FJ513704
	TRBV2.1G_T22_190719				
11.	TRBV2.1F_T10_190212	TRBV2.1W_SPL35_190212	TRBV2.1W_SPL24_190201	TRBV2.1W_T16_181214	TRBV2.1W_T19_190212
12.	TRBV2.2_FJ513707	TRBV2.2W_T07_181220	TRBV2.2W_T24_181220	TRBV2.2W_T32_181220	TRBV2.2W_T12_181220
13.	TRBV3_FJ513710	TRBV3_FJ513759	TRBV3_FJ513759	TRBV3_FJ513740	TRBV3W_T17_181214
	TRBV3W_T20_181214	TRBV3W_T13_181214	TRBV3F_T12_190201	TRBV3W_T12_181214	TRBV3W_SPL36_190212
	TRBV3F_T16_190212	TRBV3F_T07_190201	TRBV3F_SPL39_190212	TRBV3W_SPL03_190212	TRBV3W_T05_181214
	TRBV3F_T10_190201	TRBV3W_T08_181214	TRBV3F_T15_190212	TRBV3F_T11_190212	TRBV3W_T21_190212
	TRBV3F_SPL44_190212	TRBV3F_T12_190212	TRBV3F_T17_190212	TRBV3W_T18_181214	TRBV3T_T11_190207
	TRBV3T_T04_190207	TRBV3F_T08_190201	TRBV3F_SPL29_190201	TRBV3F_T09_190201	TRBV3G_T36_190719
14.	TRBV3W_T26_190212	TRBV3T_T02_190207	TRBV3T_T05_190207	TRBV3T_T07_190207	TRBV3T_T08_190207
	TRBV3T_T09_190207	TRBV3T_T10_190207	TRBV3W_SPL23_190201	TRBV3W_T13_190201	TRBV3W_SPL34_190212
	TRBV3W_T21_181214	TRBV3W_T22_181214	TRBV3W_T41_190201	TRBV3W_T09_181214	TRBV3W_T18_190201
	TRBV3W_T25_190212	TRBV3W_T11_181214	TRBV3W_SPL22_190201	TRBV3W_T10_181214	
15.	TRBV4_FJ513737	TRBV4_FJ513732	TRBV4G_T47_190710	TRBV4G_T36_190710	
16.	TRBV4T_T31_190111	TRBV4W_SPL12_181031	TRBV4T_T37_190111	TRBV4T_T41_190111	TRBV4T_T38_190111
	TRBV4T_T40_190111	TRBV4T_T12_181116	TRBV4T_T35_190111	TRBV4T_T43_190111	TRBV4T_T42_190111
	TRBV4T_T44_190111	TRBV4T_T36_190111	TRBV4F_V36_190213	TRBV4F_V21_190227	TRBV4F_V28_190227
	TRBV4F_V24_190227	TRBV4W_SPL41_190213	TRBV4F_V27_190227	TRBV4W_SPL44_190213	TRBV4F_V22_190227
17.	TRBV4F_V23_190227		TRBV4F_V32_190213		

Figure Supplement A-7. Clone sequences contained within footnotes of Supplemental Figures 1 through 4. Clones listed in a single footnote contain identical nucleotide sequences for T cell receptor (TCR) variable segments (V) of **[A]** Beta (TCRβ V) and **[B]** Gamma (TCRγ V) [Supplemental Figure 1], **[C]** Alpha/Delta (TCRαδ V) [Supplemental Figure 2], **[D]** NAR-TCR (NTCRV-STCRδV) [Supplemental Figure 3], or **[E]** IgHV-TCRδC "trans"-rearrangements (IgM-TCRδC) and TAILV [Supplemental Figure 4]. For TCRαδ V sequences, sequence colors indicate the constant (C) regions utilized by the segment (green=alpha C; blue=delta C; black=incomplete/no C).

B

18.	TRGV1.1a FJ513776	TRGV1.1a FJ513777	TRGV1.1a FJ513781	TRGV1.1a FJ513783	TRGV1.1a FJ513782
	TRGV1.1aF_S36_190227	TRGV1.1aF_S30_190227	TRGV1.1aJ_T01_150408	TRGV1.1aF_S04_190307	TRGV1.1bG_T32_190430
	TRGV1.1bW_S05_181210				
19.	TRGV1.1b FJ513789	TRGV1.1bG_T30_190430	TRGV1.1b FJ513790	TRGV1.1b_KY351639	
20.	TRGV1.2a_FJ513775	TRGV1.2a_FJ513774	TRGV1.2aG_T33_190430	TRGV1.2aW_T47_190423	
21.	TRGV1.3J_B05_150824	TRGV1.3J_T12_150415			
22.	TRGV1.4aJ_B06_150824	TRGV1.4aJ_B08_150824	TRGV1.4aJ_T02_150420	TRGV1.4aJ_T11_150415	
23.	TRGV1.4aG_T43_190430	TRGV1.4aG_T46_190430			
24.	TRGV1.4bW_T43_190423	TRGV1.4bF_S35_190227	TRGV1.4bW_T46_190423	TRGV1.4b FJ513786	TRGV1.4b FJ513787
25.	TRGV1.4bJ_B10_150824	TRGV1.4bJ_T09_150415			
26.	TRGV1.5T_S10_190424	TRGV1.5G_T39_190430	TRGV1.5F_S33_190227	TRGV1.5F_S39_190227	TRGV1.5W_S32_181130
	TRGV1.5F_S38_190227	TRGV1.5 FJ513791	TRGV1.5 FJ513779	TRGV1.5 FJ513784	
27.	TRGV1.6J_B07_150824	TRGV1.6J_B09_150824	TRGV1.6J_B09_150831	TRGV1.6J_B11_150831	TRGV1.6J_B13_150831
	TRGV1.6J_B15_150831				
28.	TRGV2.1 FJ513778	TRGV2.1T_S15_190424	TRGV2.1T_S14_190424	TRGV2.1T_S18_190424	TRGV2.1W_S12_181130
	TRGV2.1W_S08_181130				
29.	TRGV2.2G_T25_190430	TRGV2.2W_T31_190425			
30.	TRGV2.2W_T34_190425	TRGV2.2G_T29_190430	TRGV2.2T_S16_190424	TRGV2.2F_V46_190425	
31.	TRGV3.1W_T32_190307	TRGV3.1G_T48_190430			
32.	TRGV3.1J_T12_150415	TRGV3.1J_T14_150415	TRGV3.1J_T15_150415	TRGV3.1J_V02_151014	TRGV3.1J_V06_151014
	TRGV3.1J_T05_150408	TRGV3.1J_B13_150625	TRGV3.1J_B18_150722	TRGV3.1J_B22_150722	TRGV3.1J_T01_150514
	TRGV3.1J_T02_150514	TRGV3.1J_T03_150520	TRGV3.1J_T04_150520		
33.	TRGV3.2aJ_V04_151014	TRGV3.2aJ_T08_150408	TRGV3.2aJ_B19_150722	TRGV3.2aJ_B20_150625	TRGV3.2aJ_B21_150722
	TRGV3.2aJ_T04_150514	TRGV3.2aJ_B23_150722	TRGV3.2aJ_B17_150625	TRGV3.2aJ_B16_150625	TRGV3.2aJ_B14_150625
	TRGV3.2aJ_B15_150625				
34.	TRGV3.2bT_T28_190604	TRGV3.2bT_S37_190425	TRGV3.2bT_S38_190425	TRGV3.2bT_S42_190425	TRGV3.2bF_T44_190227
	TRGV3.2bF_T42_190227	TRGV3.2b FJ513792	TRGV3.2b FJ513793		
35.	TRGV3.3F_S31_190227	TRGV3.3J_V03_151014	TRGV3.3J_T07_150408	TRGV3.3J_T11_150420	TRGV3.3J_B17_150722
	TRGV3.3J_B19_150625				
36.	TRGV3.4J_V05_151014	TRGV3.4J_V07_151014	TRGV3.4J_V08_151014	TRGV3.4J_T09_150420	TRGV3.4J_T16_150415
	TRGV3.4J_T01_150520				
37.	TRGV4W_T41_190423	TRGV4W_T39_190423			
38.	TRGV4T_S23_190424	TRG V4 FJ513785			
39.	TRGV4J_B11_150824	TRGV4J_B17_150831	TRGV4J_B18_150831	TRGV4J_B19_150831	TRGV4J_B20_150831
	TRGV4J_B22_150831	TRGV4J_B23_150831	TRGV4J_B24_150831	TRGV4J_T02_150621	TRGV4J_T05_150621
	TRGV4J_T06_150621	TRGV4J_T08_150621	TRGV4J_T10_150621	TRGV4J_T04_150621	

Figure Supplement A-7. (*continued*)

C

1.	TRADV1J_V13_101310	TRADV1M_T08_060512	TRADV1J_V02_110210	TRADV1M_T13_072610	TRADV1J_V08_102810
	TRADV1J_T17_100710	TRADV1J_V09_110210	TRADV1M_T02_060512	TRADV1M_T18_072610	TRADV1M_T17_072610
	TRADV1M_T14_080912	TRADV1M_T13_060512	TRADV1J_V07_091410	TRADV1J_V18_090110	TRADV1M_T14_060512
	TRADV1M_T23_071212	TRADV1M_T21_072012	TRADV1J_V19_110210	TRADV1J_V05_101310	TRADV1M_T17_071212
2.	TRADV1J_V14_101310	TRADV1J_V16_101310			
3.	TRADV1_KY366476	TRADV1_KY366469			
4.	TRADV1_KY366470	TRADV1_KY366477			
5.	TRADV1M_T24_080212	TRADV1M_T11_070612			
6.	TRADV1B_A419.3	TRADV1J_T43_100410	TRADV1_KY189333	TRADV1_KY189334	TRADV1_KY189335
	TRADV1_KY189336				
7.	TRADV1M_T04_071212	TRADV1M_T11_072012			
8.	TRADV2M_T07_061812	TRADV2M_T09_061812			
9.	TRADV2J_B07_062512	TRADV2J_B13_072210			
10.	TRADV2M_T04_062512	TRADV2M_T09_062512			
11.	TRADV2J_T05_031210	TRADV2J_T08_032310			
12.	TRADV2_KY346724	TRADV2_KY346719	TRADV2_KY346720		
13.	TRADV2_KY346716	TRADV2_KY346717			
14.	TRADV2J_T07_030110	TRADV2J_B09_062512	TRADV2J_B14_072210	TRADV2J_T08_022510	
15.	TRADV2J_B10_062512	TRADV2J_B12_072210	TRADV2J_B12_062512		
16.	TRADV2_KY346715	TRADV2_KY346712	TRADV2_KY346718	TRADV2_KY346721	TRADV2_KY346723
17.	TRADV2J_T08_030110	TRADV2J_B06_062512	TRADV2J_B11_062512	TRADV2J_T05_022010	
18.	TRADV2M_T09_080212	TRADV2M_T02_070612			
19.	TRADV2J_T06_030110	TRADV2J_B11_072210	TRADV2J_B10_072210	TRADV2J_T06_031210	TRADV2J_B05_062512
20.	TRADV2J_B08_062512	TRADV2J_T05_022510	TRADV2J_B09_072210	TRADV2J_T07_022510	
21.	TRADV2T_T37_190409	TRADV2G_T78_190409	TRADV2W_S26_190409		
22.	TRADV2J_V12_091410	TRADV2M_T06_071212	TRADV2M_T24_071612	TRADV2M_T02_072612	
23.	TRADV2_KY346728	TRADV2_KY346729	TRADV2_KY346730	TRADV2_KY346731	TRADV2_KY346734
	TRADV2_KY346735	TRADV2_KY346736			
24.	TRADV2_KY346725	TRADV2_KY346727	TRADV2_KY346732		
25.	TRADV2M_T24_080912	TRADV2J_T14_100710			
26.	TRADV3F_V72_190308	TRADV3T_S34_190220	TRADV3W_T22_190220	TRADV3W_T18_190220	TRADV3W_S31_190119
	TRADV3T_S46_190214	TRADV3T_S48_190214	TRADV3T_S49_190214	TRADV3T_S50_190214	TRADV3T_S51_190214
	TRADV3T_S52_190214	TRADV3T_S54_190214	TRADV3T_S38_190214	TRADV3T_S39_190214	TRADV3T_S45_190214
	TRADV3T_S61_190214	TRADV3T_S68_190214	TRADV3T_S70_190214	TRADV3W_T19_190214	TRADV3W_T21_190214
	TRADV3W_T25_190214	TRADV3W_T26_190214	TRADV3G_TC91_190409	TRADV3W_T13_190409	TRADV3T_T43_190409
	TRADV3T_S34_190214	TRADV3W_S31_190409	TRADV3W_S29_190119	TRADV3 FJ513808	
27.	TRADV4B_A_419.12	TRADV4J_T11_100710	TRADV4J_T14_100410		
28.	TRADV4M_T33_051410	TRADV4M_T15A_060712	TRADV4M_T02A_060712	TRADV4M_T22_051410	TRADV4M_T10A_060712
	TRADV4M_T18A_060712	TRADV4M_T12_071712	TRADV4M_T11_071712	TRADV4M_T07_062012	TRADV4M_T21_062012
	TRADV4M_T13_062512	TRADV4M_T17_062512	TRADV4M_T22_071712	TRADV4M_T18_062012	TRADV4M_T06_062612
	TRADV4M_T11_060512	TRADV4M_T24_052410	TRADV4M_T14_070912	TRADV4M_T03_052410	TRADV4M_T15_071712
	TRADV4M_T21_062512	TRADV4M_T21_071712	TRADV4M_T03_062612	TRADV4M_T15_062012	TRADV4M_T02_071712
	TRADV4M_T03_071712	TRADV4M_T21_061812	TRADV4M_T16_071712	TRADV4M_T22B_060712	TRADV4M_T11_062012
	TRADV4M_T05_062612	TRADV4M_T21_052410	TRADV4M_T28_052410	TRADV4M_T22A_060712	TRADV4M_T16_062612
	TRADV4M_T08_061812	TRADV4M_T19_062612	TRADV4M_T24A_060712	TRADV4M_T08_062012	TRADV4M_T11_052410
	TRADV4M_T20_052410	TRADV4M_T13_062012	TRADV4M_T14_062612	TRADV4M_T04_070912	TRADV4M_T15_061812
	TRADV4M_T15_062512	TRADV4M_T06_071712	TRADV4M_T04_071712	TRADV4M_T18_071712	TRADV4M_T05_052410
	TRADV4M_T03_062512	TRADV4M_T09_071712	TRADV4M_T01_061812	TRADV4M_T01_062012	TRADV4M_T16_061812
	TRADV4M_T10_062612	TRADV4M_T12_062612	TRADV4M_T18_061812	TRADV4M_T21_060512	TRADV4M_T04_062012
	TRADV4M_T06_061812	TRADV4M_T09_052410	TRADV4M_T15_060512	TRADV4M_T03_060512	TRADV4M_T15_052410
	TRADV4M_T10_060612	TRADV4M_T14_061812	TRADV4M_T21A_060712	TRADV4M_T14_052410	TRADV4M_T17_071712
	TRADV4M_T26_052410	TRADV4M_T16_060612	TRADV4M_T01_071712	TRADV4M_T19_062012	TRADV4M_T19_071712
	TRADV4M_T07_071712	TRADV4M_T09_070912	TRADV4M_T07_062612	TRADV4M_T05_061812	TRADV4M_T14_071712
	TRADV4M_T08_062512	TRADV4M_T19_061812	TRADV4M_T23_062612	TRADV4M_T07_052410	TRADV4M_T02_062612
	TRADV4M_T21B_060712	TRADV4M_T03_060612	TRADV4M_T10_062512	TRADV4M_T22_062012	TRADV4M_T11_062612
	TRADV4M_T23_062012	TRADV4M_T16_052410	TRADV4M_T29_052410	TRADV4M_T17_062012	TRADV4M_T30_052410
	TRADV4M_T20_062612	TRADV4M_T24_061812	TRADV4M_T23_062512	TRADV4M_T19_060512	TRADV4M_T22_062512
	TRADV4M_T09_060612	TRADV4M_T10_061812	TRADV4M_T23_061812	TRADV4M_T20_061812	TRADV4M_T10_070912
	TRADV4M_T22_052410	TRADV4M_T06_062012	TRADV4M_T24_062512	TRADV4M_T20_060512	TRADV4M_T09_062012
	TRADV4M_T22_062612	TRADV4M_T23B_060712	TRADV4M_T01_062612	TRADV4M_T23_052410	TRADV4M_T25_052410
	TRADV4M_T08_071712	TRADV4M_T16_062012	TRADV4M_T01_060612	TRADV4M_T08_070912	TRADV4M_T13_052410
	TRADV4M_T24_062612	TRADV4M_T22_061812	TRADV4M_T15_062612	TRADV4M_T07_062512	TRADV4M_T27_052410
	TRADV4M_T06_060612	TRADV4M_T05_071712	TRADV4M_T05_062512	TRADV4M_T17_061812	TRADV4M_T13_061812
	TRADV4M_T18_062612	TRADV4M_T02_061812	TRADV4M_T11_062512	TRADV4M_T17_062612	TRADV4M_T01_060512
	TRADV4M_T01_062512	TRADV4M_T24_071712	TRADV4M_T24B_060712	TRADV4M_T10_062012	TRADV4M_T09_062612
	TRADV4M_T04_060512	TRADV4M_T19_052410	TRADV4M_T13_060612	TRADV4M_T02_062512	TRADV4M_T14_062512
	TRADV4M_T24_062012	TRADV4M_T12_052410	TRADV4M_T17_052410		
29.	TRADV4_KY346745	TRADV4_KY346751	TRADV4_KY346755	TRADV4_KY346753	TRADV4_KY346748
	TRADV4_KY346750	TRADV4_KY346744	TRADV4_KY346746	TRADV4_KY346757	TRADV4_KY346749
	TRADV4_KY346756	TRADV4_KY346758	TRADV4_KY346747	TRADV4_KY346754	TRADV4_KY346743
	TRADV4_KY346759				

Figure Supplement A-7. (*continued*)

30.	TRADV4J_T01_030110	TRADV4J_T04_032310	TRADV4J_T04_031210	TRADV4J_B03_072210	TRADV4J_T04_022510
	TRADV4J_T28_020210	TRADV4J_T12_020110	TRADV4J_T03_032310	TRADV4J_T03_030110	TRADV4J_T01_032310
31.	TRADV4J_T21_100710	TRADV4J_V24_102810			
32.	TRADV4M_T21_051410	TRADV4M_T24_051410	TRADV4M_T26_051410	TRADV4M_T35_051410	TRADV4M_T13A_060712
	TRADV4M_T27_051410	TRADV4M_T32_051410			
33.	TRADV4M_T28_051410	TRADV4M_T23_051410 T	RADV4M_T10_052410	TRADV4M_T17A_060712	TRADV4J_T02_020210
	TRADV4_KY346760	TRADV4_KY346768	TRADV4_KY346766		
34.	TRADV4M_T20A_060712	TRADV4M_T31_051410	TRADV4M_T16A_060712	TRADV4M_T08A_060712	TRADV4_KY346762
	TRADV4_KY346764	TRADV4_KY346765	TRADV4_KY346761		
35.	TRADV4J_V06_102810	TRADV4J_T18_100710	TRADV4J_T21_100410		
36.	TRADV4J_V19_101310	TRADV4J_V18_091010			
37.	TRADV4J_T02_032310	TRADV4J_T03_031210	TRADV4J_T01_031210	TRADV4J_T26_020210	
38.	TRADV4J_B02_072210	TRADV4J_T02_022510			
39.	TRADV6M_T08_072712	TRADV6M_T06_072712	TRADV6M_T15_072712	TRADV6M_T20_072712	TRADV6M_T02_072712
	TRADV6M_T11_072712	TRADV6M_T04_072712	TRADV6M_T23_072712	TRADV6M_T05_072712	TRADV6M_T14_072712
	TRADV6M_T17_072712	TRADV6M_T03_072712	TRADV6M_T13_072712	TRADV6M_T24_072712	TRADV6M_T10_072712
	TRADV6M_T19_072712	TRADV6M_T21_072712	TRADV6M_T07_072712	TRADV6M_T18_072712	TRADV6M_T16_080912
	TRADV6M_T03_080212				
40.	TRADV6J_T06_100410	TRADV6M_T07_071212	TRADV6M_T20_080912		
41.	TRADV7J_T39_100410	TRADV7J_T05_100710	TRADV7J_V08_091410	TRADV7J_T03_100410	
42.	TRADV8B_A_419.23	TRADV8M_T05_070912	TRADV8B_A_419.32	TRADV8W_T15_190409	TRADV8W_S34_190409
43.	TRADV8T_S11_190308	TRADV8T_S06_190308	TRADV8T_S04_190308	TRADV8T_S05_190308	TRADV8T_S03_190308
	TRADV8T_S10_190308	TRADV8T_S14_190308	TRADV8T_S15_190308	TRADV8T_S12_190308	TRADV8T_S08_190308
	TRADV8T_S09_190308	TRADV8T_S16_190308	TRADV8F_T40_190308	TRADV8T_S07_190308	TRADV8T_S13_190308
	TRADV8T_S02_190308	TRADV8T_S01_190308	TRADV8T_T61_190308	TRADV8T_T63_190308	TRADV8T_T60_190308
	TRADV8T_T58_190308	TRADV8F_T39_190308	TRADV8T_T57_190308	TRADV8T_T53_190308	TRADV8T_T51_190308
	TRADV8T_T49_190308	TRADV8T_T62_190308	TRADV8F_T38_190308	TRADV8T_T56_190308	TRADV8F_T34_190308
	TRADV8T_T55_190308	TRADV8F_T24_190308	TRADV8F_T36_190308	TRADV8T_T59_190308	TRADV8T_T50_190308
	TRADV8T_T54_190308	TRADV8W_T47_190308	TRADV8F_T35_190308	TRADV8W_T29_190308	TRADV8W_T48_190308
	TRADV8F_T33_190308	TRADV8F_T22_190308	TRADV8F_T20_190308	TRADV8F_T19_190308	TRADV8W_T44_190308
	TRADV8W_T31_190308	TRADV8W_T42_190308	TRADV8W_T28_190308	TRADV8W_T45_190308	TRADV8W_T41_190308
	TRADV8F_T17_190308	TRADV8F_V68_190308	TRADV8F_V67_190308	TRADV8F_V66_190308	TRADV8F_V64_190308
	TRADV8W_T46_190308	TRADV8W_T25_190308	TRADV8F_T18_190308		
44.	TRADV8M_T17_071612	TRADV8M_T21_072610	TRADV8J_T16_100710	TRADV8J_V21_090110	
45.	TRADV8J_V20_100809	TRADV8J_V22_100809	TRADV8J_T16_031210	TRADV8J_T11_030110	TRADV8J_T14_022510
	TRADV8J_B03_082410				
46.	TRADV8M_T04_051410	TRADV8M_T03_051410			
47.	TRADV8_KY346773	TRADV8_KY346774	TRADV8_KY346775		
48.	TRADV8J_T20_100410	TRADV8J_T04_100710	TRADV8M_T06_051410		
49.	TRADV8M_T17_080212	TRADV8M_T10_071212			
50.	TRADV8W_S24_190119	TRADV8F_V65_190308			
51.	TRADV8M_T21_080912	TRADV8M_T05_072610			
52.	TRADV8_KY346777	TRADV8_KY346781	TRADV8_KY346780		
53.	TRADV8M_T10_072012	TRADV8M_T18_071612	TRADV8J_T22_100410	TRADV8M_T12_072712	TRADV8M_T16_072712
	TRADV8M_T09_072712	TRADV8M_T22_072712			
54.	TRADV8_KY346779	TRADV8_KY346784	TRADV8_KY346786	TRADV8_KY346776	TRADV8_KY346778
	TRADV8_KY346785				
55.	TRADV8M_T11_080212	TRADV8M_T03_080912	TRADV8M_T17_080912	TRADV8M_T20_051410	
56.	TRADV8M_T13B_060712	TRADV8M_T20B_060712			
57.	TRADV8J_T13_031210	TRADV8J_T25_032310			
58.	TRADV8J_T13_022510	TRADV8J_T15_022510			
59.	TRADV8J_T01_082410	TRADV8J_T09_030110			
60.	TRADV8J_V09_090110	TRADV8J_T01_100710			
61.	TRADV8_KY346787	TRADV8_KY346792	TRADV8_KY346788	TRADV8_KY346789	TRADV8_KY346791
	TRADV8_KY346790				
62.	TRADV8_KY346772	TRADV8_KY346769	TRADV8_KY346770	TRADV8_KY346771	
63.	TRADV8J_T16_022510	TRADV8J_B01_082410	TRADV8J_T09_031210	TRADV8J_B04_082410	
64.	TRADV8J_T15_100710	TRADV8J_V11_110210			
65.	TRADV8J_B05_083110	TRADV8J_B02_082410	TRADV8J_B01_083110	TRADV8J_T24_030210	
66.	TRADV8J_V23_100809	TRADV8J_B04_083110	TRADV8J_T21_030210		

Figure Supplement A-7. (*continued*)

67.	TRADV8_KY189339	TRADV8_KY189340			
68.	TRADV9M_T11_080912	TRADV9M_T17_072012	TRADV9M_T01_080912		
69.	TRADV10F_V50_190409	TRADV10G_TH3_190403	TRADV10W_T08_190214	TRADV10W_T02_190220	TRADV10F_T23_190308
	TRADV10G_T66_190409	TRADV10W_S28_190130	TRADV10G_T72_190409	TRADV10G_TC96_190409	
70.	TRADV10_KY346800	TRADV10_KY346806	TRADV10_KY346798	TRADV10_KY346794	TRADV10_KY346812
	TRADV10_KY346813	TRADV10_KY346807	TRADV10_KY346795	TRADV10_KY346802	TRADV10_KY346809
	TRADV10_KY346816	TRADV10_KY346814	TRADV10_KY346811	TRADV10_KY346793	TRADV10_KY346808
	TRADV10_KY346801	TRADV10_KY346810	TRADV10_KY346815	TRADV10_KY346796	TRADV10_KY346804
	TRADV10_KY346799	TRADV10_KY346805	TRADV10_KY346797		
71.	TRADV11M_T23_080212	TRADV11M_T21_071212	TRADV11M_T15_080912	TRADV11M_T06_070612	TRADV11M_T02_080212
	TRADV11M_T12_070912	TRADV11J_T01_100410	TRADV11J_T46_100410	TRADV11M_T20_071212	TRADV11M_T15_072612
	TRADV11M_T14_072612	TRADV11M_T09_072612	TRADV11J_V18_102810	TRADV11_KY189343	TRADV11_KY189353
	TRADV11_KY189351				
72.	TRADV11_KY189342	TRADV11_KY189348	TRADV11_KY189345		
73.	TRADV11W_T01_190220	TRADV11W_T05_190214			
74.	TRADV11M_T01_071212	TRADV11M_T08_071612	TRADV11M_T09_080912	TRADV11M_T18_080212	TRADV11M_T19_080212
	TRADV11M_T17_070612	TRADV11M_T18_072012	TRADV11M_T13_071212	TRADV11M_T16_071212	TRADV11M_T23_072612
	TRADV11M_T20_062512	TRADV11M_T07_060612	TRADV11M_T03_062012	TRADV11M_T05_060612	

D NTRV-STRDV (NAR-TCR) sequences (Supplemental Figure 3)

1.	NTRV1_STRDV1_DQ022688	NTRV1_STRDV1W_T17_190214	NTRV1G_T88_190409	NTRV1T_T39_190409
	NTRV1W_T27_190308			
2.	NTRV1_STRDV1_DQ022689	NTRV1_STRDV1_DQ022690	NTRV1_STRDV1_DQ022691	
3.	NTRV1_STRDV1W_S41_190119	NTRV1_STRDV1W_T13_190214		
4.	NTRV1_STRDV1W_S35_190130	NTRV1_STRDV1W_S47_190119	NTRV1_STRDV1W_T01_190227	NTRV1_STRDV1W_T03_190227
	NTRV1_STRDV1W_T16_190214	NTRV1T_S36_190214		
5.	NTRV3_STRDV2_DQ022704	NTRV3_STRDV2_DQ022706	NTRV3_STRDV2W_S45_190119	NTRV3_STRDV2W_S46_190119
6.	NTRV3_STRDV2_DQ022702	NTRV3_STRDV2_DQ022696	NTRV3_STRDV2_DQ022697	NTRV3_STRDV2_DQ022699
7.	NTRV1_STRDV1_DQ022692	NTRV1_STRDV1_DQ022691	NTRV1_STRDV1W_T10_190214	NTRV1_STRDV1W_T13_190214
	STRDV1W_T30_190308	NTRV1_STRDV1W_S41_190119		
8.	NTRV1_STRDV1_DQ022689	NTRV1_STRDV1_DQ022688		
9.	NTRV1_STRDV1W_S35_190130	NTRV1_STRDV1W_T01_190227	NTRV1_STRDV1W_T03_190227	NTRV1_STRDV1W_T16_190214
	NTRV1_STRDV1W_S42_190119			
10.	NTRV1_STRDV1_DQ022694	NTRV1_STRDV1W_S48_190119	NTRV1_STRDV1W_T05_190227	NTRV1_STRDV1W_T12_190214
	NTRV1_STRDV1W_T18_190214			
11.	NTRV2_STRDV2W_S44_190119	NTRV2_STRDV2_DQ022709	NTRV2_STRDV2W_S49_190119	
12.	NTRV3_STRDV3W_S45_190119	NTRV3_STRDV3W_S43_190119	NTRV3_STRDV3W_S46_190119	NTRV3_STRDV3W_S50_190119
	NTRV3_STRDV3_DQ022704	NTRV3_STRDV3_DQ022706	NTRV3_STRDV3_DQ022702	NTRV3_STRDV3_DQ022699
	NTRV3_STRDV3_DQ022701	NTRV3_STRDV3_DQ022712	NTRV3_STRDV3_DQ022698	NTRV3_STRDV3_DQ022703
13.	STRDV3_KY346705	STRDV3_ KY346706	STRDV3_KY346707	STRDV3_ KY346708
	STRDV3_ KY346710			
14.	STRDV4_KY346738	STRDV4 _KY346739	STRDV4_KY346740	STRDV4_KY346741
	STRDV4_ KY346742			

Figure Supplement A-7. (*continued*)

E

1.	IgMV1-TRDCG_TC94_190409	IgMV1-TRDCW_T02_190130	IgMV1-TRDCG_T55_190409	IgMV1-TRDCT_T41_190409	IgMV1-TRDCF_VE1_190403
	IgMV1-TRDCW_T05_190130	IgMV1-TRDCW_T09_190214	IgMV1-TRDCW_S26_190130	IgMV1-TRDCG_TF11_190403	IgMV1-TRDCT_S35_190214
	IgMV1-TRDCW_T04_190220	IgMV1-TRDCW_T06_190130	IgMV1-TRDCG_T58_190409	IgMV1-TRDCT_S33_190214	IgMV1-TRDCF_T21_190308
	IgMV1-TRDCW_T11_190409	IgMV1-TRDCT_S31_190214			
2.	IgMV1-TRDC_JF507668	IgMV1-TRDC_JF507677	IgMV1-TRDC_JF507669	IgMV1-TRDC_JF507661	IgMV1-TRDC_JF507673
	IgMV1-TRDC_JF507665	IgMV1-TRDC_JF507685	IgMV1-TRDC_JF507692	IgMV1-TRDC_JF507663	IgMV1-TRDC_JF507697
	IgMV1-TRDC_JF507708	IgMV1-TRDC_JF507683	IgMV1-TRDC_JF507695	IgMV1-TRDC_JF507662	IgMV1-TRDC_JF507675
	IgMV1-TRDC_JF507684	IgMV1-TRDC_JF507698	IgMV1-TRDC_JF507671	IgMV1-TRDC_JF507682	IgMV1-TRDC_JF507681
	IgMV1-TRDC_JF507676	IgMV1-TRDC_JF507693	IgMV1-TRDC_JF507688	IgMV1-TRDC_JF507678	IgMV1-TRDC_JF507689
	IgMV1-TRDC_JF507701	IgMV1-TRDC_JF507666	IgMV1-TRDC_JF507694	IgMV1-TRDC_JF507679	IgMV1-TRDC_JF507674
3.	IgMV1-TRDC_JF507687	IgMV1-TRDC_JF507706	IgMV1-TRDC_JF507690		
4.	IgMV1-TRDC_JF507691	IgMV1-TRDC_JF507667			
5.	IgMV2-TRDCW_T01_190214	IgMV2-TRDCW_S28_190409	IgMV2-TRDCG_T61_190409	IgMV2-TRDCW_T06_190220	IgMV2-TRDCF_VH1_190403
	IgMV2-TRDCW_S21_190409	IgMV2-TRDCT_T52_190308	IgMV2-TRDCG_T75_190409	IgMV2-TRDCT_S30_190214	IgMV2-TRDCW_T05_190220
6.	IgMV2-TRDCT_S29_190220	IgMV2-TRDCT_S41_190220	IgMV2-TRDCT_S71_190214	IgMV2-TRDCT_S42_190220	IgMV2-TRDCT_S57_190214
	IgMV2-TRDCT_T42_190409	IgMV2-TRDCT_S43_190220	IgMV2-TRDCG_T81_190409	IgMV2-TRDCG_T60_190409	IgMV2-TRDCW_SD7_190329
	IgMV2-TRDCW_S29_190409	IgMV2-TRDCT_S44_190214	IgMV2-TRDCT_S26_190220	IgMV2-TRDCW_S32_190119	IgMV2-TRDCT_S29_190214
	IgMV2-TRDCT_S66_190214	IgMV2-TRDCT_S43_190214	IgMV2-TRDCT_S58_190214	IgMV2-TRDCT_S46_190220	IgMV2-TRDCT_S45_190220
	IgMV2-TRDCF_V71_190308	IgMV2-TRDCG_TG9_190403	IgMV2-TRDCT_S38_190220	IgMV2-TRDCT_S28_190220	IgMV2-TRDCF_V70_190308
	IgMV2-TRDCT_S31_190220	IgMV2-TRDCW_S39_190119	IgMV2-TRDCT_S39_190220	IgMV2-TRDCT_S40_190220	IgMV2-TRDCW_S25_190409
	IgMV2-TRDCG_T56_190409	IgMV2-TRDCT_S41_190214	IgMV2-TRDCT_S25_190220	IgMV2-TRDCW_S23_190409	IgMV2-TRDCT_S59_190214
	IgMV2-TRDCT_S63_190214	IgMV2-TRDCT_S33_190220	IgMV2-TRDCT_S67_190214	IgMV2-TRDCT_S36_190220	IgMV2-TRDCT_S62_190214
	IgMV2-TRDCT_S47_190220	IgMV2-TRDCT_S35_190220	IgMV2-TRDCG_T62_190409	IgMV2-TRDCT_S40_190214	IgMV2-TRDCW_S38_190119
	IgMV2-TRDCW_S33_190119				
7.	IgMV2-TRDCG_T82_190409	IgMV2-TRDCW_T06_190119	IgMV2-TRDCW_T09_190119		
8.	IgMV4-TRDCW_T43_190308	IgMV4-TRDCW_T08_190119			
9.	IgMV5-TRDCW_T07_190220	IgMV5-TRDCF_V47_190409	IgMV5-TRDCT_S72_190214	IgMV5-TRDCG_T85_190409	IgMV5-TRDCT_S27_190220
	IgMV5-TRDCW_S17_190409	IgMV5-TRDCW_S40_190119	IgMV5-TRDCG_T80_190409	IgMV5-TRDCF_V52_190409	IgMV5-TRDCG_TH5_190403
	IgMV5-TRDCF_V69_190308	IgMV5-TRDCG_T59_190409	IgMV5-TRDCG_T64_190409	IgMV5-TRDCW_T12_190220	IgMV5-TRDCW_T16_190409
	IgMV5-TRDCW_TD4_190329	IgMV5-TRDCW_TC4_190329	IgMV5-TRDCG_T54_190409	IgMV5-TRDCW_S33_190409	IgMV5-TRDCW_S24_190409
	IgMV5-TRDCW_T10_190409	IgMV5-TRDCW_T01_190119	IgMV5-TRDCW_T02_190119	IgMV5-TRDCW_T03_190220	IgMV5-TRDCT_S55_190214
	IgMV5-TRDCW_T04_190214				
10.	IgMV5-TRDCT_S60_190214	IgMV5-TRDCT_S53_190214	IgMV5-TRDCW_S36_190119		
11.	IgMV5-TRDCT_S37_190214	IgMV5-TRDCW_S32_190130	IgMV5-TRDCT_S37_190220	IgMV5-TRDCW_S35_190119	
12.	IgWV1-TRDCG_T63_190409	IgWV1-TRDCF_V46_190409			
13.	IgWV1-TRDC_JF507628	IgWV1-TRDC_JF507659	IgWV1-TRDC_JF507616	IgWV1-TRDC_JF507651	IgWV1-TRDC_JF507617
	IgWV1-TRDC_JF507612	IgWV1-TRDC_JF507647	IgWV1-TRDC_JF507615	IgWV1-TRDC_JF507614	IgWV1-TRDC_JF507607
	IgWV1-TRDC_JF507640	IgWV1-TRDC_JF507657	IgWV1-TRDC_JF507637		
14.	IgWV2-TRDC_JF507655	IgWV2-TRDC_JF507606	IgWV2-TRDC_JF507618		
15.	IgWV2-TRDC_JF507621	IgWV2-TRDC_JF507645	IgWV2-TRDC_JF507623	IgWV2-TRDC_JF507638	IgWV2-TRDC_JF507610
	IgWV2-TRDC_JF507619	IgWV2-TRDC_JF507639			
16.	IgWV3-TRDCG_T79_190409	IgWV3-TRDCG_TH12_190403	IgWV3-TRDCG_TH4_190403	IgWV3-TRDCW_S31_190130	
17.	TAILV1J_T19_100410	TAILV1m_T10_080212	TAILV1m_T20_072012	TAILV1m_T05_080212	TAILV1m_T16_072612
	TAILV1J_V10_090110	TAILV1m_T16_071612	TAILV1m_T15_080212	TAILV1m_T09_071612	

Figure Supplement A-7. (*continued*)

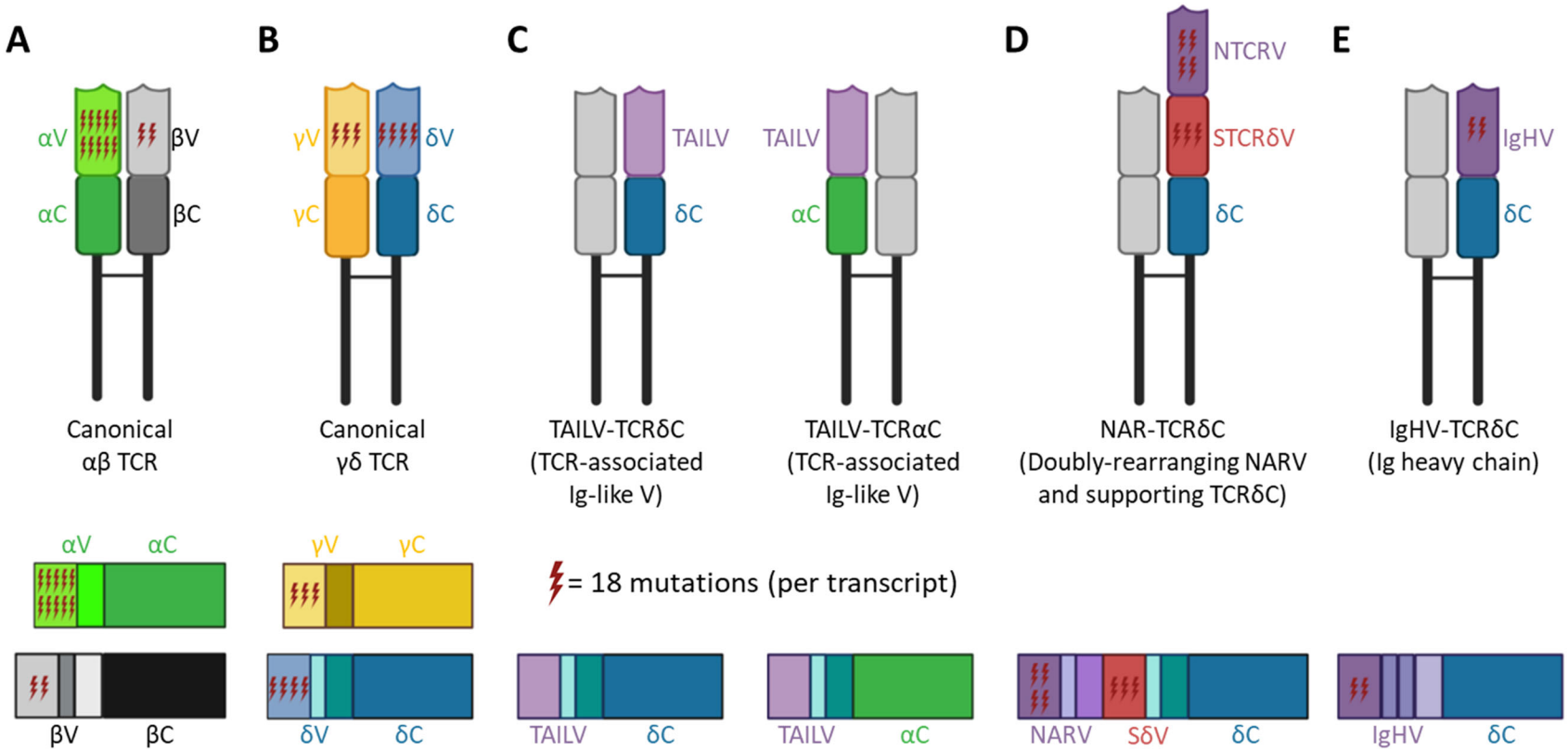

Figure Supplement A-8. Cartoon depictions of putative assembled T cell receptors (TCR) illustrate the disparate use of somatic hypermutation (SHM). Relative use of SHM in transcripts or receptors from canonical T cell chains: canonical [A] αβ (alpha: α, green; beta: β, black), and [B] γδ (gamma: γ, gold; delta: δ, blue) TCR, and non-canonical T cell chains: [C] TCR-associated Ig-like V (TAILV) associated with TCRδC or TCRαC, [D] doubly-rearranging NAR-TCR (Ig-like NTCRV, purple-gray; STCRδV, red) associated with TCRδC, and [E] IgHV associated with TCRδC,. Lightning bolts within V segments of transcripts or receptors indicate the average number of mutations to a single transcript (each bolt indicates 18 mutations). We observed no mutation to TAILV associated with TCRδC or TCRαC. Ig-like components of non-canonical T cell receptors are indicated with purple hues. [V: variable segment, C: constant region] [Figure created with BioRender.com]

Table Supplement A-1. Accession numbers of new and published sequences used in this paper. [gDNA: genomic DNA; cDNA: cDNA derived from mRNA]

V SEGMENT	ACCESSION NUMBER(S)	DNA	STUDY
TCRγV1	FJ513774-FJ513777, FJ513779, FJ513781-FJ513784, FJ513786-FJ513787, FJ513789-FJ513791, FJ513795 KY351639-KY351655 MN748821-MN748850	cDNA	[27] Criscitiello *et al.* 2010 [35] Ott *et al.* 2018 *This study*
TCRγV2	FJ513778 MN748851-MN748868	cDNA	[27] Criscitiello *et al.* 2010 *This study*
TCRγV3	FJ513792-FJ513794 KY351656-KY351693 MN748869-MN748887	cDNA	[27] Criscitiello *et al.* 2010 [35] Ott *et al.* 2018 *This study*
TCRγV4	FJ513785 KY351694-KY351707 MN748888-MN748891	cDNA	[27] Criscitiello *et al.* 2010 [35] Ott *et al.* 2018 *This study*
TCRβV1	FJ513713, FJ513722-FJ513725, FJ513727, FJ513728, FJ513731, FJ513739, FJ513741, FJ513745, FJ513753, FJ513766, FJ513768 KY351708-KY351758 MN748625-MN748702	cDNA	[27] Criscitiello *et al.* 2010 [35] Ott *et al.* 2018 *This study*
TCRβV2	FJ513704, FJ513705, FJ513707, FJ513708, FJ513711, FJ513712, FJ513736, FJ513748, FJ513752 KY351759-KY351764 MN748703-MN748742	cDNA	[27] Criscitiello *et al.* 2010 [35] Ott *et al.* 2018 *This study*
TCRβV3	FJ513710, FJ513740, FJ513759 MN748743-MN748790	cDNA	[27] Criscitiello *et al.* 2010 *This study*
TCRβV4	FJ513719, FJ513732, FJ513735, FJ513737, FJ513742, FJ513744 MN748791-MN748820	cDNA	[27] Criscitiello *et al.* 2010 *This study*
TCRαδV1	FJ513680, FJ513682, FJ513691 KY189332-KY189336, KY366469-KY366481 MN748048-MN748084	cDNA	[27] Criscitiello *et al.* 2010 [35] Ott *et al.* 2018 *This study*
TCRαδV2	KY346716, FJ513836, FJ513686, FJ513819, FJ513818, FJ513803, DQ022718 KY346711-KY346721, KY346723-KY346737, KY366482-KY366485 MN748085-MN748146	cDNA	[27] Criscitiello *et al.* 2010 [35] Ott *et al.* 2018 *This study*
TCRαδV3	FJ513802, FJ513805, FJ513808, FJ513812, FJ513816, FJ513817, FJ513822, FJ513823 MN748147-MN748174	cDNA	[27] Criscitiello *et al.* 2010 *This study*
TCRαδV4	FJ513695, FJ513697, FJ513700, FJ513701 KY346743-KY346768 MN748175-MN748387	cDNA	[27] Criscitiello *et al.* 2010 [35] Ott *et al.* 2018 *This study*
TCRαδV5	KY189337-KY189338 MN748388	cDNA	[35] Ott *et al.* 2018 *This study*
TCRαδV6	FJ513804, FJ513824 MN748389-MN748412	cDNA	[27] Criscitiello *et al.* 2010 *This study*
TCRαδV7	FJ513687, FJ513689 MN748413-MN748418	cDNA	[27] Criscitiello *et al.* 2010 *This study*
TCRαδV8	FJ513681, FJ513690, FJ513693, FJ513694, FJ513702, FJ513801, FJ513814 KY189339-KY189341, KY346769-KY346792, KY366486-KY366487 MN748419-MN748576	cDNA	[27] Criscitiello *et al.* 2010 [35] Ott *et al.* 2018 *This study*
TCRαδV9	MN748577-MN748579	cDNA	*This study*
TCRαδV10	KY346793-KY346816 MN748577-MN748579	cDNA	[35] Ott *et al.* 2018 *This study*
TCRαδV11	KY189342-KY189349, KY189350-KY189354 MN748591-MN748624	cDNA	[35] Ott *et al.* 2018 *This study*
TAILV1	MN061599 MN748038-MN748046	gDNA cDNA	[26] Deiss *et al.* 2019 *This study*
TAILV2	MN061604 MN748047	cDNA	[26] Deiss *et al.* 2019 *This study*
IgMV1	DQ857386 AY609255, AY609260	gDNA cDNA	[55] Lee *et al.* 2008 [57] Rumfelt *et al.* 2004
IgMV1-TCRδC	JF507606–JF507660, MN061605, MN061611, MN061613, MN061620, MN061624, MN061626, MN061630 MN788155-MN788172	cDNA	[26] Deiss *et al.* 2019 *This study*
IgMV2	DQ857389, DQ192492, DQ192493, DQ192494 AY609272, AY609246, AY609247	gDNA cDNA	[55] Lee *et al.* 2008 [57] Rumfelt *et al.* 2004
IgMV2-TCRδC	MN061606, MN061607, MN061609, MN061612, MN061614, MN061615, MN061618, MN061622, MN061625, MN061627, MN061631, MN061632, MN061634 MN788173-MN788240	cDNA	[26] Deiss *et al.* 2019 *This study*
IgMV3	DQ857384 AY609250, AY609253, AH013814	gDNA cDNA	[55] Lee *et al.* 2008 [57] Rumfelt *et al.* 2004
IgMV3-TCRδC	MN788241	cDNA	*This study*
IgMV4	DQ857387	gDNA	[36] Malecek *et al.* 2008
IgMV4-	MN061629	cDNA	[26] Deiss *et al.* 2019

Table Supplement A-1. (*continued*)

V SEGMENT	ACCESSION NUMBER(S)	DNA	STUDY
TCRδC	MN788242-MN788243		*This study*
IgMV5	DQ857385	gDNA	[55] Lee *et al.* 2008
	AY609248, AY609249, AY609254, AY609258, AY609265-AY609267	cDNA	[57] Rumfelt *et al.* 2004
IgMV5-TCRδC	MN061610, MN061616, MN061617, MN061619, MN061621, MN061623, MN061628, MN061633	cDNA	[26] Deiss *et al.* 2019
	MN788244-MN788280		*This study*
IgWV1	KC920791, KF192880, KF192883	gDNA	[42] Zhang *et al.* 2013
	AH013807, AY609227, AY609233, AY609244, AY524297	cDNA	[57] Rumfelt *et al.* 2004
IgWV1-TCRδC	JF507607, JF507611-JF507617, JF507620, JF507625, JF507627- JF507629, JF507637, JF507640, JF507643-JF507644, JF507646-JF507648, JF507650-JF507651, JF507653, JF507657, JF507659, JF507660	cDNA	[26] Deiss *et al.* 2019
	MN788281-MN788282		*This study*
IgWV2	KF192877, KC920789, KF192879	gDNA	[42] Zhang *et al.* 2013
IgWV2-TCRδC	JF507606, JF507608-JF507610, JF507618-JF507619, JF507621-JF507624, JF507626, JF507630-JF507636, JF507638-JF507639, JF507641, JF507645, JF507649, JF507652, JF507655-JF507656, JF507658, MN061603	cDNA	[26] Deiss *et al.* 2019
	MN788283		*This study*
IgWV3	KF192882, KF192884	gDNA	[42] Zhang *et al.* 2013
	AY609238, AY609245, AY531553	cDNA	[57] Rumfelt *et al.* 2004
IgWV3-TCRδC	MN061608	cDNA	[26] Deiss *et al.* 2019
	MN788284-MN788287		*This study*
NTCRV1-STCRδV1	DQ022688–DQ022694	cDNA	[29] Criscitiello *et al.* 2006
	MN748005-MN748024, MN748032-MN748033		*This study*
NTCRV2-STCRδV2	DQ022707, DQ022709, DQ022710, DQ022713	cDNA	[29] Criscitiello *et al.* 2006
	MN748025-MN748027, MN748034		*This study*
NTCRV2-STCRδV4	DQ022707	cDNA	[29] Criscitiello *et al.* 2006
	KY346738, KY346739, KY346740, KY346741, KY346742		[35] Ott *et al.* 2018
NTCRV3-STCRδV3	DQ022696–DQ022706, DQ022711, DQ022712	cDNA	[29] Criscitiello *et al.* 2006
	KY346705, KY346706, KY346707, KY346708, KY346709, KY346710		[35] Ott *et al.* 2018
	MN748028-MN748031, MN748035-MN748037		*This study*

Table Supplement A-2. Frequencies of somatic hypermutation in nurse shark thymus and peripheral lymphoid tissue (blood, spleen, and spiral valve). Mutation frequency was measured as the total number of nucleotide changes to a Geneious-derived consensus sequence divided by the total number of nucleotides in all sequences. Nonsynymous (N) and synonymous (S) mutations (mut) were counted separately for each framework (FR) and complementarity-determining region (CDR) for each predicted [A] TCR-alpha/delta, [B] TCR-beta, and [C] TCR-gamma V gene segment group. [FR1, FR2, FR3, CDR1, CDR2, and CDR3 refer to the first, second, or third FR or CDR region, respectively.]

[A]

Tissue Type	Mut Type	FR Mutations (#)				CDR Mutations (#)			Mutation Frequency				
		FR1	FR2	FR3	All FR	CDR1	CDR2	All CDR	FR1	CDR1	FR2	CDR2	FR3
Thymus (465 sequences)	N	5467	1425	4319	11211	2110	2155	4265	0.474	0.668	0.182	0.701	0.240
	S	2750	1022	1273	5045	489	560	1049	0.239	0.155	0.130	0.182	0.071
	ALL	8217	2447	5592	16256	2599	2715	5314	0.713	0.823	0.312	0.883	0.311
Total Nucleotides		11525	7837	17979	37341	3158	3075	6233					
Periphery (284 sequences)	N	6654	3655	9583	19892	2893	3124	6017	0.937	1.545	0.757	1.689	0.865
	S	3343	1691	3995	9029	980	574	1554	0.471	0.523	0.350	0.310	0.361
	ALL	9997	5346	13578	28921	3873	3698	7571	1.408	2.068	1.107	1.999	1.226
Total Nucleotides		7100	4828	11076	23004	1873	1850	3723					

[B]

Tissue Type	Mut Type	FR Mutations (#)				CDR Mutations (#)			Mutation Frequency				
		FR1	FR2	FR3	All FR	CDR1	CDR2	All CDR	F1	F2	F3	C1	C2
Thymus (160 sequences)	N	10	3	3	16	1	0	1	0.003	0.001	0.000	0.001	0.000
	S	1	2	1	4	0	3	3	0.000	0.001	0.000	0.000	0.004
	ALL	11	5	4	20	1	3	4	0.003	0.002	0.001	0.001	0.004
Total Nucleotides		4000	2720	6240	12960	960	840	1800					
Periphery (383 sequences)	N	18	13	18	49	4	3	7	0.027	0.028	0.020	0.025	0.037
	S	6	5	6	17	0	1	1	0.009	0.011	0.007	0.000	0.012
	ALL	24	18	24	66	4	4	8	0.036	0.039	0.026	0.025	0.049
Total Nucleotides		675	459	918	2052	162	81	243					

[C]

Tissue Type	Mut Type	FR Mutations (#)				CDR Mutations (#)			Mutation Frequency				
		FR1	FR2	FR3	All FR	CDR1	CDR2	All CDR	F1	F2	F3	C1	C2
Thymus (78 sequences)	N	1	0	10	11	2	0	2	0.000	0.000	0.002	0.002	0.000
	S	0	1	6	7	0	1	1	0.000	0.000	0.001	0.000	0.001
	ALL	1	1	16	18	2	1	3	0.000	0.000	0.003	0.002	0.001
Total Nucleotides		4000	2720	6240	12960	960	840	1800					
Periphery (250 sequences)	N	8	8	22	38	3	1	4	0.012	0.017	0.024	0.019	0.012
	S	3	4	11	18	2	0	2	0.004	0.009	0.012	0.012	0.000
	ALL	11	12	33	56	5	1	6	0.016	0.026	0.036	0.031	0.012
Total Nucleotides		675	459	918	2052	162	81	243					

Table Supplement A-3. List of forward (F) and reverse (R) primers used to generate T cell receptor (TCR) sequences. [TCRγ: gamma; TCRβ: beta; TCRα: alpha; TCRδ: delta; V: variable; C: constant]

Chain	Region	ID	F/R	Location	Nucleotide Sequence (5' to 3')	Amino Acid	Tm
TCRβ	V	998	F	TCRβV1 Leader	AGGAATTCTGGCTTC	PGILAS	39
TCRβ	V	999	F	TCRβV1 Leader	GTTGCATGAAGTCTATTATCTCA	LHEVYYL	50
TCRβ	V	1001	F	TCRβV2 Leader	CACTGTCCCAAGGAATTCTG	ALSQGIL	51
TCRβ	V	1004	F	TCRβV4 Leader	ATCTATCATCTCGTCTGGA	IYHLVW	53
TCRβ	V	1005	F	TCRβV4 Leader	GTCTGGATATTGGCACTGTC	VWILALS	50
TCRβ	V	1006	F	TCRβV5 Leader	GTTCGGTGCTCTTTCTC	FGALS	47
TCRβ	V	1007	F	TCRβV5 Leader	CATTTCCTGTTACTGCTG	HFLLLL	46
TCRβ	V	1008	F	TCRβV6 Leader	GAGCTCTTTATCTGTATTTC	GALYLYF	47
TCRβ	V	1009	F	TCRβV6 Leader	CAGGGAGTTGTGAAAAC	PGSCEN	44
TCRβ	V	1010	F	TCRβV7 Leader	GCTTTCGAGTTTATTGCAG	LSSLLQ	47
TCRβ	V	1011	F	TCRβV7 Leader	GTTCTCGTACTGATTCCATGC	VLVLIPC	52
TCRβ	C	217	R	TCRβ Constant	GTATGATGGATTCGGGGTCTGACTG	QSDPESII	67
TCRβ	C	218	R	TCRβ Constant	CTG GTG ATG GTT TGA GGA TCG TGA CT	VTILKPSP	68
TCRβ	C	923	R	TCRβ Constant	CTTTCCGTTTCTCTCTCAGCTC	ELREKRK	71
TCRβ	C	924	R	TCRβ Constant	GTACGTCATTCTGGCTGTTGT	TTARMTY	70
TCRβ	C	932	R	TCRβ Constant	GGA TCT GGA TGT TGT CGG GA	PDNIQI	61
TCRβ	C	933	R	TCRβ Constant	CTC GTG GCG CTG TAG GAT TTA	KSYSATS	61
TCRβ	C	934	R	TCRβ Constant	GGTACGTCATTCTGGCTGTTGTC	TTARMTY	65
TCRβ	C	1015	R	TCRβ Constant	GCTGTAGGATTTATTGTCTTC	QSDPESII	48
TCRβ	C	1016	R	TCRβ Constant	CATTGACGTAAAACTGCGAC	EDNKSYS	50
TCRγ	C	219	R	TCRγ Constant	GATTTGTTTCATGCTCCGCCCGGCA	CRAEHETN	69
TCRγ	C	220	R	TCRγ Constant	TCAGGAGACAGACGACGGCCGCT	AAVVCLL	70
TCRγ	C	1020	R	TCRγ Constant	GGCGACTGACCACTGAGTAGG	AYSVVSRL	65
TCRγ	C	1021	R	TCRγ Constant	CTCCGCCCGGCAGGAAATG	NISCRAE	64
TCRα	C	191	R	TCRα Constant	CATTGGTGGATAGCAAGCCCTTCGAT	SKGLLSTN	76
TCRδ	C	221	R	TCRδ Constant	CCTGTTCCACTCTTGGTCCCCAG	LGTKSGT	68
TCRδ	C	1022	R	TCRδ Constant	GCTGGCCAGACAGACTGCAGCTTGGACAG	AVQAAVCLAS	76

APPENDIX B

SOURCE DATA

Source Data B-1. CDR3 regions diversified by exonuclease activity and addition of N and P nucleotides (Figure 2-4). Alignment of nucleotides belonging to the join between variable (V) and joining (J) segments within TCRα thymocyte clones. We determined the putative ends of each V segment and putative beginning of each J segment by comparing alignments between different sharks, assuming that identical nucleotides between sharks were germline. The last number of each sequence name indicates the number of clones containing that nucleotide sequence between the V and J segments.

```
                              10        20        30
                     ....|....|....|....|....|....|....
aV1J_T17_100710_2    TGT
aV1m_T14_080912_2    TCCGCTTACG
aV1m_T21_072012_2    CCCGTAT
aV1J_V14_101310_2    CCCCCATATACT
aV1m_T17_071212_2    CCGGACT
aV1m_T13_072610_1    CCGCGCT
aV1J_V21_101310_1    CA
aV1J_V09_110210_1    TCC
aV1J_V19_110210_1    CCGG
aV1J_V07_091410_1    TCCCCT
aV1J_V05_091010_1    CCGGTA
aV1m_T08_060512_1    TCGGGA
aV1m_T13_060512_1    CCAAAAAAAT
aV1J_V08_102810_1    GCGCTGTGG
aV1J_V20_110210_1    CTCCTGCGAT
aV1J_V02_110210_1    A
aV1m_T02_060512_1    CCGCCGCAT
aV1B_A_419.40_1      CAACA
aV1m_T14_060512_1    ACCTTACT
aV1J_V24_101310_1    CGGGGCCACTACGGAACTGGTACT
aV1J_V05_102810_1    CACGGG
aV1J_V05_101310_1    CCGGGA
aV1.4J_V07_090110_1  GCCGGGTACT
aV1.4J_V08_101310_1  GCCGGGTACTGGAATA
aV1.4J_V17_090110_1  GCCGGGTACTGGAATAACT
aV1.3J_V01_091410_2  CCTATTGGGGGTTACAGC
aV1.2J_V15_091010_2  CCACGGG
aV1.2J_V23_090110_1  CCACGG
aV1.2J_V18_110210_1  CACGG
aV1.1J_V18_101310_1  CGGGGCCACTACGGAACTGGTACTGG
aV1.1J_V08_110210_1  CGGGGCCACTATGGAACTGGTACT
aV1.1J_V01_101310_1  CGGGGCCACTACGGAGATGAGACT
aV2.1m_T22_071212_5  TTCGCCGGTAT
aV2.1J_V06_091410_1  GAT
aV2.1J_T43_100410_1  CCCC
aV2.1B_A419.3        TGGCC
aV2.2m_T13_080212_1  TTTGAGAAAGAAT
aV2.2m_T04_071212_2  CGGACTGGAGGGTTTACA
aV2.2J_V22_110210_1  TACCCTGGATTCTCTGAACAA
aV2.3m_T11_070612_2  CCGGTCGAAT
aV2.3J_V12_102810_1  CCGACCT
aV2.3J_V07_110210_1  CGAGCCGCCGCATGGCTGAT
aV3m_T23_051410_1    CC
aV3B_A_419.29_1      CCGTCC
aV3m_T32_051410_1    CCCGAGATA
aV3m_T30_051410_1    TCAATGGGGAGCTGGGACT
aV3m_T20A_060712_1   CTCGGAC
aV3J_V24_110210_1    TCGAGGGGGCGTACACCAATCCTCCAC
```

Source Data B-1. (*continued*)

```
                              10        20        30
                     ....|....|....|....|....|....|....
aV3J_V21_110210_1    ATGAGGACTGGAGCTTT
aV3m_T21_051410_1    TCGGCT

aV3m_T29_051410_1    CCGAAGGAAGGAT
aV3m_T08A_060712_1   CAGACGAT
aV3B_A_419.4_1       TCCCCTAAG
aV3m_T11A_060712_1   ACCCCG
aV3m_T10_052410_1    CTCGTCGAACTTTGGAGGTAACAC
aV3m_T16A_060712_1   CTCGAGA
aV3m_T24_051410_1    CTCGATCTCGTCGCTTGC
aV3m_T04A_060712_1   CCAT
aV3J_V19_091010_4    GCG
aV3J_V06_102810_2    CTCGCTCTCG
aV3J_T21_100710_2    TCTAGTTTCT
aV3J_V16_090110_1    TCGGGCAGGCTACTCC
aV3J_V15_102810_1    GACCT
aV3m_T13A_060712_1   CCGGTCCCGCGCTAACGCC
aV3m_T31_051410_1    TGTACT
aV3J_T21_100410_1    CAA
aV3B_A_419.11_1      CCAAATACT
aV3m_T27_051410_1    C
aV3m_T23A_060712_1   TTCCACTGATTCAGGGGGTGCCACTCCAATAATA
aV3J_V06_110210_1    GTCCGGG
aV3m_T28_051410_1    GGAC
aV3m_T26_051410_1    CTCGCGTCAT
aV3m_T35_051410_1    T
aV4m_T13_070612_1    CCCGGATCGGGG
aV4m_T12_060612_1    GAACTAC
aV4J_V01_102810_1    CCTAGCGGCTAC
aV4m_T11_060612_1    CCGCCC
aV4B_A_419.33_1      CCCCGTGCTGGAGCAGCTTCT
aV4J_T10_100710_1    GCTTGGGGAT
aV4m_T12_071612_1    CGG
aV4m_T09_061812_1    CACGCGG
aV4J_V12_091410_2    TCG
aV4m_T07_061812_2    CTGCG
aV4m_T04_062512_2    CCCGGATATGAAC
aV4m_T24_080912_1    TCTCAACCCCCCG
aV4m_T09_080212_3    AAAAGCCGGGCAT
aV4m_T24_071612_3    CTCCAATGCCCG
aV4J_V10_110210_2    CCTAGTGGTT
aV4m_T04_061812_1    CCGGACGCCG
aV4J_V23_110210_1    CCTAGTGGTTAC
aV4J_V03_110210_1    CCG
aV4J_V11_101310_1    CCTAATCTATGATGC
aV4J_T14_100710_1    CTCAC
aV4B_A_419.24_1      TTTTTGGTGACTCGA
aV5B_A_419.36_1      CAC
aV5B_A_419.1_1       CCCAACG
aV5J_V02_090110_1    CGGTTGAATGCTGGAGGAAGTAAT
aV5J_T06_100410_1    TTAT
aV5J_T05_100710_1    CGCACTGGAGCAGTTCTT
aV5J_V08_091410_1    CGTCCGGCTCCGCAC
aV5J_T03_100410_1    CCAGAGGCATACTGGAGCTTGGACT
aV5J_V20_091010_2    AGACGTGAT
aV5m_T08_072712_21   AGATTAAGAAA
aV5m_T07_071212_2    GGGGCTAGGG
aV5J_T05_100410_1    AATGGCGT
aV5J_T39_100410_1    CCCGAGCCCACCGAGAGAT
```

Source Data B-1. (*continued*)

```
                              10        20        30
                     ....|....|....|....|....|....|....
aV6B_A_419.18_1      CAGCCTGGCCGGT
aV6J_V11_090110_1    CCGCGCGGCCGGGTTT
aV6B_A_419.32_1      CAGCCTCCT
aV6m_T05_070912_1    CAGCCCA
aV6m_T11_080912_3    GTCGAAT
aV6B_A_419.23_1      CCGCCGGGGGTCT
aV7m_T14B_060712_1   GCCGGCGGG
aV7m_T14_071212_1    CCCCCGGGGGT
aV7m_T14_080212_1    CCGGCC
aV7J_V21_090110_1    TGGACT
aV7B_A_419.43_1      AACCGAGTCA
aV7m_T20_080212_1    GGCGAG
aV7J_V03_101310_1    CGGCTTCTGAACTAT
aV7m_T31_052410_1    CCTATCGTAT
aV7m_T16_072712_6    GATC
aV7m_T17_080912_3    CGCCACCCC
aV7J_T15_100710_2    CCGAT
aV7m_T21_072610_2    CCCAGTCCGATGTAT
aV7m_T21_080912_2    CTG
aV7m_T10_071212_2    CTCATCG
aV7m_T04_051410_2    CCGTGGGGG
aV7J_V04_102810_2    TGAGAG
aV7J_V18_091410_1    CAATGGGAACTAT
aV7B_A_419.13_1      CCGCCGCCTAGGT
aV7m_T12_072610_1    TACAATAT
aV7m_T16_051410_1    ATACGGCTGCTGGC
aV7m_T11B_060712_1   CAGCAGCC
aV7m_T20B_060712_1   TAT
aV7m_T18_051410_1    GGGGAT
aV7m_T13B_060712_1   TCGCGAAT
aV7J_V02_102810_1    GGGGCT
aV7J_T08_100410_1    TGGACCCCCTATACT
aV7m_T12_051410_1    GGGT
aV7J_T04_100710_1    CGTCCGC
aV7J_V03_091410_1    CTAGTG
aV7J_T16_100710_1    CTATCAGATGTTGGCACC
aV7m_T17_051410_1    GTCACG
aV7J_V04_110210_1    CGGGT
aV7m_T32_052410_1    CCTATCCATC
aV7B_A_419.42_1      GGCTCTAACT
aV7J_T07_100410_1    TTAAGAA
aV7J_T20_100410_1    CTACT
aV7m_T16B_060712_1   AGTCAATAC
aV7m_T02_051410_1    TTGTCGGCTGAT
aV7J_T01_100710_1    CCTGGGTAT
aV7J_T22_100410_1    TCTATGAGGCGG
aV7J_T02_100710_1    AG
aV7m_T06_051410_1    GGGATTTC
aV7m_T19B_060712_1   GGAG
aV7J_V10_091410_1    CCGGAAT
aV1J_V18_090110_1    CCCGGCGCT
aV7m_T20_051410_1    CTAGGCTCT
aV7J_V09_090110_1    GAGTACATGGAA
aV9J_V17_091010_1    TCATCATCATG
aV9J_V20_091410_1    GAAGAGCTGG
aV9J_T06_100710_1    TTGGCGTCCCACGCGCCG
aV9J_T09_102810_1    AGTACCATCGATT
aV9J_T19_100410_1    TACT
aV9m_T05_071212_1    GCTGATATAGGAGATAGA
aV9J_V10_090110_1    CCCCCCCGGAC
aV9m_T16_071612_3    CCCAT
aV9m_T10_080212_3    GGTGATACT
```

Source Data B-1. (*continued*)

```
                              10        20        30
                     ....|....|....|....|....|....|....
aV9B_A_419.46_2      AC
aV9J_V05_110210_2    CAATGGCATTCGGAATGCTGGCACTGAC
aV9m_T16_072612_1    TGGGGTCATACT
aV9J_V12_110210_1    GAAG
aV10.1J_V04_090110_6 AAAT
aV10.1m_T01_071212_5 ATCA
aV10.1m_T09_072612_5 AATCAGAATGCTAACGCCAAC
aV10.1m_T15_080912_3 TTTCATT
aV10.1J_V03_090110_3 TCAAT
aV10.1m_T05_060612_2 AAATCGCT
aV10.1m_T17_070612_2 AAACGGCCC
aV10.1m_T13_071212_2 AACGAGCAG
aV10.1m_T20_062512_1 AGGCTGGA
aV10.1J_V13_090110_1 CTC
aV10.1J_V13_091010_1 TTGCTC
aV10.1m_T03_062012_1 GTCTGT
aV10.1m_T08_060612_1 AAGAAT
aV10.1m_T13_062612_1 TCACAC
aV10.1m_T16_071212_1 CGGAGAATG
aV10.1m_T12_070912_1 AACCAGGATATC
aV10.1J_V18_102810_1 AGTCGGACTTACGGAGCTGGTACT
aV10.1J_T01_100410_1 AAAGTGCCGGTG
aV10.1J_T46_100410_1 AAACGGGATGATT
aV10.2J_T23_100410_1 CCATCGTCA
aV10.2J_V07_101310_1 CTGAGACATCGCCCCCCTG
aV10.2J_V16_110210_1 ATTGCTC
aV10.2J_V22_091410_1 CCTGCTC
aV10.1m_T14_062012_1 CGAACGAGAACTATGCAAACAG
```

Source Data B-2. Mutation frequencies depicted in Figure 2-6. Mutability of complementarity determining regions (CDR), especially CDR1, exceeded that of framework regions (FR) for all mutations together (black bars) and for nonsynonymous mutations alone (NSYN, hatched bars). We found no statistical difference in synonymous mutations (SYN, white bars) between CDRs and FRs.

Mutation Rate

	FR1	CDR1	FR2	CDR2	FR3	CDR3
NSYN	1.63	3.56	1.29	2.01	0.95	2.31
SYN	0.65	1.72	0.57	1.01	0.69	0.95
ALL	2.28	5.29	1.87	3.02	1.64	3.27

Std Dev

	FR1	CDR1	FR2	CDR2	FR3	CDR3
NSYN	4.55	3.52	2.25	2.00	3.41	2.07
SYN	2.99	1.69	1.14	1.03	3.12	1.03
ALL	6.53	4.58	2.84	2.81	5.43	2.27

Source Data B-3. Frequency of mutation within FR and CDR of T cell receptor variable regions depicted in Figure 3-3. Mutation frequency is the total number of nucleotide (nuc) changes to a Geneious-derived consensus sequence divided by the total number of nucleotides within a region. We counted nonsynymous (N) and synonymous (S) mutations separately for each FR and CDR within [A] 422 TCR αδV-TCRAC, [B] 137 TCR αδV-TCRDC, [C] 158 TCR γV, [D] 237 TCR βV, [E] 51 NTCRV, [F] 62 STCRDV, and 275 IgHV-TCRDC sequences.

[A] Alpha: TRADV-TRAC

Total nucleotides

Region	FR1	CDR1	FR2	CDR2	FR3	Total
	31556	7982	21522	8313	48669	121359

Number mutations

Type	FR1	CDR1	FR2	CDR2	FR3	Total
N	97	74	68	47	135	461
S	46	25	34	14	61	209
All	143	99	102	61	196	670

Mutation Frequency

Type	FR1	FR2	FR3	CDR1	CDR2	Total
N	0.31	0.93	0.32	0.57	0.28	0.38
S	0.15	0.31	0.16	0.17	0.13	0.17
All	0.45	1.24	0.47	0.73	0.40	0.55

[B] Delta: TRADV-TRDC

Total nucleotides

Region	FR1	CDR1	FR2	CDR2	FR3	Total
	10275	2469	6987	2784	16023	40221

Number mutations

Type	FR1	CDR1	FR2	CDR2	FR3	Total
N	20	9	13	3	19	72
S	7	4	6	1	11	29
All	27	13	19	4	30	101

Mutation Frequency

Type	FR1	FR2	FR3	CDR1	CDR2	Total
N	0.19	0.36	0.19	0.11	0.12	0.18
S	0.07	0.16	0.09	0.04	0.07	0.07
All	0.26	0.53	0.27	0.14	0.19	0.25

[C] Gamma

Total nucleotides

Region	FR1	CDR1	FR2	CDR2	FR3	Total
	9198	3856	8057	1890	16681	42841

Number mutations

Type	FR1	CDR1	FR2	CDR2	FR3	Total
N	9	5	5	1	37	66
S	3	2	5	0	16	31
All	12	7	10	1	53	97

Mutation Frequency

Type	FR1	CDR1	FR2	CDR2	FR3	Total
N	0.10	0.13	0.06	0.05	0.22	0.15
S	0.03	0.05	0.06	0.00	0.10	0.07
All	0.13	0.18	0.12	0.05	0.32	0.23

[D] Beta

Total nucleotides

Region	FR1	CDR1	FR2	CDR2	FR3	Total
	17568	4248	12069	2979	27823	65544

Number mutations

Type	FR1	CDR1	FR2	CDR2	FR3	Total
N	26	5	14	3	21	92
S	5	0	7	1	7	34
All	31	5	21	4	28	126

Mutation Frequency

Type	FR1	CDR1	FR2	CDR2	FR3	Total
N	0.15	0.12	0.12	0.10	0.08	0.14
S	0.03	0.00	0.06	0.03	0.03	0.05
All	0.18	0.12	0.17	0.13	0.10	0.19

[E] NTRV

Total nucleotides

Region	FR1	CDR1	FR2	CDR2	FR3	Total
	1632	1530	915	2154	4317	10548

Number mutations

Type	FR1	CDR1	FR2	CDR2	FR3	Total
N	8	0	0	3	7	18
S	2	0	0	0	4	6
All	10	0	0	3	11	24

Mutation Frequency

Type	FR1	CDR1	FR2	CDR2	FR3	Total
N	0.49	0.00	0.00	0.14	0.16	0.17
S	0.12	0.00	0.00	0.00	0.09	0.06
All	0.61	0.00	0.00	0.14	0.25	0.23

[F] STRDV

Total nucleotides

Region	FR1	CDR1	FR2	CDR2	FR3	Total
	4376	1116	3162	1302	6825	16781

Number mutations

Type	FR1	CDR1	FR2	CDR2	FR3	Total
N	5	2	3	1	5	16
S	2	2	2	0	7	13
All	7	4	5	1	12	29

Mutation Frequency

Type	FR1	CDR1	FR2	CDR2	FR3	Total
N	0.11	0.18	0.09	0.08	0.07	0.10
S	0.05	0.18	0.06	0.00	0.10	0.08
All	0.16	0.36	0.16	0.08	0.18	0.17

[G] IgHV-TRDC

Total nucleotides

Region	FR1	CDR1	FR2	CDR2	FR3	Total
	16226	5545	13963	4712	31691	68540

Number mutations

Type	FR1	CDR1	FR2	CDR2	FR3	Total
N	23	17	24	27	57	214
S	7	4	10	4	17	63
All	30	21	34	31	77	280

Mutation Frequency

Type	FR1	CDR1	FR2	CDR2	FR3	Total
N	0.14	0.31	0.17	0.57	0.18	0.31
S	0.04	0.07	0.07	0.08	0.05	0.09
All	0.18	0.38	0.24	0.66	0.24	0.41

APPENDIX C

BIOINFORMATIC, STATISTICS, AND TOOLS UTILIZED

C-1. Bioinformatics

We performed all sequence analysis using a student personal license for Geneious Prime Software (BioMatters, Inc.; www.Geneious.com). We created multiple alignments using ClustalW (Chapter 2) or Clustal Omega (Chapter 3) using default settings (fast clustering, cluster size for mBed Guide trees: 100). To determine relationships between sequences (and groups of sequences), we used the Geneious Tree Builder tool to build phylogenetic trees (Jukes-Cantor genetic distance model; Neighbor-joining tree build method; no outgroup; 1000 bootstraps per tree; support threshold 60%). We then grouped sequences based on the resulting distance matrices, using an *a priori* threshold of 70% sequence identity to assign V gene segments into groups. Thus, we grouped sequences sharing 70% identity into the same V segment family.

Within these family groups, we further subdivided sequences into subfamilies based on 80% sequence identity. We allowed Geneious to determine a consensus sequence for each subfamily using a strict threshold (bases matching at least 50% of sequences). We counted nucleotide differences between clone sequences and these consensus sequences as mutation. To reduce the likelihood of counting allelic polymorphisms as mutation, we further identified putative allele groups for each V gene segment. Allele groups shared at least 95% identity, contained the same alterations to the

consensus sequence, and occurred in multiple tissues and/or individuals. We did not count any nucleotide differences we could attribute to putative allelic polymorphisms as mutation. We copied alignments for each sequence group (family and subfamily) into Microsoft Word, using dots to highlight agreements to the consensus sequence.

We identified WA/TW and (AID-favored) DGYW/WRCH ProSite motifs in forward strands of each consensus sequence using the Motif Search tool in Geneious (based on the EMBOSS 6.5.7 tool), allowing zero mismatches. We assessed mutation by examining differences between clone and consensus sequences using the SNP/variation finder tool in Geneious. We created individual annotations for each motif and mutation and exported these annotations to Microsoft Excel. We identified the position of the mutable nucleotide (A or T of each WA and TW; G or C of each DGYW and WRCH) motif within the consensus sequence. For each family or subfamily group, we identified the variant location (sequence position), consensus and altered nucleotides, effects to the codon or amino acid, and whether the mutation occurred to a mutable nucleotide inside or outside a motif.

C-2. Statistics

We assessed differences in mutation between framework regions (FR) and complementarity-determining regions (CDR) and between synonymous (S) and nonsynonymous (N) mutations using one-way Student's T-tests (df=1; $p<0.05$). Based on

previous studies examining somatic hypermutation in B cells undergoing affinity maturation, we predicted higher mutation in CDR than FR and more N mutation than S mutation. We also compared observed and expected mutations between CDR and FR and inside and outside AID-preferred motifs using Chi-square tests ($\chi^2 > 3.841$; df=1; $p>0.05$; see Table B-1 for calculations). We conducted all tests in Microsoft Excel.

Though CDR3 typically contain the greatest diversity within a B or T cell sequence (due to the joining of V(D)J gene segments and exonuclease and TdT activity during recombination), we did not analyze mutation within CDR3 regions. In Chapter 2, we comprised sequence groups based on the similarity of CDR3, assuming that clones with identical CDR3 derived from a common progenitor and thus could be used as surrogate germline sequences. As a result, analyses would underestimate mutation. However, because V(D)J recombination creates a very diverse join within CDR3 regions, it is difficult to distinguish mutation from N and P nucleotides. Thus, we did not analyze mutation to CDR3 regions in Chapter 3 either.

C-3. Tools

Real-time RT-qPCR

Using a previously isolated partial AID (activation-induced cytidine deaminase) sequence from nurse shark (courtesy of Yuko Ohta, University of Maryland School of Medicine, Baltimore), we designed forward and reverse primers in exons 1 and 2,

respectively (see Figure B-2), and amplified cDNA from nurse shark spleen, thymus, muscle, and forebrain using PCR. After confirming the presence of AID in each tissue, we used RT-qPCR to examine relative expression of AID in each tissue. We normalized results against shark muscle tissue using the reference gene β2-microglobulin (β2M) and quantified results using ΔΔCq expression fold changes (see Chapter 2, 'Real-time qPCR for AID expression" for PCR and qPCR details). We ran each sample in triplicate on a single 96-well plate at four cDNA concentrations (6.25, 12.5, 25, and 50 ng/uL) using a Roche Light Cycler PCR machine (see Table B-2). We used the second derivative max analysis option for all samples to generate Cq scores (see Table B-3). We removed one outlier data point from the spleen: β2M (6.25 ng/uL) analysis (indicated by strikethrough text in Table B-2). In Chapter 2, we report results for the 50 ng/uL concentration only.

Fluorescence in situ *hybridization (FISH)*

We embedded fresh frozen adult nurse shark ("Black") thymus tissue in OCT (optimal cutting temperature) medium using standard protocols and stored samples at -80 C until the histology lab at TAMU CVM cut tissue sections. For each tissue, two 8µm thick sections were cut and placed onto a slide using standard cryostat protocols. Every 15th slide was stained using hematoxylin and eosin (H&E). We used H&E stained slides for comparison of tissue architecture in FISH stained slides.

We used the Stellaris RNA FISH protocol for frozen tissue (Biosearch Technologies, Petaluma, CA) to identify areas of expressed TCRα and AID messages in thymus tissue. We designed custom FISH probes against the TCRα constant region for T cell identification and exons 1 and 2 of AID (see Figure B-1 for sequences) using the Stellaris RNA FISH probe designer. We hybridized TCRα with the CalFluor Red 610 fluorophore and AID with the Quasar 670 fluorophore for 16 hours and counterstained with 5ng/mL of DAPI. We photographed multiple images from adjacent sections from FISH (slide #16) and H&E (slide #15) slides at 10x, 20x, and 63x magnification and processed all images using ImageJ software (v1.47). We overlapped adjacent 10x images to create a complete thymic lobe for both FISH-labelled and H&E-stained tissue. We outlined the cortico-medullary junction on the H&E-stained lobe and then transferred this outline onto the FISH-labelled lobe to more clearly distinguish thymic cortex and medulla on the FISH images.

Dr. Caitlin Castro performed all colorimetric *in situ* hybridization experiments prior to the start of this research project.

Figure C-1. Nucleotide and amino acid sequences of nurse shark T cell receptor alpha (TCRα) constant region and activation-induced cytosine deaminase (AID) used in RT-qPCR and FISH experiments.

TCRα constant region
Nucleotide:
nAAAGGGACTCATCTGAACCATCCGTCTACATTCTCCCTCCGTACGACAGCGATACCAAAAATGCTGCATGCCTGGCGA
CTGACTATTTCCCCCAAAATGTCAGCATGGTCGTGGCAGCTGGGAATAAGAAGCAGAAGCAGGACAAATCGAAGGGC
TTGCTATCCACCAATGATCGCAGTTACAGCCTGACAGGATTCCTGGACAAATTGGAGGACCCCAATGACTTTACCTACC
ATGCTGGGAATACTGTCAAACACTTCCCAACAGCTGATCAATTGAAGTACAGCTGCATTAATGTGGAGGAA

Amino Acid:
RDSSEPSVYILPPYDSDTKNAACLATDYFPQNVSMVVAAGNKKQKQDKSKGLLSTNDRSYSLTGFLDKLEDPNDFTYHAGN
TVKHFPTADQLKYSCINVEE

Nurse Shark AID sequence, exons 1 and 2 (partial)
Nucleotide:
nGAAAGAATGTGCGCTGGGCCAAGGGGAGGCACGAGACCTACATGTTGTACATCGTCAAGCGAAGGGATAGTTCCAC
GTCCAGCTCCCTCGACTTCGGCTTCCTCCGCAACAAGCCGAGGCTCCACGCCGAGATGGTGTTCCTCGAGTACCTCGG
AGGGTGGGAGCTGGACCCTCACCGCACCTACCGCCTCACCTGGTTCACCTCCTGG

Amino Acid:
KNVRWAKGRHETYMLYIVKRRDSSTSSSLDFGFLRNKPRLHAEMVFLEYLGGWELDPHRTYRLTWFTSW

Table C-1. Calculations of χ^2 for V gene segments in Table 3-4 of Chapter 3. ($\chi^2 > 3.841$; df=1; $p>0.05$) [O=observed G/C nucleotides; E=expected G/C nucleotides; FR=framework regions; CDR=complementarity-determining regions]

V segment	Hotspot	Region	# G/C	O	E	$(O-E)^2/E$	χ^2	χ^2 *p*
TCRαδV-	DGYW/ WRCH	FR	11008	140	5.9	3033.1	4237.80	0.0000
		CDR	2997	58	2.6	1204.8		
		Inside	14005	198	8.3	4355.9	4463.11	0.0000
	Outside Motif		568931	146	335.7	107.2		
TCRαδV-	DGYW/ WRCH	FR	4012	26	2.3	240.3	295.77	0.0000
		CDR	714	6	0.5	55.4		
		Inside	4726	32	2.9	295.6	310.98	0.0000
	Outside Motif		90762	26	55.1	15.4		
TCRβV	DGYW/ WRCH	FR	5476	16	5.5	20.2	27.70	0.0000
		CDR	925	3	0.7	7.5		
		Inside	6401	19	6.2	26.5	29.94	0.0000
	Outside Motif		49435	35	47.8	3.4		
TCRγV	DGYW/ WRCH	FR	3099	13	3.4	26.5	63.98	0.0000
		CDR	581	4	0.4	37.4		
		Inside	3680	17	3.7	47.6	51.03	0.0000
	Outside Motif		51898	39	52.3	3.4		
IgHV-	DGYW/ WRCH	FR	4604	26	6.0	67.6	166.31	0.0000
		CDR	1995	22	3.5	98.8		
		Inside	6599	48	9.4	159.0	174.55	0.0000
	Outside Motif		67254	57	95.6	15.6		
NTCRV	DGYW/ WRCH	FR	644	3	1.7	1.1	1.48	0.2240
		CDR	102	3	2.1	0.4		
		Inside	746	6	3.6	1.7	2.03	0.1542
	Outside Motif		3664	15	17.4	0.3		
STCRδV	DGYW/ WRCH	FR	1019	2	1.4	0.2	0.37	0.5413
		CDR	279	0	0.1	0.1		
		Inside	1298	2	1.6	0.1	0.13	0.7161
	Outside Motif		7754	9	9.4	0.0		

Table C-2. 96-well plate sample set-up for RT-qPCR. We examined expression in triplicate samples at four cDNA concentrations (top gray row, 6.25 – 50 ng/uL) of each spleen (SPL), thymus (THY), muscle (MUS), and forebrain (FB). We normalized expression levels of activation induced cytidine deaminase (AID) against levels of β2-microglobulin (β2M). We indicate row and column labels in the top row and far left column.

	1	2	3	4	5	6	7	8	9	10	11	12
	50 ng	50 ng	50 ng	25 ng	25 ng	25 ng	12.5 ng	12.5 ng	12.5 ng	6.25 ng	6.25 ng	6.25 ng
A	SPL/B2M	SPL/B2M	SPL/B2M	SPL/B2M	SPL/B2M	SPL/B2M	SPL/B2M	SPL/B2M	SPL/B2M	SPL/B2M	SPL/B2M	SPL/B2M
B	SPL/AID	SPL/AID	SPL/AID	SPL/AID	SPL/AID	SPL/AID	SPL/AID	SPL/AID	SPL/AID	SPL/AID	SPL/AID	SPL/AID
C	THY/B2M	THY/B2M	THY/B2M	THY/B2M	THY/B2M	THY/B2M	THY/B2M	THY/B2M	THY/B2M	THY/B2M	THY/B2M	THY/B2M
D	THY/AID	THY/AID	THY/AID	THY/AID	THY/AID	THY/AID	THY/AID	THY/AID	THY/AID	THY/AID	THY/AID	THY/AID
E	MUS/B2M	MUS/B2M	MUS/B2M	MUS/B2M	MUS/B2M	MUS/B2M	MUS/B2M	MUS/B2M	MUS/B2M	MUS/B2M	MUS/B2M	MUS/B2M
F	MUS/AID	MUS/AID	MUS/AID	MUS/AID	MUS/AID	MUS/AID	MUS/AID	MUS/AID	MUS/AID	MUS/AID	MUS/AID	MUS/AID
G	FB/B2M	FB/B2M	FB/B2M	FB/B2M	FB/B2M	FB/B2M	FB/B2M	FB/B2M	FB/B2M	FB/B2M	FB/B2M	FB/B2M
H	FB/AID	FB/AID	FB/AID	FB/AID	FB/AID	FB/AID	FB/AID	FB/AID	FB/AID	FB/AID	FB/AID	FB/AID

Table C-3. RT-qPCR results.

Spleen:β2M

Plate Position	Sample/ Reference	cDNA (ng/uL)	Cq	Mean Concentration	SD
A1	Spleen/ β2M	50.00	17.84	17.76	0.073
A2			17.69		
A3			17.77		
A4		25.00	18.89	18.85	0.044
A5			18.83		
A6			18.81		
A7		12.50	20.28	20.08	0.180
A8			19.96		
A9			19.99		
A10		6.25	20.93	20.91	0.031
A11			20.89		
A12			~~29.91~~		

Spleen:AID

Plate Position	Sample/ Reference	cDNA (ng/uL)	Cq	Mean Concentration	SD
B1	Spleen/ AID	50.00	26.84	26.84	0.041
B2			26.89		
B3			26.80		
B4		25.00	28.03	28.02	0.034
B5			27.98		
B6			28.05		
B7		12.50	29.09	28.99	0.155
B8			28.81		
B9			29.06		
B10		6.25	30.24	30.06	0.161
B11			29.93		
B12			30.01		

Thymus:β2M

Plate Position	Sample/ Reference	cDNA (ng/uL)	Cq	Mean Concentration	SD
C1	Thymus/ β2M	50.00	20.62	20.51	0.101
C2			20.50		
C3			20.42		
C4		25.00	21.91	21.80	0.107
C5			21.70		
C6			21.80		
C7		12.50	22.71	22.77	0.046
C8			22.79		
C9			22.80		
C10		6.25	23.53	23.59	0.059
C11			23.62		
C12			23.63		

Thumus:AID

Plate Position	Sample/ Reference	cDNA (ng/uL)	Cq	Mean Concentration	SD
D1	Thymus/ AID	50.00	30.53	30.36	0.145
D2			30.30		
D3			30.26		
D4		25.00	31.50	31.68	0.164
D5			31.82		
D6			31.73		
D7		12.50	32.44	32.45	0.010
D8			32.46		
D9			32.45		
D10		6.25	33.17	32.97	0.169
D11			32.86		
D12			32.89		

Muscle:β2M

Plate Position	Sample/ Reference	cDNA (ng/uL)	Cq	Mean Concentration	SD
E1	Muscle/ β2M	50.00	23.15	23.09	0.101
E2			23.14		
E3			22.97		
E4		25.00	24.07	24.08	0.069
E5			24.15		
E6			24.01		
E7		12.50	25.15	25.08	0.124
E8			24.94		
E9			25.16		
E10		6.25	25.89	25.89	0.057
E11			25.83		
E12			25.95		

Muscle: AID

Plate Position	Sample/ Reference	cDNA (ng/uL)	Cq	Mean Concentration	SD
F1	Muscle/ AID	50.00	32.61	32.69	0.097
F2			32.80		
F3			32.68		
F4		25.00	33.69	33.53	0.679
F5			32.79		
F6			34.13		
F7		12.50	33.33	34.02	0.617
F8			34.51		
F9			34.22		
F10		6.25	34.46	34.80	0.587
F11			34.46		
F12			35.48		

Forebrain:β2M

Plate Position	Sample/ Reference	cDNA (ng/uL)	Cq	Mean Concentration	SD
G1	Forebrain/ β2M	50.00	20.85	20.83	0.029
G2			20.80		
G3			20.84		
G4		25.00	21.90	21.81	0.090
G5			21.81		
G6			21.72		
G7		12.50	22.73	22.68	0.048
G8			22.66		
G9			22.64		
G10		6.25	23.67	23.70	0.029
G11			23.73		
G12			23.70		

Forebrain:AID

Plate Position	Sample/ Reference	cDNA (ng/uL)	Cq	Mean Concentration	SD
H1	Forebrain/ AID	50.00	33.29	33.31	0.319
H2			33.64		
H3			33.00		
H4		25.00	34.17	34.60	0.762
H5			34.14		
H6			35.48		
H7		12.50	35.71	34.56	1.309
H8			33.14		
H9			34.84		
H10		6.25	35.01	35.53	0.560
H11			35.47		
H12			36.12		