THE UNIVERSITY OF CHICAGO

THE MOLECULAR BASIS OF SELF-ANTIGEN RECOGNITION BY REGULATORY T CELLS

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

COMMITTEE ON IMMUNOLOGY

BY

RYAN KENDALL DUNCOMBE

CHICAGO, ILLINOIS

JUNE 2021

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

A well-known scientist once wrote that “if I have seen further it is by standing on the shoulders of giants.” An appropriately grandiose statement for Newton, the discoverer of gravity and inventor of calculus, but perhaps too dramatic for the graduate student of today. Nonetheless, the underlying meaning remains relevant: scientific progress is cumulative. No scientist would succeed without those that come before them and those that collaborate with them, and I am no exception. For the purposes of space, I will keep these acknowledgments to only the mentors, collaborators, and friends who have most immediately enabled my success, though there are many more who deserve thanks. Listed here are the people upon whose shoulders I most directly stand.

I would first like to thank my mentor Dr. Erin Adams, without whom this work and my development as a scientist would not have been possible. Her scientific mentorship was invaluable to the direction of my project(s), and her command of the complexities of both immunology and biochemistry is a rare quality that I am grateful to have learned from and will continue to be inspired by. Erin’s relentlessly positive mentorship kept me afloat through many difficult times over the years. In those times, she knew what I needed to hear to keep motivation high, and it made graduate school much more enjoyable than it could have been with lesser mentorship. For this and for her scientific mentorship I will always be grateful.

I was fortunate to work closely with Dr. Peter Savage and his lab throughout my time here, and so I thank him for his close mentorship as a collaborator and as one of my committee members. I’m so grateful to have boarded the “Treg Train” years ago and to have been a part of its ongoing story. I want to thank my other committee members of the past and present, Albert Bendelac, Seungmin Hwang, Sean Crosson, and Andrew Koh, for their advice and mentorship,

which all improved this work significantly. And additional thanks to Andrew for joining my committee at the 11th hour and yet providing such valuable input.

I owe many thanks to the scientists without whom this work would not be possible. First, thanks to John Leonard and Dana Gilmore for their contributions to this story, which immediately preceded and enabled my own. Thanks to David Klawon, Jaime Chao, and Donald Rodriquez, the Savage lab members who conducted all of the mouse work described in this thesis, and conducted a marathon sort through the night to generate our 10x dataset. I have benefited tremendously from knowing and working with all of my lab mates, but my greatest thanks go to Sobhan Roy, my collaborator and friend, for his biochemical contributions to this project. The final two figures would not have been possible without him.

I want to thank the entire COI community – the administrators, the professors, staff, and students who make this community what it is. I'm also thankful for the NIH funding that enabled this work. I am grateful to live in a society that recognizes scientific advancement as essential to the collective good and funds science accordingly. Although the animal footprint of my work was only a fraction of that of most immunology projects, I am also thankful for the animal lives that were sacrificed for this project in an effort to better understand human health and disease.

I have endless thanks for the mentorship, advice, and camaraderie of my fellow Adams lab members, past and present. Adrienne Luoma, John Leonard, May Gu, Charlie Dulberger, Chris Boughter, Marta Borowska, Caitlin Castro, Kristof Nolan, Sophie Krahnke, Sobhan Roy, Augusta Broughton, Cole Ladd, Sofi Maltseva, Amrita Ramesh, Philipp Ross, Charlotte James, Sean Ryan, and Mayuri Viswanathan – each of them has been a pleasure to work with every day and have made this workplace an easy one to spend so much time in. I'm grateful to have been able to call each of them coworkers and friends. I want to give special thanks to John Leonard,

Sobhan Roy, Chris Boughter, and Caitlin Castro, four brilliant scientists I've worked with the closest and learned from most directly. Knowing each of them has made me a considerably better scientist and made my graduate career significantly more enjoyable. I'm also thankful to my undergraduate mentees, Sam Dubensky and Tianyang Mao, both of whom are far brighter than I am and probably taught me as much as I taught them. I both enjoyed and learned from our time together and look forward to seeing what they accomplish in the coming years.

Possibly my single favorite thing about graduate school has been meeting fellow graduate students – both within the Committee on Immunology and without. Never have I met so many interesting people who were so like-minded and yet so diverse in background and thought. The only shame is I never had enough time to spend with all of them. Nevertheless, I still made many, many lasting friendships. I would like to thank Kyle Cron, Brandon Lee, Sangman Kim, Toufic Mayassi, and Kevin Lei for their friendship and advisement as students who came before me. Thanks to Jen Allocco, Dave Klawon, Steven Erickson, Hailey Brown, and Tiffany Marchell for their friendship and companionship through triathlon training – a constant source of fun, health, and positivity. Thanks to Brendan MacNabb and the other members of the Windy City Red Sox for seven years of Chicago baseball. Thanks to all members, past and present, of the Costims softball team, an unlimited source of summer fun. To my all-the-time friends, Steven Erickson, Chris Stamper, Matt Zurenski, Andrew Tremain, Amelia Joslin, Jaime Chao, Jess Fessler, Hailey Brown, Alex Hoffman, Shan Kasal, Brendan MacNabb, Lydia Varesio, Liana Hernandez, Ittai Eres, Becca Barker, and Damai Vergara-Hegi, I say thanks for the support, understanding, happiness, and meaningfulness our relationships have brought me. Special thanks to the matriculants of 2014 – this close group of friends and colleagues has made all the difference and I look forward to a lifetime of friendship and accomplishments with them all.

I would also like to thank my parents Kevin and Barbara, for shaping who I am today and for nurturing my scientific interests. When I would ask endless, sometimes ridiculous questions of my father as a child ("why is the sky blue?" etc.), he would log on to our newfangled personal home computer after I was put to sleep and find the answer. He would print a summary and leave it for me on the breakfast table to read before school the next day. It may sound idealized, but I credit this habit with teaching me that answers are out there if one only looks for them, resulting in an immediate appreciation for my science classes that I did not find for other subjects until much later in life. And my mother worked hard ceaselessly and sacrificed much to ensure I had the best education I could want. I never appreciated her endless involvement in PTAs, sports teams, and her "encouragement" for me to improve my grades, but she often said I would thank her later for it all, and she was right.

I also would like to thank my other family members – the Duncombe, Durbin, and Licona clans are tight-knit groups, and a more perfect support system and loving family environment must be hard to find. My favorite thing about living in Chicago has been its proximity to Michigan, where a large contingent of my family resides or migrates to frequently. I'll never have enough time with any of them, but being closer has made holidays and the occasional weekend a time of restfulness, laughter, and love. I owe special thanks to the members of my nuclear family, Sydney, Korey, Ben, and Tash. My family is the center of my life, and if I had to rank life's pleasures, any time spent with them would be at the top of the list.

Finally, I need to thank my partner of five years, Jessica Fessler. Her enduring love and support have been a constant source of comfort, happiness, and inspiration. She is perhaps the one exception to my "standing on shoulders" metaphor, as I'm not so sure I'm standing on her shoulders as much as I am riding her coattails. I'm doing my best to keep up.

ABSTRACT

Foxp3-expressing CD4+ regulatory T (Treg) cells are critical for the prevention of autoimmunity and the maintenance of immune homeostasis. The recent identification of self-peptide/MHC class II ligands recognized by thymic-derived Treg cells by our group and the Savage group has set the foundation for new studies into self-antigen recognition by Treg cells. We analyzed the αβ T cell receptor (TCR) repertoire of Treg cells and T conventional (Tconv) cells reactive to a peptide (named "C4") derived from the prostate-specific protein Tcaf3 complexed with I-A^b using a droplet-based sequencing approach. We conducted sequencing in both C4-sufficient wild-type mice and mice with a targeted deletion of the C4 peptide, allowing us to analyze a "foreign" TCR response to the same epitope. We found negligible overlap in Treg and Tconv TCR repertoires for this single antigen and identified an abundant public C4-specific Treg cell clone in wild-type mice that is absent in C4-deficient mice. Through biochemical studies of both Treg and Tconv TCRs found in this dataset, we found that C4/I-Ab-specific Treg TCRs recognized ligand with higher affinity and longer half-life compared to TCRs expressed by C4/I-Ab-specific Tconv cells, supporting a TCR-intrinsic model of Treg cell development in which TCR-peptide/MHC-II interactions of higher stability promote Treg cell development. Transcriptional analysis of C4-specific T cells also revealed genetic markers and properties of Treg and Tconv cells elicited in both wild-type and C4-deficient mice, revealing candidate genes that may be important for optimal Treg cell fitness and function. Thus, our cumulative evidence supports a model in which TCR-self-peptide/MHC-II interactions of high stability promote Treg cell differentiation in the thymic medulla, leading to the establishment of distinct peripheral pools of antigen-specific Tconv and Treg cells.

Chapter 1: Introduction

1.01 T cells: Detecting Self vs. Non-self

The mammalian immune system is a highly complex network of cells, signaling molecules, and receptors that function together to distinguish self from non-self and maintain homeostasis. Furthermore, the immune system can distinguish which non-self agents are potentially harmful from those that aren't and permit their symbiotic relationship with the mammalian organism, such as the commensal microbiota that exists in the human gut, among other locations.

While the immune system relies upon all of its many component parts to maintain proper function, one of the principal immune cell populations responsible for its exquisite sensitivity to foreign agents are T cells. T cells are named so because they develop in the thymus, and each T cell is distinguished from other immune cells by the presence of a unique T cell receptor protein (TCR) expressed on the cell surface that can recognize small molecule antigens in the context of MHC proteins. TCR genes undergo a highly complex and refined genetic recombination process called VDJ recombination to produce one of a possible 10^{18} different T cell receptor proteins[1]. Each T cell will express one unique receptor (with some rare cells expressing two receptors), which then undergoes a complex selection process in the thymus.

T cell receptors come in two varieties, αβ T cells and γδ T cells, each of which has unique evolutionary roles. αβ T cells, the primary focus of this thesis, come in many sub-varieties themselves, the broadest classes of which are CD4+ and CD8+ αβ T cells. CD4+ T cells have TCRs that recognize antigens in the context of class II MHC proteins, whereas CD8+ T cells have TCRs that recognize antigens in the context of class I MHC proteins. Class I antigens are traditionally considered to be sourced from the cytoplasm, while class II antigens are

traditionally considered to be sourced from extracellular regions through phagocytosis. However, many exceptions to these rules exist and antigens for both classes of MHC can be derived from a myriad of sources.

Of the CD4+ αβ T cell, or helper T cells (T_H cells), there are four primary recognized inflammatory subsets. These are each specialized for a particular purpose. T_H1 cells help control intracellular bacteria and viruses. T_H2 cells primarily control extracellular parasitic infections and IgE-mediated responses and are therefore a significant component in allergies as well. T_H17 cells function to control extracellular bacteria and fungi and are especially crucial in the gut and mucosal immune compartments which interface with the microbiota. Finally, T follicular helper cells, or T_{FH} cells, function to provide B cell help in lymph nodes, a critical step in promoting germinal center responses for the generation of antibodies.

The many subsets of αβ T cells work in concert with each other and with other immune and somatic cells throughout the body to maintain homeostasis. This includes not only the defense of the organism against pathogen invaders of all kinds, but also tissue repair and regulation. Thus, the immune system must exist in a balance between pro-inflammatory and anti-inflammatory signals. Upon challenge to that homeostasis, such as in the case of infection, damage, cancer, or other insults, immune cells must act in the proper time and location to tilt the balance in the proper direction to restore homeostasis. This is often described through a metaphor of applying “gas vs. brakes” on the immune system. While most αβ T cell subsets would properly be categorized under the proinflammatory, or “gas”, section of the immune system, a population of αβ T cells called regulatory T cells function as the primary anti-inflammatory class, or “brakes”, of the adaptive immune system.

1.02 Regulatory T cells

Regulatory T cells (Treg cells) are an integral cell population for maintaining immune homeostasis and are characterized primarily by the expression of the master regulator transcription factor forkhead box P3 (Foxp3)[2,3]. Individuals with mutations in Foxp3 exhibit dysfunctional Treg populations and widespread autoimmunity with high rates of morbidity[4]. In mice, a Foxp3 mutation is responsible for "scurfy", which recapitulates the symptoms of IPEX syndrome and is lethal within the first few weeks of life[5]. Furthermore, Treg cells are required throughout life, as evidenced by murine models in which Treg depletion can be induced at specific time points. Removal of Treg cells in these models invariably results in widespread autoimmunity and death within two weeks[6].

Foxp3+ regulatory T cells comprise approximately 10% of the circulating CD4+ T cell pool and were first identified as cells expressing high levels of CD25[7]. Treg cells are a heterogenous class of cells essential in regulating immune responses to self-, foreign-, and tumor-associated antigens throughout the body[8]. They are exclusively CD4+ αβ T cells and recognize peptide antigens of diverse origins presented by MHC class II molecules. Like conventional T cells (Tconv), Tregs bind their cognate antigens using a unique, genetically recombined TCR. Signaling through this TCR is required for Treg development, recruitment to peripheral tissues, their maintenance in those tissues, and their suppressor function[9,10].

There are two primary categories of Tregs relevant to the work in this dissertation. Thymic-derived Tregs (tTregs), or natural Tregs, are directed into the Treg lineage during their development in the thymus. They then egress from the thymus and perform their suppressive function in the periphery. Peripheral Tregs (pTregs), or sometimes "induced Tregs", on the other hand, exit the thymus as Tconv cells and do not develop into Treg cells until encountering

antigen under particular environmental conditions in the periphery[11]. pTregs often develop in response to TGF-β, IL-2, and retinoic acid in the gut, where they are critical for maintaining homeostasis between the immune system and gut microbiota[12]. pTregs can also develop in reaction to inflammatory responses to foreign antigens[13]. Identifying markers that can faithfully distinguish these two Treg populations has proven difficult, as Helios and Neuropilin-1, two candidate markers, have been widely refuted as reliable markers[14,15]. While pTregs have been shown to be primarily driven to foreign antigens, there are some cases, such as tumor antigens, where they can be self-reactive[11]. Thus, while it can be difficult to distinguish pTregs from tTregs, tTregs are the primary Treg lineage relevant to the work in this dissertation and will be the default reference denoted by the term "Treg".

Traditionally, tTregs are considered a stable cell population in the periphery, though they require continuous TCR stimulation for their maintenance and suppressor function[9,10]. They accomplish this primarily through the stable expression of Foxp3, which is stabilized by the Treg-specific demethylated region (TSDR), a conserved non-coding region within the *Foxp3* locus specific to Treg cells[16]. While the TSDR maintains a stable regulatory phenotype, some plasticity does exist within the Treg lineage. Recent research has demonstrated that in certain inflammatory contexts, Tregs can acquire effector T helper-like phenotypes that enable specialized suppression of a given Th subset[17]. In these instances, Tregs may acquire expression of the master transcription factor of the Th population they are suppressing, such as T-bet for Th1, GATA3 for Th2, and RORγt for Th17 responses[18]. These are thought to aid in the permissiveness of these specialized immune responses as well as increase Treg suppressive ability by increasing colocalization with the respective inflammatory subset.

1.03 Mechanisms of Treg cell suppression

While the many mechanisms of tolerance work together to prevent autoimmunity, active Treg-mediated suppression still represent an irreplaceable and crucial mechanism of tolerance, as evidenced by the rapid onset of autoimmunity upon Treg depletion. Thus, Tregs function as a master regulator of global immune landscapes and are critical for maintaining immune homeostasis. Tregs accomplish this with the ability to suppress a wide variety of immune cells, including, but not limited to, CD4+ T cells, CD8+ T cells, B cells, NK and NKT cells, dendritic cells, and monocytes[19–21]. Treg cell-mediated suppression of autoreactive CD4+ T cell responses is critical for preventing autoimmunity throughout life. However, this suppression can also prevent proper immune responses to tumors or foreign pathogens. Therefore, Treg and Tconv cells must exist in a fine balance between inflammatory and suppressive behavior. Treg-mediated suppression of Tconv cells can occur through a myriad of mechanisms, including secretory signals, cell-cell contact, and indirect mechanisms by Treg prevention of dendritic cell maturation or activation[22,23]. The exact balance and role of these signals in the suppression of any given response *in vivo* remains the subject of much debate.

Secretory signals by which Treg cells can mediate suppression include the anti-inflammatory cytokines IL-10, TGF-β, and IL-35, among others[24–26]. TGF-β and IL-10 are both important factors for Tconv suppression as well as Treg induction. IL-35, an IL-12 family member, is a more recently described immunosuppressive cytokine that has been found to prevent early activation of Tconv cells and expand Treg cells, thereby tipping the balance of an immune response in an anti-inflammatory direction[25]. Tregs can also secrete granzyme-B and perforin to directly induce apoptosis in effector T cells[27,28], and their high expression of CD25 enables them to function as an IL-2 sink, preventing Tconv cells from receiving activation

signals[29]. However, this “IL-2 sink” function of Tregs remains controversial, as some data indicate it might contribute minimally to overall Treg suppression[30]. What is known, at least, is that Tregs require continuous IL-2 signaling for their suppressor function, so while the mechanism may be debated, IL-2 is critical for proper Treg function[31]. Thus, via this wide variety of cytokines, receptors, and secretory molecules, Treg cells can generate and maintain an anti-inflammatory cytokine milieu.

Dendritic cells (DCs) play a critical role in coordinating and facilitating immune responses. Thus, the ability of Treg cells to modulate dendritic cell function is a substantial contributor to their regulatory function. Among the primary means by which Tregs modulate DCs is through the use of CTLA-4, the blockade of which results in spontaneous autoimmunity which can be rescued by Tregs[32,33]. CTLA-4 is constitutively expressed on Tregs and binds the CD80 and CD86 costimulatory molecules expressed by dendritic cells[34]. It does so with significantly higher affinity than CD28, which is expressed by naïve T cells and Tconv cells. Thus, CTLA-4 can indirectly inhibit activation of Tconv cells through competition for these signals or by the transendocytosis of CD80 and CD86, thereby removing these receptors from dendritic cell surfaces[35]. In addition to these direct effects by which Tregs can reduce DC antigen presenting activity by DCs and other antigen presenting cells, the cytokines secreted by Tregs can also support further immunosuppressive cytokine secretion by APCs, namely by increasing IL-10 production and decreasing inflammatory cytokine production[36].

An important distinguishing factor between these mechanisms of suppression is which mechanisms are TCR specific vs. TCR non-specific. One study removed all Treg specificities save one skin-antigen specific Treg and found widespread suppression of autoimmunity compared to Treg-deficient mice[37]. While rampant autoimmunity did persist, this was a powerful

demonstration of the abilities of non-specific Treg suppression. Other studies have also shown that Tregs are able to reduce Tconv contacts with APCs[38]. This further suggested that, in general, the state of immune responses is dictated by a "numbers game", with the ratio of activated Treg cells vs. Tconv cells determining the inflammatory status of a tissue.

1.04 T cell selection

Developing T cells undergo a rigorous selection and education process in the thymus. The vast numbers of potential T cell receptors (>10^{18} receptors) that can be generated by VDJ recombination requires multiple rounds of selection to ensure that fully developed thymocytes will be maximally effective in maintaining homeostasis in the periphery. These processes include both positive selection and negative selection of developing thymocytes, and these two selection steps occur in a temporally and spatially distinct manner.

Positive selection occurs in the outer cortical region of the thymus and serves to select thymocytes with functional T cell receptors. Positive selection of αβ T cells requires peptide ligands to be presented by MHC molecules, but the affinity with which this recognition occurs is thought to be extremely low, and difficult to measure experimentally[39]. Thus, little is known about positively selecting ligands. The largest proportion of the pre-selection repertoire is removed at this step by failing to achieve positive selection, and up to 85% of thymocytes exhibit death by neglect at this stage[39–41].

Negative selection is classically considered to occur in the inner medulla of the thymus and functions to remove thymocytes with highly self-reactive TCRs that would be potentially dangerous in the periphery[42]. These could represent TCRs with high affinity MHC which might therefore exhibit high cross-reactivity[40]. Some cross-reactive cells can escape deletion by being diverted into the intraepithelial lymphocyte lineage instead[43]. The medulla is also responsible for

diverting some thymocytes into the Treg lineage, which will be expanded upon in a subsequent subchapter.

All developing T cells undergo multiple stages of development within the two major structures of the thymus – the cortex and the medulla. They begin in the cortex as "double negative" thymocytes and express neither the CD4 nor the CD8 coreceptors, thereby preventing any TCR signaling. At this stage, the TCRβ chain undergoes rearrangement and expresses on the cell surface paired to a pre-Tα chain (pTα), a truncated placeholder of the TCRα chain. This receptor complex does not require ligand for successful stimulation, although recent research indicates a possibility for some peptide-independent antigen sampling of pre-TCRs[44]. Upon successful expression of the TCRβ/pTα complex, thymocytes will then express fully rearranged TCRα genes and upregulate both coreceptors, allowing the developing cells to determine if they express a functioning TCR – if no TCR stimulation occurs at the double positive stage, then the TCR is nonfunctional and the thymocyte will die by neglect. Abundant MHC molecules of both classes are present in the cortex and present ubiquitously expressed antigens for the purposes of this "positive selection" of functional TCRs. After successful activation at the double positive stage, however, thymocytes will downregulate their coreceptors sequentially to determine if signaling occurs through either MHC I- or MHC II-dependent mechanisms. It is at this stage that thymocytes differentiate into CD8 (for MHC-1) or CD4 (for MHC-II) T cells. Negative selection of thymocytes with TCRs that strongly ligate pMHC can also occur at this stage in the thymic cortex.

Importantly, upon successful TCRβ rearrangement and expression, further TCRβ rearrangement is prevented by allelic exclusion. The RAG1/2 recombinase complex responsible for VDJ recombination is phosphorylated and degraded, preventing further TCRβ gene

rearrangement. During this time, the thymocyte undergoes cell division for several days. This ensures that each cell with a functional TCRβ rearrangement can give rise to many cells before TCRα rearrangement begins, increasing the efficiency of thymic development. Thus, a single TCRβ chain can be associated with multiple TCRα chains in a given individual.

Furthermore, rearrangement of the TCRα locus continually proceeds until positive selection occurs. Each locus can undergo multiple rearrangement events, and the two loci can rearrange independently. This means that in a small number of cases, two in-frame TCRα rearrangements can occur in a single cell. However, despite this, it is unlikely that the two chains both are stably expressed on the cell surface and functionally bind pMHC, so most dual-α expressing cells remain specific for a single antigen. Up to 10% of αβ T cells have been observed to express two TCRα chains[45]. In a very small number of cases, the second TCRβ locus can rearrange before the suppression of the RAG1/2 complex, allowing for the dual expression of T cell receptor beta chains on the cell surface.

After positive selection of cells expressing mature αβ TCRs, thymocytes exit the cortex and enter the medulla. In the medulla, thymocytes encounter a myriad of antigen presenting cells and are further selected and educated before eventually leaving the thymus. The entire thymic selection process can take more than three weeks in mice, with up to two weeks of that time spent in the medulla to allow time for ample pMHC sampling by each TCR.

Canonically, negative selection was thought to primarily occur in the medulla and to be a major mechanism of central tolerance, but recent research has demonstrated both negative selection in the cortex and less complete negative selection of self-reactive clones[46]. Early experiments investigating negative selection used both experimental and naturally occurring self-peptides injected into thymuses in TCR-transgenic mice[47,48]. These injection experiments

introduced far greater concentrations of antigen that typically present in the thymus and, combined with increased TCR presence due to transgenic expression, induced widescale deletion. Subsequent experiments in TCRβ-transgenic mice have shown much more modest frequencies of deletion[49–51]. Multiple corroborating studies have recently shown significant deletional escape of self-reactive clones in more physiological models, calling into question the dominant position to which negative selection was given in early models of central tolerance[39,51,52]. Nonetheless, it remains clear that highly self-reactive or cross-reactive TCR clones can induce deletion at the double positive and single positive stages of thymic development.

1.05 AIRE and Tissue Restricted Antigens

One of the most critical cell subsets in the medulla are the medullary thymic epithelial cells (mTECs) that produce certain antigens specifically for the purposes of thymocyte selection. The thymus faces a difficult evolutionary problem: it must educate thymocytes on antigens present throughout the body without them leaving the thymus. Evolution's solution to this problem is the transcription factor AIRE, short for "autoimmune regulator". AIRE stochastically induces transcription of mRNA for thousands of different tissue-restricted proteins from throughout the body in small amounts within mTECs[53]. Each mTEC only expresses low levels of a small number of these tissue-restricted antigens (TRAs), preventing these cells from actually acquiring the characteristics of other tissues[54]. Furthermore, the stochastic nature of Aire results in a wide variety of TRA expression across mTECs. As a population, mTECs express the widest range of genes of any cell type (>85% of all protein-coding transcripts), but any two mTECs are unlikely to express the same proteins[54,55].

These mTEC-derived TRAs are harvested by thymic antigen presenting cells and presented to the developing thymocytes. If a thymocyte experiences activation in response to these TRAs, it is known to be self-reactive to a protein found somewhere in the organism's body. It has been the subject of much debate what happens to self-reactive thymocytes in the medulla, but much evidence now indicates that many such thymocytes are diverted into the Treg cell lineage, often in an AIRE-dependent manner[52,53,55,56].

The development of many Treg specificities is dependent on the activity of AIRE. These AIRE-dependent Treg cells have been determined to play essential roles in regulating murine and human disease, as evidenced by the systemic autoimmunity observed in $AIRE^{-/-}$ mice and in human APS1 patients with loss of function mutations in AIRE[57]. Recent work by the Savage lab has demonstrated that AIRE can drive T cells expressing TCRs that are specific for self-pMHC into the Treg lineage[52,56]. These experiments showed that TCRs that are typically found on Tregs in wild-type mice are instead found on conventional T cells in $AIRE^{-/-}$ mice. These Tconv cells expressing Treg TCRs were called "Trogues" to connect their alternative development. This finding was instrumental in demonstrating that, at a minimum, Aire expression does not function exclusively as a method of negative selection, but can frequently function as a method of driving tolerance via Treg development. It also highlights the established differences between Treg and Tconv TCR repertoires, which have previously been shown to be largely distinct, with a maximum reported overlap of 10% in TCR receptor usage between the two repertoires[58]. Thus, AIRE-dependent expression of TRAs in the thymus drives the development of self-antigen reactive T cells into the Treg lineage and is critical for preventing autoimmunity.

1.06 T cell tolerance

These mechanisms of T cell selection combine with other methods to comprise multiple mechanisms of immune tolerance. These include the thymus-intrinsic, or central-tolerance, mechanisms of deletion and Treg cell development, and the thymus-extrinsic, or peripheral tolerance, mechanisms of T cell ignorance, T cell anergy, and the development of peripheral Treg cells (pTregs). Tolerance can also be divided into dominant forms of tolerance, which can actively suppress autoimmune reactions, and recessive forms of tolerance, which are achieved by the removal of autoimmune inducing cells. Dominant tolerance is best illustrated by Treg cells, which actively suppress self-reactive immune responses and can be transferred to Treg-deficient mice to induce tolerance. Recessive tolerance is best illustrated by thymic deletion, in which auto-reactive T cells are deleted from the circulating repertoire.

In addition to thymic deletion, another mechanism of central tolerance is the development of Tregs, which possess dominant immunosuppressive abilities that are crucial for preventing autoimmunity throughout life and can be transferred from mouse to mouse. Additional mechanisms of central tolerance exist for B cells and will not be discussed here.

While Tregs can actively suppress self-reactive Tconv cells, T cell ignorance is a critical tolerogenic phenomenon in which autoreactive effector T cells escape into the periphery but encounter sufficiently low levels of antigen to prevent activation and inflammation. T cell ignorance has been observed in a variety of models and is known to be a significant contributor to peripheral tolerance, though certain circumstances, such as inflammation or infection, can break this tolerance as antigen levels may increase and correlate with costimulatory and inflammatory signals[51].

Should a Tconv cell escape deletion, avoid diversion into the Treg lineage in the thymus, and bind its cognate antigen in the periphery, it will encounter yet another mechanism of tolerance in T cell anergy. T cell anergy is a quiescent state generated when Tconv cells experience activation of the TCR in the absence of costimulatory signals. In such contexts, the activated T cells are rendered unresponsive, and in some cases can develop into pTregs[59].

The synthesis of these different mechanisms of tolerance to maintain immune homeostasis was demonstrated in recent research by expressing neoantigens under multiple tissue-restricted promoters with distinct expression patterns. In one study, cre-recombinase was expressed under a ubiquitous promoter as well as multiple tissue-specific promoters[51]. The CD4+ T cell compartment was quantified with engineered cre:I-A^b tetramers. In the case of a ubiquitously expressed antigen, up to 60% of the repertoire experienced deletion – much lower than early estimates with thymically injected antigens, which likely overloaded the thymus and thus increased the observed avidity by T cells – thereby inflating deletion estimates. Tolerance, in this case, was observed to be mediated primarily through anergy and could be broken by repeated injections of antigen and adjuvant. In the case of tissue-specific antigens in this study, very little deletion was observed, and Treg suppression was found to be the primary means of maintaining tolerance, with T cell ignorance also playing a role. Importantly, this Treg suppression was antigen-specific, and deletion of Tregs induced inflammation. Thus, a variety of tolerance mechanisms work in non-redundant roles to maintain immune homeostasis throughout the body.

1.07 Treg development

In the case of Treg development and tolerance, it is unlikely that a single switch exists to determine Treg vs. Tconv development of thymocytes. Nonetheless, a model based on TCR

signaling strength has emerged as a common paradigm for explaining thymocyte selection and development[60–63]. In this model, TCR signaling strength is determined by the affinity of TCR/pMHC interactions in the thymus. In this model, high-strength TCR interactions induce negative selection by apoptosis, intermediate-strength interactions induce Treg development, and low-strength interactions induce positive selection of Tconv cells. The lowest-strength interactions, or a lack of interactions entirely, fail to induce positive selection of any kind and such thymocytes experience death by neglect.

Early models used a variety of techniques to assess the relationship between TCR affinity and Treg development. These include the use of Nur77^{GFP} mice to correlate TCR signaling strength with developmental outcome[41]. In the thymus, Treg cells in this model expressed higher levels of GFP than did Tconv cells, indicating they exhibited stronger TCR activation. Additionally, a wide variety of transgenic models have supported the affinity hypothesis. These include TCRs of varying affinity for foreign antigens which are then introduced into the thymus artificially[64]. TCR activation, as measured by CD69 and CD5 expression, and Treg development were both found to increase in thymocytes bearing TCRs of higher affinity for antigen. While these studies indicate a potential threshold of affinity in Treg development, their use of introduced antigens obscures the physiological level and precision of that threshold. Intrathymic injections or introduced promoters cannot recapitulate the sensitive tuning of natural self-antigens in the thymus.

In recent years, the affinity model has been modified and added to, as the importance of the avidity and kinetics of TCR:pMHC interactions and the context of the cytokine milieu have gained appreciation in thymic Treg development. Catch bonds, which can impart differential activation to TCRs with identical affinities, have been proposed as potential factors as well[65].

While the relative thresholds of affinity or avidity have been proposed for negative selection, Treg development, and Tconv development, it is important to note that no precise measurements have been conducted to ground these relative thresholds[60]. This is primarily due to a lack of known selecting ligands in the thymus, which has represented a significant knowledge gap in the field of thymic development. Thus, if a threshold of activation strength is present for the development of self-reactive thymocytes into Treg vs. Tconv cells, it is not only unclear where that boundary occurs, but it is also unclear how sharply demarcated the boundary would be.

Importantly, the discovery of specific naturally occurring TCRs that, when expressed in T cells via transgene, can induce regular and stable development into the Treg lineage without antigen introduction solidified TCR-intrinsic characteristics as a key determinant in Treg development[64]. However, this only occurs when the transgenic T cells are introduced into a thymus at a limited frequency, as mice with the fully transgenic repertoire exhibited minimal Treg differentiation compared to the Tconv compartment. This indicates that the Treg niche is limited in scope, possibly by IL-2 signaling, which has been shown to be critical for Treg development in the thymus. Thus, for a given self-antigen, the TCR repertoire might be determined in part by the biochemical properties of its TCRs as well as TCR-external factors such as limited cytokine signaling.

These differences in TCR development in different thymic contexts and the established minimal repertoire overlap between Treg and Tconv repertoires strongly suggest that TCRs with a specificity for self-antigen can drive development into the Treg lineage. Early suggestions included the possibility that Treg TCRs bound peptide MHC in distinct docking modes compared to Tconv TCRs[66]. In a recent paper from Beringer et. al, crystal structures were reported of an *in vitro*-induced “iTreg” TCR in complex with MHC, showing a reversed docking

mode relative to all previously published conventional αβ TCR/MHC II structures[67]. However, this TCR was derived from conventional T cells polarized to become Tregs *in vitro* and is responsive to foreign antigen, and thus it is not reflective of *bona fide* Treg TCRs that drive thymic Treg development in response to self-antigens.

Additional hypotheses proposed that Treg TCRs might have innately higher affinity for MHC. For example, a paper from Stadinski et. al used a fixed TCRβ chain with large, hydrophobic residues at positions 6 and 7 of the complementarity-determining region CDR3β to show robust development of self-reactive regulatory T cells[68]. However, the two TCRβ chains utilized for these studies were originally identified from TCRs that were reactive to a foreign antigen and induced high self-reactivity in transgenic mice, and thus are unlikely to be indicative of natural Treg behavior, which has since been shown to be highly antigen-specific[66].

In another paper from Stadinski et. al, multiple self-peptides derived from the protein Padi4 were found to be recognized by neonatal Treg cells, but the Treg cells did not persist into adulthood[50]. Multiple TCRs derived from negatively selected T cells, Tconv cells, and Treg cells specific for the Padi4 peptide presented by I-A^b were identified, crystallized, and the kinetics of their interactions were measured. The TCRs that induced Treg development were found to have moderate dwell times for pMHC (defined as the half-life of the interaction upon dissociation), whereas the negatively selected TCRs exhibited long dwell times, and the Tconv TCRs exhibited short dwell times. Affinity was not found to correlate with cell fate. Furthermore, the crystal structures of these TCRs demonstrated that Treg TCRs used conventional docking orientations and revealed subtle changes in the strength of molecular interactions that could explain the differences in dwell times. Thus, TCR dwell time, or half-life, was concluded to be the primary determinant of TCRs that drive Treg development vs. Tconv development and deletion.

While this research was instrumental in showing the conventional docking orientation of Treg TCRs and the potential for subtle kinetic differences in Treg vs. Tconv development rather than drastic differences such as reverse docking orientations or broad self-reactivity, this research was limited by a number of factors. First, the Treg cells studied in this model did not persist into adulthood, indicating the presence of an unknown thymic temporal switch driving these cells to negative selection during adulthood. Second, these studies, like those conducted previously, were conducted in mice with fixed TCRβ chains. As a result, the TCRs that were ultimately crystallized varied only by as few as one amino acid in their total CDR loops, since they bore identical TCRα chains and similar CDR3α sequences. Thus, it remains unknown whether the TCRs in this study are representative of Treg TCRs as a whole, and more studies of Treg vs. Tconv development with diverse TCRs are needed.

Another recent paper used self-antigen tetramers to track naïve and tolerant T cell repertoires and revealed a TCR hierarchy based on affinity[49]. However, this affinity was only measured via *in vitro* activation assays, and thus only informs the relative avidities of the interactions in the experiment. It does not provide an experimentally determined K_D value for comparison with other TCRs.

Affinity measurements of natural Treg TCRs with their cognate antigens will directly address this issue; if within the range of conventional TCR affinities, these results will strongly suggest that it is the presence of antigen, not innately higher affinity for MHC, that drive thymic Treg differentiation. Furthermore, kinetic measurements and crystal structures of a diversity of Treg TCRs will demonstrate if one or more biochemical factors are responsible for the majority of Treg differentiation.

1.08 Identification of Treg TCRs

Early work by Dr. Peter Savage's group used the transgenic adenocarcinoma of the mouse prostate (TRAMP) model[69,70]. Male TRAMP mice consistently and spontaneously develop prostate tumors following the onset of puberty due to the expression of the model oncogene SV40 T antigen in the prostate. Mice of this background had an enrichment of Treg cell infiltration in the prostate compared to tumor negative mice[56]. Upon TCRβ chain sequencing of prostate infiltrating Treg cells, a recurrently expressed TCRβ chain (TRBV26 – CASSLGSSYEQYF) was identified. This TCRβ chain was used to develop fixed TCRβ transgenic mice (TCRβtg). These TCRβtg mice were back-crossed onto TRAMP mice, which again exhibited significant Treg infiltration of the prostate tumors. These Tregs were then sequenced to obtain TCRα chains to identify complete, paired-chain Treg TCRs recurrently enriched in the prostates of TRAMP mice. One of these TCRs, named MJ23, was found to be reactive to a prostate-associated antigen present in tumor-free mice, and the transgenic expression of the MJ23 TCR induced Aire-dependent thymic Treg development in both male and female mice. This finding was significant in demonstrating Aire's ability to drive thymic development of self-reactive thymocytes into the Treg lineage and was crucial for enabling subsequent studies of Treg biology.

1.09 Treg ligands

Recent work by Dr. Peter Savage's group and others has made it increasingly clear that Tregs recognize self-antigens both in the thymus and the periphery. The Savage lab identified the first two naturally occurring thymic-derived Treg TCRs, MJ23 and SP33. These two clones were discovered to be recurrently enriched in experimental murine prostate tumors and in Aire-deficient mouse prostates, respectively. The Adams lab, in close collaboration with the Savage

lab, developed a candidate-antigen enrichment approach that identified 20 proteins as possible stimulatory antigens for these Treg TCRs[71]. These proteins were identified based on data determining the following three requirements for each protein: 1) each gene must not be located on the Y chromosome, 2) gene expression must be regulated by Aire in the thymus, and 3) the protein's expression must be prostate-specific. Using this approach, these two labs discovered the first two self-antigens known to drive the thymic selection of Tregs. These antigens, named C4 and F1, represent distinct epitopes of the same prostate-specific protein called Tcaf3, and, when presented in the context of the I-A^b MHC class II molecule, specifically stimulate the MJ23 and SP33 Treg TCRs, respectively. These data, along with other evidence, confirm that thymic-derived Tregs recognize self-antigens expressed in both the thymus and the periphery. However, much is still unknown about the molecular mechanisms used by Tregs to recognize self-antigens and how this drives their behavior. These two TCR/pMHC pairs represent remarkable tools for investigating Treg biology and are the foundation for the research conducted in this dissertation.

1.10 Conclusion

In the quarter-century since Tregs were originally discovered, much has been learned about their development, behavior, and function. However, much remains unknown about the antigens they recognize and the specifics of their development in the thymus. The current appreciation of the existence of self-reactive Th cells in the periphery indicates an incompleteness to negative selection and Treg differentiation for self-reactive cells. This highlights new questions and comparisons between the self-reactive CD4-restricted cells that comprise the Treg vs. Tconv lineages for a given antigen. What distinguishes these cells phenotypically, and how do their TCR repertoires compare? Are they distinct – and therefore

informative to their development and biology – or are they overlapping, indicating a stochastic nature of development for self-antigen reactive T cells. Do Treg and Tconv TCRs for a given antigen bind pMHC with distinct affinity, kinetics, or docking modes, and what implications do those similarities or differences have for their biology? And on an even more basic level, what affinity do Tregs bind pMHC with, and is it in the standard range of TCR affinity of 1-100μM?

This dissertation aims to answer these questions and discuss the implications they present for Treg development and function.

Chapter 2: Single Cell Genomics and Receptor Sequencing of Regulatory T Cells and Conventional T Cells Specific to The Same Self-Antigen in Antigen-Sufficient and -Deficient Models

2.01 Introduction

Regulatory T cells (Treg cells) are critical for the prevention of autoimmune reactions but their self-ligands have only recently begun to be defined. Recent work by the Savage and Adams labs identified two naturally occurring murine Treg cell ligands, both derived from a protein of prostate-specific origin called Tcaf3. This protein is recurrently targeted by autoantibodies in experimental models of prostatic autoimmunity, and its expression in the thymus is regulated by the transcription factor Aire. Two distinct and non-overlapping peptide epitopes from Tcaf3 were discovered to have robust Treg cell repertoires specific to them, and one of these epitopes, the C4 epitope, has been used as the basis of these studies here. Previous studies were completed in mice with a fixed transgenic TCRβ-chain to allow for bulk α-chain sequencing to identify prevalent TCRs. While operating at a lower throughput, modern single-cell sequencing techniques can be completed in mice with a fully diverse TCR repertoire, allowing for the discovery of novel paired-TCRs beyond those identified with the fixed β-chain of MJ23 and SP33, as well as allowing the development of fully physiological TCR repertoires.

Thus, the purpose of the experiments in this chapter was to identify and characterize the immunological properties of novel Treg TCR repertoires specific for a single known self-antigen. Furthermore, it sought to compare the development of TCR repertoires specific for this self-antigen in antigen-sufficient and -deficient contexts to provide insights into the effect of antigen presence on Treg development and function.

2.02 Single-cell sequencing permits the investigation of T cell biology in Treg cells and Tconv cells specific for the same antigen.

The purpose of the experiments in this chapter was to identify and characterize the immunological properties of novel Treg cell repertoires specific for a known self-antigen. This was done with a two-tiered approach: first, a single cell, paired-chain PCR based approach followed by Sanger sequencing was used, and second, single-cell droplet-based sequencing of bulk cells using a 10x Genomics-based platform was used to identify novel Treg TCRs specific for the C4 antigen in mice with fully diverse TCR repertoires.

Single-cell PCR-based sequencing was conducted in wild-type mice that were immunized with C4 peptide in CFA to expand antigen-specific cells. Then, cells from pooled spleen and lymph nodes were stained with C4/I-A^b tetramers, tetramer-positive cells were magnetically enriched, and individual tetramer-positive cells were sorted by FACS. Sorted single cells were subjected to cDNA preparation and nested PCR amplification of TCR transcripts using primers specific to all mouse TRAV and TRBV genes. Clones with successfully amplified TCR alpha and beta chains were submitted for Sanger sequencing to identify the corresponding TCR alpha/beta pair, complete with CDR3 sequences. This strategy enabled the identification of 65 new Treg TCRs from five distinct mice (Figure 2.1).

Many of these TCRs were found to be expanded significantly and several public TCRs were found, meaning the exact rearranged alpha/beta pair with identical CDR3 amino acid sequences was found in multiple mice, and demonstrating that C4-specific TCRs elicited in different mice exhibit conserved features. Two αβ V gene pairs were found to be especially enriched, TRAV 7-2/TRBV 13-2 and TRAV 7N-5/TRBV 16 (Figure 1B). Multiple TCRs with varying CDR3 sequences were found across mice using these two Vgene pairs (Table 1),

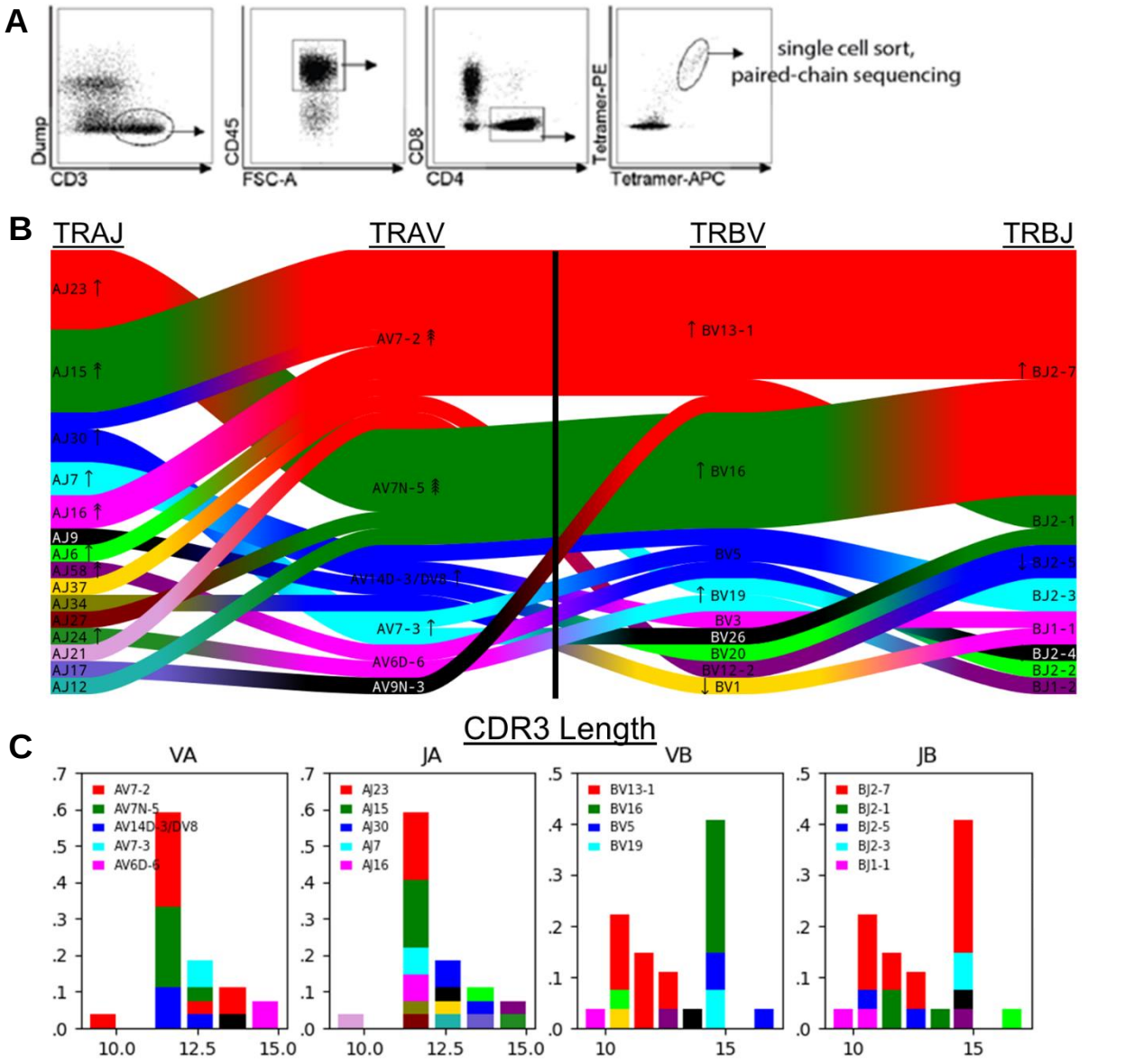

Figure 2.1 Single-cell sequencing of the C4-reactive Treg TCR repertoire in wild-type mice. **(A)** CD4+ T cells were dual-stained with C4/I-A^b tetramer and single cell FACS sorted. Representative dot plots of 1/5 mice shown. **(B)** Dual-tetramer-positive cells in (A) were amplified by PCR and subjected to Sanger sequencing. 65 paired chain TCRs were identifie and the gene pairing ribbon plot displaying the landscape of paired alpha/beta TCR sequences is shown here. Curved paths show TRAJ, TRAV, TRBV, TRBJ gene usage for each expanded TCR. Color distinguishes each gene, curves distinguish each TCR. Curve thickness indicates the number of clones sharing a particular gene segment or gene-gene pair. Arrows hatches indicate 2-fold enrichment over background frequency of gene usage, with each hatch indicating an additional fold enrichment. **(C)** CDR3 length plots of CDR3 sequences using each gene. Colors correspond to gene usage in (B).

possibly indicating some role of the germline-encoded CDR1 and CDR2 sequences in their reactivity. These TCRs used diverse Jα genes but over 90% used the same Jβ2-7 gene. While TCRs from these two V gene pairs utilize CDR3α sequences of similar length, they used CDR3β sequences of dramatically different lengths (Fig. 1C), indicating two likely distinct methods of interaction with the C4/I-A^b complex.

Table 1 displays some of the highest expanded clones found across these five mice, many of which are of the TRAV 7-2/TRBV 13-2 and TRAV 7N-5/TRBV 16 pairs. It is not determined if this expansion occurred in the baseline repertoire or following immunization. Comparing these sequences to the previously identified C4/I-A^b reactive TCRs in the fixed TCRβ transgenic model, none of the CDR3α or CDR3β sequences were identified with one exception. The TRAV7-2 gene with the CDR3 sequence of CAATSSGQKLVF was identified in two mice, each in only one clone, and with a different TRBV gene than the fixed TCRβ transgene. Given that the original prostate-enriched MJ23 TRBV gene was identified without knowledge of its antigen, it is likely that it was present in TCRs specific for an antigen other than C4.

While this single-cell PCR protocol was employed successfully, there was significant noise in the amplification process and many TCRs were lost to non-specific amplification or lack of amplification. To obtain greater sequencing depth per mouse, we decided to next undertake a more high-throughput droplet-based approach using the 10x Genomics platform. This approach would also increase throughput and accuracy, which enabled easy sequencing of C4-specific repertoires in both wild-type and C4-deficient ($C4^{KO}$) mice. The gene expression data obtained by 10x Genomics sequencing also enabled sequencing of Treg cells and Tconv cells together. Eight mice were immunized with C4 peptide in CFA two weeks prior to processing. Dual tetramer-positive cells (in APC and PE) were sorted for sequencing. Four mice were of

Previously Identified TCRs – Fixed TCRβ Background								
	Alpha Chain			Beta Chain				
Name	TRAV	CDR3α	TRAJ	TRBV	CDR3β	TRBJ	Mouse	Antigen
MJ23	14-2*01	LYYNQGKLI	23	26	ASSLGSSYEQY	2-7	Fixed TCRβ	C4
TRAV16	16*02	AMRETWSNYNV	21	26	ASSLGSSYEQY	2-7	Fixed TCRβ	C4
SP33	9D-3*02	ALSMSVNYQLI	33	26	ASSLGSSYEQY	2-7	Fixed TCRβ	F1
SP33	9D-3*02	ALSMSVNYQLI	33	26	ASSLGSSYEQY	2-7	Fixed TCRβ	F1
GAW	9D-4*04	AVGAWTGNYKYV	40	26	ASSLGSSYEQY	2-7	Fixed TCRβ	F1
TRAV7-2	7-2*02	AATSSGQKLV	16	26	ASSLGSSYEQY	2-7	Fixed TCRβ	Unconfirmed
TRAV19	19*03	AAVSNTNKVV	34	26	ASSLGSSYEQY	2-7	Fixed TCRβ	Unconfirmed
Single Cell PCR Refined Results								
	Alpha Chain			Beta Chain				
Name	TRAV	CDR3α	TRAJ	TRBV	CDR3β	TRBJ	Mouse	Clones
RKD1	7N-5*01	AVRYNQGKLI	23	16	ASSSGVGVSYEQY	2-7	167	8
RKD4	7N-5*01	AVRYNQGKLI	23	16	ASSLGTGVSYEQY	2-7	110	2
RKD6	7-2*02	AATSSGQKLV	16	13-1	ASSDRYEQY	2-7	110	2
RKD8	7N-5*01	AVRYNQGKLI	23	16	ASSPGTGVSYEQY	2-7	170	11
RKD10	7-2*02	AGNYNVLY	21	19	ASSPTGGQRGTLY	2-4	170	4
RKD11	7-2*02	AASIGNTGKLI	37	13	ASSDRYEQY	2-7	170	1
RKD13	14-3*01	AASAYMGYKLT	9	1	TCSGQTEVF	1-1	108	6
RKD16	7-2*02	AATSSGQKLV	16	13	ASSDRYEQY	2-7	108	1
RKD17	7N-5*01	AVRTNTGKLT	27	16	ASSFGLGLSYEQY	2-7	703	5

Table 1. Previously identified Treg TCRs and newly identified single-cell PCR-derived TCRs reactive to C4/I-A^b

(Top) Previously discovered Treg TCRs identified by iRepertoire TCRα chain sequencing of C4 or F1 tetramer-positive cells in fixed TCRβ transgenic mice from multiple models of prostate infiltration. Antigen specificity is noted in the far-right column. **(Bottom)** Newly identified C4/I-A^b reactive TCRs identified by single-cell PCR-based sequencing of C4-tetramer positive cells in wild-type mice two weeks post-immunization with C4 peptide in CFA. Cells were tetramer enriched and single-cell sorted before identification by PCR and Sanger sequencing. The number of cells in each mouse bearing the identical TCR nucleotide sequence is shown at far-right. **(Both)** All TCRs have TCR genes, CDR3 sequences, and names listed.

wild-type B6 background ("WT" mice) and four were of $C4^{KO}$ background ("$C4^{KO}$" mice), in which the Treg-antigenic C4 peptide of the prostatic Tcaf3 protein has been knocked out, eliminating its expression in both the thymus and peripheral tissues while maintaining normal expression of other Tcaf3-derived epitopes. C4/I-A^b tetramer-positive cells were stained with hashtag identifier antibodies before pooling into two samples containing four mice each. The hashtag antibodies allow tracing of each cell's mouse of origin, enabling comparisons of each of the eight distinct immune responses in isolation or together. Each of the two samples contained cells pooled from two WT and two $C4^{KO}$ mice to control for batch effects, and samples were submitted to the UChicago Genomics Core for sequencing. Sequencing data were obtained and processed with the 10x CellRanger software and Seurat R package.

With this strategy, paired chain TCR VDJ data, gene expression data, and hashtag identifier data for 901 C4-reactive T cells were obtained. Initial UMAP clustering of all 901 T cells revealed a few general observations (Figure 2.2A). The cells can be sorted into the following clusters: less-activated/naïve-like Tconv cells (cluster 0), activated Tconv cells (clusters 2 and 6), exhausted Tconv cells (cluster 7), cycling cells (clusters 4 and 8), and Treg cells (clusters 1 and 5). Cluster 3 contained T cells without VDJ data and was removed from the analysis. These annotations have been determined by comparing the top differentially expressed genes in each cluster, as attempts to use automated cell type annotation software such as SingleR or GO analysis were incapable of distinguishing between the subtle differences among different T cell subsets.

Interestingly, the Treg cells in these populations appear to cluster into two distinct populations (clusters 1 and 5), with some additional Treg cell presence in both cycling clusters (4 and 8) as well (Figure 2.2B). Thus, Treg cells were binned as cells expressing Foxp3 and CD25

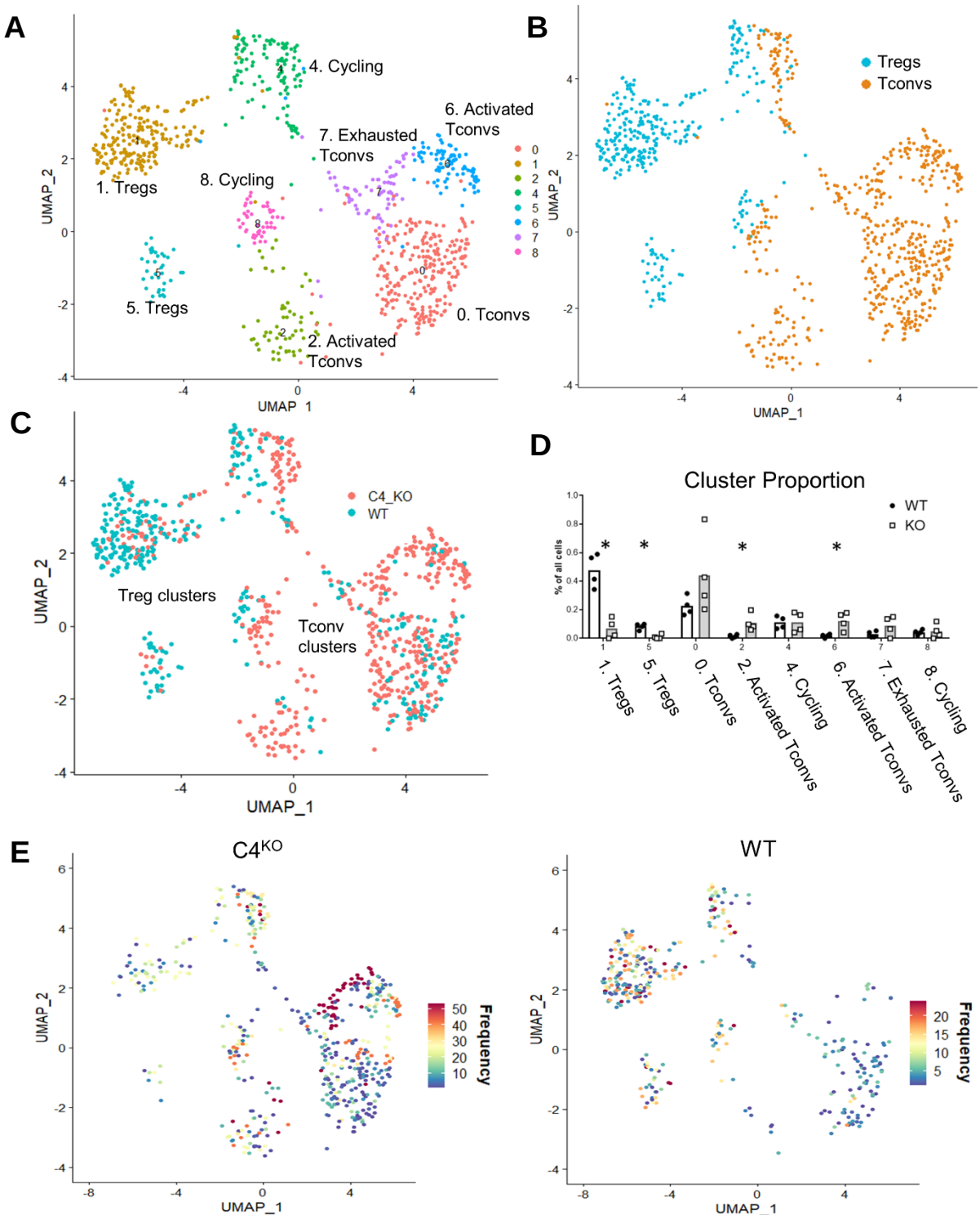

Figure 2.2 Germline deletion of the C4 antigen in mice results in Tconv-skewed repertoire
(A) UMAP reduction of 901 C4-tetramer-positive T cells identified by scRNAseq with gene expression, hashtag identifier, and VDJ data incorporated. Unsupervised clusters annotated

Figure 2.2 (continued)
manually. **(B)** Binned Treg (blue) vs. Tconv (orange) featureplots. Expression of Treg markers Foxp3 and CD25 in clusters 1, 4, 5, and 8 used to bin Treg vs. Tconv phenotype. **(C)** Featureplot of WT (blue) and $C4^{KO}$ (orange) repertoires. **(D)** Statistical analysis of clusters by genotype with Student's t-Test. * indicates $p < 0.05$. **(E)** Feature plot of clonal expansion by genotype indicated by frequency of each clone. Maximum of 53 in $C4^{KO}$ mice and 23 in WT mice.

in Clusters 1, 5, 4, and 8 to segregate from Tconv cells. Cells were compared with other Treg-specific genes to ensure accurate binning. While there are significantly differentially expressed genes between the two primary Treg cell clusters, including Helios, CD45, and CD69, they largely share a TCR repertoire, suggesting there is movement of cells between these two populations and the cycling populations. This indicates that the differences between these clusters are primarily due to the cells' stage in the cell cycle or perhaps the strength or frequency of activation rather than robust and durable differences in cell phenotype. The recency of a cell's last activation could play a role as well. Cell cycle scoring confirms clusters 4 and 8 are primarily in S phase of G2M phase (data not shown). As expected, a majority of Treg cells identified in this experiment are derived from the WT mice (74%), with Treg cells from both clusters 1 and 5 showing similar statistically significant reductions in $C4^{KO}$ mice (Figure 2.2C, D). This confirms experiments conducted in fixed beta chain transgenic[52,56] mice demonstrating a dependence on antigen presence in the thymus and/or periphery for Treg cell development.

In $C4^{KO}$ mice, the Tconv cell compartment exhibits significant activation in response to immunization at day 14, whereas the Tconv compartment in WT mice is almost entirely lacking an activation phenotype. Thus, the Treg compartment (~68% of the C4-reactive T cells post-immunization) in WT mice can either suppress or outcompete Tconv cell activation in response to C4 peptide and CFA, and the Treg compartment in $C4^{KO}$ mice (only ~17% post-immunization) is insufficient to suppress activation. This could be explained by a number of factors that will require further experiments to clarify, including but not limited to: 1) sheer

number of Treg cells and therefore a greater Treg:Tconv ratio, 2) differences in the Treg TCR repertoires enabling greater activation and/or suppressive function by WT Treg cells, and 3) different functional states attained by WT Treg cells due to the persistent presence of the C4 peptide.

By viewing the clonal expansion and clonal relatedness of these clusters, it is made clear that T cells with some TCRs are clonally expanded and found in multiple clusters, but populate one of the two broad bins of Treg cell clusters and Tconv cell clusters (Figure 2.2E). In WT mice, Tconv cells remain almost exclusively within the non-activated/naïve-like cluster (cluster 0), whereas in $C4^{KO}$ mice the Tconv cells are broadly activated and fall within multiple stages of activation, including early activation (clusters 0, 2, 6), late activation (cluster 7), and active cycling (clusters 4, 8). The clonal expansion of T cell clones was visualized by color coding the frequency of each clone's occurrence (Figure 2.2E). Doing so reveals dramatic expansion of activated Tconv cells in C4-KO mice, where highly expanded clones (in red) are located in the peripheral Tconv clusters and less so in cluster 0, where less expanded clones persist. In WT mice, the highly expanded clones were predominantly located in the Treg clusters, with the Tconv cells almost exclusively exhibiting zero to low levels of expansion. Further data, not shown here, shows TCR overlap between clusters, indicating that expanding clones can fall into multiple clusters and are possibly cycling between these clusters. Therefore, these clusters most likely represent two large classes of T cells, Tconv and Treg cells, and smaller subclasses within those two categories characterized by cell cycle stage and recency or strength of activation Further subclasses could be distinguished by "imprinting" or other phenotypic changes in response to persistent self-ligand recognition in WT mice that is absent in $C4^{KO}$ mice.

The top differentially expressed genes in each cluster are visible in Figure 2.3A, which shows the larger number of less-activated/anergic T cells in cluster 0 and Treg cells in cluster 1 compared to other clusters. Also visible is the dominance of cyclin genes in distinguishing the cells in clusters 4 and 8 over all other gene expression, including clear Treg cell markers such as Foxp3 and Helios. Markers of T cell exhaustion, such as PD-1 and Lag-3, are highly expressed in the cells from cluster 7. Clusters 2 and 6, two of the activated T cell clusters, are not easily distinguished except by phenotypic markers such as IFNg and CD45RA, so recency of activation remains as a potential explanatory factor for their clustering (with cluster 6 representing the more recently activated T cells compared to those from cluster 2).

Approximately equivalent numbers of C4-reactive T cells were obtained from each mouse, with the possible exception of mouse 5, one of the $C4^{KO}$ mice, which had fewer T cells recovered than the others (Fig 2.3B). Additionally, the hashtag identifier antibodies were able to stain cells with high reliability, and fewer than 30 cells total were lost due to being identified as doublets or as negative for hashtag antibodies. Similarly, the expansion of cells in each mouse can be seen in Figure 2.3C. Each mouse repertoire can be seen to largely replicate the overall pattern, with WT mice exhibiting expansion of the Treg repertoire and $C4^{KO}$ mice exhibiting skewing to the Tconv repertoire.

One of the strengths of this dataset lies in the ability to investigate the effect of self-ligand selection on the transcriptional signature of Treg cells expanded by immunization. By comparing gene expression profiles of Treg cells in WT mice, which were previously exposed to C4/I-Ab in the thymus and periphery, and those in C4-KO mice, which were not, the potential for “Treg training” can be investigated. To that end, the Klf2 transcription factor was identified in significant Treg cell populations in WT mice ($p = 0.0001$), but similar populations were absent

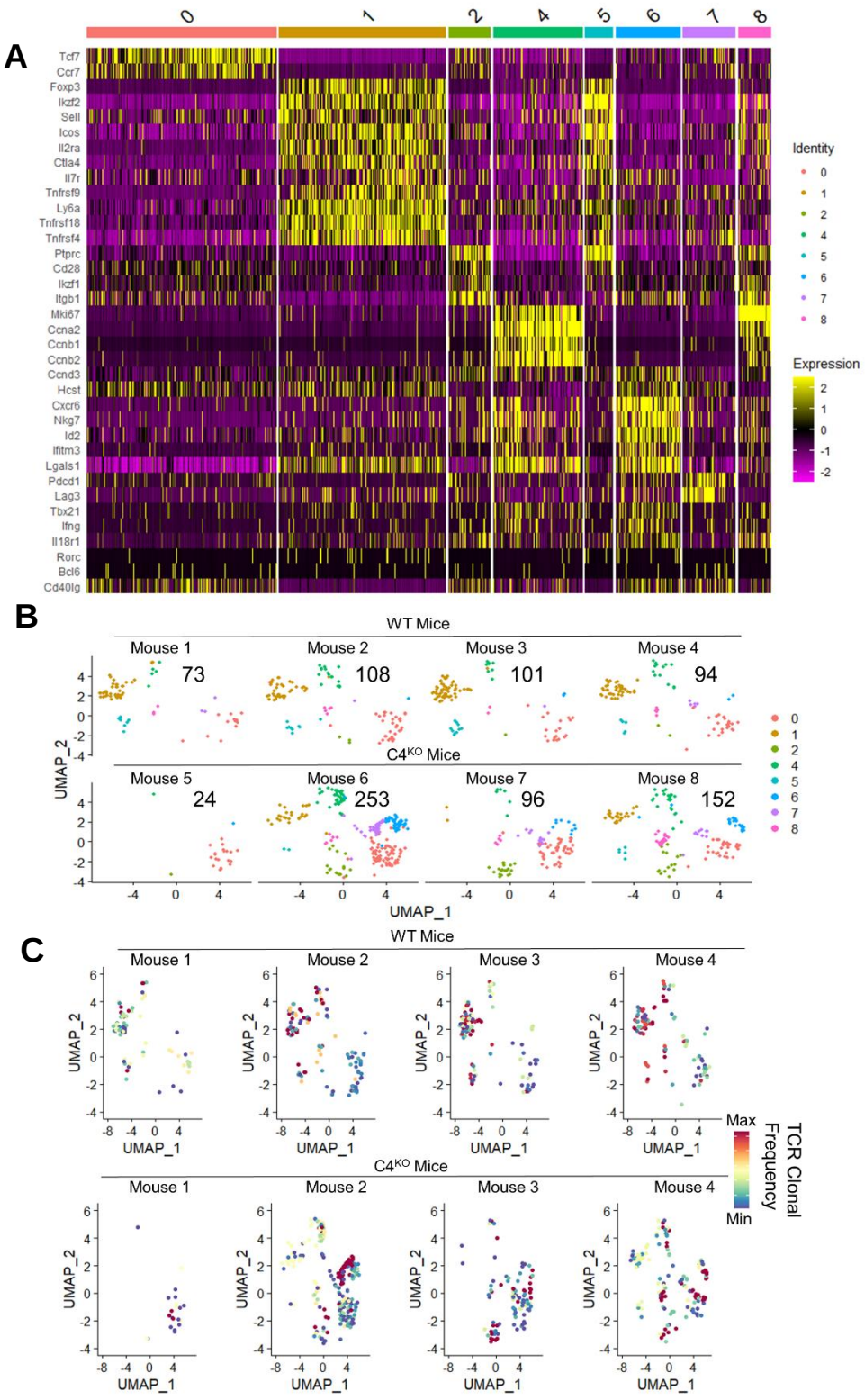

Figure 2.3 Heat map of differentially expressed genes by cluster and individual mouse data

Figure 2.3 (continued)
(A) Heat map of top differentially expressed genes of cells in figure 2.2A by cluster. Yellow indicates positive fold change, purple indicates negative fold change in proportion with the key. **(B)** Featureplots of cells per cluster in each individual mouse in Figure 2.2A. Mouse number, genotype, and cell number per mouse overlaid on each plot. **(C)** Featureplots of clonal expansion by mouse by TCR clonal frequency. The maximum frequency in each mouse is colored as maroon and the minimum is dark blue.

in C4-KO mice (Figure 2.4A, B). These genes have both been linked to suppressive Treg function previously[72,73] and could be markers of fully "licensed" Treg cells in WT mice. Further experiments will be needed to determine if this is the case or if they are markers of greater TCR-mediated activation or Treg-specific signaling occurring in immunized WT mice that do not occur in C4-KO mice.

2.03 Analysis of C4-specific TCRs reveals distinct Treg and Tconv repertoires

As mentioned above, this dataset included VDJ sequence information for paired-chain TCR identification for all 901 cells. The TCR repertoire in each mouse comprised, on average, 30 unique C4-reactive TCRs (Figure 2.5C). Given that this dataset was captured post-immunization with C4, this can be thought of as representing the number of naïve T cells and Treg cell clones circulating before immunization. To support this data, flow cytometry and dual-tetramer staining of enriched, naïve mice was conducted and identified an average of 23.5 (Figure 2.5A, B, C) naïve cells per mouse, indicating that the immunization and sequencing data was capturing a large percentage of clones, if not the entirety, within the C4-reactive repertoire.

Additionally, multiple papers have sequenced and identified the repertoire overlap between Treg and Tconv cell repertoires of entire immune repertoires before using mice with and without a fixed TCRbeta chain[50,58,74]. The data gathered in this 10x analysis allowed investigation into this question for the repertoires specific for a single antigen. Interestingly, the Treg and Tconv repertoires in WT mice were found to exhibit no sequence overlap, with zero

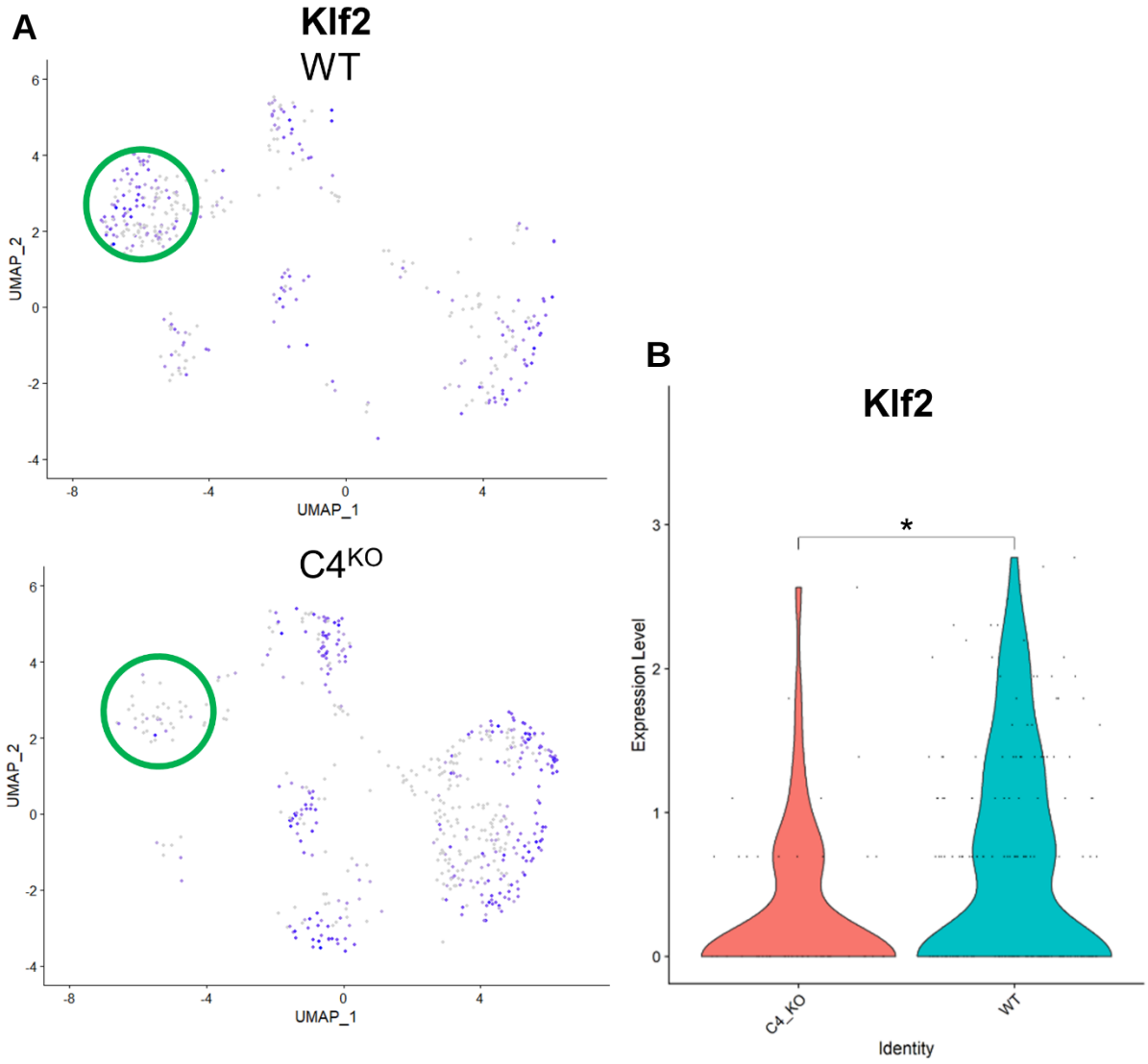

Figure 2.4 Treg-related transcription factor is upregulated in WT Treg populations
(A) Expression of Klf2 in Treg populations split by genotype (WT, top, and $C4^{KO}$, bottom). Klf2 comparisons of Treg cluster 1 in Figure 2.2A circled in green. **(B)** Violin plot of Klf2 expression in cluster 1 of WT mice vs. $C4^{KO}$ mice. Number of cells in WT mice = 176, number of cells in $C4^{KO}$ mice = 50. * = p value less than 0.05.

TCRs being found in both cell types in two different mice (Figure 2.5D). Thus, a significant population of self-reactive T cells appears to be consistently induced to develop into the same cell type across mice and few cells appear to be "agnostic" as to their developmental fate. This additionally reveals stability of Tconv and Treg subsets, as a given clone will only populate one pool or the other and will not fluctuate between the two. This reinforces the potential importance of TCR-intrinsic developmental factors, which will be discussed in greater detail in Chapter 3 of this thesis.

In the four WT mice expressing the C4 antigen, the number of unique Treg and Tconv TCR clones was found to be approximately equivalent in each mouse (Figure 2.5E, top). Importantly, this shows that at least for this single antigen, the Treg and Tconv repertoires contribute about equivalent numbers of clones without generating autoimmunity. However, post-expansion, a far greater proportion of the repertoire was derived from Treg cells, resulting from their preferential expansion prior to or following immunization (Figure 2.5E, bottom). This indicates that within a self-antigen-specific TCR repertoire, diversity may not be as meaningful towards preventing autoimmune inflammation as the overall number, activity, or location of cells in the Treg compartment compared to the Tconv compartment.

However, in the C4-KO mice, over 80% of unique clones derive from Tconv cells, with the remaining minority deriving from Treg cells pre-expansion (Figure 2.5F). This result was expected and correlates with significant previous data indicating Tconv skewing in the absence of thymic antigens important for Treg development. This increase in Tconv cells could represent the emergence of Tconv clones that are normally directed into the Treg cell lineage in WT mice (the Trogue hypothesis), or from cells that are escaping negative deletion in the thymus in C4-deficient mice. While the Tconv TCRs that appear in the C4-deficient setting do not appear in

the Treg compartment in WT mice, the variability and lack of a sufficient number of public clones in this dataset of fully diverse TCR repertoires do not allow for conclusions of either hypothesis.

Another interesting finding was made by comparing the V gene usage of the TCRα and TCRβ chains in repertoires from both backgrounds (Figure 2.6A). In WT mice, 50 of the 376 total clones sequenced shared the same Valpha and one of two Vbeta segments. In $C4^{KO}$ mice, however, six alpha chains and five beta chains were found in more than 50 clones (out of 525), indicating a greater convergence of the repertoire on a select few TCRs in WT mice.
This can be further seen in Figures 2.7 and 2.8, where each V(D)J recombination event is visualized by ribbon plot. Figure 2.7 compares the TCRs in WT mice from Treg (A) and Tconv cells (B). In Treg cells, the TRAV7-2 V gene dominates, comprising half the receptor repertoire alone. The TRBV13-1 V gene also comprised nearly half the receptors, though each of these V genes paired with other V and J genes and not only each other. The TRAV7D-2 and 14-2 were the second and third most dominant Vα genes, while numerous Vβ genes contributed to the repertoire. In Tconv cells, conversely, TRBV3 and TRBV13-2 dominated the repertoire, along with TRAV12D-3. In all, the top 50% of the TRBV genes used by WT Treg and Tconv cells were unique to each cell type. The significance of this is perhaps best demonstrated by the use of TRBV3 in the Tconv repertoire, which was not due to expansion but was recurrently found in numerous clones with distinct alpha genes (Figure 2.7B).

Comparing the full WT and $C4^{KO}$ repertoires, the TCRβ gene usage of the repertoires remained largely similar, with TRBV13-2, TRBV13-2, and TRBV3 contributing the most cells in each case. However, the TRAV repertoire experienced multiple changes. This includes the exclusion of TRAV12D-3 and TRAV14-2 in the $C4^{KO}$ repertoire, which were two of the top

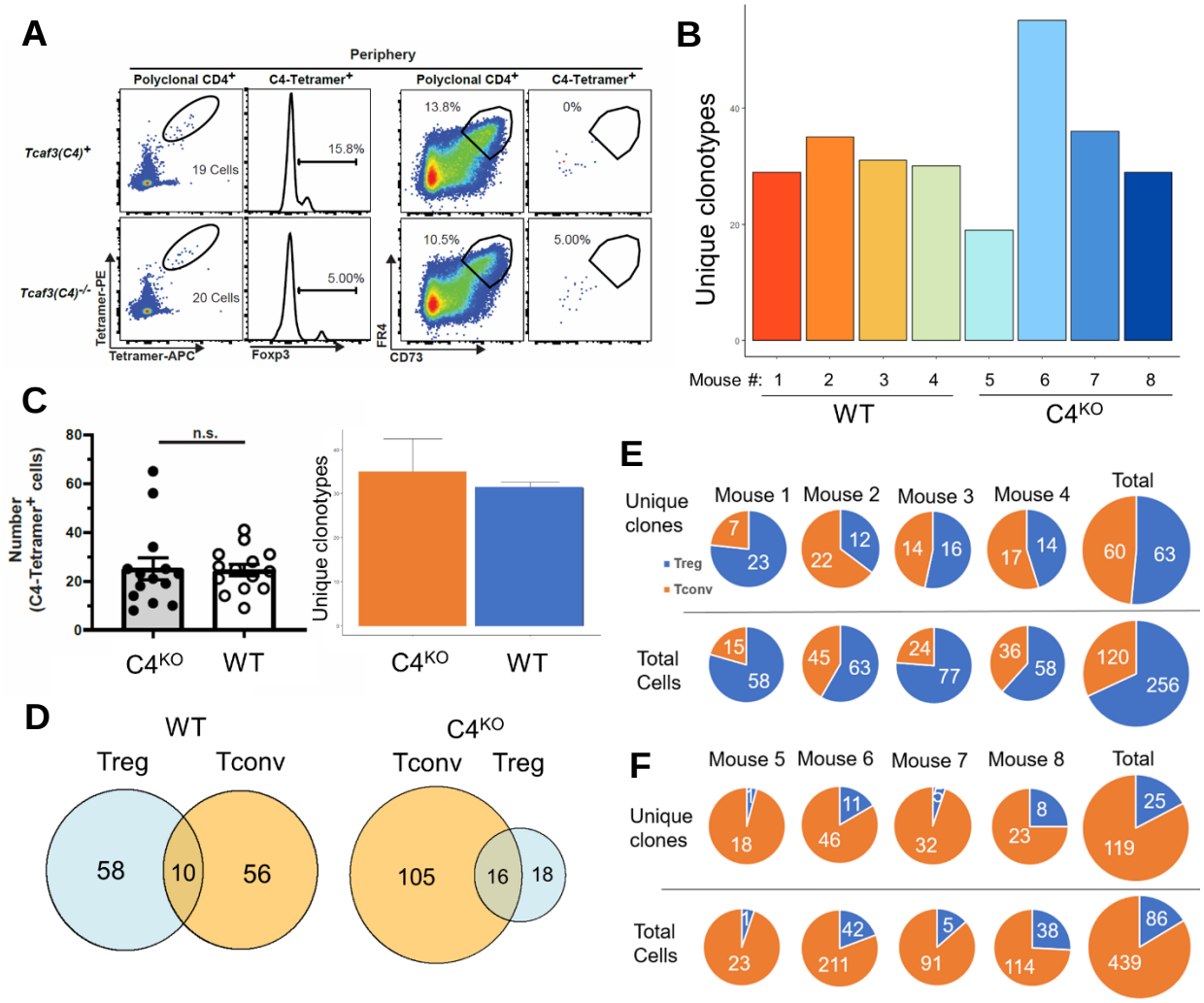

Figure 2.5 C4-reactive TCR analysis reveals distinct Treg and Tconv repertoires
(A) Representative flow plots of dual C4-tetramer positive T cells MACS enriched in naïve mice. **(B)** Quantification of unique TCR specificities in each mouse from the 10x dataset. Each genotype averaged approximately 30 specificities per mouse. **(C)** Quantification of data in (A) and (B). Data acquired from n=14 mice of each genotype in (A), which both averaged approximately 23 specificities per mouse. **(D)** TCR clonal overlap of Treg and Tconv repertoires in each genotype. **(E)** WT Treg (blue) and Tconv (orange) populations in each mouse (left four columns) and in total (right column). Top row shows the number of unique TCR specificities per mouse, bottom row shows the total number of cells of each cell type per mouse (post-clonal expansion). **(F)** Same organization as in (E) for $C4^{KO}$ mouse populations.

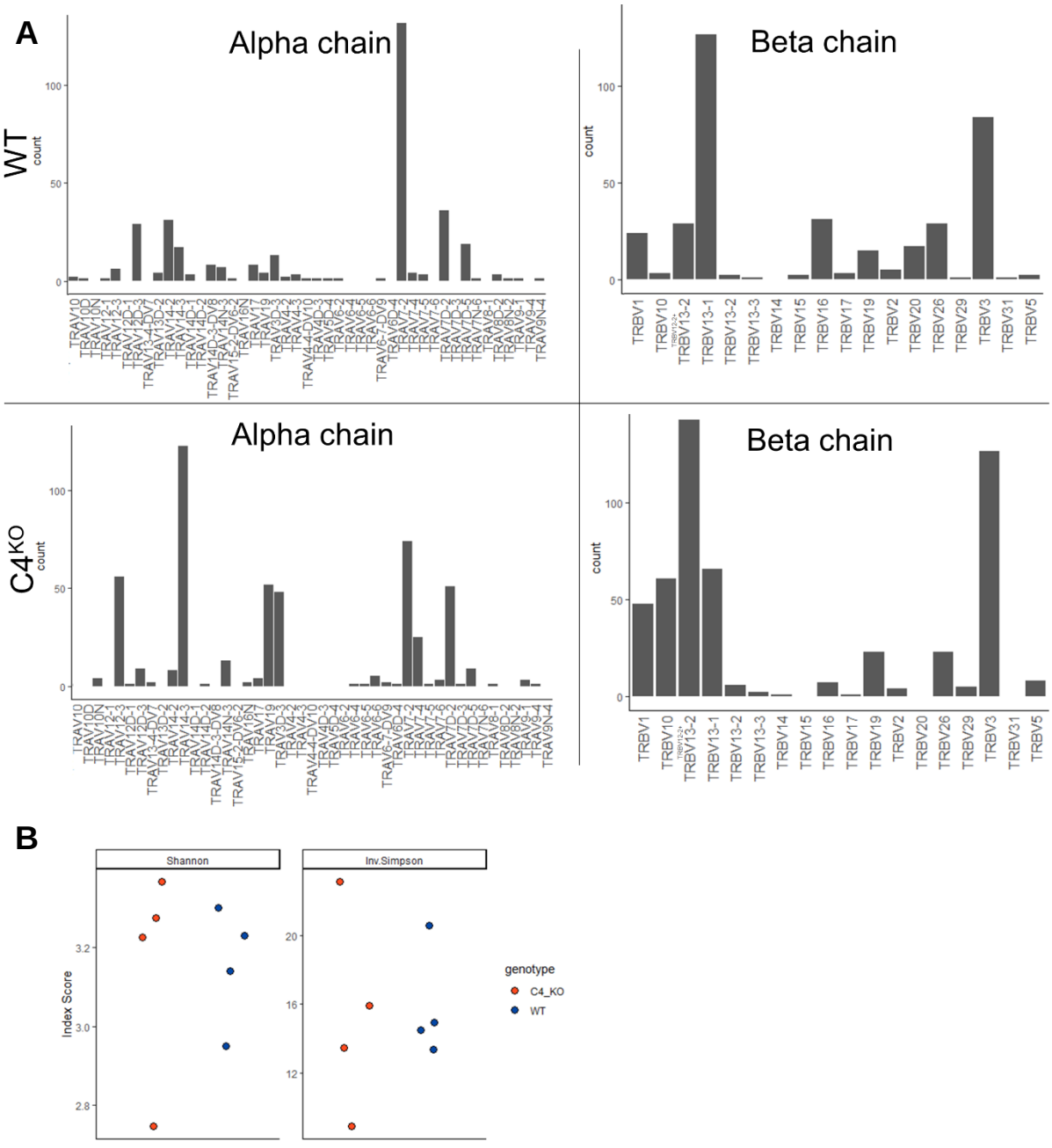

Figure 2.6 TCR Vα and Vβ usage by WT and C4^{KO} repertoires reveals appearance of new gene usage in antigen-deficient mice

(A) Number of cells bearing a TCR of each Vα (left) and Vβ (right) gene. Sample size of 376 TCRs in WT mice (top) and 525 TCRs in C4^{KO} mice (bottom). **(B)** Shannon Diversity index (left) and inverse Simpson index (right) of each mouse in both genetic backgrounds, n=4 for both backgrounds and were acquired in two independently processed experiments.

four TRAV genes in the WT repertoire. It also includes the exclusion of TRAJ16, which was the Jα gene used by a highly public WT Treg clone that will be discussed in the next subsection.

Next, all 901 TCRs in the 10x-derived repertoire were analyzed with the tcrdist3 software package[75]. TCR distances were calculated between each TCR to quantify the relatedness of each TCR to its neighbors. The differences in amino acid sequences of each CDR1, CDR2, CDR2.5 (HV4), and CDR3 loop were totaled based on their BLOSUM62 substitution penalties (which range from 0 to -4 based on observed statistical likelihood across a sample of conserved proteins). Changes in CDR3 sequence were given a 3-fold weight in scoring. Thus, if a given pair of TCRs varies by three amino acids, the distance between the two will be greater if the amino acid substitutions exhibit distinct properties (i.e. three positively charged residues being changed to three negatively charged residues), and it will be smaller if the substitutions are with amino acids of similar property. Given that CDR lengths are not uniform, the number of gaps, or spaces in a CDR sequence was taken into account by contributing a -4 score. Thus, the sequence distance between TCRs as well as the biochemical properties distinguishing them were taken into account in scoring relatedness. A resulting hierarchical tree was generated based on the clustering of CDRβ sequences while also displaying CDRα sequences (Figure 2.9A). A similar tree generated by clustering of CDRα sequences was generated and is excluded here for space considerations as it did not alter the conclusions discussed here. The tree cluster TCRs with the greatest segregation, and therefore the least distance between TCRs, at the bottom, and increases the distance between TCRs in a single cluster towards the top of the tree. The TCR distance of the bottom clusters is 0, meaning the width of each cluster at the bottom is due to identical TCRβ chains, though in a few cases the TCRα sequences differ. Overlaid on the tree are two motif

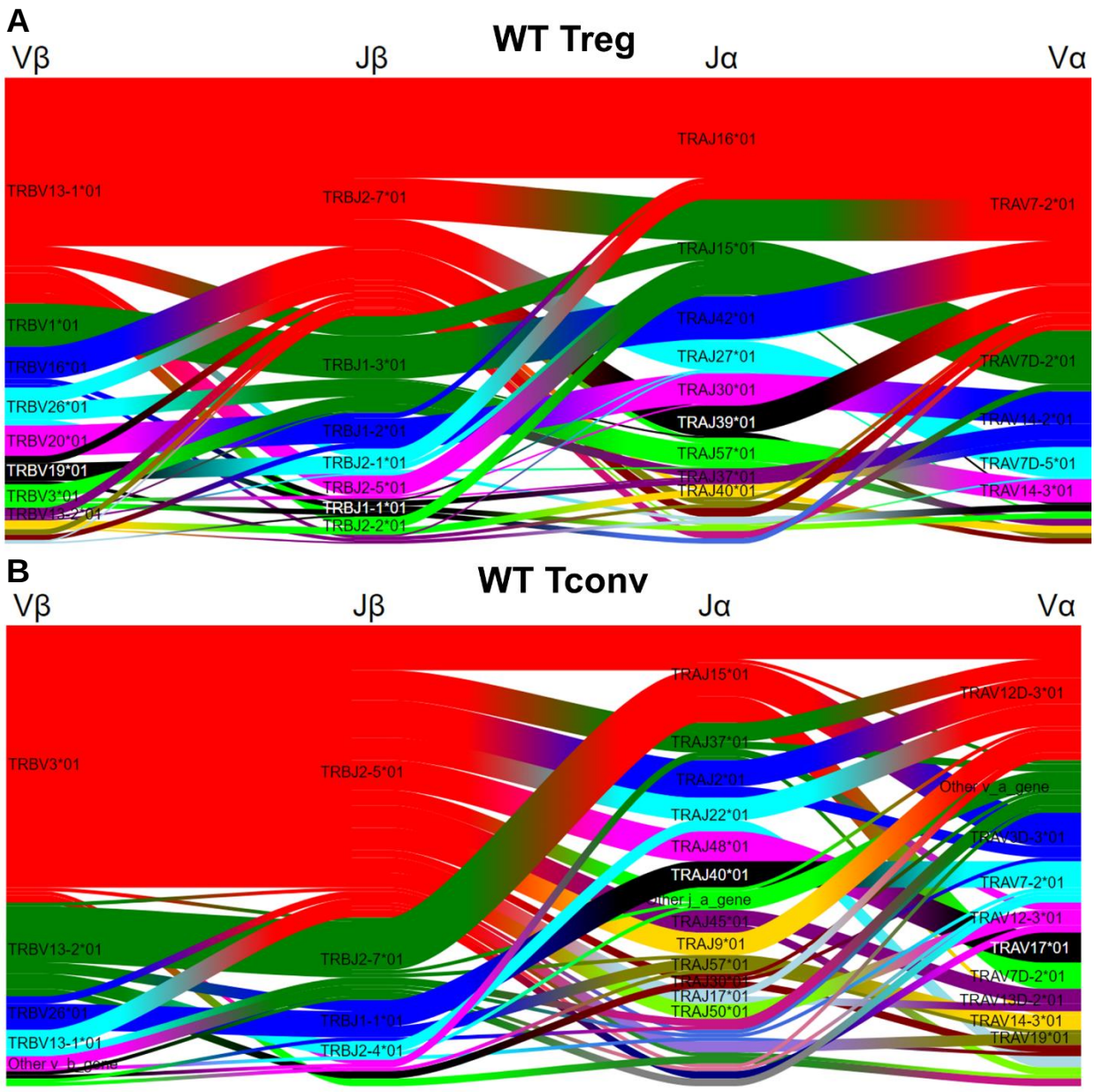

Figure 2.7 WT Treg and Tconv repertoires use distinct V(D)J rearrangements
(A) 256 WT paired chain Treg TCRs were identified and the gene pairing ribbon plot displaying the landscape of paired alpha/beta TCR sequences is shown here. Curved paths show TRAJ, TRAV, TRBV, TRBJ gene usage for each expanded TCR. Color distinguishes each gene, curves distinguish each TCR. Curve thickness indicates the number of clones sharing a particular gene segment or gene-gene pair. **(B)** 120 WT paired chain Tconv TCRs identified and displayed as in (A). n=4 for both backgrounds and were acquired in two independently processed experiments with two mice of each genotype in each sample.

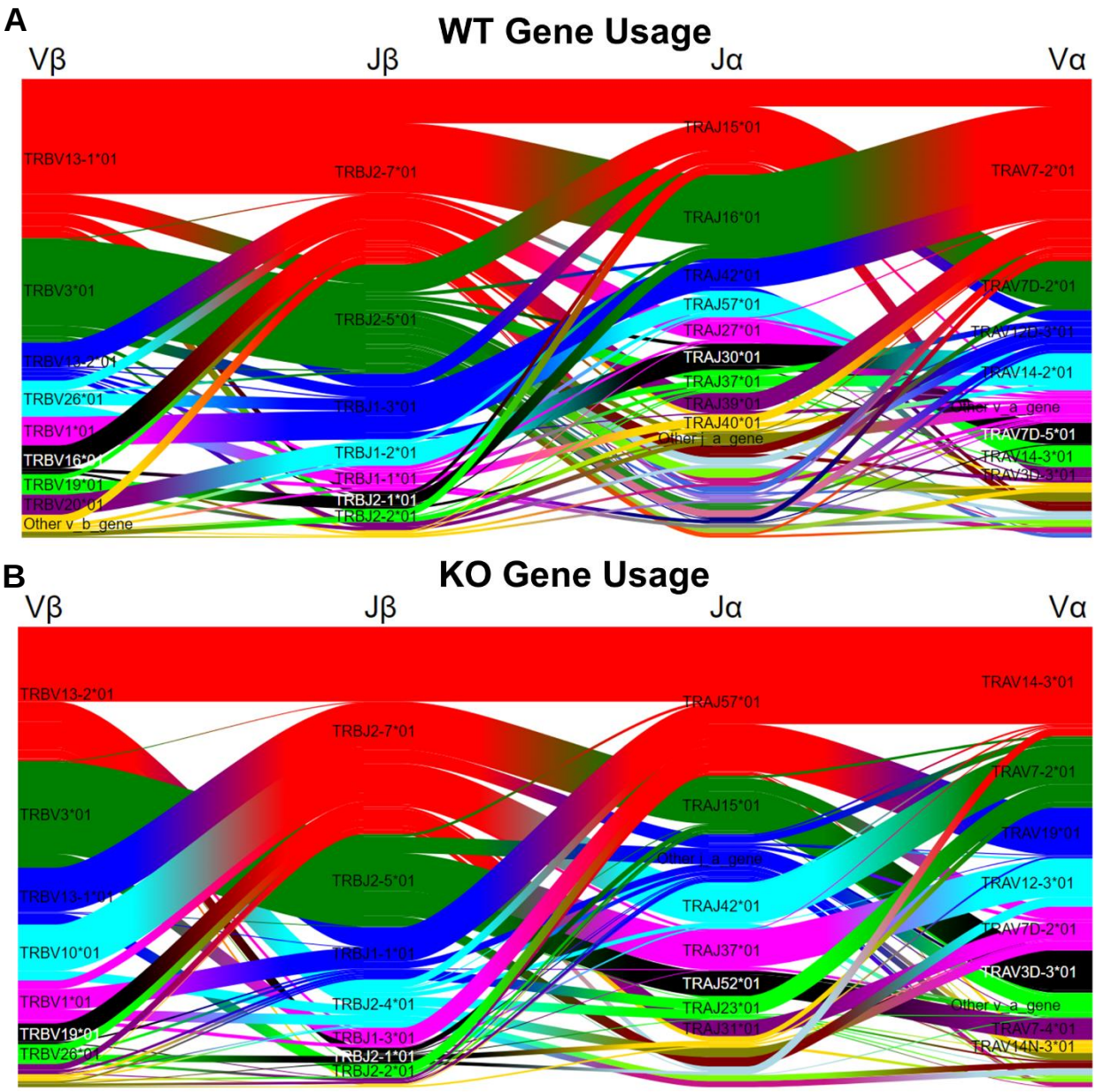

Figure 2.8 V(D)J rearrangements of WT and KO repertoires
(A) 376 WT paired chain Treg TCRs were identified and the gene pairing ribbon plot displaying the landscape of paired alpha/beta TCR sequences is shown here. Curved paths show TRAJ, TRAV, TRBV, TRBJ gene usage for each expanded TCR. Color distinguishes each gene, curves distinguish each TCR. Curve thickness indicates the number of clones sharing a particular gene segment or gene-gene pair. **(B)** 525 $C4^{KO}$ paired chain TCRs displayed as in (A). n=4 for both backgrounds and were acquired in two independently processed experiments with two mice of each genotype in each sample.

families, one in light blue for the largest WT Treg family (CDR3α sequence: CAATSSGQKLVF) and one in orange for the largest WT Tconv family (CDR3α sequence: CALITGNTGKLIF). The CDR3α sequences are shown on clustering generated by CDR3β distances to demonstrate the close relationship between their respective "neighborhoods" (i.e. the TCRα neighborhoods follow a similar pattern as the TCRβ neighborhoods).

This hierarchical tree yields several observations. First, and most importantly, the WT Treg TCRs (light blue) appear to segregate from the WT Tconv TCRs (orange). This indicates that the two cell types not only possess distinct repertoires by sequence, but that the biochemical property of the CDR loops also is likely to play a role. Furthermore, as this analysis takes into account all CDR loops, some of this segregation is due to distinct V gene usage by the two cell types, but among several WT Treg clusters there are TCRs of multiple V genes being clustered together, indicating the properties and edit distance between these V genes is a significant factor.

Second, there is a significant $C4^{KO}$ Treg population (dark blue) that clusters with the largest WT Treg cluster. This is due to its use of an identical TCRβ sequence; however, it uses a similar TRAV gene with distinct CDR3α sequences. Future studies to investigate the potential role of the shared TCRβ chain or the distinct properties of the TCRα chain in generating C4-reactive Treg specificities in $C4^{KO}$ mice are of great interest. Whether this TCR is cross-reactive to another self-ligand remains possible, and if that were the case its biochemical distinction from the WT Treg specificity would be of great interest.

Finally, the $C4^{KO}$ TCRs, in general, appear to broadly cluster with WT TCRs, with few isolated clusters of $C4^{KO}$ TCRs. If the C4 peptide were responsible for significant amounts of negative selection in the thymus, the appearance of significant new branches in $C4^{KO}$ mice would be expected. While a few such branches do appear, that can also be expected due to the

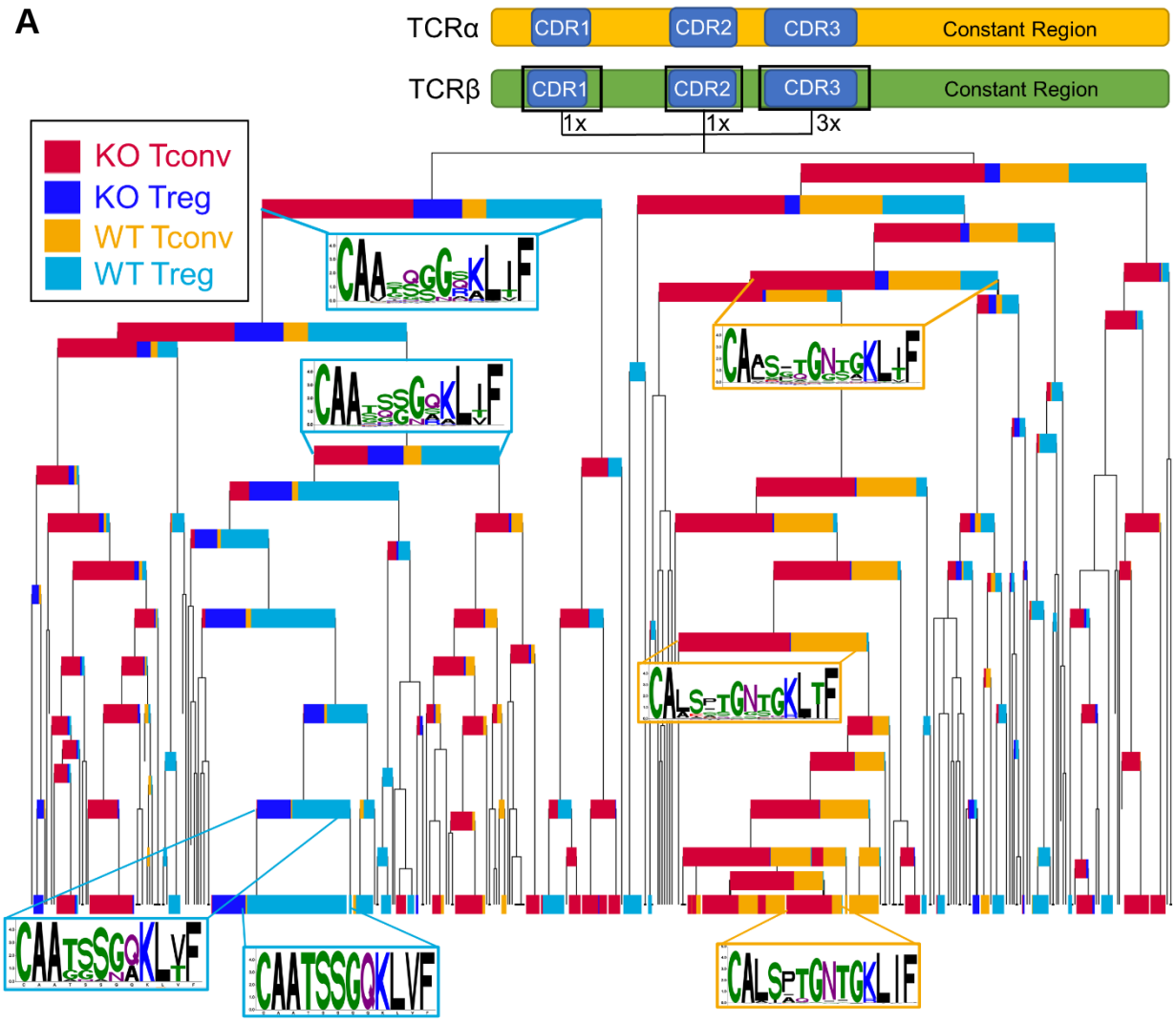

Figure 2.9 Hierarchical tree of C4-reactive TCRs reveals distinct clustering by phenotype
(A) All 901 C4-reactive TCRβ chains were analyzed and assigned distance scores by TCRdist3. Differences in amino-acid sequences of CDRs are totaled based on number of gaps and their BLOSUM62 substitution penalties with 3-fold weighting on CDR3 substitutions. The size of each cluster is determined by the number of TCRs in that cluster. The lowest level cluster exhibits a mean distance between TCRs of zero, and therefore represents clonally expanded TCRs and public TCRs. Overlaid on select clusters are CDR3α motifs of a given cluster to illustrate distance of TCR "neighbors" within each cluster. The motifs on the left are of a public Treg clone in WT mice and is correspondingly outlined in light blue. The motifs on the right represent a WT Tconv clone also found in $C4^{KO}$ Tconv cells and is outlined in orange.

diversity of unrestricted TCR repertoires across individuals, which typically might only share 10-15% of their TCRs[76].

This dataset also provided a collection of novel Treg cell TCRs for downstream biophysical analysis (see Chapter 3) in addition to allowing in-depth gene expression analysis and repertoire analysis of Treg cells and Tconv cells specific for the same antigen in antigen-sufficient and antigen-deficient contexts.

2.04 TCR repertoire analysis of C4/I-A^b specific T cells reveals an abundant public Treg clone in wild-type mice

By comparing the TCR repertoires of the WT mice alone, among the 376 T cells, 68% of C4-reactive cells are Treg cells, with the remaining third being Tconv cells. In the $C4^{KO}$ mice, however, 83% of the 525 cells are Tconv cells and the remaining 17% are Treg cells.

Interestingly, this dataset revealed ten public TCR clones that were found across mice with identical V-gene usage and CDR3 sequences. Of these, one highly expanded Treg clone was identified in all four WT mice, in addition to two out of five of the single-cell PCR-derived TCR repertoires discussed in section 2.2 (Figure 2.10A). This TCR will be referred to as WTR1 (for Wild Type Treg 1, since it comes from the WT mice), and utilizes the TRAV7-2 and TRBV 13-2 genes, bears the CDR3α sequence of CAATSSGQKLVF, and the CDR3β sequence of CASSDRYEQYF. In total, this clone was found in 51 Treg cells across the four mice, representing one of the top expanded clones in each mouse. Importantly, this clone was found in none of the C4-KO mice, indicating a dependence on C4 antigen presence in either the thymus for its selection and escape into the periphery, or, alternatively, in the periphery for the persistence of this clone through life. Without specific C4 antigen deletion experiments in the thymus and/or periphery it cannot be determined which is the correct explanation. Finally, this clone was found to exist in all

four of the Treg cell clusters (Figure 2.10B). Two clones were also found in the Tconv cell clusters cluster 0 and cluster 7 but upon closer inspection, these were revealed to be Treg cells with expression of Foxp3 and CD25 and to have been improperly binned by the UMAP algorithm, likely due to dropout of other key Treg features.

Another interesting identified clone utilized identical V and J genes as the WTR1 TCR, but with a distinct CDR3α and TRAJ gene (Figure 2.10A, B). However, this clone was found only in Treg cells in two out of four C4-KO mice, and zero of the nine WT mice sequenced in this 10x dataset and the single-cell PCR dataset, and is thus named KOR1 (for Knockout Treg 1). This TCR was found in 45 Treg cells overall in the knockout mice and in all Treg clusters (Figure 2.10B, right). The presence of this TCR in only $C4^{KO}$ mice can be explained by a number of potential factors. First, it could be negatively selected by the C4 antigen in the WT thymus. Second, it could persist as a Treg due to cross-reactivity with other ligands but be outcompeted by monospecific Treg specificities in WT mice. Both of these explanations are consistent with its development on Treg cells in the absence of C4 antigen and its absence in WT mice, and present an easily testable hypothesis for future experiments. If this TCR is found to have an affinity/kinetics for C4/I-A^b significantly above that of WT-derived Treg TCRs, it would favor the negative selection hypothesis. Conversely, if it was found to bind with similar or weaker affinity/kinetics, that would favor the cross-reactivity and competition hypothesis.

For the purposes of this thesis, discussion will be limited here to three more Treg TCRs identified in WT mice (Table 2). WTR1 was the only true public Treg TCR, but WTR2, a TCR utilizing the TRAV 7D-2 and TRBV 13-1 genes and the CDR3 sequences of CAASRGGRALIF_CASSDTRGEQYF, not only bore some similarities with WTR1, but also has two true “relatives”, defined as TCRs with identical gene usage and fewer than 4 amino acid

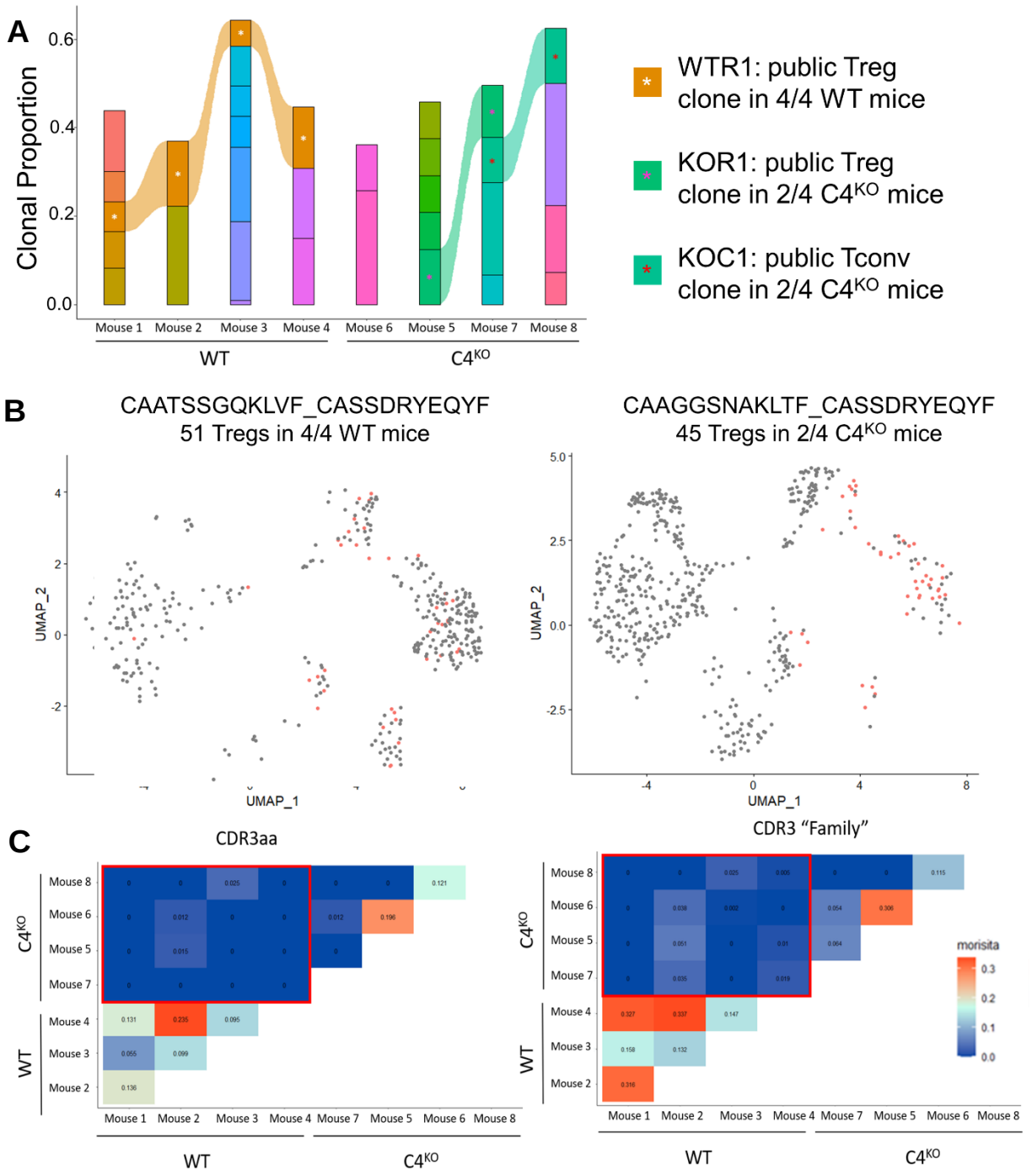

Figure 2.10 C4-reactive TCR repertoire reveals abundant public Treg clone in WT mice
(A) Alluvial plot of top 30 expanded clonotypes connecting shared TCR clones between mice ("public" TCRs). Public clones are connected between mice by similarly colored ribbons and denoted by distinctly colored asterisks. **(B)** Top two public TCR clones overlayed on UMAP by genotype. Left, top public clone in WT mice (WTR1). Right, top public clone in $C4^{KO}$ mice (KOR1). **(C)** Morisita Horn index of repertoire similarity across mice by TCR VDJ + CDR3 amino acid sequence (left) and by TCR VDJ + CDR3 "family" (right), allowing for three differences in amino acids.

differences across both CDR3 sequences. Thus, this is a “near-public” clone found in WT mice. Two additional Treg TCRs from WT mice, named WTR3 and WTR4, were also selected for further analysis.

Finally, four Tconv TCRs of interest were identified in WT mice (Table 2). Two of these clones (WTC1, WTC2) were found on Tconv cells in both WT mice and C4-KO mice and were found to be expanded in the C4-KO mice, thus indicating TCR-extrinsic factors are responsible for their lack of expansion in WT mice. Among the potential factors, Treg-mediated suppression is the likeliest factor, though through which mechanism of suppression it is occurring cannot be assessed with this data. A third public clone, WTC3, was found in Tconv cells in two out of the four WT mice, indicating that it can be recurrently developed into the Tconv lineage, as opposed to the Treg lineage. A fourth clone, WTC4, was found to have multiple “related” clones in Tconv cells in WT mice. This data, combined with the recurring Treg development of WTR1, convincingly shows that non-stochastic factors can play a role in Treg development in mice with fully diverse TCR repertoires. If a proportion of all self-reactive TCRs were to stochastically develop into Treg cells in the thymus, as some models have suggested, public self-reactive TCRs would not be expected to be found on the same cell type across mice.

Multiple other public clones were identified in only the C4-KO mice. To compare the overall similarity of TCR repertoires between these eight mice, Morisita-Horn similarity indexes were calculated (Figure 2.10C, left). Interestingly, the overlap of TCR repertoires among WT mice was high, the overlap of TCR repertoires of C4-KO mice was somewhat lower, and the overlap of WT repertoires with C4-KO repertoires was almost zero. These results were largely driven by the public TCRs discussed above, so additional similarity analysis was conducted on CDR3 “families” to identify any overlap of TCRs with the same gene usage and CDR3

WT Treg TCRs										
Genotype	TCR name	TRAV	CDR3α	TRAJ	TRBV	CDR3β	TRBJ	Mice	Celltype	Clones (by mouse)
WT	WTR1	7-2	CAATSSGQKLVF	16	13-1	CASSDRYEQYF	2-7	1,2,3,4	Treg	53 (5, 5, 13, 16)
WT	WTR2	7D-2	CAASRGGRALIF	15	13-1	CASSDTRGEQYF	2-7	2	Treg	14
WT	WTR3	7N-5	CAGNTNTGKLTF	27	16	CASSPGLGVSYEQYF	2-7	3	Treg	18
WT	WTR4	14-2	CAGDTNAYKVIF	30	20	CGARGGHSDYTF	1-2	3	Treg	17
WT Tconv TCRs										
Both	WTC1	7D-2	CAASRGQGGRALIF	15	12-2 +13-2	CASGESEQYF	2-7	2, 7	Tconv	4
Both	WTC2	14N-3	CAASGSGSWQLIF	22	12-2 +13-2	CASGDKNTLYF	2-4	2, 6	Tconv	3
WT	WTC3	12D-3	CALITGNTGKLIF	37	3	CASSKQGGQDTQYF	2-5	1, 4	Tconv	5 (4, 1)
WT	WTC4	3D-3	CAVIYQGGRALIF	15	12-2 +13-2	CASGDVLGGREQYF	2-7	2	Tconv	2
KO Treg TCRs										
$C4^{KO}$	KOR1	7-2	CAAGGSNAKLTF	42	13-1	CASSDRYEQYF	2-7	6, 8	Treg	45
$C4^{KO}$	KOR2	3D-3	CAVIYQGGRALIF	15	12-2 +13-2	CASGGLGGREQYF	2-7	5, 6	Tconv	33
Trogue TCRs										
Both	WTR5	14-2	CAGDTNAYKVIF	30	3	CASSRQGGQDTQYF	2-5	3, 8	Treg/Tconv	2
Both	WTR7	7D-2	CAASTSNNRIFF	31	12-2 +13-2	CASGKGQGSYEQYF	2-7	2, 6	Treg/Tconv	19

Table 2. Newly identified C4-reactive TCRs of distinct phenotypes

Table displaying TCR genes, CDR3 sequences, and phenotype of select TCRs from the 10x-derived C4-reactive TCR repertoire. The four sections are segregated by phenotype of cells bearing each TCR. “Trogue” TCRs are all TCRs found in the $C4^{KO}$ Tconv repertoire and WT Treg repertoire.

sequences within three amino acid differences (Figure 2.10C, right). This was to identify TCRs with potentially insignificant differences in molecular binding that would not have been captured in the previous analysis. This increased the overlap of all three comparisons by marginal amounts, with the overall trend remaining the same. Given the greater repertoire overlap among WT mice than among C4-KO mice or between the two genotypes, this indicates that without the C4 peptide presence there may be greater variability in the C4-reactive TCR repertoire, possibly due to a lack of negative selection or maintenance of T cells that persist in WT mice.

The lack of overlap between the WT and $C4^{KO}$ repertoires produces a number of interesting questions. To start, the WTR1 TCR was identified in 4/4 WT mice in the 10x dataset in addition to 2/5 of the WT mice from the single-cell PCR dataset. This represents a highly public Treg clone expanded in a majority of mice. Surprisingly, it appeared in 0/4 of the $C4^{KO}$ mice, indicating the presence of the C4 peptide in the thymus and/or periphery could be required for the positive selection, maintenance, or expansion of this clone in response to peptide. It is possible that the presence of other C4-reactive clones in $C4^{KO}$ mice can outcompete WTR1 in response to immunization. If this were the case, it could be present in the mice but not captured by 10x sequencing. However, this might be unexpected given the preliminary Treg affinity data discussed in the next chapter.

The initial hypothesis regarding the relationship between the WT and $C4^{KO}$ repertoires was that some Treg TCR clones would be found in $C4^{KO}$ mice and were likely to be in Tconv cells. Without the presence of C4 in the thymus, these cells would not be driven into the Treg lineage. However, other selecting ligands required for their positive selection should still be present, allowing for their development and egress from the thymus. Unexpectedly, none of the Treg TCR clones present in WT mice were present in $C4^{KO}$ mice. In addition, multiple other

public clones were identified in $C4^{KO}$ mice that were not present in WT mice, including KOR1 and KOR2 (Figure 2.10A, Table 2). These TCRs were each found in Treg cells in 2/4 $C4^{KO}$ mice and were highly expanded in their respective mice. These TCRs represent interesting cases, as they appear to be recurrently driven into the Treg lineage independently of the presence of C4 antigen in the thymus or periphery. It is possible that these are cross-reactive TCRs that bind to other self-epitopes, but more studies would be needed to explore that hypothesis.

To expand upon the similarity and repertoire overlap between mice, Morisita-Horn analysis was applied to multiple other subsets. These included analyzing the alpha chains alone, beta chains alone, Treg repertoires alone, and Tconv repertoires alone (Figure 2.11). By analyzing each TCR chain alone, it became apparent that there exists significantly greater overlap of exact TCRβ chain usages between mice than TCRα chains (Figure 2.11A). All mice, of both genotypes, exhibited significant TCRβ chain similarity. However, the repertoires of each genotype exhibited a much lower degree of TCRα chain overlap to explain the low degree of overall overlap.

As discussed above, when the overlap between Treg and Tconv TCRs were compared alone, there were zero Treg TCRs found to overlap between genotypes. However, when the analysis was expanded to allow for CDR3 “families”, again defined as three or fewer different amino acids between two otherwise identical TCRs, two WT Treg clones were found to have family members in $C4^{KO}$ Tconv clones (Figure 2.11B, Table 2). These represent likely “Trogue” clones – clones that occur on Treg cells in WT mice but fail to be diverted out of the Tconv lineage in $C4^{KO}$ or $Aire^{KO}$ mice, where the C4 peptide is not present in the thymus. Two of these three clones were minimally expanded or not expanded in WT mice but were dramatically expanded in $C4^{KO}$ mice in response to C4 immunization. To address these questions definitively,

these TCRs would need to be cloned and expressed in *in vivo* models of T cell selection to determine the effects of C4-antigen dependence on their development.

2.05 Summary of findings

Utilizing the recently discovered natural Treg ligand Tcaf3 and its epitope C4, this work sought to investigate the total C4-reactive TCR repertoire in mice with fully diverse TCR repertoires to improve upon previous studies completed in mice with fixed TCRβ chains. This work also sought to observe the effects of C4 antigen presence on the TCR repertoire. Using a $C4^{KO}$ mouse model developed by the Savage lab, mice could be immunized to "simulate" a foreign immune response to the C4 epitope to compare with the WT self-reactive immune response.

These data yielded many observations about the behavior of Treg and Tconv cells that corroborated work by many labs. Namely, upon immunization of the $C4^{KO}$ mice with C4 peptide, the C4-reactive T cell compartment experienced dramatic expansion and activation. However, in the WT mice, C4 immunization induced dramatic Treg expansion and activation, and almost a complete lack of expansion and activation of the Tconv compartment, likely due to Treg-mediated suppression.

However, the 10x Genomics-based approach taken in these studies allowed for many new insights into the self-reactive TCR repertoires in mice without fixed TCRβ chains. First, while several studies have compared the overall Treg and Tconv TCR repertoires and determined they exhibit minimal overlap, this study allowed for the first comparison of these repertoires specific to a single self-antigen and corroborated the other studies that there is minimal overlap. Furthermore, the Treg and Tconv repertoires specific for C4/I-A^b in WT mice were found to segregate by edit distance and biochemical property of their CDR loops. This important

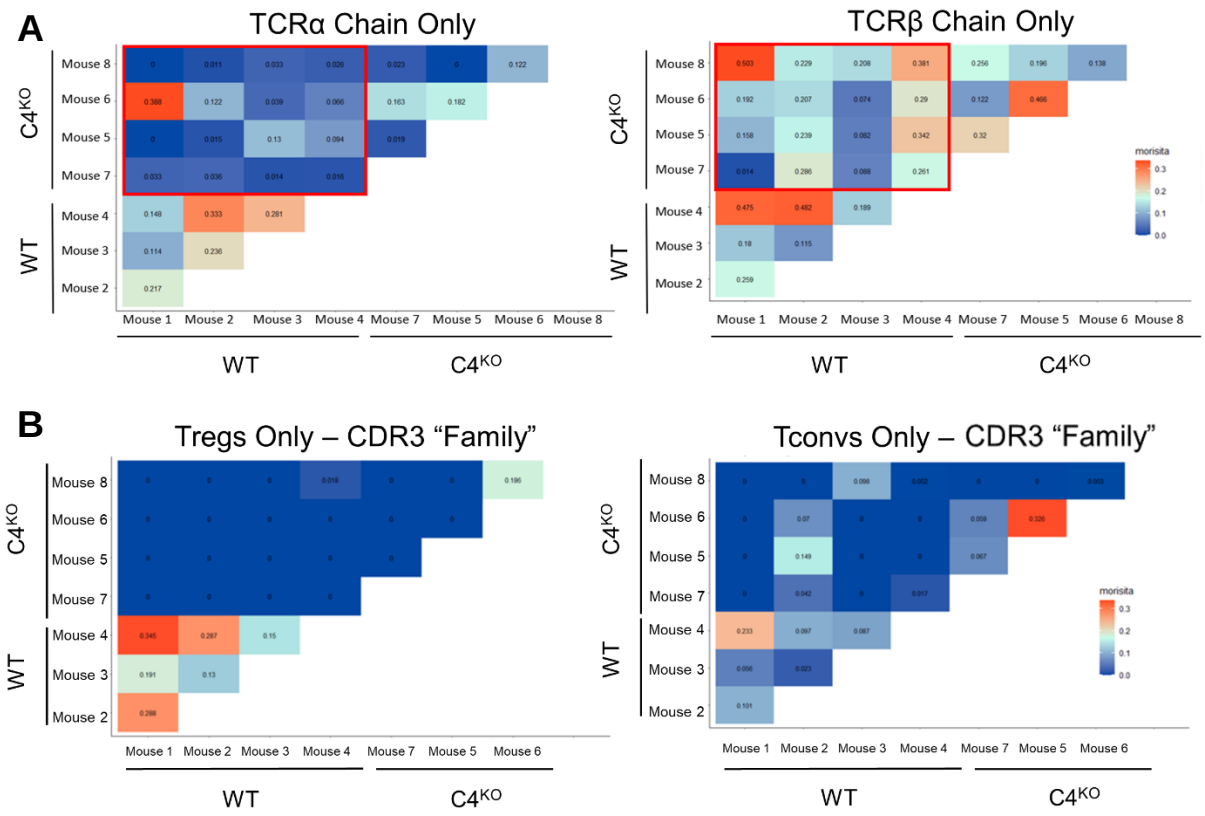

Figure 2.11 Repertoire overlap between mice by TCR chain and cell type reveals Trogue clones

(A) All individual mouse repertoires were compared by Morisita Horn index of similarity across mice by TCRα chain (left) and TCRβ chain (right) only. Redder squares indicate high overlap, bluer squares indicate lower overlap. **(B)** Morisita Horn index of repertoire similarity across mice by CDR3 "family" divided by Treg cells only (left) and Tconv cells only (right), colored as in (A).

finding adds to the support for TCR-intrinsic factors in determining Treg development.

Second, this study found multiple public clones with interesting phenotypes that can elucidate Treg biology and provide tools for future studies. The existence of TCR clones reactive to a single self-antigen that are recurrently diverted into the Treg and Tconv lineages is another important piece of evidence that TCR-intrinsic factors can determine T cell development. Additionally, the behavior of these clones between the two genotypes studies sparks further questions as to the requirement for self-antigen in thymic development. The lack of WTR1 and other Treg specificities in $C4^{KO}$ mice could indicate a role for self-antigen in positive selection or maintenance of Treg clones in the thymus or periphery, but further studies will be required to investigate this. The finding of multiple "Trogue" clones also corroborates previously published data from the Savage lab demonstrating that, in the absence of self-antigen, developing self-reactive thymocytes fail to be diverted into the Treg lineage and egress from the thymus as Tconv cells. These cells then can experience activation and expansion in the presence of their cognate ligand.

Third, this 10x dataset allows for the exploration of gene regulatory factors that could explain Treg function. The genetic regulation of Treg cells in WT mice and $C4^{KO}$ mice was found to exhibit significant differences in the expression of the transcription factor Klf2. It remains possible that peripheral encounter with antigen by Treg cells could "imprint" such cells with enhanced suppressor function. In $C4^{KO}$ mice, therefore, the Treg cells present could possess weaker suppressive capabilities compared to the Treg cells in WT mice. However, multiple confounding variables are present, including the total number of Treg cells per mouse, and further studies will be required to investigate the significance of these findings.

The comparisons of repertoire overlap conducted here must keep in mind the dramatic

variability of TCR repertoires between mice and only about 10-15% of TCRs are expected to be public across individuals[76–78]. Considering this, it is remarkable that multiple public Treg and Trogue clones were identified and these findings offer new insights into the behavior and development TCR repertoires specific to self-antigens. These results support a TCR-intrinsic model of Treg differentiation that will be further investigated in Chapter 3.

Chapter 3: Investigating the biochemical differences between TCRs of Treg and Tconv cells specific for a single antigen

3.01 Introduction

There has been significant debate surrounding the affinity with which Treg TCRs bind their cognate self-antigens and the significance of that affinity in their development and function. However, to date, there have been few true self-antigens known to be important for Treg development. The Adams lab and Savage lab recently identified a *bona fide* Treg self-antigen expressed in the prostate and in the thymus under Aire-dependent translation. The objective of the research in this chapter was to characterize the interaction of Treg and, where possible, Tconv cell TCRs specific for the C4 ligand to elucidate the molecular mechanisms by which they might differ. The primary characteristics of interest are the affinity, binding kinetics, and docking mode of the interaction between TCRs and C4/I-A^b. To answer these questions, a mix of SPR and crystallography is required.

3.02 Initial biochemical analysis attempts with iRepertoire-derived TCRs

The purpose of the experiments in this sub-chapter was to characterize the binding kinetics of Treg TCRs specific for the C4/I-A^b ligand. Work conducted in the Savage lab had used the iRepertoire sequencing strategy to previously identify a panel of TCRs reactive to C4/I-A^b, listed in Table 1. These TCRs include the previously identified MJ23 TCR. To begin the expression of these TCRs, each was cloned into the pACgp67a construct for baculovirus expression in the Hi5 insect cell line. Each construct was tested with both murine constant domains and human constant domains, and was cloned with an engineered cysteine to increase stability of the heterodimeric complex. To induce heterodimerization, as opposed to

homodimerization, which can occur in recombinantly expressed TCR chains, the alpha and beta chains of each TCR were engineered with complementary acidic and basic leucine zippers, respectively. To aid in purification, these zippers culminated in 6x histidine tags for a total of 12 histidine residues for use Ni^{2+}-resin affinity chromatography. Finally, these constructs were engineered with 3C protease cleavage sites between the C-terminus of the constant domain and the N-terminus of the leucine zippers for 3C-based removal of the zippers and histidine tags after purification. This cloning was completed for the MJ23, SP33, RET, TSS, TGN, and VSN TCRs.

Initial expression testing and optimization was conducted on MJ23 alone. This included altering the infection time from 60-72 hours, purification buffer salt concentrations and pH, the Ni^{2+}-resin incubation time from 3-24 hours, additional purification steps (adding/subtracting steps such as Ni^{2+} cleanup, ion-exchange chromatography, and avidin affinity chromatography), and construct design. Ultimately, expression of the protein was sufficient for purification, but stability of the protein through purification was a significant hindrance to final yield.

For affinity studies, the MJ23 TCR was engineered with an AviTag^{TM} site that can be covalently linked to a biotin molecule *in vivo* with the addition of biotin and the BirA biotin ligase. This can then enable the TCR to be immobilized on a streptavidin-based biosensor for biolayer interferometry studies to glean affinity and kinetic information. Additionally, the addition of the biotinylation site allowed for purification of the TCR with Avidin affinity chromatography. Avidin is a mutated form of streptavidin that reversibly binds biotin, allowing for release of protein upon addition of 2mM biotin solution. This purification protocol was found to improve the purity but decrease yield of M23 and SP33 TCRs.

Ultimately, to achieve preliminary measurements, an attempt was made to purify each of these TCRs with only Ni^{2+} chromatography and size exclusion chromatography. This was the

fastest and gentlest method of purifying these proteins, but left a significant number of contaminating proteins in addition to the TCRs (Figure 3.1A). C4/I-A^b was produced in sufficient quantity and yield via a protocol using S2 drosophila cell lines developed by previous lab member Dr. John Leonard in collaboration with Marc Jenkins' laboratory. Each of the TCRs expressed here is specific for the C4/I-A^b protein with the exception of the SP33 TCR, which is specific for F1/I-A^b and was used as a control in this experiment. Initial binding experiments were noisy for all TCRs, but the RET TCR yielded the most promising results for optimization (Figure 3.1B). The TSS and VSN TCRs appeared to not bind C4/I-A^b to any significant degree. This could have been due to low yield or contaminating proteins and it cannot be concluded that they do not bind in physiological contexts. The VSN heterodimer also appeared to exhibit some instability when analyzed via SDS-PAGE non-reducing gel (not shown). Upon producing greater amounts of RET and purifying it to a higher degree with avidin affinity chromatography, repeat measurements were obtained with the BLItz machine as seen in Figure 3.1C, D. While these measurements were cleaner and suggested a possible K_D value of approximately 43μM +/- 15μM, the ultimate range of measurements was insufficient to obtain results with high confidence.

3.03 Biophysical analysis of 10x Genomics-derived TCRs

A repeat problem observed with each of the iRepertoire-derived TCRs was that they all possess the same TCRβ chain TRBV26, which appears to dramatically limit the expression of paired TCRs independently of how well the TCRα chain expresses. Thus, additional TCRs were sought outside of the fixed transgenic TCRβ chain models that all previous experiments had been conducted in. All other Treg TCRs with identified cognate antigens[79] have been discovered in fixed TCRβ chain transgenic mice as well, so these studies represent the first characterizations of

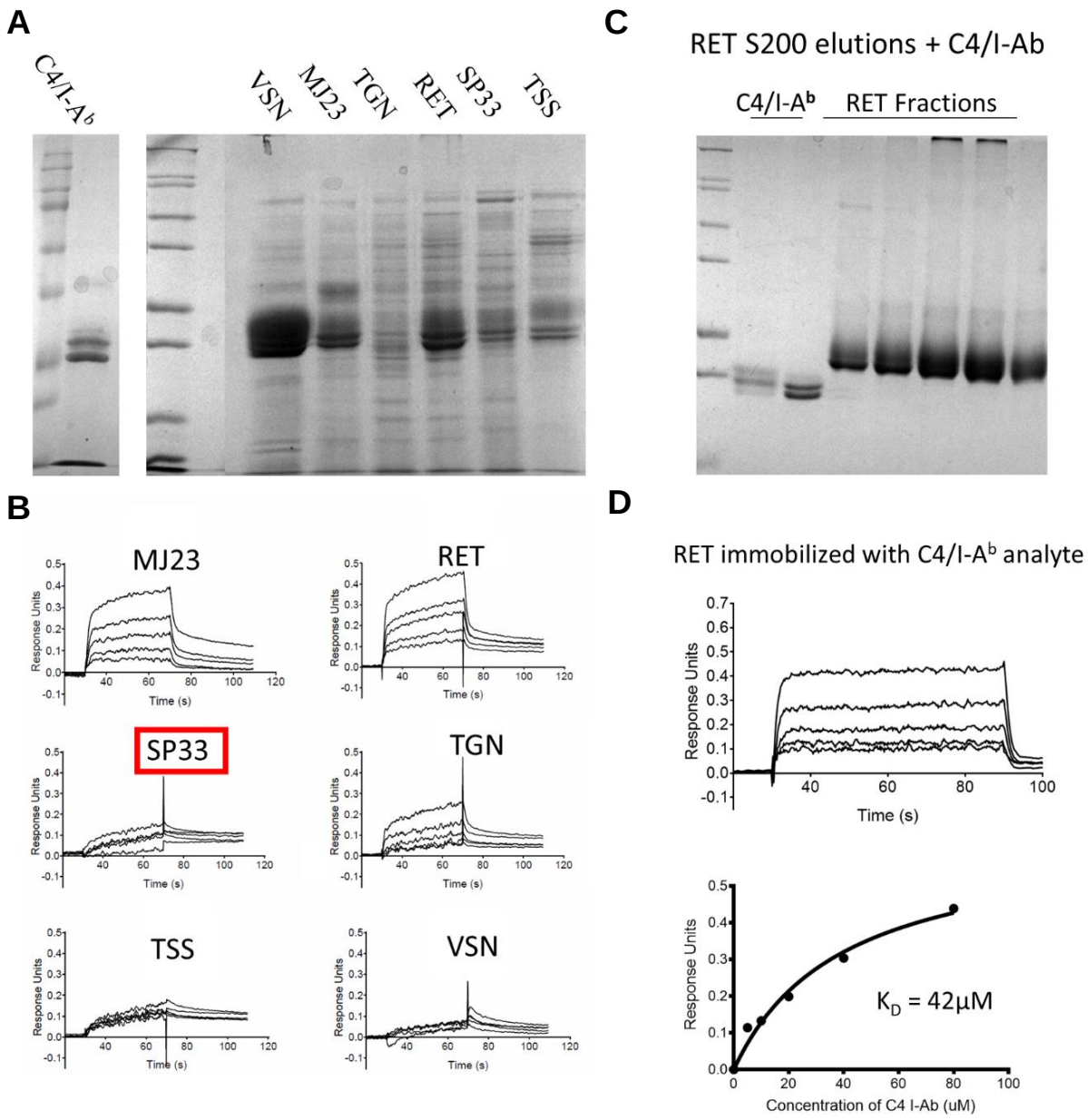

Figure 3.1 Affinity analysis of iRepertoire-derived TCRs
(A) SDS-PAGE reducing gel of iRepertoire-derived TCRs and C4/I-A^b post-size exclusion chromatography. Both MHC and both TCR chains visible. **(B)** Biolayer interferometry measurements of all iRepertoire-derived TCRs with C4/I-A^b, including SP33 as a negative control (red box). C4/I-A^b was immobilized on streptavidin biosensors and serial dilutions of each TCR were used as the analytes. **(C)** SDS-PAGE reducing gel of the RET TCR post avidin-affinity chromatography and C4/I-A^b post-size exclusion chromatography. **(D)** Biolayer interferometry of RET immobilized on the streptavidin biosensor and C4/I-A^b as the analyte.

Treg TCRs derived from fully diverse TCR repertoires.

From the VDJ data derived from the experiments described in Chapter 2, each unique TCR CDR3 sequence from wild-type mice was binned into Treg or Tconv cell types based on gene expression profiles. These constituted two distinct wild-type repertoires for the C4/I-A^b antigen, with 254 and 124 TCRs, respectively. These two repertoires were analyzed for the biophysical properties of their CDR sequences in the Automatic Immune Molecule Separator (AIMS) package developed by Dr. Chris Boughter. This package is able to analyze CDR1, CDR2, and CDR3 loops of two different repertoires for differences in charge, hydrophobicity, flexibility, bulkiness, and ten Kidera factors (such as double-bend preference, side-chain size, and flat extended preference) both on a loop-wide level and an individual amino acid level. TCRs that were clonally expanded were left overrepresented in the repertoires to properly capture their influence, though each repertoire was later analyzed with each unique TCR being represented once as well. Analyzing CDR length, AIMS found that Treg TCRs possessed longer CDR3s and CDR1α and CDR2α chains (Figure 3.2A). The difference in CDR3 length is of particular interest, as longer CDR3s are less likely to be generated via VDJ recombination[80,81], indicating possible selection for the biochemical properties of these TCRs. Analyzing the CDR loops as a whole, AIMS found minor, but significant differences between the Treg and Tconv TCR repertoires specific for the C4/I-A^b antigen (Figure 3.2B). Treg TCRs were found to be less hydrophobic, more charges, and have greater bulkiness and flexibility than Tconv TCRs.

Interestingly, when analyzed on the scale of individual amino acid positions in each CDR loop, AIMS found many significant differences. In the center of each CDR3 chain, AIMS found significant increases in charge in Treg TCRs (Figure 3.2C). This corresponded with decreases in hydrophobicity (Figure 3.2D) and a maintenance of Shannon entropy through these positions

(Figure 3.2E). The conservation of entropy indicates that there is comparable diversity of amino acids contributing to the Treg and Tconv CDR3 sequences, therefore differences in charge are likely contributed to by multiple amino acids at each position, increasing the potential for biological relevance to TCR signaling.

Another interesting observation was made regarding CDR1α and CDR2α. In each case, Shannon entropy was dramatically reduced in Tregs vs. Tconv. This is due to the large contribution of the TRAV7-2 gene to the Treg repertoire (~50%, Figure 2.7A). It appears this TRAV gene confers dramatic and significant changes in charge properties to CDR1 sequences. This is due to two aspartic acid residues and an arginine in the CDR1 sequence (full CDR1 sequence: "DRNVDY", full CDR2 sequence: "IFSNGE"). While CDR1 and CDR2α chains do not frequently exhibit large numbers of contacts with pMHC complexes,[79] there is a recent crystal structure of a Treg TCR in complex with self-pMHC that exhibited contact from a CDR1α residue with a peptide-derived residue. However, the TRAV7-2 gene has yet to be crystallized, and the direct significance of these interactions will remain unknown until alanine mutagenesis and/or crystallographic studies are conducted.

3.04 Affinity and kinetic analysis of 10x-derived Treg and Tconv TCRs specific for C4/I-A^{b}

The 10x Genomics dataset described in Chapter 2 provided a rich dataset of C4-reactive TCRs from WT mice to select from for biochemical studies. To start, eight TCRs were selected, four each from the Treg and Tconv repertoires (Table 2, top two panels). Wild type Treg TCRs were named in the format "WTR#", with "#" representing the number of each TCR. Wild type Tconv TCRs were named "WTC#", again with "#" representing each TCR's unique number. The "WT" in each name represents "wild type", the genotype of the mice these TCRs were identified

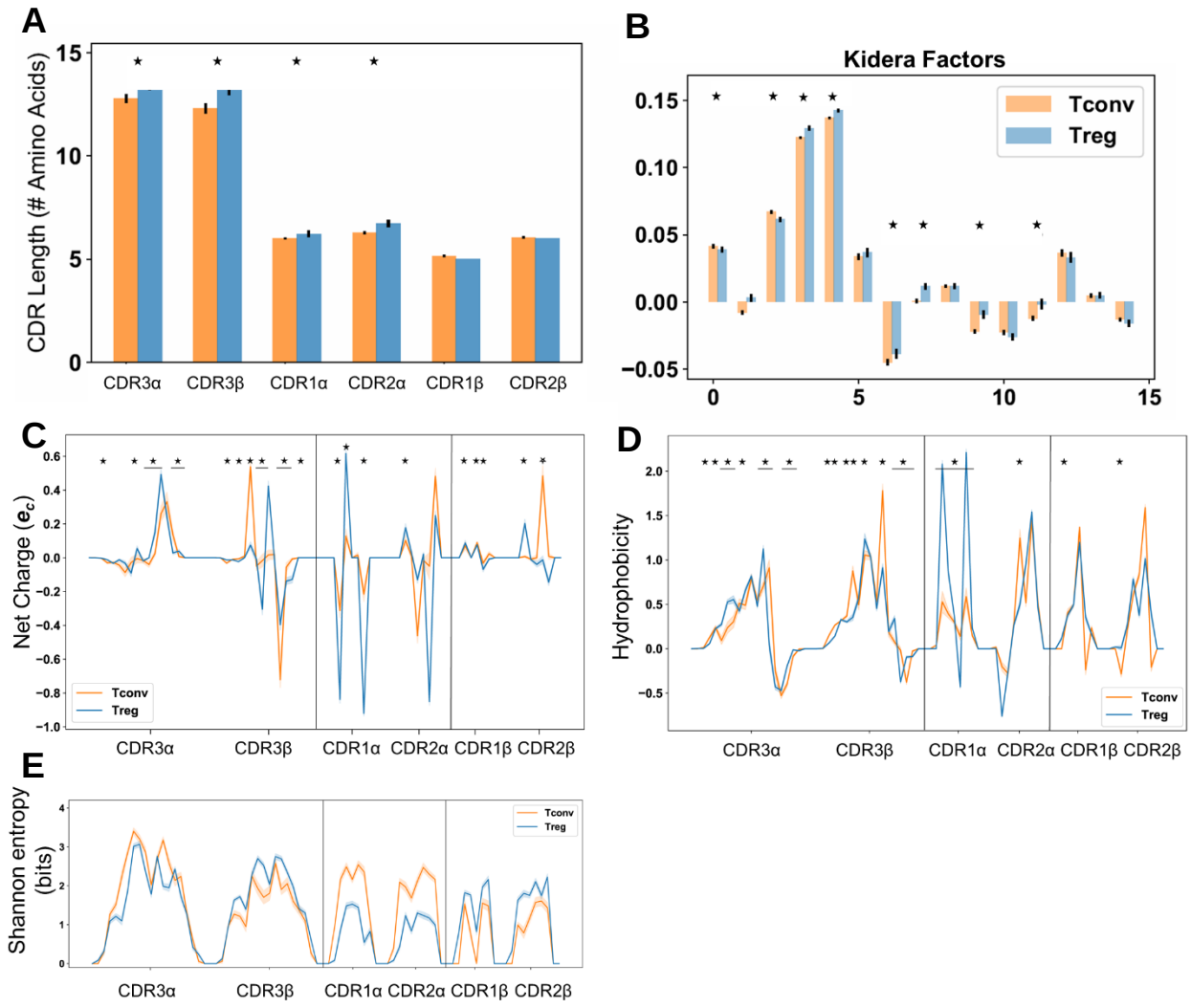

Figure 3.2 AIMS analysis of C4-reactive Treg vs. Tconv clones in WT mice reveals differences in CDR sequence charge

(A) Analysis of CDR3 length of WT Treg (blue) and Tconv (orange) TCRs by AIMS. Analysis conducted on entire CDR loops. **(B)** Kidera factors 1-15 analysis of whole CDR loops of Treg vs. Tconv TCRs by AIMS. **(C)** Positional analysis of the charge of CDR loops of WT Treg and Tconv TCRs. Positive and negative on the scale correspond to positive and negative charge at each position. **(D)** Positional analysis of hydrophobicity of CDR loops as in (C). Positive scale corresponds to less hydrophobicity and negative corresponds to increase hydrophobicity. **(E)** Positional Shannon entropy by CDR loop of WT Treg vs. Tconv TCRs. Higher values equal greater diversity, with maximally diverse sequences equating to a diversity of 4.12 bits. Treg repertoire n = 255, Tconv repertoire n = 121 TCRs.

in. The “R” and “C”, respectively, represents Treg and Tconv, the cell types in which each TCR was identified.

The selection criteria for these TCRs were the following: first, public TCRs were given highest priority. This yielded WTR1 and WTC3. Next, “near public” clones were given second priority. This included any TCRs for which related clones with three or fewer amino acid variations were found across all CDR loops. This included WTR2, WTC1, WTC3, and WTC4. Finally, highly expanded clones were given lowest priority to fill out the selection panel. This included WTR3 and WTR4. This panel is listed in full in Table 2. As can be seen in Figure 3.3, the expression of cells bearing these TCRs was varied in cell cluster by UMAP in the dataset. WTR1 was highly expanded and present in all four Treg clusters, whereas WTR3 was largely confined to cluster 1. In several cases (WTR1, WTR2, and WTR3), one or a few clones were found in Tconv cell clusters. However, these cells were still found to express Treg genes, but had one or more Treg-related genes significantly reduced. In all cases, these exceptional clones bore identical nucleotide sequences to the primary expanded clone, indicating a high likelihood of related lineage. Thus, the most likely explanation for these aberrations is cell dropout due to sequencing artifacts, rather than alternative development of a single clone in the middle of an immune response. As discussed in the previous chapter, the identification of clones in each of the four Treg clusters indicates that expanding cells are moving between different clusters as they enter different stages of the cell cycle, achieve greater time since activation, or are activated again. However, in all cases, the primary cluster these Treg cells appear to reside in is cluster 1, the largest Treg cluster.

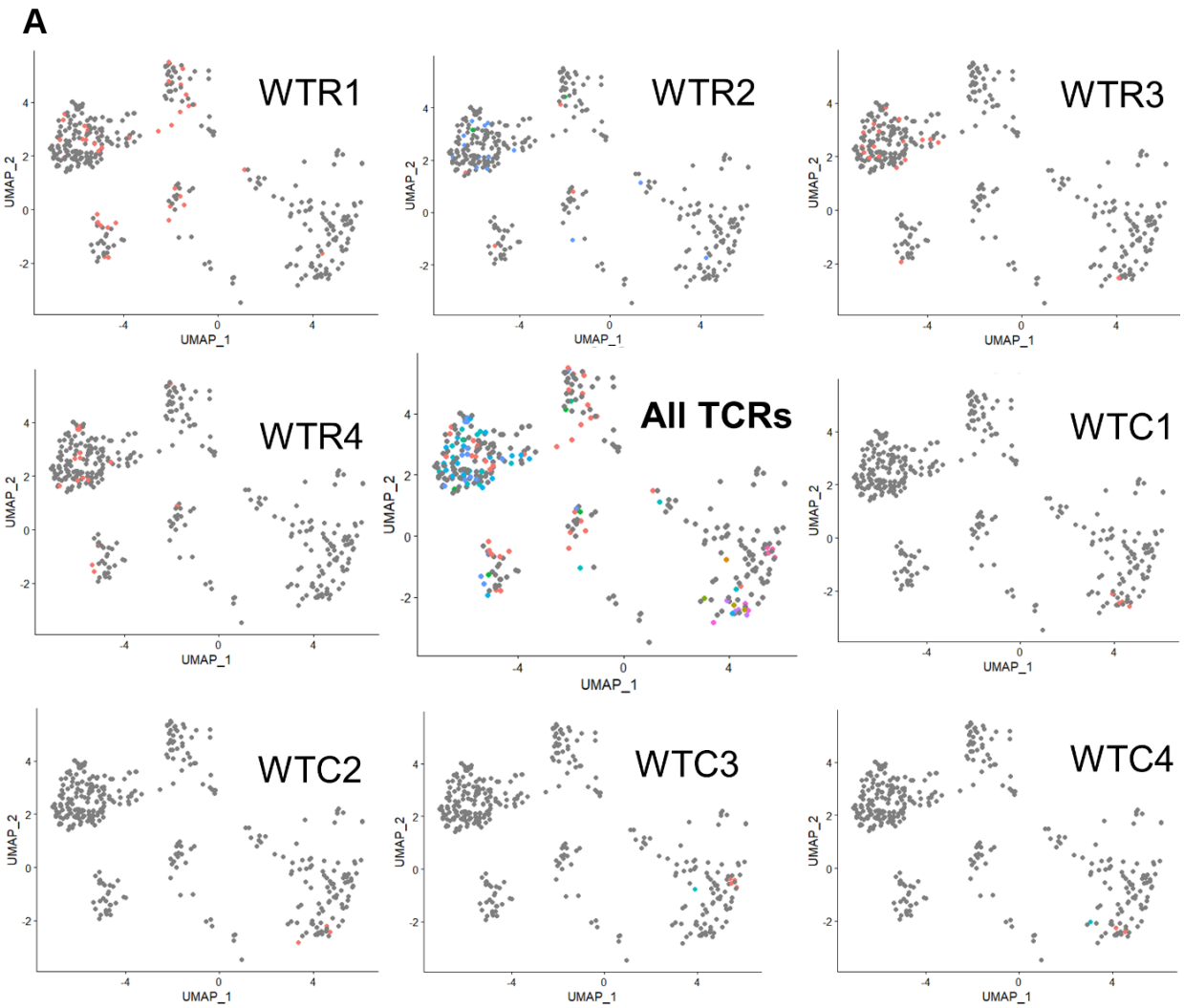

Figure 3.3 UMAP projections of eight TCRs selected for biochemical analysis
(A) UMAP overlay of all cells expressing the eight TCR specificities present in the 10x dataset selected for biochemical analysis. Each panel is expression of one TCR (labeled) and any "family" members, if applicable. Center panel includes all eight TCRs on one plot.

In the case of the selected WT Tconv cell TCRs, these were expanded to a dramatically lesser degree than the Treg TCRs. The maximally expanded Tconv clone was found only four times, compared to the 53 times for the maximally expanded Treg TCR in the selection panel. In both cases, these were representative of the overall expansion of these two cell types in wild-type mice. Interestingly, the WTC1 and WTC2 TCRs were found in only one WT mouse but they also found in one C4-KO mouse each. In the C4-KO mice, WTC1 was expanded and its clones exhibited an activated phenotype compared to the clones identified in the WT mouse, indicating its activation was likely controlled by TCR-independent forces, the most likely of which would be Treg-mediated suppression in WT mice. The WTC2 TCR, however, did exhibited neither expansion nor activation in C4-KO mice, possibly due to being outcompeted by other TCRs. WTC3 was the only true public Tconv TCR identified and was found in two mice, in addition to one closely related TCR. WTC4 was closely related to another Tconv TCR also found in WT mice. Each of the WTC1, WTC2, and WTC4 TCRs appeared to cluster closely by UMAP, likely due to the same TCRβ chain expression. However, they each express different CDR3β sequences and entirely different TCRα genes.

The location of each of these TCRs on the hierarchical tree first discussed in section 2.03 can be seen in Figure 3.4. It can be observed that these eight TCRs comprise a broad selection of branches on the tree, indicating distinct biochemical properties of their CDR loops. This is especially important due to the same V gene use (TRBV12-2+13-2, a common splice variant of TRBV12-2 and TRBV13-2) of three of the $C4^{KO}$ TCRs selected.

These TCR sequences were then identified, optimized for insect cell expression, and cloned into the identical pACgp67a plasmids described in subchapter 3.02. Unlike the iRepertoire-derived TCRs, many of these TCRs expressed to a high degree. Unfortunately,

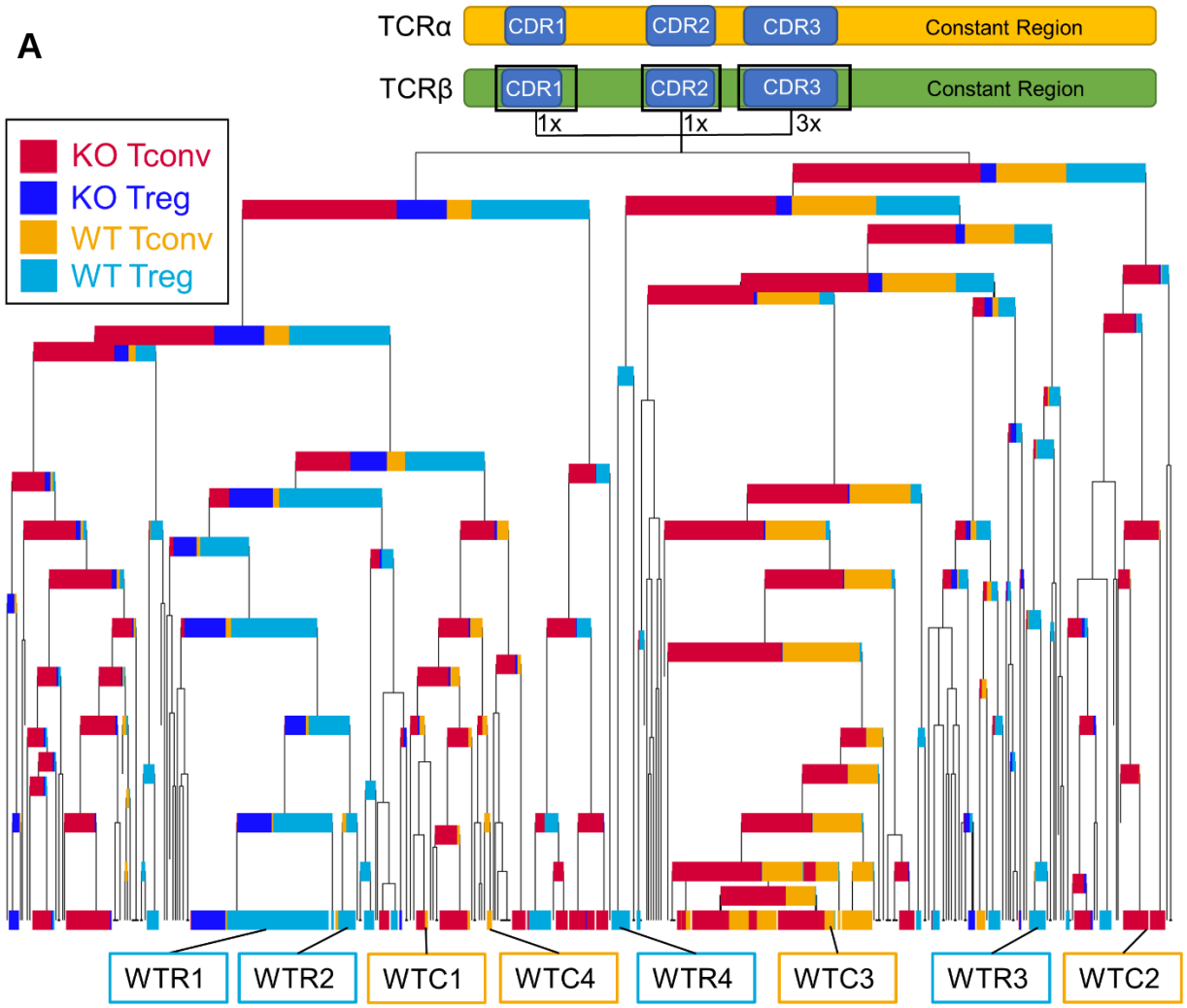

Figure 3.4 Hierarchical tree location of eight TCRs selected for biochemical analysis
(A) All 901 C4-reactive TCRβs were analyzed and given distance scores by tcrdist3. Differences in amino-acid sequences of CDRs are totaled based on number of gaps and their BLOSUM62 substitution penalties with 3-fold weighting on CDR3 substitutions as in Figure 2.9. Overlayed on bottom are the eight TCRs selected for recombinant expression colored by cell type.

WTR1, which was found in all four WT mice and therefore the most interesting TCR in the dataset, exhibited zero expression of the TCRβ chain, despite robust expression of the TCRα chain. However, WTR2 and WTR3 expressed to a robust degree, with WTR4 expressing to a lesser degree as well. Conversely, all four Tconv TCRs expressed to extremely high degrees. From initial 1L protein preps over 2 milligrams of pure protein was obtained.

Given this high expression, C4/I-A^b was the limiting factor for affinity studies. Thus, C4/I-A^b was expressed with an AviTag and biotinylated such that it could be immobilized on streptavidin biosensors, leaving the TCRs to be used as the analytes. F1/I-A^b was also expressed and biotinylated as a negative control for non-specific binding. Using the BLItz biolayer interferometry machine, C4/I-A^b was immobilized and blocked with free biotin. Then, up to seven serial dilutions of each TCR were analyzed in order of increasing concentration. Binding curves and calculated K_D affinity constants were obtained for each TCR (data not shown). Each TCR bound C4/I-A^b significantly, with the exception of WTC4, and none of the TCRs bound F1/I-A^b to any appreciable degree (Figure 3.5C). WTC4's lack of binding is consistent with previous observations in our work and in others showing some T cell specificities can be stained with pMHC tetramers without binding the same pMHC *in vitro*. Additionally, this clone was present to a very low degree in the 10x-derived dataset.

To improve the quality of affinity and kinetics measurements, SPR was conducted on the seven TCRs that expressed: WTR2, WTR3, WTR4, WTC1, WTC2, WTC3, and WTC4. These TCRs were expressed and purified in a different manner to the protein used in the BLItz studies described above, with an additional Ni^{2+} cleanup step to increase the purity of the final protein. The experiment itself had a similar setup, with the C4/I-A^b immobilized on the sensor chip and

each TCR used as the analyte. High-quality sensorgrams were obtained and fitted for each TCR (Figure 3.5A, B).

The measured curves of Figure 3.5 and the dissociation curves (Figure 3.6A) provide high confidence affinity and kinetics values for comparing Treg and Tconv TCR interactions with C4/I-A^b. The association rates of each were found to be indistinguishable (Figure 3.6B, left). In contrast, the measured affinities revealed differently trending values for the Treg and Tconv TCRs (Figure 3.6B, right). The Treg TCRs bound with an average of 22 +/- 12μM, while the Tconv TCRs bound with 79 +/- 58μM. Additionally, the k_{off} rates of the Treg TCRs were markedly slower than those of the Tconv TCRs, with average values of 0.20 +/- 0.05 (1/s) and 0.90 +/- 0.41(1/s), respectively (Figure 3.6C, left). Students t-test provides p-values of 0.17 and 0.04 for affinity and off-rates between Treg and Tconv repertoires, respectively (Figure 3.6). Calculating the half-life of the interactions, the Treg TCRs yielded a much longer dwell time of 3.7 +/- 0.8 s compared to Tconv TCRs' values of 0.9 +/- 0.5 s (Figure 3.6C, right). Finally, using K_{on} and $t_{1/2}$ to calculate the average confinement time of the interactions using standard values for dissociation[82] yielded the strongest separating variable yet for Treg vs. Tconv TCRs. Thus, with this small initial sample size, the length of interactions between TCR and pMHC appears to be of greater significance in distinguishing Treg from Tconv TCRs, though affinity still plays a large role in increasing the overall stability of the TCR/pMHC interaction and thus the expected T cell signaling strength.

This indicates that, for C4-reactive thymocytes, it is likely that thymocytes that bind C4/I-A^b with high relative affinity and low off-rates are directed into the Treg lineage, whereas thymocytes that bind C4/I-A^b with a lower affinity are recurrently driven into the Tconv lineage. While this cannot be conclusively proven for Treg development in general without further

measurements of additional TCRs in complex with self-antigens, these benchmark affinities indicate, for the first time, the potential affinities and kinetics with which a range of self-reactive TCRs can induce Treg vs. Tconv development.

3.05 Attempt at solving the crystal structure of a Treg TCR in complex with C4/I-A^b

An early goal of this project was to solve the complex crystal structure of the MJ23 TCR bound to C4/I-A^b. MJ23 was produced in Hi5 insect cells with a baculovirus transfection system and purified using a variety of techniques described in section 3.2. However, yields continued to remain low, and yields were only sufficient for a few crystal trials. A likely problem in crystallization is the significant glycosylation of the MJ23 TCR, which has five predicted N-linked glycosylation sites. I-A^b has an additional three predicted glycosylation sites. These glycosylation sites increase heterogeneity of the protein and reduce likelihood of crystallization, and thus are commonly removed via treatment with the EndoF3 glycosidase. While I-A^b responded well to treatment with EndoF, MJ23 did not respond well and precipitated out of solution upon purification to remove remaining EndoF. This EndoF treatment dramatically reduced yield to below 0.5mgs per 6L of insect cell culture. The stability of the protein was also decreased, as evidenced by high sensitivity to the concentration required for crystallization.

The Pro-Complex and PEGS II crystal screens were both utilized with the MJ23:C4/I-A^b complex, with no success. Additional PEGS II screens and specialized screens were set up with C4/I-A^b alone. The specialized screens were set up based on the conditions used in previous crystal structures of I-A^b, but again, no hits were obtained.

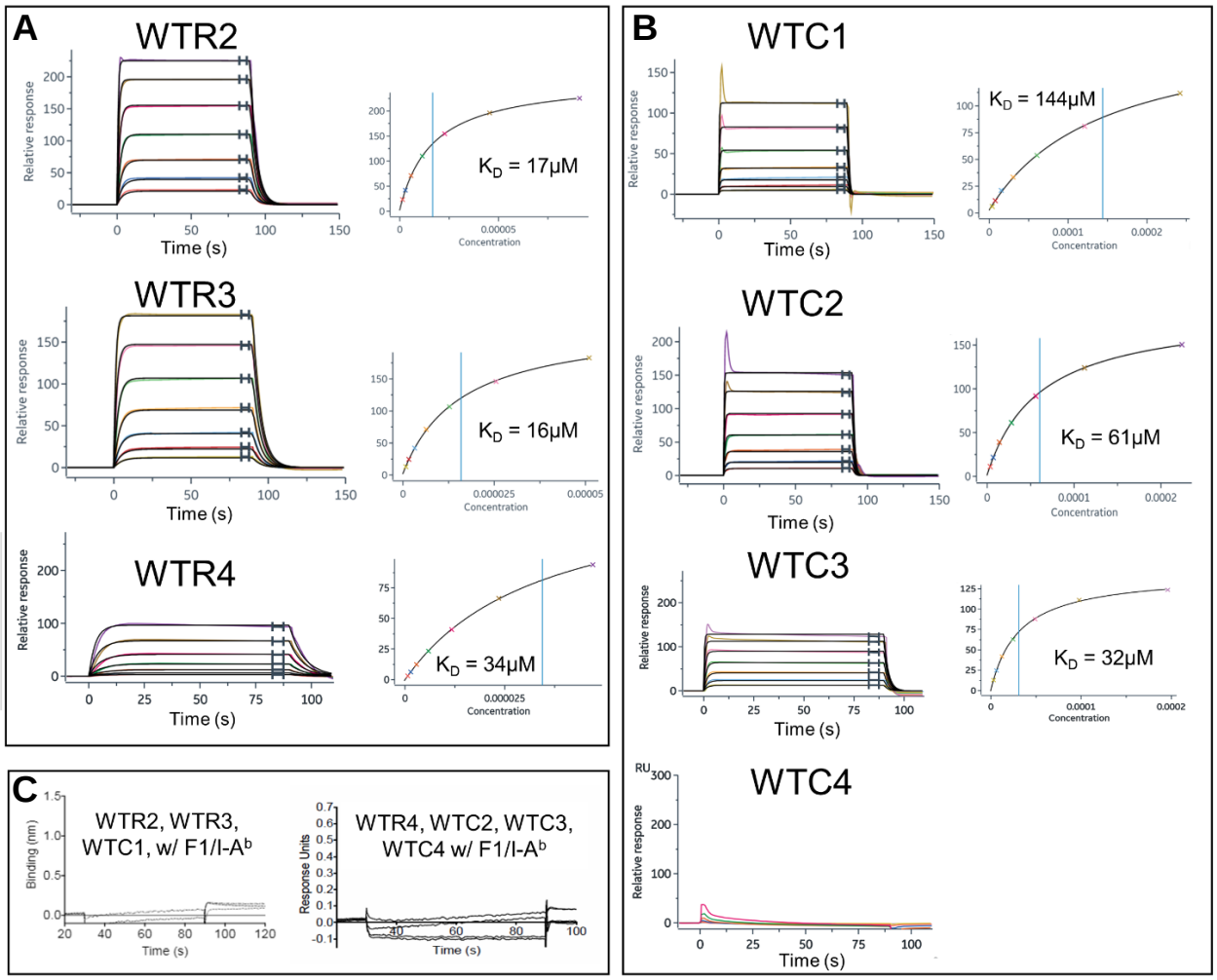

Figure 3.5 SPR affinity measurements of C4-reactive Treg and Tconv TCRs from WT background

(A) SPR of WT Treg TCRs with immobilized C4/I-A^b. C4/I-A^b was immobilized on a streptavidin chip and each TCR was used as an analyte. Fitted curves are in black, raw curves are in color. Equilibrium curves are shown on the right with calculated affinity constants. **(B)** SPR of WT Tconv TCRs with immobilized C4/I-A^b as in (A). **(C)** Biolayer interferometry of each TCR with F1/I-A^b as a negative control for specificity of binding. Each TCR was run at maximum concentration and zero binding was observed.

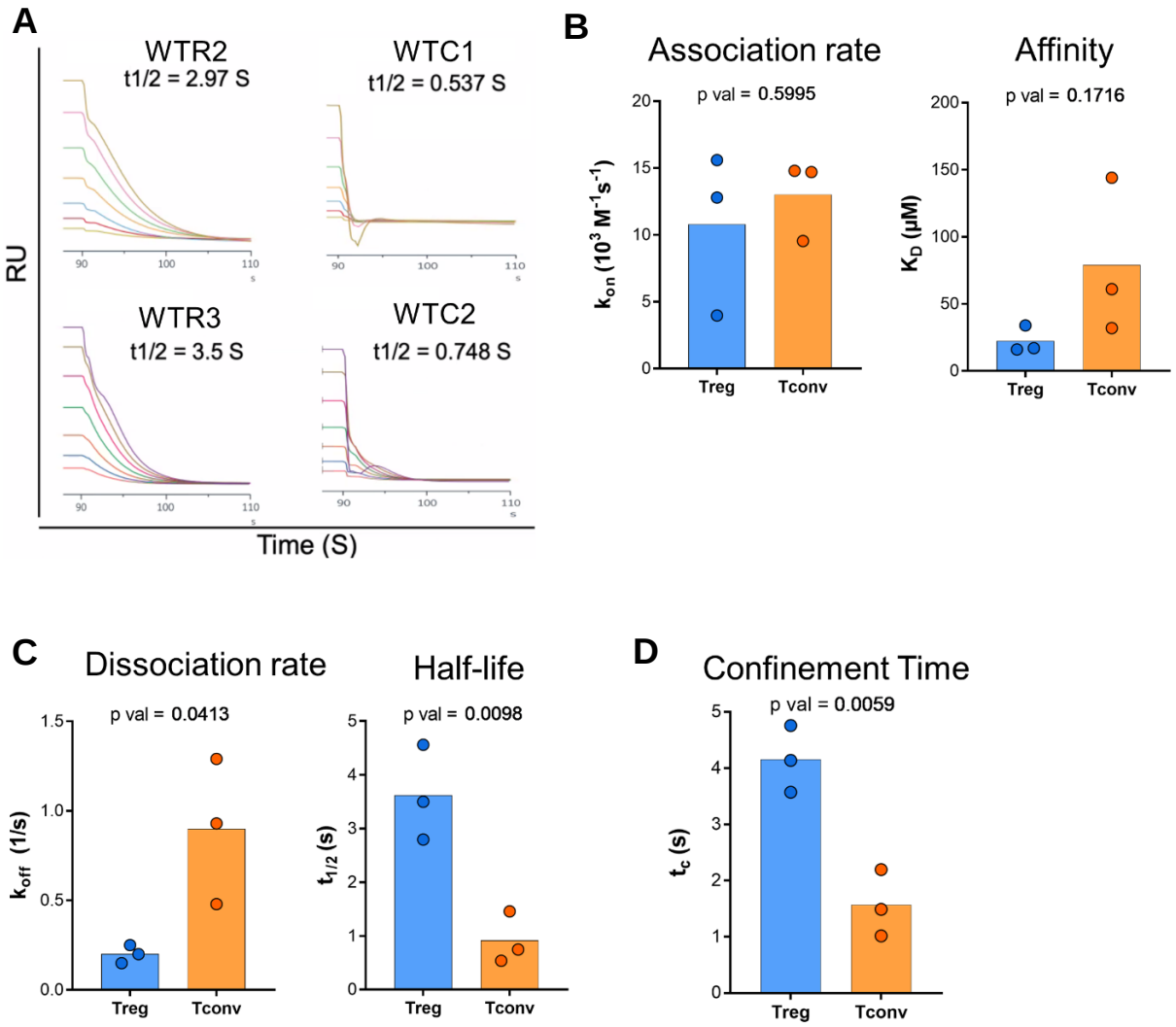

Figure 3.6 C4-reactive Treg TCRs bind with longer off-rates than Tconv TCRs
(A) Off rates visualized from select TCRs in Figure 3.5 to contrast Treg vs. Tconv off rates (units = seconds). **(B)** Association constant (k_{on}) and equilibrium constant (K_D) for all six C4-reactive TCRs measured. Treg (blue) and Tconv (orange) compared by students T-test and p-values are listed on each bar graph. **(C)** Dissociation rate (k_{off}) and half-life ($t_{1/2}$) for all C4-reactive TCRs measured and analyzed as in (B). **(D)** Confinement time (T_c) of C4-reactive TCRs, data is analyzed and displayed as in (B). n = 3 Treg TCRs and n = 3 Tconv TCRs.

Now that highly stable, highly expressing TCRs have been obtained from the 10x Genomics dataset, additional crystal trials are being conducted with the WTR2 and WTR3 TCRs. To date, a few potential hits have been obtained in the Pro-Complex screen, but sufficient time for optimization and additional trials has not yet passed.

3.06 Summary of findings

These findings provide the first benchmark measurements of diverse *bona fide* Treg and Tconv TCR affinities and kinetics for a self-antigen. They provide an important framework and basis for the kinetics with which self-reactive thymocytes encounter their antigens and are diverted into Treg or Tconv lineages. These findings indicate that recurrent, public Treg TCRs can bind their cognate ligands with higher affinity and significantly longer off-rates than recurrent Tconv TCRs specific for the same self-antigen.

Chapter 4: Discussion and Future Directions

4.01 The role of self-antigen in shaping the CD4+ T cell compartment

Bona fide Treg self-ligands and the TCRs that bind them have been unknown until recently. Therefore, it has been difficult to investigate the relationship and biology of Treg cells to specific self-antigens. This has meant that much of the research conducted to date has been conducted on tTreg and pTreg compartments as a whole, not on antigen-specific repertoires. Experimental models involving dramatic alterations to the thymus have been the primary means of investigation. Nonetheless, the strength of TCR stimuli in the thymus has emerged as a primary factor in the diversion of medullary thymocytes into the Treg lineage. This could be determined by the affinity, avidity, or kinetics of TCR/pMHC interactions in the thymus, and can further be modulated by the presence of costimulatory signals and cytokines such as IL-2 and TGF-β. Thus, if a threshold of activation strength is present for the development of self-reactive thymocytes into Treg vs. Tconv cells, it is unclear how sharply demarcated such a threshold might be.

Here, this work has established a robust sequencing dataset of gene expression data and VDJ data for self-antigen reactive CD4+ T cell repertoires in antigen-sufficient and antigen-deficient mouse models. This dataset was acquired in mice with fully diverse TCR repertoires, yielding insights into how Aire-dependent self-antigen presence shapes the TCR repertoires of Treg and Tconv cells. Preliminary analysis identified one potential transcription factor upregulated in a portion of Treg cells in WT mice that was absent in $C4^{KO}$ mice. The transcription factor Klf2 has been linked to Treg suppression previously[72,73] and could represent biologically meaningful differences in Treg cells that are induced upon persistent antigen exposure in the periphery. This antigen exposure could “imprint” Treg cells with maximal

suppressor function, leaving the Treg cells that remain in $C4^{KO}$ mice, which are representative of the foreign-reactive Treg cells present in many repertoires[13,83], less suppressive to the point of enabling a full immune response. Alternatively, this could be biologically less relevant than the total number of Treg cells specific for a given antigen, which is higher for self-antigens than for foreign antigens.

Additional experiments investigating the effects of Klf2 on C4-reactive Treg cells can determine if its presence in WT Treg cells in this dataset is meaningful. Specifically, Klf2+ Treg cells can be examined in *in vitro* suppression experiments to see if they exhibit greater suppressive capacity than their Klf2- counterparts, in WT or $C4^{KO}$ mice.

4.02 What are the repertoire characteristics of CD4+ T cells in WT and $C4^{KO}$ mice?

In the research described in this dissertation, one of the primary goals was to identify novel C4-reactive CD4+ T cell receptors in wild-type mice with fully diverse TCR repertoires. This is a natural continuation of years of work in the Savage and Adams labs to investigate natural Treg TCRs and their characteristics. The Savage lab originally identified the natural Treg TCR MJ23 as a recurrent prostate-infiltrating clone in TRAMP mice, and this TCR was then used to identify the natural Tcaf3-derived self-ligand C4/I-A^b. The work here sought to flip the mode of investigation again, now using C4/I-A^b tetramers to investigate the full C4-reactive TCR repertoire in WT mice. This work is a necessary prerequisite to determining the common characteristics of self-reactive Treg TCRs in comparison with self-reactive Tconv TCRs specific for the same antigen. While broad TCR sequencing has previously shown minimal overlap of TCR gene usage between Treg and Tconv repertoires, these experiments were designed to investigate the repertoires specific to a single self-antigen, something that previously could not be done due to the unknown nature of self-Treg ligands. Here, we were able to identify and

characterize a broad panel of C4-reactive Treg and Tconv TCRs to determine if there are shared TCRs, CDR3 motifs, gene usages, and public behavior across a sample of eight mice from antigen-sufficient and antigen-deficient backgrounds.

In corroboration of data on bulk repertoires, minimal to no overlap of the antigen-specific Treg and Tconv cell repertoires was observed. Additionally, now shared motifs were identified between Treg and Tconv repertoires, and distinct V gene usages were found for each repertoire. Multiple public TCRs of both Treg and Tconv phenotypes were discovered, which represent some of the first Treg TCR/cognate self-ligand pairs identified outside of fixed TCRβ transgenic mice. These TCRs are therefore likely to be more representative of true Treg development in fully diverse thymuses. The behavior of these public clones was highly informative. A single, dominant public Treg clone (WTR1) was identified in 6/9 mice sequenced, indicating that some characteristic of this TCR allowed for its recurrent development into the Treg compartment. Multiple public Tconv clones (WTC3, WTC4) were identified in multiple mice as well, indicating their molecular characteristics recurrently drove their development into the Tconv compartment despite their self-antigen reactivity. This introduces several questions regarding Treg biology. Why are some TCRs recurrently driven into the Treg lineage or the Tconv lineage? If it is some biochemical characteristic such as affinity for self-antigen, then how sharply demarcated is the dividing line between the two? Is there some portion of middle-ground TCRs for which development is largely stochastic, or is there a clear delineation? Importantly, this dataset did identify one TCR family that had members of the Treg and Tconv lineage in WT mice. This could be due to differences in the CDR3 loops that nudged the affinity or kinetics of the interacting TCR across the threshold of Treg/Tconv development, or it could indicate these

TCRs bind within the stochastic range of developmental affinities. True affinity measurements would shed light on this question.

Not only was there minimal overlap between the Treg and Tconv repertoires in WT mice, as expected, but there was minimal repertoire overlap between WT and $C4^{KO}$ mice as a whole. This was a surprising result and is at least partially explained by the high variability observed in mice with fully diverse TCR repertoires. This variability is the reason so many previous studies were conducted in fixed TCRβ transgenic mice, as it makes comparisons across mice much easier. However, even in this study, one TCR (WTR1) was found in all four WT mice from the 10x dataset, yet was found in 0/4 of the $C4^{KO}$ mice. This poses an interesting question regarding this TCR's development and persistence, as at least one of those characteristics appears to be C4-dependent. The expected result would have been finding the WTR1 TCR present in $C4^{KO}$ mice, but on Tconv cells instead of Treg cells. Thus, its absence from $C4^{KO}$ mice indicates one or more of a number of possibilities: 1) it requires the C4 antigen for positive selection, in addition to Treg differentiation. This would be surprising because typically positive selection is thought to occur in the thymic cortex and Treg development and Aire-dependent TRA expression occur in the thymic medulla. 2) WTR1 could require C4 antigen in the periphery for maintenance and could be deleted from the repertoire over time in $C4^{KO}$ mice. However, many naïve T cell specificities are known to persist in the periphery for long periods of time up to a lifetime, and why this specificity would require antigen and the other C4-reactive clones seen here would not is unclear. 3) WTR1 could be outcompeted by other C4-reactive clones in $C4^{KO}$ mice but not in WT mice. Given that WTR1 is highly expanded in all four WT mice, it seems to be highly reactive to C4 compared to the repertoire present in WT mice. If it were being outcompeted, it would indicate an abundance of new, highly reactive TCRs present in $C4^{KO}$ mice that might

therefore be negatively selected in WT mice. This would be a surprising result as well, as previous experiments have shown that Aire-mediated expression of self-antigens in the thymus is responsible for only minimal negative selection. 4) These results could simply be explained by the variability inherent in mice with fully diverse TCR repertoires. However, given that WTR1 was found so abundantly in the four WT mice, it appears unlikely that there is no biological significance to its absence in all four $C4^{KO}$ mice.

To address which of these explanations are responsible for this finding, a number of experiments are required. The Savage lab is currently developing mouse models that will express the C4 peptide in the thymus alone and in the prostate alone. The C4-thymus model will represent a new type of antigen exposure (similar to using a female mouse, but without hormonal and developmental caveats) that can address these questions of the role of peripheral antigen exposure on the function of Treg cells, and WTR1 specifically. First, TCR sequencing of mice with C4 expressed only in the thymus will indicate if that is sufficient for WTR1's persistence in the periphery without continued C4 exposure. Second, TCR sequencing of mice with C4 expressed only in the periphery, and not in the thymus, would answer the question of whether C4 is required for the positive selection and thymic egress of WTR1 clones. Third, affinity measurements of WTR1 and multiple $C4^{KO}$ clones with C4/I-A^b will answer the question of whether the $C4^{KO}$ clones possess a significantly greater affinity for C4 antigen, and thus could be outcompeting WTR1 for antigen post-immunization.

Furthermore, a number of public Treg clones were identified in the $C4^{KO}$ mice, providing additional questions. How were these clones recurrently driven into the Treg lineage across multiple mice? Why aren't these clones found in WT mice, in either the Treg or Tconv lineages? Conventional wisdom might suppose that these clones were negatively selected by C4 peptide in

WT mice, but might be cross-reactive to a minimal degree – enough to provide sufficient signaling for Treg development in $C4^{KO}$ mice but not excessive signaling that would induce deletion in the thymus. Perhaps another explanation is that they are cross-reactive with one additional I-A^b epitope, which would remove the potential for widespread cross-reactivity that might induce autoimmune reactions. The additional TCR sequencing experiments in C4-thymus and C4-periphery mouse models would answer these questions as well. On a more global level, flow-based experiments will be sufficient to determine if C4-immunization induces an immune response with Treg:Tconv ratios that correspond more like WT mice or $C4^{KO}$ mice, which would provide insights into whether continued antigen exposure in the periphery is required for maximal Treg suppressor function. Finally, while $F1^{KO}$ mice have not been engineered, TCR sequencing of F1-reactive T cells in WT mice would still represent a valuable dataset to compare with the C4-reactive repertoire described here. Namely, this could provide additional public Treg and Tconv clones specific for a single self-antigen to compare and contrast to attempt to further elucidate the molecular basis for Treg development.

4.03 What are the kinetic and structural differences between self-reactive TCRs of Treg and Tconv origin?

Another primary goal of the research in this dissertation was to characterize the biochemical parameters with which self-reactive Treg and Tconv TCRs bind their cognate ligands. The experiments described in the previous section provided a rich source of TCRs to select for these biochemical experiments, and in Chapter 3 of this dissertation, the 3D affinity and kinetics of self-reactive Treg and Tconv TCRs specific for a single self-antigen were quantified, providing the first "benchmark" measurements of affinities for these two populations. SPR measurements revealed a dramatic difference in affinity, with three Treg TCRs binding

C4/I-A^b with higher affinity and significantly longer off-rates than three Tconv TCRs. It is important to place these measurements in context with previous data published by Stadinski et al. in 2019, which were obtained with TCRs of a single fixed TCRβ chain. In that study, the authors argued that the difference in Treg vs. Tconv TCRs binding their self-antigen was differentiated by off-rate, or "dwell-time", rather than affinity. However, affinity also played a significant role, with an r^2 value of 0.3 (compared to off-rate's r^2 value of 0.6) for Treg/Tconv frequency. Biochemically, this makes intuitive sense, as the off-rate comprises half of the equation contributing to the affinity, or equilibrium value K_D, with the other half determined by on-rate. Thus, if affinity was not significant, but off-rate was, then there would have to be a compensatory change in on-rates as well to maintain similar affinities with varying off-rates. Thus, if off-rates or affinities for self-antigen are involved in Treg development, they are likely to correlate with each other, as a slower off-rate will be a primary contributor to a high TCR:pMHC affinity. This correlates with numerous other studies that have shown both that affinity (K_D) and dissociation rate, or off-rate (k_d) can contribute approximately equally to TCR confinement time which is the most accurate predictor of T cell signaling and activation[82,84,85]. As long as the overall confinement time of a given TCR is longer, the T cell is likely to experience stronger signaling. This is exactly what is observed in the data described in this dissertation. The affinities of the Treg TCRs are higher than the Tconv TCRs, but narrowly miss significance due to the small sample size, while the off-rates are significantly longer in Treg TCRs. Interestingly, when compared with the data in Stadinski et al., the off-rates observed in this dataset were slightly outside the range of the proposed Treg and Tconv ranges presented there. The proposed Treg range of dwell times in Stadinski et al. was 0.8-1.4s, with times of 4-7 seconds fitting those of TCRs driving negative selection. The dwell times of Treg TCRs in our

dataset were found to be 3-4.5 seconds, which is significantly above the proposed Treg range in Stadinski et al. and well into the proposed range for negative selection. The Tconv off rates in our dataset ranged from .5-1.5 seconds, slightly above Stadinski et al.'s reported Tconv range of 0.2-0.6s. These two TCR repertoires are for different self-antigens (albeit both in the context of I-A^b) which could explain the different development windows of their TCR repertoires. It remains possible that the relative affinity of TCRs specific for a single self-antigen is more important than an absolute affinity. Obtaining datasets for additional self-antigens will help answer this question. While it might be expected that a given affinity for any self-antigen would drive the development of Tregs in the thymus, the functional avidity of the interactions might play an important role as well, as it is not known which self-antigens are expressed to what degrees in the thymus, and furthermore they may be processed and presented by APCs with different efficiencies. Thus, the avidity of these interactions in the thymus may be similar, and only the relative affinity and kinetics windows of TCRs specific for the same antigen can be directly compared.

Characterizing the affinity and kinetics of additional self-reactive TCR repertoires, such as the F1-reactive repertoire, could provide further clarification of the range of permissive Treg developmental "windows" and Tconv "windows". Given the results of these first two datasets, it is clear there may be significant overlap. This leaves room for TCR-independent factors in the ultimate development of thymocytes as well. The affinity of TCRs that end up in the Treg vs. Tconv lineages may overlap some due to a number of factors. First, the avidity of their interactions in the thymus may not overlap, as discussed above. Additionally, there may be other thymus-intrinsic factors that affect cell outcome, such as availability of IL-2. It has already been shown that there is a limit to the thymic Treg niche, and IL-2 competition is likely to be a

primary factor in limiting the number of Tregs in the thymus. While the T cells with the highest affinity TCRs for self-antigen will still be most likely to receive the greatest IL-2 signaling, it does not preclude some lower affinity TCRs receiving sufficient IL-2 first, especially if they happen to stochastically develop earlier than a given high-affinity TCR. Thus, this could be one factor by which temporal regulation of Treg development could exist, in a manner distinct from that described in Stadinski et al., as this mechanism would not be expected to be consistent across mice.

While the experiments in this section yielded exciting new results, the TCR repertoires identified in Chapter 2 present myriad other questions to be investigated biochemically. First, what is the affinity with which the widely public WTR1 TCR binds C4/I-A^b? Is it similar to WTR2, WTR3, and WTR4, or is it distinct? This TCR will provide an important benchmark, as its recurrent development into the Treg lineage means it is the best representative "Treg" TCR in the dataset. Additionally, while SPR is the gold standard assay for determining biochemical affinity, *in vivo* activation assays of T cells bearing these TCRs are required to confirm activity of these TCRs in physiological models, as some data from this work and others has confirmed that some TCRs that may stain with tetramer do not activate cells *in vivo*.

Second, what are the affinities with which the highly-expanded $C4^{KO}$ TCRs bind C4/I-A^b? These TCRs are largely absent from WT mice, is it because they bind with dramatically higher affinity and are negatively selected? Or do they bind with conventional affinities, and are likely to be absent by chance? Additionally, there are four TCRs that are found in the Tconv compartment in WT and in $C4^{KO}$ mice. Two of these TCRs are dramatically expanded in the $C4^{KO}$ mice, while two are not. Affinities of these interactions would be informative as well to determine if competition for antigen is a factor in the differential expansion of these clones upon

C4-immunization in $C4^{KO}$ mice. Obtaining the affinity of the $C4^{KO}$ public Treg clones with C4/I-A^b would be of great interest, as these clones are somehow recurrently diverted into the Treg lineage across mice in the absence of the C4 peptide in the thymus and periphery. Affinity measurements of these clones with other peptide/I-A^b complexes could possibly provide some insight into whether they are cross-reactive as well. However, if they are cross-reactive with one specific epitope it would be unlikely that *in vitro* affinity measurements would discover this. Finally, there were multiple "Trogue" TCRs that were found on Treg cells in WT mice and Tconv cells in $C4^{KO}$ mice. Obtaining the affinities of these TCRs for C4/I-A^b would be of great interest to compare them with those of other WT Treg TCRs that did not occur in $C4^{KO}$ mice. Again, this, combined with other experiments described above, could inform whether stochastic repertoire variability or active selection is responsible for the development of Trogue clones.

Another area of great interest is acquiring the affinities of alanine scanning mutants of multiple TCRs in complex with C4/I-A^b. This information can map the important contact points each TCR uses to interface with C4/I-A^b and can provide insights into the general docking modes used even in the absence of complex crystal structures. Alanine scanning mutants of I-A^b and the C4 peptide can also inform as to which residues are being contacted by one or more TCRs.

Another area of importance is conducting *in vivo* activation studies to ensure that each TCR analyzed in this dataset is capable of activating T cells in physiological contexts. As shown in Sibener et. al 2018, TCR binding of pMHC is not always directly coupled with T cell activation[65]. Other work conducted by the Savage lab has previously identified TCRs that recurrently stain with tetramer but do not activate in retrogenic T cells. Indeed, as WTC4 shows, some TCRs in this dataset may not even bind C4/I-A^b.

Finally, a primary goal of future research on this project is solving a proper crystal structure of a Treg TCR in complex with C4/I-A^b or F1/I-A^b. An additional crystal structure of a Tconv TCR in complex with the same antigen would further increase our ability to directly compare and contrast Treg binding and Tconv binding to explore whether they bind with distinct characteristics, such as docking mode, bond characteristics, and peptide/MHC bias of each CDR loop in comparison with previous structures.

4.04 Summary

Together, the data described in this dissertation has provided numerous insights into the nature of self-reactive CD4+ T cell repertoires and Treg biology. We conducted next-gen sequencing of C4/I-A^b tetramer-positive T cells in mice with fully diverse TCR repertoires. This dataset is among the first sequencing experiments of a fully diverse (i.e. not in a fixed TCRβ transgenic background) self-reactive CD4+ T cell repertoire, and it provided significant insights into Treg biology worthy of future study. Furthermore, the sequencing was conducted in antigen-sufficient mice (WT) and antigen-deficient mice ($C4^{KO}$) to investigate the effect of C4 presence on the TCR repertoire. We found dramatic perturbations in the TCR repertoire that surpassed expectations. Not only was there a significant increase in the Tconv compartment in $C4^{KO}$ mice and a corresponding decrease in the Treg compartment, but the TCR sequences themselves were significantly different. The $C4^{KO}$ mice were missing multiple public clones present in WT mice and possessed multiple public clones of their own that were absent in WT mice. These clones were of both Treg and Tconv phenotypes, and provide new insights into the role of Aire-dependent self-antigens in the thymus.

We further showed the presence of almost zero overlap between the Treg and Tconv repertoires in WT mice. Further analysis of TCR distance between clonotypes revealed distinct

V-gene usage of WT Treg and Tconv repertoires, and these repertoires exhibited distinct clustering by BLOSUM62 alignment, which accounts for both sequence distance and amino acid property in assigning distances for each TCR. AIMS analysis identified increased positive charges in the central CDR3 loops of WT Treg TCRs as a key distinguishing factor from Tconv TCRs. Further biochemical analysis also revealed the potential for positively charged Treg CDR1α residues to interact with pMHC in a manner distinct from those of Tconv TCRs. This, together with the consistent developmental phenotype of the public clones (which were all Treg or all Tconv), indicates a strong ability for the thymus to discriminate Treg and Tconv clones and drive their development into the corresponding lineage.

To investigate the role of TCR:pMHC affinity and kinetics in that discrimination, we conducted SPR analysis of three Treg and three Tconv TCRs in association with C4/I-A^b and found higher affinity values and longer off rates for Treg TCRs than Tconv TCRs consistent with increased confinement times and expected TCR signaling strength. These measurements represent important benchmark values for future Treg and Tconv interaction studies with self-peptides, which will be able to more completely determine the types of interactions that inform Treg vs. Tconv development of self-reactive thymocytes.

Thus, in conclusion, the data presented here is consistent with a model (Figure 4.1) in which C4-reactive thymocytes develop into either Treg or Tconv cells in the thymus (Figure 4.1A). Those that develop into Treg cells bear TCRs with distinct V-genes from those that drive Tconv development (Figure 4.1B), and the Treg TCRs exhibit increased positive charge in their central CDR3 loops (Figure 4.1C), as well as increased positive charge in the CDR1α loop consistent with the enriched V-gene populations. These increased positive charges correlate with increased affinity and off-rates for Treg TCRs, which contribute to a longer confinement time

and expected TCR signaling strength compared to Tconv development (Figure 4.1D). We further observed zero overlap between Treg and Tconv TCRs in this model, indicating robust and consistent developmental selection of Treg cells in the thymus (Figure 4.1E). This process was further shown to be antigen-dependent by comparison to a $C4^{KO}$ model repertoire. Thus, this supports a model of TCR-intrinsic factors contributing a significant role in Treg cell development.

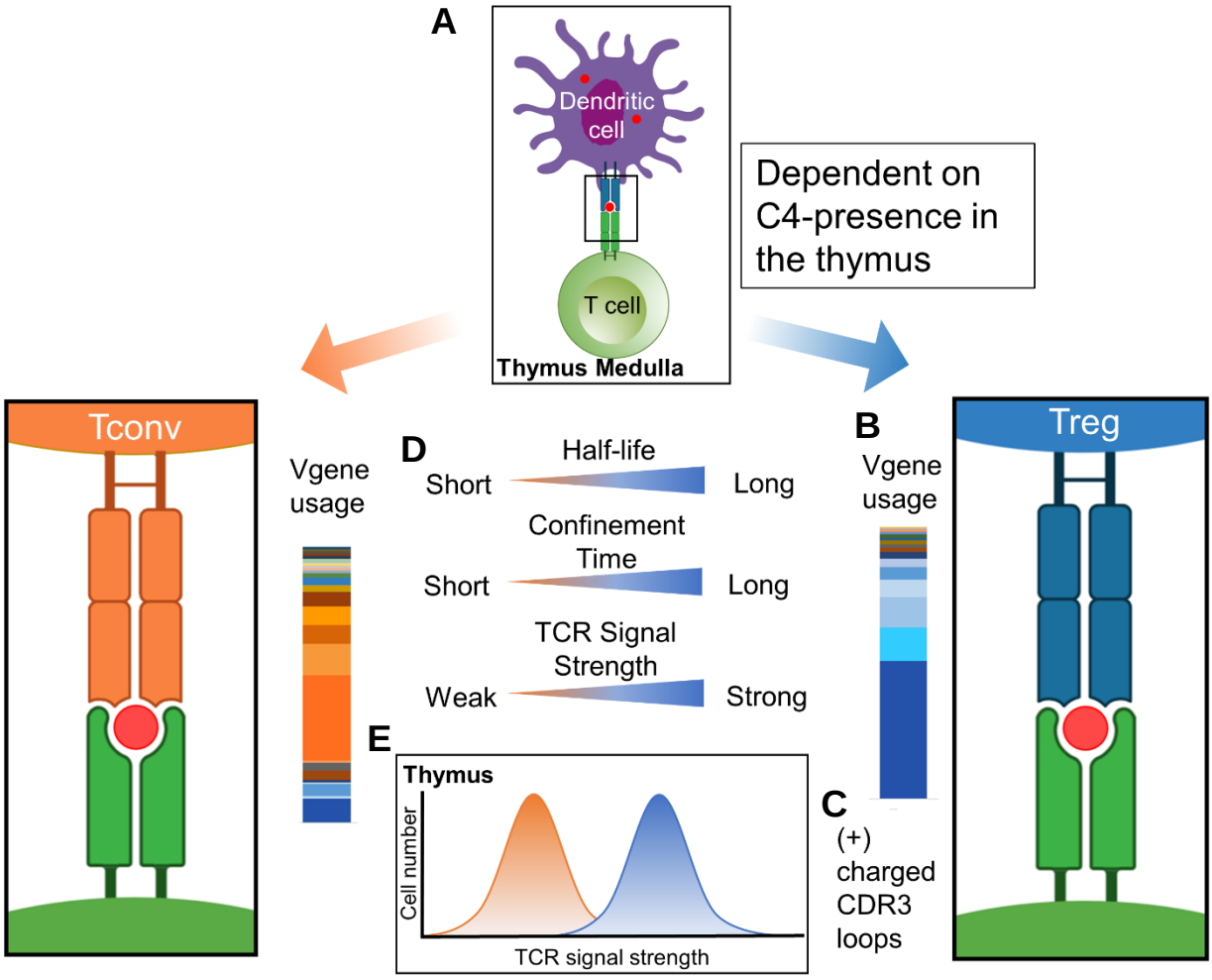

Figure 4.1 Working model of C4-reactive Treg development

Depiction of a TCR-intrinsic model of Treg cell development supported by this dissertation, in which developing thymocytes with more stable interactions with self-antigen in the thymic medulla develop into Treg TCRs. **(A)** C4-reactive thymocytes recognize C4 (red circles) presented by I-A^b in the thymus and develop into either Tconv (left, orange) or Treg (right, blue) cells. **(B)** Treg cells bear TCRs with distinct V-gene usage compared to Tconv TCRs. **(C)** Treg TCRs bear increased positive charges in their central CDR3 loops. **(D)** Treg TCRs interact with C4/I-A^b with a longer dissociation rate, confinement time, and therefore a greater expected TCR signal strength. **(E)** Repertoire sequencing supports minimal TCR overlap and suggests robust separation between these two repertoires.

Chapter 5: Materials and Methods

5.01 Single cell PCR sequencing

Foxp3^{GFP} Mice were immunized with C4 peptide and complete Freund's adjuvant two weeks prior to processing. GFP+ cells were dual stained with C4/I-A^{b} tetramers in APC and PE and single cell sorted into 96 well plates containing 10µM Tris pH 8.0 with 0.025% RNasin RNase inhibitor. Multiple random wells were left unsorted as negative controls and one well had fifty cells sorted as a positive control for subsequent PCRs. Plates were placed on ice and frozen at -80^{o}C. Plates were thawed and a master mix was added containing MAXIMA 5x Buffer Mix, 5% IgePal detergent, MAXIMA enzyme mix according to Maxima's First Strand cDNA Synthesis Kit #K1642 for cDNA production. 2uL of the resulting cDNA mix was added to two new 48 well plates, one for the TCRα reactions and one for the TCRβ reactions. To these were added 2x Master Mix DreamTaq Green and a primer mix containing TRAV or TRBV primers. Each primer mix contained one reverse primer specific from the constant region and one forward primer specific for each TRAV/TRBV gene at a final concentration of 5µM per primer. This was followed by a second round of PCR with nested primers internal to the first set before being submitted for Sanger sequencing at the UChicago Sequencing Core. Resulting chromatograms were entered into IMGT's VQuest program for TCR gene and CDR3 identification. All resulting TCRs were then analyzed with software developed by Paul Thomas's group[86].

5.02 10x Genomics sequencing

Four WT and four C4^{KO} C57BL/6 were immunized subcutaneously 2 weeks prior to cell isolation on both flanks with a total of 100uL solution containing 100ug of C4 peptide in CFA. Mice were divided into two samples for independent processing with two WT and two

$C4^{KO}$ mice in each sample. Secondary lymphoid organs were isolated and CD4+ cells were MACS enriched. Samples were incubated with destatinib for 30 minutes and incubated directly with C4 tetramer in APC and PE for 1 hour at 100nM per tetramer. Samples were enriched for tetramer using StemCell magnetic enrichment and surface stained with hashtag antibodies and a surface panel before sorting tetramer positive cells into 1mL 10x buffer before sequencing.

Sequencing for both samples was conducted with the UChicago Genomics Core's 10x Genomics Chromium processor. Libraries were constructed for gene expression data, VDJ data, and hashtag CITE-seq data for both samples using 10x Single Cell 5' R2-only chemistry.

5.03 Genomics data analysis

Raw gene expression (GE) and VDJ FASTA outputs were processed with 10x Genomics' CellRanger[87] count v3.1.0 software using 10x Genomics' preconstructed mouse references for gene expression and VDJ datasets (v3.1.0 mm10 and GRCm38) on UChicago's Midway computer cluster. Each sample was processed independently and merged with Seurat (v3.2.3)[88] in R studio (v 1.3.1093). Quality control was completed by removing cells with low feature RNA (<200 genes), high feature RNA (>5000 genes), and high mitochondrial UMI counts (>2%). Paired TCR clones were defined by V, (D), J gene usage and CDR3 nucleotide sequences for alpha and beta chains. TCR overlap was defined by V, (D), J gene usage and CDR3 amino acid sequences.

All libraries were integrated in Seurat (v. 3.2.3)[88]. The combined Seurat object UMI counts were normalized with the SCTransform function within Seurat by binomial regression of mitochondrial UMI counts. The scaled data was used for PCA analysis and the first 20 principal components were used for visualization using UMAP analysis. Clusters were identified with the FindNeighbors and FindClusters functions within Seurat, and two clusters, containing

neutrophils and VDJ-negative cells, were removed from further analysis. Gene markers were detected by comparing each cluster with all cells using the FindAllMarkers function using only positive markers with expression in a minimum of 25% of cells and a logFC threshold of 0.25.

VDJ libraries were imported into Seurat with the scRepertoire package[89] for analysis along with the circlize package[90]. TCR sequences were also loaded into the AIMS GUI for biochemical analysis[91]. TCRs with dual TCRα or, in rare cases, TCRβ chains had one TCR gene removed by scRepertoire for analysis and were analyzed as separate TCRs in AIMS. CDR3 motif analysis was performed by the glam2 function of Multiple EM for Motif Elicitation (MEME)[92]. Further repertoire analysis was conducted in Python 2.7.11 using the TCRdist3 package[75,86]

5.04 TCR Cloning and Expression

For BLItz, SPR, and crystallographic studies, all TCR constructs were designed by fusing the native variable α and β domains to the human α and β constant domains containing an engineered disulfide bond bridging the two constant domains[93]. The TCRs used for BLItz and SPR were cloned into versions of this pACgp67a construct containing a BirA biotinylation sequence upstream of the 3C protease site. Both chains were co-expressed in Hi5 cells with baculovirus transduction. The heterodimeric TCRs were purified with Nickel NTA resin (Quiagen) and further purified with size-exclusion chromatography. BirA constructs were co-expressed with BirA ligase and supplemental biotin for *in vivo* biotinylation of the BirA sequence.

5.05 I-A^b expression

Tcaf3-C4/I-A^b and Tcaf3-F1/I-A^b were produced using methods similar to those described previously[94]. Constructs were inserted into two separate pRMHa3 plasmids – one with a BirA site and no 3C protease sites for tetramer production and SPR/BLItz immobilization and one identical design with 3C sites introduced upstream of the BirA site, leucine zippers, and his tag for crystallography. The $Tcaf3_{646\text{-}658}$(648Y) C4 peptide (THYKAPWGELATD) and the $Tcaf3_{88\text{-}107}$ F1 peptide (CPGAPIAVHSSLASLVNILG) were engineered to the N-terminus of the I-A^b β chain using methods similar to those described previously. I-A^b was expressed in *Drosophila*S2 cells using separate plasmids to encode the alpha and beta chains and were co-transfected with plasmids encoding the BirA ligase and blasticidin resistance. Protein expression was induced with the addition of 0.8mM of $CuSO_4$ in the presence of 2ug/mL biotin (Sigma-Aldrich). Biotinylated I-A^b protein was purified from culture supernatant by nickel affinity chromatography with Nickel IDA resin and by avidin affinity chromatography with Pierce Monomeric Avidn UltraLink Resin (Thermo Fisher).

The extracellular domain of the I-A^b alpha chain was fused at its N terminus to a secretion signal sequence and a C terminal acidic leucine zipper and BirA sequence. The extracellular domain of the I-A^b beta chain was fused at its N terminus to a secretion signal sequence followed by the $Tcaf3_{646\text{-}658}$(S1Y) peptide and a linker sequence and a C terminal basic leucine zipper and a 6xHistidine tag.

5.06 Biolayer interferometry

All BLI measurements were performed using a ForteBio BLItz instrument at 22^{o}C. For iRepertoire-derived TCR and 10x-derived TCR measurements, biotinylated C4/I-A^b and F1/I-A^b

were immobilized on streptavidin biosensors until saturation. Unbound biotin binding sites were blocked with 2mg/mL biotin and each TCR was added in serial dilutions in HBS. An HBS-only trace was subtracted from TCR measurements. Equilibrium dissociation constants were determined by nonlinear regression with GraphPad Prism software. The orientation of the RET TCR assay was reversed, with biotinylated TCR immobilized on the biosensor and C4/I-A^b used as the analyte to confirm similar measurements.

5.07 Surface Plasmon Resonance

All SPR measurements were conducted with a Biacore 8000 instrument in UChicago's Biophysics core at 25°C. C4/I-A^b was immobilized on a renewable streptavidin sensor chip. One flow cell was left blank for reference subtraction. All flow cells were blocked with 1uM biotin. Serial dilutions of each TCR were flowed as analyte. Equilibrium dissociation constants and kinetics measurements were determined with Biacore's software program.

5.08 I-A^b tetramer production and staining

Purified and biotinylated C4/I-A^b and F1/I-A^b were diluted in HBS to remove free biotin. Monomers were mixed with streptavidin-APC or streptavidin-PE at a molar excess of I-A^b to biotin binding sites to ensure saturation. Saturation of the streptavidin conjugate was verified by non-reducing SDS-PAGE without boiling samples.

Tetramer staining was adapted from Tungatt et al. 2015[95] and performed similarly to methods described previously[71]. Cells were treated with desatinib (AdooQ Bioscience) at a final concentration of 50nM for 30 min at 37°C. PE- and APC-labeled tetramers were added directly to desatinib-treated cells at a final concentration of 100nM for 1hr at room temperature. Cells were washed and stained for flow cytometric analysis.

5.09 Crystallography

All crystal trials were conducted with Pro-Complex, PEGS II, and PEGS I screens. Screens were set up with MJ23:C4/I-A^b, C4/I-A^b alone, WTR2:C4/I-A^b, and WTR3:C4/I-A^b in wells with 75uL mother liquor and 1:1 ratios of mother liquor:protein droplets.

References

1. Arstila, T. P. *et al.* A direct estimate of the human alphabeta T cell receptor diversity. *Science* **286**, 958–961 (1999).

2. Fontenot, J. D., Gavin, M. A. & Rudensky, A. Y. Foxp3 programs the development and function of CD4+CD25+ regulatory T cells. *Nat. Immunol.* **4**, 330–336 (2003).

3. Hori, S. Control of Regulatory T Cell Development by the Transcription Factor Foxp3. *Science* **299**, 1057–1061 (2003).

4. Bennett, C. L. *et al.* The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3. *Nat. Genet.* **27**, 20–21 (2001).

5. Brunkow, M. E. *et al.* Disruption of a new forkhead/winged-helix protein, scurfin, results in the fatal lymphoproliferative disorder of the scurfy mouse. *Nat. Genet.* **27**, 68–73 (2001).

6. Kim, J. M., Rasmussen, J. P. & Rudensky, A. Y. Regulatory T cells prevent catastrophic autoimmunity throughout the lifespan of mice. *Nat. Immunol.* **8**, 191–197 (2007).

7. Sakaguchi, S., Sakaguchi, N., Asano, M., Itoh, M. & Toda, M. Immunologic self-tolerance maintained by activated T cells expressing IL-2 receptor alpha-chains (CD25). Breakdown of a single mechanism of self-tolerance causes various autoimmune diseases. *J. Immunol. Baltim. Md 1950* **155**, 1151–1164 (1995).

8. Hsieh, C.-S. *et al.* Recognition of the peripheral self by naturally arising CD25+ CD4+ T cell receptors. *Immunity* **21**, 267–277 (2004).

9. Levine, A. G., Arvey, A., Jin, W. & Rudensky, A. Y. Continuous requirement for the TCR in regulatory T cell function. *Nat. Immunol.* **15**, 1070–1078 (2014).

10. Vahl, J. C. *et al.* Continuous T cell receptor signals maintain a functional regulatory T cell pool. *Immunity* **41**, 722–736 (2014).

11. Bilate, A. M. & Lafaille, J. J. Induced CD4+Foxp3+ regulatory T cells in immune tolerance. *Annu. Rev. Immunol.* **30**, 733–758 (2012).

12. Mucida, D. *et al.* Oral tolerance in the absence of naturally occurring Tregs. *J. Clin. Invest.* **115**, 1923–1933 (2005).

13. Bacher, P. *et al.* Regulatory T Cell Specificity Directs Tolerance versus Allergy against Aeroantigens in Humans. *Cell* **167**, 1067-1078.e16 (2016).

14. Szurek, E. *et al.* Differences in Expression Level of Helios and Neuropilin-1 Do Not Distinguish Thymus-Derived from Extrathymically-Induced CD4+Foxp3+ Regulatory T Cells. *PloS One* **10**, e0141161 (2015).

15. Akimova, T., Beier, U. H., Wang, L., Levine, M. H. & Hancock, W. W. Helios expression is a marker of T cell activation and proliferation. *PloS One* **6**, e24226 (2011).

16. Schreiber, L. *et al.* The Treg-Specific Demethylated Region Stabilizes Foxp3 Expression Independently of NF-κB Signaling. *PLOS ONE* **9**, e88318 (2014).

17. Koch, M. A. *et al.* The transcription factor T-bet controls regulatory T cell homeostasis and function during type 1 inflammation. *Nat. Immunol.* **10**, 595–602 (2009).

18. Campbell, D. J. & Koch, M. A. Phenotypical and functional specialization of FOXP3+ regulatory T cells. *Nat. Rev. Immunol.* **11**, 119–130 (2011).

19. Maloy, K. J. & Powrie, F. Regulatory T cells in the control of immune pathology. *Nat. Immunol.* **2**, 816–822 (2001).

20. Levings, M. K. *et al.* Human CD25+CD4+ T Suppressor Cell Clones Produce Transforming Growth Factor β, but not Interleukin 10, and Are Distinct from Type 1 T Regulatory Cells. *J. Exp. Med.* **196**, 1335–1346 (2002).

21. Suvas, S., Kumaraguru, U., Pack, C. D., Lee, S. & Rouse, B. T. CD4+CD25+ T Cells Regulate Virus-specific Primary and Memory CD8+ T Cell Responses. *J. Exp. Med.* **198**, 889–901 (2003).

22. Shevach, E. M. Mechanisms of Foxp3+ T Regulatory Cell-Mediated Suppression. *Immunity* **30**, 636–645 (2009).

23. Schmidt, A., Oberle, N. & Krammer, P. H. Molecular Mechanisms of Treg-Mediated T Cell Suppression. *Front. Immunol.* **3**, (2012).

24. Vignali, D. A. A., Collison, L. W. & Workman, C. J. How regulatory T cells work. *Nat. Rev. Immunol.* **8**, 523–532 (2008).

25. Collison, L. W. *et al.* The inhibitory cytokine IL-35 contributes to regulatory T-cell function. *Nature* **450**, 566–569 (2007).

26. von Boehmer, H. Mechanisms of suppression by suppressor T cells. *Nat. Immunol.* **6**, 338–344 (2005).

27. Gondek, D., Lu, L.-F., Quezada, S., Sakaguchi, S. & Noelle, R. Cutting Edge: Contact-Mediated Suppression by CD4+CD25+ Regulatory Cells Involves a Granzyme B-Dependent, Perforin-Independent Mechanism. *J. Immunol. Baltim. Md 1950* **174**, 1783–6 (2005).

28. Grossman, W. J. *et al.* Human T regulatory cells can use the perforin pathway to cause autologous target cell death. *Immunity* **21**, 589–601 (2004).

29. Pandiyan, P., Zheng, L., Ishihara, S., Reed, J. & Lenardo, M. J. CD4 + CD25 + Foxp3 + regulatory T cells induce cytokine deprivation–mediated apoptosis of effector CD4 + T cells. *Nat. Immunol.* **8**, 1353–1362 (2007).

30. Szymczak-Workman, A. L., Delgoffe, G. M., Green, D. R. & Vignali, D. A. A. Cutting Edge: Regulatory T Cells Do Not Mediate Suppression via Programmed Cell Death Pathways. *J. Immunol.* **187**, 4416–4420 (2011).

31. Davidson, T. S., DiPaolo, R. J., Andersson, J. & Shevach, E. M. Cutting Edge: IL-2 is essential for TGF-beta-mediated induction of Foxp3+ T regulatory cells. *J. Immunol. Baltim. Md 1950* **178**, 4022–4026 (2007).

32. Bachmann, M. F., Köhler, G., Ecabert, B., Mak, T. W. & Kopf, M. Cutting edge: lymphoproliferative disease in the absence of CTLA-4 is not T cell autonomous. *J. Immunol. Baltim. Md 1950* **163**, 1128–1131 (1999).

33. Takahashi, T. *et al.* Immunologic self-tolerance maintained by CD25(+)CD4(+) regulatory T cells constitutively expressing cytotoxic T lymphocyte-associated antigen 4. *J. Exp. Med.* **192**, 303–310 (2000).

34. van der Merwe, P. A., Bodian, D. L., Daenke, S., Linsley, P. & Davis, S. J. CD80 (B7-1) Binds Both CD28 and CTLA-4 with a Low Affinity and Very Fast Kinetics. *J. Exp. Med.* **185**, 393–404 (1997).

35. Qureshi, O. S. *et al.* Trans-endocytosis of CD80 and CD86: a molecular basis for the cell-extrinsic function of CTLA-4. *Science* **332**, 600–603 (2011).

36. Veldhoen, M., Moncrieffe, H., Hocking, R. J., Atkins, C. J. & Stockinger, B. Modulation of Dendritic Cell Function by Naive and Regulatory CD4+ T Cells. *J. Immunol.* **176**, 6202–6210 (2006).

37. Levine, A. G. *et al.* Suppression of lethal autoimmunity by regulatory T cells with a single TCR specificity. *J. Exp. Med.* **214**, 609–622 (2017).

38. Tadokoro, C. E. *et al.* Regulatory T cells inhibit stable contacts between CD4+ T cells and dendritic cells in vivo. *J. Exp. Med.* **203**, 505–511 (2006).

39. Klein, L., Kyewski, B., Allen, P. M. & Hogquist, K. A. Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). *Nat. Rev. Immunol.* **14**, 377–391 (2014).

40. McDonald, B. D., Bunker, J. J., Erickson, S. A., Oh-Hora, M. & Bendelac, A. Crossreactive αβ T Cell Receptors Are the Predominant Targets of Thymocyte Negative Selection. *Immunity* **43**, 859–869 (2015).

41. Pobezinsky, L. A. *et al.* Clonal deletion and the fate of autoreactive thymocytes that survive negative selection. *Nat. Immunol.* **13**, 569–578 (2012).

42. Bouillet, P. *et al.* BH3-only Bcl-2 family member Bim is required for apoptosis of autoreactive thymocytes. *Nature* **415**, 922–926 (2002).

43. McDonald, B. D., Bunker, J. J., Ishizuka, I. E., Jabri, B. & Bendelac, A. Elevated T cell receptor signaling identifies a thymic precursor to the TCRαβ(+)CD4(-)CD8β(-) intraepithelial lymphocyte lineage. *Immunity* **41**, 219–229 (2014).

44. Li, X. *et al.* Pre-T cell receptors topologically sample self-ligands during thymocyte β-selection. *Science* **371**, 181–185 (2021).

45. Padovan, E. *et al.* Expression of two T cell receptor alpha chains: dual receptor T cells. *Science* **262**, 422–424 (1993).

46. Yu, W. *et al.* Clonal Deletion Prunes but Does Not Eliminate Self-Specific αβ CD8(+) T Lymphocytes. *Immunity* **42**, 929–941 (2015).

47. Bouneaud, C., Kourilsky, P. & Bousso, P. Impact of negative selection on the T cell repertoire reactive to a self-peptide: a large fraction of T cell clones escapes clonal deletion. *Immunity* **13**, 829–840 (2000).

48. Moon, J. J. *et al.* Quantitative impact of thymic selection on Foxp3+ and Foxp3- subsets of self-peptide/MHC class II-specific CD4+ T cells. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 14602–14607 (2011).

49. Hassler, T. *et al.* Inventories of naive and tolerant mouse CD4 T cell repertoires reveal a hierarchy of deleted and diverted T cell receptors. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 18537–18543 (2019).

50. Stadinski, B. D. *et al.* A temporal thymic selection switch and ligand binding kinetics constrain neonatal Foxp3+ Treg cell development. *Nat. Immunol.* **20**, 1046–1058 (2019).

51. Legoux, F. P. *et al.* CD4+ T Cell Tolerance to Tissue-Restricted Self Antigens Is Mediated by Antigen-Specific Regulatory T Cells Rather Than Deletion. *Immunity* **43**, 896–908 (2015).

52. Malchow, S. *et al.* Aire Enforces Immune Tolerance by Directing Autoreactive T Cells into the Regulatory T Cell Lineage. *Immunity* **44**, 1102–1113 (2016).

53. Anderson, M. S. Projection of an Immunological Self Shadow Within the Thymus by the Aire Protein. *Science* **298**, 1395–1401 (2002).

54. Perry, J. S. A. *et al.* Distinct contributions of Aire and antigen-presenting-cell subsets to the generation of self-tolerance in the thymus. *Immunity* **41**, 414–426 (2014).

55. Mathis, D. & Benoist, C. Aire. *Annu. Rev. Immunol.* **27**, 287–312 (2009).

56. Malchow, S. *et al.* Aire-dependent thymic development of tumor-associated regulatory T cells. *Science* **339**, 1219–1224 (2013).

57. Meyer, S. *et al.* AIRE-Deficient Patients Harbor Unique High-Affinity Disease-Ameliorating Autoantibodies. *Cell* **166**, 582–595 (2016).

58. Hsieh, C.-S., Zheng, Y., Liang, Y., Fontenot, J. D. & Rudensky, A. Y. An intersection between the self-reactive regulatory and nonregulatory T cell receptor repertoires. *Nat. Immunol.* **7**, 401–410 (2006).

59. Kalekar, L. A. *et al.* CD4(+) T cell anergy prevents autoimmunity and generates regulatory T cell precursors. *Nat. Immunol.* **17**, 304–314 (2016).

60. Hsieh, C.-S., Lee, H.-M. & Lio, C.-W. J. Selection of regulatory T cells in the thymus. *Nat. Rev. Immunol.* (2012) doi:10.1038/nri3155.

61. Aschenbrenner, K. *et al.* Selection of Foxp3+ regulatory T cells specific for self antigen expressed and presented by Aire+ medullary thymic epithelial cells. *Nat. Immunol.* **8**, 351–358 (2007).

62. Lee, H.-M., Bautista, J. L., Scott-Browne, J., Mohan, J. F. & Hsieh, C.-S. A broad range of self-reactivity drives thymic regulatory T cell selection to limit responses to self. *Immunity* **37**, 475–486 (2012).

63. Jordan, M. S. *et al.* Thymic selection of CD4+CD25+ regulatory T cells induced by an agonist self-peptide. *Nat. Immunol.* **2**, 301–306 (2001).

64. Apostolou, I., Sarukhan, A., Klein, L. & von Boehmer, H. Origin of regulatory T cells with known specificity for antigen. *Nat. Immunol.* **3**, 756–763 (2002).

65. Sibener, L. V. *et al.* Isolation of a Structural Mechanism for Uncoupling T Cell Receptor Signaling from Peptide-MHC Binding. *Cell* **174**, 672-687.e27 (2018).

66. Dai, S. *et al.* Crossreactive T Cells spotlight the germline rules for alphabeta T cell-receptor interactions with MHC molecules. *Immunity* **28**, 324–334 (2008).

67. Beringer, D. X. *et al.* T cell receptor reversed polarity recognition of a self-antigen major histocompatibility complex. *Nat. Immunol.* **16**, 1153–1161 (2015).

68. Stadinski, B. D. *et al.* Hydrophobic CDR3 residues promote the development of self-reactive T cells. *Nat. Immunol.* **17**, 946–955 (2016).

69. Greenberg, N. M. *et al.* Prostate cancer in a transgenic mouse. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 3439–3443 (1995).

70. Sharma, P. & Schreiber-Agus, N. Mouse models of prostate cancer. *Oncogene* **18**, (1999).

71. Leonard, J. D. *et al.* Identification of Natural Regulatory T Cell Epitopes Reveals Convergence on a Dominant Autoantigen. *Immunity* **47**, 107-117.e8 (2017).

72. Pabbisetty, S. K. *et al.* KLF2 is a rate-limiting transcription factor that can be targeted to enhance regulatory T-cell production. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 9579–9584 (2014).

73. Xing, S. *et al.* Tcf1 and Lef1 are required for the immunosuppressive function of regulatory T cells. *J. Exp. Med.* **216**, 847–866 (2019).

74. Pacholczyk, R. & Kern, J. The T-cell receptor repertoire of regulatory T cells. *Immunology* **125**, 450–458 (2008).

75. Mayer-Blackwell, K. *et al.* TCR meta-clonotypes for biomarker discovery with tcrdist3: identification of public, HLA-restricted SARS-CoV-2 associated TCR features. *bioRxiv* 2020.12.24.424260 (2021) doi:10.1101/2020.12.24.424260.

76. Elhanati, Y., Sethna, Z., Callan, C. G., Mora, T. & Walczak, A. M. Predicting the spectrum of TCR repertoire sharing with a data-driven model of recombination. *Immunol. Rev.* **284**, 167–179 (2018).

77. Venturi, V. *et al.* Sharing of T cell receptors in antigen-specific responses is driven by convergent recombination. *Proc. Natl. Acad. Sci.* **103**, 18691–18696 (2006).

78. Cibotti, R. *et al.* Public and private V beta T cell receptor repertoires against hen egg white lysozyme (HEL) in nontransgenic versus HEL transgenic mice. *J. Exp. Med.* **180**, 861–872 (1994).

79. Gowthaman, R. & Pierce, B. G. TCR3d: The T cell receptor structural repertoire database. *Bioinformatics* **35**, 5323–5325 (2019).

80. Sethna, Z., Elhanati, Y., Callan, C. G., Jr, Walczak, A. M. & Mora, T. OLGA: fast computation of generation probabilities of B- and T-cell receptor amino acid sequences and motifs. *Bioinformatics* **35**, 2974–2981 (2019).

81. Marcou, Q., Mora, T. & Walczak, A. M. High-throughput immune repertoire analysis with IGoR. *Nat. Commun.* **9**, 561 (2018).

82. Aleksic, M. *et al.* Dependence of T Cell Antigen Recognition on T Cell Receptor-Peptide MHC Confinement Time. *Immunity* **32**, 163–174 (2010).

83. Bacher, P. *et al.* Antigen-Reactive T Cell Enrichment for Direct, High-Resolution Analysis of the Human Naive and Memory Th Cell Repertoire. *J. Immunol.* **190**, 3967–3976 (2013).

84. van den Berg, H. A. & Rand, D. A. Quantitative theories of T-cell responsiveness. *Immunol. Rev.* **216**, 81–92 (2007).

85. van den Berg, H. A. *et al.* Cellular-Level Versus Receptor-Level Response Threshold Hierarchies in T-Cell Activation. *Front. Immunol.* **4**, (2013).

86. Dash, P. *et al.* Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature* **547**, 89–93 (2017).

87. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).

88. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).

89. Zhang, L. *et al.* Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. *Nature* **564**, 268–272 (2018).

90. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize Implements and enhances circular visualization in R. *Bioinforma. Oxf. Engl.* **30**, 2811–2812 (2014).

91. Biochemical patterns of antibody polyreactivity revealed through a bioinformatics-based analysis of CDR loops | eLife. https://elifesciences.org/articles/61393.

92. Frith, M. C., Saunders, N. F. W., Kobe, B. & Bailey, T. L. Discovering sequence motifs with arbitrary insertions and deletions. *PLoS Comput. Biol.* **4**, e1000071 (2008).

93. Lunde, E., Løset, G. Å., Bogen, B. & Sandlie, I. Stabilizing mutations increase secretion of functional soluble TCR-Ig fusion proteins. *BMC Biotechnol.* **10**, 61 (2010).

94. Moon, J. J. *et al.* Naive CD4(+) T cell frequency varies for different epitopes and predicts repertoire diversity and response magnitude. *Immunity* **27**, 203–213 (2007).

95. Tungatt, K. *et al.* Antibody stabilization of peptide-MHC multimers reveals functional T cells bearing extremely low-affinity TCRs. *J. Immunol. Baltim. Md 1950* **194**, 463–474 (2015).