2019-06-14

Conserved Features of the T Cell Receptor Repertoire Contribute to the Persistence of EBV-Specific CD8 T Cells

Larisa Kamga
University of Massachusetts Medical School

CONSERVED FEATURES OF THE T CELL RECEPTOR REPERTOIRE

CONTRIBUTE TO THE PERSISTENCE OF EBV-SPECIFIC CD8 T CELLS

A Dissertation Presented

By

Larisa Kamga

Submitted to the Faculty of the

University of Massachusetts Graduate School of Biomedical Sciences, Worcester

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Biomedical Sciences

June 14th, 2019

CONSERVED FEATURES OF THE T CELL RECEPTOR REPERTOIRE

CONTRIBUTE TO THE PERSISTENCE OF EBV-SPECIFIC CD8 T CELLS

A Dissertation Presented

By

Larisa Kamga

This work was undertaken in the Graduate School of Biomedical Sciences

Interdisciplinary Graduate Program

Under the Mentorship of

Katherine Luzuriaga, MD, Thesis Advisor

Lawrence Stern, PhD, Member of Committee

Michael Brehm, PhD, Member of Committee

Laura Gibson, MD, Member of Committee

Paul Thomas, PhD, External Member of Committee

Samuel Behar, MD, Chair of Committee

Mary Ellen Lane, PhD

Dean of the Graduate School of Biomedical Sciences

June 14th, 2019

THIS DISSERTATION IS DEDICATED TO

MY DOTING FAMILY.

Acknowledgements

I am deeply beholden to many more people than those whose names appear here. I would like to start by thanking my thesis advisor, Dr. Katherine Luzuriaga. She has been a devoted mentor, a role model and a beacon of support, guidance and encouragement. She has challenged me, has taught me to think critically and has created an environment that has fostered my professional and scientific growths as an independent investigator. I am forever grateful to her for that.

I would also like to thank current and previous members of the Luzuriaga Lab: Dr. Thomas Greenough, Dr. Maria Paz Gonzalez-Perez, Margaret McManus, Matthew Koch, Alexander Putnam, Dr. Keri Sanborn, Dr. Francisco Ramirez, Linda Lambrecht, James Coderre, Joyce Pepe, Jennifer Henderson, Anita Gautam, Ashley Sweeney and Isaac Wakiro. Thank you to Robin Brody for teaching me the principles of flow cytometry. Thank you to Drs. Eric Weiss and Mohan Somasoraduran for the scientific discussions and feedback.

I am grateful to the study participants. This work would not have been possible without them.

I would like to express my utmost gratitude to my TRAC (Drs. Liisa Selin, Lawrence Stern and Samuel Behar) and DEC (Drs. Lawrence Stern, Samuel Behar, Laura Gibson, Michael Brehm and Paul Thomas).

I would like to acknowledge my college mentors, Drs. Mark Steinberg and Jessica Seeliger, for cultivating my interests in scientific inquiry and my high-school English teacher, Ms. Naomi Pharr for not only teaching me English but for also serving as my first role model and mentor when I moved to the US; every time I put pen on paper, I am reminded of how an eloquent writer you are.

Abstract

Epstein-Barr Virus (EBV) is a ubiquitous human virus linked to several diseases, including cancers. CD8 T cells are important for controlling EBV replication. Generation and maintenance of virus-specific CD8 T cells is dependent on specific interaction between MHC-peptide complexes on the infected cell and the CD8 T cell receptor (TCR). Several lines of evidence suggest that the TCR repertoire is an essential component of the CD8 T-cell immune response. The current work focuses on delineating the **features of the TCR repertoire that drive the selection of EBV-specific CD8 T cells into the memory phase**. We used bulk and single-cell TCRαβ sequencing to analyze the TCR repertoire of human CD8 T cells specific for two immunodominant HLA-A02:01-restricted EBV-derived epitopes: $BRLF1_{109-117}$ (YVLDHLIVV) and $BMLF1_{280-288}$ (GLCTLVAML) during the acute and memory phases of primary EBV infection in humans. We showed that persistent EBV-specific clonotypes accounted for only 9% of unique clonotypes but were highly expanded in acute EBV infection and more commonly expressed identifiable features than non-persistent clonotypes. The other 91% of highly diverse unique clonotypes disappeared and were replaced in convalescence by equally diverse "de-novo" clonotypes. We provide evidence suggesting that recognition of $BRLF1_{109-117}$ may be driven by the TCRα. We identified a highly dominant and degenerate $BRLF1_{109-117}$-specific TCRα sequence, AV8.1-CAVKDTDKLIF-AJ34, that was shared by all donors studied and identified conserved residues within this sequence that were important for antigen recognition. These findings are relevant to current efforts to develop or optimize the efficacy of T cell based therapies or vaccines.

Table of Contents

Tables of Figures

List of Tables

Chapter 1 Introduction

1 The CD8 T-cell immune response

1.1 Antigen processing and presentation for MHC-I

Integral to the CD8 T-cells is their ability to sense foreign invaders. This ability is facilitated by antigen-presenting cells (APCs; *e.g.*, dendritic cells), which present antigens (Ags) in the form of short virus-derived peptides in complex with the major histocompatibility complex I (MHC-I) on their surface. MHC-I molecules are encoded by *human leukocyte antigen* (*HLA*) genes located on chromosome 6 in humans. In humans, there are three MHC-I genes: HLA-A, HLA-B and HLA-C. These Ag-presenting molecules are highly polymorphic and have multiple alleles, which are denoted with a number (*e.g.*, HLA-A2). Each allele is also subdivided into subtypes of closely related sequences (*e.g.*, HLA-A*02:01). Membrane-bound MHC-I molecules consist of a polymorphic polypeptide chain (known as the heavy chain) and a non-polymorphic β2-microglobulin (β2m) subunit [1]. The crystal structure of the HLA-A2 has revealed that the heavy chain of an MHC molecule is subdivided into three domains: α1, α2 and α3 [2]. The α1 and α2 domains are each made up of four strands of a β-sheet topped by an α-helix. These domains form an enclosed cleft where a peptide binds **(Fig 1)** [2].

APCs process Ags for presentation by MHC-I via a stepwise process starting with the proteosomal degradation of cytosolic proteins into short peptides [3]. The transporter for Ag processing (TAP) protein then relays the peptides to pre-assembled MHC-I molecules located in the lumen of the endoplasmic reticulum (ER). MHC-I molecules bind peptides that are typically 8- to 10-

mers via the interaction between conserved anchor residues in the peptide and the Ag-binding cleft of the MHC [3, 4]. The peptide-MHC-I complex (pMHC-I) then get translocated to the cell surface via the golgi apparatus where the peptide is presented to CD8 T cells.

The length of the peptide affects the topology of the pMHC-I. When bound to short peptides, the pMHC-I appears relatively flat. However, given that the MHC-I fold is enclosed, as peptides get longer, they tend to compress and bulge out of the MHC-I cleft [5, 6].

1.2 CD8 T-cell activation, proliferation and differentiation during an infection

During an infection, APCs become activated by inflammatory cytokines such as type I interferons (type I IFN) or CD4 helper T cells (T_H cells) via the CD40-CD40L signaling. This results in enhanced expression of MHC and co-stimulatory (*e.g.*, CD28, CD27) molecules, which are important for optimal T-cell activation and responses [7]. Naïve Ag-specific CD8 T cells recognize cognate APCs via the interaction between their TCRs and the pMHC complex on APCs. This stimulus enhanced by co-stimulatory signals initiates an intracellular signaling cascade that culminates in the activation, massive proliferation and differentiation of a heterogeneous group of effector cells that are equipped with various functions (*e.g.*, secretion of inflammatory cytokines - IFNγ, tumor-necrosis factor (TNF), IL-2 - and cytotoxicity) to clear or control the infection. This is known as the expansion phase of the T-cell immune response (**Fig 2**). The inflammatory milieu and duration, dose and affinity of Ag for the TCR influence T-cell activation, proliferation and differentiation [8-13]. Following control of the infection, the T-cell response is culled by apoptosis (contraction phase) and only a fraction of cells persist as memory T cells, which are poised to control viral reactivation or re-exposure to the same pathogen (memory phase) due to their higher frequencies and lower activation threshold [14]. Factors that allow T cells to avoid apoptosis and to persist as memory cells are incompletely understood. Our current understanding of these factors will be summarized below (see "Memory generation and role of the TCR" on page 5).

2 Role of T-cell receptor in controlling viral infection

2.1 Architecture and assembly of the TCR$\alpha\beta$

As discussed earlier, CD8 T cells survey APCs for the presence of nonself pMHC using the TCR. Each TCR is a membrane-bound, heterodimeric protein that is formed from two polypeptides: α and β. *TCRα* and TCRβ gene loci are located on chromosomes 14 and 7, respectively, in humans. Each locus is subdivided into multiple *variable* (*V*; *TRAV* for TCRα and *TRBV* for the TCRβ), *joining* (*J*; *TRAJ* for TCRα and *TRBJ* for the TCRβ) and *constant* (*C*; *TRAC* for TCRα and *TRBC* for the TCRβ) gene segments. The TCRβ contains an additional gene segment, *diversity* (*D*; *TRBD*). Each gene segment consists of one or more genes, which are given a number (*e.g.*, *TRAV5*). There are approximately 43-45 TRAV, 50 TRAJ, 1 TRAC, 40-48 TRBV, 2 TRBD, 12-13 TRBJ and 2 TRBC functional genes [15].

The TCR$\alpha\beta$ is assembled during naïve T-cell development in the thymus. Each *TCR* locus undergoes a process known as *V(D)J* recombination whereby *V*, *D* and *J* gene segments are randomly rearranged (**Fig 3**) [15]. The TCRβ chain rearranges first, followed by the TCRα; the TCRα chain keeps rearranging until a TCRα chain has been rearranged that is capable of successfully pairing with the TCRβ chain. The TCR is then expressed on the surface of the T cell. *V(D)J* recombination is catalyzed by the RAG protein, which binds and cuts chromosomal DNA at recombination signal sequences (RSS) flanking TCR genes; the intermediary genes are excised and the resultant adjoining genes are ligated by the DNA repair machinery. Each rearranged TCR α or β chain harbors three characteristic domains termed complementary-determining regions (CDRs). CDR1 and CDR2 regions are germline encoded by *TRAV* and TRBV genes [16]. CDR3 regions are formed at the junction between *V(D)J* gene rearrangements and are highly hypervariable due to the additions of non-templated (N) nucleotides or deletions of nucleotides at the junction. This recombination process results in a diverse pool of unique TCRα and β clonotypes. Additions or

deletions of nucleotides at the CDR3 and pairing of different TCRα and β segments further enhance the overall diversity of the TCR repertoire, estimated to range between 10^{15}-10^{20} unique potential TCRαβ clonotypes pre-thymic selection [17, 18] and 2 x 10^7 post-thymic selection [19]. This diversity enables CD8 T-cell responses to a myriad of pathogens.

2.2 Role of TCR in conferring antigenic specificity

A cardinal feature of CD8 T cells is Ag-specificity, conferred by the interaction of the TCR with pMHC on virus-infected cells [20-23]. The TCR repertoire is an important determinant of CD8 T-cell-mediated antiviral efficacy or immune-mediated pathology [17, 24-28]. Structural analyses of TCR-pMHC-I interactions have provided insights into the recognition events between a TCR and a pMHC [29-31]. According to the canonical docking geometry of the TCR-pMHC-I, a TCR binds above the axis of the pMHC-I binding groove in a diagonal orientation, with the TCRα positioned over the α2-helix and the TCRβ over the α1-helix (**Fig 4).** Regions of the peptide that are protruded out of the groove and are solvent exposed can directly contact the TCR, whereas buried residues can indirectly affect TCR binding (for example by stabilizing the Ag-binding cleft) [32]. The CDR3 loops primarily contact the solvent-exposed chains of the MHC-bound peptide (the CDR3 α and β loops engage the N- and C-terminus of the peptide, respectively). In doing so, CDR3 loops are thought to confer Ag specificity. The CDR1-2 loops typically fix the TCR to the MHC. Naturally, deviations from the canonical geometry have been noted, including the percentage of contributions from the TCR α and β chains and the strength of the interaction. The effect of these deviations in the docking geometry of TCR-pMHC interaction on CD8 T-cells signaling, functionality, and memory formation have garnered attention. There are reports that the docking geometry modulates TCR signaling and T-cell activation [11, 33]. There are also evidence supporting that the docking geometry impacts TCR selection, immunodominance and TCR bias [5, 6, 34-37]. Altogether, these studies support a role for the TCR in

shaping the CD8 T-cell response and have prompted an interest in gaining insights into structural properties of TCR repertoires that lead to TCR-dependent differences in T-cell signaling and differentiation.

2.3 Memory generation and role of the TCR

How memory CD8 T-cells are generated is still not fully understood. The inflammatory milieu, transcription factors and the strength of TCR-pMHC are thought to contribute. For example, the balance between the transcription factors Eomes and T-bet controls T-cell differentiation; while Eomes triggers memory T-cell differentiation, T-bet leads to effector T-cell differentiation [38-40]. Inflammatory cytokines such as IL-7, and IL-15 favor memory differentiation while IL-12 promotes effector CD8 T cell differentiation by modulating the balance between T-bet and Eomes expression [39-42]. In a mouse model of viral infection, mice lacking the IL-7 receptor (IL-7R) had a reduced number of virus-specific memory CD8 T cells compared to wild-type mice, indicating that signaling through the IL-7R is required to induce the formation of virus-specific memory CD8 T cells [41]. On the contrary, Kalia *et al.* [43] reported that sustained signaling through the CD25, the IL-2 receptor alpha chain, inhibits memory formation. Using a mouse model of acute LCMV infection, they showed that effector CD8 T cells expressing high levels of CD25 were prone to apoptosis and differentiated into effector cells while cells expressing low levels of CD25 took on a memory phenotype.

The lineage relationship between effector and memory CD8 T cells remains elusive. An important question is whether memory CD8 T cells descend from effectors cells or from other progenitor cells? The evidence gathered up to date lends support to multiple pathways to becoming a memory CD8 T cell. Some studies have suggested that a naïve T cell can contemporaneously give rise to an effector and memory T cells [7, 44]. Others have proposed a gradual and stepwise process of differentiation from naïve cells to effector to memory CD8 T cells. Finally, some data indicate that memory CD8 T cells can arise independently of effector differentiation [7, 10]. Signaling through the

TCR influences memory formation. Using a mouse model, Teixeiro *et al.* [12] showed that ovalbuminin-specific CD8 T cells harboring a point mutation in the transmembrane domain of their TCRβ chain had an unimpaired proliferation, effector differentiation and functional output but were impaired in their ability to differentiate into memory CD8 T cells, likely due to a decreased recruitment of TCRs to the immunological synapse and decreased NF-kB signaling. These data indicate that TCR signaling is dichotomized into effector and memory differentiation signaling and raise the possibility of TCR-dependent differences in directing cell fate. The role of signal strength in T-cell differentiation remains debatable. Some studies suggest that high-affinity TCRs are important for memory differentiation while others suggest that T cells expressing low-affinity TCRs could also differentiate into memory T cells [13, 45-48]. It has been difficult to address the lineage relationship between effector and memory CD8 T cells in humans due to the heterogeneity of the T-cell response and the lack of the ability to track the developmental process of a single T cell over the course of an infection. The advent of high-throughput and single-cell TCR sequencing makes it easier to obtain the TCR sequences of cells and to use them as molecular identifiers of T cells. Given that a T cell expresses a unique TCR, one can use the TCR sequences to label cells that derived from the same ancestor. This study aims to leverage these sequencing tools to track virus-specific CD8 T cells over the course of an infection in order to decipher their cell fate and potential features that may contribute to clonal persistence.

TCR affinity maturation, whereby you have the selection into the memory pool of TCR clones having an affinity falling within a particular range, has been documented during T-cell responses to multiple human viral infections. Abdel-Hakeem *et al.* [49] showed that during a hepatitis C virus (HCV) infection, the initial TCR repertoire evolves and narrows in breadth into the memory phase. Similar findings were observed in other infections [48, 50]. These studies suggest that there are optimal TCRs that are better fit to induce memory T-cell differentiation. On the other hand, there are reports that the initial virus-specific TCR repertoire reflects the memory virus-specific repertoire, indicating

6

that each TCR has the potential to become memory cells [51, 52]. These divergent findings likely reflect the unique inflammatory state induced by each infection and emphasize the importance to investigate the effects of the TCR repertoire on memory T-cell formation in the context of the spectrum of human infections.

2.4 Principles underlying TCR bias

TCR repertoire analyses enable us to visualize the clonotypic identity of TCRs and to glean information about important features that dictate antigenic specificity or recognition [53, 54]. Novel sequencing methods have shed important insights into the composition and organization of the TCR repertoires of common pathogens. Despite the multitude of V and J genes and virtually limitless number of TCRs that can be made from V(D)J recombination, skewing of pathogen-specific repertoires have been observed [34, 37, 49, 55-57]. These skewed repertoires have been observed in the forms of preferential usage of particular V and/or J genes, a conserved amino acid motif within the CDR3α/β, or identical CDR3α/β amino acid sequence. Additionally, some pMHC complexes select for TCRs having identical CDR3α/β amino acid sequences that are shared among multiple individuals and these are referred to as public CDR3α/β in contrast to private CDR3α/β, which are exclusively found in one individual [57, 58]. There has been a growing interest in understanding how these biases emerge, their immunological relevance, and their implications for either protection or immunopathology [24, 37, 53, 54, 59, 60].

TCR bias can be initiated during T-cell development or over the course of the immune response. The various mechanisms underlying such bias can occur during TCR transcription, splicing or as a result of structural constraints during Ag recognition. It has been shown that the promoter of each *TRBV* gene has different transcription efficiency, resulting in varying level of transcripts [61]. Moreover, McMurry *et al.* [62] have shown, albeit in the context of the TCR-δ system, that enhancer

elements can regulate which TCR genes are selected during *V(D)J* recombination. Livak *et al.* [63] reported preferential TRBV and TRBJ pairings pre-thymic selection likely due to the efficiency of different RSS sequences in recruiting RAG protein. Convergent recombination has been proposed as a mechanism for the prevalence of public TCRs [64, 65]. It is a process whereby many different recombination events converge to produce a specific nucleotide sequence and many different nucleotide sequences converge to produce a specific amino acid sequence due to codon degeneracy. These public TCRs can arise from either germline-encoded CDR3 sequences having few N-nucleotide additions or nongermline-encoded CDR3 sequences containing numerous junctional modifications (N-nucleotide additions or nucleotide deletions).

TCR biases can be a reflection of the TCR-pMHC binding modes [5, 6, 35, 37, 64, 66]. The conservation of specific amino acids tends to indicate that these residues are critical for Ag recognition. For example, the TCR repertoire of the HLA-A02:01-restricted GIL peptide from influenza virus displays a strong preservation of an arginine residue within its CDR3β region. It was determined that the conserved arginine was important for Ag recognition [35]. MHC alleles may contribute to TCR biases by preferentially interacting with specific TRAV proteins [64, 66].

3 Primary EBV infection

3.1 EBV entry and early replication

EBV is a human γ-herpesvirus with a double-stranded 1.7 kbps DNA genome [67]. EBV primarily infects B and epithelial cells and is transmitted orally. EBV uses gp350 glycoprotein to initially attach to CD21 (or CD35) on B cells [68]. Then, a trimer consisting of gp42 glycoprotein and gH/gL attaches to MHC-II on B cells, triggering a conformation change that facilitates the recruitment of gB fusion loops and eventual fusion of the viral envelope and the B cell endocytic membrane.

Given the lack of CD21 or MHC-II expression on epithelial cells, entry into these cells is thought to occur following attachment of gH/gL complex to integrins on cellular membrane. This is thought to facilitate the recruitment of the gB fusion complex and eventual fusion of the viral envelop and cellular membrane. Once inside cells, the virus hijacks the host DNA replication machinery to replicate, leading to massive virus shedding in the oral cavity. EBV establishes latency (lifelong persistence in a dormant state) in memory B cells by promoting growth-transformation of infected B cells, followed by downregulation of transforming genes [69, 70]. Intermittent reactivation of the virus from latency to the lytic cycle can occur, although the mechanism of this reactivation is still not well understood. Viral reactivation is common following T-cell immunosuppression [71].

3.2 Acute infectious mononucleosis

EBV infects almost 95 percent of the world's population by the fourth decade of life. In older children and adults, primary infection with EBV often manifests as acute infectious mononucleosis (AIM), a self-contained illness characterized by fever, pharyngitis, lymphadenopathy and malaise [72]. Activation and expansion of virus-specific CD8 T cells in the peripheral blood in AIM are among the most robust described for any viral infection. These activated cells contract in convalescence. AIM is thought to be an immunopathology that results from a high level of inflammatory cytokines produced by the markedly expanded virus-specific CD8 T cells [73, 74]. Prior infection with viruses such as influenza A virus (IAV) might exacerbate the effect of AIM likely due to cross-reactivity between IAV-specific memory T cells and EBV-derived Ags [73]. AIM is associated with an increased risk of multiple sclerosis or hodgkin lymphoma [72]. Latent EBV infection is also associated with Burkitt's lymphoma and nasopharyngeal cancer [75]. In immunocompromised individuals such as those undergoing organ transplantation, EBV reactivation can lead to lymphoproliferative malignancies such as post-transplant lymphoproliferative disorder (PTLD) [75, 76]. Unfortunately, there is currently no

vaccine to prevent EBV infection. Thus, there is great interest in the development of a vaccine to prevent EBV infection [77].

EBV-specific memory T cells provide protection from EBV reactivation as evidenced by the fact that EBV-associated post-transplant lymphoproliferative disorders can be prevented or treated by adoptive transfer of EBV-specific CD8 T-cells [76, 78-80]. The protective quality of EBV-specific memory T cells was further validated during a clinical trial of an anti-CD3 antibody therapy for type 1 diabetes [71]. EBV reactivation and increased EBV viral load were documented in participants that received the drug [71]. Altogether, these indicates that T cells are an essential gateway between viral control and reactivation and that our immune system has designed T cells equipped with protective features to tip the balance in favor of the host. Given that AIM is common, easily diagnosed and there is a robust CD8 T-cell response that is likely responsible for an initial and long-term control of viral replication, the virus-specific CD8 T-cell response elicited during AIM presents as a great model to investigate the features of a T-cell response that may be important for controlling a persistent infection in human beings.

3.3 CD8 TCR repertoires to HLA-A*02:01-restricted BRLF1$_{109-117}$- and BMLF1$_{280-288}$-derived epitopes during acute EBV infection

EBV elicits a robust CD8 T-cell response during acute infection. EBV expresses hundreds of proteins (lytic and latent, expressed in the lytic and latent phase infection, respectively). Despite the large number of EBV proteins, the CD8 T-cell response tends to be focused on a few lytic proteins such as BRLF1 and BMLF1 [68, 81]. The CD8 T-cell response to a single lytic epitope can make up to 40% of the total CD8 T-cell response [81-85]. There tends to be a predictable pattern of hierarchy in the magnitude of the response to EBV-derived epitopes that is conserved across many individuals (this is termed immunodominance) [81]. For example, during acute infection in HLA-A:02+ individuals, the EBV-specific CD8 T-cell response is dominated by responses to two lytic proteins, BRLF1

(BRLF1$_{109-117}$ epitope: YVLDHLIVV; herein referred to as YVL) and BMLF1 (BMLF1$_{280-288}$ epitope: GLCTLVAML; herein referred to as GLC) (**Fig 5**) [81, 83]. As the disease subsides, the magnitude of the response to lytic epitopes is drastically culled [81-85]. There has been a great interest to understand if and how the TCR repertoire evolves over the course of a human infection.

While the GLC-specific TCRβ repertoire has been well characterized, the TCRα (and paired TCR$\alpha\beta$ repertoire) is less well characterized and the TCR repertoire to YVL has not been well-studied [50, 53, 55, 58, 64, 86-88]. The paucity of data on the TCRα repertoire is due to technical constraints associated with the lack of anti-TRAV specific antibodies and T cells' ability to co-express two α-chains [89, 90]. An early examination of the dynamics of the TCR$\alpha\beta$ repertoire of GLC-specific CD8 TCR repertoire during acute EBV infection documented that the TCR repertoire is oligoclonal and evolves over time; *i.e.*, TCR clonotypes that dominate early on become subdominant in the memory phase [86]. One caveat of this study is that the TCR repertoire analysis was performed on cells that had been cultured *in vitro* for a long period of time and such manipulation might unpredictably alter the repertoire. Moreover, the depth of coverage was low given that very few cells were analyzed.

Analyses of the TCRβ repertoire of *ex vivo* GLC-specific CD8 T cells isolated at a single time point from chronically infected donors has uncovered a high level of bias [50, 64, 87, 88, 91]. There is conserved use of AV05 and BV20 across many individuals. More interestingly, there are a number of public, *i.e.,* shared across two or more individuals, TCRβ chains (*e.g.*, BV20.1-CSARDGTGNGYTF-BJ1.1) and they tend to dominate the response. A highly dominant and public GLC-specific TCR known as AS01 (AV5-CAEDNNARLMF-AJ31/BV20.1-CSARDGTGNGYTF-BJ1.1), has been identified [91]. The ubiquitous nature of this TCR has been postulated to occur through convergent recombination and examination of the crystal structure of AS01 has revealed that the TCR appears to

be preferentially selected because it mostly uses germline-encoded CDR1 and CDR2 residues from AV5 and BV20 to engage the GLC-HLA-A2 complex [64].

The advent of TCRα/β deep sequencing and single-cell paired TCRα β sequencing affords researchers the opportunity to analyze TCR repertoires at a high depth and coverage. Leveraging these techniques, Nguyen and colleagues reported that the repertoire of GLC-specific CD8 T cells in chronically infected EBV individuals undergoing transplantation is stable pre- and post-transplantation [92]. Dash *et al.* [53] has demonstrated that GLC-specific TCR selection in the periphery of chronically infected donors is driven by both the TCRα and TCRβ chain, consistent with previous work [64]. Comprehensive and longitudinal studies of both GLC- and YVL-specific CD8 T cells over the course of acute EBV infection are still lacking.

4   Goals and implications of thesis project

4.1 Goals

The overall goal of this thesis project is to delineate the features of the CD8 TCR repertoire that are important for the selection of EBV-specific CD8 T cells into the memory phase following primary infection with EBV in humans. We have restricted our studies to YVL and GLC epitopes because HLA-A2 is the most common allele within our study cohort and these epitopes are immunodominant, affording us the opportunity to have access to large number of cells to conduct the studies.

There is an inherent functional and phenotypic heterogeneity within and between EBV epitope-specific populations [84, 85]. Catalina *et al.* [84] showed that GLC-specific CD8 T cells consist of different cell subsets expressing various combinations of the markers CD45, CCR7. Greenough *et al.* [85] further highlighted this diversity by showing that the expression of PD-1, an exhaustion marker,

varies depending on TCR usage. Interestingly, PD-1 protein expression (which has been associated with T-cell activation, exhaustion and dysfunction [93, 94]) on YVL-specific CD8 T cells varies by Vβ gene usage [85], suggesting a TCR-dependent T-cell signaling. As discussed earlier (see "Memory formation and role of TCR" on page 6), the TCR signal strength is thought to modulate T-cell differentiation and the maintenance of the memory TCR repertoire [13, 45-48]. Affinity maturation of the CD8 TCR repertoire has been documented in various infections [48-50], suggesting that there is a selection of particular TCRs into the memory phase. This implies that different TCRs are endowed with different abilities and that different TCR may trigger different signaling cascades leading to different transcriptional, functional or phenotypical outcomes. This raises the question as to whether each TCR has a unique potential to abscond the apoptotic phase and become memory CD8 T cells.

The aforementioned differential expression of PD-1 by Vβ genes along with reports implicating TCRs as molecular determinants of T-cell function and fate [8, 9, 12, 13, 45-48, 85, 95, 96] raise the possibility that the nature of the specific pMHC-TCR interactions may regulate the fate and function of EBV-specific CD8 T cells. It prompted us to ask whether there is a selection of particular EBV-specific TCR clones into the memory phase. We have hypothesized that **features of the TCR repertoire drive the selection of EBV-specific CD8 T cells into the memory phase.** To investigate this hypothesis, we used high-throughout and single-cell TCR sequencing to probe the repertoire of YVL- and GLC-specific CD8 T cells isolated over the course of infection from individuals and to identify potential properties important for driving TCR selection and persistence. We applied a newly developed analytical tool [53] and molecular techniques to identify residues within CDR3 regions that might be important for mediating TCR selection and Ag recognition. The upcoming chapters (2 – 5) are organized as follows. In Chapter 2, we used bulk sequencing to comprehensively probe the TCR α / β repertoires of YVL- and GLC-specific CD8 T cells isolated from 3 individuals in AIM and CONV to gain deeper insights into the composition and organization of the TCR repertoire. We also investigated whether there was a selection of TCRs into CONV and potential features of the TCR

13

repertoire that may likely contribute to TCR persistence. In Chapter 3, we used single-cell sequencing to probe the paired TCR $\alpha\beta$ repertoires of YVL- and GLC-specific CD8 T cells isolated from the same individuals in AIM and CONV. We also used a recently developed analytical tool [53] to predict key residues within CDR3 regions that likely mediated Ag recognition and validated their importance using *in vitro* assays. Chapter 4 discusses our findings in the context of the current literature and the implications and limitations of the study. It also proposes future directions. Chapter 5 describes the methodologies employed.

The findings reported in Chapter 2 are part of the following manuscript:

Gil A, Kamga L, Chirravuri R, Aslan N, Ghersi D, Luzuriaga K & Selin LK. EBV epitope/MHC interaction combined with convergent recombination drive selection of diverse T cell receptor α and β repertoires.

The findings described Chapter 3 are part of the following manuscript:

Kamga L, Gil A, Song I, Brody R, Ghersi D, Aslan N, Stern LJ, Selin LK & Luzuriaga K. CDR3 α drives selection of the immunodominant EBV BRLF1-specific CD8 T cell receptor repertoire in primary infection.

4.2 Implications

This study is significant for multiple reasons. First, it addresses the limitations of past studies by using a combination of a high-throughput and single-cell TCR sequencing to gain a comprehensive understanding of the composition, organization and structural features of the TCR repertoire of not only GLC-specific CD8 T cells but also the less studied YVL. Single-cell analyses provide information on the paired TCR $\alpha\beta$, an important determinant of T-cell specificity and Ag

recognition. It is one of the first studies to provide insights into the YVL repertoire. Additionally, this is a longitudinal study that is performed in the context of a human viral infection; hence, it enables us to glean important structural components of the TCR repertoires that may facilitate Ag recognition and the persistence of human CD8 T cells into the memory phase. Improved T-cell therapies would benefit from insights derived from approaches that integrate computational biology and structural modeling to predict optimum TCR features and identify TCR specificity groups [24, 53, 54, 97-99]. These methods need to be based on an accurate and in-depth understanding of Ag-specific TCR repertoire structure and organization from studies like the one presented here. The attributes of a desirable T-cell based therapies or vaccines include magnitude of the T-cell response, functional potency, and clonal persistence. The availability of computational tools to predict TCRs that will elicit a desirable response will accelerate the discovery of optimal TCRs as well as leads to the characterization of biomarkers associated with optimal TCRs, which will empower researchers to predict the efficacy of T-cell based vaccines. The current study aims at characterizing the TCR repertoire of human-derived virus-specific CD8 T cells in an attempt to delineate properties of the TCR repertoire that may contribute to persistence of cells into the memory phase. Ultimately, this could lead to better understanding of how EBV-specific CD8 T-cells control EBV replication and facilitate the development of CD8 T-cell vaccines or T-cell based therapies to prevent primary EBV infection and the incidence of EBV-associated diseases [77-79]. There are ongoing clinical trials to test CD8 T cells as a therapy for EBV-associated cancers. For example, CMD-003 (autologous EBV-specific T cells) is being tested in a phase 2 clinical trial as a therapy for EBV-positive NK/T-cell lymphoma (study name: CITADEL; NCT01948180) and PTLD and large B cell diffuse lymphoma (CIVIC; NCT02763254). Understanding the properties of the TCR repertoires of EBV-specific CD8 T cells associated with T-cell persistence and Ag recognition will be highly relevant to these clinical enterprises and could empower researchers to predict the efficacy of potential EBV vaccine candidates.

α1

α2

β2m

α3

α1

α2
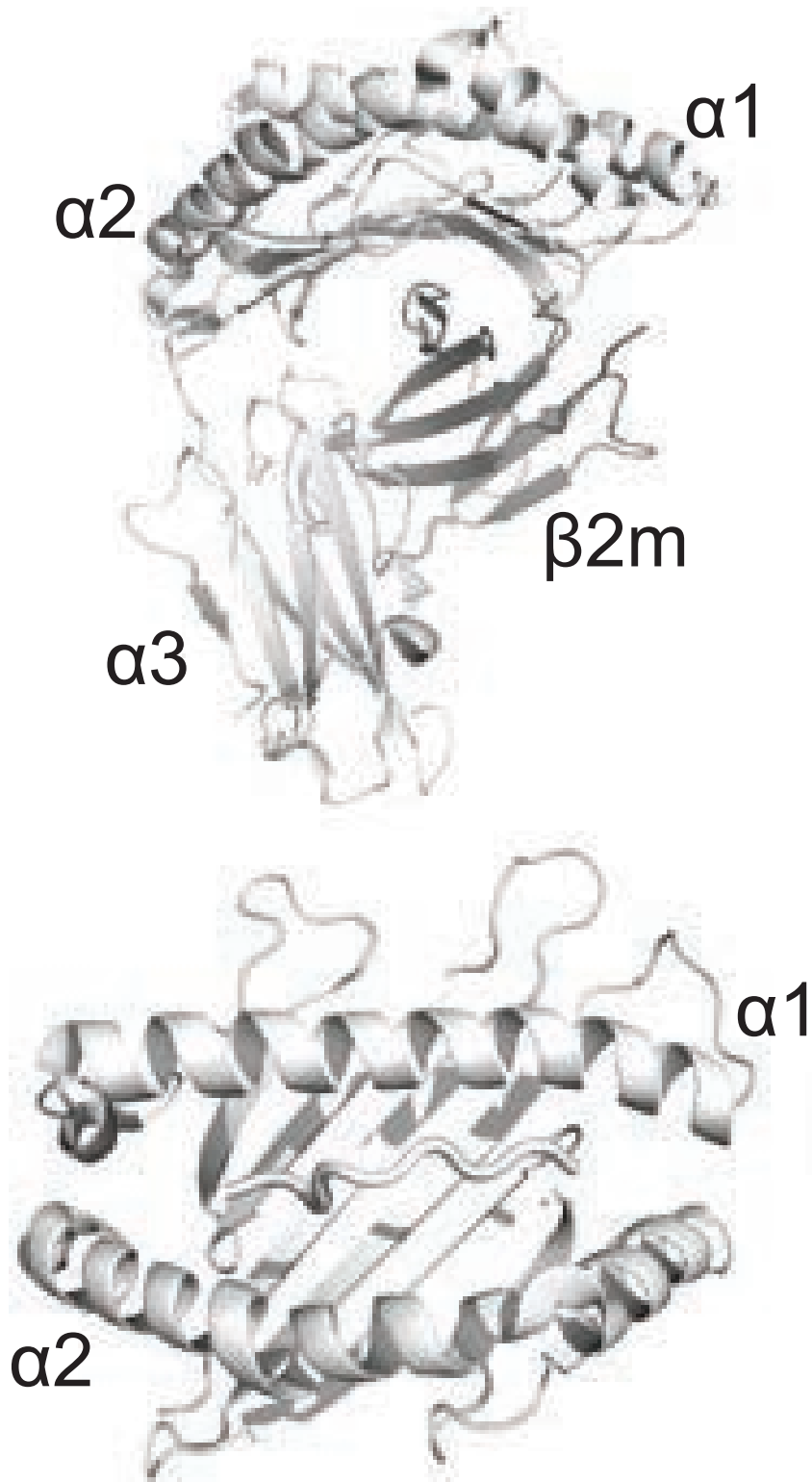
Adapted from Adams Annurev immunol 2013

Fig 1. Structure of HLA-A2. Bottom panel: close-up view of the peptide-binding cleft.
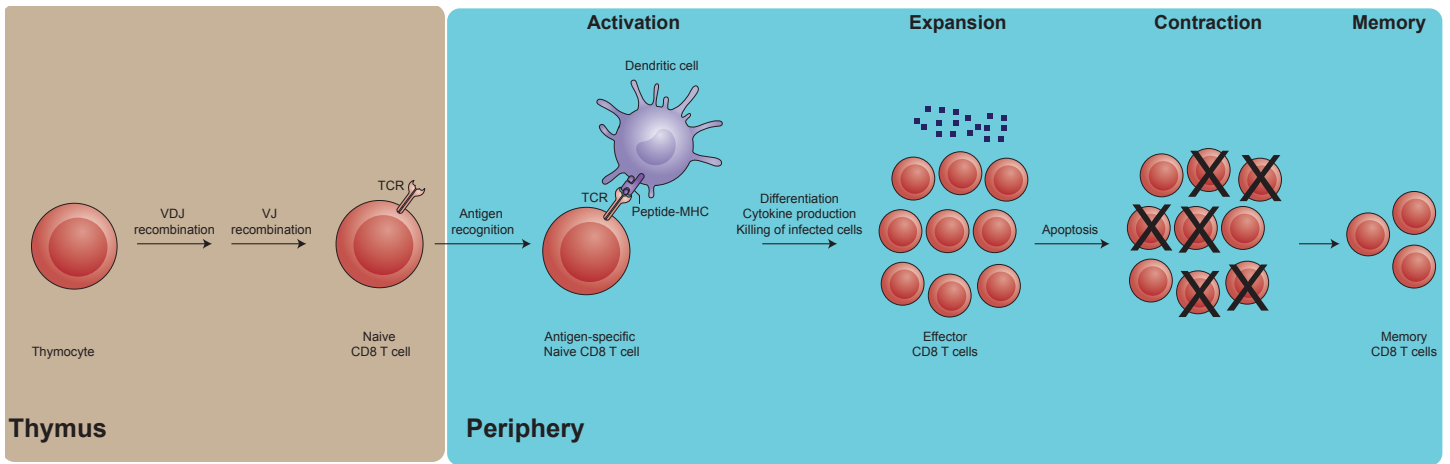
Fig 2. Phases of the T-cell immune response.
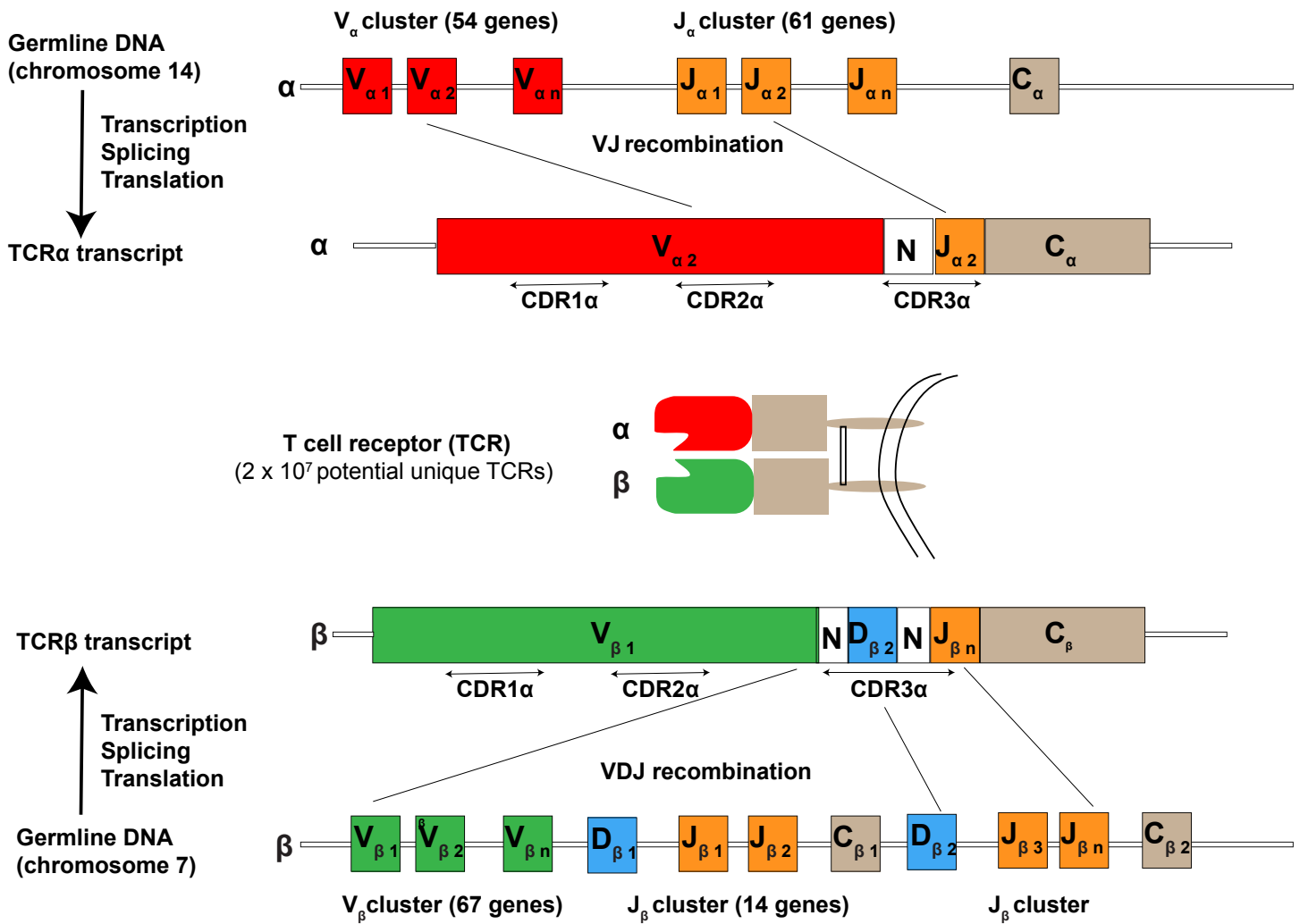
Fig 3. gDNA rearrangement of TCRα and TCRβ loci by VJ and VDJ recombinations. V: variable; J: joining; D: diversity, N: additions and/or deletions of nucleotides. 2×10^{7} is the estimated number of unique TCRs post-thymic selection (Nikolich-Zugich Nat Rev Immunol 2004).
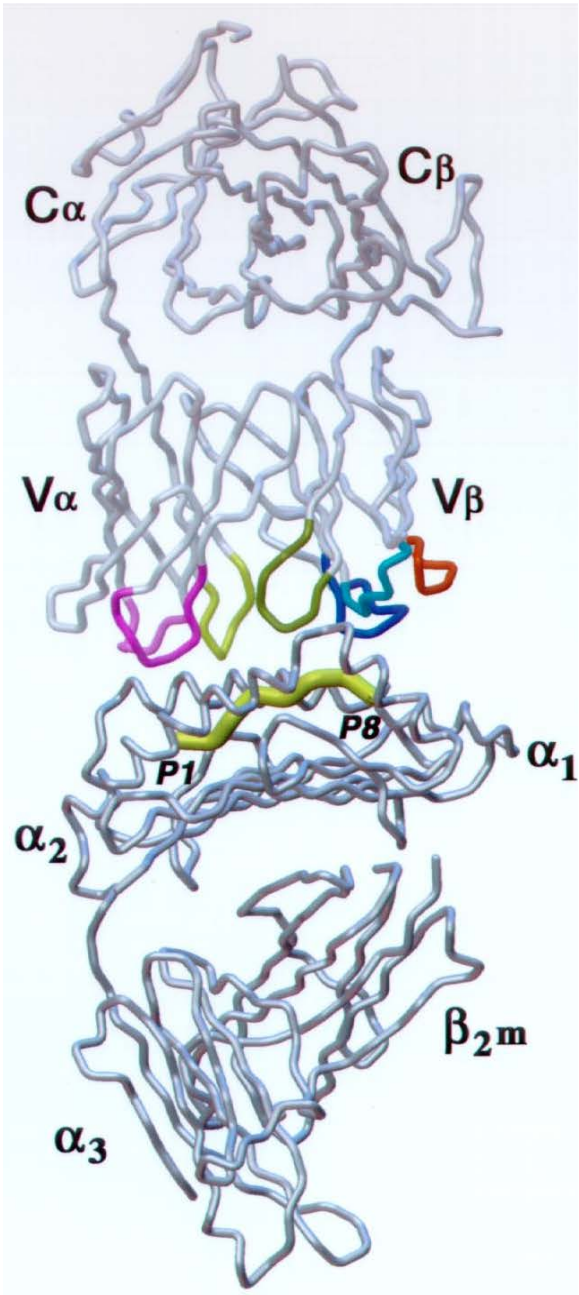
Fig 4. Canonical TCR-pMHC docking geometry.
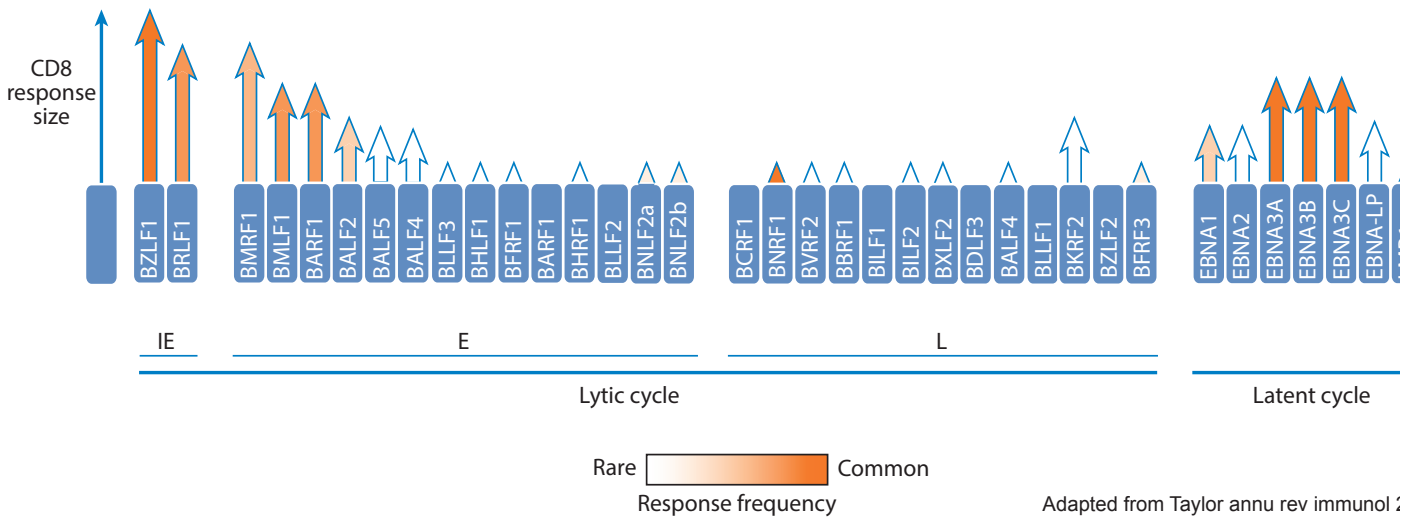
Fig 5. Immunodominance pattern in the CD8 T-cell responses to EBV-derived proteins.
IE: immediate early proteins. E: early proteins. L: late proteins.

Chapter 2 Selection of YVL- and GLC-specific TCR clonotypes from acute EBV infection into convalescence

This chapter was adapted from the following manuscript: EBV epitope/MHC interaction combined with convergent recombination drive selection of diverse T cell receptor α and β repertoires. Gil A, Kamga L, Chirravuri R, Aslan N, Ghersi D, Luzuriaga K & Selin LK.

Luzuriaga and Selin conceived the study and collaborated with clinicians to obtain the samples. Kamga and Gil contributed to study design, and were primarily responsible for cell sorting and TCR sequencing. All authors contributed to data analyses. Ghersi and Chirravuri performed all computational analyses.

1 Introduction

The CD8 TCR repertoire is an important determinant of CD8 T-cell-mediated antiviral efficacy or immune-mediated pathology [17, 24-28]. Generation and maintenance of virus-specific CD8 T cells is dependent on specific interaction between MHC-peptide complexes on the infected cell and the TCR. Defining the relationships between early and memory CD8 TCR repertoires is important to understanding structural features of the TCR repertoire that govern the selection and persistence of CD8 T-cells into memory. Application of high-throughput sequencing on bulk virus-specific CD8 T cells provides an unbiased approach to understanding Ag-specific TCR$\alpha\beta$ repertoires. [34, 37, 49, 55, 56] TCR$\alpha\beta$ repertoires of CD8 T-cell responses to common viruses (influenza, cytomegalovirus, hepatitis C virus) are highly diverse and individualized (*i.e.* "private") but "public" clonotypes (defined as the same V, J, or CDR3 aa sequences in multiple individuals) are favored for expansion, likely due to selection for optimal structural interactions [34, 37, 49, 55, 56].

To characterize TCR repertoires and evaluate molecular features of TCR sequences that likely contribute to the persistence of EBV-specific CD8 T cells over time following infection, we used direct *ex vivo* deep sequencing of TCR Vα and Vβ regions of bulk CD8 T-cells specific to two immunodominant epitopes, GLC and YVL, isolated from peripheral blood of individuals during primary EBV infection (AIM) and 6 - 8 months later in convalescence (CONV). Each TCR repertoire was diverse and individual-specific but also expressed epitope-specific elements that were shared across individuals. Persistent clonotypes accounted for only 9% of the unique clonotypes; the other 91% of unique clonotypes deployed in acute infection were replaced in convalescence by a set of *de novo* clonotypes. Additionally, persistent clonotypes expressed specific CDR3 motifs more commonly than non-persistent ones.

2 Characterization of YVL-specific TCR α and β repertoires during acute EBV infection and convalescence

Deep sequencing analyses of TCR α and β repertoires of EBV-specific CD8 T cells were conducted directly *ex vivo* on pools of tetramer-sorted CD8 T cells isolated from 3 donors (E1603, E1632 and E1655; **Tables 1 & 2**) at time points, AIM (within 0 – 2 weeks after presentation with symptoms) and CONV (6 – 8 months later). The characteristics of the TCR repertoires for each of the 3 donors were elucidated by analyzing TCR sequences for preferential use of V genes and V-J pairings and CDR3 motifs within the two most dominant CDR3 lengths. Preferential use of certain characteristics would suggest that these features are important for pMHC interaction and selection of the TCR repertoire [37, 53, 100-104].

To characterize YVL-specific TCR repertoires, we first assessed the pattern of V gene usage (**Fig 6**). Even though each individual had a unique pattern of V gene usage, there were dominant V genes that were shared in all individuals. The YVL-specific TCRα repertoire was focused on one dominant family, AV8, which was used by all donors in AIM and CONV (**Fig 6A**). By contrast, such a strong bias for a single V gene was not observed in YVL-specific TCRβ usage; there was preferential usage of multiple families, including BV6, BV20, BV28, BV29 (**Fig 6B**). The preferential use of AV8 displayed by the YVL response suggests an Ag-driven selection of TCR clonotypes and that AV8 may be an important selection factor for YVL-specific CD8 T cells. Furthermore, we observed that in AIM, CDR3α length distribution varied from 9- to 14-mers; the two most dominant CDR3α lengths were 9- and 11-mers and together they represented 49.5%±5.9 and 38.3%±8.8 of all clonotypes in AIM and CONV, respectively (**Fig 7A**). Circos plot analyses of the 9-mer CDR3α clonotypes showed that the dominant AV8.1 gene almost exclusively paired with AJ34 (**Fig 8A**) and this AV8.1-AJ34 pairing was present in all donors. CDR3α motif analyses revealed a pronounced motif, "VKDTDK", in these shorter 9-mer clonotypes, representing 13.8%±5.6 and 10.9%±6.4 of the total CD8 T-cell response in

AIM and CONV, respectively **(Fig 9A; Table 3A)**; 87%±1.7 of the clonotypes using this motif in AIM were AV8.1 and 92%±1.7 were AJ34 (**Table 3A**). Interestingly, this motif was also generated by multiple other Vα and Jα genes, including AV12, AV21 and AV3 among the most common (**Table 3A**). The fact that a dominant AV8.1-AJ34-expressing clonotypes contained a prominent motif ("VKDTDK") that was conserved in all study participants from AIM through CONV suggests that 9-mer AV8.1-VKDTDK-AJ34 expressing clones were highly selected by A2-YVL.

The YVL CDR3β length ranged between 9 and 16-mers; the two most dominant CDR3β lengths were 11- and 13-mers and together they represented 56.21%±5.3 and 58.25%±1.2 of clonotypes in AIM and CONV, respectively (**Fig 7B**). There was a preferential usage of BV20-BJ2.7 pairing within the 11-mer response **(Fig 8B)** without a CDR3β motif **(Fig 9B)**, highlighting a great degree of diversity in the amino acid composition. Within the 13-mer response, the CDR3β motif, "LLGG", was present at a relatively low level in most donors **(Fig 9; Table 3B)**. This motif arose predominantly from BV28-BJ1 pairing (donors E1603 and E1632) or BV6-BJ2 pairing (E1655). Clonotypes with this motif were only a minor part of overall responses in 2 donors (E1603 and E1655) and composed 17.4% and 6.74% of the total YVL TCRβ repertoire in E1632 in AIM and CONV, respectively. Hence, we did not identify a prominent CDR3β motif that was conserved in all three donors. Altogether, the TCRα, but not TCRβ, exhibited properties that were shared across all study participants, suggesting that the TCRα chain (*e.g.*, the shared 9-mer AV8.1-VKDTDK-AJ34 expressing clones) appeared to drive selection of YVL-specific CD8 T cells and may likely contribute to Ag recognition to a higher extent than its associated TCRβ.

3 Characterization of TCR α and β repertoires of GLC-specific CD8 T-cells during acute EBV infection and convalescence

Similar to YVL, GLC-specific TCR α and β repertoires exhibited conserved use of particular V and J gene families, consistent with prior reports [86, 105], that was maintained from AIM into CONV. All donors used AV5, AV12, BV20, BV14, BV9, BV28 and BV29 in both AIM and CONV **(Fig 10)**. There were some individual changes in the transition into CONV with AV20 becoming extinct in E1603; AV19 emerging as dominant in E1632 (**Fig 10A**). Clonotypes with 9- and 11-mer CDR3α lengths represented 63.1%±4.7 and 43.8%±5.5 of the total response in AIM and CONV, while clonotypes with longer 11- and 13-mer CDR3β lengths represented 66.6%±16.68 and 62.6%±6.1 in AIM and CONV of total response (**Fig 7)**. Circos plot analysis of the 9-mer CDR3α length clonotypes revealed a conserved and dominant AV5-AJ31 pairing in all 3 donors **(Fig 11B)**. A prominent motif, "EDNNA", was identified within 9-mer clonotypes and predominantly resulted from the AV5-AJ31 recombination event (**Fig 12A; Table 4A**). This CDR3α motif was used by 2.8%±1.7 and 4.9%±2.5 of all unique clonotypes recognizing GLC in the 3 donors in AIM and CONV, respectively (**Table 4A**). The AV12 family dominated the 11-mer response and paired with multiple different Jα genes depending on the donor. The 11-mer CDR3β BV14-BJ2 pairing exhibited a conserved, previously reported public motif, "SQSPGG" [92], which represented 26% and 40% of the total GLC-specific response in donors E1632 and E1655 in AIM, respectively **(Fig 12; Table 4B)**. Within the CDR3β 13-mer responses, a conserved BV20-BJ1 pairing, including the previously reported public motif, "SARD", was used by all 3 donors, and represented 11%±6 and 13.1%±5.3 of the total GLC-specific response in AIM and CONV **(Figs 11-12; Table 4B)**. Within the 13-mer CDR3β response, there was also a consensus motif, "SPTSG" present in all 3 donors, which was derived from multiple different Vβ families and represented 20% and 2% of the total response in donors E1632 and E1655, respectively in AIM and 22.4% and 0.2% in CONV **(Figs 11-12; Table 4B)**. Altogether, both the TCR

α and β repertoires of GLC-specific CD8 T cells exhibited CDR3 motifs that were conserved across all study participants suggesting that both TCR α and β chains appeared to drive selection of GLC-specific CD8 T cells.

4 Persistent clonotypes represent a small fraction of unique clonotypes

To examine features that drive selection of YVL- and GLC-specific TCRs into the memory phase, YVL and GLC TCR α/β repertoires were compared between AIM and CONV. Each unique TCR α or TCR β clonotype (defined as a unique DNA rearrangement) elicited during AIM that was also detected during CONV was defined as a "persistent" clonotype. Clonotypes were regarded as "non-persistent" or "*de novo*" if they were present only during AIM or CONV, respectively. A high level of TCR diversity was maintained from AIM to CONV; however, the overlap between the number of unique clonotypes detected during AIM and CONV was small (**Fig 13**). Only a small fraction, 6.6±2.2 - 9.1±4.2% of the TCRα/β YVL- and GLC-specific unique clonotypes, respectively, present in AIM were maintained at 8.7±4.9 - 18.5±5.6% during CONV. However, they occupied 57.5±26.2 - 75.5±12% of the magnitude (takes into account sequence reads) of the total epitope-specific response in AIM and 35.8±10.2 - 55.8±13.4% in CONV. To account for the inevitable confounding effects of sampling errors, we leveraged the Chao1 method [106], a nonparametric statistical analysis based on bootstrapping that has been used in human studies [107] to predict the degree of TCR clonotypes sharing between different TCR repertoires. To perform Chao1 calculations, we used the online tool EstimateS [108] in order to predict the number of TCR clonotypes shared between AIM and CONV (overlap) for each epitope and to compare it to the observed overlap. The observed overlap was consistent with the estimated overlap as obtained from the Chao1 analyses, indicating that sampling errors were likely negligible (**Fig 14**).

While the clonotypic composition of GLC- and YVL-specific CD8 T cells changed over the course of primary infection (**Figs 15 & 16**), dominant TCR clonotypes detected during AIM tended to persist into CONV. For example, the YVL-specific clonotype, CAVKDTDKLIF, was present in AIM in E1603, E1632 and E1655 at a frequency of 21.9%, 10.1% and 5.9%, respectively. In CONV, this frequency was somewhat constant (22.5% in E1603 and 9.7% in E1632) except in E1655 where it substantially decreased to 0.4%. Altogether, these data indicate that persistent clonotypes made up only a small percentage of unique clonotypes but were highly expanded in AIM and CONV. Surprisingly, the vast majority (≥91%) of unique clonotypes completely disappeared following AIM and were replaced with *de novo* clonotypes in CONV.

5 Features associated with YVL-specific persistent clonotypes

The TCR repertoires of persistent and non-persistent clonotypes in AIM were examined in order to identify factors that potentially governed TCR persistence. Persistent YVL TCR clonotypes maintained expression of the features that were identified and described earlier (**Figs 17-20**). Although some features were shared by all 3 TCR subsets, there were clear structural differences in these repertoires. Specifically, persistent clonotypes used significantly more of the shorter 9-mer CDR3α and more of 10-, 11- and 12-mer CDR3β than the non-persistent. In contrast, the *de novo* clonotypes favored 12-mer CDR3α and 11-mer CDR3β length (**Fig 17**).

The YVL non-persistent CDR3α clonotypes used AV8.1 but it was paired with many more Jα genes **(Fig 18A**). Moreover, AV8.1-VKDTDK-AJ34 clonotypes, which were present in 42±20% or 19±11% of all persistent clonotypes during AIM or CONV, respectively, were present in the non-persistent response at a much lower mean frequency (6±1%; **Fig 20; Table 5B)**. The clonal composition of the CDR3β non-persistent response varied greatly in Vβ usage between donors and

lacked clear motifs, suggesting that for YVL clones expressing AV8.1-VKDTDK-AJ34 to persist, there may be some TCRβ characteristics that make them better fit.

De novo clonotypes appeared to exhibit slightly different properties. For instance, in the YVL 9-mer *de novo* clonotypes, use of AV8.1-AJ34 was maintained in 2/3 donors and a variant motif, "VKNTDK", emerged (**Figs 18A & 20**). The *de novo* 11-mer CDR3α response had increased use of AV12 in all 3 donors (**Fig 18A**). In *de novo* BV clonotypes, the pattern of Vβ-Jβ gene usage changed compared to that observed in AIM. Similarly, *de novo* 13-mer CDR3β clonotypes were also different with use of a new motif, "SALLGX", in 2/3 donors **(Fig 20; Table 5C)**.

6 Features associated with GLC-specific persistent clonotypes

Within the GLC TCR repertoire, significant changes in CDR3 length (**Fig 17**) were observed. The persistent clonotypes preferentially used 9- and 11-mer CDR3α while *de novo* preferred longer 12- and 14-mer CDR3α. The persistent clonotypes also preferentially used 11- and 13-mer CDR3β, while *de novo* preferred 12-mer CDR3β. The persistent GLC TCRα clonotypes maintained the features that were identified and described earlier (**Figs 17-20**) with the 9-mer "EDNNA" motif, which strongly associated with AV5-1-AJ31, being present at a mean of 5±3.7% or 10±8.6% at AIM or CONV, respectively (**Fig 20; Table 6A)**. The fact that clonotypes using this motif were not present in non-persistent clonotypes suggests that this motif and not just the gene family may be important in determining persistence of GLC-specific clonotypes. Additionally, persistent GLC-clonotypes expressed the 11-mer "SARD" motif, which associated with BV20.1-BJ1. This motif was present at a mean of 16±9.9% and 24±13.7% of all persistent clonotypes at AIM and CONV, respectively. Two of the donors had the 11-mer "SQSPGG" motif at a mean of 40±8% and 30±25% of all persistent clonotypes at AIM and CONV, respectively. Only the "SARD" motif clonotypes appeared in non-

persistent TCRβ clonotypes at AIM but at a low mean frequency of 3±1% (**Fig 20; Table 6B)**. The *de novo* clonotypes appeared to slightly vary from persistent ones. Although there was much greater diversity and more variation between donors in *de novo* clonotypes (each donor was private) with recruitment of Vα genes such as AV41 or AV24 in E1632 and E1655, there was still a preferential usage by 2/3 donors of AV5.1 (**Fig 18)** and the appearance in 2/3 donors of a new 11-mer CDR3α motif "ELDGQ", which associated with AV5.1-AJ16.1 (**Fig 20; Table 6C)**. They also used common Vβ families such as BV20 and expressed the "SARD" motif at a mean of 5%±2.9 (**Figs 18B & 20B; Table 6C)**.

7 Conclusion

We characterized the TCRα/β repertoires of EBV-specific CD8 T cells isolated over the course of acute EBV infection from three individuals using bulk deep sequencing and examined potential features of TCR repertoires that may drive selection of T cells into the memory phase. We observed that these repertoires evolved. Persistent EBV-specific clonotypes accounted for only 9% of unique clonotypes but were highly expanded in acute EBV infection. They tended to express particular CDR3 motifs more commonly than non-persistent clonotypes, suggesting that these motifs may play a role in T-cell persistence into convalescence. The other 91% of unique clonotypes disappeared and were replaced in convalescence by "de-novo" clonotypes.
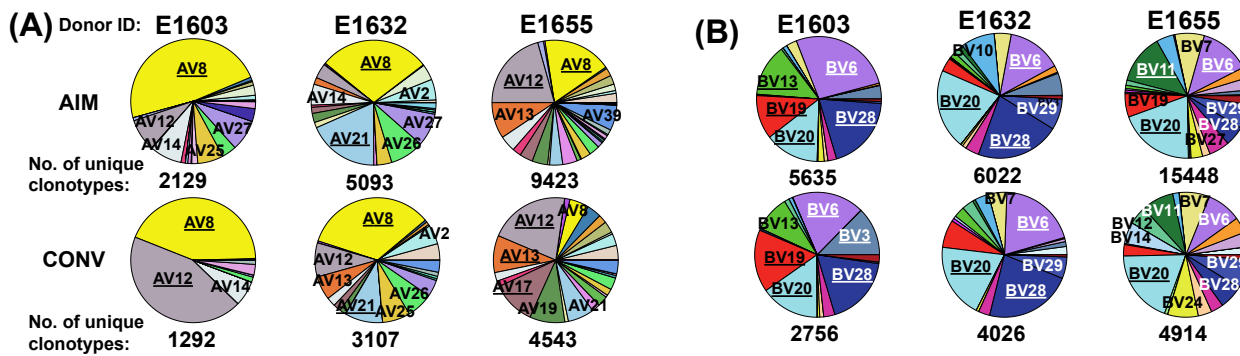
Fig 6. Preferential use of AV8, AV12, BV6, BV20, and BV28 by YVL-specific CD8 T cells in AIM and CONV. YVL-specific TCRα (A) and TCRβ (B) repertoires were analyzed for 3 AIM donors (E1603, E1632, E1655) during the acute (within two weeks of onset of symptoms; primary response; AIM) and convalescent (6 months later; memory response; CONV) phase of EBV infection by deep sequencing. Frequency of each TRAV (A) and TRBV (B) in total YVL-specific TCR-repertoire is shown in pie charts. The pie plots are labeled with gene families having a frequency ≥10% (dominant, underlined) or between 5% and 10% (subdominant; not underlined). The total numbers of unique clonotypes in each donor is shown below the pie charts. The pie plots are color-coded by V genes and the color scheme is consistent throughout this dissertation. The analyses take into consideration the magnitude of the TCR sequences (*i.e.*, number of sequence reads).
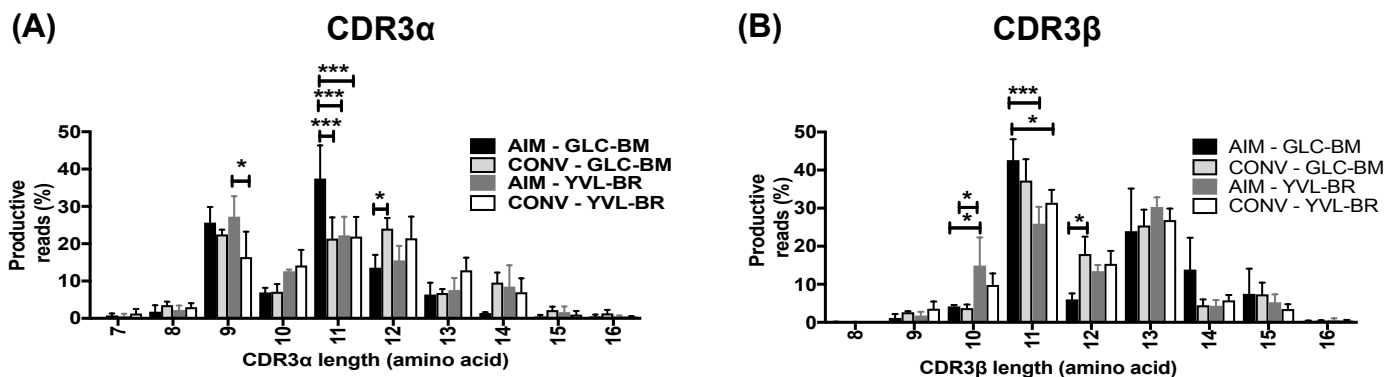
Fig 7. CDR3 α and β length distribution of YVL- and GLC-specific TCR repertoires in AIM and CONV. The CDR3 α (A) and β (B) amino acid length distribution of TCR repertoires obtained by deep sequencing of tetramer-sorted cells isolated from 3 EBV-infected donors during AIM and CONV. The conserved cysteine 104 and phenylalanine 118 in the rearranged CDR3 are not included when determining length. The analyses take into consideration the magnitude of the TCR sequences (*i.e.*, number of sequence reads). Data was analyzed by two-way ANOVA multivariant analysis with correction for multiple comparisons. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$. Error bars are SEM.
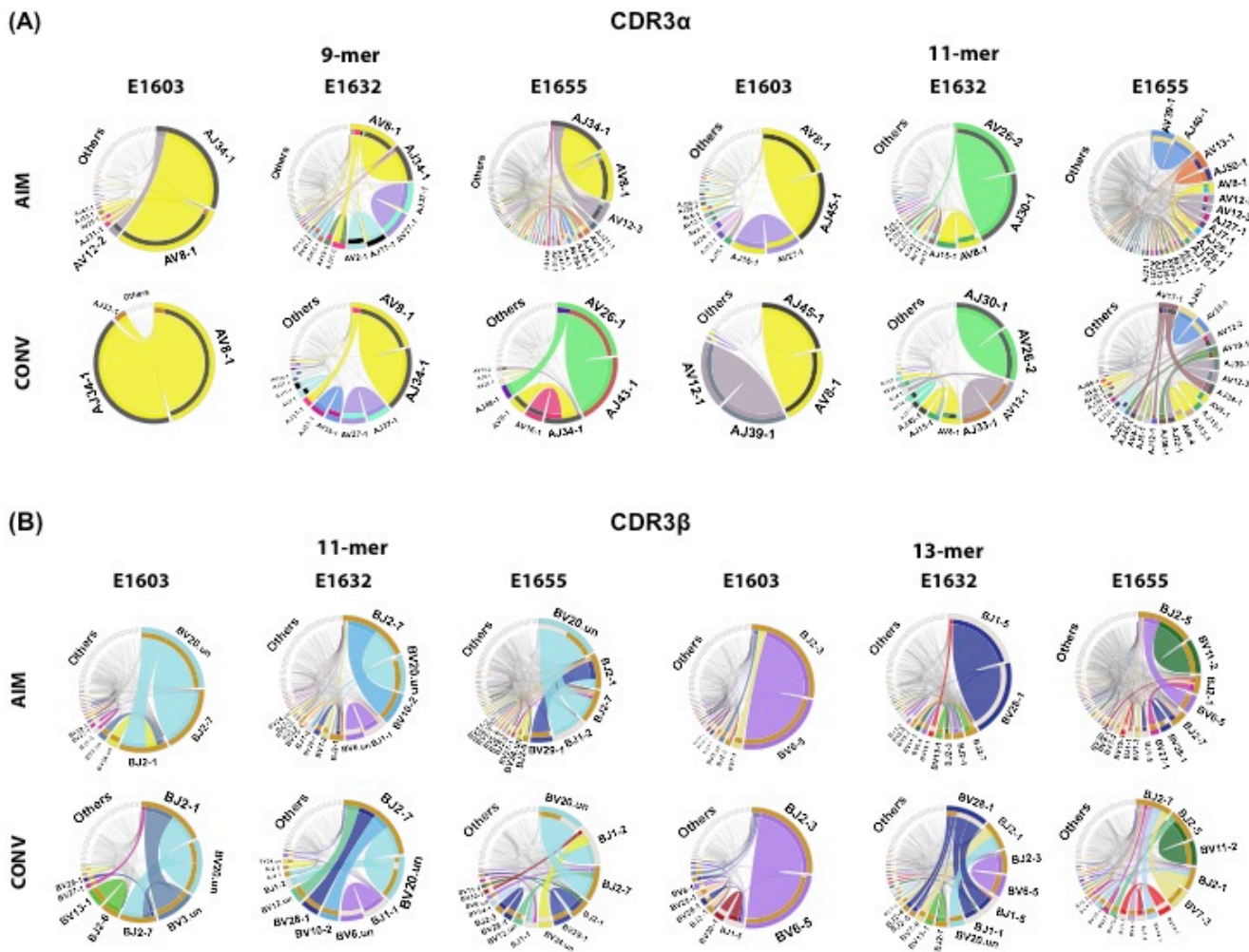
Fig 8. Conservation of 9-mer AV8.1-AJ34.1- and 11-mer BV20-BJ2.7-expressing clonotypes within YVL TCR repertoires in AIM and CONV. YVL-specific TCRα (A) and TCRβ (B) repertoires were analyzed for 3 AIM donors (E1603, E1632, E1655) during the acute (within two weeks of onset of symptoms; primary response; AIM) and convalescent (6 months later; memory response; CONV) phase of EBV infection by deep sequencing. Circos plots depicting V-J gene pairing of clonotypes within the two most dominant CDR3 α (A) and β (B) lengths are presented for AIM and CONV. The frequency of each V or J is represented by its arc length and that of the V-J pairing by the width of the arc. The arches of the circos plots are color-coded by V genes and the color scheme is consistent throughout this dissertation. The analyses take into consideration the magnitude of the TCR sequences (*i.e.*, number of sequence reads). ".un" denotes V families where the exact gene names were unknown.

Fig 9. CDR3 motif analyses of YVL-specific CD8 T cells in AIM and CONV. YVL-specific TCRα (A) and TCRβ (B) repertoires were analyzed for 3 AIM donors (E1603, E1632, E1655) during the acute (within two weeks of onset of symptoms; primary response; AIM) and convalescent (6 months later; memory response; CONV) phase of EBV infection by deep sequencing. Sequence logos for the two most dominant CDR3 α (A) and β (B) lengths are shown for AIM (top) and CONV (bottom). The analyses take into consideration the magnitude of the TCR sequences (*i.e.*, number of sequence reads). n: number of unique clonotypes.
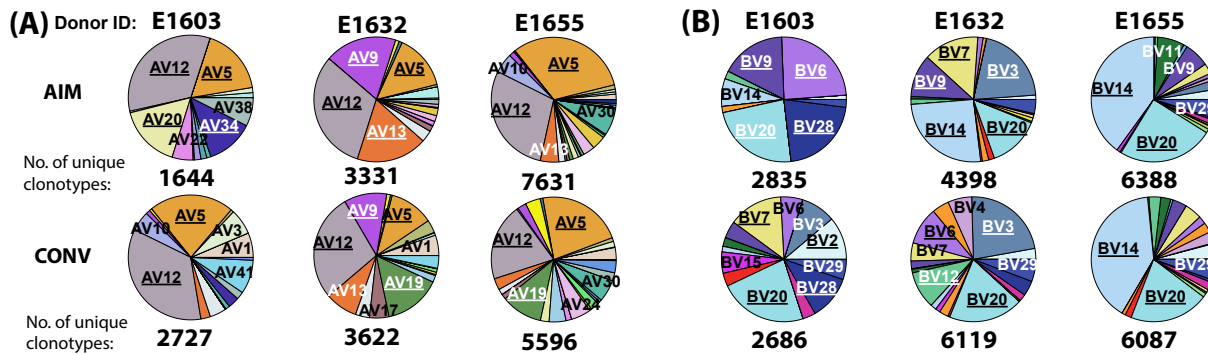
Fig 10. Preferential use of AV5, AV12 and BV20 by GLC-specific CD8 T cells in AIM and CONV.
GLC-specific TRAV (A) and TRBV (B) repertoires were analyzed for 3 AIM donors (E1603, E1632, E1655) during the acute (within two weeks of onset of symptoms; primary response; AIM) and convalescent (6 months later; memory response; CONV) phase of EBV infection by deep sequencing. Frequency of each TRAV (A) and TRBV (B) in total GLC-specific TCR-repertoire is shown in pie plots. The pie plots are labeled with gene families having a frequency ≥10% (dominant, underlined) or between 5% and 10% (subdominant; not underlined). The total numbers of unique clonotypes in each donor is shown below the pie plots. The pie plots are color-coded by V genes and the color scheme is consistent throughout this manuscript. The analyses take into consideration the magnitude of the TCR sequences (*i.e.,* number of sequence reads).
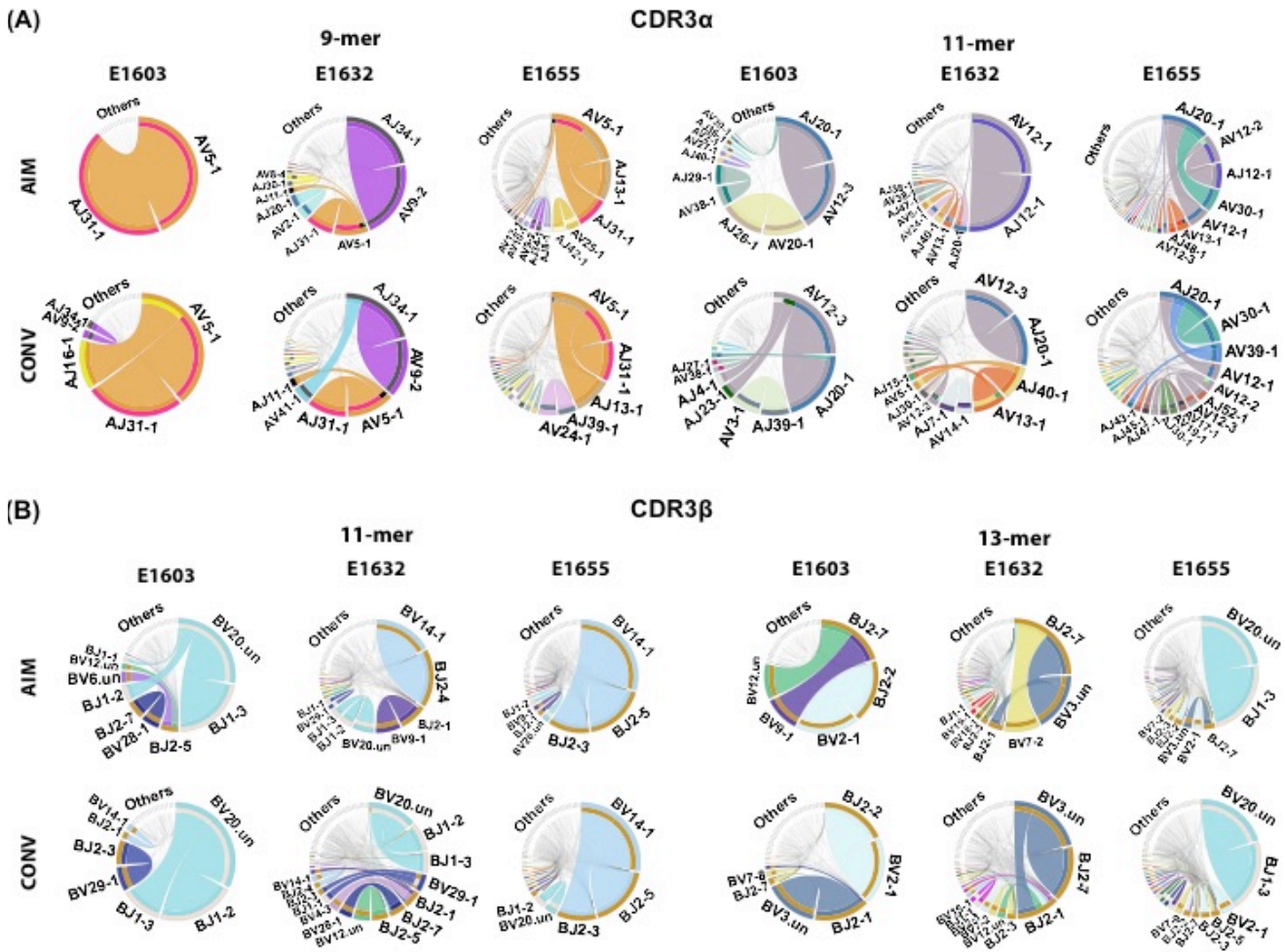
Fig 11. Conservation of 9-mer AV5-AJ31- and 11-mers BV20-BJ1- and BV14-BJ2-expressing clonotypes within GLC TCR repertoires in AIM and CONV. GLC-specific TCRα (A) and TCRβ (B) repertoires were analyzed for 3 AIM donors (E1603, E1632, E1655) during the acute (within two weeks of onset of symptoms; primary response; AIM) and convalescent (6 months later; memory response; CONV) phase of EBV infection by deep sequencing. Circos plots depicting V-J gene pairing of clonotypes within the two most dominant CDR3 α (A) and β (B) lengths are presented for AIM and CONV. The frequency of each V or J is represented by its arc length and that of the V-J pairing by the width of the arc. The arches of the circos plots are color-coded by V genes and the color scheme is consistent throughout this dissertation. The analyses take into consideration the magnitude of the TCR sequences (*i.e.*, number of sequence reads). ".un" denotes V families where the exact gene names were unknown.

(A) CDR3α

9-mer **11-mer**

Vα Jα Vα Jα

AIM

n = 3,841 n = 2,777

CONV

n = 2,956 n = 1,830

(B) CDR3β

11-mer **13-mer**

Vβ Jβ Vβ Jβ

AIM

n = 4,337 n = 2,863

CONV

n = 3,662 n = 2,900

Fig 12. CDR3 motif analyses of GLC-specific CD8 T cells in AIM and CONV. GLC-specific TCRα (A) and TCRβ (B) repertoires were analyzed for 3 AIM donors (E1603, E1632, E1655) during the acute (within two weeks of onset of symptoms; primary response; AIM) and convalescent (6 months later; memory response; CONV) phase of EBV infection by deep sequencing. Sequence logos for the two most dominant CDR3 α (A) and β (B) lengths are shown for AIM (top) and CONV (bottom). The analyses take into consideration the magnitude of the TCR sequences (*i.e.*, number of sequence reads). n: number of unique clonotypes.
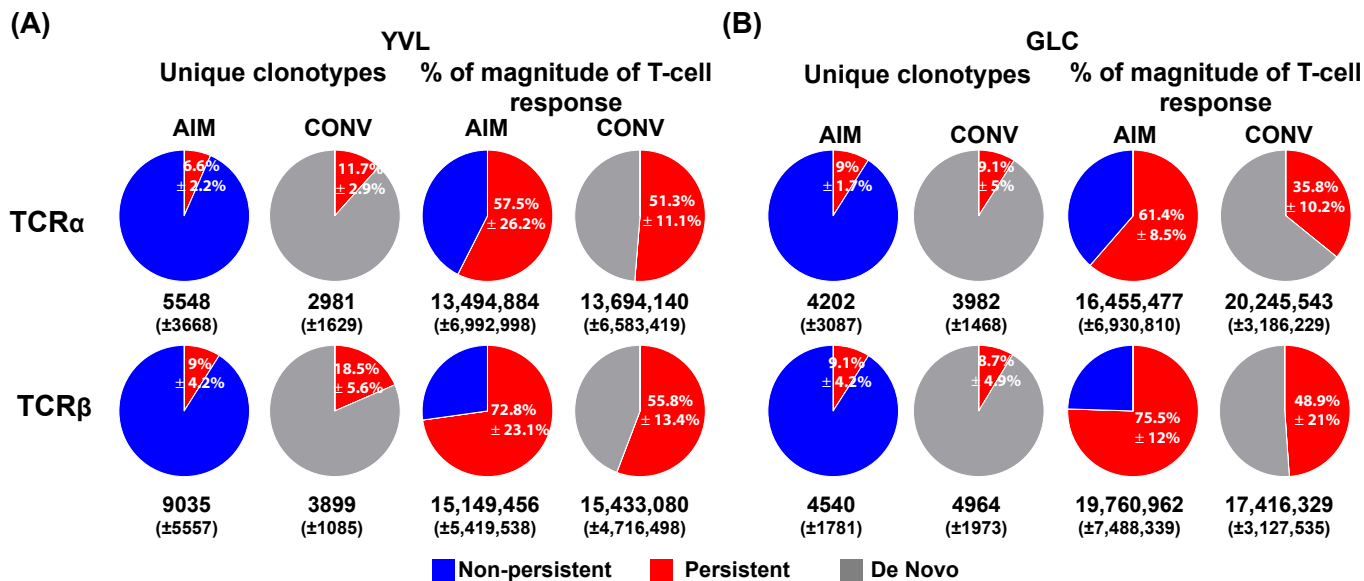
Fig 13. Persistent dominant clonotypes represent a small fraction of unique clonotypes, with TCR α and β repertoire richness maintained by the development of *de novo* clonotypes. Clonotypes that persist from the acute phase into memory represent only 6-18% of the unique clonotypes, but contribute 35-75% of the magnitude of YVL (A) or GLC (B) T-cell response. The highly diverse non-persistent clonotypes are replaced by new (*de novo*) highly diverse clonotypes, which were not present in the acute response. The average frequency (±SEM) of unique clonotypes that persist into the memory phase is shown for each epitope in the pie plots labelled unique clonotypes. The average number (±SEM) of total unique clonotypes from the 3 donors are shown below these pie plots. Also shown in pie plots is the percentage these unique clonotypes contribute to the magnitude (sequence reads) of the T-cell response for each epitope. The average numbers (±SEM) of sequence reads is shown below the pie plots labelled % of magnitude of T-cell response. % of magnitude of T-cell response takes into account the abundance of sequence reads.
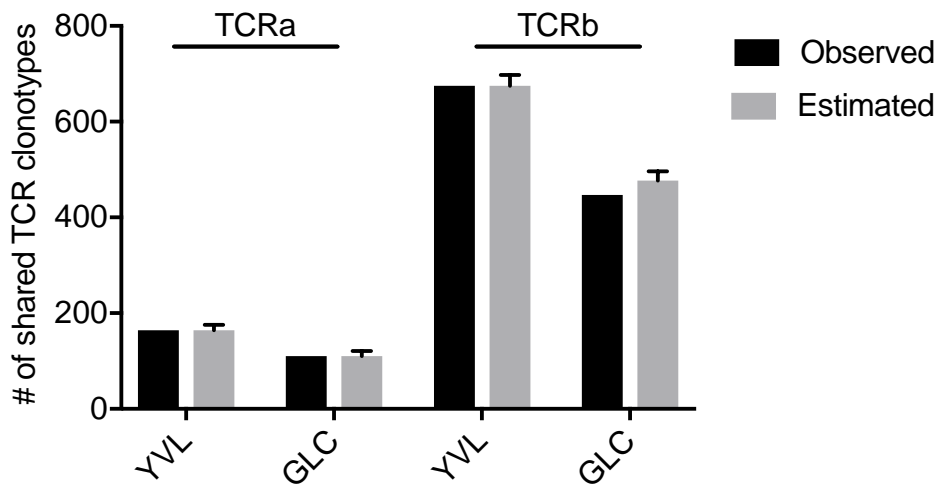
Fig 14. Comparison between estimated and observed number of TCR clonotypes shared between AIM and CONV. YVL- and GLC-specific TCRα/β repertoires were analyzed during the acute (within two weeks of onset of symptoms; primary response; AIM) and convalescent (6 months later; memory response; CONV) phase of EBV infection by deep sequencing. The plot displays the observed total number of TCRα or TCRβ clonotypes shared between AIM and CONV for each epitope and compared it to the estimated shared clonotype counts derived from the non-parametric Chao1 statistical analysis. Chao1 uses bootstrapping to predict the degree of clonotypes sharing between different TCR repertoires. Error bars represent 95% confidence intervals. Calculations were performed using EstimateS from TCRα or TCRβ nucleotide sequences specific for YVL or GLC from donor E1603. We wanted to perform the analyses by pooling sequences from all donors; however, the online tool EstimateS could not handle the large dataset derived from pooling and so we decided to perform it using a donor.
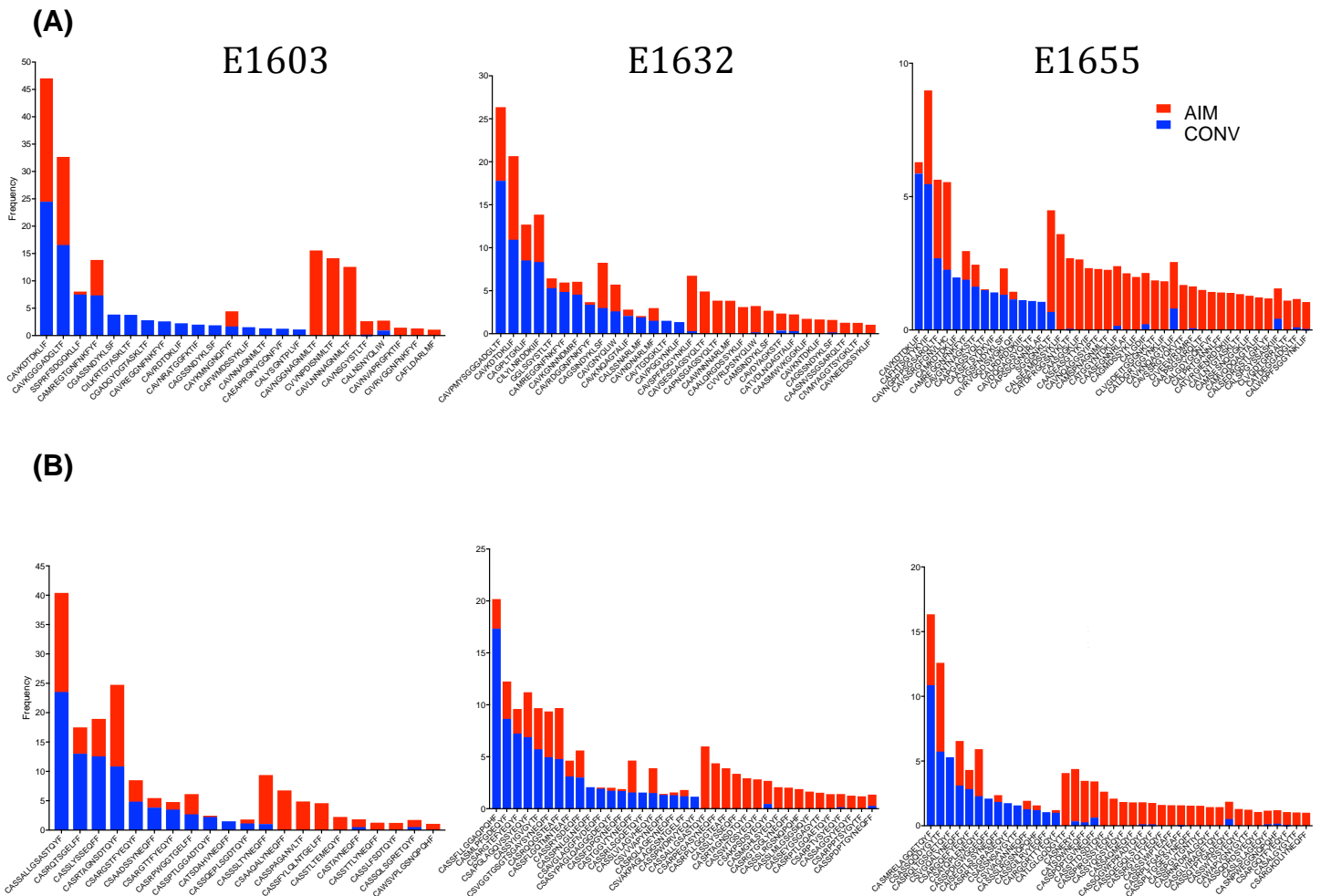
Fig 15. Clonotype distribution of YVL-specific CD8 T cells in AIM and CONV. YVL-specific TCRα (A) and TCRβ (B) repertoires were analyzed for 3 AIM donors (E1603, E1632, E1655) during the acute (within two weeks of onset of symptoms; primary response; AIM) and convalescent (6 months later; memory response; CONV) phase of EBV infection by deep sequencing. Frequency of each TCRα clonotype (A) and TCRβ clonotype (B) in total YVL-specific TCR-repertoire during AIM and CONV is shown in stacked histograms. The plots display clonotypes having a frequency ≥1%. The analyses take into consideration the magnitude of the TCR sequences (*i.e.*, number of sequence reads).
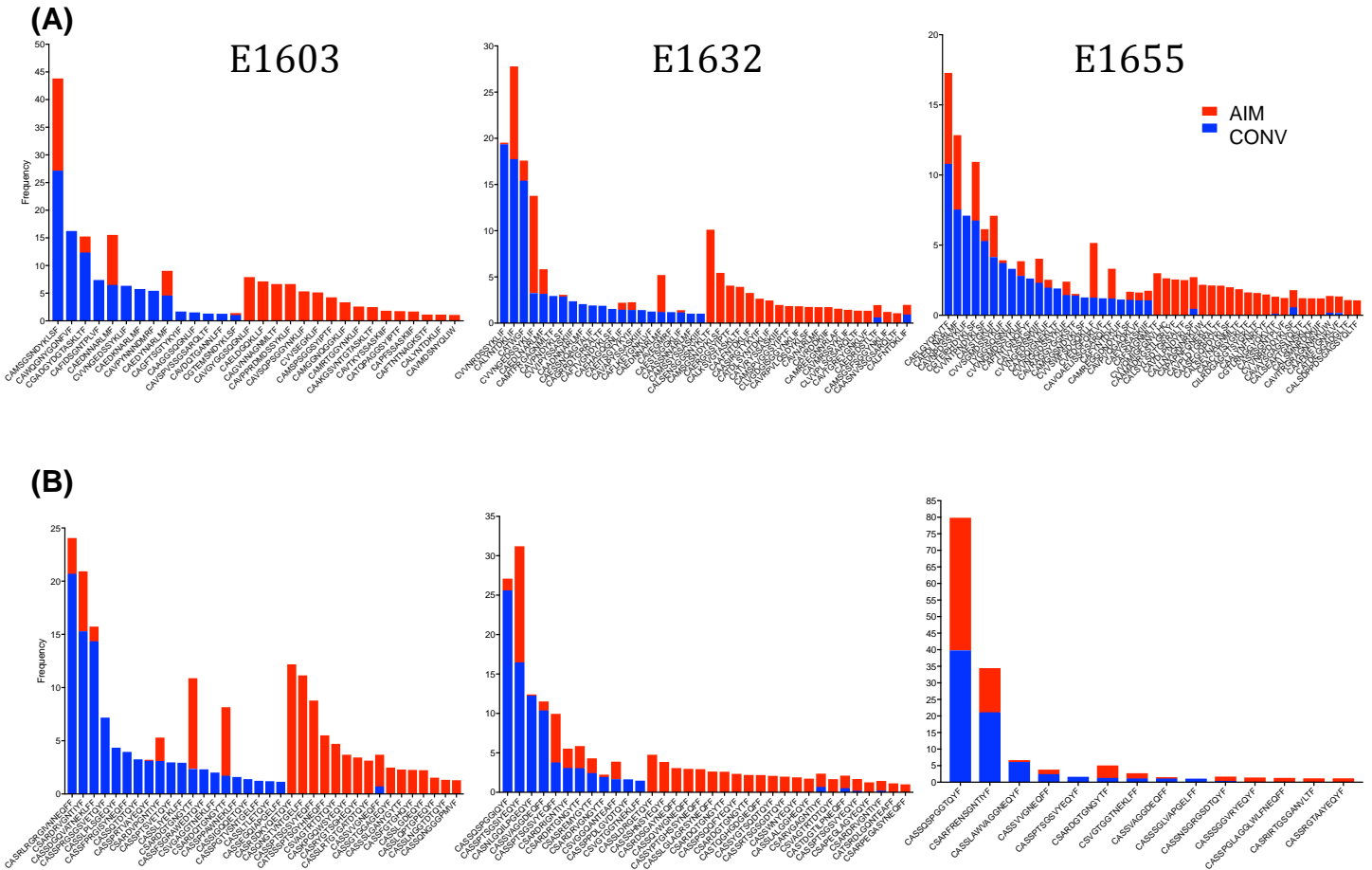
Fig 16. Clonotype distribution of GLC-specific CD8 T cells in AIM and CONV. GLC-specific TCRα (A) and TCRβ (B) repertoires were analyzed for 3 AIM donors (E1603, E1632, E1655) during the acute (within two weeks of onset of symptoms; primary response; AIM) and convalescent (6 months later; memory response; CONV) phase of EBV infection by deep sequencing. Frequency of each TCRα clonotype (A) and TCRβ clonotype (B) in total GLC-specific TCR-repertoire during AIM and CONV is shown in stacked histograms. The plots display clonotypes having a frequency ≥1%. The analyses take into consideration the magnitude of the TCR sequences (*i.e.*, number of sequence reads).

(A)

(B)
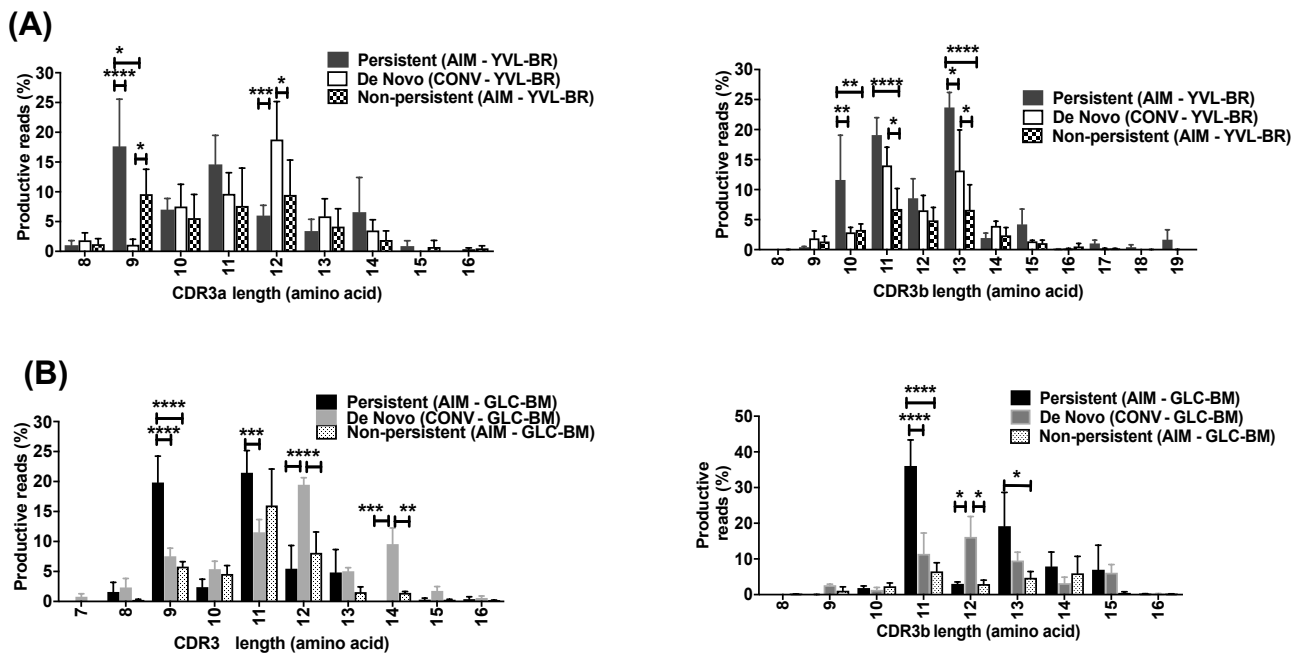
Fig 17. CDR3 α and β length distribution of YVL- and GLC-specific persistent, *de novo* and non-persistent clonotypes. The CDR3 α (left) and β (right) amino acid length distribution of TCR repertoires obtained by deep sequencing of YVL- (A) and GLC- (B) specific cells isolated from 3 EBV-infected donors during AIM and CONV. TCR clonotypes of each donor were assorted into 3 groups: those that persist from AIM into CONV (persistent), those that do not persist (non-persistent) and those that emerge in CONV (*de novo*) clonotypes. The conserved cysteine 104 and phenylalanine 118 in the rearranged CDR3 are not included when determining length. The analyses take into consideration the magnitude of the TCR sequences (*i.e.*, number of sequence reads). Data was analyzed by two-way ANOVA multivariant analysis with correction for multiple comparisons. * $p<0.05$, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$. Error bars are SEM.

Fig 18A. Circos plot analyses of persistent, non-persistent and *de novo* YVL-specific TCRα clonotypes. YVL-specific TCRα repertoires were analyzed for 3 AIM donors (E1603, E1632, E1655) during the acute (within two weeks of onset of symptoms; primary response; AIM) and convalescent (6 months later; memory response; CONV) phase of EBV infection by deep sequencing. Circos plots depicting Vα-Jα gene pairing of clonotypes within the two most dominant CDR3α lengths. The frequency of each V or J is represented by its arc length and that of the V-J pairing by the width of the arc. The arches of the circos plots are color-coded by V genes and the color scheme is consistent throughout this dissertation. The analyses take into consideration the magnitude of the TCR sequences (*i.e.*, number of sequence reads). ".un" denotes V families where the exact gene names were unknown.

Fig 18B. Circos plot analyses of persistent, non-persistent and *de novo* YVL-specific TCRβ clonotypes. YVL-specific TCRβ repertoires were analyzed for 3 AIM donors (E1603, E1632, E1655) during the acute (within two weeks of onset of symptoms; primary response; AIM) and convalescent (6 months later; memory response; CONV) phase of EBV infection by deep sequencing. Circos plots depicting Vβ-Jβ gene pairing of clonotypes within the two most dominant CDR3β lengths. The frequency of each V or J is represented by its arc length and that of the V-J pairing by the width of the arc. The arches of the circos plots are color-coded by V genes and the color scheme is consistent throughout this dissertation. The analyses take into consideration the magnitude of the TCR sequences (*i.e.*, number of sequence reads). ".un" denotes V families where the exact gene names were unknown.

Fig 19A. Circos plot analyses of persistent, non-persistent and *de novo* GLC-specific TCRα clonotypes. GLC-specific TCRα repertoires were analyzed for 3 AIM donors (E1603, E1632, E1655) during the acute (within two weeks of onset of symptoms; primary response; AIM) and convalescent (6 months later; memory response; CONV) phase of EBV infection by deep sequencing. Circos plots depicting Vα-Jα gene pairing of clonotypes within the two most dominant CDR3α lengths. The frequency of each V or J is represented by its arc length and that of the V-J pairing by the width of the arc. The arches of the circos plots are color-coded by V genes and the color scheme is consistent throughout this dissertation. The analyses take into consideration the magnitude of the TCR sequences (*i.e.*, number of sequence reads). ".un" denotes V families where the exact gene names were unknown.

Fig 19B. Circos plot analyses of persistent, non-persistent and *de novo* GLC-specific TCRβ clonotypes. GLC-specific TCRβ repertoires were analyzed for 3 AIM donors (E1603, E1632, E1655) during the acute (within two weeks of onset of symptoms; primary response; AIM) and convalescent (6 months later; memory response; CONV) phase of EBV infection by deep sequencing. Circos plots depicting Vβ-Jβ gene pairing of clonotypes within the two most dominant CDR3β lengths. The frequency of each V or J is represented by its arc length and that of the V-J pairing by the width of the arc. The arches of the circos plots are color-coded by V genes and the color scheme is consistent throughout this dissertation. The analyses take into consideration the magnitude of the TCR sequences (*i.e.*, number of sequence reads). ".un" denotes V families where the exact gene names were unknown.

Fig 20. CDR3 motif analyses for persistent, non-persistent and *de novo* clonotypes for YVL- and GLC-specific CD8 T cells. YVL-specific TCRα (A) and TCRβ (B) repertoires were analyzed for 3 AIM donors (E1603, E1632, E1655) during the acute (within two weeks of onset of symptoms; primary response; AIM) and convalescent (6 months later; memory response; CONV) phase of EBV infection by deep sequencing. Sequence logos for the two most dominant CDR3 α and β lengths of YVL- (A) and GLC- (B) specific persistent, non-persistent and *de novo* clontypes. The analyses take into consideration the magnitude of the TCR sequences (*i.e.,* number of sequence reads). n: number of unique clonotypes.
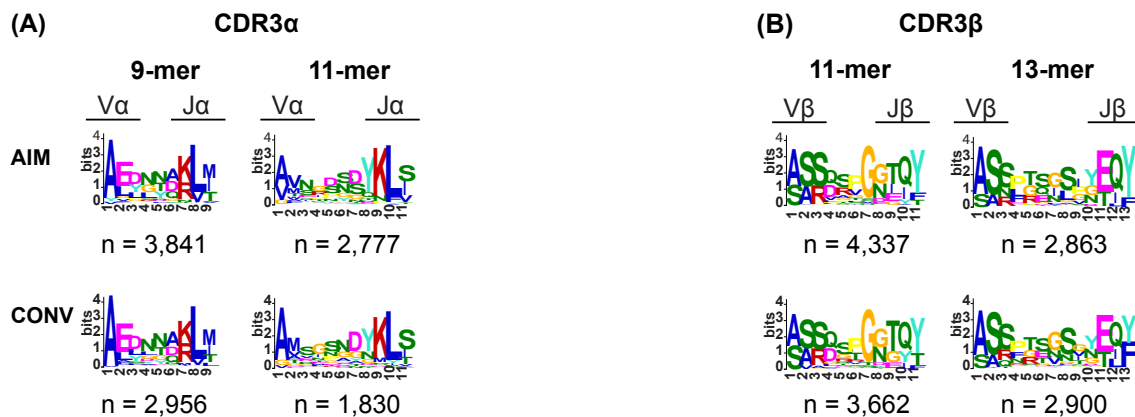
Table 1. Characteristics of study population.

Donor ID[1]	Gender	Time point	Time (months)[2]	EBV viral load (log$_{10}$ copies/million B cells)	GLC[3] (%)	YVL[3] (%)
E1603	M	AIM	0	2.03	0.6	2.9
		CONV	6	1.91	0.5	0.7
E1632	F	AIM	0	4.36	1.1	2
		CONV	7	No data[4]	0.2	0.1
E1655	F	AIM	0	5.05	1.6	1.3
		CONV	5	3.3	0.2	0.1
E1651	F	AIM	0	4.14	1.7	4.5
		CONV	8	Not detected	0.1	0.1

[1]Single-cell paired TCR α β sequencing was performed on tetramer sorted CD8 T cells of all four donors at presentation with AIM and 5-8 months later. However, TCR deep sequencing was performed in all donors except E1651 due to a limited number of PBMCs.
[2]Time elapsed between AIM and CONV.
[3]Frequency of HLA-A2 restricted GLC or YVL tetramer+ cells within CD3+ CD8+ T cells in PBMCs isolated from each respective donor.
[4]B cells were not available from this donor to perform a viral load assay.
AIM: acute infectious mononucleosis; CONV: convalescence; M: male; F: Female.

Table 2. TCR Sequencing depth and counts of productive DNA rearrangements by donor, epitope-specificity and time point.

Donor ID	Epitope-specificity	Time Point	Locus	Total reads[1]	Productive reads[2]	Fraction productive reads	Total DNA rearrangements[3]	Productive DNA rearrangements	Frequency of the most abundant DNA rearrangements
E1603	GLC	AIM	α	10,054,626	9,292,966	0.92	2,616	1,644	27.10
E1603	GLC	CONV	α	22,973,724	17,423,275	0.76	5,092	2,727	16.65
E1632	GLC	AIM	α	22,097,431	16,944,815	0.77	7,168	3,331	17.04
E1632	GLC	CONV	α	22,859,210	19,612,623	0.86	7,454	3,622	10.11
E1655	GLC	AIM	α	28,202,652	23,128,651	0.82	16,303	7,631	10.74
E1655	GLC	CONV	α	28,962,599	23,700,732	0.82	10,620	5,596	6.46
E1603	YVL	AIM	α	7,910,653	5,660,732	0.72	4,756	2,129	17.19
E1603	YVL	CONV	α	9,104,849	6,329,264	0.70	2,679	1,292	19.75
E1632	YVL	AIM	α	23,535,422	15,717,394	0.67	10,826	5,093	17.76
E1632	YVL	CONV	α	21,440,524	15,745,519	0.73	6,880	3,107	8.55
E1655	YVL	AIM	α	23,058,186	19,106,529	0.83	19,844	9,423	5.46
E1655	YVL	CONV	α	24,136,082	19,007,637	0.79	9,792	4,543	3.76
E1603	GLC	AIM	β	27,465,189	26,612,904	0.97	4,440	2,835	20.66
E1603	GLC	CONV	β	18,100,236	17,244,686	0.95	4,490	2,686	12.13
E1632	GLC	AIM	β	21,780,262	20,902,638	0.96	6,777	4,398	25.50
E1632	GLC	CONV	β	15,288,247	14,378,151	0.94	9,054	6,119	14.39
E1655	GLC	AIM	β	13,121,531	11,767,345	0.90	11,862	6,388	21.52
E1655	GLC	CONV	β	22,723,664	20,626,151	0.91	10,152	6,087	22.16
E1603	YVL	AIM	β	23,441,696	21,388,055	0.91	9,171	5,635	22.82
E1603	YVL	CONV	β	22,613,004	20,588,768	0.91	4,928	2,756	15.94
E1632	YVL	AIM	β	14,253,678	11,604,360	0.81	10,671	6,022	17.28
E1632	YVL	CONV	β	15,662,887	14,374,943	0.92	6,846	4,026	5.52
E1655	YVL	AIM	β	12,882,103	12,455,953	0.97	23,621	15,448	10.39
E1655	YVL	CONV	β	11,684,624	11,335,528	0.97	8,007	4,914	6.85

[1]Total reads are the sum of productive and non-productive reads.
[2]Productive reads are reads that do not contain premature stop codons.
[3]Total DNA rearrangements are the total number of unique reads (productive and non-productive).

Table 3. CDR3α (A) and CDR3β (B) motifs for YVL-specific CD8 T cells during AIM and CONV.

(A)

T cell receptor:	Valpha											
Motif sequence:	"VKDTDK" (9-mer)											
Sample name:	E1603				E1632				E1655			
Phase of infection:	AIM		CONV		AIM		CONV		AIM		CONV	
Frequency of all copies (reads):	24,48%(1,385,735 / 5,660,732)		22.5 % (1,426,914 / 6,329,264)		10.9% (1,718,846 / 15,717,394)		9.7% (1,530,475 / 15,745,519)		5.86% (120,799 / 19,106,529)		0.43% (81,323 / 19,007,637)	
Frequency of all unique clonotypes:	4.1% (195/4,756)		6.8% (182/2,679)		2.67% (290 /10,826)		2.74% (189/6,880)		1.1% (221/19844)		0.56% (55/9792)	
Family usage (%):	TRAV	TRAJ	TRAV	TRAJ	TRAV	TRAJ	TRAV	TRAJ	TRAV	TRAJ	TRAV	TRAJ
(% of usage of all other families detected for this sample)	AV8 (85.13%)	AJ34 (94.87%)	AV8 (98.9%)	AJ34 (99.45%)	AV8 (85.52%)	AJ34 (91.04%)	AV8 (96.29%)	AJ34 (91%)	AV8 (90.5%)	AJ34 (90.95%)	AV8 (65.45%)	AJ34 (87.27%)
	AV12 (11.28%)	AJ30 (2.06%)	AV35 (0.55%)	AJ45 (0.55%)	AV12 (8.97%)	AJ32 (1.72%)	AV39 (1.06%)	AJ30 (3.17%)	AV12 (3.62%)	AJ10 (1.8%)	AV12 (23.64%)	AJ50 (7.26%)
	AV3 (1.54%)	AJ32 (1.54%)	AV12 (0.55%)		AV21: 1.724%	AJ46 (1.03%)	AV2 (1.06%)	AJ46 (1.59%)	AV21 (1.36%)	AJ32 (0.9%)	AV21 (5.45%)	AJ32 (1.82%)
	AV6 (1.03%)	AJ45 (0.51%)			AV14 (1.38%)	AJ37 (1.03%)	AV25 (0.53%)	AJ32 (1.06%)	AV39 (0.9%)	AJ30 (0.9%)	AV19 (1.82%)	AJ26 (1.82%)
	AV25 (0.51%)	AJ39 (0.51%)			AV3 (1.03%)	AJ31 (1.03%)	AV14 (0.53%)	AJ59 (0.53%)	AV30 (0.9%)	AJ26 (0.9%)	AV17 (1.82%)	AJ3 (1.82%)
	AV21 (0.51%)	AJ20 (0.51%)			AV25 (0.34%)	AJ39 (0.69%)	AV3 (0.53%)	AJ45 (0.53%)	AV5 (0.9%)	AJ20 (0.9%)	AV6 (1.82%)	
					AV23 (0.34%)	AJ30 (0.69%)		AJ38 (0.53%)	AV3 (0.9%)	AJ50 (0.9%)		
					AV16 (0.34%)	AJ20 (0.69%)		AJ37 (0.53%)	AV29 (0.45%)	AJ59 (0.45%)		
					AV6 (0.34%)	AJ3 (0.69%)		AJ20 (0.53%)	AV23 (0.45%)	AJ46 (0.45%)		
						AJ59 (0.34%)		AJ3 (0.53%)	AV19 (0.45%)	AJ37 (0.45%)		
						AJ44 (0.34%)			AV13 (0.45%)	AJ36 (0.45%)		
						AJ27 (0.34%)				AJ23 (0.45%)		
						AJ14 (0.34%)				AJ3 (0.45%)		

(B)

T cell receptor:	V beta											
Motif sequence:	"LLGG" (13-mer)											
Sample name:	E1603				E1632				E1655			
Phase of infection:	AIM		CONV		AIM		CONV		AIM		CONV	
Frequency of all copies (reads):	0.003% (559/21,388,055)		0.59% (121,787 / 20,588,768)		17.37% (2,016,477/11,604,360)		6.74% (969,516/14,374,943)		0.84% (104,619/12,455,953)		0% (0/11,335,528)	
Frequency of all unique clonotypes:	0.053% (3/5,635)		1.16% (32 / 2,756)		6.4% (388/6,022)		2.73% (110/4,026)		0.36% (56/15,448)		0% (0/4,914)	
Family usage (%):	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ
(% of usage of all other families detected for this sample)	BV28 (33.33%)	BJ1 (66.67%)	BV28 (62.5%)	BJ1 (100%)	BV28 (72.4%)	BJ1 (80.92%)	BV28 (65.45%)	BJ1 (100%)	BV6 (41%)	BJ2 (73.2%)	not found	not found
	BV27 (33.33%)	BJ2 (33.33%)	BV6 (21.87%)		BV3 (19.07%)	BJ2 (19.07%)	BV6 (17.2%)		BV19 (26.78%)	BJ1 (26.8%)		
	BV6 (33.33%)		BV30 (3.1%)		BV6 (2.8%)		BV27 (2.72%)		BV28 (5.35%)			
			BV27 (3.1%)		BV27 (1.8%)		BV7 (2.72%)		BV7 (3.57%)			
			BV25 (3.1%)		BV19 (1.03%)		BV4 (2.72%)		BV3 (3.57%)			
			BV3 (6.24%)		BV7 (1.03%)		BV13 (1.8%)		BV30 (1.78%)			
					BV10 (1.02)%		BV11 (1.8%)		BV27 (1.78%)			
					BV25 (0.51%)		BV26 (0.9%)		BV25 (1.78%)			
					BV26 (0.25%)		BV25 (0.9%)		BV24 (1.78%)			
							BV19 (0.9%)		BV13 (1.78%)			
							BV12 (0.9%)		BV12 (1.78%)			
							BV10 (0.9%)		BV11 (1.78%)			
							BV3 (0.9%)		BV10 (1.78%)			
									BV5 (1.78%)			
									BV4 (1.78%)			
									BV2 (1.78%)			

Table 4. CDR3α (A) and CDR3β (B) motifs for GLC-specific CD8 T cells during AIM and CONV.

(A)

T cell receptor: V alpha

Motif sequence: "EDNNA" (9-mer)

	E1603 AIM		E1603 CONV		E1632 AIM		E1632 CONV		E1655 AIM		E1655 CONV	
Sample name / Phase:	AIM		CONV		AIM		CONV		AIM		CONV	
Frequency of all copies (reads):	6.5% (604,202/9,292,966)		9% (1,574,269/17,423,275)		1.2% (203,838/16,944,815)		4% (785,776/19,612,623)		0.6% (142,712/23,128,651)		0.8% (198,339/23,555,245)	
Frequency of all unique clonotypes:	6.4% (105/1,644)		9.2% (252/2,727)		2.2% (74/3,331)		4.8% (173/3,622)		1.4% (113/16,303)		0.7% (69/10,620)	
Family usage (%):	TRAV	TRAJ	TRAV	TRAJ	TRAV	TRAJ	TRAV	TRAJ	TRAV	TRAJ	TRAV	TRAJ
(% of usage of all other families detected for this sample)	AV5 (99.05%) AV23 (0.95%)	AJ31 (98.09%) AJ32 (1.9%)	AV5 (92.46%) AV23 (2.78%) AV14 (1.58%) AV13 (1.58%) AV19 (0.79%) AV9 (0.79%)	AJ31 (97.61%) AJ31 (1.98%) AJ34 (0.39%)	AV5 (62.5%) AV13 (19.44%) AV14 (5.55%) AV19 (5.55%) AV6 (5.55%) AV19 (1.38%)	AJ31 (98.61%) AJ13 (1.38%)	AV5 (81.81%) AV13 (5.45%) AV14 (3.03%) AV19 (2.42%) AV9 (2.42%) AV12 (1.81%) AV6 (1.81%) AV1 (1.21%)	AJ31 (96.97%) AJ13 (1.21%) AJ13 (1.21%) AJ44 (0.6%)	AV5 (92.23%) AV13 (3.88%) AV23 (1.94%) AV14 (0.97%) AV14 (0.97%)	AJ31 (98.05%) AJ13 (1.94%)	AV5 (74.62%) AV13 (10.44%) AV23 (7.46%) AV19 (4.47%) AV9 (1.49%)	AJ31 (95.52%) AJ13 (2.98%) AJ13 (1.49%)

Motif sequence: "MSGSN" (11-mer)

	E1603 AIM		E1603 CONV		E1632 AIM		E1632 CONV		E1655 AIM		E1655 CONV	
Frequency of all copies (reads):	27% (2,523,606/9,292,966)		16.7% (2,906,306/17,423,275)		0% (0/16,944,815)		5.43% (1,066,456/19,612,623)		0% (0/23,128,651)		0% (0/23,555,245)	
Frequency of all unique clonotypes:	14.2% (234/1,644)		7.73% (211/2,727)		0% (0/3,331)		2.24% (81/3,622)		0% (0/16,303)		0% (0/10,620)	
Family usage (%):	TRAV	TRAJ	AV	AJ	AV	AJ	AV	AJ	AV	AJ	AV	AJ
(% of usage of all other families detected for this sample)	AV12 (95.3%) AV06 (1.3%) AV20 (0.86%) AV23 (0.86%) AV05 (0.43%) AV10 (0.43%) AV27 (0.43%) AV34 (0.43%)	AJ20 (99.6%) AJ03 (0.4%)	AV12 (96.7%) AV10 (1.4%) AV05 (0.96%) AV23 (0.48%) AV30 (0.48%)	AJ20 (95.05%) AJ03 (0.95%) AJ04 (0.5%) AJ34 (0.5%) AJ35 (0.5%) AJ39 (0.5%)	not found	not found	AV12 (82.6%) AV14 (7.4%) AV21 (3.7%) AV01 (2.5%) AV06 (2.5%) AV05 (1.3%)	AJ20 (96.2%) AJ34 (2.5%) AJ39 (1.3%)	not found	not found	not found	not found

Motif sequence: "VNGSN" (11-mer)

	E1603 AIM		E1603 CONV		E1632 AIM		E1632 CONV		E1655 AIM		E1655 CONV	
Frequency of all copies (reads):	0.0001% (9/9,292,966)		0% (0/17,424,275)		0% (0/16,944,815)		0.00004% (8/19,612,623)		0% (0/23,128,651)		0% (0/23,555,245)	
Frequency of all unique clonotypes:	0.04% (1/1,644)		0% (0/2,727)		0% (0/3,331)		0.03% (1/3,622)		0% (0/16,303)		0% (0/10,620)	
Family usage (%):	AV	AJ	AV	AJ	AV	AJ	AV	AJ	AV	AJ	AV	AJ
(% of usage of all other families detected for this sample)	AV12 (100%)	AJ2 (100%)	not found	not found	not found	not found	AV12 (100%)	AJ2 (100%)	not found	not found	not found	not found

(B)

T cell receptor: V beta

Motif sequence: "SQSPGG" (11-mer)

	E1603 AIM		E1603 CONV		E1632 AIM		E1632 CONV		E1655 AIM		E1655 CONV	
Frequency of all copies (reads):	0% (0/26,612,904)		0% (0/17,244,686)		25.6% (5,353,658/20,902,638)		1.49% (214,503/14,378,151)		39.83% (4,687,375/11,767,345)		40.04% (8,258,795/20,626,151)	
Frequency of all unique clonotypes:	0% (0/2,835)		0% (0/2,686)		6.07% (267/4,398)		0.47% (29/6,119)		9.65% (616/6,388)		9.72% (592/6,087)	
Family usage (%):	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ
(% of usage of all other families detected for this sample)	not found	not found	not found	not found	BV14 (87.64%) BV7 (1.87%) BV3 (1.87%) BV9 (1.49%) BV6 (1.12%) BV5 (1.12%) BV27 (0.75%) BV21 (0.75%) BV12 (0.75%) BV11 (0.75%) BV2 (0.75%) BV25 (0.37%) BV18 (0.37%) BV4 (0.37%)	BJ2.04 (97.7%) BJ2.07 (1.1%) BJ2.01 (0.4%) BJ2.03 (0.4%) BJ2.05 (0.4%)	BV14: 48.276% BV6 (13.79%) BV5 (10.34%) BV28 (3.45%) BV27 (3.45%) BV18 (3.44%) BV12 (3.44%) BV11 (3.44%) BV9 (3.44%) BV7 (3.44%) BV3 (3.44%)	BJ2.04 (90.1%) BJ2.01 (3.3%) BJ2.05 (3.3%) BJ2.07 (3.3%)	BV14 (79.7%) BV7 (5.19%) BV6 (1.95%) BV9 (1.78%) BV3 (1.78%) BV4 (1.46%) BV11 (1.29%) BV27 (0.97%) BV23 (0.97%) BV21 (0.97%) BV25 (0.81%) BV12 (0.81%) BV28 (0.48%) BV2 (0.48%) BV13 (0.32%) BV5 (0.64%) BV16 (0.16%) BV10 (0.16%)	BJ2.05 (67%) BJ2.03 (32%) BJ2.01 (0.8%) BJ2.07 (0.2%)	BV14 (75.33%) BV7 (6.25%) BV6 (2.03%) BV4 (2.03%) BV9 (1.86%) BV11 (1.68%) BV27 (1.52%) BV12 (1.35%) BV23 (1.18%) BV21 (1.18%) BV25 (1.01%) BV18 (0.84%) BV5 (0.84%) BV03 (0.84%) BV2 (0.84%) BV13 (0.51%) BV28 (0.17%) BV24 (0.17%) BV10 (0.34%)	BJ2.05 (68%) BJ2.03 (31%) BJ2.01 (0.8%) BJ2.07 (0.2%)

Motif sequence: "SARD" (13-mer)

	E1603 AIM		E1603 CONV		E1632 AIM		E1632 CONV		E1655 AIM		E1655 CONV	
Frequency of all copies (reads):	23.42% (6,232,244/26,612,904)		22.79% (3,931,496/17,244,686)		8.64% (1,805,070/20,902,638)		12.15% (1,747,217/14,378,151)		2.01% (236,726/11,767,345)		4.41% (910,160/20,626,151)	
Total frequency:	15.34% (435/2,835)		10.31% (277/2,686)		4.05% (178/4,398)		2.65% (162/6,119)		2.81% (180/6,388)		1.85% (113/6,087)	
Family usage (%):	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ
(% of usage of all other families detected for this sample)	BV20 (98.85%) BV29 (1.15%)	BJ1 (99.31%) BJ2 (0.69%)	BV20 (97.83%) BV29 (2.17%)	BJ1 (98.92%) BJ2 (1.08%)	BV20 (98.31%) BV29 (1.69%)	BJ1: 97.75% BJ2: 2.25%	BV20: 99.38% BV29: 0.62%	BJ1 (90.74%) BJ2 (9.26%)	BV20 (98.89%) BV29 (1.1%)	BJ1 (99.44%) BJ2 (0.56%)	BV20 (100%)	BJ1 (97.34%) BJ2 (2.66%)

Motif sequence: "SPTSG" (13-mer)

	E1603 AIM		E1603 CONV		E1632 AIM		E1632 CONV		E1655 AIM		E1655 CONV	
Frequency of all copies (reads):	0.00006% (16/26,612,904)		9.01% (1,554,025/17,244,686)		20.4% (4,265,886/20,902,638)		22.42% (3,223,740/14,378,151)		1.92% (225,991/11,767,345)		0.2% (41,172/20,626,151)	
Total frequency:	0.18% (5/2,835)		4.2% (114/2,686)		11.18% (492/4,398)		8.75% (536/6,119)		1.72% (110/6,388)		0.74% (45/6,087)	
Family usage (%):	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ
(% of usage of all other families detected for this sample)	BV6 (100%)	BJ2.05 (100%)	BV3 (71.05%) BV6 (7.89%) BV7 (5.26%) BV9 (4.38%) BV11 (3.51%) BV28 (1.75%) BV27 (1.75%) BV14 (1.75%) BV21 (0.88%) BV19 (0.88%) BV15 (0.88%)	BJ2 (99.12%) BJ1 (0.88%)	BV3 (72.76%) BV9 (5.28%) BV7 (4.47%) BV6 (3.86%) BV12 (2.85%) BV14 (2.03%) BV27 (1.42%) BV23 (1.22%) BV21 (1.22%) BV5 (1.22%) BV11 (1.02%) BV2 (0.82%) BV4 (0.61%) BV25 (0.41%) BV19 (0.41%) BV18 (0.41%)	BJ2.07 (72.4%) BJ2.01 (25%) BJ2.04 (1.7%) BJ2.05 (0.9%)	BV3 (58.21%) BV6 (9.14%) BV7 (8.02%) BV12 (4.1%) BV4 (3.54%) BV9 (3.36%) BV11 (2.79%) BV27 (1.86%) BV18 (1.12%) BV14 (1.12%) BV19 (0.93%) BV21 (0.75%) BV2 (0.75%) BV10 (0.19%) BV23 (0.19%)	BJ2.07 (70%) BJ2.01 (26%) BJ2.03 (2.3%) BJ2.05 (2%) BJ2.04 (0.35%) BJ2.02 (0.2%)	BV3 (45.45%) BV7 (11.82%) BV11 (7.27%) BV9 (6.36%) BV6 (5.45%) BV12 (3.64%) BV21 (2.73%) BV14 (2.73%) BV4 (2.73%) BV2 (2.73%) BV28 (1.82%) BV27 (1.82%) BV25 (1.82%) BV13 (1.82%) BV15 (0.91%) BV5 (0.91%)	BJ2.07 (99.1%) BJ2.05 (0.9%)	BV3 (35.55%) BV7 (13.33%) BV11 (11.11%) TBV12 (6.67%) BV4 (6.67%) BV9 (4.44%) BV6 (4.44%) BV2 (4.44%) BV27 (2.22%) BV25 (2.22%) BV21 (2.22%) BV19 (2.22%) BV18 (2.22%) BV14 (2.22%)	BJ2.07 (100%)

Table 5. YVL-specific motifs by CDR3 length for persistent (A), non-persistent (B), and *de novo* (C) clonotypes.

(A)

T cell receptor:	Valpha											
Motif sequence:	Persistent "VKDTDK" (9-mer)											
Sample name:	E1603				E1632				E1655			
Phase of infection:	AIM		CONV		AIM		CONV		AIM		CONV	
Frequency of all copies (reads):	32.37% (1,238,327/3,824,852)		42.29% (1,426,296/3,372,293)		12.66% (1,533,671/12,114,592)		15.87% (1,529,404/9,637,298)		81.82% (328,595/401,597)		1.08% (81,139/ 7,472,186)	
Frequency of all unique clonotypes:	21.95% (36/164)		21.95% (36/164)		15.71% (71/452)		15.71% (71/452)		6.21% (25/402)		6.21% (25/402)	
Family usage (%):	TRAV	TRAJ	TRAV	TRAJ	TRAV	TRAJ	TRAV	TRAJ	TRAV	TRAJ	TRAV	TRAJ
(% of usage of all other families detected for this sample)	AV8-1 (97.22%)	AJ34-1 (94.44%)	AV8-1 (97.22%)	AJ45-1 (5.56%)	AV8-1 (95.77%)	AJ34-1 (88.73%)	AV8-1 (95.77%)	AJ34-1 (88.73%)	AV8-1 (80%)	AJ34-1 (76%)	AV8-1 (80%)	AJ34-1 (76%)
	AV12-2 (2.78%)	AJ45-1 (5.56%)	AV12-2 (2.78%)		AV25 (1.4%)	AJ32-1 (2.81%)	AV25 (1.4%)	AJ32-1 (2.81%)	AV12-2 (16%)	AJ50-1 (8%)	AV12-2 (16%)	AJ50-1 (8%)
					AV14-1 (1.4%)	AJ30-1 (5.66%)	AV14-1 (1.4%)	AJ30-1 (5.66%)	AV21-1 (4%)	AJ32-1 (4%)	AV21-1 (4%)	AJ32-1 (4%)
					AV3-1 (1.4%)	AJ59 (1.4%)	AV3-1 (1.4%)	AJ59-1 (1.4%)		AJ26-1 (4%)		AJ26-1 (4%)
						AJ37 (1.4%)		AJ37-1 (1.4%)		AJ3-1 (8%)		AJ3-1 (8%)

T cell receptor:	Vbeta											
Motif sequence:	Persistent "SALLGX" (13-mer)											
Sample name:	E1603				E1632				E1655			
Phase of infection:	AIM		CONV		AIM		CONV		AIM		CONV	
Frequency of all copies (reads):	26.71% (5,002,042 / 18,722,687)		23.65% (3,472,087/14,679,574)		not found		not found		not found		not found	
Frequency of all unique clonotypes:	15.55% (105/675)		15.55% (105/675)		not found		not found		not found		not found	
Family usage (%):	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ
(% of usage of all other families detected for this sample)	BV6-5 (86.66%)	BJ2-3 (100%)	BV6-5 (86.66%)	BJ2-3 (100%)	not found	not found	not found	not found	not found	not found	not found	not found
	BV28-1 (1.9%)		BV28-1 (1.9%)									
	BV27-1 (1.9%)		BV27-1 (1.9%)									
	BV3un (1.9%)		BV3un (1.9%)									
	BV25-1 (0.95%)		BV25-1 (0.95%)									
	BV21-1 (0.95%)		BV21-1 (0.95%)									
	BV19-1 (0.95%)		BV19-1 (0.95%)									
	BV13-1 (0.95%)		BV13-1 (0.95%)									
	BV12 (0.95%)		BV12 (0.95%)									
	BV10-1 (0.95%)		BV10-1 (0.95%)									
	BV7-3 (0.95%)		BV7-3 (0.95%)									
	BV2 (0.95%)		BV2 (0.95%)									

(B)

Motif sequence:	Non-Persistent "VKDTDK" (9-mer)											
Sample name:	E1603				E1632				E1655			
Phase of infection:	AIM		CONV		AIM		CONV		AIM		CONV	
Frequency of all copies (reads):	8.02% (147,408/1,835,880)		not applied		5.14% (185,175/3,602,802)		not applied		5.74% (792,204/13,797,945)		not applied	
Frequency of all unique clonotypes:	8.09% (159/1965)		not applied		4.71% (219/4,641)		not applied		2.17% (196/9,021)		not applied	
Family usage (%):	TRAV	TRAJ	TRAV	TRAJ	TRAV	TRAJ	TRAV	TRAJ	TRAV	TRAJ	TRAV	TRAJ
(% of usage of all other families detected for this sample)	AV8-1 (82.39%)	AJ34-1 (94.96%)	not applied		AV8-1 (82.19%)	AJ34-1 (90.84%)	not applied		AV8-1 (91.84%)	AJ34-1 (92.86%)	not applied	
	AV12-2 (13.20%)	AJ32 (1.88%)			AV12 (11.87%)	AJ46 (1.37%)			AV12 (2.04%)	AJ30 (1.02%)		
	AV3 (1.88%)	AJ30 (1.25%)			AV21-1 (2.28%)	AJ32 (1.37%)			AV39-1 (1.02%)	AJ20 (1.02%)		
	AV6 (1.25%)	AJ39 (0.62%)			AV14-1 (1.37%)	AJ31-1 (1.37%)			AV21 (1.02%)	AJ10 (1.02%)		
	AV25-1 (0.62%)	AJ20 (1.24%)			AV3 (0.91%)	AJ39-1 (0.91%)			AV5 (1.02%)	AJ59 (0.51%)		
	AV21-1 (0.62%)				AV23-1 (0.45%)	AJ37-1 (0.91%)			AV3 (1.02%)	AJ46 (0.51%)		
					AV16-1 (0.45%)	AJ20 (0.91%)			AV29 (0.51%)	AJ37 (0.51%)		
					AV6 (0.45%)	AJ3 (0.91%)			AV23 (0.51%)	AJ36 (0.51%)		
						AJ44 (0.45%)			AV19 (0.51%)	AJ32 (0.51%)		
						AJ27-1 (0.45%)			AV13 (0.51%)	AJ26-1 (0.51%)		
						AJ14-1 (0.45%)				AJ23 (1.2%)		

T cell receptor:	Vbeta											
Motif sequence:	Non Persistent "SALLGX" (13-mer)											
Sample name:	E1603				E1632				E1655			
Phase of infection:	AIM		CONV		AIM		CONV		AIM		CONV	
Frequency of all copies (reads):	1.06% (28,269/2,665,368)		not applied		not found		not applied		not found		not applied	
Frequency of all unique clonotypes:	12.27% (609/4,960)		not applied		not found		not applied		not found		not applied	
Family usage (%):	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ
(% of usage of all other families detected for this sample)	BV6 (94.74%)	BJ2 (99.84%)	not applied		not found	not found	not applied		not found	not found	not applied	
	BV7 (2.45%)	BJ1 (0.16%)										
	BV12 (0.65%)											
	BV27-1 (0.33%)											
	BV24 (0.33%)											
	BV11 (0.33%)											
	BV5 (0.33%)											
	BV2 (0.33%)											
	BV14 (0.16%)											
	BV4 (0.16%)											
	BV3 (0.16%)											

(C)

T cell receptor:	Valpha											
Motif sequence:	De novo "VKNTDK" (9-mer)											
Sample name:	E1603				E1632				E1655			
Phase of infection:	AIM		CONV		AIM		CONV		AIM		CONV	
Frequency of all copies (reads):	not applied		0.0035% (65/1,835,880)		not applied		0.00094% (34/3,602,802)		not applied		0.00025% (34/13797945)	
Frequency of all unique clonotypes:	not applied		0.4% (8/1,965)		not applied		0.24% (11/4,641)		na		0.055% (5/9,021)	
Family usage (%):	TRAV	TRAJ	TRAV	TRAJ	TRAV	TRAJ	TRAV	TRAJ	TRAV	TRAJ	TRAV	TRAJ
(% of usage of all other families detected for this sample)	not applied		AV8-1 (100%)	AJ34-1(100%)	not applied		AV8-1 (100%)	AJ34-1 (90.9%)	not applied		AV8-1 (100%)	AJ34-1(100%)
								AJ37-1 (9.1%)				

T cell receptor:		Vbeta										
Motif sequence:		De Novo "SALLGX" (13-mer)										
Sample name:		E1603				E1632				E1655		
Phase of infection:		**AIM**	**CONV**		**AIM**		**CONV**		**AIM**		**CONV**	
Frequency of all copies (reads):		not applied	**0.042%** (2,494/5,909,194)		not applied		**11.64%** (861,364/7,396,900)		not applied		not found	
Frequency of all unique clonotypes:		not applied	**11.72%** (244/2,081)		not applied		**4.3%** (145/3,368)		not applied		not found	
Family usage (%):	TRBV	TRBJ	TRBV	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ	
(% of usage of all other families detected for this sample)	not applied		*BV6 (97.95%)*	*BJ2 (99.59%)*	not applied		*BV6 (82.06%)*	*BJ2 (100%)*	not applied	not found	not found	
			BV27 (0.82%)	BJ1 (0.41%)			BV10 (3.45%)					
			BV28-1 (0.41%)				BV7-6 (2.76%)					
			BV10 (0.41%)				BV28-1 (2.07%)					
			BV9 (0.41%)				BV27-1 (1.38%)					
							BV13 (1.38%)					
							BV3 (1.38%)					
							BV25 (0.69%)					
							BV24 (0.69%)					
							BV19-1 (0.69%)					
							BV12 (0.69%)					
							BV11 (0.69%)					
							BV5 (0.69%)					
							BV2 (1.38%)					

Table 6. GLC-specific motifs by CDR3 length for persistent (A), non-persistent (B), and *de novo* (C) clonotypes.

(A)

T cell receptor:	Valpha											
Motif sequence:	Persistent "EDNNA" (9-mer)											
Sample name:	E1603				E1632				E1655			
Phase of infection:	AIM		CONV		AIM		CONV		AIM		CONV	
Frequency of all copies (reads):	12.45% (603,951/4,847,739)		27.39% (1573,133/5,742,713)		1.55% (181,290/1,676,908)		1.55% (181,290/1,676,908)		0.91% (133,331/14,630,436)		1.77% (198,068/11,163,342)	
Frequency of all unique clonotypes:	24.57% (29/118)		24.57% (29/118)		7.87% (26/330)		7.87% (26/330)		2.6% (21/810)		2.6% (21/810)	
Family usage (%):	TRAV	TRAJ	TRAV	TRAJ	TRAV	TRAJ	TRAV	TRAJ	TRAV	TRAJ	TRAV	TRAJ
(% of usage of all other families detected for this sample)	AV5-1(96.55%)	AJ31-1 (100%)	AV5-1(96.55%)	AJ31-1 (100%)	AV5-1 (61.54%)	AJ31-1 (96.15%)	AV5-1 (61.54%)	AJ31-1 (96.15%)	AV5-1 (80.95%)	AJ31-1 (95.23%)	AV5-1 (80.95%)	AJ31-1 (95.23%)
	AV23-1 (3.44%)		AV23-1 (3.44%)		AV13-1 (15.38%)	AJ13 (3.84%)	AV13-1 (15.38%)	TCRAJ13: 3.846%	AV13-1 (14.28%)	AJ13-1 (4.76%)	AV13-1 (14.28%)	AJ13-1 (4.76%)
					AV14-1 (7.69%)		AV14-1 (7.69%)		AV23-1 (4.77%)		AV23-1 (4.77%)	
					AV9-1 (7.69%)		AV9-1 (7.69%)					
					AV6 (7.69%)		AV6 (7.69%)					

T cell receptor:	Vbeta											
Motif sequence:	Persistent "SQSPGG" (11-mer)											
Sample name:	E1603				E1632				E1655			
Phase of infection:	AIM		CONV		AIM		CONV		AIM		CONV	
Frequency of all copies (reads):	not found		not found		31.22% (5,349,199/17,132,769)		4.26% (214,440/5,029,504)		47.86% (4,665,421/9,746,758)		8254039 / 15050633 = 54.841806%	
Frequency of all unique clonotypes:	not found		not found		5.35% (19/355)		5.35% (19/355)		22.95% (202/880)		202 / 880 = 22.954545%	
Family usage (%):	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ
(% of usage of all other families detected for this sample)	not found	not found	not found	not found	BV14-1 (47.36%)	BJ2-4 (100%)	BV14-1 (47.36%)	BJ2-4 (100%)	BV14-1 (57.92%)	BJ2-5 (100%)	BV14-1 (57.92%)	BJ2-5 (100%)
					BV5-1 (10.52%)		BV5-1 (10.52%)		BV7.2 (8.91%)		BV7.2 (8.91%)	
					BV27 (5.26%)		BV27 (5.26%)		BV9-1 (4.45%)		BV9-1 (4.45%)	
					BV18 (5.26%)		BV18-1 (5.26%)		BV6 (4.45%)		BV6 (4.45%)	
					BV12 (5.26%)		BV12 (5.26%)		BV4-1 (3.46%)		BV4-1 (3.46%)	
					BV11 (5.26%)		BV11 (5.26%)		BV27 (2.47%)		BV27 (2.47%)	
					BV9-1 (5.26%)		BV9-1 (5.26%)		BV25 (2.47%)		BV25 (2.47%)	
					BV7-2 (5.26%)		BV7-2 (5.26%)		BV23 (2.47%)		BV23 (2.47%)	
					BV6 (5.26%)		BV6 (5.26%)		BV21 (2.47%)		BV21 (2.47%)	
					BV3-1 (5.26%)		BV3-1 (5.26%)		BV12 (2.47%)		BV12un (2.47%)	
									BV11 (2.47%)		BV11 (2.47%)	
									BV3 29.47%)		BV3 29.47%)	
									BV2 (1.48%)		BV2 (1.48%)	
									BV13 (0.99%)		BV13 (0.99%)	
									BV28 (0.49%)		BV28 (0.49%)	
									BV10 (0.49%)		BV10 (0.49%)	

T cell receptor:	Vbeta											
Motif sequence:	Persistent "SARD" (11-mer)											
Sample name:	E1603				E1632				E1655			
Phase of infection:	AIM		CONV		AIM		CONV		AIM		CONV	
Frequency of all copies (reads):	35.04% (5,743,446/16,388,249)		51.22% (3,413,622/6,664,314)		9.86% (1,690,729/17,132,769)		15.87% (798,484/5,029,504)		2.23% (217,520/9,746,758)		909494 / 15050633 = 6.042895%	
Frequency of all unique clonotypes:	41.55% (64/154)		41.55% (64/154)		8.4% (30/355)		8.4% (30/355)		4.54% (40/880)		40 / 880 = 4.545455%	
Family usage (%):	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ
(% of usage of all other families detected for this sample)	BV20un (95.32%)	BJ1-2 and 1-3 (96.87%)	BV20un (95.32%)	BJ1-2 and 1-3 (96.87%)	BV20un (100%)	BJ01 (96.67%)	BV20un (100%)	BJ01 (96.67%)	BV20un (100%)	BJ1 (100%)	BV20un (100%)	BJ1 (100%)
	BV29-1 (4.68%)	BJ2 (3.12%)	BV29-1 (4.68%)	BJ2 (3.12%)		BJ2 (3.33%)		BJ2 (3.33%)				

(B)

T cell receptor:	Vbeta											
Motif sequence:	Non Persistent "SARD" (11-mer)											
Sample name:	E1603				E1632				E1655			
Phase of infection:	AIM		CONV		AIM		CONV		AIM		CONV	
Frequency of all copies (reads):	488798 / 10224655 = 4.780582%		not applied		114341 / 3769869 = 3.033023%		not applied		19206 / 2020587 = 0.950516%		not applied	
Frequency of all unique clonotypes:	371 / 2681 = 13.838120%		not applied		148 / 4043 = 3.660648%		not applied		140 / 5508 = 2.541757%		not applied	
Family usage (%):	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ	TRBV	TRBJ
(% of usage of all other families detected for this sample)	BV20un (99.46%)	BJ1-2 (99.73%)	not applied		BV20un (98.65%)	BJ01 (97.97%)	not applied		BV20un (99.28%)	BJ1-3 (99.28%)	not applied	
	BV29-1 (0.54%)	BJ2 (0.27%)			BV29-1 (1.35%)	BJ2 (2.03%)			BV29-1 (0.72%)	BJ2 (0.71%)		

(C)

T cell receptor:	Valpha											
Motif sequence:	De Novo "ELDGQ" (9-mer)											
Sample name:	E1603				E1632				E1655			
Phase of infection:	AIM		CONV		AIM		CONV		AIM		CONV	
Frequency of all copies (reads):	not applied		10.64% (1,243,532/11,680,562)		not applied		0.0012% (177/14,259,876)		not applied		not found	
Frequency of all unique clonotypes:	not applied		8.31% (217/2,609)		not applied		0.03% (1/3,292)		not applied		not found	
Family usage (%):	TRAV	TRAJ	TRAV	TRAJ	TRAV	TRAJ	TRAV	TRAJ	TRAV	TRAJ	TRAV	TRAJ
(% of usage of all other families detected for this sample)	not applied		AV5-1 (94%)	AJ16-1 (99.54%)	not applied		AV5-1 (100%)	AJ16-1 (100%)	not applied		not found	not found
			AV23-1 (1.38%)	AJ39-1 (0.46%)								
			AV13-1 (1.38%)									
			AV19-1 (0.92%)									
			AV14-1 (0.92%)									
			AV38-1 (0.46%)									
			AV9-1 (0.92%)									

T cell receptor:	Vbeta											
Motif sequence:	De Novo "SARD" (11-mer)											
Sample name:	E1603				E1632				E1655			
Phase of infection:	AIM		CONV		AIM		CONV		AIM		CONV	
Frequency of all copies (reads):	not applied		4.9% (517,874/10,580,372)		not applied		10.14% (948,733/9,348,647)		not applied		0.012% (666/ 5,575,518)	
Frequency of all unique clonotypes:	not applied		8.41% (213/2,532)		not applied		2.3% (132/5,764)		not applied		1.4% (73/5,207)	
Family usage (%):	TRAV	TRAJ	TRBV	TRBJ	TRAV	TRAJ	TRBV	TRBJ	TRAV	TRAJ	TRBV	TRBJ
(% of usage of all other families detected for this sample)	not applied		BV20-1 (98.59%)	BJ1-3 (99.53%)	not applied		BV20-1 (98.48%)	BJ1 (89.39%)	not applied		BV20-1 (100%)	BJ1 (95.89%)
			BV29-1(1.41%)	BJ2-3 (0.47%)			BV29-1 (1.52%)	BJ2-3 (10.6%)				BJ2-1 (4.11%)

Chapter 3 CDR3α appears to drive selection of the immunodominant YVL-specific CD8 TCR repertoire in primary EBV infection

This chapter was adapted from the following manuscript: CDR3 α drives selection of the immunodominant EBV BRLF1-specific CD8 T cell receptor repertoire in primary infection. Kamga L, Gil A, Song I, Brody R, Ghersi D, Aslan N, Stern LJ, Selin LK & Luzuriaga K.

Luzuriaga and Selin conceived the study and collaborated with clinicians to obtain the samples. Kamga contributed to study design, and was primarily responsible for cell sorting, TCR sequencing and expression. All authors contributed to data analyses. Ghersi performed all computational analyses. Song and Stern performed and analyzed the data from the crystallographic studies.

1 Introduction

In the previous chapter, we characterized both the TCR α and β repertoires of the immunodominant YVL- and GLC-specific CD8 T cells using bulk high-throughput sequencing. We identified a dominant YVL-specific TCRα chain, AV8.1-CAVKDTDKLIF-AJ34, that was shared by the three study participants. Here, we sought to understand a potential basis for the selection of this shared chain. We hypothesized that enriched residues within the CDR3 region might mediate selection of CD8 T cells expressing AV8.1-CAVKDTDKLIF-AJ34. To investigate this hypothesis, we performed single-cell TCR$\alpha\beta$ sequencing of YVL-specific CD8 T cells directly *ex vivo* over the course of primary infection and applied a newly developed analytical tool [53] for the identification of significantly enriched features in epitope-specific TCR repertoires. This confirmed selective use of particular V genes and provided evidence that recognition of the YVL epitope appeared to be driven by the TCRα chain, as well as identified novel pairing relationships between the α and β TCR chains. We observed that AV8.1-CAVKDTDKLIF-AJ34 TCRα paired with multiple different TCRβ chains within the same donor (range: 1-9). Newly developed analytical tool [53] identified a lysine at position 4 of the CDR3α that was highly conserved and likely important for Ag recognition. Mutating the CDR3α lysine led to an ablation of YVL-tetramer staining and reduced T-cell activation (using CD69 upregulation as a surrogate marker) by mutant TCR-transduced cells, suggesting that this residue is important for Ag recognition.

2 The TCR repertoire is individualized and qualitative features distinguish YVL- and GLC-specific TCR repertoires

To better understand the EBV-specific CD8 TCR repertoire, we performed single-cell paired TCR sequencing of tetramer-sorted, epitope-specific CD8 T cells from four donors during AIM and

CONV **(Table 1)**. A total of 65 and 64 (YVL; AIM and CONV) and 48 and 52 (GLC; AIM and CONV) productive paired TCR$\alpha\beta$ sequences were generated **(Tables 7 & 8)**. Circos plots were used to examine pairing relationships between Vα and Vβ gene segments by individual, epitope and time point (**Fig 21**). These analyses revealed that each individual had a unique repertoire. For example, the pattern of Vα-Vβ pairing for YVL TCRs was more intricate in some individuals (E1655 and E1651), as demonstrated by the dense interaction web, than in others (E1603 and E1632) (**Fig 21A**). However, despite the uniqueness of the repertoire, there were prominent features shared across individuals and these features were peculiar to each epitope. Such features included the use of AV8.1 by YVL TCRs in all individuals in both AIM and CONV. This gene was overrepresented in three of the four donors (E1603, E1655, and E1651) and paired with multiple Vβ genes **(Fig 21A)**. The pairing of AV8.1 with multiple Vβ genes was more pronounced in E1651 and E1655. The TCRβ repertoire of YVL used multiple different families that differed between the donors. With respect to GLC, although there were differences between the donors, AV5 and BV20 were common in all four patients in AIM and CONV. The public AV5-BV20 pairing was conserved in all individuals **(Fig 21B),** consistent with previous reports [86, 105]**.**

3 Selection of YVL-specific repertoire appears to be driven by TCRα

The qualitative analysis of the repertoire discussed above suggested the presence of generalizable features associated with epitope-specific TCRs. To quantitatively evaluate these features, we used a TCR$\alpha\beta$ sequence-based analytical tool recently developed by Dash and colleagues [53]. This tool identifies features that are significantly enriched in an epitope-specific repertoire by quantifying different properties of an epitope-specific TCR repertoire and comparing them against a collection of publicly available human TCR sequences. The Shannon-Jensen

divergence is used to identify significantly enriched gene segments (**Fig 22A**) and to score the magnitude of preferential gene usage (**Fig 22B**) within an epitope-specific repertoire, then the adjusted mutual information is used to measure the magnitude of gene usage correlations within and across TCRα/β (**Fig 22C**). In order to work with a large dataset, we merged the data from all patients, per time point and per epitope. We first examined the pattern of Vα-Jα/Vβ-Jβ gene segment pairing by ribbon plots (**Fig 22A**). There were identifiable enriched gene segments and pairings. For example, the AV8.1/AJ34 pairing was significantly enriched in the YVL repertoire in AIM and CONV but paired with multiple different Vβ/Jβ genes (**Fig 22A**). In contrast, the AV5/AJ31 combination paired with BV20.1/ BJ1.3 was significantly enriched in the GLC repertoire in AIM and CONV (**Fig 22A**).

To quantitatively evaluate the degree of bias in the repertoire, we evaluated the gene preference score. We observed that YVL-specific CD8 T cells exhibited a strong preferential usage of particular Vα and Jα in both AIM and CONV. By contrast, GLC-specific TCRs exhibited strong preferential usage of particular Vα, Jα, Vβ and Jβ **(Fig 22B)**. This is consistent with previous structural studies demonstrating that both the TCR α and β chains are important for recognition of the GLC epitope [64]. Altogether, these data suggest that the TCRα chain may be important for selection of YVL-specific CD8 T cells whereas both the TCR α and β chains together are important for selection of GLC-specific CD8 T cells. Moreover, quantification of the degree of gene usage correlations within and across TCRα/β revealed that GLC-specific TCR α and β repertoires were rigid; every gene association except for two in AIM (Vα-Jα and Vα-Jβ pairings) and one in CONV (Vα-Jβ) were enriched and thus important for selection into the repertoire **(Fig 22C)**. In contrast, the YVL repertoire was highly flexible in gene pairings. No obvious gene associations emerged for YVL-specific TCR in AIM, whereas only the Vα-Jα association emerged as important in CONV **(Fig 22C)**.

These data highlight the differences in selection of the YVL and GLC TCR repertoires at the clonal level.

4 The public and dominant CDR3α AV8.1-CAVKDTDKLIF-AJ34 pairs with multiple different TCRβ chains within the YVL repertoire

To dissect the clonal composition of the TCRαβ repertoires, we clustered similar clones into groups (Figs 23-25) by performing a number of analyses including 2D kernel principal component analysis (kPCA) projections (Figs 25C-D) and hierarchical clustering with dendograms (Figs 23-24) for YVL- (Figs 25C) and GLC- (Figs 25D) specific responses (n =4 donors pooled in AIM and CONV). In the 2D kPCA projections, the color correlates with gene usage. The hierarchical clustering is presented as a dendogram of the paired TCRαβ clones and also derives TCR logo representations showing gene usages and frequencies and CDR3 amino acid sequences of specific clusters. For the YVL response, clustering was driven by the TCRα chain, particularly the dominant AV8.1-KDTDKL-AJ34 expressing clones; this TCRα chain was detected in all individuals and resulted from an obligate pairing between AV8.1 and AJ34 (Fig 23). More importantly, this public AV8.1-KDTDKL-AJ34 TCR appeared to be so important for selection of the YVL TCR repertoire that there was a large number of clones (range: 1-9) where this one TCRα chain paired with multiple different TCRβ chains within a single donor (Fig 23 and Table 9). Additionally, there was also a redundancy in nucleotide sequences encoding this clonotype (Fig 26). Together, this degeneracy and redundancy suggest that this TCRα might be favored by its interaction with the A2-YVL complex. In contrast, in the GLC TCR repertoire there was little evidence of such pairing of a single public TCRα chain being paired with multiple different TCRβ chains. Unlike YVL, the clustering of GLC-specific TCRs was driven by dominant interactions with both the TCR α and β chains (Figs 24 & 25D).

58

5   Enriched residues within the CDR3α appear to drive selection of the shared AV8.1-CAVKDTDKLIF-AJ34 expressing clones

To better understand the factors driving the selection of AV8.1-CAVKDTDKLIF-AJ34-expressing clones, we analyzed CDR3 sequences for motifs or conserved residues that may determine Ag recognition [53]. The analytical tool reported by Dash and colleagues [53] is a predictive algorithm for the identification of key enriched residues within CDR3 regions that may be important in Ag recognition. It calculates motif enrichment against a background dataset of naïve non-YVL or GLC-specific CD8 TCRs **(Fig 25)**.

With this useful algorithm, for GLC we identified a known [53, 64] highly public dominant CDR3α motif "AEDNNA", with a conserved Asp(D) residue within these public AV5-AJ31 expressing clones (**Fig 25Bi**). We also identified two known [53, 64] CDR3β motifs, "RDxTGN" within the public BV20.1-expressing clones in CONV and "S/P,T/P,S/G,G" within the public BV14-expressing clones in AIM (**Fig 25Bii**) as potentially important for GLC recognition. We should note that CDR3 motifs identified within the GLC response during AIM were similar to those in CONV.

In the YVL-specific TCR repertoire, we discerned a strong CDR3α motif and somewhat surprisingly, the presence of CDR3β motifs **(Fig 25A)**. Within the public AV8.1/AJ34 clonotypes, CAVKDTKL was prevalent **(Fig 25A)** from AIM to CONV and contained the highly conserved "KD" motif, which is partially non-germline **(Fig 26)**, suggesting that they may provide a selective advantage and play a critical role in A2-YVL recognition. There was a second CDR3α motif ("CAXRxGGN") that included a conserved "GG" pair predominantly resulting from the AV14-AJ21 recombination event. Interestingly, although many different Vβ genes were used in the YVL response, the algorithm identified two potential CDR3β motifs, "xAxL", present in AIM and "GTSx" present in CONV, resulting from multiple Vβ-Jβ recombination events; this suggests that TCRβ may play some role in selection of YVL-specific TCRs but it differs between AIM and CONV.

To experimentally test whether Lys(K) at position 4 in the AV8.1-CAVKDTDKLIF-AJ34 chain and present in the highly conserved amino acid pair, "KD", was important for YVL recognition, we cloned a YVL-specific TCR expressing AV8.1-CAVKDTDKLIF-AJ34 (wild-type, WT,TCR; AV8.1-CAVKDTDKLIF-AJ34BV24.1-CATSDWDDSTGELFF-BJ2.2). We confirmed the antigenic specificity of the cloned WT TCR by expressing it in TCR-null CD8α-expressing J76 cells; the TCR-transduced cells stained with YVL tetramer (**Figs 27Ai, B**). To determine whether the Lys(K) within the conserved "KD" amino acid pair was involved in Ag recognition, we mutated the WT TCR (K113A TCR). This mutation abrogated YVL tetramer staining (**Fig 27Aii, B**), consistent with this Lys(K) contributing to A2-YVL recognition. We confirmed that the mutated TCR was expressed by measuring CD3 upregulation (**Fig 28**), which excludes the possibility that the lack of tetramer staining was due to a lack of TCR expression. Increasing the amount of YVL tetramer did not result in tetramer staining to the mutated TCR (K113A; **Fig 27B**). Additionally, we observed functional signaling through the wild-type TCR as measured by the upregulation of CD69 following stimulation of the WT TCR-transduced J76 cells with peptide-pulsed A2-expressing T2 cells acting as APCs **(Fig 27C & D)**. Introduction of the K113A mutation decreased CD69 upregulation (a proxy for T-cell activation), further indicating that Lys(K) contributed to A2-YVL recognition **(Fig 27C)**. Of interest, a functional response was still inducible even though tetramer no longer bound to this mutated TCR, consistent with previous studies demonstrating induction of functional response but no tetramer binding when using variant ligands [28, 109-111].

6 Conclusion

To investigate the potential basis for the selection of the YVL-specific TCRα sequence, AV8.1-CAVKDTDKLIF-AJ34, we performed single-cell sequencing of paired TCR α and β chains of EBV-

specific CD8 T cells isolated at two time points (primary infection and six months later in convalescence) from four individuals undergoing acute EBV infection. We described a highly degenerate AV8.1-CAVKDTDKLIF-AJ34 TCRα and provided evidence suggesting that selection of YVL-specific CD8 T cells appeared to be driven by the TCRα chain. Sequence analyses predicted that enriched residues, "KD" at positions 4-5, within this sequence likely contributes to viral recognition. This was validated using site-directed mutagenesis and *in vitro* assays.
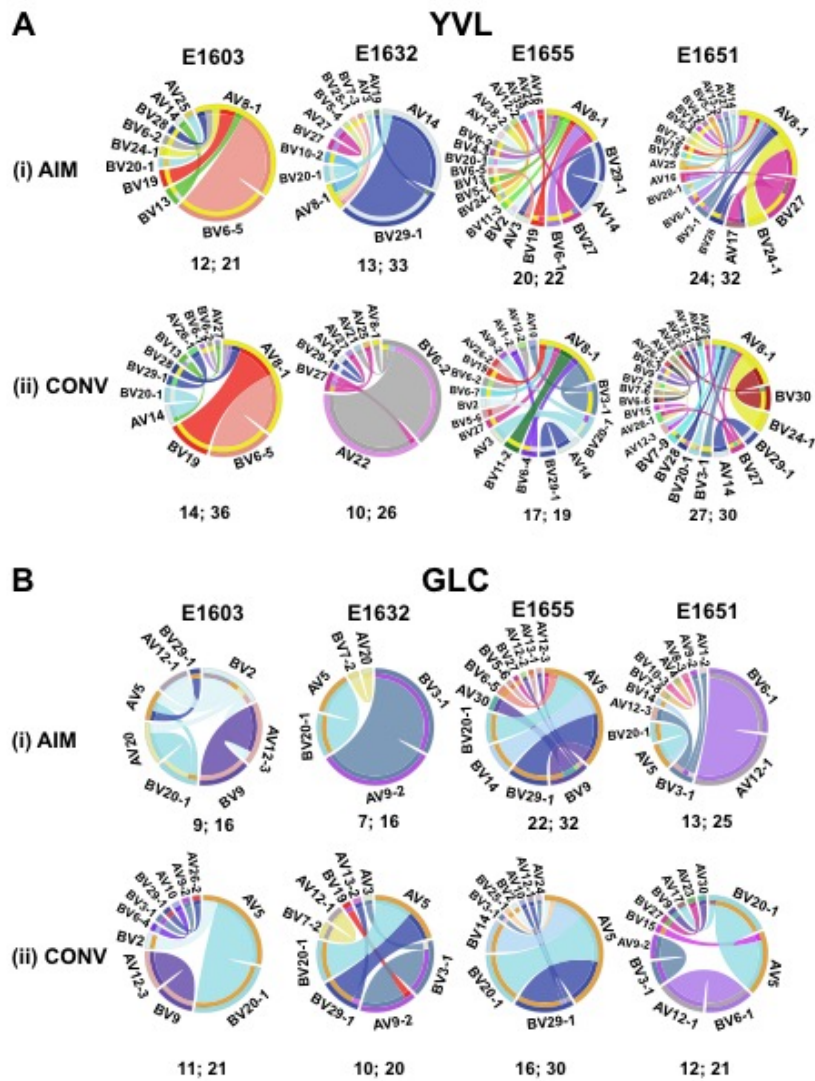
Fig 21. Patterns of Vα-Vβ gene pairings by YVL- and GLC-specific CD8 T-cells as revealed by single-cell TCR$\alpha\beta$ sequencing. Vα-Vβ gene pairings in four donors during AIM (i) and CONV (ii) for YVL- (A) and GLC- (B) specific TCR$\alpha\beta$ repertoires are displayed in circos plots, with frequency of each V gene represented by its arc length and that of the Vα-Vβ pairing by the width of the arc. The numbers of unique and productive paired TCR$\alpha\beta$ clonotypes as well as the total numbers of sequences for each donor are shown below the pie charts (# of unique TCR$\alpha\beta$ clonotypes; total # of sequences).
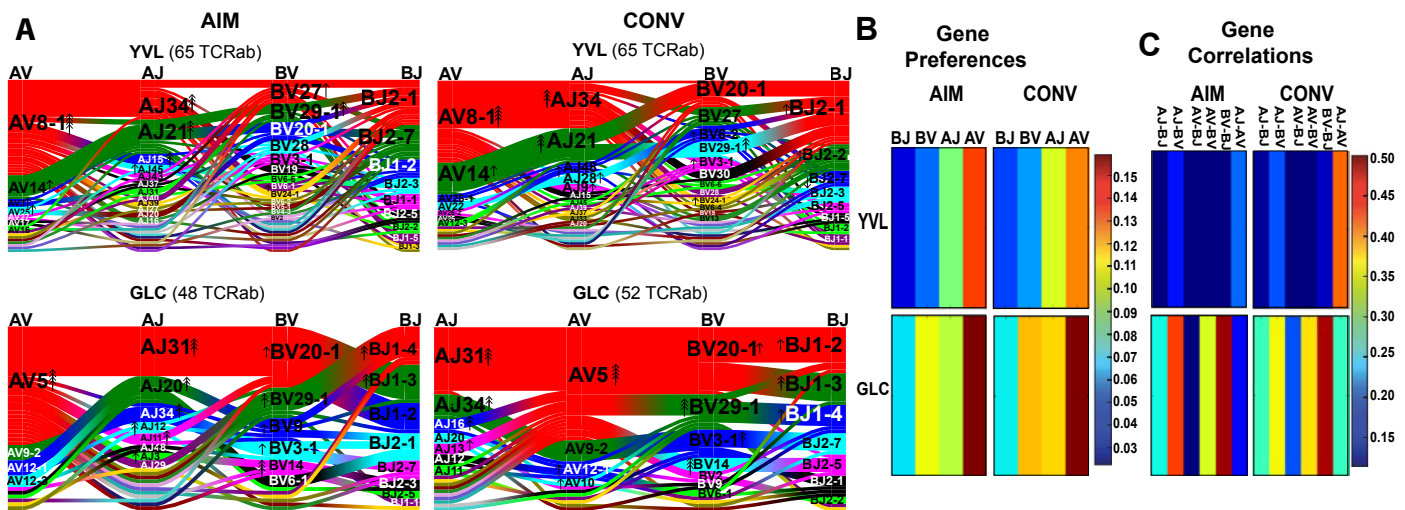
Fig 22. Single-cell paired TCR$\alpha\beta$ sequencing suggests that the selection of YVL-specific repertoire may be driven by TCRα. (A) Gene segment usage and gene–gene pairing landscapes for all four donors, combined, during AIM and CONV for YVL- (Top) and GLC-specific (Bottom) TCR $\alpha\beta$ repertoires are illustrated using four vertical stacks (one for each V and J segment) connected by curved paths whose thickness is proportional to the number of TCR clones with the respective gene pairing (each panel is labeled with the four gene segments atop their respective color stacks and the epitope identifier in the top middle). Genes are colored by frequency within the repertoire, which begins red (most frequent), green (second most frequent), blue, cyan, magenta, and black. The enrichment of gene segments relative to background frequencies is indicated by up or down arrows with an arrowhead number equal to the log2 of the fold change. (B), Jensen–Shannon divergence between the observed gene frequency distributions and background frequencies, normalized by the mean Shannon entropy of the two distributions (higher values reflect stronger gene preferences). (C), Adjusted mutual information of gene usage correlations between regions (higher values indicate more strongly covarying gene usage). Analyses are based on Dash *et al.* [53].
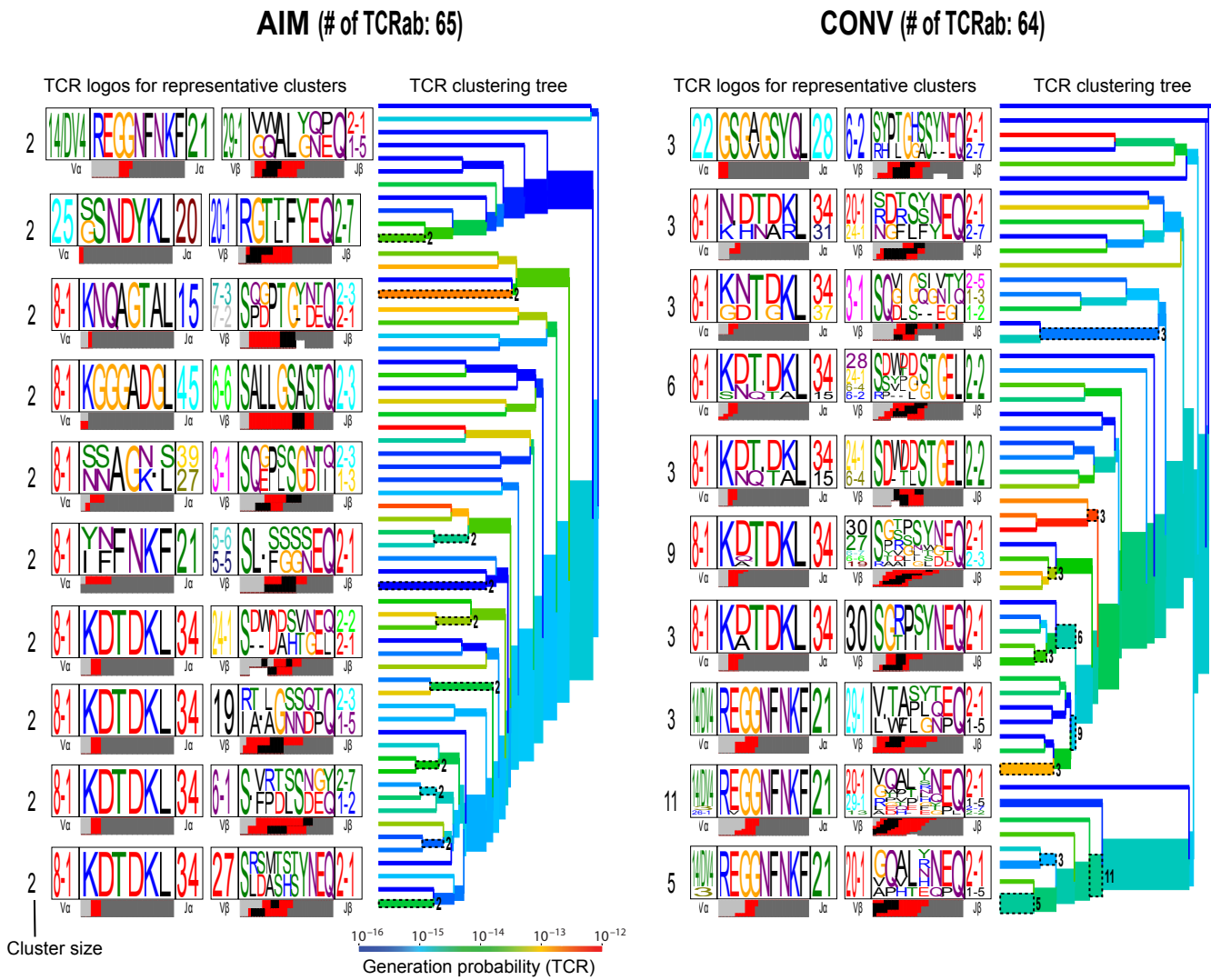
Fig 23. Hierarchical clustering of YVL-specific TCRs highlights the structural features required for interaction with pMHC of paired TCRα/β. (**A-B**) Hierarchical TCR α β clustering along with corresponding TCR logos for YVL-specific CD8 T-cell responses in AIM **(A)** and CONV **(B).** Number on the branches and next to TCR logos depicts number of TCRs contributing to the cluster. Color of the branches indicates the TCR probability generation scores. The bar at the bottom of the CDR3 logo is color-coded by the source of the nucleotide. Light grey, red, black, and dark grey denote that the nucleotides encoding those amino acid residues originate from the V, N, D and J regions, respectively. Analyses are based on Dash *et al.* [53].
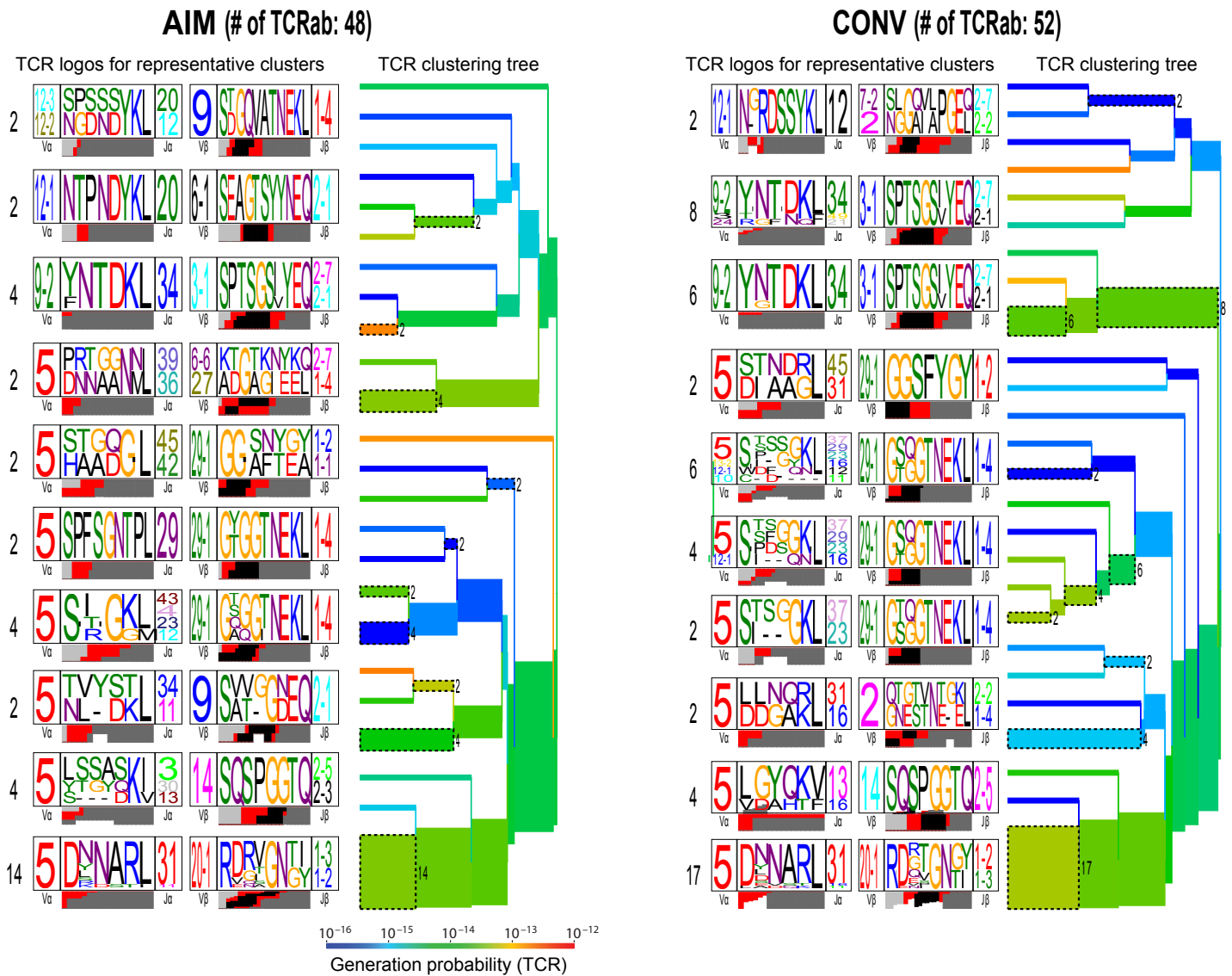
Fig 24. Hierarchical clustering of GLC-TCRs highlights the structural features required for interaction with pMHC of paired TCRα/β. (**A-B**) Hierarchical TCR $\alpha\beta$ clustering along with corresponding TCR logos for GLC-specific CD8 T-cell responses in AIM **(A)** and CONV **(B).** Number on the branches and next to TCR logos depicts number of TCRs contributing to the cluster. Color of the branches indicates the TCR probability generation scores. The bar at the bottom of the CDR3 logo is color-coded by the source of the nucleotide. Light grey, red, black, and dark grey denote that the nucleotides encoding those amino acid residues originate from the V, N, D and J regions, respectively. Analyses are based on Dash *et al.* [53].
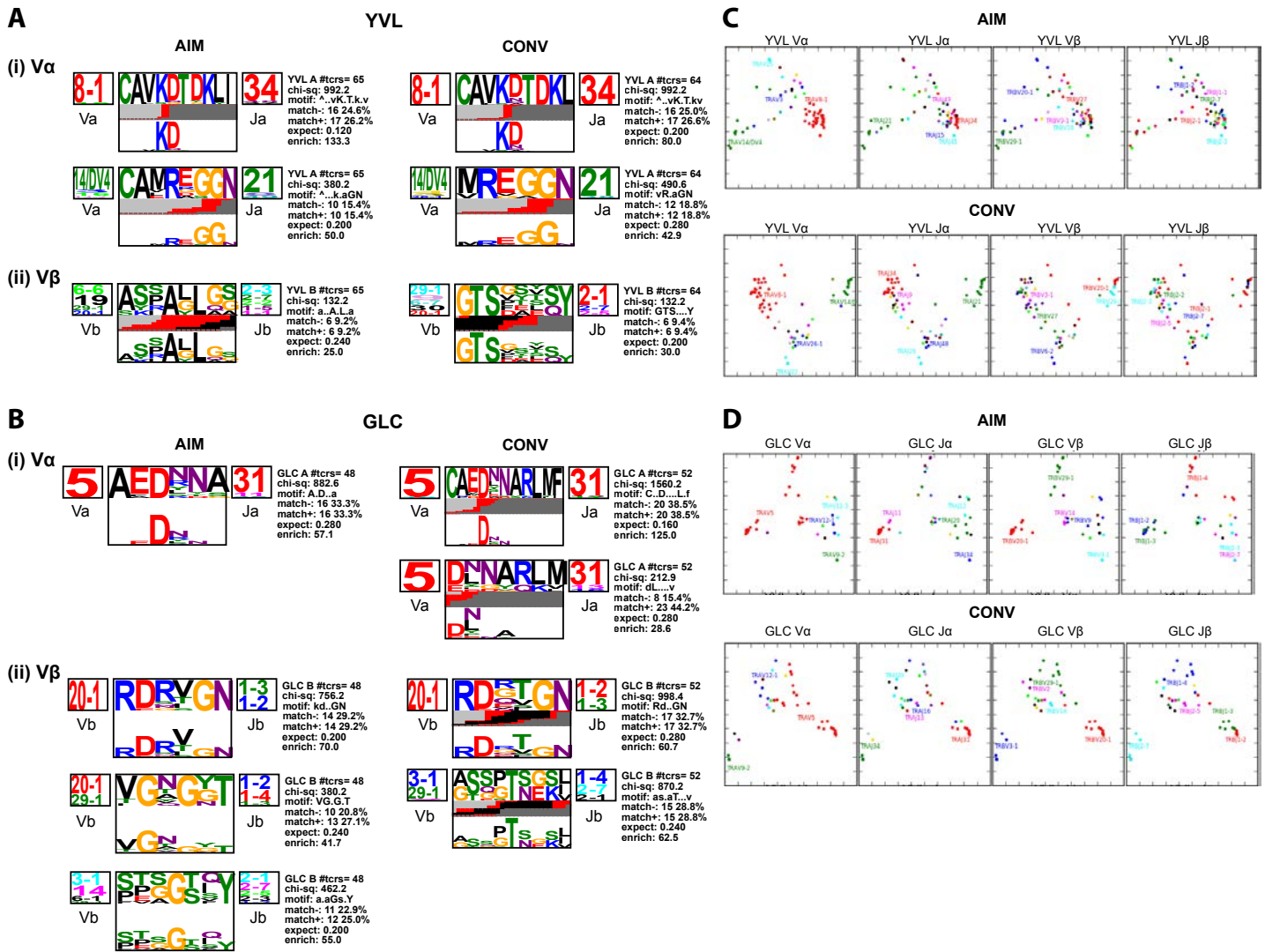
Fig 25. Single-cell CDR3α (i) and β (ii) motif analyses of YVL- (A) and GLC- (B) specific CD8 T cells during AIM and CONV. CDR3 α or β sequences were pooled from all 4 patients during AIM and CONV. Using TCRlogo in TCRdist algorithm, CDR3 sequence logos and top-scoring CDR3β and CDR3α motifs are shown for each group. V and J gene usage are indicated on the left and right side of logos, respectively. The bottom panel highlights enriched amino acids (motif) relative to a background dataset of naïve non-YVL- or GLC- specific CD8 TCRs. The bar at the bottom of the logo is color-coded by the source of the nucleotide. Light grey, red, black, and dark grey denote that the nucleotides encoding those amino acid residues originate from the V, N, D and J regions, respectively. **(C-D)** 2D kernel principal component analysis (kPCA) projections for YVL- (C) and GLC- (D) specific responses (*n* =4 donors pooled in AIM and CONV). Color correlates with gene usage. Most prevalent gene usages are mentioned within the plots. Each column depicts a 2D kPCA projection for each of the four gene segment usages (Vα, Jα, Vβ, and Jβ). Analyses are based on Dash *et al.* [53].

A

CDR3a aa ⟶   C   A   V   **K**   **D**   T   D   K   L   I   F

CDR3a nt ⟶ tgt   gcc   gtg   aa**g**   **g**ac   acc   gac   aag   ctc   atc   ttt

        AV8.1   AJ34

B

CDR3a aa ⟶   C   A   V   **K**   **D**   T   D   K   L   I   F

CDR3a nt ⟶ tgt   gcc   gtg   aa**a**   **g**ac   acc   gac   aag   ctc   atc   ttt

        AV8.1   AJ34

C

CDR3a aa ⟶   C   A   V   **K**   **D**   T   D   K   L   I   F

CDR3a nt ⟶ tgt   gcc   gtg   aa**a**   **gat**   acc   gac   aag   ctc   atc   ttt

        AV8.1   AJ34
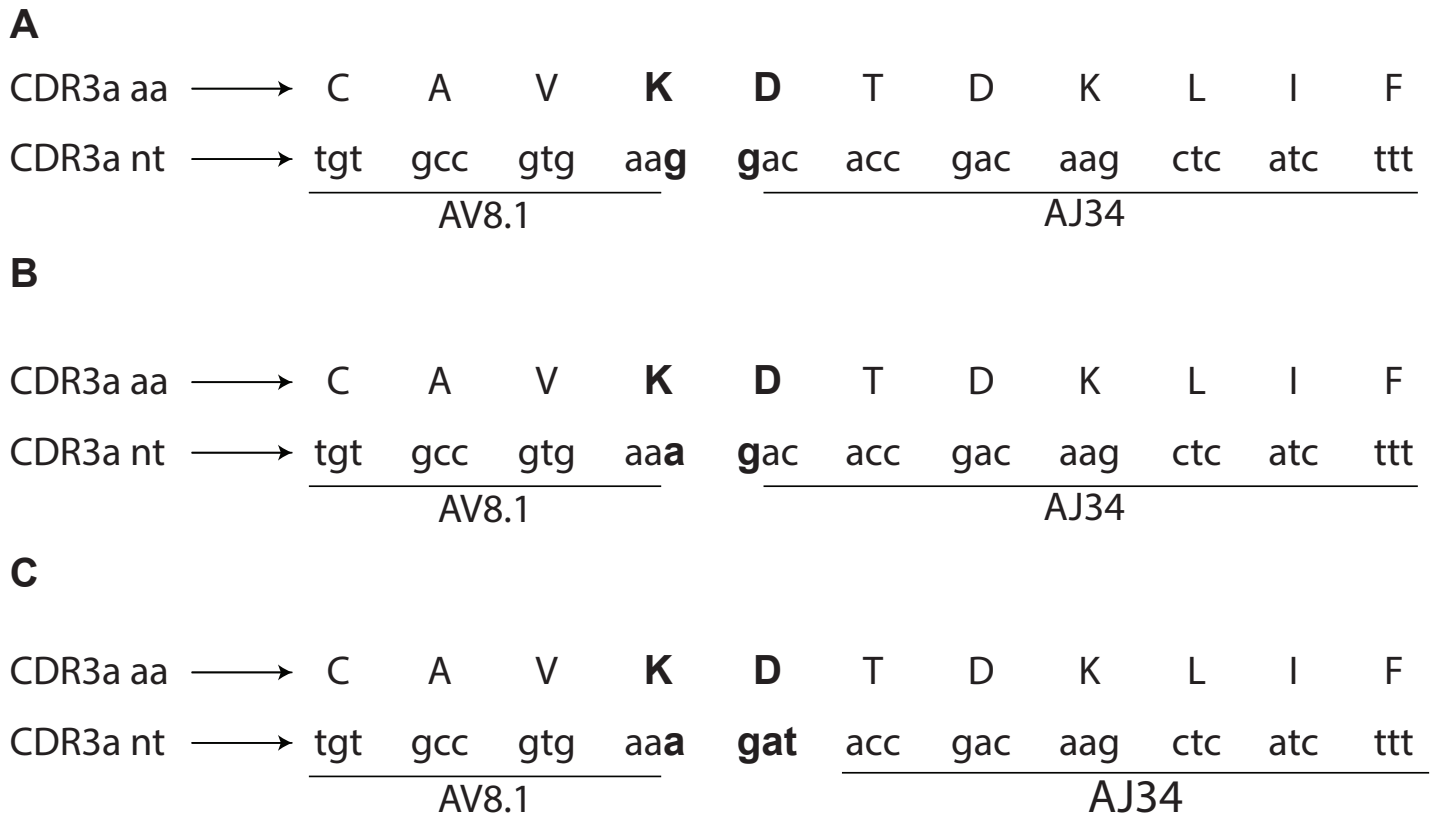
Fig 26. The KD motif is partially non-germline and is encoded by different nucleotide sequences (A & B). aa: amino acid; nt: nucleotide; bold: N nucleotide additions.
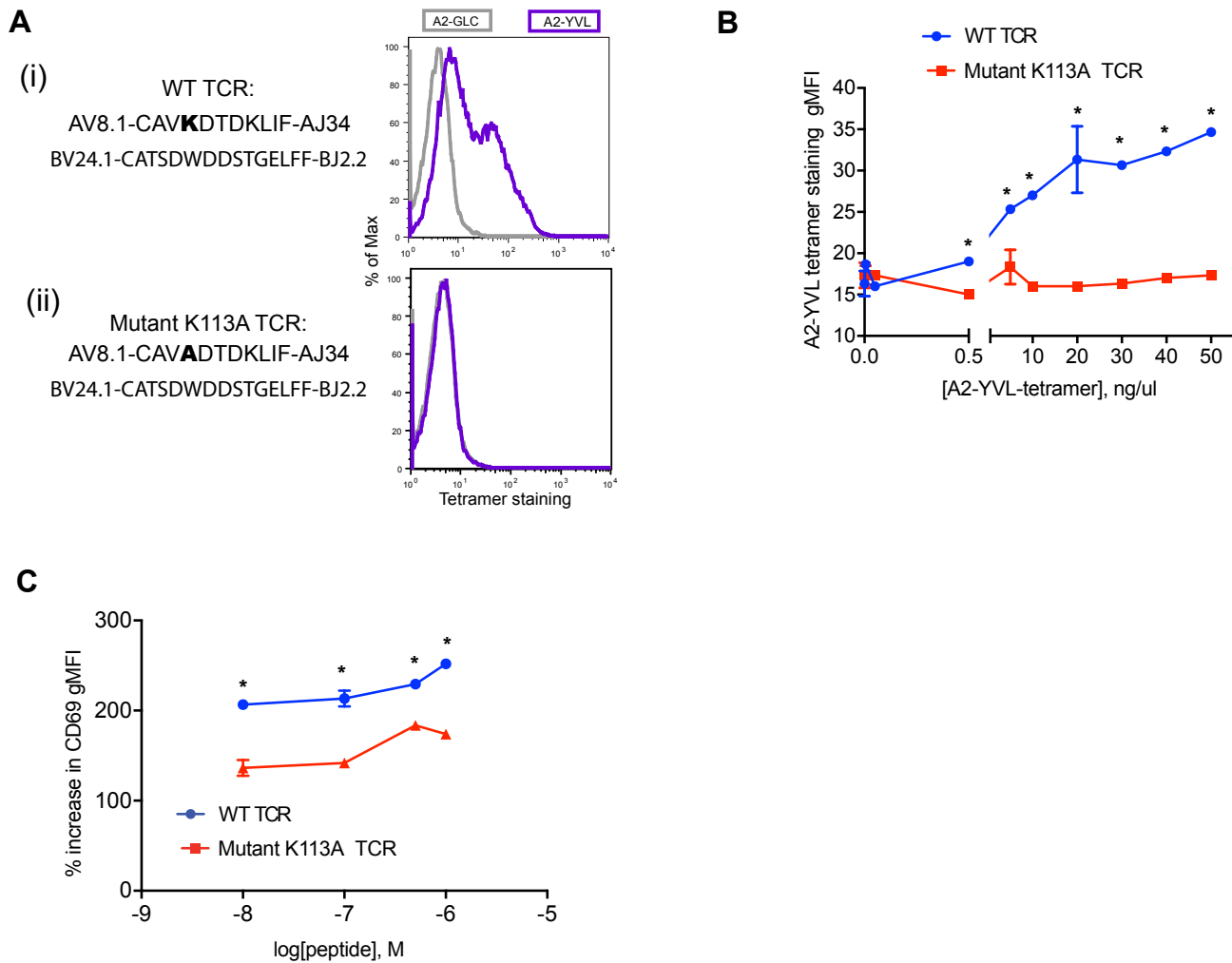
Fig 27. Site-directed mutagenesis demonstrates that Lys(K) in the "KD" motif of the public YVL-specific clonotype AV8.1-CAVKDTDKLIF-AJ34 is important for recognition of the A2-YVL complex. (A) Tetramer staining of J76-cells transduced with either the WT (i) or mutant K113A TCR (ii) and stained with the cognate A2-YVL- (purple) or an irrelevant A2-GLC- (grey) tetramer. CDR3α/β sequences are noted below each TCR name. The bold residue in the CDR3α denotes the one that has been mutated. (B) Tetramer titration of J76-cells transduced with either the WT (blue) or mutant K113A TCR (red). Assay was performed in three technical replicates and mean ± SD is shown. (C) CD69 upregulation of J76-cells transduced with either the WT (blue) or mutant K113A TCR (red). CD69 expression was measured on cells double positive for high expression of CD3 and CD8. Assay was performed in two technical replicates and mean ± SD is shown. Statistical analyses were done using multiple t tests. *: p value < 0.001. WT: wild-type.
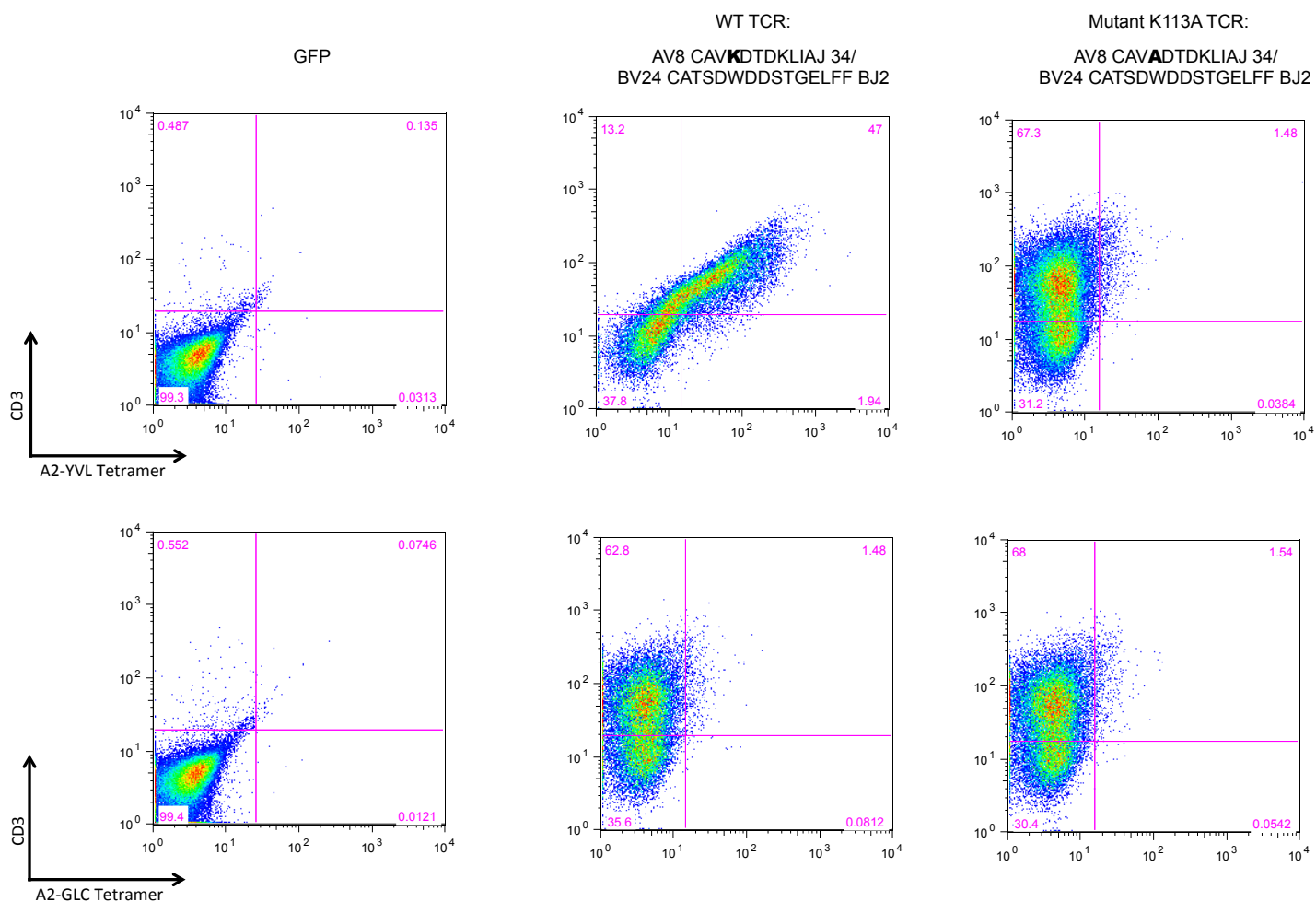
Fig 28. Flow cytometry analysis of J76 cells transduced with WT or mutant K113A TCRs and GFP, as a mock control, and stained with the cognate A2-YVL (top) or an irrelevant A2-GLC (bottom) tetramer and CD3 antibody. The CDR3 α and β amino acid sequences of the TCRs are shown. Bold: residue that has been mutated.

Table 7. Paired single-cell YVL-specific TCR sequences.

Donor ID	CDR3α[1]	CDR3β[2]	Count AIM	Count CONV
1632	AV8CAVKDTDKLIF AJ34	BV10CASMLPFGDEQYF BJ1	1	
1632	AV8CAVKDTDKLIF AJ34	BV10CASMLPFGDEQYF BJ2	1	
1655	AV8CAAPGAGSYQLTF AJ28	BV11CASMRELAGQETQYF BJ2		2
1655	AV8CNASGAGSFHFTF AJ28	BV11CASMRELAGQETQYF BJ2	1	
1603	AV14CAMREGTGNFNKFYF AJ21	BV13CASRQTSGELFF BJ2	1	1
1603	AV27SSPRFSDGQKLLF AJ16	BV13CASRQTSGELFF BJ2		1
1603	AV8CAVKGGGADGITF AJ45	BV13CASRQTSGELFF BJ2	1	
1655	AV8CLIGQAGFTLIF AJ15	BV13CASSSPRGTGGRDTGELFF BJ2	1	
1651	AV12CAMSASNFGNEKLTF AJ48	BV15CATSSTARDSSYSNQPQHF BJ1		1
1651	AV8CALSGGSQGNLIF AJ42	BV15CATSTGLAGIHEQYF BJ2	1	
1603	AV8CAVKDTDKLIF AJ34	BV19CASIAYLGSNQPQHF BJ1	1	
1655	AV8CAVNVPDGQKLLF AJ16	BV19CASRALLGGATEAFF BJ1	1	1
1603	AV14CAMRGGVNNHNKFIF AJ21	BV19CASRTAGNSDTQYF BJ2	1	
1603	AV8CAVKDTDKLIF AJ34	BV19CASRTAGNSDTQYF BJ2	1	10
1655	AV16CGVRNRDDKIIF AJ30	BV19CASSIGFDIETQYF BJ2	1	
1651	AV8CAVKDTDKLIF AJ34	BV19CASSSLLISEAFF BJ1	1	
1603	AV14CAMREGGNFNKFYF AJ21	BV20CSAAQALYNEQFF BJ2		1
1655	AV3CAVREGGNFNKFYF AJ21	BV20CSAGQALRNEQFF BJ2		1
1603	AV14CAMREGGNFNKFYF AJ21	BV20CSAGQVLEQPQHF BJ1		1
1651	AV12CAMSASNFGNEKLTF AJ48	BV20CSANDRSYNEQFF BJ2		1
1655	AV3SVVSTGARFNKFFF AJ21	BV20CSARDLAGNTGELFF BJ2		1
1655	AV9CALRDTSGSRLTF AJ58	BV20CSARDSRDLLRGYTEAFF BJ1		1
1651	AV25CAAGSNDYKLSF AJ20	BV20CSARGTLFYEQYF BJ2	1	
1651	AV8CAVNDNARLMF AJ31	BV20CSARGTSFYEQYF BJ2		1
1603	AV25CAGSSNDYKLSF AJ20	BV20CSARGTTFYEQYF BJ2	1	
1655	AV22CAVGPLVRF AJ43	BV20CSASSNSAYGYTF BJ1	1	
1632	AV14CAMREGGNFNKFYF AJ21	BV20CSASYPAGLQAGGGDEQYF BJ2	2	
1651	AV14CAMRAGGNFNKFYF AJ21	BV20CSATIPPDNEQFF BJ2	1	
1632	AV3CAVRDGGNFNKFYF AJ21	BV20CSAVGLAGGFIVDEQFF BJ1	1	
1632	AV3CAVRDGGNFNKFYF AJ21	BV20CSAVGLAGGFIVDEQFF BJ2	1	
1603	AV14CAMREGGNFNKFYF AJ21	BV20CSGGQALHNEQFF BJ2		1
1603	AV8CAVKDTDKLIF AJ34	BV24CATSDAHVNEQFF BJ2	1	
1651	AV8CAVNAHTDKLIF AJ34	BV24CATSDFLSNEQFF BJ2		1
1655	AV38CASLNSGGGADGLTF AJ45	BV24CATSDPTRGRRNNEQFF BJ2	1	
1651	AV8CAVKDTDKLIF AJ34	BV24CATSDWDDSTGELFF BJ2	4	2
1651	AV8CAVKNQAGTALIF AJ15	BV24CATSDWDDSTGELFF BJ2		1
1632	AV8CAVKNQAGTALIF AJ15	BV25CASSEWGLGEAFF BJ1	1	
1632	AV25CAGATIGFGNVLHC AJ35	BV27CASGRLAGYNEQFF BJ2		1
1632	AV22WAAGSGVGSFQLTF AJ28	BV27CASSFLGQTMNTEAFF BJ1		1
1632	AV27CAGPLTGKLIF AJ37	BV27CASSFLGQTMNTEAFF BJ1	3	1
1655	AV39CAVMRTYKYIF AJ40	BV27CASSLRGGGYEQYF BJ2	1	

	AV	BV		
1651	AV26CIVRGRNFGNEKLTF AJ48	BV27CASSLSITRVYEQYF BJ2		1
1651	AV8CAVKDTDKLIF AJ34	BV27CASSLSMTHTYNEQFF BJ2	1	
1655	AV8CAVKQTDKLIF AJ34	BV27CASSPDIGWGGEQFF BJ2	1	
1655	AV8CAVKDTDKLIF AJ34	BV27CASSPSLSDYFTDTQYF BJ2		1
1651	AV16CALRGSGNQFYF AJ49	BV27CASSPTTMLETQYF BJ2	1	
1651	AV8CAVKDTDKLIF AJ34	BV27CASSRAASSSYNEQFF BJ2		1
1651	AV25CAGLQGANNIFF AJ36	BV27CASSRDASSSYNEQFF BJ2		1
1651	AV8CAVKDTDKLIF AJ34	BV27CASSRDASSSYNEQFF BJ2	3	
1651	AV8CAVKGTYKYIF AJ40	BV27CASSSNPMLETQYF BJ2	2	
1651	AV8SIGKGPDKLIL AJ34	BV27CASSSNPMLETQYF BJ2	1	
1651	AV17CATDADYGQNFVF AJ26	BV27CASSTVPGHQPQHF BJ1	3	
1655	AV8CALPGLNNDMRF AJ43	BV27CASSTYSGRATEQYF BJ2	1	
1603	AV8CAVNNQAGTALIF AJ15	BV28CAGRPLLGGGSPLHF BJ1		1
1651	AV8CAVIAGGYQKVTF AJ13	BV28CANVMGGPEGGYTF BJ1	1	
1603	AV8CALKDTDKLIF AJ34	BV28CASRPWGGTGELFF BJ2		1
1603	AV8CAVKNQAGTALIF AJ15	BV28CASRSSYNSPLHF BJ1	1	
1651	AV17CALNTGGFKTIF AJ9	BV28CASSLSGSRSEQFF BJ2	1	
1651	AV8CAVKDTDKLIF AJ34	BV28CASSSYPGLSTGELFF BJ2	1	
1651	AV8CAVSNTDKLIF AJ34	BV28CASSSYPGLSTGELFF BJ2		1
1651	AV8CAVAGTGNQFYF AJ49	BV28CPAVPKGDSAPGTVFF BJ2		1
1632	AV14CAMREGGNFNKFYF AJ21	BV29CSVAKPAGLAGGSNTGELFF BJ2	6	
1632	AV14CAMREGGNFNKFYF AJ21	BV29CSVGGTGSGTSGAYSYNEQFF BJ2	13	1
1632	AV14GAMREGGNFNEFYC AJ21	BV29CSVGGTGSGTSGAYSYNEQFF BJ2	1	
1651	AV14CAMREGGNFNKFYF AJ21	BV29CSVGGTSGSVSYNEQFF BJ2		1
1655	AV14CAMREGGNFNKFYF AJ21	BV29CSVGQALYNEQFF BJ2	1	
1651	AV14CAMREGGNFNKFYF AJ21	BV29CSVLDPTFSYNEQFF BJ2		1
1632	AV19CALSSNARLMF AJ31	BV29CSVREAANYGYTF BJ1	1	
1655	AV14CAMREGGNFQKLVF AJ8	BV29CSVVAGANNEQFF BJ2	2	
1655	AV14CAMREGGNFNKFYF AJ21	BV29CSVVAPLWNEQFF BJ2	1	
1603	AV26CIVRVGGNFNKFYF AJ21	BV29CSVVEPPYNEQFF BJ2		1
1651	AV14CAMREGGNFNKFYF AJ21	BV29CSVVTAPLTEQFF BJ2		1
1655	AV14CAMREGGNFNKFYF AJ21	BV29CSVVWALGQPQHF BJ1	1	1
1655	AV14KSYGGGASYGKLTF AJ52	BV29CSVVWALGQPQHF BJ1		1
1655	AV3CAVNNARLMF AJ31	BV29CSVVWALGQPQHF BJ1	1	
1603	AV26CIVRVGGNFNKFYF AJ21	BV29PSVVEPPYNEQFF BJ2		1
1655	AV1CAVRRGSTLGRLYF AJ18	BV2CASSAPGGTGGRNTEAFF BJ1	1	
1655	AV14CAMSAPPTSGSVRQLTF AJ22	BV2CASSSPLAESPAGELFF BJ2		1
1651	AV8CAVKDTDKLIF AJ34	BV30CAWSGRPSYNEQFF BJ2		2
1651	AV8CAVKATDKLIF AJ34	BV30CAWSGTPIYNEQFF BJ2		2
1651	AV23CAASIPHFQAGTDLIF AJ15	BV30CAWSGTSPSSSYNEQFF BJ2		1
1651	AV8CAVAGTGNQFYF AJ49	BV30CSGRGQSTSEQYF BJ2		1
1651	AV8CAVKNTDKLIF AJ34	BV3CASSQDLGQIETQYF BJ2		1
1655	AV10CVVSAPTGANKLIF AJ20	BV3CASSQDRAAYGYTF BJ1		1
1651	AV8CAQGDAGNMLTF AJ39	BV3CASSQEPGSGETQYF BJ2		1
1603	AV8CAVNNAGNMLTF AJ39	BV3CASSQEPLSGDTQYF BJ2	1	

71

	CDR3α[1]	CDR3β[2]		
1655	AV8CAVGNTGKLIF AJ37	BV3CASSQGISSGNTIYF BJ1		2
1651	AV13CAERRGGNFNKFYF AJ21	BV3CASSQGLADGETQYF BJ2	1	
1651	AV27CARGNEKLTF AJ48	BV3CASSQGLSGRAHEQFF BJ2	1	
1651	AV8CAVSSAGKSTF AJ27	BV3CASSQGPSSGNTIYF BJ1	1	
1655	AV8CAVKDTDKLIF AJ34	BV3CASSQVIGVGYTF BJ1		1
1655	AV8CAFGGYNKLIF AJ4	BV4CASSDVTGTYGYTF BJ1	1	
1651	AV8CAVKDTDKLIF AJ34	BV4CASSNTAGVLGDEQFF BJ2	1	
1655	AV26CILRDSHTGTASKLTF AJ44	BV5CASKSLSSYEQYF BJ2		1
1655	AV12CAVGNTNAGKSTF AJ27	BV5CASSLAAREDEQYF BJ2	1	
1651	AV25CAYYNFNKFYF AJ21	BV5CASSLELAGANSYEQYF BJ2	1	
1632	AV8CAGYNFNKFYF AJ21	BV5CASSLFSGNEQFF BJ2	1	
1651	AV8CAVIFFNKFYF AJ21	BV5CASSLGASGSSEQFF BJ2	1	
1655	AV8CAVKDTDKLIF AJ34	BV6CARSGTSDTLSYNEQFF BJ2		1
1603	AV8CALNSNYQLIW AJ33	BV6CASRALSGGGQPQHF BJ1		1
1632	AV22CAAGSGAGSYQLTF AJ28	BV6CASRHPLGGASEQYF BJ2		1
1651	AV8CAVNSDYKLSF AJ20	BV6CASRQLTGGAQPQHF BJ1		1
1632	AV8CAVGNYQLIW AJ33	BV6CASRQQGSTEAFF BJ1		1
1603	AV8CAGKGGGADGLTF AJ45	BV6CASSALLGSASTQYF BJ2	1	
1603	AV8CAVKGGGADGLTF AJ45	BV6CASSALLGSASTQYF BJ2	10	14
1655	AV1CAVNNDYKLSF AJ20	BV6CASSASPGGAGNEQFF BJ2		1
1655	AV8CAVKDTDKLIF AJ34	BV6CASSDTLSTGELFF BJ2		1
1651	AV16CALKDTDKLIF AJ34	BV6CASSEFSMYEAFF BJ1	1	
1603	AV8CAFYHAGNMLTF AJ39	BV6CASSGPGWDEQYF BJ2	1	
1655	AV3CAVNNARLMF AJ31	BV6CASSPDSYEQYF BJ2	1	
1655	AV8CAVKDTDKLIF AJ34	BV6CASSSFRDSSNEQYF BJ2	2	
1655	AV12CAVNGPPPSGSARQLTF AJ22	BV6CASSSTYPGSVGETQYF BJ2		1
1603	AV8CAVKDTDKLIF AJ34	BV6CASSSVLDFLGTGELFF BJ2		1
1655	AV8CAVKDTGKLIF AJ37	BV6CASSVIDTQYF BJ2	1	
1651	AV8CAVKDTDKLIF AJ34	BV6CASSVPTLSDGYTF BJ1	1	
1632	AV21CAVPMYSGGGADGLTF AJ45	BV6CASSYIGVGYTF BJ1		1
1632	AV22CAAGSGAGSYQLTF AJ28	BV6CASSYPTGHSSYNEQFF BJ2		17
1632	AV22CAAGSGVGSYQLTF AJ28	BV6CASSYPTGHSSYNEQFF BJ2		1
1632	AV22RDAGCGAGSYQFTF AJ28	BV6CASSYPTGHSSYNEQFF BJ2		1
1603	AV8CAVKDTDKLIF AJ34	BV6CASSYVGLLGDTQYF BJ2		1
1651	AV26CILRGPPLGNEKLTF AJ48	BV6CASTGIQGNTGELFF BJ2		1
1651	AV8CAVNAVGDMRF AJ43	BV7CASSAGQGSGNTIYF BJ1	1	
1651	AV26CIVPSTSGTYKYIF AJ40	BV7CASSLAGINYGYTF BJ1		1
1632	AV8CAVKNQAGTALIF AJ15	BV7CASSPDPTGYNEQFF BJ2	1	
1651	AV8CAVKNQAGTALIF AJ15	BV7CASSQGPTGDTQYF BJ2	1	
1651	AV8CAVSEGTNAGKSTF AJ27	BV7CASSYTGRALEAFF BJ1		1
1651	AV8CAVVTGGFKTIF AJ9	BV7CASSYTTGSADTQYF BJ2		1
1651	AV12CVVNRLDNAGNMLTF AJ39	BV9CASSVAGTSVETQYF BJ2		1

[1]Color-coded by CDR3 α clones. [2]Color-coded by CDR3 β clones.

Table 8. Paired single-cell GLC-specific TCR sequences.

Donor ID	CDR3α[1]	CDR3β[2]	Count AIM	Count CONV
E1603	AV5CAEDNNARLMF AJ31	BV20CSARDGTGNGYTF BJ1		3
E1603	AV5CAEDNNARLMF AJ31	BV20CSARDQTGNGYTF BJ1		4
E1603	AV5CAELDVQKLVS AJ16	BV20CSARDRVGNTIYF BJ1		1
E1603	AV5CAEDYNARLMF AJ31	BV20CSARDRVGNTIYF BJ1		2
E1603	AV5CAEDKNARLMF AJ31	BV20CSARDRVGNTIYF BJ1	1	
E1603	AV20CAWQGNYGQNFVF AJ26	BV20CSARDRVGNTIYF BJ1	2	
E1603	AV20CARNGYNDYKLSF AJ20	BV20CSARDRVGNTIYF BJ1	1	
E1603	AV5CAEDENARLMF AJ31	BV20CSARDVPGNTIYF BJ1		1
E1603	AV5CAESGRGKLIF AJ12	BV29CSVGAGGTNEKLFF BJ1	1	
E1603	AV26CILTGGGNKLTF AJ10	BV29CSVSPVDYYTQYF BJ2		1
E1603	AV5CAEVDAHTFLF AJ16	BV2CASQNGTVNTGELFF BJ2	1	
E1603	AV5CAELDGQKLLF AJ16	BV2CASQNGTVNTGELFF BJ2		1
E1603	AV20CARKGNNGYKLVF AJ26	BV2CASSEGQIAPGELFF BJ2	1	
E1603	AV12SAMSGYTDYKLSF AJ20	BV2CASSEGQIAPGELFF BJ2	1	
E1603	AV12CVVNGEDSSYKLIF AJ12	BV2CASSEGQIAPGELFF BJ2	3	
E1603	AV9CAGYNTDKLIF AJ34	BV3CASSPTSGSVYEQFF BJ2		1
E1603	AV10CVVSEGKLIF AJ34	BV6CASELWTGHNEQFF BJ2		1
E1603	AV12CAMSGSNDYKLSF AJ20	BV9CASSDGQVATNEKLFF BJ1	5	5
E1632	AV9CASYGTDKLIF AJ34	BV19CCRSPRLARRRIYEQYF BJ2		1
E1632	AV5CAEDNNARLMF AJ31	BV20CSARDQTGNGYTF BJ1		1
E1632	AV5CAEDLNARLMF AJ31	BV20CSARDRIGNTIYF BJ1	2	4
E1632	AV5CAEDNNARLMF AJ31	BV20CSARDRTGNGYTF BJ1	1	
E1632	AV5CAEDRDSTLTF AJ11	BV20CSARVGAGNTIYF BJ1	1	
E1632	AV5CAEHAGQGNLIF AJ42	BV29CSVGGQANTEAFF BJ2		1
E1632	AV5CAEHAGQGNLIF AJ42	BV29CSVGGQANTEAFF BJ1		3
E1632	AV13CATWGF AJ11	BV29CSVGTGGTNEKLFF BJ1		1
E1632	AV9CASYGTDKLIF AJ34	BV3CASSPTSGSIYEQYF BJ2		1
E1632	AV9CALYNTDKLIV AJ34	BV3CASSPTSGSIYEQYF BJ2	1	
E1632	AV9CALYNTDKLIF AJ34	BV3CASSPTSGSIYEQYF BJ2	9	5
E1632	AV3CAVRDGGNFNKFYF AJ21	BV3CASSPTSGSIYEQYF BJ2		1
E1632	AV9CALFNTDKLIF AJ34	BV3CASSPTSGSVYEQFF BJ2	1	
E1632	AV20WGGNWDGSYKLIF AJ12	BV7CASNLGQILPGEQYF BJ2	1	
E1632	AV12CVVNRDSSYKLIF AJ12	BV7CASNLGQILPGEQYF BJ2		2
E1651	AV8CAVVGTGNQFYF AJ49	BV10CAISDDRGPYEQYF BJ2	1	
E1651	AV5CAELSDKIIF AJ30	BV14CARSQSPGGTQYF BJ2	1	
E1651	AV5CAEAMAGNMLTF AJ39	BV15CATSITSGSQTQYF BJ2		1
E1651	AV5CAVDNNARLMF AJ31	BV20CSARDETGNGYTF BJ1		2
E1651	AV5CAEDADSTLTF AJ11	BV20CSARDGTGNTIYF BJ1		1
E1651	AV5CAEDNNARLMF AJ31	BV20CSARDQIGNGYTF BJ1	1	

E1651	AV5CAEDNNARLMF AJ31	BV20CSARDRTGNGYTF BJ1		4
E1651	AV17CALYNTDKLIF AJ34	BV20CSARDRTGNGYTF BJ1		1
E1651	AV12CVVNVPNDYKLSF AJ20	BV20CSARDRTGNGYTF BJ1		1
E1651	AV5CAEDRYSTLTF AJ11	BV20CSARDSTGNGYTF BJ1	1	
E1651	AV5WTEDQRGKLMF AJ23	BV20CSVRDRVGNTIYF BJ1	1	
E1651	AV23CAVTGGAANKLIF AJ32	BV27CASSPGTSGYYNEQFF BJ2		1
E1651	AV12CAISNFGNEKLTF AJ48	BV3CASSFTSGSIYEQYF BJ2	2	
E1651	AV9CALYNTDKLIF AJ34	BV3CASSPTSGSIYEQFF BJ2		2
E1651	AV9CAFYNTDKLIF AJ34	BV3CASSPTSGSIYEQFF BJ2	1	
E1651	AV9CALYNTDKLIF AJ34	BV3CASSPTSGSVYEQFF BJ2		1
E1651	AV1VYVSDKDSTDKLIF AJ34	BV3CASSPTSGSVYEQYF BJ2	1	
E1651	AV12CVVNTPNDYKLSF AJ20	BV6CASSEAGLATTMSSSS BJ2	1	
E1651	AV12RVVNTPGNDKFTF AJ20	BV6CASSEAGTSYYNEQFF BJ2	2	
E1651	AV12CVVNTPNDYKLTF AJ20	BV6CASSEAGTSYYNEQFF BJ2	1	
E1651	AV12CVVNTPNDYKLSF AJ20	BV6CASSEAGTSYYNEQFF BJ2	11	4
E1651	AV12CVVNRPNDYKLSF AJ20	BV6CASSESAWVAGGSDTQYF BJ2		1
E1651	AV12CVVNMEGYSTLTF AJ11	BV6CASSESPMWDPRYGYTF BJ1		1
E1651	AV4CLVVNDYKLSF AJ20	BV7CASSLAFSGLLTDTQYF BJ2	1	
E1651	AV30CGTEILNDYKLSF AJ20	BV9CASSEGQLSSGNTIYF BJ1		1
E1655	AV5CAELGYQKVTF AJ13	BV14CANSQSPGGTQFF BJ2		1
E1655	AV5CAVYSSASKIIF AJ3	BV14CASSQSPGGTQYF BJ2	1	
E1655	AV5CAESTSASKIIF AJ3	BV14CASSQSPGGTQYF BJ2	1	
E1655	AV5CAELGYQKVTF AJ13	BV14CASSQSPGGTQYF BJ2	3	5
E1655	AV5CADRQSQKDTF AJ13	BV14CASSQSPGGTQYF BJ2	1	
E1655	AV5CAVSTSYGKLTF AJ52	BV20CSAPRAGGGQETQYF BJ2		1
E1655	AV5CAEDRMPEIMF AJ31	BV20CSARDGTGNGYTF BJ1		1
E1655	AV5CAEDNNARLMF AJ31	BV20CSARDGTGNGYTF BJ1		1
E1655	AV5CAFDNNARLMF AJ31	BV20CSARDGVGNGYTF BJ1	1	
E1655	AV5CAEDLNARLMF AJ31	BV20CSARFRENSGNTIYF BJ1	5	7
E1655	AV5CAEDNNARLMF AJ31	BV25CASSGARDTQYF BJ2		1
E1655	AV5CAEDRTGANNLFF AJ36	BV27CASADGAKIYEQYF BJ2	1	
E1655	AV5CPGDINARLMF AJ31	BV29CSVGGSFYGYTF BJ1		1
E1655	AV5CAESTADGLTF AJ45	BV29CSVGGSFYGYTF BJ1	2	1
E1655	AV5CAESIGGMRF AJ43	BV29CSVGQGGTNEKLFF BJ1	1	
E1655	AV5CAESPFSGNTPLVF AJ29	BV29CSVGSGGTNEKLFF BJ1		3
E1655	AV5CAESISGGKLIF AJ23	BV29CSVGSQGINEKLFF BJ1	1	
E1655	AV5VSESISGRKLIF AJ23	BV29CSVGSQGTNEKLFF BJ1	1	
E1655	AV5CAESISGGKLIF AJ23	BV29CSVGSQGTNEKLFF BJ1		1
E1655	AV12CVVSWFSDGQKLLF AJ16	BV29CSVGSQGTNEKLFF BJ1		1
E1655	AV10CVVCRMDSSYKLIF AJ12	BV29CSVGSQGTNEKLFF BJ1		1
E1655	AV5CAESTGKLIF AJ4	BV29CSVGTGGTNEKLFF BJ1	2	
E1655	AV5CAESTGKLIF AJ37	BV29CSVGTGGTNEKLFF BJ1		3
E1655	AV5CAESPFSGNTPLVF AJ29	BV29CSVGTGGTNEKLFF BJ1	1	
E1655	AV5CAEDLNARLMF AJ31	BV2CAVGTESTNEKLFF BJ1		1
E1655	AV24CAFTNTGNQFYF AJ49	BV3CASSPTSGSVYEQYF BJ2		1

E1655	AV13CAASQIGNEKLTF AJ48	BV5CASSPWDRGATNEKLFF BJ1	1
E1655	AV5CAEPNNAGNMLTF AJ39	BV6CASKTGTGNEKLFF BJ1	1
E1655	AV5CSESNKDGNYQLIW AJ33	BV6SSSADLRTRWLLIKPSL BJ1	1
E1655	AV5CAENLYSTLTF AJ11	BV9CASSATGGDEQFF BJ2	1
E1655	AV12CAMGGSNDYKLSF AJ20	BV9CASSGNPQTGPMNTEAFF BJ1	1
E1655	AV5GTGTDDYKLSF AJ20	BV9CASSPGLVSSGELFF BJ2	1
E1655	AV30CGTEIPHDYKLSF AJ20	BV9CASSTGQLSSGNTIYF BJ1	3
E1655	AV12CAVNPDSSYKLIF AJ12	BV9CASSTGQVATNEKLFF BJ1	1
E1655	AV5CAETVDKLIF AJ34	BV9CASSVVGNEQFF BJ2	1

[1]Color-coded by CDR3 α clones. [2]Color-coded by CDR3 β clones.

Table 9. TCR AV8-CAVKDTDKLIF-AJ34 pairs with multiple different TCRβ within the same individual.

Donor ID	AV	CDR3α sequence (AA)	AJ	CDR3α length	BV	CDR3β sequence (AA)	BJ	CDR3β length
E1603	AV08	CAVKDTDKLIF	AJ34	9	BV24	CATSDAHVNEQFF	BJ02	11
					BV06	CASSYVGLLGDTQYF	BJ02	13
					BV06	CASSSVLDFLGTGELFF	BJ02	15
					BV19	CASRTAGNSDTQYF	BJ02	12
E1632	AV08	CAVKDTDKLIF	AJ34	9	BV10	CASMLPFGDEQYF	BJ01	11
E1651	AV08	CAVKDTDKLIF	AJ34	9	BV30	CAWSGRPSYNEQFF	BJ02	12
					BV24	CATSDWDDSTGELFF	BJ02	13
					BV06	CASSVPTLSDGYTF	BJ01	12
					BV28	CASSSYPGLSTGELFF	BJ02	14
					BV19	CASSSLLISEAFF	BJ01	11
					BV27	CASSRDASSSYNEQFF	BJ02	14
					BV27	CASSRAASSSYNEQFF	BJ02	14
					BV04	CASSNTAGVLGDEQFF	BJ02	14
					BV27	CASSLSMTHTYNEQFF	BJ02	14
E1655	AV08	CAVKDTDKLIF	AJ34	9	BV06	CASSSFRDSSNEQYF	BJ02	13
					BV03	CASSQVIGVGYTF	BJ01	11
					BV27	CASSPSLSDYFTDTQYF	BJ02	15
					BV06	CASSDTLSTGELFF	BJ02	12
					BV06	CARSGTSDTLSYNEQFF	BJ02	15

Chapter 4 Discussion

This is one of a few studies to use deep sequencing to comprehensively analyze the dynamics of both the TCR α and β repertoires to two different epitope-specific CD8 T-cell responses over the course of a primary human EBV infection. Our studies are unique since they examined both CD8 TCR α and β directly *ex vivo* from the peripheral blood of individuals during primary EBV infection (AIM) and again 6-8 months later in convalescence (CONV). Bulk deep sequencing analyses revealed considerable biases. There was prevalent usage of particular gene families such as AV8 within the YVL-specific responses and AV5, AV12 and BV14, BV20 within the GLC-specific populations. Previous reports indicated that the bias for particular gene families can be explained in part by the fact that the specificity of TCR for a pMHC is determined by contacts made between the germline-encoded regions within a V segment and the MHC [64, 112].

Longitudinal analyses of the EBV-specific TCR repertoires over the course of primary infection revealed that the repertoires evolved, consistent with previous reports [86]. There were TCR clonotypes that did or did not persist or were newly recruited. While the TCR repertoires of each epitope were diverse and individualized (*i.e.*, each donor studied had a unique TCR-repertoire), they also consisted of features that were dominant and shared across individuals. These features were more commonly associated with persistent than non-persistent clonotypes. Persistent clonotypes accounted for only 9% of unique clonotypes, but predominated in acute and convalescence phases by accounting for about 57%±4 of the total epitope-specific response. It was previously found that TCR immunodominance patterns seem to scale with the number of specific interactions required between pMHC and TCR [37]. It would seem that TCRs that find simpler solutions to being generated and to recognizing Ag involving fewer specific amino acids are easier to evolve and come to dominate

the T-cell response. Surprisingly, the other 91% of unique clonotypes disappeared following AIM and were replaced in convalescence by "*de-novo*" clonotypes, which accounted for the other 43%±5 of the total response. The fact that only a small subset of TCRs expressing dominant features persisted suggests that these persistent clones may be the best fit TCRs to recognize the pertinent pMHC clonotypes. In fact, Abdel-Hakeem and colleagues [49] examined and compared TCR repertoires between HCV-infected individuals that cleared or did not clear infection and demonstrated that viral clearance was associated with recruitment of highly expanded TCRs. Hence, the features that have been identified to be associated with persistent clonotypes may play a role in asserting the fitness of these clones. In contrast, non-persistent clonotypes that were activated in AIM during high viral load and increased inflammation may not have fitted as well and perhaps did not receive TCR signals that led to survival into memory. Interestingly, 6-8 months after the initial infection, a completely new (*de novo)* and TCR repertoire expanded. This may be due to continued antigenic exposure as may be the case during persistent EBV infection. The recruitment of new TCRs over the course of infection is consistent with previous reports examining TCR repertoire following hepatitis C infection [49]. Altogether, these studies suggest that the pMHC structure may drive selection of particular shared and dominant clonotypes. The broad fluctuating repertoires may lend plasticity to Ag recognition, perhaps assisting in early cross-reactive CD8 T-cell responses to heterologous new pathogens while at the same time potentially protecting against T-cell clonal loss and viral escape [113].

The persistence of a select TCRs may also reflect the fact that EBV establishes a persistent infection leading to recurrent immunological events and hence exhaustion of cells. TCR affinity-dependent compartmentalization of Ag-specific T cells whereby T cells expressing high-affinity TCRs tended to migrate to non-lymphoid tissues to become resident-memory T cells has been reported [45, 47]. In light of this, it is also plausible that cells expressing non-persistent TCRs may have migrated to other tissues and differentiated into resident-memory cells. Although these observations derived from the deep sequencing study could be influenced by sampling and deep-sequencing errors, which are

inherent in most human studies and deep-sequencing assays, we have minimized these errors to the best of our ability by maximizing cell input and reads of deep sequencing and using high-fidelity enzymes during sample preparation. Additionally, we demonstrated that the observed number of clonotypes shared between AIM and CONV was similar with that predicted using the Chao1 statistical analysis, indicating that there were minimal sampling errors in the deep-sequencing TCR dataset.

Despite the loss of the vast majority of the initial pool of clones deployed during acute infection, clonotypic diversity was unaffected and remained high in memory as a result of the recruitment of a diverse pool of new clonotypes. A previous mouse study has shown analogous results. Adoptive transfer of epitope-specific CD8 T-cells of known Vβ family from a single virus-infected mouse to a naive mouse, then challenging with virus resulted in altering the hierarchy of the clonotypes with recruitment of new clonotypes thus maintaining diversity [114]. A highly diverse repertoire should mitigate the effects of T-cell clonal loss [115, 116]. Moreover, a large pool of TCR clonotypes could provide increased resistance to viral escape mutants common in persistent virus infections [113]. Different TCRs may also activate Ag-specific cell functions differently, leading to a more functionally heterogeneous and more complete pool of memory cells [117].

These studies have also uncovered a role for TCRα-driven selection of the YVL-repertoire. AV8.1 was a public gene family, dominating the conserved 9-mer response, with an obligate pairing with AJ34 and a predominant CDR3α motif "VKDTDK", representing 42% and 19% of the total persistent response in AIM and CONV, respectively. In contrast, the TCRβ response was highly diverse without evidence of strongly conserved or shared properties, suggesting that AV8.1-VKDTDK-AJ34 could pair with multiple different TCRβ chains and still successfully be selected by YVL-MHC as seen in the single-cell TCRαβ data. This is consistent with other studies that have shown a TCRα-driven repertoire and provided a structural basis for it [5, 6]. A2-YVL crystal structure showed a highly protuberant negatively charged aspartic acid (D) at position 4 of the YVL epitope

(work of Inyoung Song and reported in Kamga *et al.* 2019, manuscript in review), the side of the pMHC that a TCRα would likely have to accommodate. The fact that YVL-repertoire displayed a preference for TCRs expressing AV8.1-VKDTDK-AJ34, containing the positively charged K in position 2 would be consistent with this requirement. Future structural analysis would be important to ascertain whether the YVL TCRα contributes the majority of contacts with the pMHC. Unlike YVL, the selection of GLC-specific TCRs appeared to be driven by interactions with both TCRα/β, such as AV5.1-EDNNA-AJ31, BV14-SQSPGG-BJ2 and BV20.1-SARD-BJ1, which are previously identified shared features [86, 105, 118]. This observation is consistent with the structural examination of a GLC-specific TCR in complex with its ligand, which demonstrates that recognition of GLC is driven by both TCR chains [64].

Given that the deep sequencing data suggest that the CD8 T-cell repertoire to the immunodominant human viral epitope YVL appeared to be biased towards TCRα usage and revealed a ubiquitous and public TCRα clonotype, AV8.1-VKDTDK-AJ34, that appeared to be a selection factor for persistence into the memory phase, we went on to investigate the potential basis for the high selection of this clonotype. We applied single-cell TCR$\alpha\beta$ sequencing on YVL-specific CD8 T cells to study the TCR$\alpha\beta$ repertoires at the clonal level and compared the repertoire to that of GLC in the same donors. Consistent with deep sequencing, single-cell analyses confirmed that each epitope behaved differently and there were prominent features that were significantly enriched and shared across individuals. Moreover, YVL was significantly biased towards TCRα usage and the ubiquitous and public TCRα chain, AV8.1-VKDTDK-AJ34, was highly degenerate and paired with multiple different TCRβ chains in the same donor. In contrast and consistent with prior reports, GLC-specific CD8 T-cell repertoires were notably skewed in both TCRα and TCRβ usage with a bias towards use of the public AV5-AJ31-BV20-BJ1.2 combination.

Prominent biases have been noted in the TCR repertoires of various common virus-specific CD8 T cells (*e.g.* influenza A virus, CMV, hepatitis C virus) [34, 37, 49, 53, 56, 58, 60]. For example, Chen and colleagues [34] have studied the TCR repertoires of two HLA-A2-restricted epitopes: $pp65_{495-503}$ (NLV) of cytomegalovirus and $M1_{58-67}$ (GIL) of influenza A virus. They have shown that NLV-specific TCRs predominantly used AV26-2, AV24, BV7-6 and BV12-3 while GIL-specific TCRs predominantly used BV19. GIL-specific TCRs using BV19 were highly degenerate and could paired with multiple different TCRα chains. Structural analyses revealed that this degeneracy is likely underpinned by the fact that the TCRβ chain makes the most contacts with the ligand. T-cell responses to some EBV and yellow fever Ags have a Vα repertoire that is more restricted than their Vβ repertoire, and in some instances contains shared CDR3α motifs, which would suggest a role for the CDR3 in Ag recognition. For example, the TCR repertoire of the EBV BZLF1-derived 13-mer peptide (LPE) restricted by HLA-B*35:08 is biased towards AV19 in many individuals and displays a strong preservation of a "GF" motif within its CDR3α region that pair with different TCRβ sequences [5, 6]. Examination of the crystal structure of this TCR with its ligand has provided a structural basis for the selection of this public CDR3α motif [6]. The LPE peptide bulges out of the cleft of the HLA-B*35:08 causing the TCRα chain to make the most contributions to the pMHC. Interestingly, despite being highly conserved, the "GF" motif does not make direct contact with the peptide and instead stabilizes the peptide-binding cleft. Similar to B35/LPE, the HLA-A*02-restricted epitope LLWWNGPMAV (A2/LLW) from the yellow fever virus has a biased Vα repertoire that is focused on the AV12.2 [119]. TCR modeling of an A2/LLW-specific TCR with its ligand suggests that this high preference for the AV12.2 is likely due to pronounced interactions between the CDR1α and the pMHC.

In contrast, the TCRβ repertoire of the HLA-A02:01-restricted $M1_{58-67}$ (GIL) epitope from influenza A virus is more restricted than its TCRα repertoire and displays a high bias towards the

BV19 in many individuals with a strong preservation of a dominant xRSx CDR3β motif. Crystal structures of TCR specific to this epitope have revealed that the TCR is β-centric with residues of the BV19-encoded CDR1 and CDR2 loops engaging pMHC and the conserved arginine in the CDR3β loop being inserted into a pocket formed between the peptide and the α2-helix of the HLA-A02:01 [35, 37]. The TCRα is not as important as the TCRβ in pMHC engagement and this helps explain the high degree of sequence conservation in the TCRβ and the variability in the TCRα. Similarly, studies using EBV virus GLC-specific CD8 T cells have documented that TCR-pMHC binding modes also contribute to TCR biases. Miles and colleagues [64] showed that the highly public AS01 TCR, which is specific to the HLA-A*02:01-restricted EBV-derived GLC epitope, was highly selected by the GLC epitope because of a few very strong interactions of its TRAV5- and TRBV20-encoded CDR3 loops with the peptide/MHC.

Given the aforementioned studies, we reasoned that the differences in constraints in the TCR repertoires of YVL and GLC may reflect an essential requirement of Ag recognition and that the topology of the pMHC may provide some structural insights into the mechanisms underlying these constraints. This alpha-centricity displayed by the YVL repertoire might be grounded in the fact that the TCRα chain potentially makes more pronounced contact with its ligand, the pMHC, compared to the TCRβ chain. In light of this apparent TCRα bias in the YVL repertoire and the substantial evidence that TCR-pMHC binding modes also contribute to TCR biases, we hypothesized that the bias for TCRα in the YVL repertoire most likely reflects an inherent feature of how the YVL peptide lies in the HLA-A2 groove or the topology of the pMHC. To understand why the YVL epitope drives the selection of this highly conserved public TCRα, we performed single-cell paired TCRαβ sequencing and applied the Dash *et al.* algorithm [53] to identify key amino acid residues that might be important for Ag recognition. We uncovered the highly conserved "KD" amino acid pair within the AV8.1-VKDTDK-AJ34 clonotype as being potentially critical for recognition of the YVL epitope.

Conserved residues suggest a structural basis for Ag recognition. Structural analysis of A2-YVL (work of Inyoung Song and reported in Kamga *et al.* manuscript in review) revealed that the MHC-bound YVL bulged at position 4, in a region of the peptide that TCRα would have to accommodate, exposing a negatively charged Asp(D). We cloned a TCR expressing the dominant public TCRα chain AV8.1-VKDTDK-AJ34 and using mutagenesis, we provided evidence indicating that the positively charged Lys(K) residue at position 2 in the CDR3α of the public TCRα chain was important for A2-YVL recognition. Hence, we propose that the selection of AV8.1-VKDTDK-AJ34, which contains the positively charged (Lys)K in position 2, is potentially driven by an electrostatic interaction between this Lys(K) and the solvent-exposed Asp(D) on the peptide. Future structural analyses of this TCRαβ with its ligand would be important to validate the existence of this electrostatic interaction and to confirm whether the TCRα contributes the majority of contacts with the pMHC. This apparent preference for TCRα may create a large repertoire of different memory TCRβ that could potentially cross-react with other ligands such as IAV-M1$_{58}$, which predominantly interact with the TCRβ chain [28, 37, 53].

Others and we have been interested in understanding the factors that leads to the persistence of CD8 T cells into the memory phase. Multiple factors have been shown to contribute to this. For example, the inflammatory milieu is thought to play a role in the generation of memory CD8 T cells; cytokines such as IL7 and IL15 have been shown to favor memory formation while IL12 inhibits it [39-42]. The transcriptions factors EOMES promotes memory formation whereas T-bet promotes effector differentiation [38-40]. The TCR signal strength, which is dictated by the TCR sequence, could also influences memory formation. For example, Teixeiro *et al.* [12] showed that a point mutation in the TCR of ovalbuminin-specific CD8 T cells impaired their ability to differentiate into memory CD8 T cells but maintained the ability to become effector cells, likely due to a decreased recruitment of TCRs to the immunological synapse and decreased signaling. These data raise the possibility of TCR-dependent differences in directing cell fate. Moreover, TCR affinity maturation, whereby you have the

selection into the memory pool of TCR clones having an affinity falling within a particular range, has been documented during T-cell responses to multiple human viral infections. Abdel-Hakeem *et al.* [49] showed that during a hepatitis C virus (HCV) infection, the initial TCR repertoire evolves and narrows in breadth into the memory phase. Similar findings were observed in other infections [48, 50]. These studies suggest that there may be optimal TCRs, which may be better fit to induce memory T-cell differentiation. The current work focuses on investigating features of the TCR repertoire that leads to the persistence of EBV-specific CD8 T cells into the memory phase. To address this matter, we examined the TCR repertoire at two time points over the course of infection: between AIM (which falls within the expansion phase of the immune response) and CONV, which falls within the memory phase for two HLA-A2-restricted EBV-derived epitopes, YVL and GLC. Ag-driven selection of TCR clonotypes can lead to a skewed V gene distribution. We sought to determine whether this was the case for YVL and GLC responses. We observed that the YVL response displayed a skewed Vα gene distribution with preferential use of AV8.1 whereas the GLC response displayed a skewed Vα and Vβ gene distributions with preferential use of AV5, AV12 and BV14, BV20. The preferential use of a restricted set of genes across individuals provided support for an Ag-driven selection of YVL- and GLC-specific TCR clonotypes and suggest that these genes may be important for TCR selection and recognition. This phenomenon has been observed with others epitopes such as HLA-A2-restricted epitopes pp65$_{495-503}$ (NLV) of cytomegalovirus, M1$_{58-67}$ (GIL) of influenza A virus and LLW of yellow fever and B48-restricted BZLF1 of EBV [34, 37, 53, 119]. NLV-specific TCRs predominantly used AV26-2, AV24, BV7-6 and BV12-3 while GIL-specific TCRs predominantly used BV19 [34, 37]. Furthermore, Ag-driven selection can also leads to conservation of amino acid in the CDR3 regions. Indeed, we observed that the YVL response during AIM and CONV displayed dominant and conserved motifs within its CDR3α (AV8.1-VKDTDKL-AJ34) but a diverse CDR3β. On the other hand, the GLC response displayed dominant and conserved motifs within both its CDR3α (*e.g.* AV5.1-EDNNA-AJ31) and CDR3β (*e.g.* BV20.1-SARD-BJ1).

To uncover features that likely contributed to persistence into the memory phase, we compared TCR repertoires between AIM and CONV. Longitudinal analyses of the EBV-specific TCR repertoires over the course of primary infection revealed that the repertoires evolved, consistent with previous reports [86]. There were TCR clonotypes that did (persistent) or did not (non-persistent) persist or were newly recruited (*de novo*). Persistent TCRs more commonly expressed the dominant CDR3 motifs that were identified during AIM and CONV compared to non-persistent TCRs. For example, persistent YVL-specific TCR α clonotypes expressed the CDR3α motif AV8.1-VKDTDK-AJ34 more commonly than non-persistent ones. In contrast, the TCRβ response was diverse without evidence of strongly conserved or shared properties, suggesting that AV8.1-VKDTDK-AJ34 may contribute to the persistence of T cells into the memory phase, likely by mediating optimal Ag-recognition and TCR signaling leading to persistence. This is consistent with other studies showing a TCRα-driven selection of repertoire [5, 6]. For example, BZLF1 response displayed a TCRα-centricity and this was underpinned by the fact that the chain was more involved in Ag-recognition compared to the TCR β chain [5, 6]. In light of this, we posit that the structural constraints in the TCRα may reflect a requirement in Ag-recognition, leading to a signaling cascade optimal for clonal persistence. On the other hand, the persistence of GLC-specific TCRs appeared to be driven by a combination of both TCRα/β features. Persistent GLC-specific TCRα/β clonotypes expressed the CDR3 motifs AV5.1-EDNNA-AJ31, BV14-SQSPGG-BJ2 and BV20.1-SARD-BJ1 more commonly than non-persistent ones. This corroborates an earlier study demonstrating that recognition of GLC is indeed driven by both TCR chains [64]. So we posit that in the case of GLC, both chains maybe important to form the optimal connections and signaling leading to clonal persistence. In a study comparing TCR repertoires between individuals that controlled or did not control a chronic infection, it was found that viral control following re-exposure to the pathogen was associated with the recruitment from the memory pool of clonally expanded clones that were polyfunctional. These findings are consistent with our observations that clones expressing dominant features tended to persist into the memory phase

and prompted us to speculate that these clonally expanded clonotypes might elicit a more polyfunctional response than the diverse non-persistent clonotypes. High-affinity TCRs have been linked to T-cell exhaustion [120]. We speculate that the observed clonal loss may be a reflection of TCR affinity and that non-persistent cells likely express high-affinity TCRs than persistent ones. The emergence of new TCR clonotypes may reflects the fact that EBV establishes a persistent infection leading to continued antigenic exposure and recurrent immunological events. The recruitment of new TCRs over the course of infection is consistent with previous reports examining TCR repertoire following hepatitis C infection [49]. The aforementioned speculations will need to be investigated. In the future, we would like to test the role of TCR affinity in clonal persistence. Do persistent and non-persistent TCRs display distinct affinities for their cognate ligands and/or elicit different functional profiles?

In summary, these data reveal that apparent molecular constraints are associated with TCR selection and persistence in the context of controlled viral replication following primary EBV infection. The gathered evidence provide support for the proposed model that persistence of YVL-specific TCR clonotypes into the memory phase is driven by TCRα chain whereas that of GLC is driven by both chains (**Fig 29**). The ability to design efficacious T-cell based therapies or vaccines is predicated on our ability to identify markers associated with clonal persistence. Hence, our findings will contribute to growing interests and efforts (*e.g.* autologous EBV-specific T cells are being tested in a phase 2 clinical trial as a therapy for EBV-positive NK/T-cell lymphoma (study name: CITADEL; NCT01948180)) focused on the fundamental goal of rationally designing better T-cell based therapies and vaccines by empowering researchers to predict treatments that are more likely to elicit desirable and potent responses [24, 53, 54, 97-99].

1 The challenge of mitigating errors in deep-sequencing TCR dataset

High-throughput sequencing is a multi-step process involving PCR amplification, DNA sequencing and in some cases cDNA synthesis. The error-prone nature of the enzymes involves in some of the steps renders high-throughput sequencing inherently susceptible to introducing errors in the generated nucleotide sequences. These errors are inevitable and can lead to an overestimation of the richness of the TCR repertoire. There are currently no reliable and conventional approaches to purge sequences dataset of erroneous sequences. Preventive measures to mitigate the occurrence of these errors are taken during sample preparation and sequencing. This includes the use of high-fidelity enzymes having low error rate and filtering and trimming sequence reads having a low quality score. It is important to note that the quality score-based filtering mostly gets rid of errors introduce during base calling of DNA sequencing but will not pick up potential errors introduced in earlier steps during PCR amplification or cDNA synthesis. To deal with these precocious errors, some have proposed to filter sequences based on frequency threshold, *i.e.*, excluding low-frequency TCRs because these are more likely to contain error [121]. This method has been shown to be ineffective in regards to TCR sequencing for multiple reasons: (1) most TCRs are rare; therefore this method would eliminate these scarce TCRs and deflate the richness of the repertoire; (2) an error introduced early on during PCR amplification will be multiplied exponentially and become dominant; hence this approach is biased in favor of high-frequency TCRs and will fall short of excluding erroneous sequences having a high frequency; (3) There is no empirical way to derive a threshold and this value will vary from sample to sample. Given these reasons, we decided not to filter out our TCR sequence dataset based on a frequency dataset.

An innovative and reliable strategy to mitigate the occurrence of errors in deep sequencing includes the addition of unique molecular identifiers (UMIs; A UMI is a short and unique stretch of DNA added to a piece of DNA template during library preparation and prior to PCR amplification to track DNAs sequences that arise from the same parental template based on the fact that they share

the same UMI) [122]. UMIs are also advantageous because they can be used for precise quantification of TCR sequences. Unfortunately, we did not incorporate UMIs in our deep sequencing work to be able to take advantages of its benefits of identifying and filtering out erroneous sequences. Future TCR repertoire analyses of EBV-specific T cells, or any T-cell populations in general, should consider integrating such innovative approach to not only mitigate sequencing errors but to also provide a more accurate quantification of TCR clonotypes.
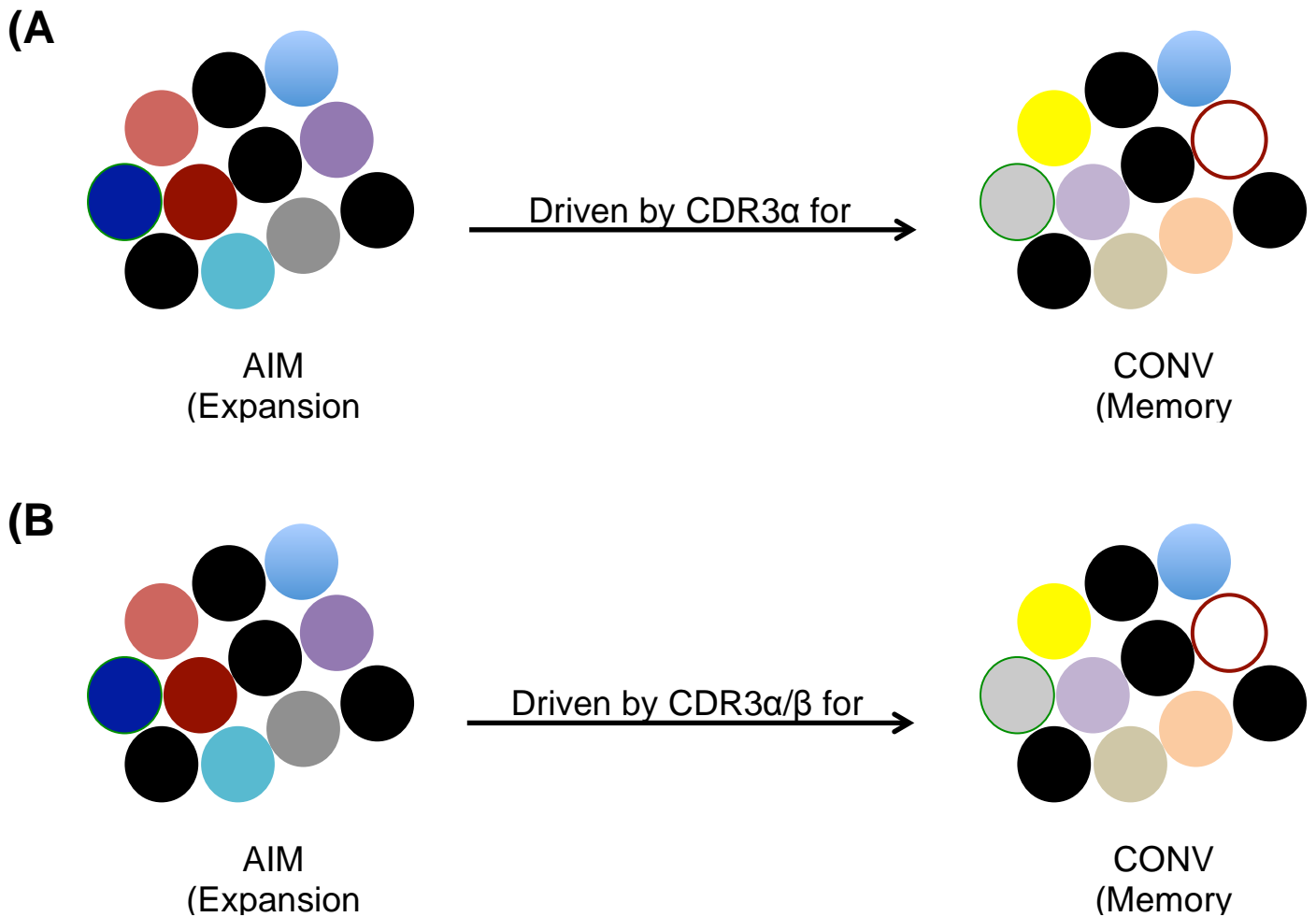
Fig 29. Proposed model for the persistence of YVL- and GLC-specific CD8 T cells into the memory phase following acute EBV infection. Following EBV infection (AIM), a diverse pool of YVL- (A) or GLC- (B) specific CD8 T cells expressing unique TCRs are deployed to control viral replication. However, only ~9% of these unique TCRs persist into the convalescence phase 6-8 months later (CONV; memory phase) and the other 91% disappear and are replaced with new clonotypes. Persistent cells are highly expanded and persistent TCRs display Ag-driven features. These features are more prevalent in the CDR3α region for YVL-specific cells (A) and in both the CDR3α and CDR3β regions for GLC-specific cells suggesting that the TCRα chain drives the persistence of YVL-specific cells whereas both chains drive that of GLC. The emergence of a new set of TCRs is likely due to ongoing antigenic stimulation. Each circle depicts an epitope-specific cell and is color-coded for expressing a unique TCR. Black circles denote persistent TCRs (present during both AIM and CONV) and the other colors denote either non-persistent (present only during AIM) or new TCRs (present only during CONV).

7 Materials & Methods

1 Study population

Four HLA-A*02:01+ individuals presenting with symptoms and laboratory studies (positive serum heterophile antibody and the detection of EBV viral capsid antigen (VCA)-specific IgM) consistent with AIM were studied (**Table S1**) at initial clinical presentation (AIM) and 6-8 months later (Convalescence; CONV)). Blood samples were collected in heparinized tubes at clinical presentation with AIM symptoms (acute phase) and six months later (memory phase). PBMC were extracted by Ficoll-Paque density gradient media and cryopreserved until needed. Direct tetramer staining of peripheral blood revealed that 2.7%±0.7 (mean±SEM) and 1.3%±0.3 of CD8 T-cells were YVL- and GLC-specific, and declined to 0.3%±0.7 and 0.3%±0.1, respectively, in CONV. Mean blood EBV load was 3.9±0.7 $\log_{10}$ and 1.7±1.0 $\log_{10}$ genome copies/10^6 B cells during AIM and CONV, respectively.

The Institutional Review Board of the University of Massachusetts Medical School approved these studies. All participants were adults and provided written informed consent.

2 EBV DNA quantitation in B cells

B cells were purified from whole blood using the RosetteSep human B-cell enrichment cocktail according to the manufacturer's recommendations (StemCell Technologies, Vancouver BC, Canada). Cellular DNA was extracted using QIAGEN DNeasy Blood & Tissue Kit (Valencia, CA). Each DNA sample was diluted to 5ng/ul and the Roche LightCycler EBV Quantitation Kit (Roche Diagnostics, Indianapolis, IN) was used to quantify EBV DNA copy number in the samples as recommended by the manufacturer. Reactions were run in duplicate. B cell counts in each sample were determined

using a previously described PCR assay to quantify the copy number of the gene encoding CCR5 (two copies per diploid cell) [123]. Samples were normalized to B cell counts and EBV DNA copy number was calculated as DNA copy per 10^6 B cells.

3 Identification and isolation of YVL- and GLC-specific CD8 T cells by tetramer staining and FACS

The percentages of peripheral blood antigen-specific CD8 T-cells were enumerated by tetramer staining and flow cytometry. Antibodies were purchased from BD Biosciences and included: anti-CD3-FITC, anti-CD4-AF700 and anti-CD8-BV786, 7AAD and PE-conjugated HLA-A*02:01-peptide tetramers (BRLF-1$_{109-117}$: **YVL**DHLIVV, A2-YVL; BMLF-1$_{280-288}$: **GLC**TLVAML, A2-GLC). Tetramers were made in-house and underwent quality assurance, as previously described [124]. Total CD8 T-cells were enriched from PBMCs by positive selection using MACS technology (Miltenyi Biotec, Auburn, CA) according to the manufacturer's protocol. The cells were then stained with anti-CD3, anti-CD4, anti-CD8, 7AAD, and GLC- or YVL-loaded tetramers for 30 minutes on ice in staining buffer (1% BSA in PBS). Single-cell (into 384-well PCR plates) or bulk sorting of live CD3+, CD8+, and GLC- or YVL-tetramer+ cells using FACSAria III (BD) was performed for subsequent TCR sequencing.

4 Analyses of the TCRα/β repertoires of bulk EBV-specific CD8 T cells using deep sequencing

mRNA was isolated from the bulk sorted EBV-specific CD8 T cells and cDNA synthesized. The cDNA samples were sent to Adaptive Biotechnologies for TCR α and β sequencing using the ImmunoSeq platform as per the vendor's protocols and standards for sequencing and error correction. Briefly, amplification of CDR3 regions was done by multiplex PCR using primers that

annealed to the V and J regions. PCR products were used to prepare a library for sequencing. Sequences were corrected for errors using known error rates of sequencers. Sequences of CDR3 regions were identified according to the definition founded by the International ImMunoGeneTics collaboration. Deep sequencing data of TCR α and β repertoires were analyzed using ImmunoSEQ Analyzer versions: 2.0 and 3.0, which were provided by Adaptive Biotechnologies. Only productively (without stop codon) rearranged TCR α and TCR β sequences were used in analyses.

Circos plots and motif analysis. The V and J gene segment combinations were illustrated as circos plots [125] across different CDR3 aa sequence lengths. Motif analysis was performed using the Multiple EM for Motif Elicitation (MEME) framework [126]. Consensus motifs were acquired across different CDR3 lengths and statistics on those motifs were computed with an in-house program called motifSearch and available at http://github.com/thecodingdoc/motifSearch.

5 Single-cell analyses of the paired TCR $\alpha\beta$ repertoires of EBV-specific CD8 T-cells

To examine TCR α and TCR β pairing relationships, we conducted an *ex vivo* single-cell analysis of the paired TCR $\alpha\beta$ repertoire of YVL and GLC-specific CD8 T-cells from PBMCs of 4 donors in AIM and CONV. Single tetramer-positive CD8 T cells were sorted into a 384-well plate containing 0.5 ul/well of lysis buffer, consisting of a mixture of 0.03ulTriton X 100 (Sigma, St. Louis, MO), 0.01ul RNase inhibitor (Clontech, Mountain View, CA), 0.12ul 3' SMART CDS Primer IIA (Clontech, Mountain View, CA), 0.35ul water, using the FACSARIA (BD Biosciences, San Jose, CA). The cells were lysed by incubating at 72 ^{o}C for 3min, 4 ^{o}C for 10min, 25 ^{0}C for 1min. Reverse transcription and global amplification of full-length cDNA were performed using the SMARTer kit (Clontech, Mountain View, CA). Briefly, reverse transcription reaction mixture consisted of the following: the lysed sample, 0.03ul C1 loading reagent, 0.23ul 5X first-strand buffer, 0.03ul DTT,

0.12ul dNTP mix, 0.12ul SMARTer IIA oligonucleotide, 0.03ul Rnase inhibitor and 0.12ul SMARTScribe reverse transcriptase. The reaction mixture was incubated at 42 ^{o}C for 90min and 72 ^{o}C for 10min. We proceeded to global amplification of full-length cDNA by adding 3ul water, 0.5ul 10X Advantage 2 PCR Buffer, 0.2ul 50X dNTP Mix, 0.1ul IS PCR primer and 50X Advantage 2 Polymerase mix. The samples were incubated as follows: 1 cycle of [1min at 95 ^{o}C], 5 cycles of [20sec at 95 ^{o}C, 4min at 58 ^{o}C and 6min at 68 ^{o}C], 9 cycles of [20sec at 95 ^{o}C, 30sec at 64 ^{o}C and 6min at 68 ^{o}C], 7 cycles of [30sec at 95 ^{o}C, 30sec at 64 ^{o}C and 7min at 68 ^{o}C] and a final elongation at 72 ^{o}C for 10min. The synthesized cDNA samples were diluted by adding 2.3ul of water.

Amplification of paired CDR3α and CDR3β regions was performed as previously described [127] by multiplex-nested using the synthesized cDNA as template and Taq Polymerase kit (Qiagen). 1st round of PCR reaction mixture consisted of 0.5ul 20uM of TRAV external primers (**Table 10**), 0.5ul 5uM TRAC external primers (**Table 10**), 0.5ul 20uM of TRBV external primers (**Table 11**), 0.5ul 5uM TRBC external primers (**Table 11**), 0.5ul 10mM dNTP, 2.5ul 10X buffer, 1ul of the diluted cDNA sample, 0.15ul 5U/ul Taq polymerase and water to a final volume of 25ul. The reaction mixture was incubated as follows: 1 cycle of [2min at 95 ^{o}C], 35 cycles of [20sec at 95 ^{o}C, 20sec at 52 ^{o}C and 45sec at 72 ^{o}C] and a final elongation at 72 ^{o}C for 10min. A second round of PCR was done to amplify the CDR3α or CDR3β separately using the same reaction mixture described above and using as template 2.5ul of 1st round of PCR reaction and the respective internal primers (**Tables 10-11**). CDR3α and CDR3β products were purified using Exonuclease I (Affymetrix) according to the manufacturer᾽s guidelines and then Sanger sequencing. Sequences were analyzed according to the IMGT/V-QUEST web-based tool [15]. Only productively rearranged CDR3α and CDRβ sequences without stop codons will be used for repertoire analyses.

6 TCR cloning

Full-length TCR α or TCR β chains of the WT TCR (AV8.1-CAVKDTDKLIF-JA34/BV24.1-CATSDWDDSTGELFF-BJ2.2) were amplified using the synthesized cDNA (described above) as template, a TRAV8.1- or TRAV24.1-specific forward primer (specific to the corresponding V α or V β gene segment identified by single-cell CDR3 sequencing) and a TRAC- or TRBC-specific reverse primer (**Table 12**). The PCR reaction was performed using the Primestar Polymerase kit (Clontech) and the reaction mixture consisted of 10ul of 2X Primestar buffer, 0.4ul 10mM dNTPs, 0.6ul of each 10uM primer, 0.1ul 2.5U/ul Primestar polymerase and water up to 20ul. The PCR reaction was run as follows: 1 cycle of [1min at 98 ^{o}C] and 30 cycles of [10sec at 98 ^{o}C, 5sec at 55 ^{o}C and 1min at 72 ^{o}C]. The alpha chain was linked to the beta chain via the viral P2A sequence for stoichiometric expression and subcloned by homologous recombination into the pscALPS lentiviral vector (a gift from Dr. Jeremy Luban) using GeneArt Assembly Kit (Invitrogen) according to the manufacturer's recommendations. The mutant TCR was synthesized by site-directed mutagenesis of the WT TCR.

7 Lentivirus transduction of TCR

293T (CRL-3216, ATCC) and J76 cells (a TCR-deficient CD8α-expressing Jurkat cell line [128]; a gift from Dr. Wolfgang Uckert) were maintained in DMEM and RPMI, respectively, containing 10% FBS, HEPES and L-glutamine. Viral packaging was performed in 293T cells using the following three vectors (gifts from Dr. Jeremy Luban): recombinant lentiviral vector containing the cloned TCR gene, an HIV-1 Gag-Pol packaging plasmid (psPAX2) and a VSV-G plasmid (pMD2.G), and the Lipofectamine 2000 Transfection Reagent Kit (Invitrogen, Carlsbad, CA) as recommended by the manufacturer. J76 cells were re-suspended at a density of 1 million cells per 100ul of media and 100ul of cell suspension were transferred into each well of a 12-well plate. Infection was performed by adding 250ul of the harvested virus supernatant and the cells were placed in an incubator (37 ^{o}C, 5%

CO_2) for at least 6 hours, following which 2 mL of media was added for overnight growth. The following day, the cells were harvested, washed and re-infected as described above. The cells were expanded and tetramer staining of the TCR-transduced J76 cells was performed 5 days post infection to validate the specificity of the overexpressed TCRs.

8 CD69 upregulation

HLA-A2-expressing T2 Ag-presenting cells (CRL-1992, ATCC) were maintained in RPMI containing 10% FBS, HEPES and L-glutamine. T2 cells were pulsed with the desired peptide concentration overnight at 37°C. The cells were then washed to remove excess peptide. Peptide-pulsed T2 cells and TCR-transduced J76 cells were co-cultured at a 1:1 ratio (for a total of 500,000 cells in 200ul of cell culture media per well of a 96-well plate) for 90 minutes at 37 ^{0}C. Cells were harvested and stained with anti-CD8, anti-CD3 and anti-CD69 for 30 minutes on ice in staining buffer. Cells were washed twice and analyzed via flow cytometry using a Calibur (BD Biosciences). Assay was performed in duplicate.

9 Statistics

GraphPad Prism (GraphPad Software, La Jolla, CA) was used for all statistical analyses. The online tool EstimateS (version 9.0.0; available at http://viceroy.eeb.uconn.edu/estimates/) was used to predict the number of TCR clonotypes shared between AIM and CONV.

Table 10. Primers for single-cell CDR3α amplification

TRA gene(s) targeted by primer	External primer sequence	TRA gene(s) targeted by primer	Internal primer sequence
TRAV1	5' AACTGCACGTACCAGACATC 3'	TRAV1-Nest	5' GCACCCACATTTCTKTCTTAC 3'
TRAV2	5' GATGTGCACCAAGACTCC 3'	TRAV2-Nest	5' CACTCTGTGTCCAATGCTTAC 3'
TRAV3	5' AAGATCAGGTCAACGTTGC 3'	TRAV3-Nest	5' ATGCACCTATTCAGTCTCTGG 3'
TRAV4	5' CTCCATGGACTCATATGAAGG 3'	TRAV4-Nest	5' ATTATATCACGTGGTACCAACAG 3'
TRAV5	5' CTTTTCCTGAGTGTCCGAG 3'	TRAV5-Nest	5' TACACAGACAGCTCCTCCAC 3'
TRAV6	5' CACCCTGACCTGCAACTATAC 3'	TRAV6-Nest	5' TGGTACCGACAAGATCCAG 3'
TRAV7	5' AGCTGCACGTACTCTGTCAG 3'	TRAV7-Nest	5' ACAATTTGCAGTGGTACAGG 3'
TRAV8-1	5' CTCACTGGAGTTGGGATG 3'	TRAV8-1-Nest	5' GTCAACACCTTCAGCTTCTC 3'
TRAV8-2, 8-4	5' GCCACCCTGGTTAAAGG 3'	TRAV8-2, 8-4-Nest	5' AGAGTGAAACCTCCTTCCAC 3'
TRAV8-3	5' CACTGTCTCTGAAGGAGCC 3'	TRAV8-3-Nest	5' TTTGAGGCTGAATTTAAGAGG 3'
TRAV8-6	5' GAGCTGAGGTGCAACTACTC 3'	TRAV8-6-Nest	5' AACCAAGGACTCCAGCTTC 3'
TRAV8-7	5' CTAACAGAGGCCACCCAG 3'	TRAV8-7-Nest	5' ATCAGAGGTTTTGAGGCTG 3'
TRAV9-1, 9-2	5' TGGTATGTCCAATATCCTGG 3'	TRAV9-1, 9-2-Nest	5' GAAACCACTTCTTTCCACTTG 3'
TRAV10	5' CAAGTGGAGCAGAGTCCTC 3'	TRAV10-Nest	5' GAAAGAACTGCACTCTTCAATG 3'
TRAV12-1, 12-2, 12-3	5' CARTGTTCCAGAGGGAGC 3'	TRAV12-1, 12-2, 12-3-Nest	5' AAGATGGAAGGTTTACAGCAC 3'
TRAV13-1	5' CATCCTTCAACCCTGAGTG 3'	TRAV13-1-Nest	5' TCAGACAGTGCCTCAAACTAC 3'
TRAV13-2	5' CAGCGCCTCAGACTACTTC 3'	TRAV13-2-Nest	5' CAGTGAAACATCTCTCTCTGC 3'
TRAV14	5' AAGATAACTCAAACCCAACCAG 3'	TRAV14-Nest	5' AGGCTGTGACTCTGGACTG 3'
TRAV16	5' AGTGGAGCTGAAGTGCAAC 3'	TRAV16-Nest	5' GTCCAGTACTCCAGACAACG 3'
TRAV17	5' GGAGAAGAGGATCCTCAGG 3'	TRAV17-Nest	5' CCACCATGAACTGCAGTTAC 3'
TRAV18	5' TCCAGTATCTAAACAAAGAGCC 3'	TRAV18-Nest	5' TGACAGTTCCTTCCACCTG 3'
TRAV19	5' AGGTAACTCAAGCGCAGAC 3'	TRAV19-Nest	5' TGTGACCTTGGACTGTGTG 3'
TRAV20	5' CACAGTCAGCGGTTTAAGAG 3'	TRAV20-Nest	5' TCTGGTATAGGCAAGATCCTG 3'
TRAV21	5' TTCCTGCAGCTCTGAGTG 3'	TRAV21-Nest	5' AACTTGGTTCTCAACTGCAG 3'
TRAV22	5' GTCCTCCAGACCTGATTCTC 3'	TRAV22-Nest	5' CTGACTCTGTGAACAATTTGC 3'
TRAV23	5' TGCTTATGAGAACACTGCG 3'	TRAV23-Nest	5' TGCATTATTGATAGCCATACG 3'
TRAV24	5' CTCAGTCACTGCATGTTCAG 3'	TRAV24-Nest	5' TGCCTTACACTGGTACAGATG 3'
TRAV25	5' GGACTTCACCACGTACTGC 3'	TRAV25-Nest	5' TATAAGCAAAGGCCTGGTG 3'
TRAV26-1	5' GCAAACCTGCCTTGTAATC 3'	TRAV26-1-Nest	5' CGACAGATTCACTCCCAG 3'
TRAV26-2	5' AGCCAAATTCAATGGAGAG 3'	TRAV26-2-Nest	5' TTCACTTGCCTTGTAACCAC 3'
TRAV27	5' TCAGTTTCTAAGCATCCAAGAG 3'	TRAV27-Nest	5' CTCACTGTGTACTGCAACTCC 3'
TRAV29	5' GCAAGTTAAGCAAAATTCACC 3'	TRAV29-Nest	5' CTGCTGAAGGTCCTACATTC 3'
TRAV30	5' CAACAACCAGTGCAGAGTC 3'	TRAV30-Nest	5' AGAAGCATGGTGAAGCAC 3'
TRAV34	5' AGAACTGGAGCAGAGTCCTC 3'	TRAV34-Nest	5' ATCTCACCATAAACTGCACG 3'
TRAV35	5' GGTCAACAGCTGAATCAGAG 3'	TRAV35-Nest	5' ACCTGGCTATGGTACAAGC 3'
TRAV36	5' GAAGACAAGGTGGTACAAAGC 3'	TRAV36-Nest	5' ATCTCTGGTTGTCCACGAG 3'
TRAV38-1, 38-2	5' GCACATATGACACCAGTGAG 3'	TRAV38-1, 38-2-Nest	5' CAGCAGGCAGATGATTCTC 3'
TRAV39	5' CTGTTCCTGAGCATGCAG 3'	TRAV39-Nest	5' TCAACCACTTCAGACAGACTG 3'
TRAV40	5' GCATCTGTGACTATGAACTGC 3'	TRAV40-Nest	5' GGAGGCGGAAATATTAAAGAC 3'
TRAV41	5' AATGAAGTGGAGCAGAGTCC 3'	TRAV41-Nest	5' TTGTTTATGCTGAGCTCAGG 3'
TRAC	5' GACCAGCTTGACATCACAG 3'	TRAC-Nest	5' TGTTGCTCTTGAAGTCCATAG 3'

Table 11. Primers for single-cell CDR3β amplification

TRB gene(s) targeted by primer	External primer sequence	TRB gene(s) targeted by primer	Internal primer sequence
TRBV2	5' TCGATGATCAATTCTCAGTTG 3'	TRBV2-Nest	5' TTCACTCTGAAGATCCGGTC 3'
TRBV3-1	5' CAAAATACCTGGTCACACAG 3'	TRBV3-1-Nest	5' AATCTTCACATCAATTCCCTG 3'
TRBV4-1, 4-2, 4-3	5' TCGCTTCTCACCTGAATG 3'	TRBV4-1, 4-2, 4-3-Nest	5' CCTGCAGCCAGAAGACTC 3'
TRBV5-1, 5-3, 5-4	5' GATTCTCAGGKCKCCAGTTC 3'	TRBV5-1, 5-3, 5-4-Nest	5' CTTGGAGCTGGRSGACTC 3'
TRBV5-5, 5-6, 5-7, 5-8	5' GTACCAACAGGYCCTGGGT 3'	TRBV5-5, 5-6, 5-7, 5-8-Nest	5' TCTGAGCTGAATGTGAACG 3'
TRBV6-1, 6-2, 6-3, 6-5, 6-6, 6-7, 6-8, 6-9	5' ACTCAGACCCCAAAATTCC 3'	TRBV6-1, 6-2, 6-3, 6-5, 6-6, 6-7, 6-8, 6-9-Nest	5' GTGTRCCCAGGATATGAACC 3'
TRBV6-4	5' ACTGGCAAAGGAGAAGTCC 3'	TRBV6-4-Nest	5' TGGTTATAGTGTCTCCAGAGC 3'
TRBV7-1, 7-2, 7-3	5' TRTGATCCAATTTCAGGTCA 3'	TRBV7-1, 7-2, 7-3-Nest	5' TCYACTCTGAMGWTCCAGCG 3'
TRBV7-4, 7-6, 7-7, 7-8, 7-9	5' GSWTCTYTGCAGARAGGCC 3'	TRBV7-4, 7-6, 7-7, 7-8, 7-9-Nest	5' TGRMGATYCAGCGCACA 3'
TRBV9	5' GATCACAGCAACTGGACAG 3'	TRBV9-Nest	5' GTACCAACAGAGCCTGGAC 3'
TRBV10-1, 10-2, 10-3	5' TGTWCTGGTATCGACAAGACC 3'	TRBV10-1, 10-2, 10-3-Nest	5' TCCYCCTCACTCTGGAGTC 3'
TRBV11-1, 11-2, 11-3	5' CGATTTTCTGCAGAGACGC 3'	TRBV11-1, 11-2, 11-3-Nest	5' GACTCCACTCTCAAGATCCA 3'
TRBV12-3, 12-4, 12-5	5' ARGTGACAGARATGGGACAA 3'	TRBV12-3, 12-4, 12-5-Nest	5' CYACTCTGARGATCCAGCC 3'
TRBV13	5' AGCGATAAAGGAAGCATCC 3'	TRBV13-Nest	5' CATTCTGAACTGAACATGAGC 3'
TRBV14	5' CCAACAATCGATTCTTAGCTG 3'	TRBV14-Nest	5' ATTCTACTCTGAAGGTGCAGC 3'
TRBV15	5' AGTGACCCTGAGTTGTTCTC 3'	TRBV15-Nest	5' ATAACTTCCAATCCAGGAGG 3'
TRBV16	5' GTCTTTGATGAAACAGGTATGC 3'	TRBV16-Nest	5' GAAAGATTTTCAGCTAAGTGCC 3'
TRBV17	5' CAGACCCCCAGACACAAG 3'	TRBV17-Nest	5' TGTTCACTGGTACCGACAG 3'
TRBV18	5' CATAGATGAGTCAGGAATGCC 3'	TRBV18-Nest	5' CGATTTTCTGCTGAATTCC 3'
TRBV19	5' AGTTGTGAACAGAATTTGAACC 3'	TRBV19-Nest	5' TTCCTCTCACTGTGACATCG 3'
TRBV20-1	5' AAGTTTCTCATCAACCATGC 3'	TRBV20-1-Nest	5'ACTCTGACAGTGACCAGTGC 3'
TRBV23-1	5' GCGATTCTCATCTCAATGC 3'	TRBV23-1-Nest	5' GCAATCCTGTCCTCAGAAC 3'
TRBV24-1	5' CCTACGGTTGATCTATTACTCC 3'	TRBV24-1-Nest	5' GATGGATACAGTGTCTCTCGA 3'
TRBV25-1	5' ACTACACCTCATCCACTATTCC 3'	TRBV25-1-Nest	5' CAGAGAAGGGAGATCTTTCC 3'
TRBV27, 28	5' TGGTATCGACAAGACCCAG 3'	TRBV27, 28-Nest	5' TTCYCCCTGATYCTGGAGTC 3'
TRBV29-1	5' TTCTGGTACCGTCAGCAAC 3'	TRBV29-1-Nest	5' TCTGACTGTGAGCAACATGAG 3'
TRBV30	5' TCCAGCTGCTCTTCTACTCC 3'	TRBV30-Nest	5' AGAATCTCTCAGCCTCCGAC 3'
TRBC	5' TAGAACTGGACTTGACAGCG 3'	TRBC-Nest	5' TTCTGATGGCTCAAACACAG 3'

Table 12. Primers for TCR amplification

	Primer's name	Primer's sequence
Forward primer (specific for TRAV8.1)	pscALPSKoVa8-1fwd	5' CGC CCG GGG GTC TAG AAGC CCACC ATG CTC CTG TTG CTC ATA C 3'
Reverse primer (specific for TRAC)	CaNoStopSGSGP2ARvs	5' ACG TCG CCG GCC TGC TTC AGC AGG CTG AAG TTG GTG GC GCC GCT GCC GCT GCT GGA CCA CAG CCG CAG 3'
Forward primer (specific for TRBV 24.1)	P2AVb24-1fwd	5' GAA GCA GGC CGG CGA CGT GGA GGA GAA CCC CGG CCCC ATG GCC TCC CTG CTC TTC 3'
Reverse primer (specific for TRBC 1)	Cb1StoppscALPSRvs	5' TCG AGA ATT CTG GCC AGTT TCA GAA ATC CTT TCT CTT GAC 3'
Reverse primer (specific for *TRBC* 2)	Cb2StoppscALPSRvs	5' TCG AGA ATT CTG GCC AGTT CTA GCC TCT GGA ATC CTT TC 3'

Primer names or sequences are color-coded or formatted according to the short DNA fragments that make up the respective primers. Primers were designed by linking short DNA fragments specific to the lentivirial vector (pscALPS), Kozak sequence (Ko), the viral P2A peptide and/or the SGSG linker peptide.

Table 13. List of symbols, abbreviations & acronyms

Symbols	Definitions
aa	Amino acid
Ag	Antigen
AIM	Acute infectious mononucleosis
AJ	TRAJ
APC	Antigen-presenting cells
AV	TRAV
B2m	β_2-microglobulin
BJ	TRBJ
BSA	Buried surface area
BV	TRBV
CDR	Complementary-determining region
CMV	Cytomegalovirus
CONV	Convalescence
DC	Dendritic cells
EBNA	EBV nuclear antigen
EBV	Epstein-Barr virus
GLC	HLA-A*02:01-restricted EBV protein BMLF1$_{280-288}$-derived epitope
HCV	Hepatitis C virus
HLA	Human leukocyte antigen
IAV	Influenza A virus
MHC-I	Major histocompatibility complex class I
N-nucleotide	Non-templated nucleotide
PBMC	Peripheral blood mononuclear cell
pMHC	Peptide-MHC complex
RSS	Recombination signal sequences
TAP	Transporter for Ag processing
TCR	T cell receptor
TRAC	T cell receptor alpha constant
TRAJ	T cell receptor alpha joining genes
TRAV	T cell receptor alpha variable genes
TRBC	T cell receptor beta constant
TRBD	T cell receptor beta diversity genes
TRBJ	T cell receptor beta joining genes
TRBV	T cell receptor beta variable genes
VCA	Viral capsid antigen
YVL	HLA-A*02:01-restricted EBV protein BRLF1$_{109-117}$-derived epitope

References

1.	Bjorkman PJ, Parham P. Structure, function, and diversity of class I major histocompatibility complex molecules. Annu Rev Biochem. 1990;59:253-88. Epub 1990/01/01. doi: 10.1146/annurev.bi.59.070190.001345. PubMed PMID: 2115762.

2.	Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, Wiley DC. Structure of the human class I histocompatibility antigen, HLA-A2. Nature. 1987;329(6139):506-12. Epub 1987/10/08. doi: 10.1038/329506a0. PubMed PMID: 3309677.

3.	Yewdell JW, Reits E, Neefjes J. Making sense of mass destruction: quantitating MHC class I antigen presentation. Nat Rev Immunol. 2003;3(12):952-61. Epub 2003/12/04. doi: 10.1038/nri1250. PubMed PMID: 14647477.

4.	Adams EJ, Luoma AM. The adaptable major histocompatibility complex (MHC) fold: structure and function of nonclassical and MHC class I-like molecules. Annu Rev Immunol. 2013;31:529-61. Epub 2013/01/10. doi: 10.1146/annurev-immunol-032712-095912. PubMed PMID: 23298204.

5.	Tynan FE, Borg NA, Miles JJ, Beddoe T, El-Hassen D, Silins SL, et al. High resolution structures of highly bulged viral epitopes bound to major histocompatibility complex class I. Implications for T-cell receptor engagement and T-cell immunodominance. J Biol Chem. 2005;280(25):23900-9. Epub 2005/04/26. doi: 10.1074/jbc.M503060200. PubMed PMID: 15849183.

6.	Tynan FE, Burrows SR, Buckle AM, Clements CS, Borg NA, Miles JJ, et al. T cell receptor recognition of a 'super-bulged' major histocompatibility complex class I-bound peptide. Nat Immunol. 2005;6(11):1114-22. Epub 2005/09/28. doi: 10.1038/ni1257. PubMed PMID: 16186824.

7.	Kaech SM, Cui W. Transcriptional control of effector and memory CD8+ T cell differentiation. Nat Rev Immunol. 2012;12(11):749-61. Epub 2012/10/20. doi: 10.1038/nri3307. PubMed PMID: 23080391; PubMed Central PMCID: PMCPMC4137483.

8.	Moon JJ, Chu HH, Pepper M, McSorley SJ, Jameson SC, Kedl RM, et al. Naive CD4(+) T cell frequency varies for different epitopes and predicts repertoire diversity and response magnitude. Immunity. 2007;27(2):203-13. Epub 2007/08/21. doi: 10.1016/j.immuni.2007.07.007. PubMed PMID: 17707129; PubMed Central PMCID: PMCPMC2200089.

9.	Prlic M, Hernandez-Hoyos G, Bevan MJ. Duration of the initial TCR stimulus controls the magnitude but not functionality of the CD8+ T cell response. J Exp Med. 2006;203(9):2135-43. Epub 2006/08/16. doi: 10.1084/jem.20060928. PubMed PMID: 16908626; PubMed Central PMCID: PMCPMC2118397.

10.	Manjunath N, Shankar P, Wan J, Weninger W, Crowley MA, Hieshima K, et al. Effector differentiation is not prerequisite for generation of memory cytotoxic T lymphocytes. J Clin Invest. 2001;108(6):871-8. Epub 2001/09/19. doi: 10.1172/JCI13296. PubMed PMID: 11560956; PubMed Central PMCID: PMCPMC200936.

11.	Sibener LV, Fernandes RA, Kolawole EM, Carbone CB, Liu F, McAffee D, et al. Isolation of a Structural Mechanism for Uncoupling T Cell Receptor Signaling from Peptide-MHC Binding. Cell. 2018;174(3):672-87 e27. Epub 2018/07/28. doi: 10.1016/j.cell.2018.06.017. PubMed PMID: 30053426; PubMed Central PMCID: PMCPMC6140336.

12.   Teixeiro E, Daniels MA, Hamilton SE, Schrum AG, Bragado R, Jameson SC, et al. Different T cell receptor signals determine CD8+ memory versus effector development. Science. 2009;323(5913):502-5. Epub 2009/01/24. doi: 10.1126/science.1163612. PubMed PMID: 19164748.

13.   Zehn D, Lee SY, Bevan MJ. Complete but curtailed T-cell response to very low-affinity antigen. Nature. 2009;458(7235):211-4. Epub 2009/02/03. doi: 10.1038/nature07657. PubMed PMID: 19182777; PubMed Central PMCID: PMCPMC2735344.

14.   Cui W, Kaech SM. Generation of effector CD8+ T cells and their conversion to memory T cells. Immunol Rev. 2010;236:151-66. doi: 10.1111/j.1600-065X.2010.00926.x. PubMed PMID: 20636815; PubMed Central PMCID: PMCPMC4380273.

15.   Lefranc MP. IMGT, the international ImMunoGeneTics database. Nucleic Acids Res. 2003;31(1):307-10. PubMed PMID: 12520009; PubMed Central PMCID: PMCPMC165532.

16.   Davis MM, Bjorkman PJ. T-cell antigen receptor genes and T-cell recognition. Nature. 1988;334(6181):395-402. Epub 1988/08/04. doi: 10.1038/334395a0. PubMed PMID: 3043226.

17.   Miles JJ, Douek DC, Price DA. Bias in the alphabeta T-cell repertoire: implications for disease pathogenesis and vaccination. Immunol Cell Biol. 2011;89(3):375-87. Epub 2011/02/09. doi: 10.1038/icb.2010.139. PubMed PMID: 21301479.

18.   Nikolich-Zugich J, Slifka MK, Messaoudi I. The many important facets of T-cell repertoire diversity. Nat Rev Immunol. 2004;4(2):123-32. doi: 10.1038/nri1292. PubMed PMID: 15040585.

19.   Arstila TP, Casrouge A, Baron V, Even J, Kanellopoulos J, Kourilsky P. A direct estimate of the human alphabeta T cell receptor diversity. Science. 1999;286(5441):958-61. Epub 1999/11/05. PubMed PMID: 10542151.

20.   Yanagi Y, Yoshikai Y, Leggett K, Clark SP, Aleksander I, Mak TW. A human T cell-specific cDNA clone encodes a protein having extensive homology to immunoglobulin chains. Nature. 1984;308(5955):145-9. PubMed PMID: 6336315.

21.   Zinkernagel RM, Doherty PC. Restriction of in vitro T cell-mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semiallogeneic system. Nature. 1974;248(5450):701-2. PubMed PMID: 4133807.

22.   Hedrick SM, Cohen DI, Nielsen EA, Davis MM. Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. Nature. 1984;308(5955):149-53. PubMed PMID: 6199676.

23.   La Gruta NL, Gras S, Daley SR, Thomas PG, Rossjohn J. Understanding the drivers of MHC restriction of T cell receptors. Nat Rev Immunol. 2018;18(7):467-78. doi: 10.1038/s41577-018-0007-5. PubMed PMID: 29636542.

24.   Attaf M, Sewell AK. Disease etiology and diagnosis by TCR repertoire analysis goes viral. Eur J Immunol. 2016;46(11):2516-9. doi: 10.1002/eji.201646649. PubMed PMID: 27813075.

25.   Furman D, Jojic V, Sharma S, Shen-Orr SS, Angel CJ, Onengut-Gumuscu S, et al. Cytomegalovirus infection enhances the immune response to influenza. Science translational medicine. 2015;7(281):281ra43. Epub 2015/04/04. doi: 10.1126/scitranslmed.aaa2293. PubMed PMID: 25834109; PubMed Central PMCID: PMCPMC4505610.

26.   Kloverpris HN, McGregor R, McLaren JE, Ladell K, Harndahl M, Stryhn A, et al. CD8+ TCR Bias and Immunodominance in HIV-1 Infection. Journal of immunology (Baltimore, Md : 1950). 2015;194(11):5329-45. Epub 2015/04/26. doi: 10.4049/jimmunol.1400854. PubMed PMID: 25911754; PubMed Central PMCID: PMCPMC4433859.

27.   Watkin L, Gil A, Mishra R, Aslan N, Ghersi D, Luzuriaga K, et al. Potential of unique influenza A crossreactive memory CD8 TCR repertoire to protect against Epstein Barr virus (EBV) seroconversion. J Allergy Clin Immun. 2017;in press.

28.   Aslan N, Watkin LB, Gil A, Mishra R, Clark FG, Welsh RM, et al. Severity of Acute Infectious Mononucleosis Correlates with Cross-Reactive Influenza CD8 T-Cell Receptor Repertoires. MBio. 2017;8(6). doi: 10.1128/mBio.01841-17. PubMed PMID: 29208744; PubMed Central PMCID: PMCPMC5717389.

29. Garboczi DN, Ghosh P, Utz U, Fan QR, Biddison WE, Wiley DC. Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. Nature. 1996;384(6605):134-41. Epub 1996/11/14. doi: 10.1038/384134a0. PubMed PMID: 8906788.

30. Garcia KC, Degano M, Pease LR, Huang M, Peterson PA, Teyton L, et al. Structural basis of plasticity in T cell receptor recognition of a self peptide-MHC antigen. Science. 1998;279(5354):1166-72. Epub 1998/03/21. PubMed PMID: 9469799.

31. Garcia KC, Degano M, Stanfield RL, Brunmark A, Jackson MR, Peterson PA, et al. An alphabeta T cell receptor structure at 2.5 A and its orientation in the TCR-MHC complex. Science. 1996;274(5285):209-19. Epub 1996/10/11. PubMed PMID: 8824178.

32. Theodossis A, Guillonneau C, Welland A, Ely LK, Clements CS, Williamson NA, et al. Constraints within major histocompatibility complex class I restricted peptides: presentation and consequences for T-cell recognition. Proc Natl Acad Sci U S A. 2010;107(12):5534-9. Epub 2010/03/10. doi: 10.1073/pnas.1000032107. PubMed PMID: 20212169; PubMed Central PMCID: PMCPMC2851776.

33. Adams JJ, Narayanan S, Liu B, Birnbaum ME, Kruse AC, Bowerman NA, et al. T cell receptor signaling is limited by docking geometry to peptide-major histocompatibility complex. Immunity. 2011;35(5):681-93. Epub 2011/11/22. doi: 10.1016/j.immuni.2011.09.013. PubMed PMID: 22101157; PubMed Central PMCID: PMCPMC3253265.

34. Chen G, Yang X, Ko A, Sun X, Gao M, Zhang Y, et al. Sequence and Structural Analyses Reveal Distinct and Highly Diverse Human CD8+ TCR Repertoires to Immunodominant Viral Antigens. Cell Rep. 2017;19(3):569-83. doi: 10.1016/j.celrep.2017.03.072. PubMed PMID: 28423320; PubMed Central PMCID: PMCPMC5472051.

35. Ishizuka J, Stewart-Jones GB, van der Merwe A, Bell JI, McMichael AJ, Jones EY. The structural dynamics and energetics of an immunodominant T cell receptor are programmed by its Vbeta domain. Immunity. 2008;28(2):171-82. Epub 2008/02/16. doi: 10.1016/j.immuni.2007.12.018. PubMed PMID: 18275829.

36. Motozono C, Kuse N, Sun X, Rizkallah PJ, Fuller A, Oka S, et al. Molecular basis of a dominant T cell response to an HIV reverse transcriptase 8-mer epitope presented by the protective allele HLA-B*51:01. J Immunol. 2014;192(7):3428-34. Epub 2014/03/07. doi: 10.4049/jimmunol.1302667. PubMed PMID: 24600035; PubMed Central PMCID: PMCPMC3962895.

37. Song I, Gil A, Mishra R, Ghersi D, Selin LK, Stern LJ. Broad TCR repertoire and diverse structural solutions for recognition of an immunodominant CD8+ T cell epitope. Nat Struct Mol Biol. 2017;24(4):395-406. doi: 10.1038/nsmb.3383. PubMed PMID: 28250417; PubMed Central PMCID: PMCPMC5383516.

38. Banerjee A, Gordon SM, Intlekofer AM, Paley MA, Mooney EC, Lindsten T, et al. Cutting edge: The transcription factor eomesodermin enables CD8+ T cells to compete for the memory cell niche. J Immunol. 2010;185(9):4988-92. Epub 2010/10/12. doi: 10.4049/jimmunol.1002042. PubMed PMID: 20935204; PubMed Central PMCID: PMCPMC2975552.

39. Joshi NS, Cui W, Chandele A, Lee HK, Urso DR, Hagman J, et al. Inflammation directs memory precursor and short-lived effector CD8(+) T cell fates via the graded expression of T-bet transcription factor. Immunity. 2007;27(2):281-95. Epub 2007/08/29. doi: 10.1016/j.immuni.2007.07.010. PubMed PMID: 17723218; PubMed Central PMCID: PMCPMC2034442.

40. Takemoto N, Intlekofer AM, Northrup JT, Wherry EJ, Reiner SL. Cutting Edge: IL-12 inversely regulates T-bet and eomesodermin expression during pathogen-induced CD8+ T cell differentiation. J Immunol. 2006;177(11):7515-9. Epub 2006/11/23. PubMed PMID: 17114419.

41. Kaech SM, Tan JT, Wherry EJ, Konieczny BT, Surh CD, Ahmed R. Selective expression of the interleukin 7 receptor identifies effector CD8 T cells that give rise to long-lived memory cells. Nat Immunol. 2003;4(12):1191-8. Epub 2003/11/20. doi: 10.1038/ni1009. PubMed PMID: 14625547.

42.   Raeber ME, Zurbuchen Y, Impellizzieri D, Boyman O. The role of cytokines in T-cell memory in health and disease. Immunol Rev. 2018;283(1):176-93. Epub 2018/04/18. doi: 10.1111/imr.12644. PubMed PMID: 29664568.

43.   Kalia V, Sarkar S, Subramaniam S, Haining WN, Smith KA, Ahmed R. Prolonged interleukin-2Ralpha expression on virus-specific CD8+ T cells favors terminal-effector differentiation in vivo. Immunity. 2010;32(1):91-103. Epub 2010/01/26. doi: 10.1016/j.immuni.2009.11.010. PubMed PMID: 20096608.

44.   Becattini S, Latorre D, Mele F, Foglierini M, De Gregorio C, Cassotta A, et al. T cell immunity. Functional heterogeneity of human memory CD4(+) T cell clones primed by pathogens or vaccines. Science. 2015;347(6220):400-6. Epub 2014/12/06. doi: 10.1126/science.1260668. PubMed PMID: 25477212.

45.   Frost EL, Kersh AE, Evavold BD, Lukacher AE. Cutting Edge: Resident Memory CD8 T Cells Express High-Affinity TCRs. J Immunol. 2015;195(8):3520-4. Epub 2015/09/16. doi: 10.4049/jimmunol.1501521. PubMed PMID: 26371252; PubMed Central PMCID: PMCPMC4592826.

46.   Maru S, Jin G, Schell TD, Lukacher AE. TCR stimulation strength is inversely associated with establishment of functional brain-resident memory CD8 T cells during persistent viral infection. PLoS Pathog. 2017;13(4):e1006318. Epub 2017/04/15. doi: 10.1371/journal.ppat.1006318. PubMed PMID: 28410427; PubMed Central PMCID: PMCPMC5406018.

47.   Sanecka A, Yoshida N, Kolawole EM, Patel H, Evavold BD, Frickel EM. T Cell Receptor-Major Histocompatibility Complex Interaction Strength Defines Trafficking and CD103(+) Memory Status of CD8 T Cells in the Brain. Front Immunol. 2018;9:1290. Epub 2018/06/21. doi: 10.3389/fimmu.2018.01290. PubMed PMID: 29922298; PubMed Central PMCID: PMCPMC5996069.

48.   Busch DH, Pamer EG. T cell affinity maturation by selective expansion during infection. J Exp Med. 1999;189(4):701-10. Epub 1999/02/17. PubMed PMID: 9989985; PubMed Central PMCID: PMCPMC2192934.

49.   Abdel-Hakeem MS, Boisvert M, Bruneau J, Soudeyns H, Shoukry NH. Selective expansion of high functional avidity memory CD8 T cell clonotypes during hepatitis C virus reinfection and clearance. PLoS Pathog. 2017;13(2):e1006191. doi: 10.1371/journal.ppat.1006191. PubMed PMID: 28146579; PubMed Central PMCID: PMCPMC5305272.

50.   Price DA, Brenchley JM, Ruff LE, Betts MR, Hill BJ, Roederer M, et al. Avidity for antigen shapes clonal dominance in CD8+ T cell populations specific for persistent DNA viruses. J Exp Med. 2005;202(10):1349-61. doi: 10.1084/jem.20051357. PubMed PMID: 16287711; PubMed Central PMCID: PMCPMC2212993.

51.   Blattman JN, Sourdive DJ, Murali-Krishna K, Ahmed R, Altman JD. Evolution of the T cell repertoire during primary, memory, and recall responses to viral infection. J Immunol. 2000;165(11):6081-90. PubMed PMID: 11086040.

52.   Sourdive DJ, Murali-Krishna K, Altman JD, Zajac AJ, Whitmire JK, Pannetier C, et al. Conserved T cell receptor repertoire in primary and memory CD8 T cell responses to an acute viral infection. J Exp Med. 1998;188(1):71-82. PubMed PMID: 9653085; PubMed Central PMCID: PMCPMC2525546.

53.   Dash P, Fiore-Gartland AJ, Hertz T, Wang GC, Sharma S, Souquette A, et al. Quantifiable predictive features define epitope-specific T cell receptor repertoires. Nature. 2017;547(7661):89-93. doi: 10.1038/nature22383. PubMed PMID: 28636592; PubMed Central PMCID: PMCPMC5616171.

54.   Glanville J, Huang H, Nau A, Hatton O, Wagar LE, Rubelt F, et al. Identifying specificity groups in the T cell receptor repertoire. Nature. 2017;547(7661):94-8. doi: 10.1038/nature22976. PubMed PMID: 28636589; PubMed Central PMCID: PMCPMC5794212.

55.   Klarenbeek PL, Remmerswaal EB, ten Berge IJ, Doorenspleet ME, van Schaik BD, Esveldt RE, et al. Deep sequencing of antiviral T-cell responses to HCMV and EBV in humans reveals a stable repertoire that is maintained for many years. PLoS Pathog. 2012;8(9):e1002889. doi:

10.1371/journal.ppat.1002889. PubMed PMID: 23028307; PubMed Central PMCID: PMCPMC3460621.

56. Sant S, Grzelak L, Wang Z, Pizzolla A, Koutsakos M, Crowe J, et al. Single-Cell Approach to Influenza-Specific CD8(+) T Cell Receptor Repertoires Across Different Age Groups, Tissues, and Following Influenza Virus Infection. Front Immunol. 2018;9:1453. doi: 10.3389/fimmu.2018.01453. PubMed PMID: 29997621; PubMed Central PMCID: PMCPMC6030351.

57. Turner SJ, Doherty PC, McCluskey J, Rossjohn J. Structural determinants of T-cell receptor bias in immunity. Nat Rev Immunol. 2006;6(12):883-94. Epub 2006/11/18. doi: 10.1038/nri1977. PubMed PMID: 17110956.

58. Venturi V, Price DA, Douek DC, Davenport MP. The molecular basis for public T-cell responses? Nat Rev Immunol. 2008;8(3):231-8. doi: 10.1038/nri2260. PubMed PMID: 18301425.

59. Watkin LB, Mishra R, Gil A, Aslan N, Ghersi D, Luzuriaga K, et al. Unique influenza A cross-reactive memory CD8 T-cell receptor repertoire has a potential to protect against EBV seroconversion. J Allergy Clin Immunol. 2017;140(4):1206-10. Epub 2017/06/21. doi: 10.1016/j.jaci.2017.05.037. PubMed PMID: 28629751; PubMed Central PMCID: PMCPMC5669360.

60. Bradley P, Thomas PG. Using T Cell Receptor Repertoires to Understand the Principles of Adaptive Immune Recognition. Annu Rev Immunol. 2019. Epub 2019/01/31. doi: 10.1146/annurev-immunol-042718-041757. PubMed PMID: 30699000.

61. Chen F, Rowen L, Hood L, Rothenberg EV. Differential transcriptional regulation of individual TCR V beta segments before gene rearrangement. J Immunol. 2001;166(3):1771-80. Epub 2001/02/13. PubMed PMID: 11160223.

62. McMurry MT, Hernandez-Munain C, Lauzurica P, Krangel MS. Enhancer control of local accessibility to V(D)J recombinase. Mol Cell Biol. 1997;17(8):4553-61. Epub 1997/08/01. PubMed PMID: 9234713; PubMed Central PMCID: PMCPMC232309.

63. Livak F, Burtrum DB, Rowen L, Schatz DG, Petrie HT. Genetic modulation of T cell receptor gene segment usage during somatic recombination. J Exp Med. 2000;192(8):1191-6. Epub 2000/10/18. PubMed PMID: 11034609; PubMed Central PMCID: PMCPMC2195867.

64. Miles JJ, Bulek AM, Cole DK, Gostick E, Schauenburg AJ, Dolton G, et al. Genetic and structural basis for selection of a ubiquitous T cell receptor deployed in Epstein-Barr virus infection. PLoS Pathog. 2010;6(11):e1001198. doi: 10.1371/journal.ppat.1001198. PubMed PMID: 21124993; PubMed Central PMCID: PMCPMC2987824.

65. Venturi V, Kedzierska K, Price DA, Doherty PC, Douek DC, Turner SJ, et al. Sharing of T cell receptors in antigen-specific responses is driven by convergent recombination. Proc Natl Acad Sci U S A. 2006;103(49):18691-6. doi: 10.1073/pnas.0608907103. PubMed PMID: 17130450; PubMed Central PMCID: PMCPMC1693724.

66. Miles JJ, Borg NA, Brennan RM, Tynan FE, Kjer-Nielsen L, Silins SL, et al. TCR alpha genes direct MHC restriction in the potent human T cell response to a class I-bound viral epitope. J Immunol. 2006;177(10):6804-14. Epub 2006/11/04. PubMed PMID: 17082594.

67. Hammerschmidt W, Sugden B. Replication of Epstein-Barr viral DNA. Cold Spring Harb Perspect Biol. 2013;5(1):a013029. Epub 2013/01/04. doi: 10.1101/cshperspect.a013029. PubMed PMID: 23284049; PubMed Central PMCID: PMCPMC3579399.

68. Shannon-Lowe C, Rowe M. Epstein Barr virus entry; kissing and conjugation. Curr Opin Virol. 2014;4:78-84. Epub 2014/02/21. doi: 10.1016/j.coviro.2013.12.001. PubMed PMID: 24553068.

69. Thorley-Lawson DA. Epstein-Barr virus: exploiting the immune system. Nat Rev Immunol. 2001;1(1):75-82. Epub 2002/03/22. doi: 10.1038/35095584. PubMed PMID: 11905817.

70. Tracy SI, Kakalacheva K, Lunemann JD, Luzuriaga K, Middeldorp J, Thorley-Lawson DA. Persistence of Epstein-Barr virus in self-reactive memory B cells. J Virol. 2012;86(22):12330-40. Epub 2012/09/07. doi: 10.1128/JVI.01699-12. PubMed PMID: 22951828; PubMed Central PMCID: PMCPMC3486485.

71.   Keymeulen B, Vandemeulebroucke E, Ziegler AG, Mathieu C, Kaufman L, Hale G, et al. Insulin needs after CD3-antibody therapy in new-onset type 1 diabetes. N Engl J Med. 2005;352(25):2598-608. Epub 2005/06/24. doi: 10.1056/NEJMoa043980. PubMed PMID: 15972866.

72.   Luzuriaga K, Sullivan JL. Infectious mononucleosis. N Engl J Med. 2010;362(21):1993-2000. doi: 10.1056/NEJMcp1001116. PubMed PMID: 20505178.

73.   Clute SC, Watkin LB, Cornberg M, Naumov YN, Sullivan JL, Luzuriaga K, et al. Cross-reactive influenza virus-specific CD8+ T cells contribute to lymphoproliferation in Epstein-Barr virus-associated infectious mononucleosis. J Clin Invest. 2005;115(12):3602-12. Epub 2005/11/26. doi: 10.1172/JCI25078. PubMed PMID: 16308574; PubMed Central PMCID: PMCPMC1288832.

74.   Hadinoto V, Shapiro M, Greenough TC, Sullivan JL, Luzuriaga K, Thorley-Lawson DA. On the dynamics of acute EBV infection and the pathogenesis of infectious mononucleosis. Blood. 2008;111(3):1420-7. Epub 2007/11/10. doi: 10.1182/blood-2007-06-093278. PubMed PMID: 17991806; PubMed Central PMCID: PMCPMC2214734.

75.   Crawford DH. Biology and disease associations of Epstein-Barr virus. Philos Trans R Soc Lond B Biol Sci. 2001;356(1408):461-73. doi: 10.1098/rstb.2000.0783. PubMed PMID: 11313005; PubMed Central PMCID: PMCPMC1088438.

76.   Loren AW, Porter DL, Stadtmauer EA, Tsai DE. Post-transplant lymphoproliferative disorder: a review. Bone Marrow Transplant. 2003;31(3):145-55. Epub 2003/03/07. doi: 10.1038/sj.bmt.1703806. PubMed PMID: 12621474.

77.   Cohen JI, Mocarski ES, Raab-Traub N, Corey L, Nabel GJ. The need and challenges for development of an Epstein-Barr virus vaccine. Vaccine. 2013;31 Suppl 2:B194-6. doi: 10.1016/j.vaccine.2012.09.041. PubMed PMID: 23598481; PubMed Central PMCID: PMCPMC3636506.

78.   Bollard CM, Gottschalk S, Torrano V, Diouf O, Ku S, Hazrat Y, et al. Sustained complete responses in patients with lymphoma receiving autologous cytotoxic T lymphocytes targeting Epstein-Barr virus latent membrane proteins. J Clin Oncol. 2014;32(8):798-808. doi: 10.1200/JCO.2013.51.5304. PubMed PMID: 24344220; PubMed Central PMCID: PMCPMC3940538.

79.   Bollard CM, Rooney CM, Heslop HE. T-cell therapy in the treatment of post-transplant lymphoproliferative disease. Nat Rev Clin Oncol. 2012;9(9):510-9. doi: 10.1038/nrclinonc.2012.111. PubMed PMID: 22801669; PubMed Central PMCID: PMCPMC3743122.

80.   Luzuriaga K, Sullivan JL. Infectious mononucleosis. N Engl J Med. 2010;362(21):1993-2000. doi: 362/21/1993 [pii];10.1056/NEJMcp1001116 [doi].

81.   Taylor GS, Long HM, Brooks JM, Rickinson AB, Hislop AD. The immunology of Epstein-Barr virus-induced disease. Annu Rev Immunol. 2015;33:787-821. Epub 2015/02/24. doi: 10.1146/annurev-immunol-032414-112326. PubMed PMID: 25706097.

82.   Callan MF. The evolution of antigen-specific CD8+ T cell responses after natural primary infection of humans with Epstein-Barr virus. Viral Immunol. 2003;16(1):3-16. Epub 2003/05/03. doi: 10.1089/088282403763635401. PubMed PMID: 12725684.

83.   Catalina MD, Sullivan JL, Bak KR, Luzuriaga K. Differential evolution and stability of epitope-specific CD8(+) T cell responses in EBV infection. J Immunol. 2001;167(8):4450-7. PubMed PMID: 11591771.

84.   Catalina MD, Sullivan JL, Brody RM, Luzuriaga K. Phenotypic and functional heterogeneity of EBV epitope-specific CD8+ T cells. J Immunol. 2002;168(8):4184-91. PubMed PMID: 11937579.

85.   Greenough TC, Campellone SC, Brody R, Jain S, Sanchez-Merino V, Somasundaran M, et al. Programmed Death-1 expression on Epstein Barr virus specific CD8+ T cells varies by stage of infection, epitope specificity, and T-cell receptor usage. PLoS One. 2010;5(9):e12926. doi: 10.1371/journal.pone.0012926. PubMed PMID: 20886079; PubMed Central PMCID: PMCPMC2944873.

86.   Annels NE, Callan MF, Tan L, Rickinson AB. Changing patterns of dominant TCR usage with maturation of an EBV-specific cytotoxic T cell response. J Immunol. 2000;165(9):4831-41. PubMed PMID: 11046006.

87.   Iancu EM, Corthesy P, Baumgaertner P, Devevre E, Voelter V, Romero P, et al. Clonotype selection and composition of human CD8 T cells specific for persistent herpes viruses varies with differentiation but is stable over time. J Immunol. 2009;183(1):319-31. Epub 2009/06/23. doi: 10.4049/jimmunol.0803647. PubMed PMID: 19542443.

88.   Iancu EM, Gannon PO, Laurent J, Gupta B, Romero P, Michielin O, et al. Persistence of EBV antigen-specific CD8 T cell clonotypes during homeostatic immune reconstitution in cancer patients. PLoS One. 2013;8(10):e78686. doi: 10.1371/journal.pone.0078686. PubMed PMID: 24205294; PubMed Central PMCID: PMCPMC3808305.

89.   Padovan E, Casorati G, Dellabona P, Meyer S, Brockhaus M, Lanzavecchia A. Expression of two T cell receptor alpha chains: dual receptor T cells. Science. 1993;262(5132):422-4. Epub 1993/10/15. PubMed PMID: 8211163.

90.   Naumov YN, Naumova EN, Yassai MB, Kota K, Welsh RM, Selin LK. Multiple glycines in TCR alpha-chains determine clonally diverse nature of human T cell memory to influenza A virus. J Immunol. 2008;181(10):7407-19. doi: 181/10/7407 [pii].

91.   Venturi V, Chin HY, Asher TE, Ladell K, Scheinberg P, Bornstein E, et al. TCR beta-chain sharing in human CD8+ T cell responses to cytomegalovirus and EBV. J Immunol. 2008;181(11):7853-62. PubMed PMID: 19017975.

92.   Nguyen TH, Bird NL, Grant EJ, Miles JJ, Thomas PG, Kotsimbos TC, et al. Maintenance of the EBV-specific CD8+ TCRalphabeta repertoire in immunosuppressed lung transplant recipients. Immunol Cell Biol. 2017;95(1):77-86. doi: 10.1038/icb.2016.71. PubMed PMID: 27507557; PubMed Central PMCID: PMCPMC5214975.

93.   Trautmann L, Janbazian L, Chomont N, Said EA, Gimmig S, Bessette B, et al. Upregulation of PD-1 expression on HIV-specific CD8+ T cells leads to reversible immune dysfunction. Nat Med. 2006;12(10):1198-202. Epub 2006/08/19. doi: 10.1038/nm1482. PubMed PMID: 16917489.

94.   Wei F, Zhong S, Ma Z, Kong H, Medvec A, Ahmed R, et al. Strength of PD-1 signaling differentially affects T-cell effector functions. Proc Natl Acad Sci U S A. 2013;110(27):E2480-9. Epub 2013/04/24. doi: 10.1073/pnas.1305394110. PubMed PMID: 23610399; PubMed Central PMCID: PMCPMC3703988.

95.   Corse E, Gottschalk RA, Allison JP. Strength of TCR-peptide/MHC interactions and in vivo T cell responses. J Immunol. 2011;186(9):5039-45. Epub 2011/04/21. doi: 10.4049/jimmunol.1003650. PubMed PMID: 21505216.

96.   Corse E, Gottschalk RA, Krogsgaard M, Allison JP. Attenuated T cell responses to a high-potency ligand in vivo. PLoS Biol. 2010;8(9). doi: 10.1371/journal.pbio.1000481. PubMed PMID: 20856903; PubMed Central PMCID: PMCPMC2939023.

97.   DeWitt WS, Smith A, Schoch G, Hansen JA, Matsen FA, Bradley P. Human T cell receptor occurrence patterns encode immune history, genetic background, and receptor specificity. Elife. 2018;7. doi: 10.7554/eLife.38358. PubMed PMID: 30152754.

98.   Emerson RO, DeWitt WS, Vignali M, Gravley J, Hu JK, Osborne EJ, et al. Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. Nat Genet. 2017;49(5):659-65. doi: 10.1038/ng.3822. PubMed PMID: 28369038.

99.   Attaf M, Huseby E, Sewell AK. alphabeta T cell receptors as predictors of health and disease. Cell Mol Immunol. 2015;12(4):391-9. doi: 10.1038/cmi.2014.134. PubMed PMID: 25619506; PubMed Central PMCID: PMCPMC4496535.

100.   Yang X, Gao M, Chen G, Pierce BG, Lu J, Weng NP, et al. Structural Basis for Clonal Diversity of the Public T Cell Response to a Dominant Human Cytomegalovirus Epitope. J Biol Chem.

2015;290(48):29106-19. doi: 10.1074/jbc.M115.691311. PubMed PMID: 26429912; PubMed Central PMCID: PMCPMC4661422.

101. Pymm P, Illing PT, Ramarathinam SH, O'Connor GM, Hughes VA, Hitchen C, et al. MHC-I peptides get out of the groove and enable a novel mechanism of HIV-1 escape. Nat Struct Mol Biol. 2017;24(4):387-94. doi: 10.1038/nsmb.3381. PubMed PMID: 28218747.

102. Yang X, Chen G, Weng NP, Mariuzza RA. Structural basis for clonal diversity of the human T-cell response to a dominant influenza virus epitope. J Biol Chem. 2017;292(45):18618-27. doi: 10.1074/jbc.M117.810382. PubMed PMID: 28931605; PubMed Central PMCID: PMCPMC5683187.

103. Singh NK, Riley TP, Baker SCB, Borrman T, Weng Z, Baker BM. Emerging Concepts in TCR Specificity: Rationalizing and (Maybe) Predicting Outcomes. J Immunol. 2017;199(7):2203-13. doi: 10.4049/jimmunol.1700744. PubMed PMID: 28923982.

104. Lu J, Van Laethem F, Bhattacharya A, Craveiro M, Saba I, Chu J, et al. Molecular constraints on CDR3 for thymic selection of MHC-restricted TCRs from a random pre-selection repertoire. Nat Commun. 2019;10(1):1019. Epub 2019/03/06. doi: 10.1038/s41467-019-08906-7. PubMed PMID: 30833553; PubMed Central PMCID: PMCPMC6399321.

105. Callan MF, Fazou C, Yang H, Rostron T, Poon K, Hatton C, et al. CD8(+) T-cell selection, function, and death in the primary immune response in vivo. J Clin Invest. 2000;106(10):1251-61. doi: 10.1172/JCI10590. PubMed PMID: 11086026; PubMed Central PMCID: PMCPMC381436.

106. Shen TJ, Chao AS, Lin JF. Predicting the number of new species in further taxonomic sampling. Ecology. 2003;84(3):798–804.

107. Shugay M, Bolotin DA, Putintseva EV, Pogorelyy MV, Mamedov IZ, Chudakov DM. Huge Overlap of Individual TCR Beta Repertoires. Front Immunol. 2013;4:466. Epub 2014/01/09. doi: 10.3389/fimmu.2013.00466. PubMed PMID: 24400005; PubMed Central PMCID: PMCPMC3872297.

108. Colwell RK. EstimateS: statistical estimation of species richness and shared species from samples. Version 9.0.0 User's Guide and application published at: http://viceroy.eeb.uconn.edu/estimates. 2005.

109. Watkin LB, Gil A, Mishra R, Aslan N, Ghersi D, Luzuriaga K, et al. Potential of influenza A memory CD8+ T-cells to protect against Epstein Barr virus (EBV) seroconversion. Journal of Allergy and Clinical Immunology. 2017;140(4):1206-10.

110. Cornberg M, Clute SC, Watkin LB, Saccoccio FM, Kim SK, Naumov YN, et al. CD8 T cell cross-reactivity networks mediate heterologous immunity in human EBV and murine vaccinia virus infections. J Immunol. 2010;184(6):2825-38. doi: jimmunol.0902168 [pii];10.4049/jimmunol.0902168 [doi].

111. Clute SC, Watkin LB, Cornberg M, Naumov YN, Sullivan JL, Luzuriaga K, et al. Cross-reactive influenza virus-specific CD8+ T cells contribute to lymphoproliferation in Epstein-Barr virus-associated infectious mononucleosis. J Clin Invest. 2005;115(12):3602-12. doi: 10.1172/JCI25078 [doi].

112. Davis MM, McHeyzer-Williams M, Chien YH. T-cell receptor V-region usage and antigen specificity. The cytochrome c model system. Ann N Y Acad Sci. 1995;756:1-11. PubMed PMID: 7645810.

113. Wolfl M, Rutebemberwa A, Mosbruger T, Mao Q, Li HM, Netski D, et al. Hepatitis C virus immune escape via exploitation of a hole in the T cell repertoire. J Immunol. 2008;181(9):6435-46. PubMed PMID: 18941234; PubMed Central PMCID: PMCPMC2742502.

114. Cukalac T, Chadderton J, Handel A, Doherty PC, Turner SJ, Thomas PG, et al. Reproducible selection of high avidity CD8+ T-cell clones following secondary acute virus infection. Proc Natl Acad Sci U S A. 2014;111(4):1485-90. doi: 10.1073/pnas.1323736111. PubMed PMID: 24474775; PubMed Central PMCID: PMCPMC3910643.

115. Gil A, Yassai MB, Naumov YN, Selin LK. Narrowing of human influenza A virus-specific T cell receptor alpha and beta repertoires with increasing age. J Virol. 2015;89(8):4102-16. doi: 10.1128/JVI.03020-14. PubMed PMID: 25609818; PubMed Central PMCID: PMCPMC4442365.

116. Benn CS, Netea MG, Selin LK, Aaby P. A small jab - a big effect: nonspecific immunomodulation by vaccines. Trends Immunol. 2013;34(9):431-9. doi: 10.1016/j.it.2013.04.004. PubMed PMID: 23680130.

117. Han A, Glanville J, Hansmann L, Davis MM. Linking T-cell receptor sequence to functional phenotype at the single-cell level. Nat Biotechnol. 2014;32(7):684-92. doi: 10.1038/nbt.2938. PubMed PMID: 24952902; PubMed Central PMCID: PMCPMC4337815.

118. Levitsky V, de Campos-Lima PO, Frisan T, Masucci MG. The clonal composition of a peptide-specific oligoclonal CTL repertoire selected in response to persistent EBV infection is stable over time. J Immunol. 1998;161(2):594-601. PubMed PMID: 9670932.

119. Bovay A, Zoete V, Dolton G, Bulek AM, Cole DK, Rizkallah PJ, et al. T cell receptor alpha variable 12-2 bias in the immunodominant response to Yellow fever virus. Eur J Immunol. 2018;48(2):258-72. Epub 2017/10/05. doi: 10.1002/eji.201747082. PubMed PMID: 28975614; PubMed Central PMCID: PMCPMC5887915.

120. Osuna CE, Gonzalez AM, Chang HH, Hung AS, Ehlinger E, Anasti K, et al. TCR affinity associated with functional differences between dominant and subdominant SIV epitope-specific CD8+ T cells in Mamu-A*01+ rhesus monkeys. PLoS Pathog. 2014;10(4):e1004069. Epub 2014/04/20. doi: 10.1371/journal.ppat.1004069. PubMed PMID: 24743648; PubMed Central PMCID: PMCPMC3990730.

121. Warren RL, Freeman JD, Zeng T, Choe G, Munro S, Moore R, et al. Exhaustive T-cell repertoire sequencing of human peripheral blood samples reveals signatures of antigen selection and a directly measured repertoire size of at least 1 million clonotypes. Genome Res. 2011;21(5):790-7. Epub 2011/02/26. doi: 10.1101/gr.115428.110. PubMed PMID: 21349924; PubMed Central PMCID: PMCPMC3083096.

122. Rosati E, Dowds CM, Liaskou E, Henriksen EKK, Karlsen TH, Franke A. Overview of methodologies for T-cell receptor repertoire analysis. BMC Biotechnol. 2017;17(1):61. Epub 2017/07/12. doi: 10.1186/s12896-017-0379-9. PubMed PMID: 28693542; PubMed Central PMCID: PMCPMC5504616.

123. Precopio ML, Sullivan JL, Willard C, Somasundaran M, Luzuriaga K. Differential kinetics and specificity of EBV-specific CD4+ and CD8+ T cells during primary infection. J Immunol. 2003;170(5):2590-8. PubMed PMID: 12594286.

124. Luzuriaga K, McManus M, Catalina M, Mayack S, Sharkey M, Stevenson M, et al. Early therapy of vertical human immunodeficiency virus type 1 (HIV-1) infection: control of viral replication and absence of persistent HIV-1-specific immune responses. J Virol. 2000;74(15):6984-91. PubMed PMID: 10888637; PubMed Central PMCID: PMCPMC112215.

125. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19(9):1639-45. doi: 10.1101/gr.092759.109. PubMed PMID: 19541911; PubMed Central PMCID: PMCPMC2752132.

126. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol. 1994;2:28-36. PubMed PMID: 7584402.

127. Wang GC, Dash P, McCullers JA, Doherty PC, Thomas PG. T cell receptor alphabeta diversity inversely correlates with pathogen-specific antibody levels in human cytomegalovirus infection. Sci Transl Med. 2012;4(128):128ra42. doi: 10.1126/scitranslmed.3003647. PubMed PMID: 22491952; PubMed Central PMCID: PMCPMC3593633.

128. Liang X, Weigand LU, Schuster IG, Eppinger E, van der Griendt JC, Schub A, et al. A single TCR alpha-chain with dominant peptide recognition in the allorestricted HER2/neu-specific T cell repertoire. J Immunol. 2010;184(3):1617-29. Epub 2010/01/01. doi: 10.4049/jimmunol.0902155. PubMed PMID: 20042572.