THE UNIVERSITY
of EDINBURGH

Decoding malaria T-cell responses using adaptive immune receptor repertoire sequencing

Natasha Louise Smith

THE UNIVERSITY *of* EDINBURGH

A thesis submitted for the degree of

Doctor of Philosophy

The University of Edinburgh

School of Biological Sciences

September 2021

Declaration

I declare that this thesis contains my own work and any collaborative work has been explicitly stated in the text. This work has not been submitted, as a whole or in part, for any other degree to any other university or educational institution.

Natasha Louise Smith

Acknowledgements

Firstly, I would like to thank my supervisor Dr. Graeme Cowan for allowing me the opportunity to work on this project. His patience, guidance and support has been invaluable throughout this experience. I'm sure I will be the first of many PhD students in his group. I would also like to thank Prof. David Gray and Dr. Joanne Thompson for their time and support as members of my committee. I am also very grateful to my collaborators Dr. Jason Mooney, Dr. Wiebke Nahrendorf and Dr. Phil Spence, who helped facilitate much of the data collection for this project.

To my HPGH cohort - Flo, Áine, Frank, Florian and James – I would like to say a big thank you for all the friendship and fun they brought to this adventure. Having such a wonderful cohort to work and discuss issues with certainly lightened any demanding times and has given me some amazing friends for life. I am also indebted to Florian, Diana and Flo for the many malaria immunology discussions we had, which helped shape this thesis. A special thank you also to honorary HPGH cohort member Deirdre, for her endless positivity, smiles and support, as well as the daily coffee walks we did during the long months of lockdown.

None of this would have been possible without the support of my family. A special thank you to my Dad who has always supported and encouraged me in all my endeavours. I am forever grateful to him and to my stepmother Christelle, for their never-ending love and guidance over the years, and for setting an example of where hard work and perseverance can take you. Finally, to my partner Jack, thank you for being a steady rock of support and encouragement throughout this journey.

Abstract

Malaria continues to be a serious public health problem in many parts of the world, and progress in reducing the global malaria burden has stalled in recent years. Despite decades of research, current vaccine candidates have low efficacy, and major challenges in achieving long-lasting immune-mediated elimination of malaria remain. This is in part due to a lack of knowledge regarding how both clinical and anti-parasite immunity develops during a malaria infection. Epidemiologically, immunity to severe clinical disease develops after only a limited number of infections, whilst anti-parasite immunity requires years of repeat exposure. T-cells are known key functional mediators of the developing immune response during a first *Plasmodium* infection, undergoing extensive activation and splenic expansion during the acute phase. However aberrant T-cell responses have also been implicated in the pathogenesis of severe disease. The clonality and clonal composition of the T-cell response during a first malaria infection, and how this varies following repeat exposure, has not previously been described. Here, I have used T-cell receptor repertoire sequencing as a novel tool to address this knowledge gap. Firstly, I sequenced the splenic $CD4^+$ T-cell receptor repertoires generated over the time-course of a murine *P. chabaudi* infection. Profiling the response using bulk TCRβ repertoire sequencing, single-cell RNA-seq, and analyses of independent RNA-seq data, I determined that following a first infection - within a highly polyclonal expansion - murine T-effector repertoires are consistently dominated by a specific TCRβ signature. This conserved T-cell response was consistently a hallmark of a first infection, but not expanded upon re-challenge. Determining the host or parasite factors driving this conserved response may uncover novel immune targets for malaria therapeutic purposes. Secondly, to resolve if similar dynamics occur in human *P. falciparum* infections, I sequenced the peripheral TCRβ repertoires generated longitudinally over the time-course of a controlled human malaria infection model, including following re-challenge. No clonally expanded or conserved populations were evident in response to either a first or second *P. falciparum* infection. However, non-specific recruitment of established T-cell clones from the peripheral circulation was evident, a dynamic repeated in homologous re-challenge infections. Understanding the consequences of this non-specific trafficking, and whether or not it shapes an individual's response to a *Plasmodium* infection, warrants further investigation. Overall, the results presented here demonstrate the

utility of TCR repertoire sequencing when applied to antigenically complex infections and provide novel insights that deepen our understanding of the multifaceted immune response elicited by the parasite. Findings also set up a myriad of future research directions to address the question of how immune-mediated protection against malaria is achieved.

Lay Summary

Malaria is a mosquito-borne disease caused by the *Plasmodium* parasite. It is a serious public health problem in many parts of the world, causing almost half a million deaths a year, and there is currently no vaccine that provides effective protection against the disease. This is in part due to the fact that we still know very little about how an individual living in a country with malaria, naturally becomes immune to the disease. T-cells are part of the immune response that play key roles against the parasite, although excessive T-cell responses have also been known to cause severe disease. The parts of the parasite that drive these T-cell responses remain poorly defined. Un-activated T-cells in an individual each have a unique T-cell receptor on their surface, the DNA sequence of which makes them a specific match for part of a particular pathogen. When activated by the part of the pathogen they are specific for, T-cells respond by making identical copies (clones) of themselves, enabling them to launch a large response against an invading pathogen. By studying the DNA sequences that encode the T-cell receptors of these responding clones, we can determine what specific T-cell receptors are activated by the pathogen and whether or not individuals share the same responses. In-turn, we hope to be able to determine what parts of the pathogen are driving the T-cell response. Here, I first studied the T-cell receptors that are expanded during a mouse infection of malaria. Although the T-cell responses were very broad, I found that some of the T-cell receptors activated across different mice were identical, indicating that a shared response between mice was generated. Uncovering the part of the parasite that drives this shared response may help us better understand what immune targets on the parasite could be used to help inform effective vaccine or malaria therapeutics design. I next studied the T-cell receptors of T-cells that are expanded in human volunteers infected with malaria. Although no shared responses between volunteers were found, I discovered that T-cells specific for other diseases appeared to respond to the malaria parasite in a non-specific fashion. This over-zealous response may disrupt the developing immune response or cause damage to the host, and further work is required to determine the consequence of this non-specific activation of T-cells during a human malaria infection, and whether or not it affects the outcome of an infection.

Contents Page

List of Figures

List of Tables

List of Abbreviations

AIRR	Adaptive immune receptor repertoire
AMA-1	Apical membrane antigen 1
ANOVA	Analysis of variance
BCR	B-cell receptor
BMI	Body mass index
BSA	Bovine serum albumin
CCR	Chemokine receptor
CCVTM	Centre for Clinical Vaccinology and Tropical Medicine - Oxford
CFA	Complete Freund's adjuvant
CHMI	Controlled human malaria infection
CIDR	Cysteine rich inter-domain region
CMV	Cytomegalovirus
CTLA-4	Cytotoxic T-lymphocyte associated protein 4
CXCL	C-X-C Motif Chemokine Ligand
EBV	Epstein-Barr virus
ECM	Experimental cerebral malaria
FDR	False discovery rate
GLIPH	Grouping of lymphocyte interactions by paratope hotspots
HBV	Hepatitis B virus
HCV	Hepatitis C virus
HIV	Human immunodeficiency virus
HLA	Human leukocyte antigen
IFN	Interferon
IG	Immunoglobulin
IL	Interleukin
IP	Intraperitoneal
IGMM	Institute of genetics and molecular medicine - Edinburgh
LAG	Lymphocyte activation gene
MAIT	Mucosal associated invariant T-cell
MBP	Maltose binding protein
MCMV	Murine cytomegalovirus
MSP	Merozoite surface protein

MT	Mosquito transmitted
NK	Natural killer cells
NT	Nucleotide
OLGA	Optimised likelihood estimate of immunoglobulin Amino-acid sequences
OVA	Ovalbumin
PAGE	Polyacrylamide gel electrophoresis
PAMP	Pathogen-associated molecular pattern
PCA	Principal components analysis
PD-1	Programmed cell death 1
PMR	Parasite multiplication rate
QC	Quality Control
RACE	Rapid amplification of cDNA ends
RAG	Recombination activating gene
RNA	Ribonucleic acid
RSV	Respiratory syncytial virus
SARS	Severe acute respiratory syndrome
SBP	Serial blood-passaged
SEM	Standard error of the mean
TB	Tuberculosis
Tcm	Central memory T-cell
TCR	T-cell receptor
Te	Effector T-cell
Tem	Effector memory T-cell
Tfh	Follicular helper T-cell
Tn	Naïve T-cell
TNF	Tumour necrosis factor
UMAP	Uniform manifold approximation and projection
UMI	Unique molecular identifier
VSA	Variant surface antigen
WHO	World Health Organisation

Chapter 1: Introduction

1.1. Adaptive immunity

All organisms have evolved to protect themselves against infectious agents, their toxins, and the pathology they may cause through a wide variety of mechanisms, that together make up an individual's immune system. In jawed vertebrates, this system is typically divided in to two functionally distinct groups: innate immunity and adaptive immunity. Effector cells of the innate immune response provide non-specific first line defence mechanisms that are able to respond to the presence of a pathogen immediately, triggered by detection through invariant innate recognition receptors. Such sensors recognise common pathogen-associate molecular patterns (PAMPs), that are not part of the host's own make-up, for example peptidoglycans and lipopolysaccharides of bacterial cell walls. Working in conjunction with innate immunity is the adaptive immune response. Adaptive immune responses, also referred to as acquired immunity, are highly specific to the particular pathogen that induced them, being triggered in an antigen-specific manner. A defining principle of adaptive immunity is the formation of immunological memory, which allows enhanced specific recall responses upon re-challenge with a particular antigen. T and B-cell lymphocytes are the key cellular effectors of this adaptive system. They display receptors on their surface that give them unique specificity. B-cells express B-cell receptors (BCRs), of which antibodies are a secreted form, whilst T-cells – which are the focus of this thesis – express T-cell receptors (TCRs). Each T and B-cell expresses many exact copies of a single variant BCR or TCR that has its own distinct antigen-binding site to confer specificity.

TCR's are expressed on the T-cell surface, and are comprised of 2 polypeptide chains. The majority of these consist of a heterodimer of an α and β chain, though a second group of T-cells which have a TCR composed of a γ and δ chain also exists, the latter constituting ~2-5% of the human T-cell repertoire (Dantzler and Jagannathan, 2018). $\alpha\beta$ TCR's largely constitute two groups - $CD4^+$ and $CD8^+$ T-cells - which recognise processed peptides presented on major histocompatibility complex (MHC) II and I respectively. These groups are also functionally distinct, with $CD4^+$ commonly

described as helper cells which upon activation secrete cytokines that are instrumental in orchestrating the developing adaptive response, including through provision of essential help to B-cell responses for germinal centre formation and affinity maturation (Crotty, 2014). $CD8^+$ T-cells are also known as 'killer' T-cells, due to their direct cytotoxic abilities through upregulation of effector molecules such as granzyme B and perforin amongst others (Zhang and Bevan, 2011). $\gamma\delta$ T-cells have a semi-invariant TCR, and are thought to help bridge the innate and adaptive arms of the immune system. Although a comprehensive knowledge of the antigenic targets of $\gamma\delta$ T-cells remains elusive (Willcox and Willcox, 2019), they have been shown to recognise a broad repertoire of ligands which includes amongst others recognition of phophoantigens, heat shock proteins and lipid antigens (Vavassori *et al.*, 2013; Deseke and Prinz, 2020; Junqueira *et al.*, 2021; Wang *et al.*, 2021). Recognition involves the TCR but does not require antigen processing and is not constrained by presentation by MHC molecules (Dantzler and Jagannathan, 2018; Ribot, Lopes and Silva-Santos, 2021).

1.1.1. V(D)J recombination

The clonal selection theory of adaptive immunity requires the immune system of an individual to be able to produce a large and diverse repertoire of T-cells, with each cell expressing a TCR with different antigenic specificity. The vast number of potential pathogens that a host may encounter necessitates an almost infinite number of distinct TCRs, in order for a specific response to be available for selection from a naïve T-cell pool. Indeed, it is estimated that there are ~10^8 to 10^{10} unique T-cell clones in a human body at any given time (Qi et al., 2014; Lythe et al., 2016; Krishna et al., 2020), and ~2 x10^6 in mice (Zarnitsyna et al., 2013). It was recognised in the 1960s that such vastly diverse repertoires could not be directly encoded in the genome due to size limitations. Instead, loci encoding each TCR chain comprise numerous non-contiguous sections of DNA that are recombined to form functional sequences, in a process known as V(D)J recombination, first identified in B-cells (Hozumi and Tonegawa, 1976). TCR sequence diversity is thus generated through somatic diversification brought about by the quasi-random recombination of a Variable (V), Diversity (D) and Joining (J) receptor gene segment during T-cell development. A fully assembled variable region of a TCR chain is produced by combining either a V and J

gene segments for α chain, or V, (D) and J for β chains, a process mediated by recombinases RAG-1 and RAG-2 (Schatz, Oettinger and Baltimore, 1989; Oettinger et al., 1990) (Figure 1.1). TCR sequence diversity is further increased by the addition and subtraction of non-templated nucleotides at the gene segment joining junctions. This junctional diversity results in the third complementarity-determining region (CDR3) of the TCR, which bridges the VDJ junction, being the most variable region of the receptor (Xu and Davis, 2000) (Figure 1.1C). Further variance in the TCR is added through combinational diversity created from the different possibilities of $\alpha\beta$ chain pairings. Overall, V(D)J recombination results in repertoires of T-cells with immune receptors with nearly infinite distinct antigen specificities, with the potential diversity of an individual's repertoire estimated at ~$2x10^{19}$ TCRs (Dupic *et al.*, 2019). Given that the TCR binding to its cognate peptide:MHC represents a key interface between immune system and pathogen, understanding the molecular architecture of TCR repertoires and their clonal kinetics in disease states can provide insight in to specific TCR recognition and in turn the antigenic drivers of these dynamics, with repertoires continuously shaped by clonal expansion of antigen responsive cells in the periphery.

Figure 1.1

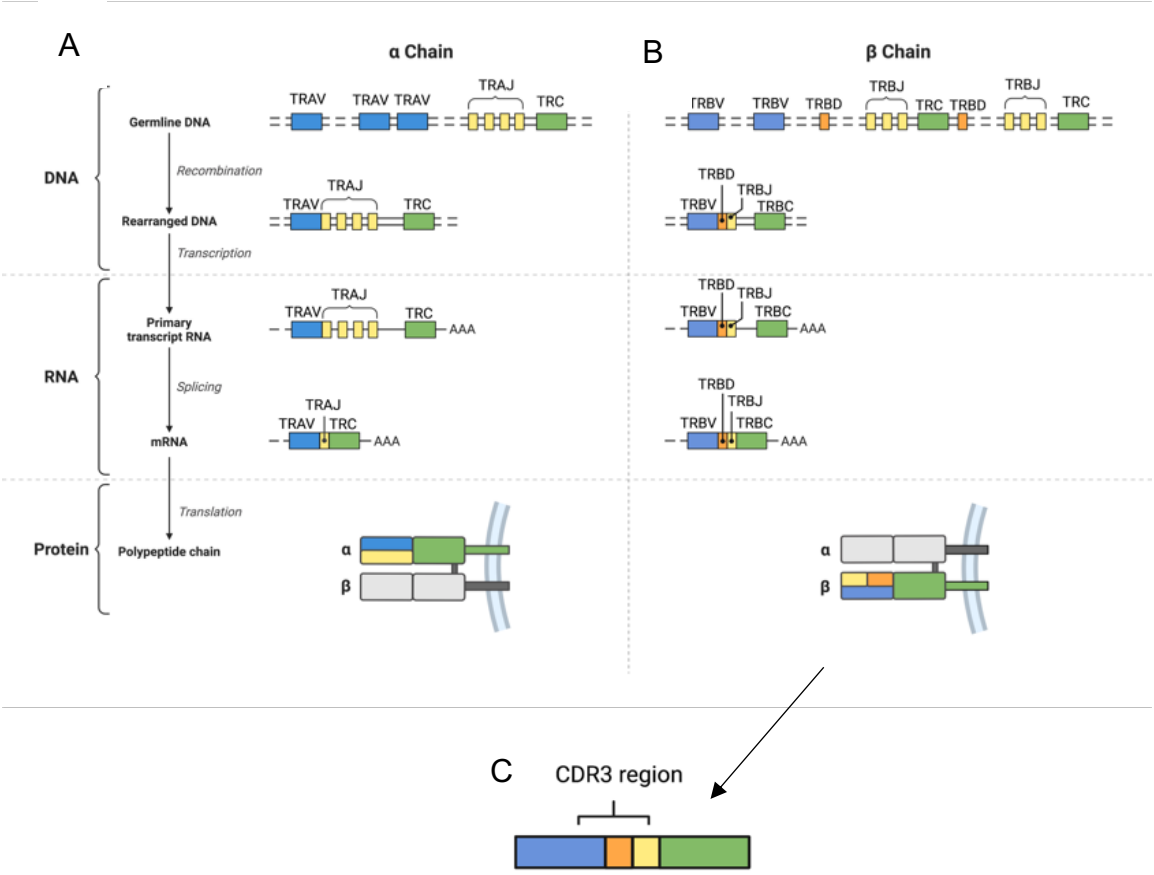


Figure 1.1: Graphical representation of V(D)J recombination for development of a T-cell receptor (A) α-chain and (B) β-chain. The CDR3 region of the TCRβ chain shown in (C) is the most diverse region of the TCR and mediates epitope specificity. Figure created using BioRender™.

1.1.2. Adaptive immune receptor repertoire sequencing (AIRR-seq): background

For decades, the complexity of these immune repertoires limited their analysis. Early studies that undertook TCRβ length analysis through spectratyping revealed biased enrichment of TCRs with identical V-gene and CDR3 length in antigen-responsive T-cell populations (Gorski *et al.*, 1994), and use of tetramers combined with FACS sorting on individual TRBV gene allowed for exhaustive sequencing of dominant TCRβ chains (McHeyzer-Williams *et al.*, 1999; Turner *et al.*, 2004). Such analyses were however low-resolution and couldn't provide insight into the real diversity and molecular clonal sequence architecture of a pathogen-specific response. A breakthrough in studying TCR repertoires without the need for prior knowledge of

epitope or dominant TRBV gene, came with the advent of long read high-throughput sequencing (HTS) techniques, when methods were developed that sequenced single receptor chains using multiplexed primers against individual TRBV genes, allowing entire unsorted repertoires to be analysed (Robins *et al.*, 2009; Dash *et al.*, 2011). The field was then further advanced by the development of techniques such as rapid amplification of cDNA ends (RACE), which uses a template-switching effect to introduce a universal adapter sequence during cDNA synthesis (Mamedov *et al.*, 2013; Heather *et al.*, 2016; Oakes *et al.*, 2017), negating the need for multiplexed primers that can cause bias in PCR amplification. Further refinement through addition of unique molecular identifiers (UMIs) – short stretches of random nucleotides incorporated during cDNA synthesis – was first introduced by Mamedov *et al.* (2013), with software developed for processing these by Shugay *et al.* (2014), though UMIs have only recently become more commonly used. Use of UMIs means that each template that enters amplification has a unique molecular tag, allowing PCR amplification bias and sequencing error correction, for accurate quantification of clones.

Many TCR-seq methods have now been developed, broadly making use of either DNA or RNA based methods, with the latter categorised into multiplex-PCR or RACE-PCR (with or without UMIs) techniques. All have study-dependent advantages and limitations (Liu *et al.*, 2016; Rosati *et al.*, 2017). With regards to starting material, the greater stability of gDNA and its' single template per cell allows for robust and accurate quantification of TCR clones without influence of expression level (Dziubianau *et al.*, 2013; Rosati *et al.*, 2017). However, the presence of introns can skew results when gDNA is used as input. As this methods also requires the use of multiplex primer sets for use at both V and J segment ends, amplification biases may be introduced that can lead to erroneous quantification and the loss of some rarer clonotypes (Mamedov *et al.*, 2013; Barennes *et al.*, 2020). In contrast, RNA-based techniques capture the final TCR product and are more sensitive due to the presence of multiple transcript copies per cell. They do however require consideration of the quality and quantity of available material as well as the cell-to-cell variation in mRNA expression levels(Turchaninova *et al.*, 2016). TCR expression levels do not vary by significant orders of magnitude between different T-cell subtypes, however B-cell and in particular plasma cell BCR expression levels may vary by 30-fold (Shi *et al.*, 2015), necessitating appropriate

separation by subtype for meaningful BCR repertoire analyses (Turchaninova *et al.*, 2016). A further advantage of using RNA as starting material is that RACE-PCR techniques can be used which negate the need for bias-inducing multiplex primer sets. RACE-PCR also allows incorporation of UMIs, which are used to accurately correct for inherent PCR and sequencing errors, in order to precisely quantify TCR clonotype frequency. (Kivioja *et al.*, 2012; Shugay *et al.*, 2014; Egorov *et al.*, 2015). Interpretation of TCR sequencing is particularly sensitive to PCR and sequencing errors, given a specific clone may differ from another by a single nucleotide. The use of UMIs was developed to overcome this issue and allow true biological difference to be distinguished from technical error. Whilst the use of UMIs is therefore recommended and is the method of choice for accurate normalisation and comparative diversity analyses (Egorov *et al.*, 2015; Rosati *et al.*, 2017; Izraelson *et al.*, 2018), it does have some limitations. Inclusion of UMIs complicates sequencing library preparation and their use also decreases the sensitivity of clonotype detection. Barennes *et al.* (2020) recently showed that RACE-PCR techniques without UMIs could detect clonotypes at a 10-fold lower frequency than those incorporating UMIs, thus maximal diversity is not captured. This is thought to be due to the increase in sequencing coverage (reads per UMI) required for appropriate error correction. The comparative analyses of 9 commonly used TCR repertoire profiling methods by Barennes *et al.* (2020) in general found marked differences in repertoires generated using different techniques, concluding that RACE UMI-based methods were most accurate for quantifying abundance of specific clonotypes, though they may still miss rarer clones due to their lower throughput. Overall, satisfactory methods exist for techniques based on both DNA and RNA input, but consideration must be given to the study objectives, material available and lab capabilities when choosing a suitable method for TCR repertoire profiling.

Methods to mine TCR repertoires from whole transcriptome RNA-seq data are also now available (Bolotin *et al.*, 2017; Li *et al.*, 2017), and although such techniques only superficially profile a repertoire, they have proved accurate in quantifying abundant clones in sorted T-cell subsets (Bai *et al.*, 2018), and non-targeted samples can also be informative. More recently, advances in single-cell technology such as the 10x Genomics™ platform, have allowed the capture of cognate TCR $\alpha\beta$ chain pairings, combined with the transcriptome of the cell, allowing TCR sequence to be linked to

cell phenotype (Zheng *et al.*, 2017; Hashimoto *et al.*, 2019). Recent theoretical advances in repertoire analysis through modelling intrinsic biases in V(D)J recombination have also led to the understanding that TCR's have different generation probabilities (PGen), which can now be taken in to consideration when quantifying repertoire dynamics (Murugan *et al.*, 2012; Marcou, Mora and Walczak, 2018; Sethna *et al.*, 2018; Dupic *et al.*, 2019).

1.1.3. AIRR-seq: applications

Given these rapid developments and advances in the field, AIRR-seq is now a powerful tool that has been applied to many areas of immunology to provide insight and further understanding of the immune status of a healthy individual (Britanova *et al.*, 2014; Chu *et al.*, 2019) or in altered conditions such as cancer (Li *et al.*, 2020), autoimmunity (Alves Sousa *et al.*, 2019; Servaas *et al.*, 2021), immunodeficiency (Wong *et al.*, 2017), infectious disease (Heather *et al.*, 2016; Huang *et al.*, 2020; Niu *et al.*, 2020) and vaccination (Pogorelyy, Minervina, Touzel, *et al.*, 2018; Miyasaka *et al.*, 2019). It has also been used to answer more theoretical questions such as the link between TCR and determinism of T-cell fate (Lönnberg *et al.*, 2017). The applications - both potential and realised - of AIRR-seq are extremely broad and ever-increasing as the field rapidly develops.

With regard to infectious diseases, a focus for this thesis, AIRR-seq has to date primarily been used to decipher immune responses against viral infections and following vaccination. Following pathogen challenge, activation and expansion of T-cells with a TCR specific to the pathogen occurs. Sequencing of these specific repertoires thus permits insights including the elucidation and quantification of repertoire clonality in direct response to the pathogen (Heather *et al.*, 2016; Wolf, Hether, Gilchuk, Kumar, *et al.*, 2018), including in some instances identifying cases of by-stander activation rather than antigen-driven clonal expansion (Ritvo *et al.*, 2018). It also allows the identification and documentation of specific TCR sequences which can be used as bio-markers of disease (Huth *et al.*, 2019), as well as their vast potential use in facilitating engineered TCRs for both therapeutic and epitope-discovery purposes. Although tetramer binding combined with AIRR-seq has identified and allowed the curation of many epitope-specific TCR sequences (Tickotsky *et al.*, 2017; Corrie *et al.*, 2018; Shugay *et al.*, 2018), a major goal of the field has been to

detect antigen-specific responding TCR's from AIRR-seq data, without the use of tetramers, as these require prior knowledge of specific epitope as well as an individual's HLA type. Two landmark studies (Dash *et al.*, 2017a; Glanville *et al.*, 2017) showed that epitope-specific TCRβ CDR3 aa sequences were often highly similar to one another, differing by as little as 1 amino acid mismatch, as well as being enriched for shared sequence motifs between those who did not share HLA-type. Based on these observations, methods to detect pathogen-specific clones from unsorted repertoires either post-challenge or from individuals with the same condition, have now been developed (Dash *et al.*, 2017b; Glanville *et al.*, 2017; Pogorelyy *et al.*, 2019; H. Zhang *et al.*, 2020; Huang *et al.*, 2020). These detect shared clones based on sequence similarity or shared enriched motifs. Such shared 'public' convergent responses between unrelated individuals have since been identified in response to a number of pathogens, including for example CMV, yellow-fever and COVID-19 (Emerson *et al.*, 2017; Pogorelyy, Minervina, Touzel, *et al.*, 2018; Schultheiß *et al.*, 2020; Shomuradova *et al.*, 2020). Interest in shared clones is high, as it is thought they may aid rational vaccine design, as they indicate there may be a common immunodominant epitope that can be used to generate and enhance T-cell responses across a wide population (Hill *et al.*, 2019). Such clinical applications of AIRR-seq are however still in their relative infancy, but this aspect of the field is expected to rapidly progress as much larger repertoire databases are generated (Fink, 2019). As the commonality of public clones post-pathogen exposure suggests responses to immunodominant antigens, their expression and panning can also facilitate *de novo* epitope discovery (Joglekar and Li, 2020), an exciting application of AIRR-seq that holds much promise, and is more thoroughly reviewed in Chapter 4 of this thesis.

Tracking of particular clonotypes including 'private' responses within an individual, has also now become tractable given increases in sequencing depth, though to date only a handful of truly longitudinal TCR-seq studies exist (Robinson *et al.*, 2016; Pogorelyy *et al.*, 2018; Minervina *et al.*, 2020; Minervina *et al.*, 2020; Schultheiß *et al.*, 2020; Servaas *et al.*, 2021), and methods for longitudinal analyses are not yet standardised. Such studies have the added advantage of being able to reveal novel clonal kinetics through unbiased clone trajectory analyses, for example the recent detection of a two-wave T-cell clone response in response to a COVID-19 infection (Minervina *et al.*, 2020), as well as documenting the antigen-specific TCRs involved.

Overall, AIRR-seq has become a highly effective tool to study the developing immune response to a pathogen and its ability to elucidate clonal dynamics, identify shared clonal responses that can aid rational vaccine design and the potential for de novo epitope discovery, make it an ideal technique to apply to malaria research.

1.2. Malaria

1.2.1. General background and global disease burden of *P. falciparum*

Malaria is a mosquito-transmitted disease caused by infection with different species of the protozoan parasite *Plasmodium*. Of these, *P. falciparum* is considered the most important, as it causes the vast majority of malaria deaths. It is the causative agent of falciparum malaria, a disease responsible for ~200 million cases of malaria a year and ~386,000 deaths in 2019, with most fatalities occurring in children under 5 years old (WHO, 2020). The majority of these cases occur in low and middle-income countries in sub-Saharan Africa, where it poses both a significant economic and public health burden (Shretta, Avanceña and Hatefi, 2016; Weiss *et al.*, 2019). Control measures are currently reliant on a combined toolbox of vector control strategies and anti-malarial drug treatment, and although much progress in the reduction of malaria cases and deaths worldwide have been made since the turn of the century, worrying reports from the World Health Organisation (WHO) indicate that progress in reducing the global malaria burden has stalled since 2015. As such, the reduction, elimination and eradication targets set for the next decade are thought to no longer be feasible with current control strategies (WHO Strategic Advisory Group on Malaria Eradication, 2019; WHO, 2020). This highlights the essential need to develop effective new tools to help combat the heavy burden of disease, and both innovative vaccines and novel drug treatments are cited as potential tools that could positively impact eradication programmes. However, the most advanced vaccine candidate, Mosquirix®, has low efficacy (RTSS Clinical Trials Partnership, 2015), and although more promising candidates may be on the horizon (Datoo *et al.*, 2021; Minassian *et al.*, 2021), major challenges to achieving long-lasting immune-mediated elimination of malaria still remain. This is in part due to the large gaps in our understanding regarding how both clinical and anti-parasite immunity develops during a malaria infection. Greater knowledge regarding both mechanisms and mediators of protective responses, and

the key targets of immunity are essential for the development of effective immune-mediated tools (Langhorne *et al.*, 2008; Beeson *et al.*, 2019).

1.2.2. Naturally acquired immunity to *P. falciparum* malaria

In endemic settings, naturally acquired immunity against falciparum malaria develops over infancy and childhood, through repeat exposure, with different hallmarks of protection presenting sequentially as the response develops. Firstly, protection against severe disease can be acquired after only a limited number of infections, with very few children experiencing more than 1 severe episode (Snow *et al.*, 1998; Gupta *et al.*, 1999; Gonçalves *et al.*, 2014). This appears to be independent of controlling parasitaemia load (Gonçalves *et al.*, 2014). This is followed by the acquisition of clinical immunity (asymptomatic) over the next decade or so if exposure is maintained, usually by adolescence. Coinciding with this, non-sterile anti-parasite immunity gradually develops. Thus, adults in endemic areas are protected against both severe and uncomplicated malaria, but sterile immunity is rarely if ever achieved, with asymptomatic submicroscopic infections highly prevalent in endemic areas (Langhorne *et al.*, 2008; Galatas, Bassat and Mayor, 2016; Nguyen *et al.*, 2018; Andrade *et al.*, 2020).

The key role of antibodies in naturally acquired anti-parasite immunity was first established following demonstration of parasite clearance after passive transfer of immunoglobulins from immune adults to children (Cohen, McGregor and Carrington, 1961), and this humoral response is thought to target multiple stages of the parasite (reviewed in (Gonzales *et al.*, 2020)). However, many major questions surrounding this humoral response still remain; for example why the acquisition of this immunity is so slow, why immunity is not long-lived and whether or not strain-transcending antibodies are ever generated. For the pathogenic blood-stage of malaria - which are the infection models used in this thesis - principal known antigenic-targets of parasites include parasite variant surface antigens (VSA's) expressed on the cell surface of infected erythrocytes, as well as merozoite surface proteins and merozoite-secreted proteins utilised during erythrocyte invasion. It is thought antigenic variation through switching expression of the VSA's displayed by the parasite on the infected red cell surface contributes to slow acquisition of immunity by allowing immune evasion (Smith

et al., 1995; reviewed in (Deitsch and Dzikowski, 2017)), and their extreme sequence diversity also hampers effective vaccine development. High titres of antibodies against merozoite proteins in endemic areas have been demonstrated (King *et al.*, 2015), but these have not always correlated with protection (Stanisic *et al.*, 2015; Ssewanyana *et al.*, 2017), and merozoite protein blood-stage vaccine candidates including MSP family proteins and AMA-1 have to date shown poor efficacy (Thera *et al.*, 2011; Sheehy *et al.*, 2012) without heterologous protection, with antigen polymorphisms likely playing a role in the latter. Thus, determining which antigens are key, the means by which antibodies against these confer protection and/or if conserved responses are generated remain important active areas of research (Gonzales *et al.*, 2020). Dysregulation of the B-cell compartment through atypical B-cell formation has also been associated with chronic malaria infection and is thought to cause humoral deficiencies that affect the ability to generate and maintain effective memory responses (Portugal *et al.*, 2015, 2017; Holla *et al.*, 2021).

Anti-parasite humoral responses have been linked to the acquisition of immunity against severe disease, with a major hypothesis in the field positing that generation of antibodies against the more virulent variants of parasite genes, for example those which mediate pathogenic processes including rosetting by the VSA PfEMP1 as the reason for protection from severe disease in the presence of high parasite-burden (Ofori *et al.*, 2002; Doumbo *et al.*, 2009; Duffy *et al.*, 2016; Chan *et al.*, 2019). A fuller discussion of the anti-parasite humoral responses generated during a malaria infection is beyond the scope of this thesis, but brief comments here serve to highlight the complexity and many unknowns of this response and the antigenic drivers involved, and advocate the need for further research, including BCR repertoire analysis of infected individuals, to help untangle these complicated responses and identify antigen-specific responses if present.

1.2.3. The role of T-cells in malaria

A more recent theory that decouples immunity to severe disease from anti-parasite immunity is the host-centric idea that an individual learns to tolerate their parasite burden, by controlling their inflammatory response to limit the pathological

consequences and collateral damage of the host response to infection (Muñoz Sandoval et al., (In Prep), Nahrendorf, Ivens and Spence, 2021)). This has recently been proposed to be mediated through modified CD4+ T-cell activation when re-challenged, reducing the cytotoxic effectors associated with auto-inflammatory disease and tissue damage (Muñoz Sandoval et al., (In Prep)). The exact mechanisms that orchestrate this have not yet been defined, but it is proposed that bystander activation of non-specific cross-reactive T-cells may occur in a first infection, but that this is then prevented from occurring in re-infection, despite identical pathogen load. This theory places a greater emphasis on T-cell activity rather than on B-cell functionality in the prevention of severe disease, however the two theories are unlikely to be mutually exclusive, as reducing inflammation could prevent the conditions that facilitate endothelium activation, and thus the selection of pathogenic variants.

T-cell responses have been shown to be crucial to surviving a first malaria infection in murine models of disease, where athymic mice succumb to lethal infection with *P. chabaudi adami* and *P. berghei*, but can be rescued with adoptive transfer of splenocytes or thymic graft (Grun and Weidanz, 1981; Cavacini, Long and Weidanz, 1986). Here, I briefly summarise what is known about the major malaria T-cell subsets that are examined in this thesis, with insights gained from both murine and human field studies.

$CD4^+$ T-cells orchestrate key aspects of both innate and adaptive immunity during a malaria infection, differentiating into several distinct effector subsets. T-helper 1 cells (T_H1) are the main sources of IFN-γ during the acute phase of infection, which has been shown to play a central role in protective immunity during blood-stage malaria in murine models (Su and Stevenson, 2000) and been shown to be positively associated with protection in field studies of African children (Luty *et al.*, 1999; Reece *et al.*, 2004; Walther *et al.*, 2009) and following vaccination in naïve adults with the RTS/S vaccine candidate (Sun *et al.*, 2003; Kester *et al.*, 2009). Other cellular sources of IFN-γ such as CD8+ cells, NK cells and $\gamma\delta$ T-cells have also been reported (King and Lamb, 2015). IFN-γ mediates its effector function through optimal activation of CD8+ T-cells, B-cells and macrophages, enhancing the cytotoxic destruction, opsonisation and phagocytosis of infected cells respectively. T_H1 cells also produce other cytokines including IL-2, which aids T-cell expansion and activation of NK cells, and TNF-α which

is thought to limit parasite growth (Kumar *et al.*, 2019). Although these T_H1 pro-inflammatory anti-parasitic functions are important for protection, there is a fine balance between an appropriate immune response that limits parasite replication, and avoiding a response that damages the host if left unchecked (King and Lamb, 2015). Such cytokines are implicated in the pathology of severe malaria; in the *P. berghei* ANKA experimental cerebral malaria (ECM) model of disease, mice were protected from developing ECM following anti-IFN-γ antibody treatment or IFN-γ gene knock out (Grau *et al.*, 1989; Yañez *et al.*, 1996; Hunt *et al.*, 2014). Similarly, inflammatory imbalances have been demonstrated in studies of severe falciparum malaria (Mahanta *et al.*, 2015), with higher levels of pro-inflammatory cytokines including IFN-γ and TNF-α observed in children with severe malaria versus those with uncomplicated disease (Lyke *et al.*, 2004; Awandare *et al.*, 2006; Mandala *et al.*, 2017), and it is thought that severe malaria develops as a consequence of pro-inflammatory cytokine driven excessive inflammation, alongside the microvascular obstruction caused by sequestration of parasitised erythrocytes (Crompton *et al.*, 2014). IL-10, which is secreted by many different immune cell types including T_H1, is recognised as an important regulatory anti-inflammatory molecule for suppressing this excessive inflammation (Couper, Blount and Riley, 2008), though it too has been implicated in exacerbating disease if unchecked, through the suppression of cell-mediated immunity (Kumar, Ng and Engwerda, 2019).

T follicular helper cells (T_FH) are another major $CD4^+$ T-cell subset in malaria and are essential for promoting antibody responses. They secrete cytokines including IL-4 and IL-21, which provide crucial support for isotype-switching and affinity maturation of germinal centre B-cells. They have been shown to be critical to antibody-mediated resolution of parasitaemia in the *P. chabaudi* and *P. yoelii* murine models (Pérez-Mazliah *et al.*, 2015, 2017; Surette *et al.*, 2021), as well as promoting protective antibody responses in humans (Obeng-Adjei *et al.*, 2015; Figueiredo *et al.*, 2017). Paradoxically, $CD4^+$ responses have also been suggested to impede effective humoral immunity and the generation of long-term memory with $PD1^+$ and $LAG3^+$ T_{CM} and T_{EM} phenotypes reported in cohorts with chronic malaria exposure (Illingworth *et al.*, 2013), and $PD1^+$ $CTLA4^+$ T-cell populations in cases of acute imported malaria (Mackroth *et al.*, 2016). These markers are associated with T-cell exhaustion and impairment of T-cell effector function, and have been suggested to be one reason for the inefficient

clearance of parasitaemia (Frimpong *et al.*, 2018). Both T_H1 and T_{FH} can transition from effector to memory populations, though the mechanisms through which these orchestrate secondary responses in malaria remains undefined (Kurup, Butler and Harty, 2019).

Cytotoxic $CD8^+$ T-cells are important effectors in the pre-erythrocytic stages of malaria, where in murine models they recognise infected hepatocytes (Weiss *et al.*, 1988; Romero *et al.*, 1989; Spencer *et al.*, 2017), and although antigen-specific $CD8^+$ T-cells have correlated with efficacy following experimental human vaccination (Ewer *et al.*, 2013), very high numbers appear to be required to mediate protection (Schmidt *et al.*, 2010). They primarily function through production of cytotoxic molecules including granzyme-B and perforin, as well as producing IFN-γ and TNF-α. Their role in controlling the blood-stage of malaria is thought to be limited. $CD8^+$ T-cells have however also been implicated in the pathophysiology of severe disease, being shown to sequester in the brain in murine models during development of cerebral malaria (Belnoue *et al.*, 2002; Swanson *et al.*, 2016), secreting high-levels of IFN-γ, granzyme B and perforin. Similarly, in human cerebral malaria, recent field studies have shown an association between increased $CD8^+$ T-cells expressing granzyme B in the brain vasculature and children who died from cerebral malaria (Barrera *et al.*, 2019; Riggle *et al.*, 2020).

Evidently, T-cells play important but conflicting roles in the pathogenesis of malaria, being essential for protection but also contributing to severe disease, and potentially slowing the acquisition of anti-parasite humoral immunity. Deep profiling these TCR repertoires with novel AIRR-seq techniques combined with single-cell RNA-seq can provide new insights in to the molecular architecture of these complex responses, matching clonality and TCR sequence with phenotype. By analysing TCR repertoires in a first infection and following re-challenge, the evolution of the response following exposure can also be assessed. The complex epidemiology of acquisition of immunity against severe disease after sometimes only a first infection, makes studying these responses in re-challenge also essential to properly understanding how T-cell responses differ between sequential infections as the different hallmarks of immunity develop.

1.2.4. *P. chabaudi* and CHMI as models of malaria

Murine models have proved invaluable to malaria research, due to the ability to control every aspect of the infection as well as being able to easily harvest immunologically important tissues such as spleen and bone marrow. However, *Plasmodium* species are mostly host-specific, and human *Plasmodium* species therefore cannot be used in murine models. In this thesis, TCR repertoires generated from both the *P. chabaudi* murine model as well as the human *P. falciparum* CHMI model are used. Given this, I felt it important to briefly highlight here the commonalities, but also some major differences between the two models and stress that findings from one model although informative for the other, may not be directly transferable between each other nor to endemic settings.

P. chabaudi is widely used as a murine model of malaria, as it induces an infection with many similarities to *P. falciparum* (Stephens, Culleton and Lamb, 2012). Firstly, it induces a chronic infection, with parasites persisting at a low to submicroscopic level (Achtman *et al.*, 2007). *P. chabaudi* also sequesters on vascular endothelium and induces rosetting (Mackinnon, Walker and Rowe, 2002), key mechanisms associated with pathology in falciparum malaria. *P. chabaudi* also uses variant surface antigens in immune evasion, although the multigene family used in cytoadherence is not homologous to the *var* gene family in *P. falciparum.* Clinically, acute *P. chabaudi* infection can induce many of the presentations associated with falciparum malaria including anaemia, splenomegaly, hypoglycaemia and placental malaria, although mice develop hypothermia rather than the classic cyclic pyrexia seen in falciparum. Notably however, *P. chabaudi* does not sequester in the brain and does not induce cerebral malaria, except in genetically susceptible backgrounds, for example $IL10^{-}$ /$IL10^{-}$ (Sanni *et al.*, 2004). Immunologically, the most importance difference between the two models is that sterile or near-sterile immunity can be achieved against the *P. chabaudi* parasite after only 1 infection. The initial development of the immune response and memory formation have however provided useful hypotheses for further investigation in human studies, with observation of both T and B cell dynamics fairly consistent between the two models (Stephens, Culleton and Lamb, 2012). Antigenically, although the two species have diverged, many orthologous genes

between the two have been identified, for example those encoding the merozoite surface proteins, meaning the drivers of any detected epitope-specific responses in murine models may in some cases be extrapolated to human-infective *Plasmodium* species.

CHMI studies have many advantages over field studies in that they too can be controlled to some extent; standardising parasite genotype, inoculation dose and time-course of infection allows accurate comparative observations between individuals to be drawn. Volunteers can also be pre-screened to ensure they are malaria-naïve, and baseline values immediately prior to challenge known as well as data provided from regular sampling over the course of infection, including duration of infection and parasitaemia. However, the CHMI model is potentially limited in what can be extrapolated to endemic areas. Ethical considerations require volunteers to be drug-treated before severe disease is allowed to develop, thus this aspect of the disease cannot be directly studied. Using mostly adult European volunteers may also potentially effect extrapolation of results to endemic areas, where age as well as genetic polymorphisms may influence the host response. Comparisons with ongoing CHMI trials in endemic areas will provide more clarity on this in due course, and ultimately the insights that can be gained using these controlled infections make it a highly effective and informative model of falciparum malaria.

1.3. Aims of thesis

In sum, much remains unknown regarding how protective immunity to malaria develops. Dissecting these T-cell responses using AIRR-seq provides a powerful and novel tool with which to determine clonal dynamics of responses and determine if shared signatures of disease that could be driven by immunodominant epitopes are present. Similarly, determining if by-stander activation is induced in a first infection would provide supporting evidence for the disease tolerance theory in human malaria (Muñoz Sandoval *et al.*, (In Prep)). In this thesis, firstly the TCR repertoires elicited during a murine *P. chabaudi* model will be sequenced to demonstrate utility of the technique and determine if antigen specific TCR signatures are elicited. This should not only be novel and highly informative for the malaria field but will also provide proof of principle that such signatures can be detected following challenge with such a highly

antigenically complex parasite. Secondly, the TCR repertoires generated longitudinally following a novel homologous *P. falciparum* rechallenge model, will be sequenced and analysed to characterise and compare T-cell clonal kinetics during an individual's first and second falciparum infection. This rich data set will be used to investigate whether shared *Plasmodium*-specific clonal responses between individuals are elicited or if responses bear the hallmarks of a non-specific response.

Thus, the overall aims of this thesis are as follows:

1. To sequence the splenic TCR repertoires elicited over the time-course of a murine malaria *P. chabaudi* infection using AIRR-seq techniques:
    a. Determine if conserved clonal TCR signatures are generated in the acute phase of infection by sequencing the $CD4^+$ effector population
    b. Profile the splenic TCR memory repertoires to determine if shared responses enter memory populations
    c. Compare and contrast the splenic repertoires elicited in a comparative *P. chabaudi* model of serially-blood passaged or recently mosquito transmitted parasite, in both effector and memory populations
    d. Reveal the function and phenotype of any conserved signatures detected through single cell RNA-seq combined with AIRR-seq.

2. To sequence and characterise the T-cell repertoires generated in a first *P. falciparum* infection as well as following homologous re-challenge:
    a. Determine the kinetics and dynamics of T-cell clones longitudinally over the course of a first infection and a re-challenge infection
    b. Capture longitudinal changes in T-cell clonality, and compare these over the course of a first and re-challenge infection
    c. Reveal whether pathogen-specific responses can be tracked through time within an individual to reveal private clonal dynamics
    d. Compare longitudinal responses between individuals to determine whether conserved public responses between individuals are generated.

Chapter 2: Materials and methods

Both murine and human samples were used to generate the TCR repertoire data analysed in this thesis. In this chapter, I have outlined the main methods used to generate, pre-process and analyse the TCR sequence data for both. Specific materials and methods for each experiment have been included in relevant chapters.

2.1. Murine model of malaria

2.1.1. Mouse infections

All procedures were carried out in accordance with UK Home Office regulations (Animals Scientific Procedures Act, 1986; project license number 70/8,546 and P04ABDCAA) and were approved by the Ethical Review Body of the University of Edinburgh. C57Bl/6 mice were bred and housed under specific pathogen free conditions at the University of Edinburgh and subjected to regular pathogen monitoring by sentinel screening. Mice were housed with at least one companion in individually ventilated cages furnished with autoclaved woodchip, fun tunnel and tissue paper at 21°C ± 2°C under a reverse light-dark cycle (light, 19.00 – 07.00; dark, 07.00 – 19.00) at a relative humidity of 55% ± 10%. *P. chabaudi* (AS) parasites were obtained from the European Malaria Reagent Repository at the University of Edinburgh. Eight- to 10-week-old C57Bl/6 female mice were infected with *P. chabaudi* (AS) by intra-peritoneal injection of $1x10^5$ parasitised erythrocytes that had either been maintained by serial blood-passage over a high number of generations (SBP) or undergone a single passage following mosquito transmission (MT) as per Spence et al. (2012). For all mouse work included in this thesis, infections were carried out by Dr Wiebke Nahrendorf (University of Edinburgh), Dr Jason Mooney (University of Edinburgh) or myself as stated. Specific timings and numbers used for each experiment are included in relevant chapters.

2.1.2. Isolation of murine splenocytes

2.1.1.1. List of materials

Reagent	Supplier	Catalogue Number
RPMI	Sigma-Aldrich	R0883
Cell-strainer (70μm)	Falcon	352350
Red Cell Lysis Buffer	Sigma-Aldrich	R7757
Foetal Calf Serum (FCS)	Sigma-Aldrich	C8056
Casyton Buffer	Sedna Scientific	43003
Casyton Tubes	Sedna Scientific	43001

Following humane euthanasia, spleens were immediately transferred into RPMI on ice, before being gently homogenised using a glass mortar and pestle. This suspension was filtered through a cell strainer (70μm) with 10mL 10% FCS in RPMI, before centrifuging for 10 minutes at 300g and 4°C. Once supernatant was discarded, the resulting pellet was resuspended and treated with 1mL Red Cell Lysis Buffer for 1 minute, before the reaction was neutralised by adding 19mL FCS 10% in RPMI. This was then centrifuged for 10 minutes at 300g and 4°C. Cell counts per mL were determined using a Casy cell counter.

2.1.3. Cell sorting

2.1.3.1. List of materials

Reagent	Supplier	Catalogue Number
CD62L FITC clone MEL-14	BioLegend	104406
CD4 APC clone GK15	BioLegend	100411

CD44 A700 clone IM7	BioLegend	103026
CD127 BV421 clone A7R34	BioLegend	135024
CD3 PECy7 clone 17A2	BioLegend	100219
Dulbecco's PBS	Sigma-Aldrich	D8537

5 x 10^7 splenocytes were immune-stained in the dark for 20 minutes at 4°C, to facilitate sorting, with the following staining pool per sample:

- CD62L FITC - 10μL
- CD4 APC - 3μL
- CD44 A700 - 3μL
- CD127 BV421 - 3μL
- CD3 PE/Cy7 - 3μL

$CD4^+$ splenic T-cell populations of interest were then isolated by FACS using a BD FACSAria III instrument, according to gates described by Spence *et al*. (2013) as follows: all $CD3^+$ $CD4^+$ - T-naïve ($CD62L^+$ $CD127^+$), T-effector ($CD62L^-$ $CD127^-$), T-effector memory (CD44HI $CD127^+$ $CD62L^-$) and T-central memory (CD44HI $CD127^+$ $CD62L^+$). Cell sorting was carried out by Dr Martin Waterfall (University of Edinburgh). Cells were sorted into 50μL FACS buffer and stored at -80°C until being processed. An example of gating strategy used is presented in Chapter 3 (Appendix 3.1).

For the single-cell experiment, cells were only immune-stained with CD4 APC and CD3 PE/Cy7, to minimise any potential effects of staining on transcriptome. $CD4^+$ cells were then sorted into 100μL 0.04% BSA in PBS, to generate the single-cell suspension required for 10x Genomics sequencing.

2.1.4. Murine TCRβ bulk amplification

2.1.4.1. List of materials

Reagent	Supplier	Catalogue Number
Dynabeads mRNA Purification Kit	ThermoFisher Scientific	61011
RNAse Inhibitor	Clontech	2313A
SMARTScribe™ Reverse Transcriptase	Clontech	639537
First Strand Buffer	Clontech	639537
DTT	Clontech	639537
dNTP Mix 10mM	ThermoFisher Scientific	R0192
Uracyl DNA glycosylase (5 U/μL)	New England BioLabs	M0280S
Phusion flash high-fidelity PCR mastermix	ThermoFisher Scientific	F548L
Monarch PCR Gel-extraction Kit	New England BioLabs	T1020L

All RNA extractions were carried out with appropriate control measures to minimise risk of contamination, including being carried out in a separate 'clean' room, where no amplified PCR products are handled. RNA was extracted from the sorted $CD4^+$ splenic T-cell populations of interest using Dynabeads mRNA purification kits. cDNA was synthesised from each RNA preparation by adding the following to each sample: 4μL First-strand Buffer, 2μL 10mM dNTP mix, 0.5μL 20mM DTT, 2μL 10μM SMART-PTO2 oligo, 0.5μL RNase inhibitor, and 2μL (100U/μL) SMARTScribe™ reverse transcriptase. For repeat experiments (indicated in relevant chapters), unique molecular identifiers (UMIs) (Kivioja *et al.*, 2012) were incorporated during cDNA synthesis by replacing the template-switch oligo with 2μL 10μM SMARTNNN oligo. Samples were then incubated at 42°C for 50 minutes, before the reaction was terminated by heating at 70°C for 10 minutes. cDNA synthesised with SMARTNNN oligo were treated with 1μl of Uracyl DNA glycosylase (5U/μl) and incubated for 15 minutes at 37°C. PCR was then used to generate TCRβ V-region amplicons, using

indexed forward primers composed of the SMART synthesis oligo sequence fused to a P7 Illumina tag and a reverse primer within the TCR-C region fused to a P5 Illumina tag (Table 2.1). A full list of the indexed forward primers used is presented at the end of this Chapter in Appendix 2.1. The following amplification protocol was used:

- Phusion flash high-fidelity mastermix - 10µL
- cDNA – 2µL
- P5-mTCRBrev3 (12µM) - 2µL
- P7-SMART-Index (10µM) - 2µL
- Nuclease free water - 4µL

Reactions were heated to 98°C for 2 minutes, followed by 5 cycles of amplification using the following program: 98°C for 5s, 60°C for 10s and 72°C for 15s; followed by 25 cycles of 98°C for 5s, 65°C for 10s and 72°C for 15s; final elongation at 72°C for 2 minutes.

Table 2.1

Primer	**Function**	**Primer Sequence**
SMART-PTO2	Template switch	AAGCAGTGGTATCAACGGAGAGTACATrG$_{(3)}$
SMARTNNN	Template switch	AAGCAGUGGTAUCAACGCAGAGUNNNNUNNNNUNNNNUCTTrG$_{(3)}$
P7-SMART-Index	Forward	CAAGCAGAAGACGGCATACGAGATXXXXXXGGCGAAGCAGTGGTATCAACGCAGAGT
P5-mTCRBrev3	Reverse	AATGATACGGCGACCACCGAGATCTACACCTTGGGTGGAGTCACATTTCT

Table 2.1: Table of murine TCRβ cDNA synthesis and amplification primers

Amplified products were pooled and purified by extraction from excised agarose gel bands, with the Monarch PCR gel extraction kit.

2.1.4.2. Murine TCRβ bulk sequencing

Single end 1x400bp or asymmetric 400bp+100bp (to incorporate UMIs) sequencing was performed on an Illumina MiSeq v2 platform, using the custom read primers presented in Table 2.2. Library quality and concentration was assessed by Nanodrop spectrophotometer prior to submission to Edinburgh Genomics, who carried out all bulk sequencing runs presented in this thesis. Edinburgh Genomics provided further quality control and fragment size analysis using an Agilent Bioanalyzer, as well as assessing read quality post-sequencing using inbuilt MiSeq software. Quality was further assessed after receiving the raw fastq files, using FastQC (Andrews, 2010).

Table 2.2

Primer Name	Function	Primer Sequence
R1-mTCRB	Read 1	CGAGATCTACACCTTGGGTGGAGTCACATTTCT
UMI-read2	Read 2 (for UMI use only)	GGCGAAGCAGTGGTATCAACGCAGAGT
Index-smartrev	Index	ACTCTGCGTTGATACCACTGCTTCGCC

Table 2.2: Table of murine TCRβ sequencing read primers.

2.1.5. Murine single-cell sequencing

Single-cell sequencing was undertaken at the MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, using 10X Genomics sequencing technology: two barcoded cDNA libraries were prepared from FACS sorted samples using the Chromium Single Cell 5' Library Kit v2 (Zheng *et al.*, 2017). Full length V(D)J segments were enriched from amplified cDNA with primers specific to the TCR constant region using the Chromium Single Cell V(D)J Enrichment Kit – Mouse T-Cell. Sequencing was performed using the High-Output v2.5 Kit on a NextSeq 550 platform.

2.1.5.1. List of materials

Reagent	Supplier	Catalogue Number
Chromium Single Cell 5' Library Kit	10x Genomics	1000011
Chromium Next GEM Single Cell 5' Library Construction Kit v2	10x Genomics	1000020
Chromium Single Cell V(D)J Enrichment Kit – Mouse T Cell	10x Genomics	1000071
Agilent Bioanalyser DNA HS Kit	Agilent Technologies	5067-4626
Qubit dsDNA HS assay Kit	ThermoFischer Scientific	Q32851
NextSeq 500/550 High-Output v2.5 (150 cycles) Kit		20024904

2.1.6. Murine MSP-1 ELISA & avidity ELISA

2.1.6.1. List of materials

Reagent	Supplier	Catalogue Number
BL21 (DE3) competent *E.coli*	New England BioLabs	C2527H
MBTrap HP Maltose Binding Column	GE Healthcare	28918779
Benzonase DNase	Sigma	E1014
NuPAGE LDS Loading Buffer	ThermoFisher Scientific	NP0007
Instant Blue™ Coomassie protein stain	Abcam	ab119211
Greiner 96 Well Microtitre Plates	Bio-One	655945
HRP-conjugated rabbit anti-mouse Ig	Dako	P026002-2

HRP-conjugated goat anti-mouse IgG	Abcam	ab6789
SureBlue™ TMB Substrate	Insight Biotechnology Ltd	53-00-01

2.1.6.2. MSP-1 protein expression & purification

Recombinant *P. chabaudi* MSP-1_{21} protein was produced for use as the antigenic target in development of an indirect ELISA. Recombinant maltose-binding protein fused MSP-1_{21} was expressed in BL21 (DE3) *E. coli* cells, transformed with a previously constructed pMAL-p2X(AJ)-MSP-1_{21} plasmid vector previously constructed by Liz Stevenson (University of Edinburgh). For transformation, 1µL of pMAL-p2X(AJ)-MSP-1_{21} was added to 100µL *E. coli* cell solution and incubated on ice for 10 minutes, before being heat-shocked at 42°C for 45 seconds. Transformed cells were plated on LB agar containing 0.1mg/mL ampicillin and incubated overnight at 37°C. A single colony from each was selected and used to inoculate starter cultures by adding them to 10mL LB medium containing 0.1mg/mL ampicillin. These starter cultures were incubated initially at 37°C then turned down to 30°C overnight in an incubator shaking at 180 rpm. These were then centrifuged and bacterial pellets obtained were re-suspended in 5mL LB medium before being added to flasks containing 1L LB medium, 0.1mg/mL ampicillin and 10mL w/v 1% filter-sterilised (0.22µm filter) glucose solution. Flasks were then incubated at 37°C and turbidity monitored using a spectrophotometer at wavelength 600nm. Once turbidity had reached 1.0 (mid-log growth), cultures were induced by addition of Isopropyl β-D-1-thiogalactopyranoside (ITPG) solution at a final concentration of 1mM. Flasks were incubated overnight at 25°C, shaking at 150rpm. Cultures were then centrifuged at 5000g, supernatant decanted and resulting bacterial pellet frozen at -20°C. Cells were lysed using 3 x freeze/thaw cycles. For each cycle, cells were thawed on ice for 10 minutes then frozen on dry ice for 2 minutes. Each pellet was then suspended in 20mL binding buffer (20mM Tris HCl, 200mM NaCl & 1mM EDTA, pH 7.4) and 1µL Benzonase DNase (100 U/µL) added to each once re-suspended. 100µL samples were taken before and after centrifugation of product for control purposes and supernatant kept for affinity chromatography.

Purification of protein was undertaken using affinity chromatography on amylase resin using a 5mL maltose-binding column with an NGCTM medium-pressure liquid chromatography system (Bio-rad). The column was equilibrated with binding buffer before loading lysate samples. All buffers used were vacuum filter sterilised (0.22µm filter) prior to use. Untagged proteins (flow through) were washed from the column using binding buffer and then MBP-tagged protein were eluted from the column using elution buffer (binding buffer + 10mM maltose solution). Elutes were collected in 5mL fractions, with a single peak of A280 for both the MBP-control and MSP-121 visualised in first fraction (fraction 1) only. The column was then regenerated using an alkaline regeneration buffer (0.5M NaOH).

An SDS polyacrylamide gel was run to visualise expected protein size and confirm protein purification. For this, 30µL of whole cell, lysate, flow through and elute of both control and the MSP-1_{21} transfect obtained during the expression and chromatography process, were combined with 10µL of loading buffer and 1µL of DTT and denatured at 70°C for 90 seconds to linearise proteins present. These were loaded into a preformed acrylamide gel (Bis-tris 10%) run in MOPS SDS. Electrophoresis was carried out at 200V for 50 minutes before the gel was placed in Blue™ Coomassie protein stain and placed on a shaker plate for 1 hour before being visualised.

2.1.6.3. MSP1_{21} Ig ELISA

A chequerboard titration was initially undertaken to optimise the ELISA. Following this, 100µL of MSP-1_{21} at 0.5µg/mL protein in carbonate bicarbonate buffer (15mM Na2CO3 35mM NaHCO3 pH 9.3) was coated on 96 well microtitre plates and incubated humidified over- night at 4°C. Plates were then washed 4 times in washing buffer (0.05% Tween-20 in PBS) using an automated plate washer to remove unbound antigen. Blocking of free binding sites was undertaken by adding 200µL blocking buffer (1% Marvel® skimmed milk in PBS) to each well for 5 hours at room temperature, before plates were washed 4 times. Sera diluted 1:100 in blocking buffer was added in duplicate to wells and serially titrated by tripling dilution up to 1:72900 and plates were incubated overnight at 4°C. Plates were again washed 4 times, before 100µL of HRP-conjugated polyclonal rabbit anti-mouse immunoglobulin diluted 1:2000 in washing buffer was added to each well. Plates were then washed 4 times to remove unbound secondary antigen, before the reaction was developed by adding 100µL

SureBlue™ TMB substrate to each well. The reaction was stopped by adding 25µL 2M H_2SO4 to each well after 5 minutes and optical density was measured at 450 nm (Labsystems Multiscan Ascent microtitre plate reader software). Duplicated titrations of pooled positive sera, as well as 2 blank wells, were included on each plate for standardisation. Titres for each sample were calculated using Prism 7.0 (GraphPad). The A50 value of each sample was determined through interpolation of a 4-parameter logistic regression curve and an arbitrary titre then calculated relative to the A50 value of the standard curve generated from the titration of *P. chabaudi* positive pooled sera included on each plate. MBP purified from *E. coli* transfected with pMAL-p2X(AJ) alone, was used as a control to determine non-specific binding of mice immunoglobulin to MBP.

2.1.6.4. $MSP1_{21}$ IgG ELISA and antibody avidity assay

An ELISA affinity assay based on the dissociation of antigen-antibody interactions by chaotropic anions was undertaken. Sera from *P. Chabaudi (AS)* infected mice was assayed in the presence and absence of the mild chaotropic agent ammonium thiocyanate (NH_4SCN), the tolerance to which has been demonstrated to be proportional to antigen-antibody avidity (MacDonald, Hosking and Jones, 1988; Ferreira and Katzin, 1995; Almanzar *et al.*, 2013). A modified version of the assay used by Ferreira & Katzin (1995) was utilised. The optimum concentration of NH_4SCN for the protocol was determined by assaying known *P. chabaudi* positive pooled sera in the presence and absence of varying concentrations of NH_4SCN, indicating that 1M gave an appropriate level of dissociation for experimental requirements. A modified version of the above ELISA was then undertaken. Briefly, following incubation of diluted serum overnight at 4°C, wells were washed 4 times before 100µL of 1M NH_4SCN diluted in PBS was added to each test well for 15 minutes at room temperature. 100µL of PBS was added to control wells at the same time. Test samples with and without NH_4SCN were analysed in duplicate and run on the same plate to minimise variability. Bound IgG was detected using HRP-conjugated goat anti-mouse IgG and titres of IgG for each sample were again calculated using a 4-parameter logistic regression method.

2.2. Controlled Human Malaria Infections (CHMI)

2.2.1. Study background and ethical approvals

Human PBMC samples utilised in this thesis were obtained as part of a collaboration with Dr Phil Spence (University of Edinburgh), as part of a wider collaboration with Dr Simon Draper at the Jenner Institute, University of Oxford. Samples were obtained from the clinal trial: “A Phase I/IIa clinical trial to assess the safety, immunogenicity and efficacy of the blood-stage *P. falciparum* malaria vaccine candidate RH5.1/AS01”, ClinicalTrials.gov Identifier: NCT02927145, VAC063 study. This trial was designed and conducted by teams led by Dr Simon Draper and Dr Angela Minassian of the Jenner Institute, University of Oxford, and samples were obtained between November 2017 and June 2018 at the Centre for Clinical Vaccinology and Tropical Medicine (CCVTM) in Oxford. Ethical approval for this study was granted by the United Kingdom National Health Service Research Ethics Service (Oxfordshire Research Ethics Committee A, reference 16/SC/0345) and was conducted in full concordance with the Declaration of Helsinki 2008.

2.2.2. Volunteers

The above clinical trial was designed to assess the vaccine candidate RH5.1/AS01. As part of this, a group of volunteers who underwent both a primary and secondary CHMI, but did not receive the vaccine, were included as a control group. These two CHMI trials were categorised as VAC63A (primary CHMI) and VAC63B (secondary CHMI) respectively. As the objective of this thesis was to examine TCRβ repertoires following a malaria infection in unvaccinated individuals, only samples from a sub-group of individuals within this control group were utilised. This comprised of 7 individuals, who underwent a secondary homologous challenge 4 months following their primary infection, inside a framework to study boosting of the vaccine candidate. All recruited volunteers were between 20-50 years old and were healthy malaria naïve males or non-pregnant females. Exclusion criteria during volunteer recruitment included no history of any previous malaria infection, seropositivity for HBV, HIV or HCV, recent (within 30 days) malaria chemoprophylaxis, G6PD deficiency or any

known haemoglobinopathy. Detailed volunteer information is presented in Chapter 5, Appendix 5.7, with HLA type presented in Appendix 5.8.

2.2.3. Volunteer infections

Volunteers underwent a blood stage infection only. The inoculum used for this came from cryopreserved vials of 3D7 *Plasmodium falciparum* infected erythrocytes kept at the QIMR in Brisbane, Australia (Cheng *et al.*, 1997). All infections were undertaken with a homologous 3D7 *P. falciparum* recently transmitted line, which came from a single volunteer previously infected by mosquito bite, cryopreserved after 3 asexual blood cycles following liver egress (Cheng *et al.*, 1997). The inoculum was administered to volunteers in 5mL 0.9% saline intravenously at the CCVTM in Oxford, at an estimated dose of 1,000 parasitised erythrocytes per volunteer. Following the initiations of infections for each trial, a limiting dilution assay was used to determine viability of each inoculum, indicating a viability of 452 ring stage infected erythrocytes in the first infection and 857 in the second infection. Blood was collected daily from individuals starting from the day following first challenge (C+1), to monitor parasitaemia. Volunteers were diagnosed and treated if they were positive by thick blood smear, and/or if they had a parasite density greater than 5000 parasites per mL as quantified by qPCR and/or 1000 parasites per mL and were symptomatic. Volunteers were then treated with either Riamet® or Malarone®, dependent on any drug use contraindications.

2.2.4. Blood count monitoring

Full blood counts were performed at the Churchill or John Radcliffe Hospitals, Oxford, by NHS staff. For these, blood samples were taken at baseline (C-1), C+6, day of diagnosis, and at a post-treatment timepoint (either C+28, C+45 and/or C+90). This count data included cell counts per mL of blood for lymphocytes, eosinophils, monocytes and neutrophils.

2.2.4 Isolation and cryopreservation of PBMCs

2.2.4.1. List of materials

Reagent	Supplier	Catalogue Number
Tempus	Thermo Fisher Scientific	4342792
Lymphoprep	STEMCELL Technologies	7811
PBS	Life Technologies	20012-019
Fetal Bovine Serum (FBS)	Thermo Fisher Scientific	16000044
Sepmate-15™	STEMCELL Technologies	85415
CoolCell®	Corning	432006

5mL of whole blood was diluted in a 1:1 ratio with 2% FBS in PBS in a sterile 15mL falcon tube and mixed gently. This was carefully added with a serological pipette to a Sepmate-15™ (STEMCELL Technologies) tube, prepared with 3.5mL of Lymphoprep™. The Sepmate-15™ tubes were then centrifugated at 1200g for 10 mins at room temperature. After centrifugation, the top layer of supernatant was discarded using a sterile pipette, without disturbing the PBMC layer, and the remaining supernatant with PBMCs was poured off into a new 15 mL falcon tube and 10 mL of 2% FBS in PBS was added. For the first wash, the tube was centrifuged at 300g for 11 mins at room temperature. After centrifugation, the supernatant was again carefully removed with a sterile pipette and the PBMCs were re-suspended by flicking gently, and 10 mL of 2% FBS in PBS was added for a second wash. The tube was centrifuged at 300g for 11 mins at room temperature and supernatant removed. PBMCs were then re-suspended and diluted in a final volume of 1 mL of 2% FBS in PBS. Cells were then counted using a CASY counter. Following cell counting, a final wash was conducted by adding 10mL of 2% FBS in PBS and centrifuged at 300g for 11 minutes at room temperature, with all supernatant removed. PBMCs were then resuspended in neat FBS at a final concentration of 10^7 cells per mL and incubated at 4ºC for 30 minutes. An equal volume of ice-cold freezing media (20% DMSO in FBS) was then added,

diluting the number of cells to 5 x 10^6 cells per mL. PBMCs were then gently mixed by pipetting and immediately aliquoted in volumes of 1mL into labelled cryovials. These were then placed in CoolCell® containers and transferred immediately to a -80ºC freezer, before being transferred to liquid nitrogen for longer term storage.

2.2.4.2. FACS staining for T-cells

For VAC63B only (secondary infection), whole blood collected in EDTA tubes at C-1 and day of diagnosis, was processed for cell staining. 40 mL of 1X erythrocyte lysis buffer was added to 3mL of whole blood in a 50mL falcon tube. This was incubated for 10-15 minutes at room temperature, with constant mixing by inversion. The tube was then centrifuged at 400g for 5 minutes. The supernatant was then removed and the cell pellet resuspended in 1mL FACS buffer. This was transferred to a new 15 mL falcon tube and washed with 13 mL FACS buffer followed by centrifugation at 400g for 5 minutes. The supernatant was then removed and the sample resuspended in 300μL FACS buffer. After adjusting for a concentration of 1 x 10^7 cells / 300 μL the staining protocol was started. First, 20 μL of FC block was added, gently mixed and incubated at room temperature for 5 minutes, before 80 μL of the following antibody mix was added, gently mixed and incubated at 4°C for 20 minutes:

- CD3 BV421 clone OKT3 - 8μL
- CD38 BV510 clone HIT2 - 8μL
- CCR7 BV605 clone G043H7 - 20μL
- HLADR FITC clone L243 - 8μL
- CD4 PerCP Cy5.5 clone OKT4 - 12μL
- CD127 PE clone A019D5 - 8μL
- CD45RA PE Cy7 clone HI100 - 8μL
- CD25 APC clone M-A251 - 8μL

The stained sample was then washed with 3 mL of cold FACS buffer spun down at 400g for 5 minutes at 4°C. This washing step was repeated before the sample was resuspended in 400 μL FACS buffer and kept cool until cell sorting. Cell population frequencies were determined using a FACSAria–III instrument, by Dr Wiebke Nahrendorf (University of Edinburgh).

2.2.5. Human TCRβ amplification

2.2.5.1. List of materials

Reagent	Supplier	Catalogue Number
Quick-RNA MiniPrep Plus Kit	Zymo Research	R1057
RNAse Inhibitor	Clontech	2313A
SMARTScribe™ Reverse Transcriptase	Clontech	639537
First Strand Buffer	Clontech	639537
DTT	Clontech	639537
dNTP Mix 10mM	ThermoFisher Scientific	R0192
Uracyl DNA glycosylase (5U/ μL)	New England BioLabs	M0280S
Phusion flash high-fidelity PCR mastermix	ThermoFisher Scientific	F548L
Monarch PCR Gel-extraction Kit	New England BioLabs	T1020L

All RNA extractions were carried out with appropriate control measures to minimise risk of contamination, including being carried out in a separate 'clean' room, where no concentrated PCR products are handled. RNA was extracted from cryo-preserved PBMC samples using the Quick-RNA MiniPrep Plus Kit (Zymogen) as per kit protocol, including the optional DNAse step. RNA was eluted from columns in 50μL nuclease free water. A modified protocol from Mamedov *et al.* (2013) was then used to

synthesise cDNA and amplify the TCRβ region: 2μL of 20μM BC1R (primer for cDNA synthesis) was added to 8μL of each RNA preparation, heated to 72°C for 2 minutes and then cooled to 42°C for at least 3 minutes. cDNA was then synthesised by adding the following to each sample: 4μL First-strand Buffer, 2μL 10mM dNTP mix, 0.5μL 20mM DTT, 2μL of 10μM SMARTNNN template switch oligo, 0.5μL RNase inhibitor, and 2μL (100U/μL) SMARTScribe™ reverse transcriptase. Samples were then incubated at 42°C for 50 minutes, before the reaction was terminated by heating at 70°C for 10 minutes. cDNA was then treated with 1μl of Uracyl DNA glycosylase (5 U/μl) and incubated for 15 minutes at 37°C. A nested PCR was then used to generate TCRβ V-region amplicons, with outer primers using indexed forward primers composed of the SMART synthesis oligo sequence fused to a P7 Illumina tag (P7-SMART-Index), and a reverse primer within the TCR-C region fused to a P5 Illumina tag (P5-BCJ) (Table 2.3). A full list of the indexed forward primers used is presented in Appendix 2.1. All PCR amplifications were performed on the same day as RNA extractions and cDNA synthesis, according to the following optimised protocol:

- First PCR amplification:
    - Phusion flash high-fidelity mastermix - 10μL
    - cDNA - 2μL
    - Smart_stepout_1 primer (10μM) - 2μL
    - BC2R primer (10μM) - 2μL
    - Nuclease free water - 4μL

Reactions were heated to 98°C for 2 minutes, followed by 18 cycles of amplification using the following program: 98°C for 5s, 68°C for 10s and 72°C for 15s; final elongation at 72°C for 4 minutes.

- Second PCR amplification:
    - Phusion flash high-fidelity mastermix - 10μL
    - First PCR product - 2μL
    - P7-SMART-Index (10μM) - 2μL
    - P5-BCJ (10μM) - 2μL
    - Nuclease free water - 4μL

Reactions were heated to 98°C for 2 minutes, followed by 12 cycles of amplification using the following program: 98°C for 5s, 68°C for 10s and 72°C for 15s; final elongation at 72°C for 4 minutes.

Table 2.3

Primer	Application	Sequence
BC1R	cDNA synthesis	CAGTATCTGGAGTCATTGA
SMART-NNN	Template switch	AAGCAGUGGTAUCAACGCAGAGUNNNNUNNNNUNNNNUCTTrG$_{(3)}$
Smart_stepout1	Nested forward 1	CACTCTATCCGACAAGCAGTGGTATCAACGCAG
BC2R	Nested reverse 1	TGCTTCTGATGGCTCAAACAC
P7-SMART-Index	Forward	CAAGCAGAAGACGGCATACGAGATXXXXXXGGCGAAGCAGTGGTATCAACGCAGAGT
P5-BCJ	Reverse	AATGATACGGCGACCACCGAGATCTACACACACSTTKTTCAGGTCCTC

Table 2.3: Table of human TCRβ cDNA synthesis and amplification primers.

Amplified products were then stored at -20°C. Amplified products were pooled and purified by extraction from excised agarose gel bands using a PCR-gel extraction kit.

2.2.6. Human TCRβ sequencing

Asymmetric 400bp+100bp (to incorporate UMIs) sequencing was performed initially on an Illumina MiSeq nano platform to assess primer design, before being sequenced on an Illumina NovaSeq platform to ensure adequate depth of sequencing. Custom read primers were used, presented in Table 2.4. Library quality and concentration was assessed by Nanodrop prior to submission to Edinburgh Genomics. Edinburgh Genomics provided further quality control and fragment size analysis using a Bioanalyzer, as well as assessing read quality post-sequencing using inbuilt NovaSeq

software. Quality was further assessed after receiving the raw fastq files, using FastQC (Andrews, 2010).

Table 2.4

Primer	Purpose	Primer Sequence (5' to 3')
TCR_read_1 primer	Read 1	CGAGATCTACACACACSTTKTTCAGGTCCTC
TCR_read_2_primer	Read 2	GGCGAAGCAGTGGTATCAACGCAGAGT
TCR&BCR_index1	Index	ACTCTGCGTTGATACCACTG

Table 2.4: Table of human TCRβ custom sequencing read primers.

2.3. TCR repertoire analysis

2.3.1. General considerations

The majority of bioinformatic analyses in this thesis were performed using Python (v3.7.6) scripts, with plotting performed using Seaborn (v0.10.0) and Matplotlib (v3.1.0). Pandas (v1.0.0) was used for data analysis and manipulation. Large data sets were processed using node-distributed parallel computing resources, which were provided by the Edinburgh Compute and Data Facility (ECDF) (http://www.ecdf.ed.ac.uk/).

2.3.2. Raw data processing

Sample demultiplexing and adapter trimming was performed by Edinburgh Genomics. Raw repertoire sequencing data that incorporated UMIs were then further demultiplexed to the UMI level using MiGEC (Shugay *et al.*, 2014). All sequences were aligned using MiXCR software (Bolotin *et al.*, 2015), utilising IMGT nomenclature (Lefranc, 1999). For TCRβ amplicons, the MiXCR analyze amplicon command was used, and for publicly available RNA-seq data sets, the MiXCR analyze shotgun command was used. MiGEC and MiXCR standard error correction thresholds were used. These include using a minimal nucleotide quality score of 20 within the target

gene region. Custom pipelines of Python scripts and VDJtools software (Shugay *et al.*, 2015) (where stated) were then used to analyse and plot the MiXCR output. Only in-frame functional CDR3 sequences, as determined by MiXCR alignment, were included in analyses.

2.3.3. Calculation of diversity and sharing indices

2.3.3.1. Jaccard index

The Jaccard index is a normalised sharing metric between two sets of items. It is defined as the size of the intersection of those two sets (how many items are found in both sets) divided by the size of the union (how many different items are found in total between the two sets). In this thesis, Jaccard indices were calculated using a custom Python function, and calculated both weighted (taking into account clone frequency) and unweighted as stated:

$$Jaccard\ (A,B) = \frac{|A \cap B|}{|A \cup B|}$$

The Jaccard Index ranges from 0 to 1: an index of 1 indicates identical sets, whilst an index of 0 indicates no overlap at all between sets.

2.3.3.2. Similarity networks

Similarity network analysis was undertaken using the python-Levenshtein module in Python. The Levenshtein distance can be considered as an 'edit distance' and represents the distance between two TCRβ CDR3 amino acid sequences, i.e. the minimum number of mutations, deletions and insertions needed to transform the first sequence into the second sequence. Networks were created both within individual and within replicate groups, with nodes representing unique CDR3 amino acid sequences, and an edge drawn between two nodes if the CDR3 amino acid sequences they represent are within a Levenshtein distance of 0 (identical) or 1 (1 amino acid

mismatch) of each other. All network visualisation and analyses were carried out using Gephi software (Bastian, Heymann and Jacomy, 2009).

2.3.3.3. Swarm clustering

To cluster similar CDR3 amino acid sequences within 1 amino-acid mismatch of each other, a modified Swarm algorithm (Mahe *et al.*, 2014) was used. This algorithm was customised and accuracy tested for use with immune repertoire data by Noramon Dron (University of Edinburgh MSc Bioinformatics Student, unpublished data). To ensure sensitivity and prevent over-grouping, the program was set to perform at a threshold of only 1 amino-acid mismatch to the initial seed sequence of each cluster (with initial cluster seed determined by abundance).

2.3.3.4. Simpson's diversity index

To determine diversity of TCR repertoires, Simpson's diversity index was used. For TCR repertories, this represents the probability that an identical TCR (as determined by identical CDR3 amino acid sequence) will be drawn from the repertoire with two independent draws. Repertoire diversity was calculated using Simpson's diversity index only on size-matched repertoires, by subsampling an equal number of UMI-labelled cDNA molecules, for the precise normalisation required for comparing such diversity data (Izraelson *et al.*, 2018). Simpsons Diversity Index was calculated using a custom Python function as follows:

$$SD = \sum_{i=1}^{a} \frac{ni(ni-1)}{n(n-1)}$$

Where *ni* is the clone size of the *i*th clonotype (count of each clonotype), *a* is the number of different clonotypes in the TCR sample, and n is the total number of TCR sequences for that sample. Simpson's Diversity Index ranges between 0 and 1: an

index of 0 indicates infinite diversity within a repertoire, whilst an index of 1 indicates no diversity at all.

2.3.3.5. Renyi entropy

The Renyi entropy was used to create a diversity profile for each of the CHMI repertoires (Rempala and Seweryn, 2013, Greiff et al., 2015;). Renyi's entropy is defined as:

$$H_\alpha = \frac{\log \sum_{i=1}^{s} p_i^\alpha}{1-\alpha}$$

Where p_i is the frequency of each clone, and s the total number of clones in the repertoire. The parameter α can be varied from 0 to ∞ to generate a diversity profile. At $\alpha = 0$, the entropy will be $\log s$ – simply a count of the number of clones present, a measure of the richness of the repertoire. As α is increased, greater weight is applied to more frequent clones. As $\alpha \to \infty$, the measure tends to only consider the proportional number of sequences in the largest clonotype.

2.3.4. Data normalisation

Where UMI's were incorporated during sequencing and where stated, for depth-dependent analyses, repertoires have been normalised by down-sampling to a set number of unique UMI-labelled reads. VDJtools software (Shugay *et al.*, 2015) was used to either down-sample to an exact number of reads, or to select the top *n* most abundant clones: for down-sampling, the DownSample function was used, which randomly selects a pre-defined number of reads following demultiplexing at the UMI-level. Where the top *n* most abundant clones were required, the SelectTop function was used.

2.3.5. Statistical analyses

Unless stated otherwise, all statistical analyses was performed using SciPy Python software (Jones, Oliphant and Peterson, 2001).

2.3.5.1. Murine

For comparisons between different time-points but within replicate group, either a parametric Students t-test or non-parametric unpaired Mann-Whitney U-test was used, dependent on data distribution. For comparing groups between different time-points of infection, one-way ANOVA was used.

2.3.5.2. CHMI

As the CHMI TCRβ repertoires obtained were longitudinal, to determine changes in TCRβ metrics through time, linear mixed-effects models were fitted for each metric. These were undertaken using the *lme4* package (Bates *et al.*, 2015) in R, with post-hoc pairwise analyses carried out using the *emmeans* package (Lenth, 2021). Specific details regarding models used for each analysis are given in Chapter 5.

2.4. Appendices

Appendix 2.1: Table of TCRβ indexing primers (P7-SMART-Index)

Primer Name	Primer Sequence
P7-SMARTamp1	CAAGCAGAAGACGGCATACGAGATAGCTCTGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp2	CAAGCAGAAGACGGCATACGAGATGATCCTGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp3	CAAGCAGAAGACGGCATACGAGATCTAGCTGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp4	CAAGCAGAAGACGGCATACGAGATTCGACTGGCGAAGCAGTGGTATCAACGCAGAGT

P7-SMARTamp5	CAAGCAGAAGACGGCATACGAGATCAGTGTGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp6	CAAGCAGAAGACGGCATACGAGATTGACGTGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp7	CAAGCAGAAGACGGCATACGAGATACTGGTGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp8	CAAGCAGAAGACGGCATACGAGATGTCAGTGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp9	CAAGCAGAAGACGGCATACGAGATTACGATGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp10	CAAGCAGAAGACGGCATACGAGATGACTTCGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp11	CAAGCAGAAGACGGCATACGAGATAGTCTCGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp12	CAAGCAGAAGACGGCATACGAGATTCAGTCGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp13	CAAGCAGAAGACGGCATACGAGATCTGATCGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp14	CAAGCAGAAGACGGCATACGAGATCCTTCCGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp15	CAAGCAGAAGACGGCATACGAGATAAGGCCGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp16	CAAGCAGAAGACGGCATACGAGATGGAACCGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp17	CAAGCAGAAGACGGCATACGAGATATATGCGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp18	CAAGCAGAAGACGGCATACGAGATTGGTACGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp19	CAAGCAGAAGACGGCATACGAGATCAACACGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp20	CAAGCAGAAGACGGCATACGAGATGTTGACGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp21	CAAGCAGAAGACGGCATACGAGATACCAACGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp22	CAAGCAGAAGACGGCATACGAGATACGTTGGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp23	CAAGCAGAAGACGGCATACGAGATGTACTGGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp24	CAAGCAGAAGACGGCATACGAGATCTAGTGGGCGAAGCAGTGGTATCAACGCAGAGT

P7-SMARTamp25	CAAGCAGAAGACGGCATACGAGATTGCATGGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp26	CAAGCAGAAGACGGCATACGAGATGCCGCGGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp27	CAAGCAGAAGACGGCATACGAGATATTACGGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp28	CAAGCAGAAGACGGCATACGAGATGGTTGGGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp29	CAAGCAGAAGACGGCATACGAGATAACCGGGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp30	CAAGCAGAAGACGGCATACGAGATCCAAGGGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp31	CAAGCAGAAGACGGCATACGAGATCTCTAGGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp32	CAAGCAGAAGACGGCATACGAGATTCTCAGGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp33	CAAGCAGAAGACGGCATACGAGATAGAGAGGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp34	CAAGCAGAAGACGGCATACGAGATGAGAAGGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp35	CAAGCAGAAGACGGCATACGAGATCGATTAGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp36	CAAGCAGAAGACGGCATACGAGATGCTATAGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp37	CAAGCAGAAGACGGCATACGAGATGTGTCAGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp38	CAAGCAGAAGACGGCATACGAGATACACCAGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp39	CAAGCAGAAGACGGCATACGAGATTGTGCAGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp40	CAAGCAGAAGACGGCATACGAGATCACACAGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp41	CAAGCAGAAGACGGCATACGAGATTCCTGAGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp42	CAAGCAGAAGACGGCATACGAGATCTTCGAGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp43	CAAGCAGAAGACGGCATACGAGATGAAGGAGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp44	CAAGCAGAAGACGGCATACGAGATAGGAGAGGCGAAGCAGTGGTATCAACGCAGAGT

P7-SMARTamp45	CAAGCAGAAGACGGCATACGAGATAATTAAGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp46	CAAGCAGAAGACGGCATACGAGATGGCCAAGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp47	CAAGCAGAAGACGGCATACGAGATCCGGAAGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp48	CAAGCAGAAGACGGCATACGAGATTAGCTAGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp49	CAAGCAGAAGACGGCATACGAGATATCGTAGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp50	CAAGCAGAAGACGGCATACGAGATTAATCGGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp51	CAAGCAGAAGACGGCATACGAGATCGGCCGGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp52	CAAGCAGAAGACGGCATACGAGATGCGCGCGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp53	CAAGCAGAAGACGGCATACGAGATCGCGGCGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp54	CAAGCAGAAGACGGCATACGAGATTATAGCGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp55	CAAGCAGAAGACGGCATACGAGATGCATATGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp56	CAAGCAGAAGACGGCATACGAGATATGCATGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp57	CAAGCAGAAGACGGCATACGAGATCGTAATGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp58	CAAGCAGAAGACGGCATACGAGATGCCTCTGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp59	CAAGCAGAAGACGGCATACGAGATGCGAGTGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp60	CAAGCAGAAGACGGCATACGAGATGATGCCGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp61	CAAGCAGAAGACGGCATACGAGATCGAGTCGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp62	CAAGCAGAAGACGGCATACGAGATACGTACGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp63	CAAGCAGAAGACGGCATACGAGATCGCTCAGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp64	CAAGCAGAAGACGGCATACGAGATATCTGAGGCGAAGCAGTGGTATCAACGCAGAGT

P7-SMARTamp65	CAAGCAGAAGACGGCATACGAGATATCTATGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp66	CAAGCAGAAGACGGCATACGAGATGCTAGAGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp67	CAAGCAGAAGACGGCATACGAGATATCAGGGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp68	CAAGCAGAAGACGGCATACGAGATTAGAGCGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp69	CAAGCAGAAGACGGCATACGAGATCGAAGTGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp70	CAAGCAGAAGACGGCATACGAGATCTGACAGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp71	CAAGCAGAAGACGGCATACGAGATTCAACGGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp72	CAAGCAGAAGACGGCATACGAGATAGTACCGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp73	CAAGCAGAAGACGGCATACGAGATGACACTGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp74	CAAGCAGAAGACGGCATACGAGATGTAATCGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp75	CAAGCAGAAGACGGCATACGAGATGTCGAAGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp76	CAAGCAGAAGACGGCATACGAGATACTGAGGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp77	CAAGCAGAAGACGGCATACGAGATTGAGACGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp78	CAAGCAGAAGACGGCATACGAGATCAGGATGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp79	CAAGCAGAAGACGGCATACGAGATAGGGGAGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp80	CAAGCAGAAGACGGCATACGAGATCTTGGCGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp81	CAAGCAGAAGACGGCATACGAGATTCCGGTGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp82	CAAGCAGAAGACGGCATACGAGATTATGCAGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp83	CAAGCAGAAGACGGCATACGAGATCCAGTAGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp84	CAAGCAGAAGACGGCATACGAGATTTGGTGGGCGAAGCAGTGGTATCAACGCAGAGT

P7-SMARTamp85	CAAGCAGAAGACGGCATACGAGATAACGTCGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp86	CAAGCAGAAGACGGCATACGAGATGGTGTTGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp87	CAAGCAGAAGACGGCATACGAGATCGTCAAGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp88	CAAGCAGAAGACGGCATACGAGATTACCAGGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp89	CAAGCAGAAGACGGCATACGAGATTAGCACGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp90	CAAGCAGAAGACGGCATACGAGATGCACATGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp91	CAAGCAGAAGACGGCATACGAGATGGCCGCGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp92	CAAGCAGAAGACGGCATACGAGATAATCGTGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp93	CAAGCAGAAGACGGCATACGAGATACCCCAGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp94	CAAGCAGAAGACGGCATACGAGATGTTCCGGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp95	CAAGCAGAAGACGGCATACGAGATTGGCCTGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp96	CAAGCAGAAGACGGCATACGAGATGAGCTAGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp97	CAAGCAGAAGACGGCATACGAGATAGACTGGGCGAAGCAGTGGTATCAACGCAGAGT
P7-SMARTamp98	CAAGCAGAAGACGGCATACGAGATTCTCTCGGCGAAGCAGTGGTATCAACGCAGAGT

Chapter 3: Analysis of the splenic TCRβ repertoires generated during a murine malaria infection.

This chapter contributes to the following publication in Frontiers in Immunology: N.L. Smith, W. Nahrendorf, C. Sutherland, J.P. Mooney, J. Thompson, P.J. Spence, G.J.M. Cowan. 2020. A Conserved TCRβ signature dominates a highly polyclonal T-cell expansion during the acute phase of a murine malaria infection. Frontiers in Immunology 11: doi 10.3389/fimmu.2020.587756

3.1. Abstract

CD4$^+$ αβ T-cells are key mediators of the immune response to a first *Plasmodium* infection, undergoing extensive activation and splenic expansion during the acute phase of an infection. However, the clonality and clonal composition of this expansion has not previously been described. Here, using a comparative infection model, I analysed the splenic CD4$^+$ T-cell receptor repertoires generated over the time-course of a *Plasmodium chabaudi (AS)* infection. Through both repeat replicate experiments and analysis of publicly available RNA-seq data, I found that following a first infection – within a highly polyclonal expansion – acute T-effector repertoires are consistently dominated by TRBV3 gene usage. Clustering by sequence similarity, the same dominant clonal signature was found to be expanded across replicates in the acute phase of an infection, revealing a conserved T-cell response that is consistently a hallmark of a first infection. Determining the host or parasite factors driving this conserved response may uncover novel immune targets for malaria therapeutic purposes.

3.2. Introduction

Although protective natural immunity against clinical malaria is slow to develop and requires years of repeated exposure (Ryg-Cornejo, Ly and Hansen, 2016), protection

against severe disease is obtained after a more limited number of symptomatic infections (Gupta *et al.*, 1999; Gonçalves *et al.*, 2014). The acquisition of this naturally acquired immunity is mediated by both antibody (reviewed in (Doolan, Dobaño and Baird, 2009; Ly and Hansen, 2019)) and T-cell responses (Nlinwe *et al.*, 2018); the latter being crucial for B-cell class switching and affinity maturation. As well as guiding the humoral response, $CD4^+$ T-cells play a key role in restricting the growth and pathogenesis of blood-stage *Plasmodium* through cytokine secretion and macrophage activation (reviewed in (Kurup, Butler and Harty, 2019)). However, the antigenic drivers and developmental dynamics underlying this naturally acquired immunity remain poorly understood, presenting major challenges for effective vaccine design.

In animal models of malaria, a *Plasmodium* infection in previously unexposed individuals initially produces a massive expansion of $CD4^+$ T-cells in the spleen (Spence *et al.*, 2013; Opata and Stephens, 2017), a major site of the developing immune response (del Portillo *et al.*, 2012a). The size of this response, together with the generation of a highly diverse range of cellular responses, suggests that the splenic expansion of $CD4^+$ populations is highly polyclonal, as opposed to the expansion of a minor (oligoclonal) subset of the repertoire. However, it is not known whether this expansion is primarily a non-specific response, such as a result of cytokine-driven bystander activation, or whether it is dominated by antigen-specific responses generated through classic TCR-engagement mediated clonal expansion. Spectratyping (CDR3 length analysis) of T-cell receptor (TCR) β chain repertoires induced by the rodent malaria *Plasmodium berghei* has previously detected a unique TCRβ CDR3 length signature enhanced over the course of infection, suggesting that there may be a clonal response to specific antigenic peptides (Mariotti-Ferrandiz *et al.*, 2016). In agreement with this, an expanded fraction of $CD4^+$ T-cells and fast-responding cytokine secretors that respond to a secondary challenge has been observed following a *Plasmodium chabaudi (AJ)* infection in mice, indicating initial priming by the parasite, and the presence of pathogen-specific T-cells within the $CD4^+$ T-cell population (Opata and Stephens, 2017). Alternatively, there is evidence from *P. falciparum,* that the PfEMP1 binding domain, CIDR-1α, stimulates $CD4^+$ T-cells non-specifically through TCR-independent pathways (Ndungu *et al.*, 2006), and that regulatory T-cell (Treg) proliferation during an infection can be induced in an antigen

non-specific manner (Scholzen *et al.*, 2009). Non-specific proliferation of T-cells due to cross-reactivity in response to *P. falciparum* antigens has also been reported (Wipasa *et al.*, 2011). Overall, proliferation is likely to be a combination of activation dynamics. However, whether a detectable clonal malaria-specific $CD4^+$ T cell response that is conserved between individuals, and thus a potential focal target for therapeutics, is induced, has not previously been demonstrated.

Advances in high-throughput TCR repertoire sequencing techniques now allow deep profiling of immune responses. This approach has been used to ascertain clonality of T-cell responses, identify expanded T-cell clones and determine if conserved or 'public' responses between individuals are generated following antigenic stimulation (reviewed in (Bradley and Thomas, 2019)). Repertoire sequencing thus provides a novel, immune-focused approach to delineate the clonality of the developing immune response to a malaria infection.

Here, using bulk TCRβ repertoire sequencing I examine the dynamics and clonal structure of the splenic $CD4^+$ T-cell repertoires generated during infection with the well-established mouse malaria model *Plasmodium chabaudi (AS)*. By comparing serially blood passaged (SBP) and recently mosquito-transmitted (MT) *P. chabaudi* infections, Spence *et al.* (2013) demonstrated that vector transmission of *P. chabaudi* intrinsically modified parasite gene expression in asexual blood-stage parasites, eliciting an altered host immune response that in turn regulates parasite virulence. In this model, infection with SBP parasites leads to hyperparasitaemia with more severe disease during the acute phase of infection. In contrast, mosquito transmission (MT) attenuates parasite growth and virulence, through a mechanism associated with epigenetic reprogramming of the expression of the subtelomeric multigene families, including the variant surface antigen (VSA) family. I have used this comparative model to compare TCR repertoires integral to an immune response that rapidly controls parasite growth, against a less effective response that fails to control parasite replication and induces immunopathology (Spence *et al.*, 2013; Spence, Brugat and Langhorne, 2015). I sequenced the T-naïve (T_N), T-effector (T_E), T-effector memory (T_{EM}) and T-central memory (T_{CM}) $CD4^+$ splenic TCRβ repertoires elicited in mice over the time-course of both MT and SBP *P. chabaudi* infections. For both infection types, the T_E expansion

seen during the acute phase of a *P. chabaudi* infection is highly polyclonal. However, within this diverse expansion, a conserved pathogen-specific response characterised by TRBV3 gene usage consistently dominates the effector repertoire following a first infection.

3.3. Methods

3.3.1. Mice infections

All mouse infections were carried out according to Methods 2.1.1. Data presented in this chapter was generated from 3 independent experiments, PC01, PC02 and PC03.

3.3.1.1. Experiment PC01

For the first experiment presented in this chapter (PC01), 50 C57BL/6 mice were infected with *P. chabaudi* (AS) either by intra-peritoneal (IP) injection of ~10^5 serially blood passaged (SBP) parasitised erythrocytes (n=25) or by IP injection of ~10^5 parasitised erythrocytes from a mouse previously infected by mosquito transmission (MT, n=25) (Spence *et al*, 2012). Mice (n=5 from each infection group) were humanely euthanased on days 6, 10, 20, 40 and 60 post- infection, and at each time point, serum was collected, spleens harvested and Giemsa-stained thin blood films prepared. Parasitaemia levels were determined from thin blood films at all available timepoints. Splenic CD4+ cell populations of interest to be sequenced - T_N ($CD62L^+$ $CD127^+$), T_E ($CD62L^-$ $CD127^-$), T_{CM} (CD44HI $CD127^+$ $CD62L^+$), and T_{EM} (CD44HI $CD127^+$ $CD62L^-$) were isolated by FACS according to Methods 2.1.2. An example of gating strategy used is presented in Appendix 3.1.

3.3.1.2. Experiment PC02

The second experiment, PC02, was undertaken to confirm key findings from PC01, and to include a finer precision of time-points in the acute phase of infection. Only MT infections were included in this repeat experiment, representing the most natural route of infection. Mice (n=12) were infected by IP injection of ~10^5 recently MT *P. chabaudi (AS)-GFP* parasitised erythrocytes only, with challenged mice (n=4) euthanised at days 4, 7 and 11 post infection, and unchallenged mice (n=6) used for control purposes. For this follow-up experiment, only T_N ($CD62L^+$ $CD127^+$) and T_E ($CD62L^-$

CD127$^-$) cell populations were isolated for TCRβ repertoire sequencing, by FACS according to Methods 2.1.2.

3.3.1.3. Experiment PC03

The third experiment presented in this chapter, PC03, was an experiment designed to determine TCRβ splenic repertoire composition following a homologous *P. chabaudi (AS)* re-challenge. For this experiment, mice (n=12) were infected with MT *P. chabaudi (AS)-GFP* by intra-peritoneal injection of 1x10^5 parasitised erythrocytes. Mice (n=8), were euthanised at days 7 and day 67 post-infection. Mice (n=4) underwent a homologous re-challenge at day 60 post infection and were euthanised 7 days post re-challenge (day 67). Control mice consisted of unchallenged mice (n=8), with 4 unchallenged mice euthanised with the day 7 cohort, and 4 with the day 67 cohort, to ensure age-matched controls. To eliminate chronic infection before re-challenge, 0.288mg/ml of chloroquine diphosphate salt (Sigma), supplemented with glucose for palatability, was dissolved in drinking water daily for 10 days, from day 30 to day 40 post-infection (Lewis, Pfeil and Mueller, 2011). For this follow-up experiment only T_N (CD62L$^+$ CD127$^+$) and T_E (CD62L$^-$ CD127$^-$) cell populations were isolated for TCRβ repertoire sequencing, by FACS according to Methods 2.1.2.

3.3.2. TCRβ repertoire generation

All TCRβ splenic repertoires were generated according to Methods 2.1.4. Briefly, at each time point, spleens were harvested and CD4$^+$ splenic T-cells populations of interest were isolated by FACS before RNA was extracted, reverse transcribed, and TCRβ chains were amplified and sequenced on an Illumina MiSeq platform. All data generated and presented in this chapter has been made publicly available on the European Nucleotide Archive (https://www.ebi.ac.uk/ena), accession PRJEB40867. Clone counts for experiments PC01, PC02 and PC03 are presented in Appendices 3.2 to 3.7.

3.3.2.1. Use of UMIs for PCR error correction

For experiment PC01, a previously utilised 5-RACE technique was used to generate the TCRβ amplicon libraries (Sood *et al.*, 2016). This technique avoids the library

amplification bias of certain V genes inherent in multiplex PCR techniques. However PCR error and base-calling errors are still inevitable in HTS techniques, and can lead to irregular clonal counts that do not reflect sample input. Unique Molecular Identifiers (UMIs) – short stretches of random nucleotides incorporated during cDNA synthesis - were not previously commonly used in TCR repertoire studies at the start of this project, but have become increasingly utilised, and currently represent the gold-standard method to ensure unbiased quantitative analysis of TCR repertoires. Therefore, for the follow-up experiments PC02 and PC03, a 12 nucleotide UMI was incorporated during cDNA synthesis (5-RACE-UMI technique). This is used to identify, post-amplification, which sequences originated from the same starting mRNA molecule. Demultiplexing at the UMI level was undertaken using MiGEC (Shugay *et al.*, 2014), a software specifically designed for UMI-based analysis of immune repertoire data, which collapses TCRs with the same UMI, as well as removing or collapsing erroneous sequences that have an UMI sequence with 1 mismatch from a TCR sequence with the same consensus sequence.

3.3.2.2. Bulk TCR repertoire analyses

Bulk TCR sequence data was initially processed using MiGEC (Shugay *et al.*, 2014) and MiXCR (Bolotin *et al.*, 2015) software with default settings. Samples were excluded from further analyses if the repertoire contained fewer than 10,000 total reads after processing, as this was indicative of poor sample preservation or preparation. A combination of custom pipelines of Python scripts and VDJtools software (Shugay *et al.*, 2015) was used to analyse and plot the MiXCR output. All statistical analyses were performed using SciPy Python software (Jones, Oliphant and Peterson, 2001). A TCR clone was defined by 100% amino acid sequence identity of the CDR3 region, and IMGT nomenclature used for gene usage. Only in-frame (functional) CDR3s were analysed. A modified version of the Swarm algorithm (Mahe *et al.*, 2014) - altered to accommodate protein rather than just nucleotide sequence - was used to cluster highly homologous CDR3 amino sequences, with identical V-gene usage, within 1 amino-acid mismatch of each other. The Gliph2 package (Huang *et al.*, 2020) was also used to identify enriched amino acid motifs within the contact region of CDR3 sequences; the unchallenged T-naïve repertoires were used to make custom murine reference files for this. Network analyses was undertaken using Gephi

(Bastian, Heymann and Jacomy, 2009) software (v0.9.2). Generation probability of TCRs (Pgen) was calculated using OLGA (Sethna *et al.*, 2018).

3.3.3. Publicly available RNA-seq data

To verify findings from my own data generation, I mined publicly available RNA-seq data sets obtained from the spleens of C57Bl/6 mice infected with blood-stage *Plasmodium chabaudi (AS)* and *Plasmodium chabaudi (CB).* Raw fastq files were downloaded from the ArrayExpress archive (https://www.ebi.ac.uk/arrayexpress/), accession numbers ENA-ERP004042 (data set E-ERAD 221) and ENA-ERP005730 (data-set E-ERAD 289) respectively. Both of these data sets were generated by Professor Jean Langhorne's research group (The Francis Crick Institute), and were made publicly available in 2013 and 2014 respectively. These data sets were aligned using MiXCR software (Bolotin *et al.*, 2015), using the analyze shotgun command, which has been specifically designed to extract immune receptor transcripts from whole transcriptome data. Only functional in-frame TCRβ sequences were included in analyses.

3.3.4. MSP-1_{21} ELISA and affinity ELISA

Previously, enhanced immune responses including increased antibody responses in MT infections compared to SBP have been described (Spence *et al.*, 2013). To determine if such differences existed in PC01, before comparing immune repertoires, an indirect Ig, IgG and avidity ELISA were undertaken using the known immunogenic MSP-1_{21} protein as the antigenic target. Immunoglobulin (Ig) titres were measured longitudinally over the time-course of infection using an ELISA, as presented in Methods 2.1.6. Serum from naïve C57BL/6 mice (n=5) was used for control purposes. Recombinant MSP-1_{21} protein was produced for use in the indirect ELISA. MSP-1_{21} is the 21-kDa C-terminal region of the merozoite surface protein 1 of *P. chabaudi* - known to be highly immunogenic and a candidate antigen for inclusion in a blood- stage malaria vaccine (Cheesman *et al.*, 2009; Stephens and Langhorne, 2010). To determine IgG-specific titres and to compare the functional affinity of IgG for MSP-1_{21} between the SBP and MT *P. chabaudi* infections, an ELISA affinity assay based on the dissociation of antigen-antibody interactions by chaotropic anions was undertaken. Sera from the infected mice was assayed in the presence and absence of the mild

chaotropic agent ammonium thiocyanate (NH_4SCN), the tolerance to which has been demonstrated to be proportional to antigen-antibody avidity (Ferreira and Katzin, 1995; Achtman *et al.*, 2007; Almanzar *et al.*, 2013). A modified version of the assay used by Ferreira & Katzin (1995) was utilised. An avidity index (AI) – defined as the proportion of antibody bound after exposure to 1M NH_4SCN (AI = [A50 value following incubation with NH_4SCN/A50 value without NH_4SCN]) – was calculated at days 20, 40 and 60 post-infection.

3.4. Results

3.4.1. Infection dynamics

Parasite dynamics for both infection types were in agreement with what has previously been observed with this comparative infection model – mice with SBP infections were hyperparasitaemic but cleared the infection more rapidly than mice with MT infections, who displayed a very low-grade chronic parasitaemia. Consistent with previously published data for *P. chabaudi* (Spence *et al.*, 2013; Mamedov *et al.*, 2018), $CD4^+$ splenic T-effectors (T_E) reached maximum levels of expansion in the acute phase of infection, increasing by up to 10-fold. Expansion coincided with the peak of parasitaemia and contracted back to pre-challenged levels between days 20 and 40 post-infection (Figure 3.1).

Figure 3.1

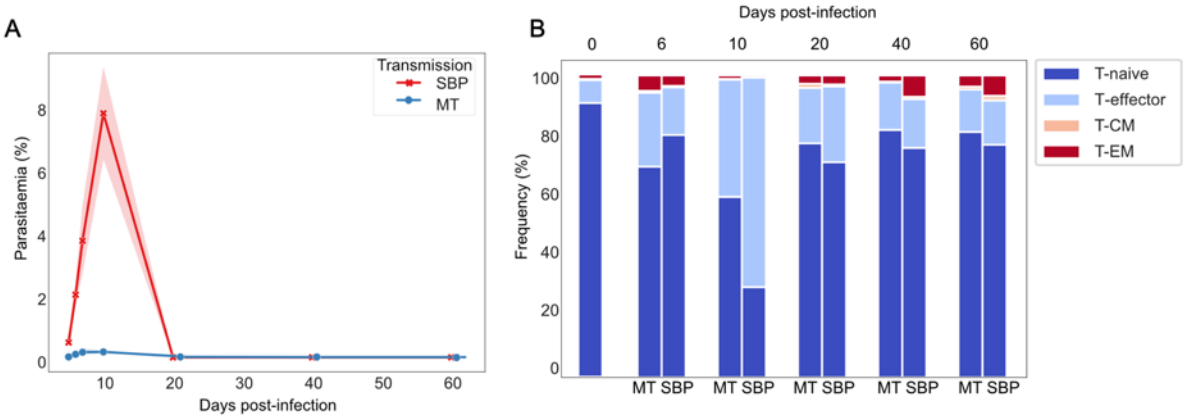


Figure 3.1: Dynamics of *P. chabaudi* infection: T-effector expansion coincides with peak parasiatemia. A) Parasitaemia of C57Bl/6 mice infected with either SBP (red) or recently MT (blue) 5 x10^5 *P. chabaudi* parasitised erythrocytes, n=5 mice per infection type per time point, shaded area depicts 95% CI. B) Phenotypic profiling of splenic $CD4^+$ T cells as determined by FACS. Representative frequencies over the time course of infection of T-naïve $CD4^+$ T cells ($CD62L^+$ $CD127^+$), T-effector $CD4^+$ T cells (T_E) ($CD62L^-$ $CD127^-$), effector memory (T_{EM}) (CD44HI $CD127^+$ $CD62L^-$) and central memory (T_{CM}) (CD44HI $CD127^+$ $CD62L^+$) $CD4^+$T cells.

3.4.2. ELISA Results

3.4.2.1. Protein expression

Recombinant maltose-binding protein (MBP) fused MSP-1_{21} was expressed in BL21 (DE3) *E. coli* cells, transformed with a previously constructed pMAL- p2X (AJ)-MSP-1_{21} plasmid vector and purified using affinity chromatography (Methods 2.1.6.2). Purification was confirmed by visualising expected protein product size using SDS polyacrylamide gel electrophoresis. MBP purified from *E. coli* transfected with pMAL-p2X (AJ) only was used for control purposes throughout the expression and purification process. An SDS polyacrylamide gel was run to visualise expected protein size and confirm protein purification. Expected band size for MBP was 42kDa, and for MBP-MSP-121 combined, 63kDa (Figure 3.2).

Figure 3.2

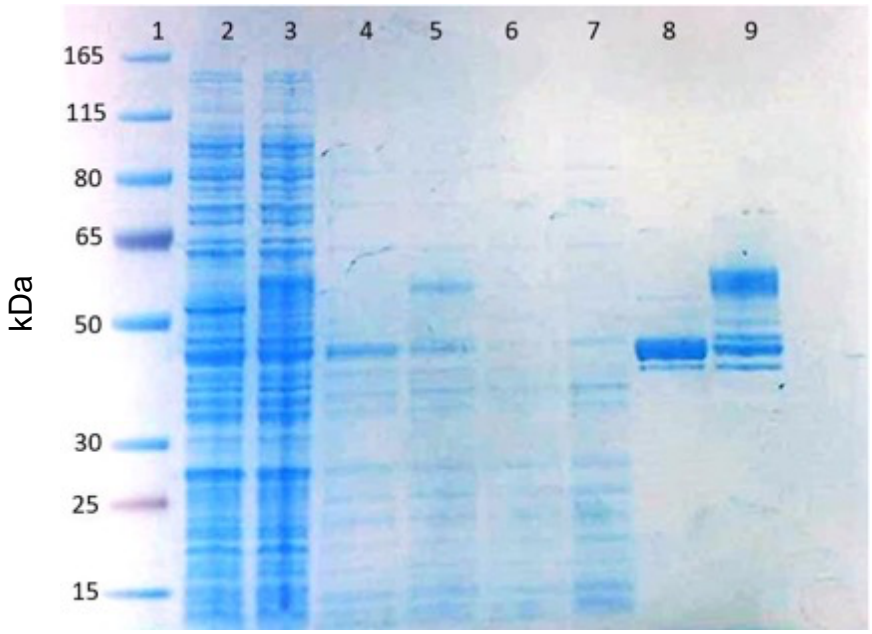


Figure 3.2: SDS PAGE analysis stained with Coomassie blue to confirm purification of MSP-1_{21} protein following affinity chromatography. Plasmid pMAL-p2X (AJ) encoding MBP but not MSP-1_{21} was used for control purposes throughout the protein expression and purification process. Lane 1: protein ladder (PageRuler Plus SM1811), 2: whole cell protein (pMAL-p2X (AJ)), 3: whole cell protein (pMAL-p2X(AJ)-MSP-1_{21}), 4: lysate (pMAL-p2X(AJ)), 5: lysate (pMAL-p2X(AJ)- MSP-1_{21}), 6: chromatography flow through (pMAL-p2X(AJ)), 7: chromatography flow through (pMAL-p2X(AJ)-MSP-1_{21},) 8: fraction 1 (pMAL-p2X (AJ)), 9: fraction 1 (p2X(AJ)- MSP-1_{21}). Expected band size for MBP is 42kDa, MBP-MSP-1_{21} 63 kDa.

3.4.2.2. Ig ELISA results

Immunoglobulin (Ig) titres between transmission types at each time-point were compared using Mann-Whitney U test. Results from the ELISA revealed higher Ig production against MSP-1_{21} in mice with MT *P. chabaudi* infections compared to SBP infections, with a significant difference in Ig titres between the two transmission types present at day 20, 40 and 60 post-infection (Figure 3.3A, Table 3.1).

Figure 3.3

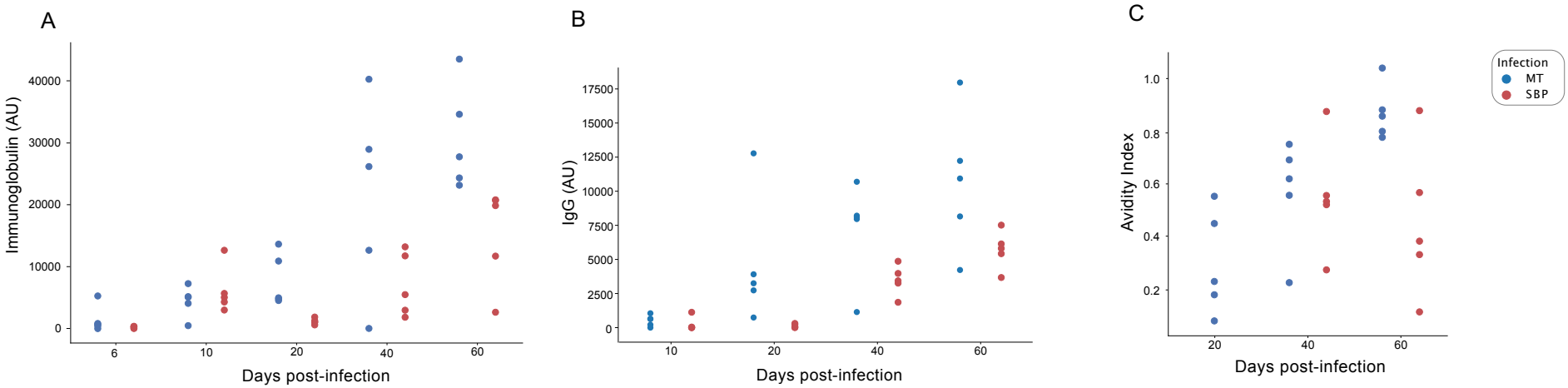


Figure 3.3: Serum MSP-1_{21} Ig (A) and IgG (B) titres following SBP (red) or MT (blue) P. chabaudi (AS) infections at days 6, 10, 20 , 40 and 60 days post-infection. Data is presented as arbitrary units relative to known *P. chabaudi (AS)* positive pooled serum. C) Avidity index (AI) values for MSP-1_{21} IgG from at day 20, 40 and 60 post-infection. AI = [A_{50} value following incubation with NH4SCN/A_{50} value without NH4SCN].

Table 3.1

Days post-infection	Mean Ig Titre (AU±SEM)	p value	Mean IgG Titre (AU±SEM)	p value	Mean Avidity Index (±SEM)	p value
Day 6: SBP	214.1 (±69.9)		< min	NA	NA	NA
MT	1436.6 (±967.7)	0.117	< min	NA	NA	
Day 10: SBP	6128.4 (±1688.6)		236.66 (± 222.87)		NA	NA
MT	4406.51 (±1111.3)	0.754	509.18(± 183.62)	0.347	NA	
Day 20: SBP	1163.3 (±205.5)		130.7 (±59.9)		NA	NA
MT	7765.3 (±1888.9)	0.008	4682.8 (±2091.6)	0.009	0.298 (± 0.089)	
Day 40: SBP	7043.4 (±2303.4)		3479.3 (±492.3)		0.552 (±0.097)	
MT	21603.6 (±6957.5)	0.009	7220.9 (±1601.2)	0.001	0.569 (±0.093)	0.902
Day 60: SBP	15135.78 (±3557.3)		5711.4 (±619.6)		0.454 (±0.130)	
MT	30659.3 (±3778.6)	0.009	10701.3 (±2276.3)	0.017	0.876 (± 0.047)	0.016

Table 3.1: Summary of ELISA results for Ig (yellow), IgG (green) and avidity (orange) to MSP-1_{21} following *P. chabaudi* infection at days 6, 10, 20, 40 and 60 post-infection with either SBP or MT parasites. Data is presented as mean ± SEM. Titres and avidity index between transmission type at each time point were compared using Mann-Whitney U tests.

3.4.2.3. IgG ELISA and avidity assay results

Similarly for IgG, MT infections had significantly higher titres than SBP infections, at days 20, 40 and 60 post-infection (Figure 3.3B, Table 3.1). For the avidity assay, firstly for MT infections, the mean avidity index (AI) increased from 0.298 (± 0.089) on day 20, to 0.569 (± 0.093) by day 40 and 0.876 (± 0.047) by day 60. For the SBP samples, an AI could not be calculated for day 20 as the IgG titres were too low. From days 40 to 60, the mean AI of the SBP samples decreased from 0.552 (± 0.097) to 0.454 (± 0.130), although this change was not significant (p=0.561). Between SBP and MT samples there was a significant difference in AI at day 60 (p=0.016), but this was not detected at day 40 (p=0.902) (Figure 3.3C, Table 3.1).

The results from the MSP-1_{21} ELISA for both total immunoglobulin and IgG mirror what was previously reported by Spence *et al* (2013), who used total-parasite extract as their ELISA antigen. Results indicate enhanced antibody production against MSP-1_{21} in the chronic phase of infection in mice infected with recently MT parasites compared to those infected following SBP and highlight the immunogenicity of this protein. As such, I confirm that in our experiment, similar differences in immune response between the comparative infection model that have previously been documented are present. Antibody avidity measurements are novel for this infection model and indicate a progressive increase in avidity index in MT infections from days 20 to 60, compared to a more temporally static response from SBP infections, which also show a significantly reduced AI at day 60 compared to MT infections. In part agreement with this, Achtman *et al* (2007) previously demonstrated that half of the antibodies produced against MSP-1_{21} in a SBP *P. chabaudi* infection were of low avidity, but no previously published avidity data for *P. chabaudi* MT infections is available for comparative purposes. Similarly, although affinity assays using chaotropic agents are commonly cited in the literature, methods vary, and have not previously been conclusively validated with other methods for measuring antibody-antigen avidity such as serum plasmon resonance (Reddy *et al.*, 2012). As such, these results should be interpreted as a qualitative marker of difference in avidity rather than a quantifiable measure. Nonetheless, they are again in agreement with enhanced immune

responses in MT infections compared to SBP infections, being present in experiment PC01.

3.4.3. TCRβ sequencing: 5'RACE vs 5'RACE-UMI corrected

To ensure adequate PCR error correction, for PC02, unique molecular identifiers (UMIs) were incorporated during cDNA synthesis and demultiplexed at the UMI level using MiGEC (Shugay *et al.*, 2014). To determine the effect of UMI-based correction on repertoire clonal composition, data from PC02 with and without UMI-based correction was analysed. A strong overall correlation between CDR3 proportions between the two techniques was evident ($r=0.87$, $p<0.001$), particularly between highly expanded clones (Figure 3.4A). Analysis of the raw clone count data (Figure 3.4B) also indicated a strong positive correlation between the two techniques ($r=0.75$, $p<0.001$), but highlighted how not using UMI-based correction can over-estimate the number of clones present in a repertoire. However, it is mostly lower frequency clones that are over-estimated. At the TCRB V-gene level, correlation between UMI-corrected and not UMI-corrected data was even stronger ($r=0.94$, $p<0.001$). This is in agreement with Rosati *et al.* (2017) who compared 5'RACE vs 5'RACE-UMI corrected techniques, and found a strong overlap between expanded clones and good agreement between these two techniques compared to others. Hence, despite not having used UMIs in PC01, I am confident that using the 5-RACE technique alone without UMI-correction will not have affected conclusions drawn. The reduction of counts by ~10 fold once UMI-corrected is also in line with previously published studies using the stringent error correction employed by MiGEC (Izraelson *et al.*, 2018; Alves Sousa *et al.*, 2019). Where indicated, data is presented for both PC01 and PC02, with all data for PC02 UMI-corrected.

Figure 3.4

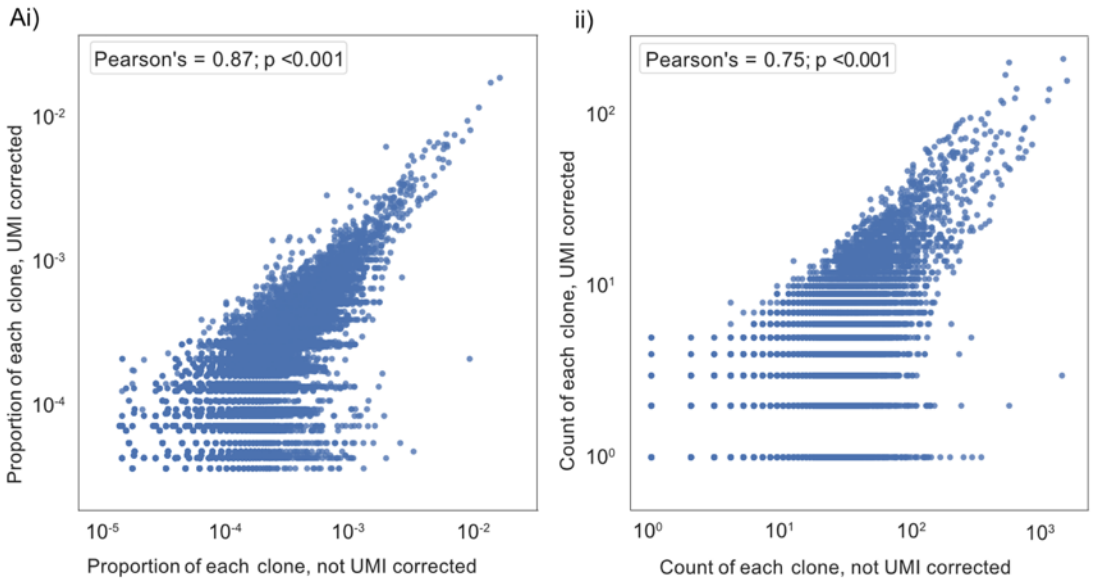


Figure 3.4: Scatterplots show correlation between Ai) the proportion of each CDR3 sequence from challenged repertoires from PC02, from libraries that were either UMI corrected using MiGEC software (y axis) or not UMI corrected (x axis), and ii) the raw clonal counts from the same data.

3.4.4. TCRβ T-effector repertoire sequencing results

3.4.4.1. TRBV gene usage

I first hypothesised that if the T_E expansion in the acute phase of infection was solely the result of non-specific activation, V gene usage and V/J allele usage within the T_E repertoire would mirror that of T_N repertoires, despite the vast cellular proliferation. Thus, there would be no change in the distribution of V or V/J allele usage post parasite-challenge. However, a distinct increase in TRBV3 gene usage was observed during the acute phase of infection, differentiating challenged T_E repertoires from both unchallenged T_N and T_E repertoires, and from challenged T_E repertoires at later time-points (Figure 3.5A). For MT infections during the acute phase, TRBV3 encoded on average 23.7% (± 2.03 95% CI) of the effector repertoire at day 6 and 21.6% (± 2.21 95% CI) at day 10 post-infection, compared to only 7.6% (± 0.47, 95% CI) of the unchallenged naïve repertoire (Figure 3.6). This increase in TRBV3 usage was more delayed in SBP challenged repertoires, not apparent until day 10 post-challenge, where at its peak it encoded 17.2% (±2.54, 95% CI) of the T_E repertoire. This pattern of increased TRBV3 gene usage during the acute phase was repeatable in the UMI-corrected second experiment PC02 by day 7, but was not increased compared to

unchallenged levels at the earlier day 4 time-point included in this follow-up experiment (Figure 3.7).

Figure 3.5

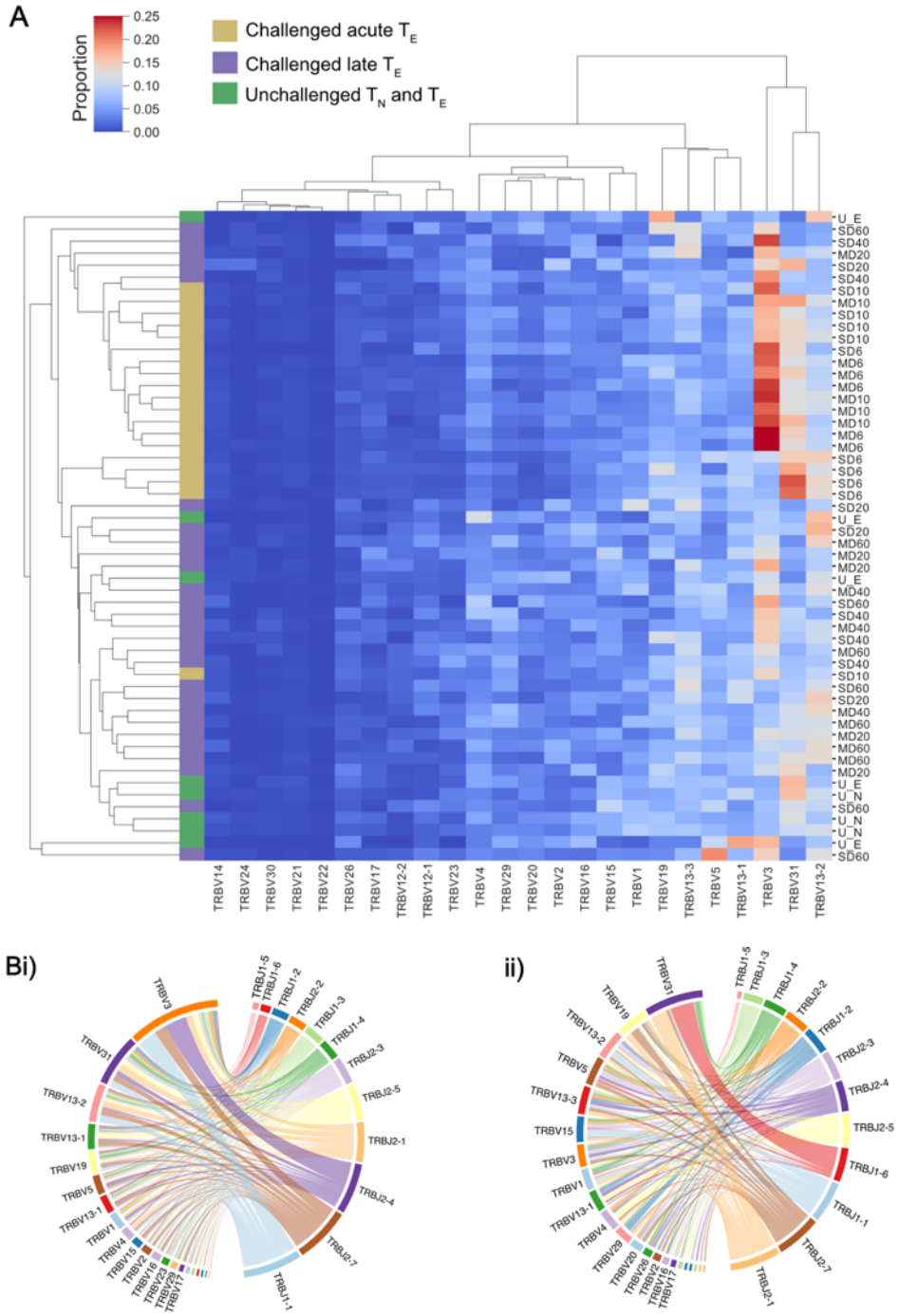


Figure 3.5: T_E repertoires have dominant TRBV3 gene usage during the T_E expansion in acute phase of a *P. chabaudi* infection. A) Clustermap displays TRBV gene proportional usage for individual splenic CD4$^+$ TCR repertoires from challenged acute T_E repertoires (gold), challenged late phase T_E repertoires (purple) and unchallenged T_N and T_E repertoires (green). Each column is a unique TRBV gene and each row an individual repertoire. B) Circos plots show V/J gene usage from a representative repertoire of i) a challenged MT T_E repertoire at day 6 post-infection and ii) an unchallenged T_N repertoire. Band width is proportional to usage frequency.

Figure 3.6

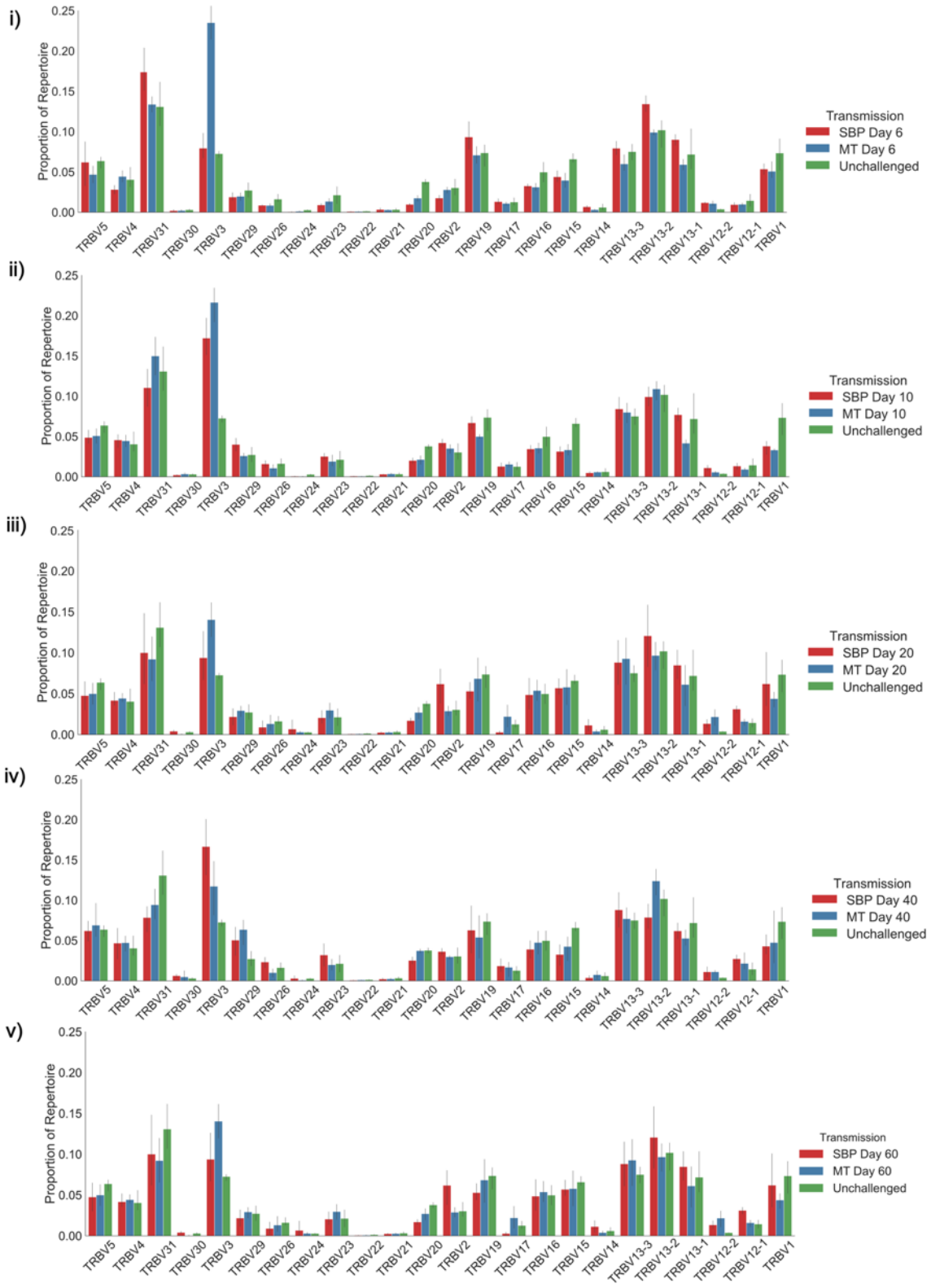


Figure 3.6: Proportion of TRBV-gene usage in unchallenged TN repertoires (green), and TE repertoires of mice infected with SBP parasites (red) or recently MT parasites (blue) at days 6 (i), 10 (ii), 20 (iii), 40 (iv) and 60 (v) days post-infection.

Figure 3.7

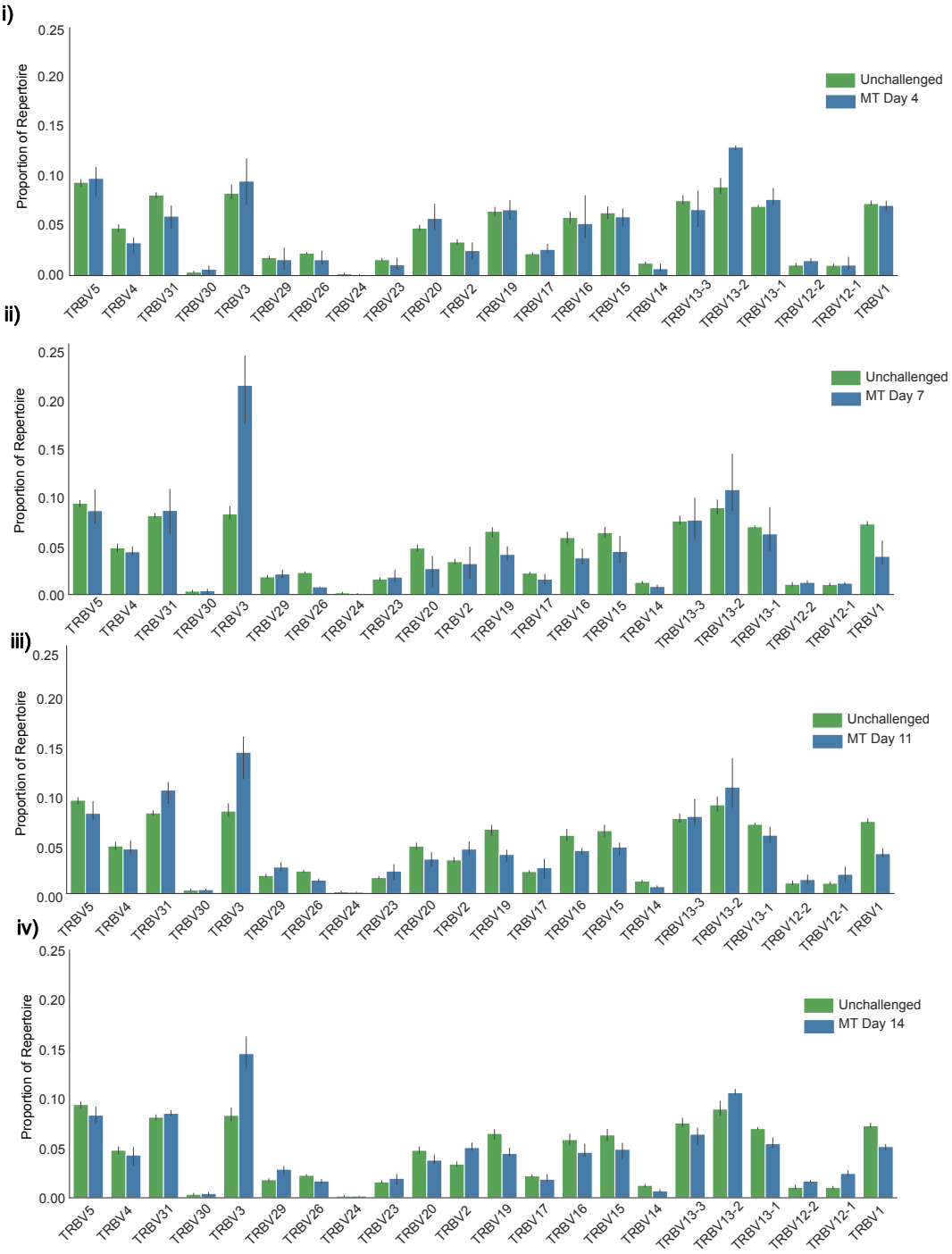


Figure 3.7: Proportion of TRBV-gene usage in unchallenged TN repertoires (green) and TE repertoires of mice infected with recently MT parasites (blue) in a second independent experiment (PC02), utilising UMIs, at days 4 (i), 7 (ii), 11 (iii) and 14 (iv) post-infection.

3.4.4.2. TRBV/J gene usage

For V/J allele combination usage, an overall strong positive correlation between challenged T_E and unchallenged T_N repertoires was evident during the effector expansion, indicating a highly polyclonal response with broad expansion of the naïve precursor pool. However, three specific TRBV3/J allelic combinations,TRBV3-TRBJ1-1, TRBV3-TRBJ2-4 and TRBV3-TRBJ2-7, were disproportionately increased in challenged T_E repertoires of mice infected with MT parasites at both days 6 and day 10, and for SBP infections by day 10 post-infection (Figure 3.8A). This specific V/J usage was conserved across all individual replicate mice infected with MT parasites during the acute phase of infection (Figure 3.8Bii) and was evident in the repeat experiment PC02 (Figure 3.8Biii). During the late phase of infection, as the T_E population contracted, this conserved V/J signature was lost in both infection types (Figure 3.8Bi, ii).

Figure 3.8

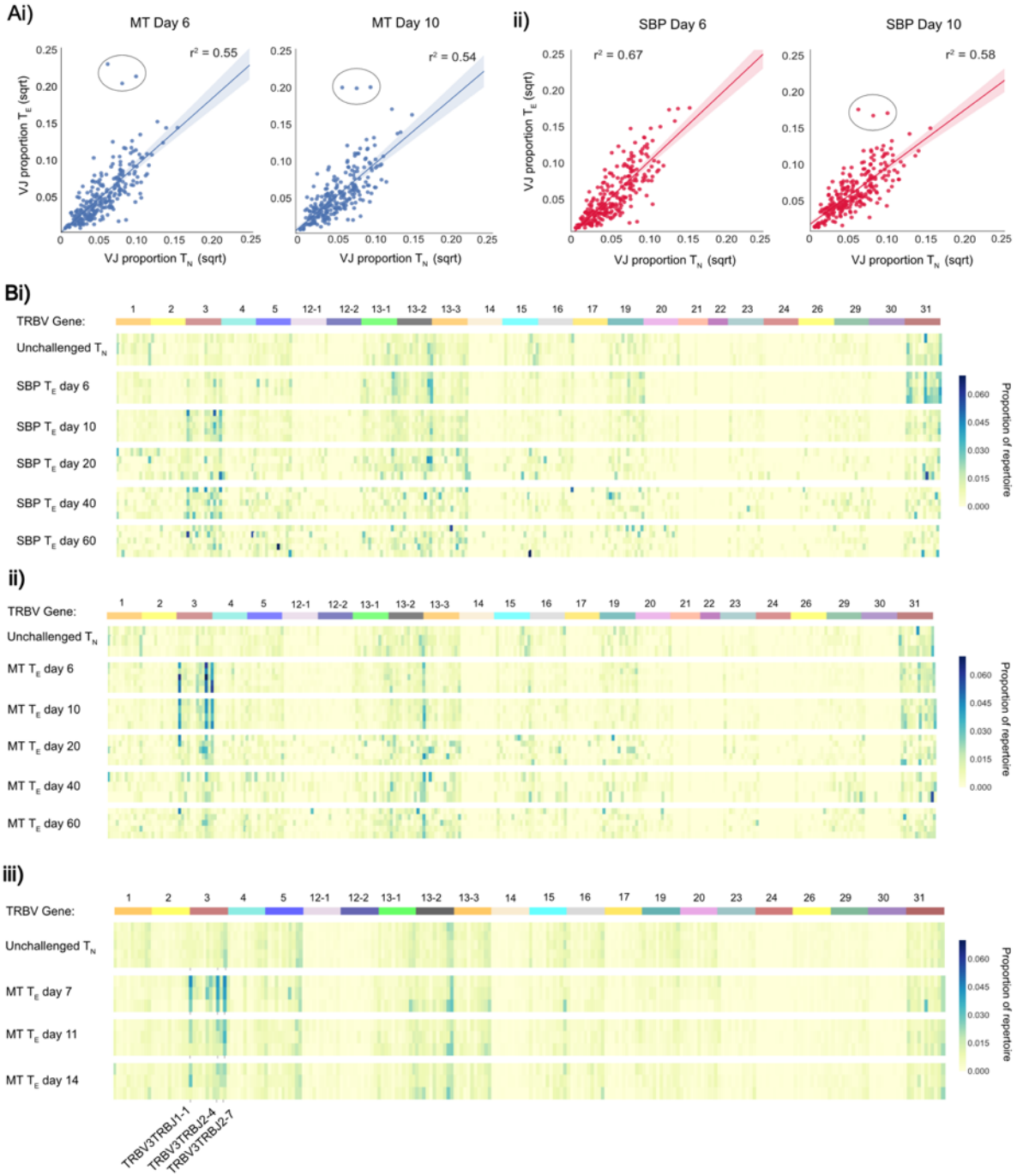


Figure 3.8: A) Mean proportion of each V/J allelic combination in unchallenged T_N repertoires versus challenged T_E repertoires at days 6 and 10 post-infection for mice infected with (i) MT parasites (blue) and (ii) SBP parasites (red). Each point represents a unique V/J combination. B) Heatmaps depict proportional usage of each V/J allelic combination (columns) for individual replicate mice (rows) for (i) unchallenged T_N repertoires and T_E SBP repertoires and (ii) unchallenged T_N repertoires and T_E MT repertoires and iii) unchallenged TN repertoires and TE MT repertoires from a second independent experiment (PC02). Horizontal colour bar indicates TRBV gene used in the V/J combination.

3.4.4.3. Clonality of response

Changes in diversity of a TCR repertoire following pathogen exposure are indicative of the extent to which clonal expansion within a repertoire has occurred (Venturi *et al.*, 2007). To assess the diversity of the acute CD4+ T_E expansion response at the clonal level, repertoire diversity was calculated using Simpson's diversity index on age-matched (Britanova *et al.*, 2014) and size-matched repertoires from the repeat experiment. This allowed us to sample an equal number of UMI-labelled cDNA molecules for the precise normalisation required for comparing diversity metrics (Izraelson *et al.*, 2018). Although unchallenged T_N repertoires were, as expected, significantly more diverse than challenged T_E repertoires (Figure 3.9A, $p<0.01$), the T_E repertoires after challenge were still highly polyclonal – with the most abundant clone taking up on average only 0.72% (± 0.11%) of the repertoire compared to 0.112% (± 0.01%) in T_N unchallenged repertoires (Figure 3.9B).

Figure 3.9

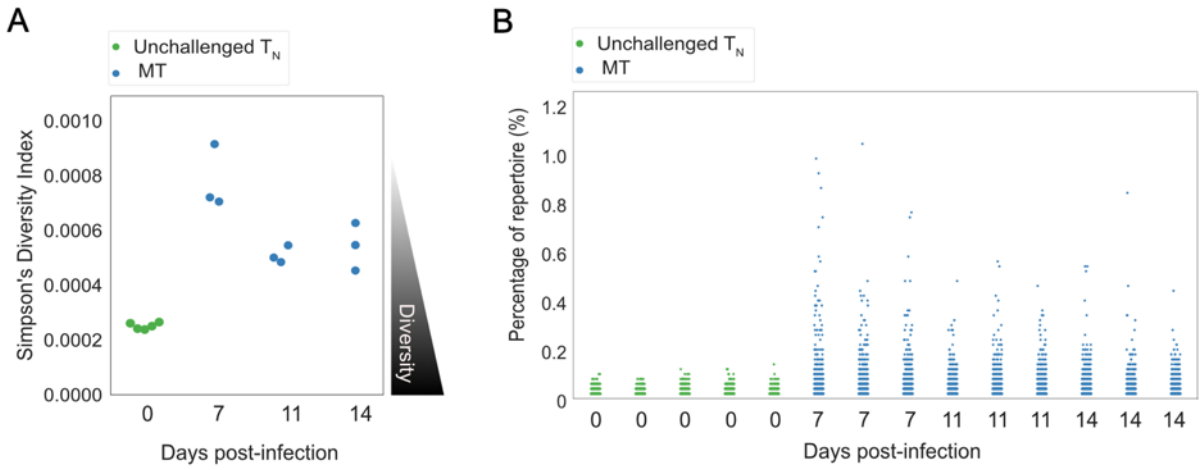


Figure 3.9: Splenic T_E repertoires elicited by *Plasmodium chabaudi* are highly diverse. A) Simpson's Diversity Index (SI) for UMI-size matched unchallenged T_N (green) and acute MT T_E repertoires (blue). Data shown is from PC02 only. SI varies from 0 to 1, and for TCR repertoires represents the probability that identical TCRs (as determined by identical CDR3 amino acid sequence) will be drawn from the repertoire with two independent draws. SI of 0 therefore represents maximal diversity. B) Each point in the strip plot represents a clone and the percentage of repertoire they occupy, for individual replicate mice for T_N unchallenged repertoires (green) and MT T_E repertoires (blue). Data for A) and B) was normalised by down-sampling to 5000 UMI.

3.4.4.4. Similarity of responses

If a pathogen-specific response is elicited at the clonal level, one would expect challenged T_E repertoires of replicate mice to contain an increased number of shared clones, and to therefore be more similar to each other than unchallenged repertoires. To examine the degree of repertoire sharing between replicate mice, the Jaccard index, a similarity or 'overlap' metric was used, matching at the CDR3 amino acid sequence level. Over the course of infection, for each infection type, similarity between replicate groups was significantly altered (one-way ANOVA, MT: $p<0.001$, SBP: $p<0.001$) (Figure 3.10Ai), with replicates being more similar to each other in the acute phase of infection than at later time-points. MT repertoires were also more similar to each other during the acute phase than SBP infections (day 6: $t=2.6$, $p=0.016$, day 10: $t=7.2$, $p<0.001$). For both infection types, during the acute phase, replicate repertoires were more similar to each other than to unchallenged T_N (day 6: MT: $t=15.13$, $p<0.001$, SBP: $t=11.7$, $p<0.001$, day 10: MT: $t=13.4$, $p<0.001$, SBP: $t=9.2$, $p<0.001$) and unchallenged T_E repertoires (day 6: MT: $t=15.1$, $p<0.001$, SBP: $t=11.7$, $p<0.001$, day 10: MT: $t=13.4$, $p<0.001$, SBP: $t=9.2$, $p<0.001$). Randomly sampling the same number of sequences (10^4) to produce size-matched repertoires did not alter this pattern of results, nor did size-matching the repeat UMI data (Figure 3.10Aiii). Further exploration using multi-dimensional scaling (MDS) of size-matched repertoires, also indicated clustering of acute T_E repertoires for both infection types according to Jaccard similarity, with MT repertoires at day 6 and 10 most tightly co-clustered (Figure 3.10Aii).

Figure 3.10

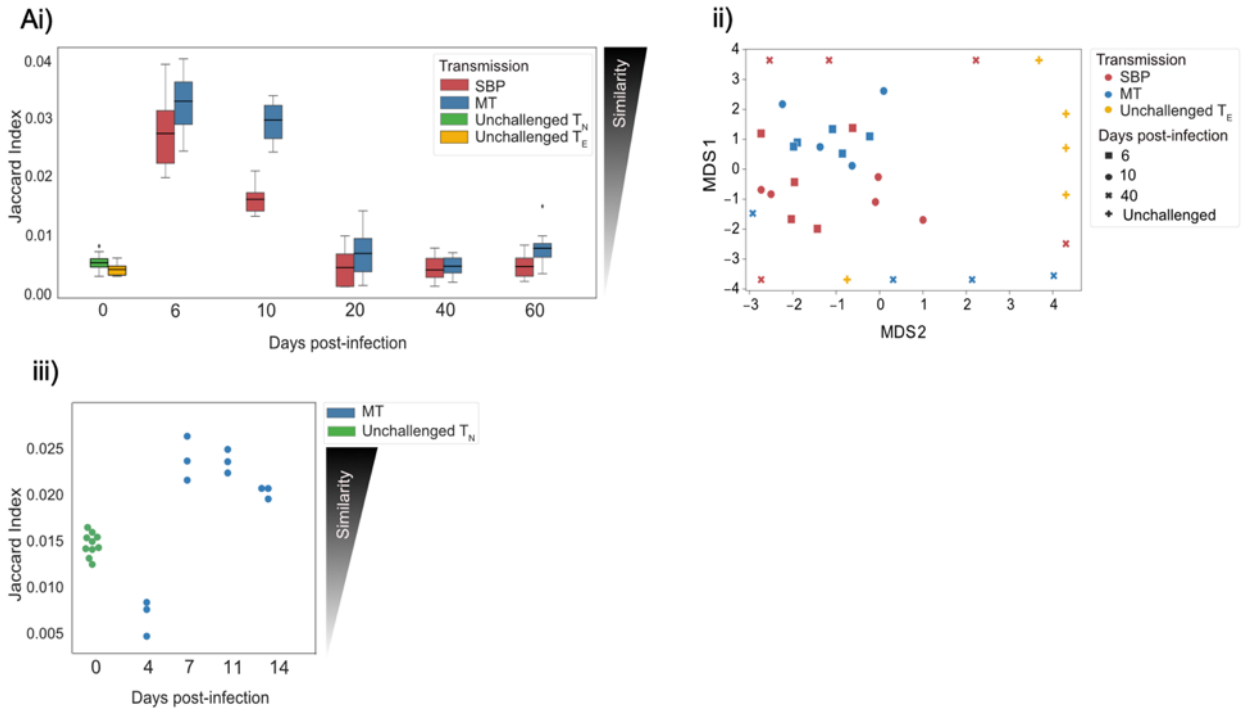


Figure 3.10: Ai) Similarity between replicate repertoires is demonstrated using the Jaccard index, a normalized sharing metric which represents the degree of clonal overlap between repertoires. Jaccard index varies form 0 (no similarity) to 1 (identical repertoires). T_E replicate repertoires are more similar to each other in the acute phase of infection for both infection types. ii) Multi-dimensional-scaling (MDS) analysis using Jaccard similarity index of T_E repertoires for unchallenged (yellow), MT (blue) and SBP (red) repertoires. Data for (i) and (ii) was calculated on weighted data to include clonotype frequency. iii) Jaccard index of unchallenged TN repertoires (green) and challenged T_E repertoires from PC02. For (iii), data was normalised by down-sampling to 5000 UMI.

3.4.4.5. CDR3 network analyses

The TCR repertoire is a dynamic network, so examining similarity solely at the clonal level can fail to take in to account the degree of extended clonal networks that may be present. Despite not undergoing somatic hypermutation, T-cell repertoires have been shown to contain networks generated by sequence similarity, with CDR3 sequence similarity and thus network connectivity increased in antigen-experienced repertoires (Glanville *et al.*, 2017; Madi *et al.*, 2017) . To explore connectivity between CDR3s in the T_E repertoires, network analysis was undertaken between the top 100 most abundant clones, using Levenshtein distance. Networks were constructed between replicate repertoires, by creating an edge between unique CDR3 sequences (nodes) if they were within a Levenshtein distance of one of each other (Figure 3.11A). Node degree (the average number of edges per node within network) for each individual

indicated a higher degree of connectivity for both infection types in the acute stage of infection compared to unchallenged T_E repertoires (Mann-Whitney U, (day 6: MT: $p<0.01$, SBP: $p=0.022$, day10: MT: $p<0.01$, SBP: $p=0.045$) (Figure 3.11Avi). MT repertoires also tended to have a higher node degree than SBP repertoires at days 6, 10 and 20 post-infection, although significance was not detected at day 10 (Mann-Whitney U, day 6: $p<0.01$, day 10: $p=0.088$, day 20: $p=0.039$). Within each individual, there was also an increased frequency of CDR3 pairwise comparisons that differed by a distance of one or two for MT repertoires in the acute phase of infection compared to both SBP and unchallenged T_E repertoires, highlighting their greater repertoire similarity (Figure 3.11B).

Figure 3.11

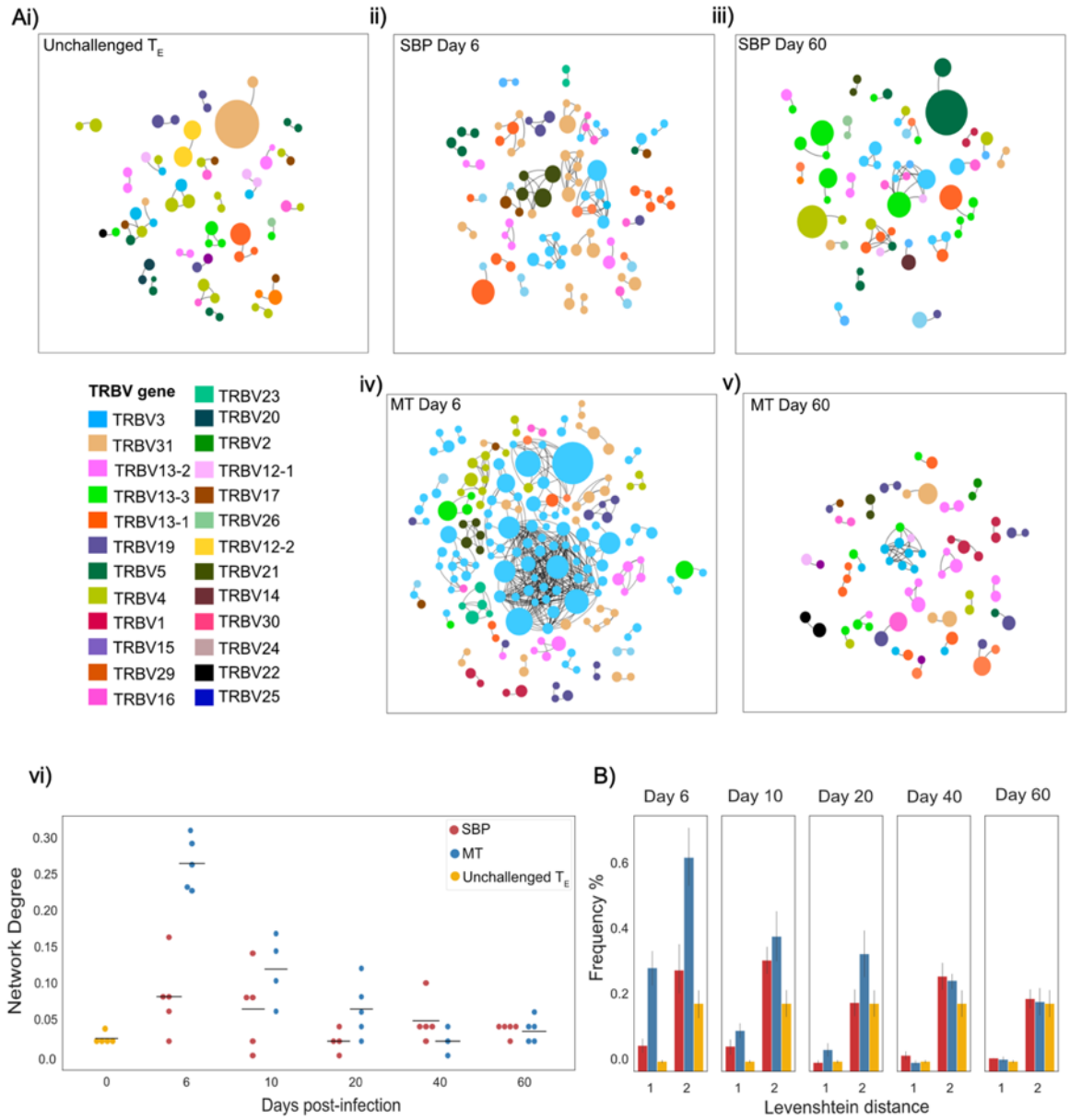


Figure 3.11: Ai-v) Networks showing the top 100 most abundant CDR3 amino acid sequences in replicate repertoires within a Levenshtein distance of 1 of each other. For TCRs a Levenshtein distance of 1 represents 1 amino acid mismatch (insertion, deletion or substitution) between CDR3 sequences. Each node represents a TCR clone as defined by CDR3 amino acid sequence, with node size indicating proportion of repertoire occupied by clone. Nodes are coloured according to TRBV-gene usage. An edge is drawn between nodes if within a Levenshtein distance of 1, with unconnected nodes not depicted. vi) network degree (mean number of edges per node) for each individual T_E repertoire network. B) Frequency (%) of individual CDR3 sequence pairwise-comparisons, within the top 100 most abundant CDR3 sequences of each individual T_E repertoire, that are within a Levenshtein distance of 1 and 2.

3.4.4.6. Clustering analyses using a modified swarm algorithm

Given the increased connectivity in challenged repertoires, and the knowledge that TCRs recognizing the same antigen typically have a high global similarity to each other (Dash *et al.*, 2017a; Glanville *et al.*, 2017), I clustered CDR3 sequences of individual repertoires within one amino acid mismatch of each other using a modified Swarm algorithm (Mahe *et al.*, 2014). This identified two clusters of highly similar CDR3 sequences, hereafter referred to as OTU1 and OTU2, that dominated T_E repertoires and were conserved across replicates in the acute phase of infection (Figure 3.12). These were either not present or present at a low level in unchallenged repertoires in both PC01 and PC02 (Table 3.2). For PC02, near-identical cluster to both OTU1 and OTU2 were also found to be expanded and OTU1 was dominant in the majority of challenged repertoires at day 7 and 11 post-infection, but not at the earlier day 4 time-point included in PC02 (Figure 3.13).

Figure 3.12

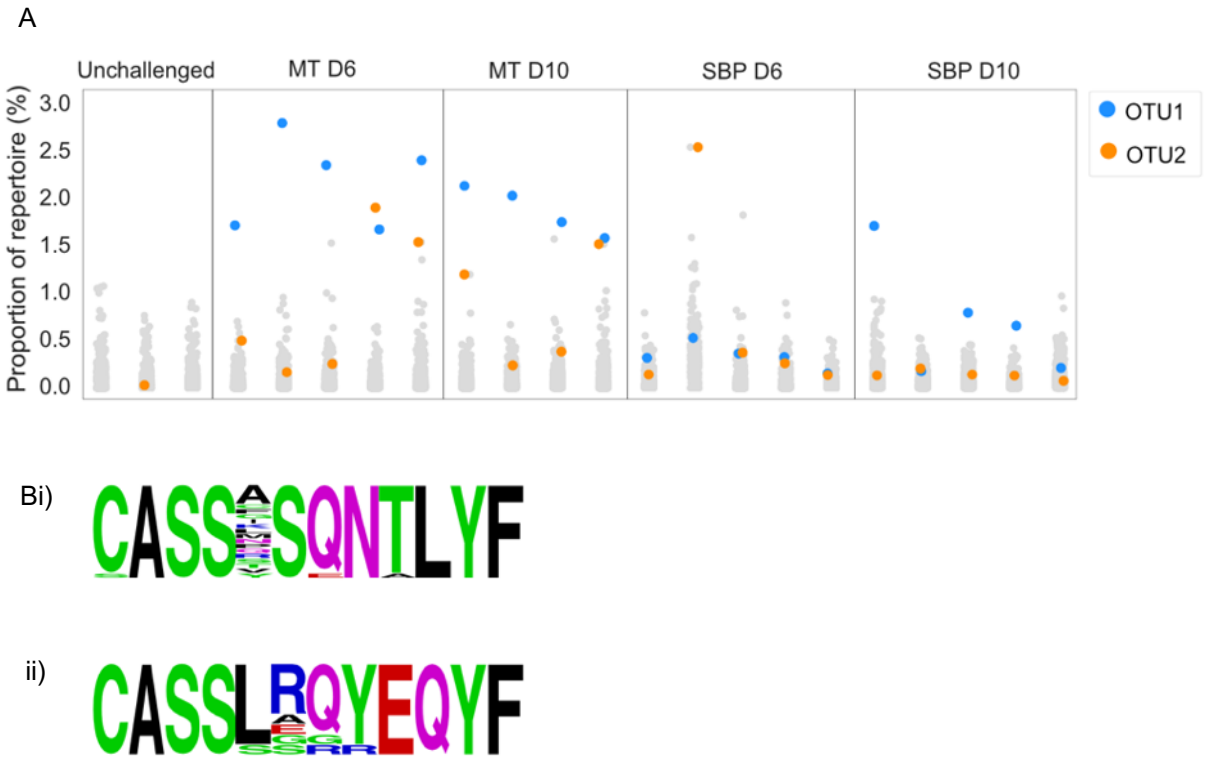


Figure 3.12: Acute MT T_E repertoires are dominated by the same cluster of clones. Repertoires were clustered using a modified Swarm algorithm, to cluster CDR3 sequences within 1 amino-acid mismatch of each other, with identical V-gene usage. A) Strip plots display the proportion of repertoire occupied by each cluster in individual repertoires for unchallenged T_N and MT and SBP T_E repertoires at day 6 and 10 post-infection. Each column is an individual repertoire, with each point representing a unique cluster. Cluster OTU1 is displayed in blue and OTU2 in orange. B) Representative amino acid sequence logos of clusters OTU1 (i) and OTU2 (ii).

Table 3.2

Group	OTU1 % (± SEM)	No. of mice	OTU2 % (± SEM)	No. of mice	Experiment
MT6	2.44 (0.24)	5/5	0.98 (0.4)	5/5	PC01
MT10	2.09 (0.14)	4/4	0.94 (0.35)	4/4	PC01
MT20	0.84 (0.27)	5/5	0.28 (007)	3/3	PC01
MT40	0.47 (0.16)	3/3	0	0/3	PC01
MT60	0.44 (0.13)	5/5	0.16 (0.12)	2/5	PC01
SBP6	0.38 (0.07)	5/5	0.77 (0.52)	5/5	PC01
SBP10	0.8 (0.31)	5/5	0.16 (0.02)	5/5	PC01
SBP20	0.19 (0.17)	3/5	0.26 (0.1)	3/5	PC01
SBP40	0.55 (0.2)	4/5	0.9 (0.41)	2/5	PC01
SBP60	0.66 (0.54)	4/5	0.1 (0.05)	3/5	PC01
Unchallenged	0	0/3	0.44 (0)	1/3	PC01
MD4	0.05 (0.01)	2/3	0.03 (0.01)	2/3	PC02
MD7	1.35 (0.45)	3/3	0.37 (0.09)	3/3	PC02
MD11	0.54 (0.06)	3/3	0.26 (0.13)	3/3	PC02
MD14	0.44 (0.11)	3/3	0.16 (0.09)	3/3	PC02
Unchallenged	0.03 (0.01)	4/6	0.04 (0.0)	4/6	PC02

Table 3.2: mean repertoire proportion (% ± SEM) of OTU1 and OTU2 from experiments PC01 and PC02, present in challenged T_E repertoires and unchallenged T_N repertoires.

Figure 3.13

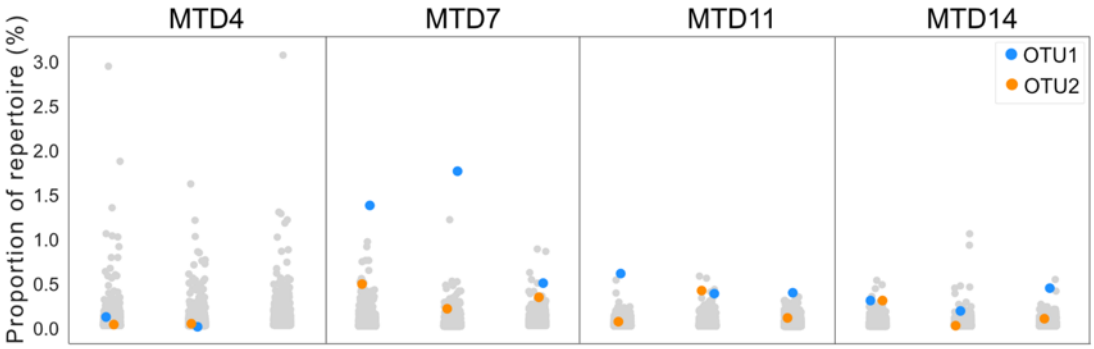


Figure 3.13: Strip plots display the proportion of repertoire occupied by each cluster in individual repertoires for challenged repertoires from PC02, at days 4, 7, 11 and 14 days post-infection. Each column is an individual repertoire, with each point representing a unique cluster. Cluster OTU1 is displayed in blue and OTU2 in orange.

3.4.4.7. Clustering analyses using Gliph2

The recently published Gliph2 algorithm (Huang *et al.*, 2020) was also applied to the PC01 data. Gliph2 is designed to identify TCRs recognizing the same antigen by clustering sequences with enriched amino acid motifs in the high-contact-probability region of CDR3β (IMGT positions 107-116). It identified significant clusters that corresponded to both OTU1 and OTU2 in the acute phase of MT infections, as well several other clusters which were not as dominant nor as well-conserved between replicates in their response (Figure 3.14, Table 3.3).

Figure 3.14

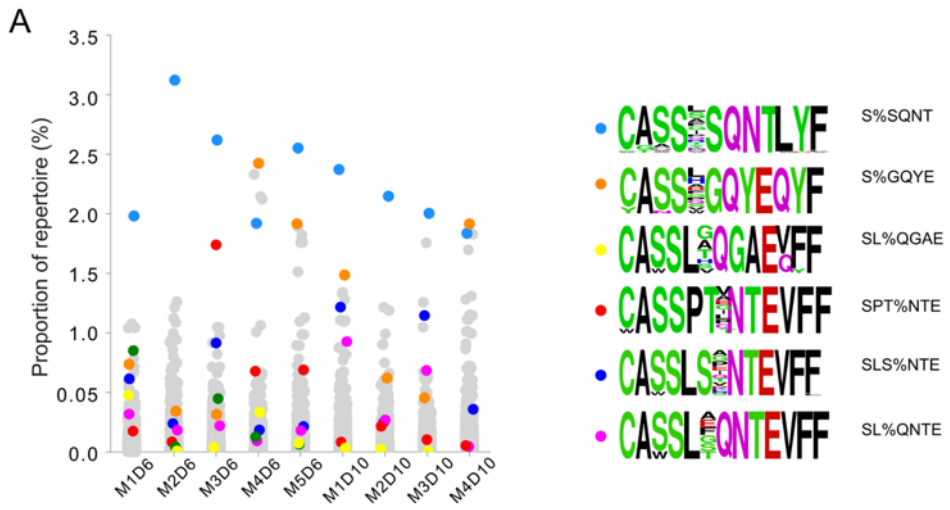


Figure 3.14: Strip plots indicate proportion of individual repertoires taken up by clusters identified as significantly enriched in acute MT TE repertoires using GLIPH2, with representative sequence logos for each cluster.

Table 3.3

Infection type	Time point	No. of clusters meeting criteria	Cluster_ID	Pattern	Final_score	
MT	D6	5	14	S%SQNT	8.50E-14	OTU1
			18	SPT%NTE	3.10E-13	
			10	SLS%NTE	4.60E-13	
			4	SLTQ	4.60E-12	
			7	SL%QGAE	8.50E-12	
MT	D10	4	21	S%SQNT	7.70E-14	OTU1
			32	S%GQYE	4.40E-13	OTU2
			5	SLS%NTE	1.70E-12	
			23	SL%QNTE	6.10E-12	
MT	D20	0				
MT	D40	0				
MT	D60	0				
SBP	D6	0				
SBP	D10	1	16	S%GTTSAET	1.10E-09	
SBP	D20	0				
SBP	D40	0				
SBP	D60	0				

Table 3.3: Summary results of significant clusters from GLIPH2 analysis. Filter criteria used: CDR3 sequence present in at least 75% of replicate group; vb_score <=0.05; expansion_score <=0.05; Fisher_score <=0.001. Pattern indicates the motif, with '%' indicating an amino acid mismatch.

3.4.4.8. Publicly available RNA-seq data

To determine if TRBV3 and the conserved clonal signatures detected are consistently expanded in the acute phase of a *P. chabaudi* infection, TCR sequences were extracted from publicly available whole transcriptome RNA-seq data, generated form the splenocytes of mice infected with *P. chabaudi (AS)* (data set E-ERAD-221) and *P. chabaudi (CB)* (data set E-ERAD-289). Clonotype frequencies extracted by MiXCR from RNA-seq data have previously been shown to correlate well with TCR amplicon data for enriched clones (Bolotin *et al.*, 2017; Bai *et al.*, 2018). However, repertoires are often only superficially profiled, and less abundant TCRs cannot be reliably quantified due to presumed under-sampling of TCR transcripts, as only a small percentage of genome-wide reads may cover the TCR loci (Bai *et al.*, 2018). Data set E-ERAD-221 consisted of 24 unchallenged samples and 120 MT challenged samples over days 3 to 12 post-infection. The mean number of unique functional clones extracted from each unchallenged sample was 203.9 (±19.4, 95% CI), and 152.9 (±6.9, 95% CI) from each challenged sample. Data set E-ERAD-289 consisted of 10 unchallenged mice and 22 challenged mice, 11 of which were infected with MT parasites and 11 with SBP parasites. The mean number of unique clones extracted from each unchallenged sample was 239.3 (±81.5, 95% CI) and 273.45 (±33.96, 95% CI) from each challenged sample. Analysis of E-ERAD-221 confirmed a temporal expansion of TRBV3 during the acute phase of infection (Figure 3.15A & B), and a search for clones present in OTU1 and OTU2 confirmed that sequences from these clusters were present and showed a similar temporal pattern of expansion as seen in the bulk TCR amplicon data (Figure 3.15Ci & ii). For E-ERAD-289, no exact time-point data was available, but comparing challenged versus unchallenged repertories again indicated a conserved increase in TRBV3 gene usage in challenged repertoires (Figure 3.16A & B), with clones from OTU1 and OTU2 again found to be present and at a higher proportion in challenged compared to unchallenged repertoires (Figure 3.16Ci & ii).

Figure 3.15

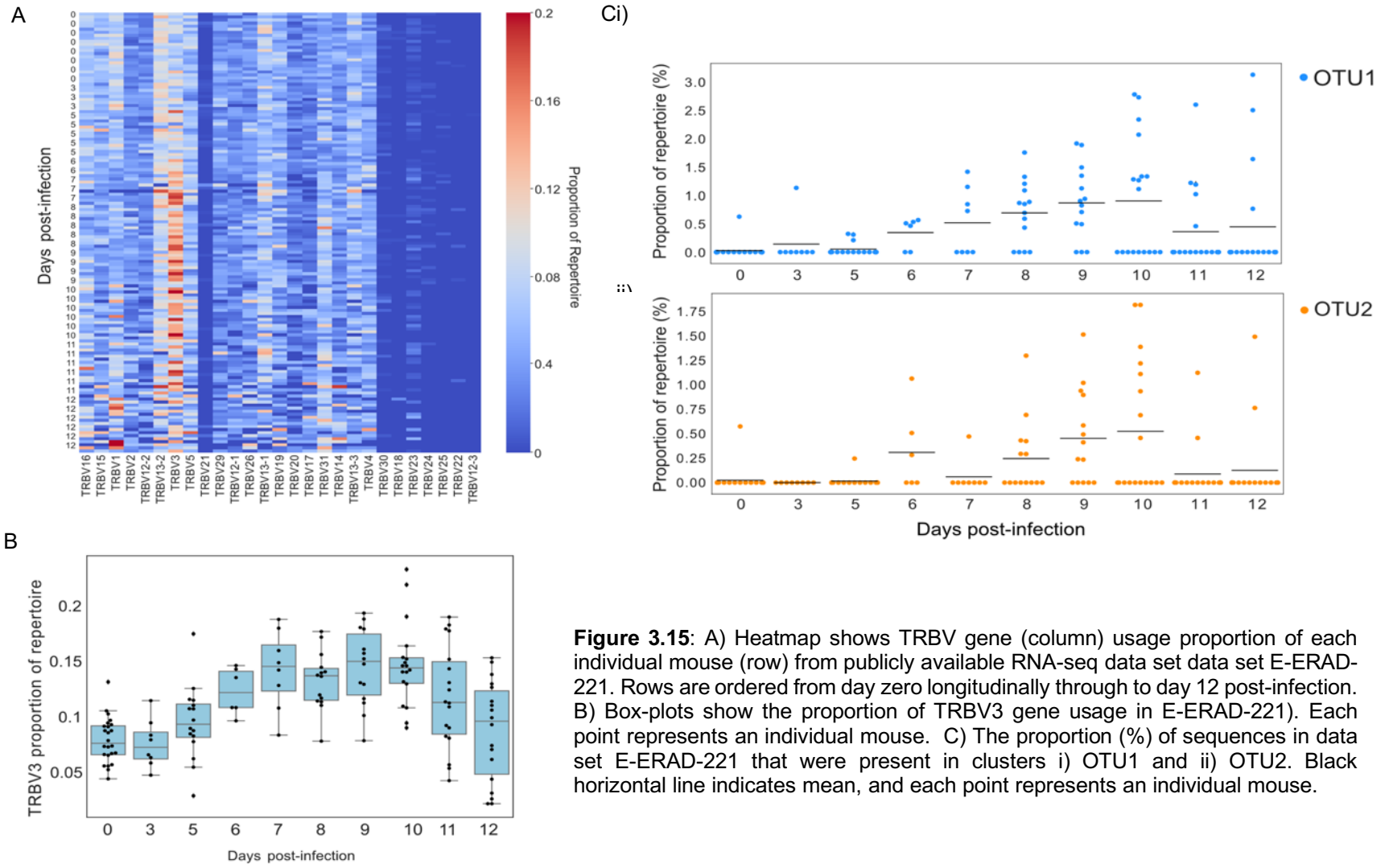


Figure 3.15: A) Heatmap shows TRBV gene (column) usage proportion of each individual mouse (row) from publicly available RNA-seq data set data set E-ERAD-221. Rows are ordered from day zero longitudinally through to day 12 post-infection. B) Box-plots show the proportion of TRBV3 gene usage in E-ERAD-221). Each point represents an individual mouse. C) The proportion (%) of sequences in data set E-ERAD-221 that were present in clusters i) OTU1 and ii) OTU2. Black horizontal line indicates mean, and each point represents an individual mouse.

Figure 3.16

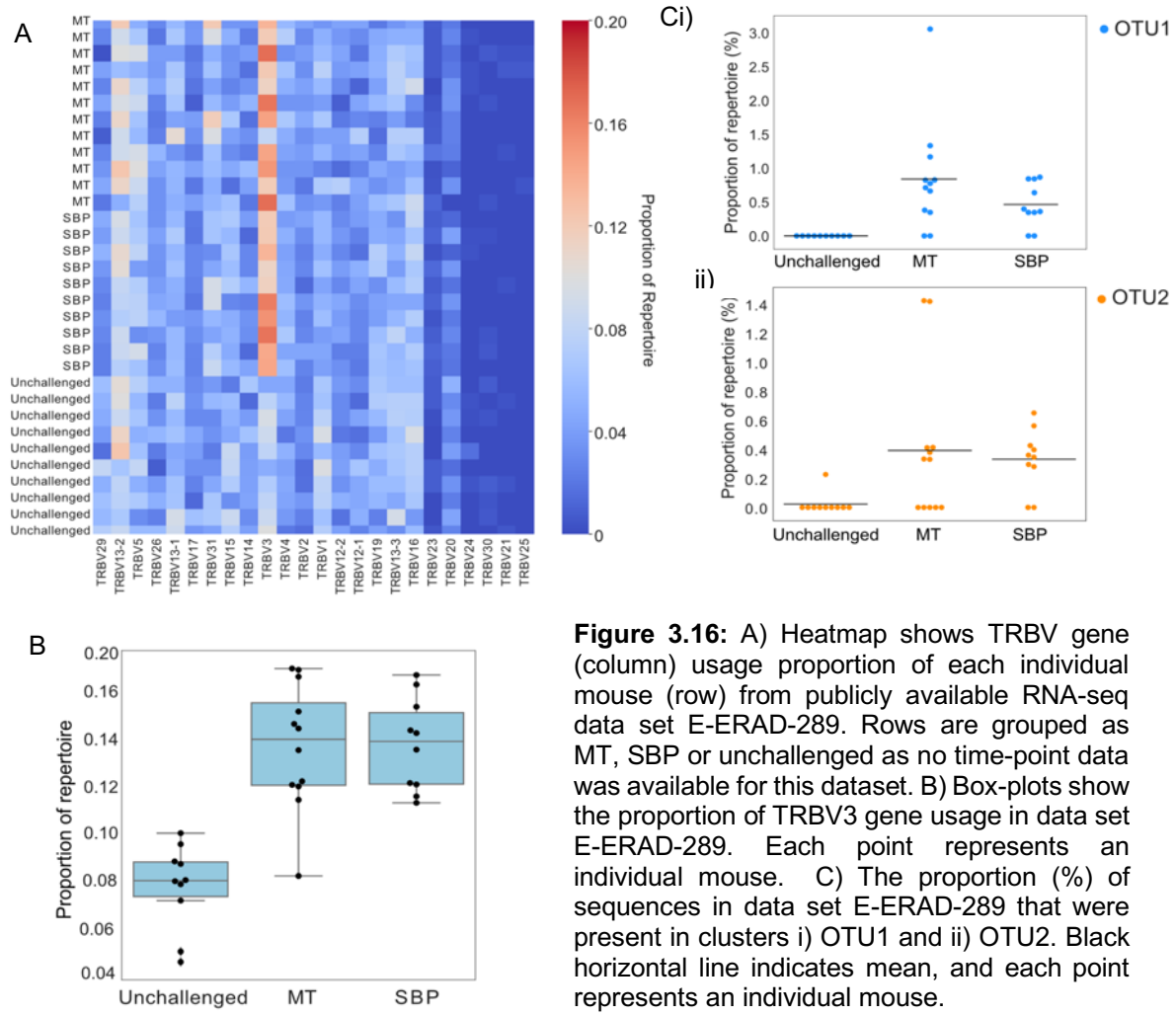


Figure 3.16: A) Heatmap shows TRBV gene (column) usage proportion of each individual mouse (row) from publicly available RNA-seq data set E-ERAD-289. Rows are grouped as MT, SBP or unchallenged as no time-point data was available for this dataset. B) Box-plots show the proportion of TRBV3 gene usage in data set E-ERAD-289. Each point represents an individual mouse. C) The proportion (%) of sequences in data set E-ERAD-289 that were present in clusters i) OTU1 and ii) OTU2. Black horizontal line indicates mean, and each point represents an individual mouse.

3.4.4.9. Properties of OTU1

Given the conserved nature of OTU1 between individual mice, I hypothesised that the CDR3 sequences it contains would share similar properties with other known 'public' CDR3 sequences, defined simply as TCR clones shared by different individuals. Public TCR clones have been detected in numerous T-cell responses in multiple species, and although their functional significance remains unknown, they have been shown to be expanded in response to antigenic stimulation (Covacu *et al.*, 2016), viral infection (Miles *et al.*, 2010; Benati *et al.*, 2016) and associated with regulatory self-immunity (Madi *et al.*, 2014). In some previous studies, public sequences have been shown to have minimal alterations to germline V, D and J gene sequences. In agreement with this, I found fewer recombination events in the CDR3 sequences in OTU1, with the mean number of randomly inserted nucleotides in the CDR3 sequences in these clusters significantly lower than that for CDR3 sequences in both challenged ($t=-61.5$, $p<0.001$) and unchallenged repertoires ($t=-58.7$, $p<0.001$) (Figure 3.17Ai). The mean number of nucleotides lying between the V and J gene segment sequences was also significantly lower (unchallenged: $t=-22.1$, $p=0.002$, challenged: $t=-83.7$, $p<0.001$) (Figure 3.17Aii). This cluster also showed a greater degree of convergent recombination – considered an important mechanism of public TCR generation (Venturi *et al.*, 2006; Covacu *et al.*, 2016) - with a higher average number of unique CDR3 nucleotide sequences that code for the same CDR3 amino acid sequence (Figure 3.17Aiii) compared to CDR3 sequences in unchallenged ($t=4.4$, $p<0.001$) and challenged ($t=5.3$, $p<0.001$) repertoires. Consequently, CDR3 amino acid sequences in OTU1 had a higher probability of generation (Pgen) (Sethna *et al.*, 2018) than CDR3 sequences in unchallenged T_N repertoires and challenged repertoires (Figure 3.17Bi). In a scenario of non-specific polyclonal expansion, a higher Pgen could indicate sharing and detection of this cluster incidentally due to higher abundance in the pre-selection pool, rather than as a result of common specificity (convergent selection) or function (Pogorelyy, Minervina, Chudakov, *et al.*, 2018). However, the cluster was either not found or was present at a low level in unchallenged T_N and T_E repertoires, and our use of UMI-corrected data for the follow-up experiment PC02 confirmed that the CDR3s within this cluster are clonally expanded in acute T_E repertoires. Further, a large proportion of CDR3s with a similar

high Pgen in unchallenged T_N populations, were either not present in T_E repertoires, or present but not expanded, demonstrating that the public cluster is not simply dominant due to a higher Pgen, and thus is likely to be pathogen-associated (Figure 3.17Bii).

Figure 3.17

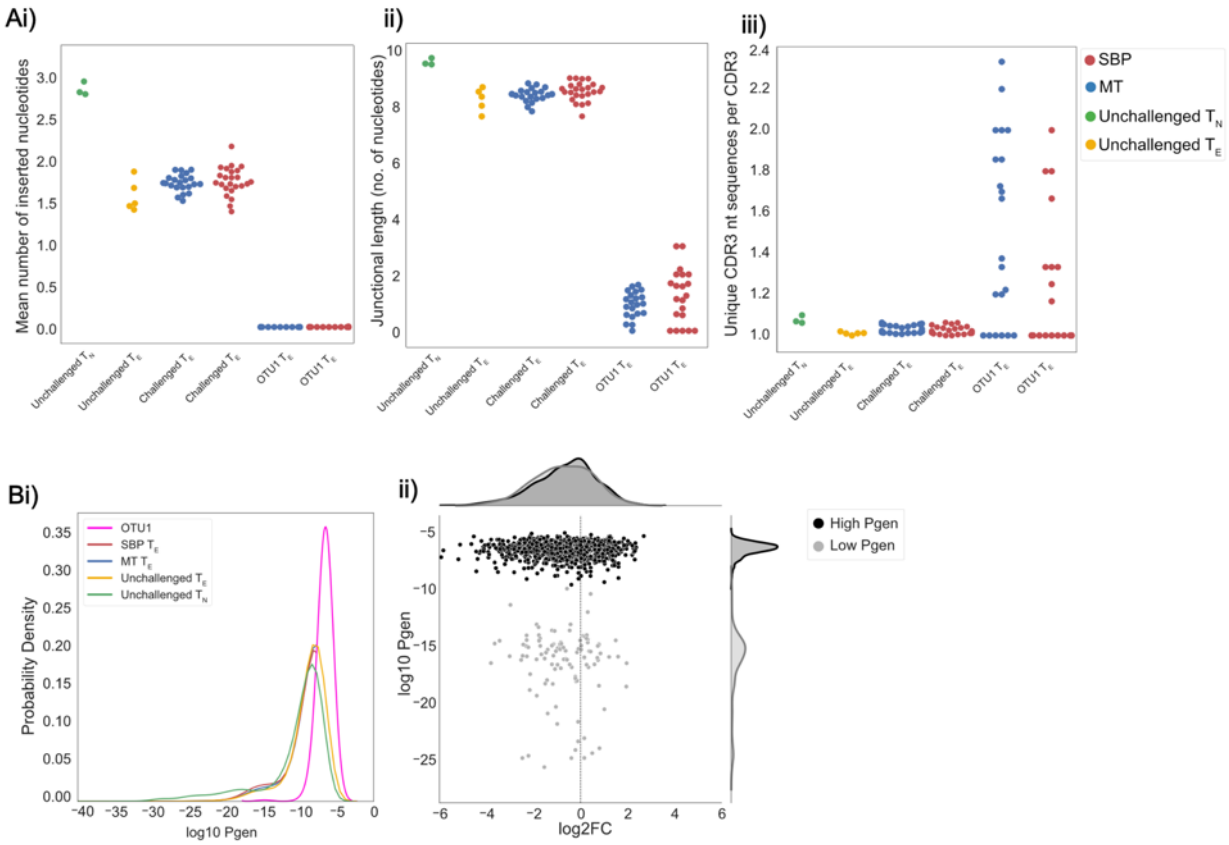


Figure 3.17: OTU1 CDR3 sequences have hallmarks of a public response: Ai) mean number of inserted random nucleotides in CDR3 sequences, ii) mean number of nucleotides lying between V and J gene segment sequences and iii) convergence (mean number of unique nucleotide sequences that encode a particular CDR3 sequence), for unchallenged T_N (green), unchallenged T_E repertoires (yellow), all MT T_E repertoires (blue) and all SBP T_E repertoires (red) and for cluster OTU1 in all MT and SBP T_E repertoires. Bi) Distributions of the generation probabilities (log10) of CDR3 nucleotide sequences for unchallenged T_N (green) and T_E repertoires (yellow), SBP (red) and MT (blue) T_E repertoires and CDR3 nucleotide sequences of cluster OTU1 (pink). ii) Log2 fold change of clones present in unchallenged T_N repertoires versus challenged T_E repertoires. Each point represents an individual clone, and Pgen is separated in to high (>-10) (black) and low (<-10) (grey). Log2FC was calculated on UMI normalised data from PC02.

3.4.5. TCRβ memory repertoires

Splenic CD4$^+$ T_{EM} and T_{CM} populations have been shown to moderately expand following a *P. chabaudi* infection (Stephens and Langhorne, 2010; Spence *et al.*, 2013; Opata and Stephens, 2017). To determine if a conserved splenic memory response was generated following a *P. chabaudi (AS)* infection, the T_{CM} and T_{EM} repertoires from experiment PC01 were also sequenced.

3.4.5.1. T_{CM} repertoires

For the T_{CM} repertoires, unchallenged T_{CM} cell numbers were too low to generate sequencing libraries, and thus repertoires were only available for the challenged repertoires. Thus, a comparison to unchallenged T_{CM} repertoires could not be undertaken. Replicate repertoires had similar diverse TCRBV gene usage, which did not alter for either transmission type over the time-course of infection (Figure 3.18A). At the clonal level, similarity between replicate repertoires also did not increase over the time-course of infection for either transmission type (Figure 3.19A). For MT infections a difference in similarity over the time-course was detected (one-way ANOVA, $p<0.002$), but this was due to day 6 replicates being more similar to each other than other time-points were (post-hoc Tukey HSD). Clustering amino acid CDR3 sequences using the modified swarm algorithm did not alter this pattern of similarity for either transmission type, indicating that a shared response between replicates was not detectable in the T_{CM} repertoires over the time-course of infection.

3.4.5.2. T_{EM} repertoires

For T_{EM} repertoires, at day 6 for both transmission types, TRBV19 was increased compared to unchallenged (encoding 18.7% (±5.2) for MT repertoires and 18.7% (± 3.1) for SBP repertoires, compared to 3.5% (± 2.5) for unchallenged (Figure 3.18B). However, TRBV19 was not increased at any other time-point, and was otherwise not significantly altered over the time-course of infection. At the clonal level, similarity between replicate repertoires also did not increase over the time-course of infection for either transmission type (Figure 3.19B), with a similar pattern seen when clustering amino acid CDR3 sequences using the modified swarm algorithm. This indicates that for T_{EM} repertoires, a conserved response between replicates was also not detectable over the time-course of infection.

Figure 3.18

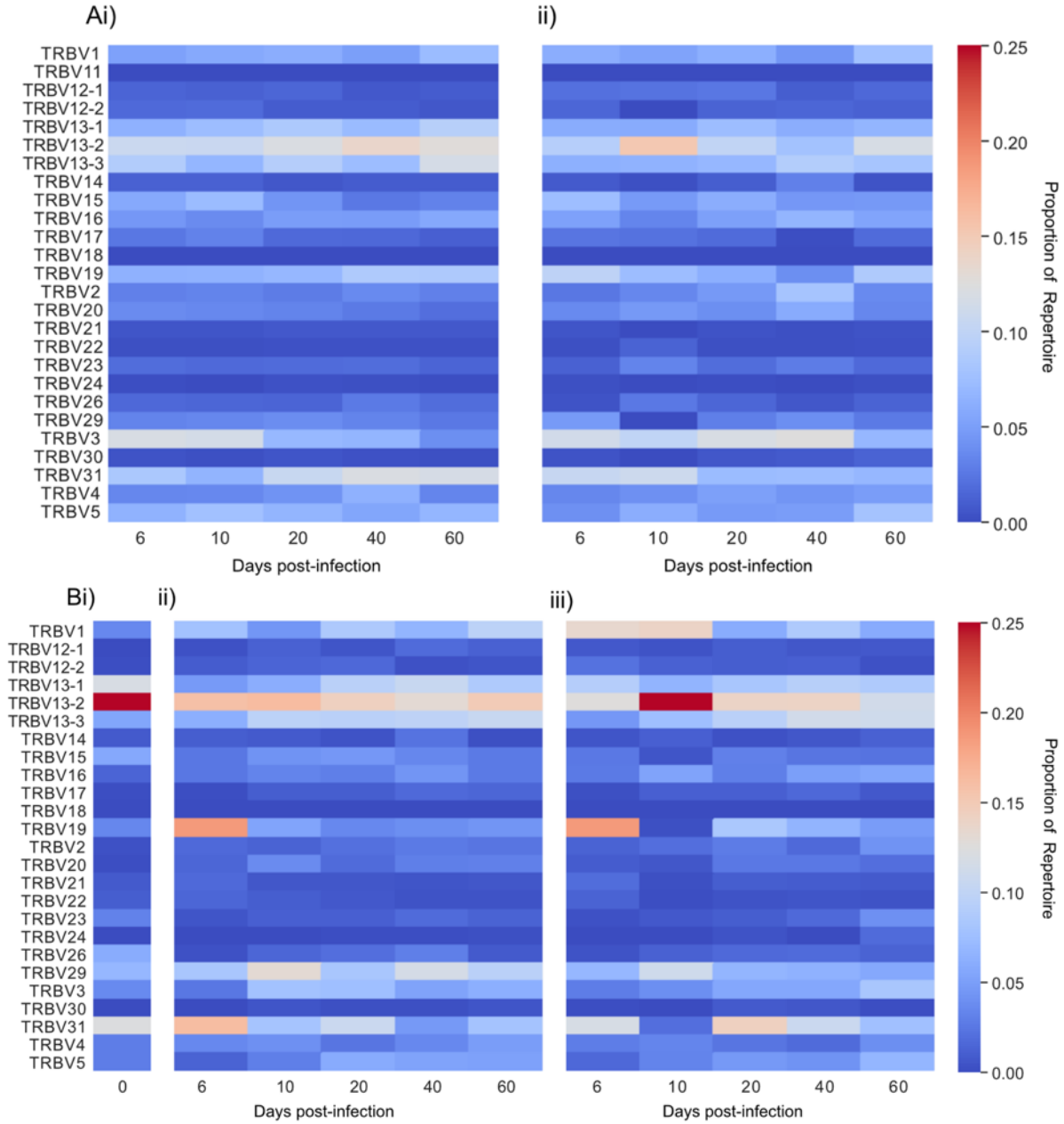


Figure 3.18: Heatmaps shows the mean V-gene usage of A)T_{CM} replicate reperotires per day post-infection for mice infected with (i) MT parasites or (ii) SBP parasites and B) T_{EM} replicate repertoires per day post-infection for i) unchallenged repertoires, ii) mice infected with MT parasites and iii) mice infected with SBP parasites.

Figure 3.19

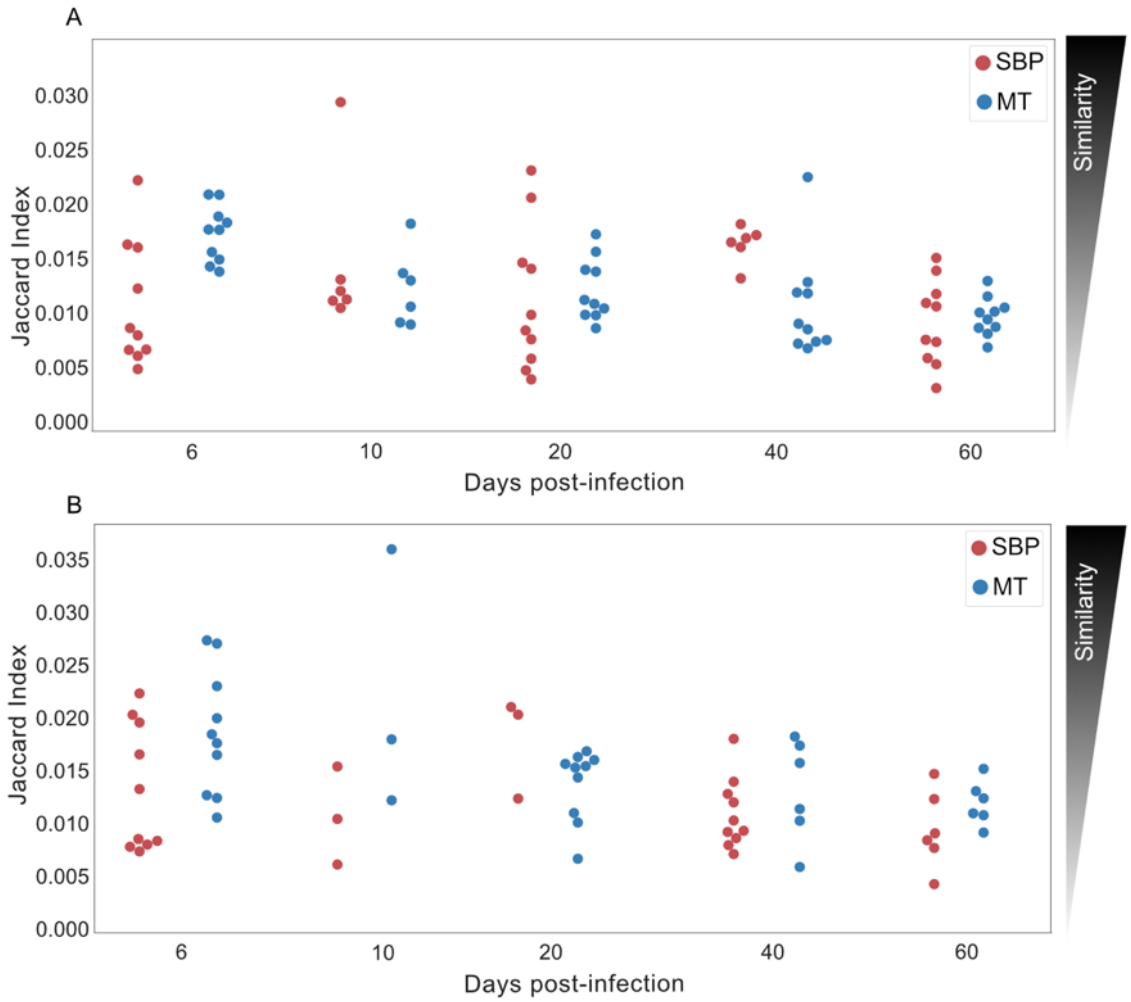


Figure 3.19: Jaccard (similarity) index for replicate mice infected with MT (blue) or SBP (red) parasites, for A) TCM repertoires and B) TEM repertoires.

3.4.5.3. Overlap between T_E, T_{CM} and T_{EM} repertoires

To determine the degree of clonal sharing between the clones that dominated the challenged T_E repertoires and the memory populations, the overlap between the top 100 clones in T_E, T_{CM} and T_{EM} repertoires was calculated. CDR3 overlap analysis was conducted unweighted, to avoid biases from effector clonal expansion. Little to no sharing was found between either the T_{EM} or T_{CM} replicate repertoires themselves for either infection type, nor between the T_{EM}, T_{CM} and T_E populations (Figure 3.20). TRBV3 dominated acute TE repertoires, but was not increased over the time-course of infection in TCM or TEM repertoires, for either transmission type. In agreement with this, although some clones found in OTU1 were detected in some T_{CM} and T_{EM} populations, the cluster was not conserved across replicates, and if present only made

up a small proportion of repertoires, consisting of single-figure unique clone counts (Table 3.4).

Figure 3.20

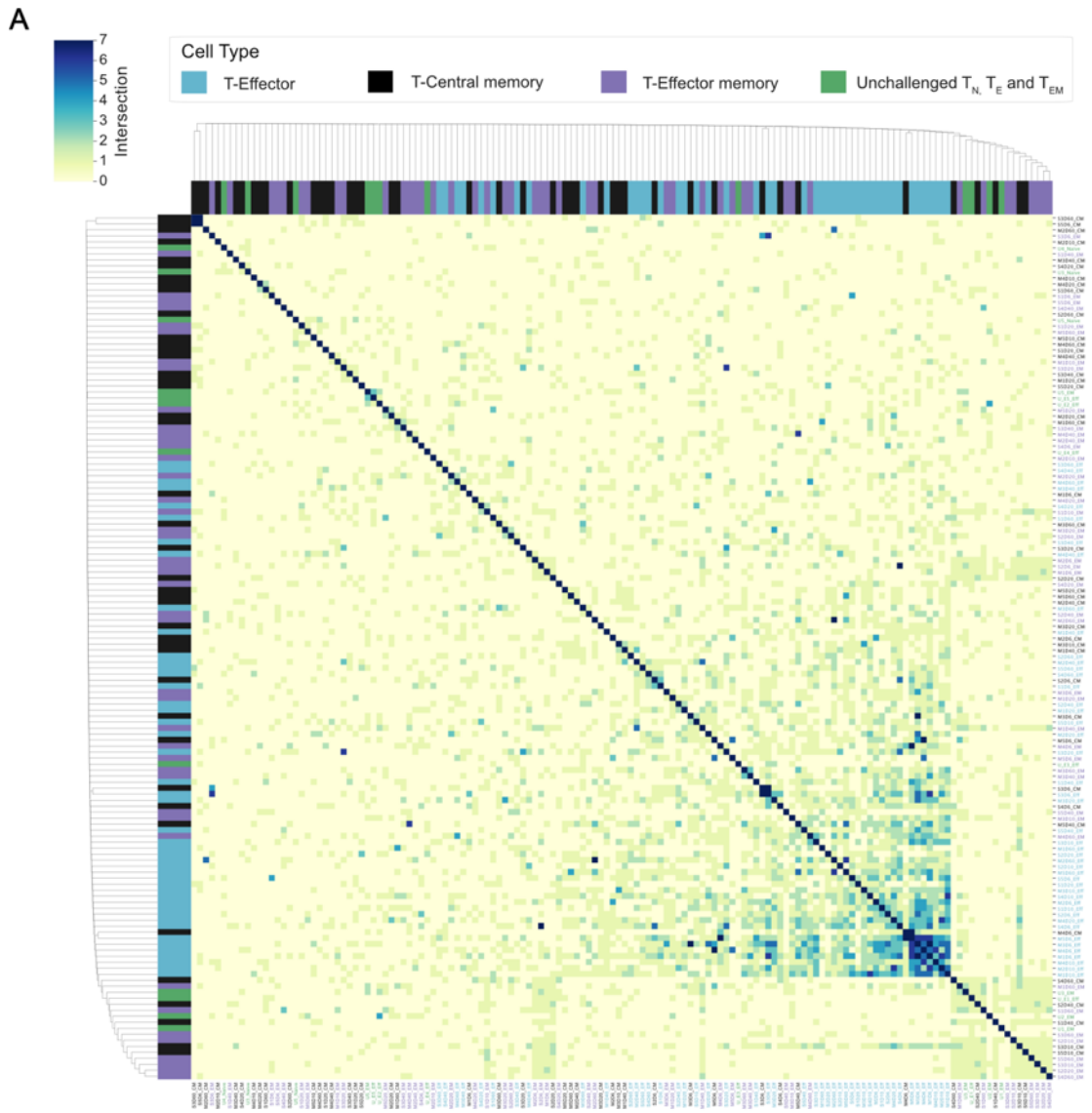


Figure 3.20: A) Cluster map shows the pairwise TCR repertoire overlap of the top 100 CDR3 sequences of individual repertoires, of challenged T_E (blue), T_{CM} (black) and T_{EM} (purple) repertoires, and unchallenged T_N, T_E and T_{EM} repertoires (green). The size of the pairwise intersection between each repertoire (number of shared CDR3 amino acid sequences) is displayed, with greatest overlap evident between acute T_E repertoires.

Table 3.4

Group	OTU1 Proportion (%) (±SEM)	Unique OTU1 clones (±SEM)	No.of mice	Cell type
MT day 6	0.35 (±0.05)	4.2 (±0.97)	5/5	TCM
MT day 10	0.16 (±0.09)	1 (±0.71)	2/5	TCM
MT day 20	0.06 (±0.04)	0.8 (±0.37)	3/5	TCM
MT day 40	0.29 (±0.15)	2 (±1.3)	3/5	TCM
MT day 60	0.04 (±0.04)	0.4 (±0.4)	1/5	TCM
SBP day 6	0.17 (±0.11)	1.2 (±0.8)	2/5	TCM
SBP day 10	0.92 (±0.92)	0.25 (±0.25)	1/4	TCM
SBP day 20	0 (0)	0 (0)	0/5	TCM
SBP day 40	0 (0)	0 (0)	0/4	TCM
SBP day 60	0 (0)	0 (0)	0/5	TCM
MT day 6	0.21 (±0.21)	0.4 (0.24)	1/5	TEM
MT day 10	0.04 (±0.04)	0.33 (0.33)	1/3	TEM
MT day 20	0.44 (±0.15)	2.6 (0.6)	4/5	TEM
MT day 40	0.4 (±0.4)	1 (1)	1/4	TEM
MT day 60	0.58 (±0.31)	1.5 (0.29)	3/4	TEM
SBP day 6	0.04 (±0.04)	0.2 (0.2)	1/5	TEM
SBP day 10	0 (0)	0 (0)	0/4	TEM
SBP day 20	0.08 (±0.02)	1.33 (0.33)	3/4	TEM
SBP day 40	0.02 (±0.01)	0.6 (0.4)	2/5	TEM
SBP day 60	0.02 (±0.02)	0.25 (0.25)	1/4	TEM
Unchallenged	0 (0)	0 (0)	0	TEM

Table 3.4: Mean proportion (%) of OTU1 sequences present in replicate TCM and TEM repertoires. Data is presented as the mean of each replicate group at each time-point (±SEM), and the number of mice within the replicate group that the cluster was found in is also displayed.

3.4.6. Rechallenge results

As a conserved response was not detected within memory populations, a homologous re-challenge infection was undertaken to determine if a similar expansion of TRBV3, as seen in the acute phase of the primary infections, was observed upon secondary challenge. Mice were re-challenged at day 60 post-primary infection, and then sacrificed at day 67 (day 7 post re-challenge). Following re-challenge, mice did not develop a detectable parasitaemia, but a splenic expansion of an activated effector population was evident, though not as marked as following primary challenge (Figure 3.21).

Figure 3.21

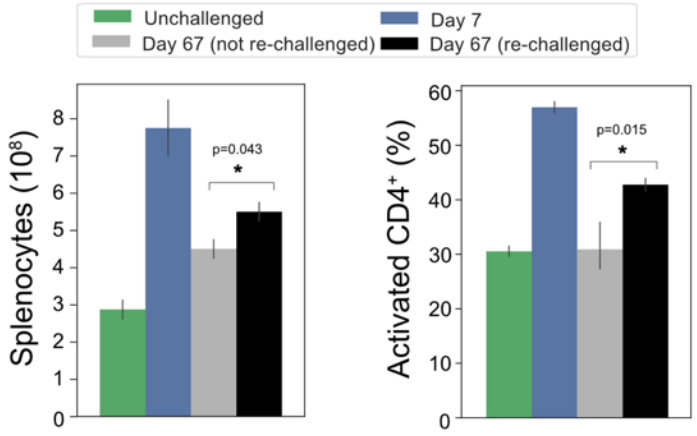


Figure 3.21: A) Splenocyte numbers and frequency of activated CD4$^+$T cells (CD44HI) of unchallenged mice (green), day 7 post-primary infection (blue), day 67 post primary infection (grey, not re-challenged) and 7 days following re-challenge (black, re-challenged day 67). n=3-4 mice per group.

Although TRBV3 was again confirmed to be expanded 7 days after primary challenge, encoding 21.3% (± 3.6, 95% CI) compared to 6.0% (± 2.1, 95% CI) of unchallenged T_N repertoires, no re-expansion of TRBV3 was detected following re-challenge. TRBV3 encoded 11.5% (± 0.4, 95% CI) of the re-challenged T_E repertoires, compared to 10.5% (± 1.7, 95% CI), for mice who had not undergone a secondary challenge, and re-challenged T_E TRBV gene repertoires clustered with later time-points from PC01 rather than acute repertoires (Figure 3.22A). At the clonal level, re-challenged repertoires were found to be as dissimilar to each other as unchallenged repertoires (Mann Whitney U, p=0.29) (Figure 3.22B), with little to no sharing of the 100 most abundant CDR3 amino acid sequences (Figure 3.22C).

Figure 3.22

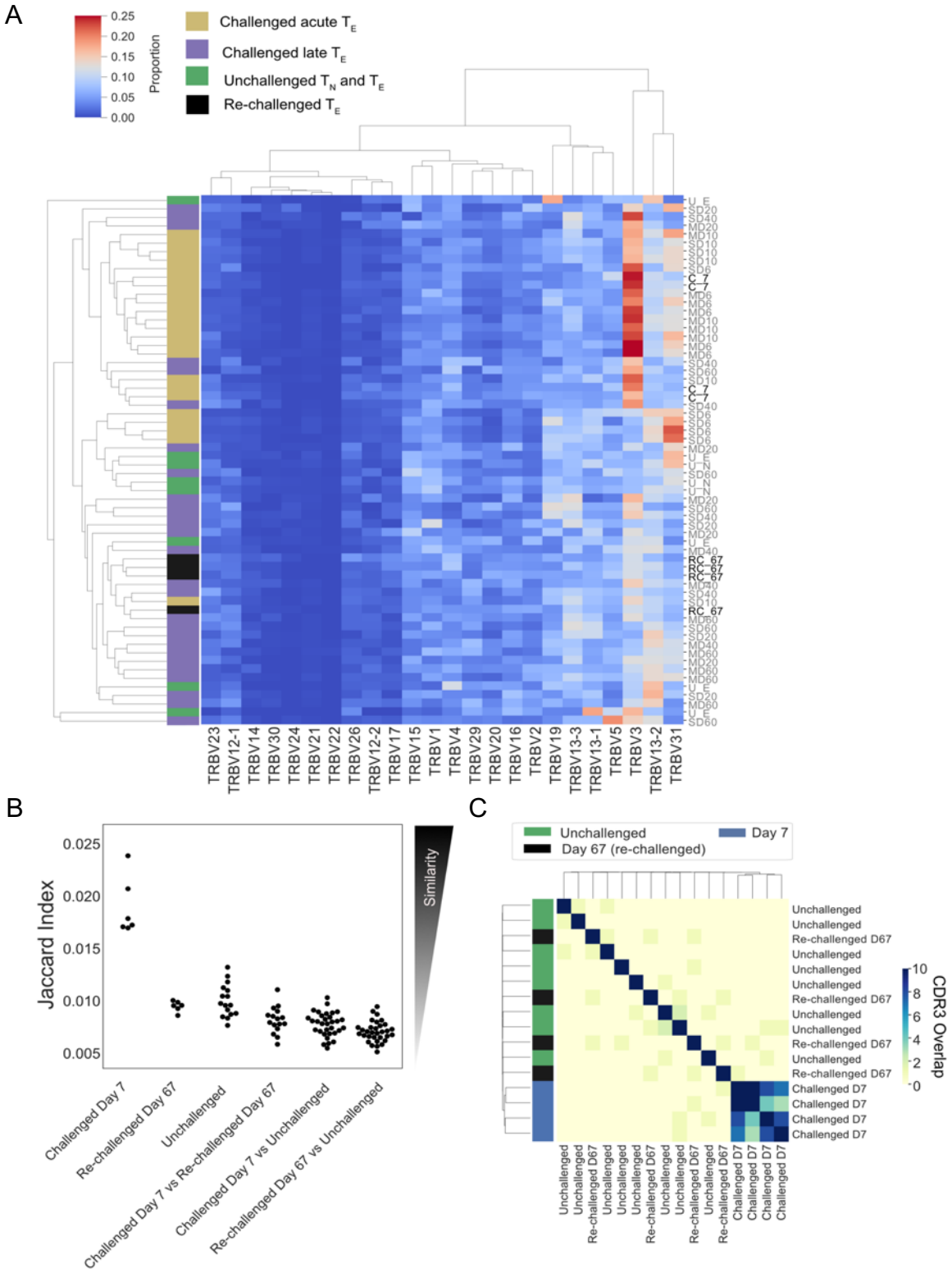


Figure 3.22: A) Clustermap displays the TRBV gene proportion of each repertoire for challenged acute T_E repertoires (gold), challenged late phase T_E repertoires (purple) and unchallenged T_N and T_E repertoires (green) from the first experiment, with re-challenged T_E repertoires (black) also included. B) Jaccard (similarity) index of replicate repertoires (normalized by down-sampling to 5000 UMI) and B) Clustermap depicts the pairwise overlap of the top 100 most abundant CDR3 amino acid sequences of repertoires from the re-challenge experiment. The size of the pairwise intersection between each repertoire (number of shared CDR3 amino acid sequences) is displayed. CDR3 overlap was calculated unweighted by clonotype frequency.

3.5. Discussion

$CD4^+$ T-cells play a critical role in the immune response against the pathological blood-stage of malaria (reviewed in (Nlinwe *et al.*, 2018; Kurup, Butler and Harty, 2019)). However there is a lack of deep mechanistic understanding regarding the development of T-cell mediated immunity against *Plasmodium* (Kurup, Butler and Harty, 2019). To my knowledge, this is the first study to use bulk TCRβ deep-sequencing to examine the composition of splenic $CD4^+$ T-cell repertoires induced by a *Plasmodium* infection. I found that in both the MT and SBP infection models, which show differential control of parasite growth and degrees of immunopathology (Spence *et al.*, 2013; Spence, Brugat and Langhorne, 2015), T_E repertoires elicited upon infection are highly diverse and polyclonal. Despite the upto10-fold increase in splenic T_E cell numbers, the V/J gene usage frequencies of acute challenged T_E repertoires are strongly positively correlated with that of unchallenged T_N repertoires, indicating a highly polyclonal broad expansion of the naïve repertoire, consistent with the massive cellular expansion seen. The degree of correlation between V/J usage frequencies likely indicates a degree of non-specific expansion, although a highly heterogenous response to the vast number of potential antigens expressed by the parasite, argued as a potential cause for the preponderance seen in *P. falciparum* IgG responses (Weiss *et al.*, 2010), cannot be ruled out. However, within this highly polyclonal effector proliferation, an oligo-clonal conserved response is observed in the acute phase following a first infection. Here, repertoires are skewed to TRBV3 gene usage, have a higher degree of clonal sharing, and show increased amino acid sequence similarity of the CDR3 region between dominant clones. Thus, I demonstrate that despite the antigenic complexity of *P. chabaudi,* T_E repertoires bear the hallmarks of a specific response (Dash *et al.*, 2017a; Glanville *et al.*, 2017), mirroring that observed with less antigenically diverse organisms (Emerson *et al.*, 2017; Shugay *et al.*, 2018; Wolf, Hether, Gilchuk, Ahn, *et al.*, 2018). This conserved response is evident in both infection models, although it is delayed and less marked in mice infected with SBP parasites, and does not become evident until parasitaemia has peaked. Adaptive immune responses of lower magnitude have previously been documented in SBP infections compared to MT (Spence *et al.*, 2013) and whilst the precise reasons for this difference remain speculative and represent a complex scenario, the fact that the conserved response does eventually become evident suggests that the drivers of this

response are likely to be common to both MT and SBP parasites despite apparent distinct kinetics in the TRBV3 response. Sequencing of total parasite RNA for both MT and SBP parasites undertaken by Spence et al (2013), indicated the parasites' *cir* gene family - believed to encode a large set of variable antigens displayed on parasitized erythrocytes - was differentially expressed between the two infection models. MT parasites upregulated expression of many genes within this family equally, re-setting broad expansion of the antigenic repertoire, whilst SBP selected for a limited but dominant *cir* expression. Our results are therefore inconsistent with proteins encoded by *cir* genes driving the detected conserved response, which would instead have been expected to result in a more dominant response in SBP infections. I hypothesise that the differences observed in the timing and magnitude of the TRBV3-restricted shared response seen between the two infection models are a result of the systemic inflammation induced by SBP parasites (Spence *et al.*, 2013) disrupting or delaying the formation of an appropriate T-cell response. Future work to determine MHC-presentation pathway and ligands or peptides involved is ultimately required to determine what specific parasite epitope elicits the conserved response.

Public TCR sequences are shared between multiple individuals either due to biases in V(D)J recombination, and/or convergent selection by a common antigen (Emerson *et al.*, 2017; Pogorelyy, Minervina, Chudakov, *et al.*, 2018). The CDR3 amino acid sequences in the most dominant and conserved cluster detected (OTU1) had similar features to other previously observed public TCR responses. These include a reduced number of nucleotide addition and deletion events during VDJ recombination and a greater degree of recombinant convergence (different nucleotide sequences encoding the same amino acid CDR3 sequence). Such recombinational biases during T-cell development mean these TCRs will have a higher probability of generation (PGen) during VDJ recombination and are therefore more likely to be present in a naïve pool. However, the degree of clonal expansion observed in OTU1 sequences was greater than many other sequences of equal or higher Pgen, indicating that these clones were truly expanded and not simply found to be of high frequency as a result of recombinational biases. It has been hypothesised that public TCR responses may provide rapid cross-reactive immunity (Khosravi-Maharlooei *et al.*, 2019) to cope with diverse antigenic challenge, allowing time for more specific private responses to develop (Miles, Douek and Price, 2011; Covacu *et al.*, 2016). Thus, during a *P.*

chabaudi infection, a public response that is mobilised rapidly due to high Pgen and/or a higher chance of positive selection if cross-reactive, may act as a first line of defence against the parasite before more specific responses become effective. In agreement with this, OTU1 appears to be temporally associated with enhanced control of parasitaemia. It arises earlier and is more dominant in MT infections compared with SBP, where rapid parasite growth is observed alongside a delayed and less marked conserved response. Mice infected with MT parasites also show reduced disease severity (Spence *et al.*, 2013). Despite these positive associations, whether this conserved response is truly beneficial to the host, remains unknown. There are reports of public TCR responses being implicated in self-related immunity (Madi *et al.*, 2014), and in *P. berghei*, the presence of conserved pathogenic $CD8^+$ T-cells has been used to predict cerebral malaria (Mariotti-Ferrandiz *et al.*, 2016). Determining whether or not the response is specific to a *P. chabaudi* infection, or is simply a non-specific response to inflammation, is also crucial to determine the importance of the shared response.

No expanded conserved response was evident in memory populations at the timepoints tested. Using a transgenic TCR, Soon *et al.* (2020) recently demonstrated that transition from effector to memory occurred gradually over several weeks in a *P. chabaudi* infection, rather than a linear model of naïve straight to memory. It is known that precursors of memory T-cells derive from an earlier T_E population ($CD127^-$ $CD62L^{hi}$) that precedes the terminally differentiated $CD62L^-$ $CD127^-$ T_E population captured by our T_E gating strategy (Opata *et al.*, 2015). While that precursor population successfully expands to dominate the T_E repertoire, since memory populations preferentially favour TCRs with high avidity, it is possible that the public TCR clones identified lack sufficient avidity required to enter the memory pool (Gasper, Tejera and Suresh, 2014), functioning primarily as short-lived effectors only.

A low level of overlap between replicate T_M repertoires indicates that splenic memory responses are diverse and mostly private to each individual. V-gene usage in both memory populations was diverse, and was not obviously altered over the time-course of infection, although TRBV19 was over-represented at day 6 post-infection in T_{EM} repertoires for both transmission types. Similarly, in the re-challenged T_E compartment, no increase in specific V-gene usage, including TRBV3, was detected

compared to later time points that had not undergone re-infection, and at the clonal level responses had as little overlap between replicates as unchallenged repertoires. This does not imply lack of a memory/primed response; mice are protected from re-infection and have previously been shown to have primed secondary responses (Opata and Stephens, 2017). Heterologous responses could be due to diverse T-cells more specific for the parasite than the TRBV3+ public response surviving the contraction phase/entering memory to be able to respond, whilst the public response appears to be a short-lived effector response only (Opata *et al.*, 2018). Diverse or 'private' responses have previously been seen with other complex pathogens such as *mycobacteria* (Thomas *et al.*, 2014; Sun *et al.*, 2017) and would not be unexpected given the antigenic complexity of *Plasmodium*. However, primed private secondary responses cannot be confirmed without experiments that sample the same individuals longitudinally over the course of a primary and secondary infection, and this hypothesis therefore remains untested and an avenue for future investigation.

These results also do not exclude the possibility that a low-level re-expansion of TRBV3 does occur upon re-challenge, given only a small increase in the number of activated cells was observed. This may not have constituted a large enough proportion of the repertoire, given the number of animals used in re-infection, to detect subtle differences in re-challenged T_E V-gene usage. This is supported by work done by Opata et al (2017), who demonstrated that T-cells specific to *P. chabaudi* MSP-1 that survive the contraction phase, do not re-expand upon secondary challenge due to increased levels of apoptosis. Such clones that are responding but do not accumulate/re-expand would be difficult to discern from baseline controls using bulk repertoire sequencing techniques. Future re-infection studies may benefit from including single-cell transcriptomics to help differentiate such responding cells based on phenotype.

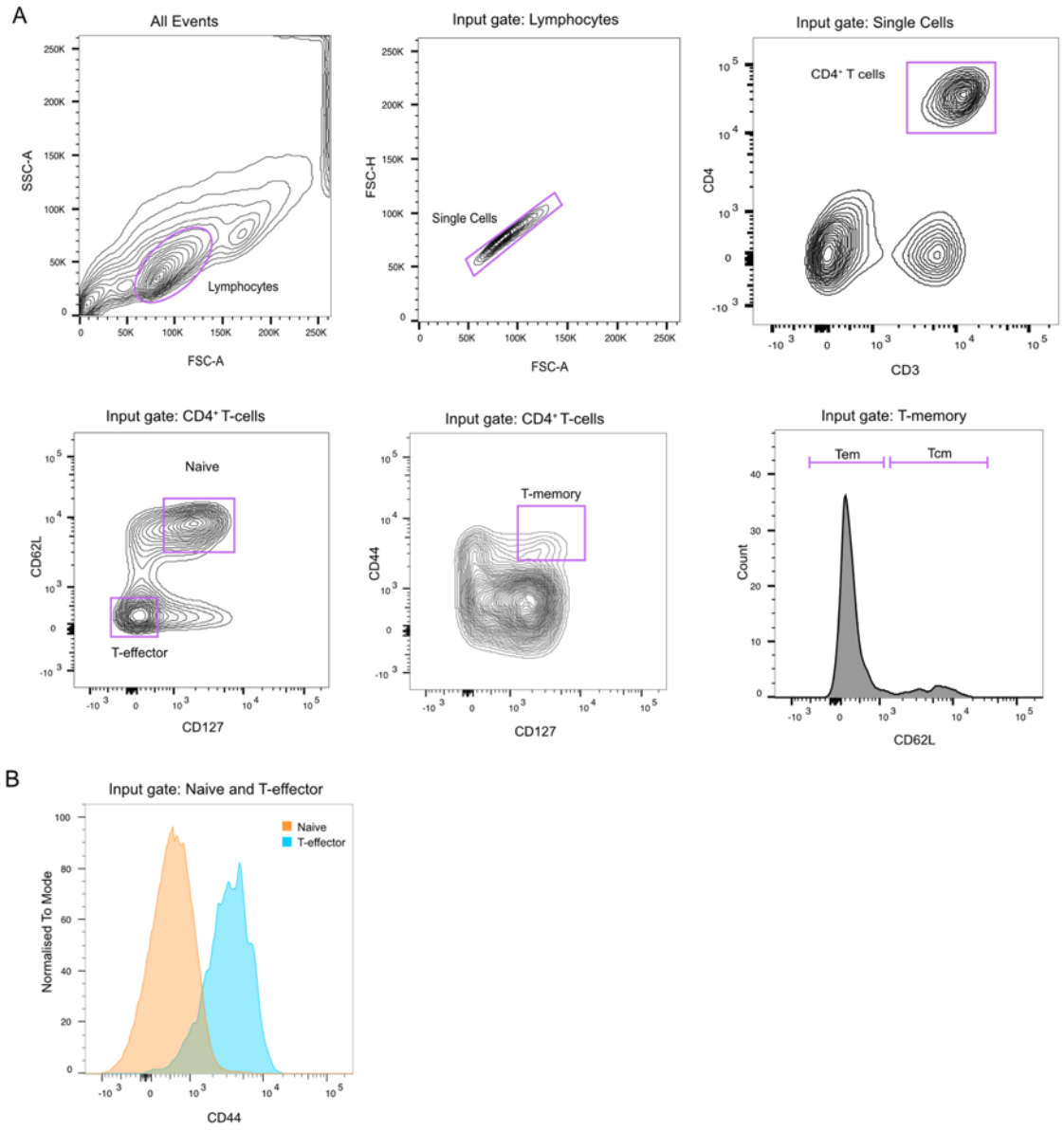
Future studies are also required to determine whether the conserved response plays a critical role in protection during a first infection. MT infections develop chronic recrudescing infection, so although the response does not fully clear infection, it is temporally associated with control of parasite growth and reduced disease severity (Spence *et al.*, 2013). Short-lived T-effectors present at the peak of parasitaemia have also previously been shown to provide a higher degree of protection on adoptive

transfer than memory T-cells from later time-points (Opata *et al.*, 2018), though the use of transgenic models in such previous studies means they are not directly comparable to experiments presented here. Our own adoptive transfer experiments of TRBV3^{+} cells from the peak of infection, although desirable, are currently hindered by a lack of monoclonal agents against TRBV3, but represent a future line of investigation. Whether entirely favourable to the host or not, I hypothesise the response may allow time for mechanisms that govern the formation of more specific private responses and subsequent immunity to the parasite to develop. Mice are known to develop highly effective strain-specific anti-parasite immunity after a single malaria episode, whilst this takes years to develop in people. Therefore, if some degree of first line protection were shown, such a response could be a novel target to promote in malaria naïve individuals potentially through vaccination, providing initial cover whilst more specific but slower private responses develop, conveying higher levels of protection. Even if only partially protective, the response is conserved between individuals and receptor sequences have a high probability of generation. Targeting these is predicted therefore to increase vaccine success rate (Sethna *et al.*, 2018; Hill *et al.*, 2019).

In summary, I have demonstrated that a conserved TCRβ signature encoded by TRBV3 is consistently expanded in response to a first *Plasmodium chabaudi* infection. In contrast, memory formation appears to be a more private response to the individual, in agreement with a lack of a shared response following homologous re-challenge. Understanding the specificity, antigenic drivers and contribution to protection (or pathogenesis) of this conserved signature, that is consistently a hallmark of a first infection, is ultimately required to determine if it should be promoted or mitigated for malaria therapeutic purposes.

3.6. Appendices

Appendix 3.1: Murine Gating Strategy



Appendix 3.1: A) Gating strategy for murine T-cell populations, illustrated from a representative mouse infected with SBP *P. chabaudi (AS)* at day 6 post-infection. B) Histogram showing CD44 level of T-naïve and T-effector populations.

Appendix 3.2: T-Effector metadata for experiment PC01

File	Cell_type	Filtered	Sample_ID	Sample_ID_Eff	Infection	Days_pi	Unique_CDR3	Functional TCR
180309_M01270_0440_000000000-BNH7G_1_11253CG0001L01.fastq.gz	T_E	N	S1D60	S1D60_Eff	SBP	60	454	29109
180309_M01270_0440_000000000-BNH7G_1_11253CG0002L01.fastq.gz	T_E	N	S2D60	S2D60_Eff	SBP	60	246	43003
180309_M01270_0440_000000000-BNH7G_1_11253CG0003L01.fastq.gz	T_E	N	S3D60	S3D60_Eff	SBP	60	950	36126
180309_M01270_0440_000000000-BNH7G_1_11253CG0004L01.fastq.gz	T_E	N	S4D60	S4D60_Eff	SBP	60	325	33393
180309_M01270_0440_000000000-BNH7G_1_11253CG0005L01.fastq.gz	T_E	N	S5D60	S5D60_Eff	SBP	60	672	39727
180309_M01270_0440_000000000-BNH7G_1_11253CG0006L01.fastq.gz	T_E	N	S1D40	S1D40_Eff	SBP	60	240	34088
180309_M01270_0440_000000000-BNH7G_1_11253CG0007L01.fastq.gz	T_E	N	S2D40	S2D40_Eff	SBP	40	330	34286
180309_M01270_0440_000000000-BNH7G_1_11253CG0008L01.fastq.gz	T_E	N	S3D40	S3D40_Eff	SBP	40	277	38148
180309_M01270_0440_000000000-BNH7G_1_11253CG0009L01.fastq.gz	T_E	N	S4D40	S4D40_Eff	SBP	40	890	49349
180309_M01270_0440_000000000-BNH7G_1_11253CG0010L01.fastq.gz	T_E	N	S5D40	S5D40_Eff	SBP	40	468	39118
180309_M01270_0440_000000000-BNH7G_1_11253CG0011L01.fastq.gz	T_E	Y	S1D20	S1D20_Eff	SBP	40	21	4260
180309_M01270_0440_000000000-BNH7G_1_11253CG0012L01.fastq.gz	T_E	N	S2D20	S2D20_Eff	SBP	20	710	51474
180309_M01270_0440_000000000-BNH7G_1_11253CG0013L01.fastq.gz	T_E	N	S3D20	S3D20_Eff	SBP	20	388	27681
180309_M01270_0440_000000000-BNH7G_1_11253CG0014L01.fastq.gz	T_E	N	S4D20	S4D20_Eff	SBP	20	1183	50516
180309_M01270_0440_000000000-BNH7G_1_11253CG0015L01.fastq.gz	T_E	N	S5D20	S5D20_Eff	SBP	20	208	38333
180309_M01270_0440_000000000-BNH7G_1_11253CG0016L01.fastq.gz	T_E	N	S1D10	S1D10_Eff	SBP	20	1301	39123
180309_M01270_0440_000000000-BNH7G_1_11253CG0017L01.fastq.gz	T_E	N	S2D10	S2D10_Eff	SBP	10	2365	36983
180309_M01270_0440_000000000-BNH7G_1_11253CG0018L01.fastq.gz	T_E	N	S3D10	S3D10_Eff	SBP	10	1663	46015
180309_M01270_0440_000000000-BNH7G_1_11253CG0019L01.fastq.gz	T_E	N	S4D10	S4D10_Eff	SBP	10	3029	35158
180309_M01270_0440_000000000-BNH7G_1_11253CG0020L01.fastq.gz	T_E	N	S5D10	S5D10_Eff	SBP	10	1169	36072
180309_M01270_0440_000000000-BNH7G_1_11253CG0021L01.fastq.gz	T_E	N	S1D6	S1D6_Eff	SBP	10	3031	29434
180309_M01270_0440_000000000-BNH7G_1_11253CG0022L01.fastq.gz	T_E	N	S2D6	S2D6_Eff	SBP	6	390	29667
180309_M01270_0440_000000000-BNH7G_1_11253CG0023L01.fastq.gz	T_E	N	S3D6	S3D6_Eff	SBP	6	7330	44038
180309_M01270_0440_000000000-BNH7G_1_11253CG0024L01.fastq.gz	T_E	N	S4D6	S4D6_Eff	SBP	6	8175	42450
180309_M01270_0440_000000000-BNH7G_1_11253CG0025L01.fastq.gz	T_E	N	S5D6	S5D6_Eff	SBP	6	6157	27633

180309_M01270_0440_000000000-BNH7G_1_11253CG0026L01.fastq.gz	T_E	N	U_1	U_1_Eff	Unchallenged	6	107	10342
180309_M01270_0440_000000000-BNH7G_1_11253CG0027L01.fastq.gz	T_E	N	U_2	U_2_Eff	Unchallenged	0	377	48668
180309_M01270_0440_000000000-BNH7G_1_11253CG0028L01.fastq.gz	T_E	N	U_3	U_3_Eff	Unchallenged	0	419	45967
180309_M01270_0440_000000000-BNH7G_1_11253CG0029L01.fastq.gz	T_E	N	U_4	U_4_Eff	Unchallenged	0	354	42171
180309_M01270_0440_000000000-BNH7G_1_11253CG0030L01.fastq.gz	T_E	N	U_5	U_5_Eff	Unchallenged	0	180	31736
180309_M01270_0440_000000000-BNH7G_1_11253CG0031L01.fastq.gz	T_E	N	M1D6	M1D6_Eff	MT	0	4833	38570
180309_M01270_0440_000000000-BNH7G_1_11253CG0032L01.fastq.gz	T_E	N	M2D6	M2D6_Eff	MT	6	2860	35127
180309_M01270_0440_000000000-BNH7G_1_11253CG0033L01.fastq.gz	T_E	N	M3D6	M3D6_Eff	MT	6	4645	42820
180309_M01270_0440_000000000-BNH7G_1_11253CG0034L01.fastq.gz	T_E	N	M4D6	M4D6_Eff	MT	6	3418	42589
180309_M01270_0440_000000000-BNH7G_1_11253CG0035L01.fastq.gz	T_E	N	M5D6	M5D6_Eff	MT	6	2647	42155
180309_M01270_0440_000000000-BNH7G_1_11253CG0036L01.fastq.gz	T_E	N	M1D10	M1D10_Eff	MT	6	2852	47418
180309_M01270_0440_000000000-BNH7G_1_11253CG0037L01.fastq.gz	T_E	N	M2D10	M2D10_Eff	MT	10	2146	45815
180309_M01270_0440_000000000-BNH7G_1_11253CG0038L01.fastq.gz	T_E	N	M3D10	M3D10_Eff	MT	10	2395	36258
180309_M01270_0440_000000000-BNH7G_1_11253CG0039L01.fastq.gz	T_E	N	M4D10	M4D10_Eff	MT	10	1646	51373
180309_M01270_0440_000000000-BNH7G_1_11253CG0040L01.fastq.gz	T_E	Y	M5D10	M5D10_Eff	MT	10	8	369
180309_M01270_0440_000000000-BNH7G_1_11253CG0041L01.fastq.gz	T_E	N	M1D20	M1D20_Eff	MT	10	336	42044
180309_M01270_0440_000000000-BNH7G_1_11253CG0042L01.fastq.gz	T_E	N	M2D20	M2D20_Eff	MT	20	265	38307
180309_M01270_0440_000000000-BNH7G_1_11253CG0043L01.fastq.gz	T_E	N	M3D20	M3D20_Eff	MT	20	502	42629
180309_M01270_0440_000000000-BNH7G_1_11253CG0044L01.fastq.gz	T_E	N	M4D20	M4D20_Eff	MT	20	893	44536
180309_M01270_0440_000000000-BNH7G_1_11253CG0045L01.fastq.gz	T_E	N	M5D20	M5D20_Eff	MT	20	876	41371
180309_M01270_0440_000000000-BNH7G_1_11253CG0046L01.fastq.gz	T_E	N	M1D40	M1D40_Eff	MT	20	325	43278
180309_M01270_0440_000000000-BNH7G_1_11253CG0047L01.fastq.gz	T_E	Y	M2D40	M2D40_Eff	MT	40	1	1
180309_M01270_0440_000000000-BNH7G_1_11253CG0048L01.fastq.gz	T_E	N	M3D40	M3D40_Eff	MT	40	778	44117
180309_M01270_0440_000000000-BNH7G_1_11253CG0049L01.fastq.gz	T_E	N	M4D40	M4D40_Eff	MT	40	559	43357
180309_M01270_0440_000000000-BNH7G_1_11253CG0050L01.fastq.gz	T_E	Y	M5D40	M5D40_Eff	MT	40	48	14438
180309_M01270_0440_000000000-BNH7G_1_11253CG0051L01.fastq.gz	T_E	N	M1D60	M1D60_Eff	MT	40	398	25397

180309_M01270_0440_000000000-BNH7G_1_11253CG0052L01.fastq.gz	T_E	N	M2D60	M2D60_Eff	MT	60	513	24439
180309_M01270_0440_000000000-BNH7G_1_11253CG0053L01.fastq.gz	T_E	N	M3D60	M3D60_Eff	MT	60	536	27618
180309_M01270_0440_000000000-BNH7G_1_11253CG0054L01.fastq.gz	T_E	N	M4D60	M4D60_Eff	MT	60	1443	45390
180309_M01270_0440_000000000-BNH7G_1_11253CG0055L01.fastq.gz	T_E	N	M5D60	M5D60_Eff	MT	60	2423	35515

Appendix 3.3: Metadata for experiment PC02

File	Cell_type	Filtered	Sample_ID	Infection	Days_pi	Unique_CDR3	Unique_UMIS	Functional TCR
190226_M05898_0061_000000000-C83MC_1_11641CG0023L01_1.fastq.gz	T_N	N	U1_D7	Unchallenged	0	930	942	24210
190226_M05898_0061_000000000-C83MC_1_11641CG0024L01_1.fastq.gz	T_N	N	U2_D7	Unchallenged	0	11448	16051	71930
190226_M05898_0061_000000000-C83MC_1_11641CG0025L01_1.fastq.gz	T_N	N	U3_D7	Unchallenged	0	15127	21236	110211
190226_M05898_0061_000000000-C83MC_1_11641CG0026L01_1.fastq.gz	T_N	N	U4_D14	Unchallenged	0	9791	14117	81962
190226_M05898_0061_000000000-C83MC_1_11641CG0027L01_1.fastq.gz	T_N	N	U5_D14	Unchallenged	0	12106	18094	120803
190226_M05898_0061_000000000-C83MC_1_11641CG0028L01_1.fastq.gz	T_N	N	U6_D14	Unchallenged	0	13848	19356	113475
190226_M05898_0061_000000000-C83MC_1_11641CG0029L01_1.fastq.gz	T_E	N	M1_D4	MT	4	3345	8227	67530
190226_M05898_0061_000000000-C83MC_1_11641CG0030L01_1.fastq.gz	T_E	N	M2_D4	MT	4	3512	8781	77292
190226_M05898_0061_000000000-C83MC_1_11641CG0031L01_1.fastq.gz	T_E	Y	M3_D4	MT	4	302	457	27895
190226_M05898_0061_000000000-C83MC_1_11641CG0032L01_1.fastq.gz	T_E	N	M4_D4	MT	4	2281	5664	62279
190226_M05898_0061_000000000-C83MC_1_11641CG0033L01_1.fastq.gz	T_E	N	M1_D7	MT	7	6256	14236	102005
190226_M05898_0061_000000000-C83MC_1_11641CG0034L01_1.fastq.gz	T_E	N	M2_D7	MT	7	7848	18216	84921
190226_M05898_0061_000000000-C83MC_1_11641CG0035L01_1.fastq.gz	T_E	Y	M3_D7	MT	7	381	471	56496
190226_M05898_0061_000000000-C83MC_1_11641CG0036L01_1.fastq.gz	T_E	N	M4_D7	MT	7	4026	7562	103405
190226_M05898_0061_000000000-C83MC_1_11641CG0037L01_1.fastq.gz	T_E	N	M1_D11	MT	11	5979	11637	98248
190226_M05898_0061_000000000-C83MC_1_11641CG0038L01_1.fastq.gz	T_E	N	M2_D11	MT	11	7803	17441	97835
190226_M05898_0061_000000000-C83MC_1_11641CG0039L01_1.fastq.gz	T_E	Y	M3_D11	MT	11	840	993	85592
190226_M05898_0061_000000000-C83MC_1_11641CG0040L01_1.fastq.gz	T_E	N	M4_D11	MT	11	9368	21610	125426
190226_M05898_0061_000000000-C83MC_1_11641CG0041L01_1.fastq.gz	T_E	N	M1_D14	MT	14	7966	18261	108951
190226_M05898_0061_000000000-C83MC_1_11641CG0042L01_1.fastq.gz	T_E	N	M2_D14	MT	14	7977	16375	89269
190226_M05898_0061_000000000-C83MC_1_11641CG0043L01_1.fastq.gz	T_E	Y	M3_D14	MT	14	1122	1332	110932
190226_M05898_0061_000000000-C83MC_1_11641CG0044L01_1.fastq.gz	T_E	N	M4_D14	MT	14	6148	10757	130407

Appendix 3.4: Metadata for experiment PC03

File	Cell_type	Filtered	Sample_ID	Infection	Days_pi	Unique_CDR3	Unique_UMIS	Functional TCR
190909_M05898_0108_000000000-CGCRN_1_11895CG0033L01_1.fastq.gz	T_E	N	C_7.1	MT	7	7619	14896	100839
190909_M05898_0108_000000000-CGCRN_1_11895CG0034L01_1.fastq.gz	T_E	N	C_7.2	MT	7	7719	11830	162081
190909_M05898_0108_000000000-CGCRN_1_11895CG0035L01_1.fastq.gz	T_E	N	C_7.3	MT	7	12147	30813	110514
190909_M05898_0108_000000000-CGCRN_1_11895CG0036L01_1.fastq.gz	T_E	N	C_7.4	MT	7	12036	26025	138365
190909_M05898_0108_000000000-CGCRN_1_11895CG0037L01_1.fastq.gz	T_E	N	RC_67.1	MT, re-challenged	67	6662	10913	213145
190909_M05898_0108_000000000-CGCRN_1_11895CG0038L01_1.fastq.gz	T_E	N	RC_67.2	MT, re-challenged	67	5876	9353	191620
190909_M05898_0108_000000000-CGCRN_1_11895CG0039L01_1.fastq.gz	T_E	N	RC_67.3	MT, re-challenged	67	8197	13787	186053
190909_M05898_0108_000000000-CGCRN_1_11895CG0040L01_1.fastq.gz	T_E	N	RC_67.4	MT, re-challenged	67	6218	9301	162864
190909_M05898_0108_000000000-CGCRN_1_11895CG0001L01_1.fastq.gz	T_N	N	U_67.1	Unchallenged	0	19283	23979	82110
190909_M05898_0108_000000000-CGCRN_1_11895CG0002L01_1.fastq.gz	T_N	N	U_67.2	Unchallenged	0	8074	10442	143464
190909_M05898_0108_000000000-CGCRN_1_11895CG0003L01_1.fastq.gz	T_N	N	U_67.3	Unchallenged	0	8065	10480	110581
190909_M05898_0108_000000000-CGCRN_1_11895CG0004L01_1.fastq.gz	T_N	N	U_67.4	Unchallenged	0	9456	11824	107811
190909_M05898_0108_000000000-CGCRN_1_11895CG0005L01_1.fastq.gz	T_N	N	U_7.1	Unchallenged	0	11525	13456	59812
190909_M05898_0108_000000000-CGCRN_1_11895CG0006L01_1.fastq.gz	T_N	N	U_7.2	Unchallenged	0	12447	14937	84610
190909_M05898_0108_000000000-CGCRN_1_11895CG0007L01_1.fastq.gz	T_N	N	U_7.3	Unchallenged	0	3808	5007	46595
190909_M05898_0108_000000000-CGCRN_1_11895CG0008L01_1.fastq.gz	T_N	N	U_7.4	Unchallenged	0	18555	23354	139783

Appendix 3.5: Metadata for PC01 unchallenged T-naïves

File	Cell_type	Filtered	Sample_ID	Sample_ID_TN	Infection	Days_pi	Unique_CDR3	Functional TCR
180420_M05898_0004_000000000-BPT25_1_11316CG0026L01_1.fastq.gz	T_N	Y	U_1	U_1_N	Unchallenged	0	261	7759
180420_M05898_0004_000000000-BPT25_1_11316CG0027L01_1.fastq.gz	T_N	Y	U_2	U_2_N	Unchallenged	0	111	3542
180420_M05898_0004_000000000-BPT25_1_11316CG0028L01_1.fastq.gz	T_N	N	U_3	U_3_N	Unchallenged	0	1678	34313
180420_M05898_0004_000000000-BPT25_1_11316CG0029L01_1.fastq.gz	T_N	N	U_4	U_4_N	Unchallenged	0	1657	26703
180420_M05898_0004_000000000-BPT25_1_11316CG0030L01_1.fastq.gz	T_N	N	U_5	U_5_N	Unchallenged	0	1290	18453

Appendix 3.6: Metadata for PC01 T-central memory

File	Cell_type	Filtered	Sample_ID	Sample_ID_CM	Infection	Days_pi	Unique_CDR3	Functional TCR
181114_M05898_0037_000000000-C59F6_1_11519CG0001L01.fastq.gz.clones.txt	T_CM	N	S1D60	S1D60_CM	SBP	60	474	26700
181114_M05898_0037_000000000-C59F6_1_11519CG0002L01.fastq.gz.clones.txt	T_CM	N	S2D60	S2D60_CM	SBP	60	279	42862
181114_M05898_0037_000000000-C59F6_1_11519CG0003L01.fastq.gz.clones.txt	T_CM	N	S3D60	S3D60_CM	SBP	60	316	39241
181114_M05898_0037_000000000-C59F6_1_11519CG0004L01.fastq.gz.clones.txt	T_CM	N	S4D60	S4D60_CM	SBP	60	173	32463
181114_M05898_0037_000000000-C59F6_1_11519CG0005L01.fastq.gz.clones.txt	T_CM	N	S5D60	S5D60_CM	SBP	60	90	14603
181114_M05898_0037_000000000-C59F6_1_11519CG0006L01.fastq.gz.clones.txt	T_CM	N	S1D40	S1D40_CM	SBP	40	109	32508
181114_M05898_0037_000000000-C59F6_1_11519CG0007L01.fastq.gz.clones.txt	T_CM	N	S2D40	S2D40_CM	SBP	40	137	35685
181114_M05898_0037_000000000-C59F6_1_11519CG0008L01.fastq.gz.clones.txt	T_CM	N	S3D40	S3D40_CM	SBP	40	209	39994
181114_M05898_0037_000000000-C59F6_1_11519CG0009L01.fastq.gz.clones.txt	T_CM	N	S4D40	S4D40_CM	SBP	40	41	10528
181114_M05898_0037_000000000-C59F6_1_11519CG0010L01.fastq.gz.clones.txt	T_CM	Y	S5D40	S5D40_CM	SBP	40	2	3
181114_M05898_0037_000000000-C59F6_1_11519CG0011L01.fastq.gz.clones.txt	T_CM	N	S1D20	S1D20_CM	SBP	20	542	48294
181114_M05898_0037_000000000-C59F6_1_11519CG0012L01.fastq.gz.clones.txt	T_CM	N	S2D20	S2D20_CM	SBP	20	151	41333
181114_M05898_0037_000000000-C59F6_1_11519CG0013L01.fastq.gz.clones.txt	T_CM	N	S3D20	S3D20_CM	SBP	20	249	46665
181114_M05898_0037_000000000-C59F6_1_11519CG0014L01.fastq.gz.clones.txt	T_CM	N	S4D20	S4D20_CM	SBP	20	726	46072
181114_M05898_0037_000000000-C59F6_1_11519CG0015L01.fastq.gz.clones.txt	T_CM	N	S5D20	S5D20_CM	SBP	20	154	57236
181114_M05898_0037_000000000-C59F6_1_11519CG0016L01.fastq.gz.clones.txt	T_CM	N	S1D10	S1D10_CM	SBP	10	143	18516
181114_M05898_0037_000000000-C59F6_1_11519CG0017L01.fastq.gz.clones.txt	T_CM	N	S2D10	S2D10_CM	SBP	10	136	18503
181114_M05898_0037_000000000-C59F6_1_11519CG0018L01.fastq.gz.clones.txt	T_CM	N	S3D10	S3D10_CM	SBP	10	156	29974
181114_M05898_0037_000000000-C59F6_1_11519CG0019L01.fastq.gz.clones.txt	T_CM	Y	S4D10	S4D10_CM	SBP	10	30	9714
181114_M05898_0037_000000000-C59F6_1_11519CG0020L01.fastq.gz.clones.txt	T_CM	N	S5D10	S5D10_CM	SBP	10	150	20802
181114_M05898_0037_000000000-C59F6_1_11519CG0021L01.fastq.gz.clones.txt	T_CM	N	S1D6	S1D6_CM	SBP	6	110	10921
181114_M05898_0037_000000000-C59F6_1_11519CG0022L01.fastq.gz.clones.txt	T_CM	N	S2D6	S2D6_CM	SBP	6	771	53419
181114_M05898_0037_000000000-C59F6_1_11519CG0023L01.fastq.gz.clones.txt	T_CM	N	S3D6	S3D6_CM	SBP	6	1369	37430
181114_M05898_0037_000000000-C59F6_1_11519CG0024L01.fastq.gz.clones.txt	T_CM	N	S4D6	S4D6_CM	SBP	6	276	36951
181114_M05898_0037_000000000-C59F6_1_11519CG0025L01.fastq.gz.clones.txt	T_CM	N	S5D6	S5D6_CM	SBP	6	209	37801

181114_M05898_0037_000000000-C59F6_1_11519CG0026L01.fastq.gz.clones.txt	T_CM	Y	U_1CM	U_1CM_CM	Unchallenged	0	2	2
181114_M05898_0037_000000000-C59F6_1_11519CG0027L01.fastq.gz.clones.txt	T_CM	Y	U_2CM	U_2CM_CM	Unchallenged	0	5	1365
181114_M05898_0037_000000000-C59F6_1_11519CG0028L01.fastq.gz.clones.txt	T_CM	Y	U_3CM	U_3CM_CM	Unchallenged	0	3	714
181114_M05898_0037_000000000-C59F6_1_11519CG0029L01.fastq.gz.clones.txt	T_CM	Y	U_4CM	U_4CM_CM	Unchallenged	0	3	200
181114_M05898_0037_000000000-C59F6_1_11519CG0030L01.fastq.gz.clones.txt	T_CM	Y	U_5CM	U_5CM_CM	Unchallenged	0	10	1870
181114_M05898_0037_000000000-C59F6_1_11519CG0031L01.fastq.gz.clones.txt	T_CM	N	M1D6	M1D6_CM	MT	6	1818	69145
181114_M05898_0037_000000000-C59F6_1_11519CG0032L01.fastq.gz.clones.txt	T_CM	N	M2D6	M2D6_CM	MT	6	920	76329
181114_M05898_0037_000000000-C59F6_1_11519CG0033L01.fastq.gz.clones.txt	T_CM	N	M3D6	M3D6_CM	MT	6	819	94011
181114_M05898_0037_000000000-C59F6_1_11519CG0034L01.fastq.gz.clones.txt	T_CM	N	M4D6	M4D6_CM	MT	6	1365	82032
181114_M05898_0037_000000000-C59F6_1_11519CG0035L01.fastq.gz.clones.txt	T_CM	N	M5D6	M5D6_CM	MT	6	1693	75405
181114_M05898_0037_000000000-C59F6_1_11519CG0036L01.fastq.gz.clones.txt	T_CM	Y	M1D10	M1D10_CM	MT	10	23	8575
181114_M05898_0037_000000000-C59F6_1_11519CG0037L01.fastq.gz.clones.txt	T_CM	N	M2D10	M2D10_CM	MT	10	478	88092
181114_M05898_0037_000000000-C59F6_1_11519CG0038L01.fastq.gz.clones.txt	T_CM	N	M3D10	M3D10_CM	MT	10	1000	72054
181114_M05898_0037_000000000-C59F6_1_11519CG0039L01.fastq.gz.clones.txt	T_CM	N	M4D10	M4D10_CM	MT	10	261	82716
181114_M05898_0037_000000000-C59F6_1_11519CG0040L01.fastq.gz.clones.txt	T_CM	N	M5D10	M5D10_CM	MT	10	204	77895
181114_M05898_0037_000000000-C59F6_1_11519CG0041L01.fastq.gz.clones.txt	T_CM	N	M1D20	M1D20_CM	MT	20	2134	51761
181114_M05898_0037_000000000-C59F6_1_11519CG0042L01.fastq.gz.clones.txt	T_CM	N	M2D20	M2D20_CM	MT	20	1243	51031
181114_M05898_0037_000000000-C59F6_1_11519CG0043L01.fastq.gz.clones.txt	T_CM	N	M3D20	M3D20_CM	MT	20	1667	49448
181114_M05898_0037_000000000-C59F6_1_11519CG0044L01.fastq.gz.clones.txt	T_CM	N	M4D20	M4D20_CM	MT	20	1518	50463
181114_M05898_0037_000000000-C59F6_1_11519CG0045L01.fastq.gz.clones.txt	T_CM	N	M5D20	M5D20_CM	MT	20	714	64565
181114_M05898_0037_000000000-C59F6_1_11519CG0046L01.fastq.gz.clones.txt	T_CM	N	M1D40	M1D40_CM	MT	40	201	52070
181114_M05898_0037_000000000-C59F6_1_11519CG0047L01.fastq.gz.clones.txt	T_CM	N	M2D40	M2D40_CM	MT	40	171	55127
181114_M05898_0037_000000000-C59F6_1_11519CG0048L01.fastq.gz.clones.txt	T_CM	N	M3D40	M3D40_CM	MT	40	1248	51847
181114_M05898_0037_000000000-C59F6_1_11519CG0049L01.fastq.gz.clones.txt	T_CM	N	M4D40	M4D40_CM	MT	40	872	63677
181114_M05898_0037_000000000-C59F6_1_11519CG0050L01.fastq.gz.clones.txt	T_CM	N	M5D40	M5D40_CM	MT	40	1735	50106
181114_M05898_0037_000000000-C59F6_1_11519CG0051L01.fastq.gz.clones.txt	T_CM	N	M1D60	M1D60_CM	MT	60	1370	54091

181114_M05898_0037_000000000-C59F6_1_11519CG0052L01.fastq.gz.clones.txt	T_CM	N	M2D60	M2D60_CM	MT	60	1251	53035
181114_M05898_0037_000000000-C59F6_1_11519CG0053L01.fastq.gz.clones.txt	T_CM	N	M3D60	M3D60_CM	MT	60	565	50359
181114_M05898_0037_000000000-C59F6_1_11519CG0054L01.fastq.gz.clones.txt	T_CM	N	M4D60	M4D60_CM	MT	60	493	79895
181114_M05898_0037_000000000-C59F6_1_11519CG0055L01.fastq.gz.clones.txt	T_CM	N	M5D60	M5D60_CM	MT	60	667	46043

Appendix 3.7: Metadata for PC01 T-effector memory

File	Cell_type	Filtered	Sample_ID	Sample_ID_EM	Infection	Days_pi	Unique_CDR3	Functional TCR
180822_M05898_0017_000000000-C2L4H_1_11438CG0001L01.fastq.gz	T_EM	N	S1D60	S1D60_EM	SBP	60	180	17484
180822_M05898_0017_000000000-C2L4H_1_11438CG0002L01.fastq.gz	T_EM	N	S2D60	S2D60_EM	SBP	60	231	31124
180822_M05898_0017_000000000-C2L4H_1_11438CG0003L01.fastq.gz	T_EM	N	S3D60	S3D60_EM	SBP	60	123	14148
180822_M05898_0017_000000000-C2L4H_1_11438CG0004L01.fastq.gz	T_EM	Y	S4D60	S4D60_EM	SBP	60	25	5066
180822_M05898_0017_000000000-C2L4H_1_11438CG0005L01.fastq.gz	T_EM	N	S5D60	S5D60_EM	SBP	60	175	16053
180822_M05898_0017_000000000-C2L4H_1_11438CG0006L01.fastq.gz	T_EM	N	S1D40	S1D40_EM	SBP	40	1725	36994
180822_M05898_0017_000000000-C2L4H_1_11438CG0007L01.fastq.gz	T_EM	N	S2D40	S2D40_EM	SBP	40	1860	40718
180822_M05898_0017_000000000-C2L4H_1_11438CG0008L01.fastq.gz	T_EM	N	S3D40	S3D40_EM	SBP	40	414	32336
180822_M05898_0017_000000000-C2L4H_1_11438CG0009L01.fastq.gz	T_EM	N	S4D40	S4D40_EM	SBP	40	572	45596
180822_M05898_0017_000000000-C2L4H_1_11438CG0010L01.fastq.gz	T_EM	N	S5D40	S5D40_EM	SBP	40	461	36179
180822_M05898_0017_000000000-C2L4H_1_11438CG0011L01.fastq.gz	T_EM	N	S1D20	S1D20_EM	SBP	20	1414	45082
180822_M05898_0017_000000000-C2L4H_1_11438CG0012L01.fastq.gz	T_EM	Y	S2D20	S2D20_EM	SBP	20	23	7631
180822_M05898_0017_000000000-C2L4H_1_11438CG0013L01.fastq.gz	T_EM	N	S3D20	S3D20_EM	SBP	20	1352	35571
180822_M05898_0017_000000000-C2L4H_1_11438CG0014L01.fastq.gz	T_EM	N	S4D20	S4D20_EM	SBP	20	1279	35643
180822_M05898_0017_000000000-C2L4H_1_11438CG0015L01.fastq.gz	T_EM	Y	S5D20	S5D20_EM	SBP	20	1	1
180822_M05898_0017_000000000-C2L4H_1_11438CG0016L01.fastq.gz	T_EM	N	S1D10	S1D10_EM	SBP	10	186	32320
180822_M05898_0017_000000000-C2L4H_1_11438CG0017L01.fastq.gz	T_EM	Y	S2D10	S2D10_EM	SBP	10	58	2
180822_M05898_0017_000000000-C2L4H_1_11438CG0018L01.fastq.gz	T_EM	N	S3D10	S3D10_EM	SBP	10	184	22714
180822_M05898_0017_000000000-C2L4H_1_11438CG0019L01.fastq.gz	T_EM	N	S5D10	S5D10_EM	SBP	10	145	10363

180822_M05898_0017_000000000-C2L4H_1_11438CG0020L01.fastq.gz	T_EM	N	S1D6	S1D6_EM	SBP	6	850	10852
180822_M05898_0017_000000000-C2L4H_1_11438CG0021L01.fastq.gz	T_EM	N	S2D6	S2D6_EM	SBP	6	330	18337
180822_M05898_0017_000000000-C2L4H_1_11438CG0022L01.fastq.gz	T_EM	N	S3D6	S3D6_EM	SBP	6	641	22546
180822_M05898_0017_000000000-C2L4H_1_11438CG0023L01.fastq.gz	T_EM	N	S4D6	S4D6_EM	SBP	6	456	25110
180822_M05898_0017_000000000-C2L4H_1_11438CG0024L01.fastq.gz	T_EM	N	S5D6	S5D6_EM	SBP	6	233	12649
180822_M05898_0017_000000000-C2L4H_1_11438CG0025L01.fastq.gz	T_EM	Y	U_EM1	U_EM1_EM	Unchallenged	0	106	5312
180822_M05898_0017_000000000-C2L4H_1_11438CG0026L01.fastq.gz	T_EM	N	U_EM2	U_EM2_EM	Unchallenged	0	109	20975
180822_M05898_0017_000000000-C2L4H_1_11438CG0027L01.fastq.gz	T_EM	Y	U_EM3	U_EM3_EM	Unchallenged	0	141	5203
180822_M05898_0017_000000000-C2L4H_1_11438CG0028L01.fastq.gz	T_EM	N	U_EM4	U_EM4_EM	Unchallenged	0	225	12659
180822_M05898_0017_000000000-C2L4H_1_11438CG0029L01.fastq.gz	T_EM	N	U_EM5	U_EM5_EM	Unchallenged	0	399	29931
180822_M05898_0017_000000000-C2L4H_1_11438CG0030L01.fastq.gz	T_EM	N	M1D6	M1D6_EM	MT	6	340	15525
180822_M05898_0017_000000000-C2L4H_1_11438CG0031L01.fastq.gz	T_EM	N	M2D6	M2D6_EM	MT	6	396	18044
180822_M05898_0017_000000000-C2L4H_1_11438CG0032L01.fastq.gz	T_EM	N	M3D6	M3D6_EM	MT	6	950	25644
180822_M05898_0017_000000000-C2L4H_1_11438CG0033L01.fastq.gz	T_EM	N	M4D6	M4D6_EM	MT	6	769	27034
180822_M05898_0017_000000000-C2L4H_1_11438CG0034L01.fastq.gz	T_EM	N	M5D6	M5D6_EM	MT	6	873	27030
180822_M05898_0017_000000000-C2L4H_1_11438CG0035L01.fastq.gz	T_EM	Y	M1D10	M1D10_EM	MT	10	23	3436
180822_M05898_0017_000000000-C2L4H_1_11438CG0036L01.fastq.gz	T_EM	N	M2D10	M2D10_EM	MT	10	435	47270
180822_M05898_0017_000000000-C2L4H_1_11438CG0037L01.fastq.gz	T_EM	N	M3D10	M3D10_EM	MT	10	349	34391
180822_M05898_0017_000000000-C2L4H_1_11438CG0038L01.fastq.gz	T_EM	N	M4D10	M4D10_EM	MT	10	269	47897
180822_M05898_0017_000000000-C2L4H_1_11438CG0039L01.fastq.gz	T_EM	Y	M5D10	M5D10_EM	MT	10	3	3
180822_M05898_0017_000000000-C2L4H_1_11438CG0040L01.fastq.gz	T_EM	N	M1D20	M1D20_EM	MT	20	186	32466
180822_M05898_0017_000000000-C2L4H_1_11438CG0041L01.fastq.gz	T_EM	N	M2D20	M2D20_EM	MT	20	1084	39060
180822_M05898_0017_000000000-C2L4H_1_11438CG0042L01.fastq.gz	T_EM	N	M3D20	M3D20_EM	MT	20	1157	41354
180822_M05898_0017_000000000-C2L4H_1_11438CG0043L01.fastq.gz	T_EM	N	M4D20	M4D20_EM	MT	20	1056	41850
180822_M05898_0017_000000000-C2L4H_1_11438CG0044L01.fastq.gz	T_EM	N	M5D20	M5D20_EM	MT	20	770	47507
180822_M05898_0017_000000000-C2L4H_1_11438CG0045L01.fastq.gz	T_EM	N	M1D40	M1D40_EM	MT	40	176	19146

180822_M05898_0017_000000000-C2L4H_1_11438CG0046L01.fastq.gz	T_EM	N	M2D40	M2D40_EM	MT	40	200	43225
180822_M05898_0017_000000000-C2L4H_1_11438CG0047L01.fastq.gz	T_EM	N	M3D40	M3D40_EM	MT	40	500	39424
180822_M05898_0017_000000000-C2L4H_1_11438CG0048L01.fastq.gz	T_EM	Y	M4D40	M4D40_EM	MT	40	1	1
180822_M05898_0017_000000000-C2L4H_1_11438CG0049L01.fastq.gz	T_EM	N	M5D40	M5D40_EM	MT	40	434	45847
180822_M05898_0017_000000000-C2L4H_1_11438CG0050L01.fastq.gz	T_EM	Y	M1D60	M1D60_EM	MT	60	102	6422
180822_M05898_0017_000000000-C2L4H_1_11438CG0051L01.fastq.gz	T_EM	N	M2D60	M2D60_EM	MT	60	770	32745
180822_M05898_0017_000000000-C2L4H_1_11438CG0052L01.fastq.gz	T_EM	N	M3D60	M3D60_EM	MT	60	320	34378
180822_M05898_0017_000000000-C2L4H_1_11438CG0053L01.fastq.gz	T_EM	N	M4D60	M4D60_EM	MT	60	271	52637
180822_M05898_0017_000000000-C2L4H_1_11438CG0054L01.fastq.gz	T_EM	N	M5D60	M5D60_EM	MT	60	611	42688

Chapter 4: Investigating the role of TRBV3 in the acute phase of a murine malaria infection

This chapter contributes to the following publication in Frontiers in Immunology: N.L. Smith, W. Nahrendorf, C. Sutherland, J.P. Mooney, J. Thompson, P.J. Spence, G.J.M. Cowan. 2020. A Conserved TCRβ signature dominates a highly polyclonal T-cell expansion during the acute phase of a murine malaria infection. Frontiers in Immunology 11: doi 10.3389/fimmu.2020.587756

4.1. Abstract

A conserved TRBV3 encoded TCRβ signature was previously shown to be consistently expanded in the acute phase of a murine *Plasmodium chabaudi* infection, but not re-expanded following re-challenge. To determine the significance and explore the specificity of this conserved response further, publicly available TCR repertoire data from mice challenged with a variety of other pathogens was analysed, and the response was also further profiled using single cell RNA-seq. The conserved TRBV3 encoded signature was not found to be expanded by any other pathogens examined, and single-cell transcriptomics revealed the response to be part of the broad adaptive response against the parasite, rather than a distinct innate cell subtype. Together, these results suggest the conserved response is specific to the *Plasmodium* parasite, and capture of the α/β cognate chain pairings involved now facilitates future experiments to determine the antigenic drivers of this response.

4.2. Introduction

The main objective of work undertaken in the first phase of this project, was to generate and analyse the TCRβ splenic repertoires generated during a *P. chabaudi* infection, to deepen understanding of the early adaptive immune responses elicited against the parasite, and also to determine if a conserved response that could be a focal target for malaria therapeutic purposes was generated. Bulk TCRβ amplicon sequencing of the splenic repertoires and analysis of independent publicly available

RNA-seq data, presented in Chapter 3, indicated a conserved response encoded by TRBV3 was consistently expanded during the acute phase of a first infection. However, splenic memory compartments were not dominated by TRBV3, and appeared to be more private responses to the individual. In agreement, the TRBV3 signature detected was not found to be re-expanded upon homologous re-challenge. As discussed in Chapter 3, this could indicate that either the conserved response is non-specific to the parasite, and/or potentially represents an innate cell population, or that the response is potentially cross-reactive given its' public nature, and lacks the required avidity/specificity to enter the memory pool. However, given that the response is conserved and temporally associated with control of parasitaemia, further investigation of the response is warranted to determine the significance of the response regarding the potential for it to be exploited as a target for malaria therapeutic purposes. Two key questions to address are: 1) is the detected TRBV3 response specific to a *Plasmodium* infection (and if so can we determine the antigenic driver of this response?) and 2) does the conserved response represent a specific immune cell subpopulation?

The first question of specificity to a *Plasmodium* infection itself can be partly addressed through mining bulk TCRβ repertoires elicited by other known immunogens and pathogens in C57Bl/6 mice. Such an approach is possible in these syngeneic mice, as they have highly similar baseline V-J segment usage frequencies (Ndifon *et al.*, 2012; Oftedal *et al.*, 2017; Izraelson *et al.*, 2018). As a result, even data with limited available T-cell counts, such as those mined from RNA-seq data sets, can be used to make clear conclusions concerning the characteristics and similarity of responses between mice (Izraelson *et al.*, 2018).

Although bulk TCRβ repertoire sequencing can reveal TCR clones of interest post-pathogen exposure, it does not provide any information regarding the specific antigenic determinants of these clonotypes. Determining the antigen specificity of a given TCR has long been a goal of the immune-repertoire sequencing community, and although recent advances have led to a surge in new methods, this area still remains challenging (reviewed in (Bradley and Thomas, 2019; Joglekar and Li, 2020)). Three broad approaches have been used to address the questions of specificity and T-cell

antigen discovery: antigen-directed, computational and TCR-directed approaches (Joglekar and Li, 2020). Antigen-directed approaches most commonly use peptide:MHC multimers to isolate specific T-cell clones of interest. This has led to the generation of large data sets of epitope-specific TCR sequences (for example, Klarenbeek *et al.*, 2012; Link *et al.*, 2016; Wolf *et al.*, 2018), now searchable in three principle curated databases – VDJdb (Shugay *et al.*, 2018), iReceptor (Corrie *et al.*, 2018) and McPAS-TCR (Tickotsky *et al.*, 2017). As such databases grow in size, it is thought that eventually predictive models of TCR:pMHC specificity may become possible from sequence data alone. However, currently determining specificity from this approach requires foreknowledge of epitopes and their post-translational modifications to generate appropriate multimers, and suffers from low throughput when considering pathogens with large genomes such as *Plasmodium* (Huang *et al.*, 2020). The sensitivity of this approach is also limited by T-cell promiscuity, with each TCR capable of recognising a multitude of different peptides (reviewed in (Sewell, 2012)). Computational approaches that can predict TCR ligands have been developed, including GLIPH2 (Glanville *et al.*, 2017; Huang *et al.*, 2020), TCRdist (Dash *et al.*, 2017a) and iSMART (H. Zhang *et al.*, 2020). However, without prior p:MHC multimer sorting, each programme only identifies clusters of TCR sequences that are enriched compared to an unchallenged repertoire, that are predicted to bind the same epitope. iSMART is also not currently functional (author correspondence). TCR-directed approaches general employ strategies to screen a particular T-cell clone against a large library of p:MHC complexes. The most promising of these include yeast display libraries (Gee *et al.*, 2018), which can exhibit degenerate libraries of billions of p:MHC complexes and more recently mammalian cell surface display methods (Joglekar *et al.*, 2019; Sharma, Rive and Holt, 2019). Both can be used without the need for priori knowledge of the epitope. However, to pan T-cell clones of interest against such assays requires either isolation of the clones of interest, or construction of transgenic clones which requires knowledge of the TCRα/β chain pairing.

Given the size of the *Plasmodium* genome and vast amount of potential epitopes that may be produced, in Chapter 3 I initially used computational methods to identify TCRs of interest, identifying TRBV3 and conserved clusters OTU1 and OTU2. If antigen-driven, to determine *de novo* what epitopes elicit this response, cell-based display

methods to screen custom libraries present the most promising method of choice. This technology is however not yet widely available and such platforms - although in development in our lab - are not yet functional. Similarly, there is currently a lack of monoclonal agents against TRBV3 which prohibits isolating our clones of interest. As such, a transgenic TCR would be required, necessitating single-cell analysis to obtain paired TCRα/β chain pairings.

Understanding cellular heterogeneity and identifying novel cellular subsets in immunological studies has traditionally relied upon flow cytometry, and more recently mass cytometry. Although mass cytometry has increased the number of cellular markers that can be examined, both methods still only provide a coarse level of cellular resolution, that is vastly superseded by recent single cell RNA-seq technologies that can accurately capture the whole transcriptome of a cell in an unbiased and hypothesis-free manner. The recent advances in single cell RNA-seq technologies have now opened up a myriad of new perspectives for immune receptor analysis (De Simone, Rossetti and Pagani, 2018). Particularly, platforms that combine single cell RNA-seq with V(D)J enrichment allow joint profiling of transcriptome and capture of the TCR or BCR (including obtaining clonal chain pairings) of individual cells (Zheng *et al.*, 2017; Singh *et al.*, 2019). This enables characterisation of the whole cellular phenotype to be linked to the immune receptor, and has already been used to describe the specific clonality of novel immune cell subtypes and uncover insights in to the phenotypes of activated or responding clones in auto-immunity (Penkava *et al.*, 2020), vaccinology (Waickman *et al.*, 2019; A. Minervina *et al.*, 2020) and infectious disease (J. Y. Zhang *et al.*, 2020; Liao *et al.*, 2020; Liu *et al.*, 2020).

Plasmodium studies that combine single cell RNA-seq of immune cell populations with targeted immune receptor capture are currently lacking. Specifically for *P. chabaudi* infections, to date single cell RNA-seq has been used to understand the T_H1 and T_H2 fate bifurcation during acute infection (Lönnberg *et al.*, 2017) and to explore transcriptome dynamics during memory development (Soon *et al.*, 2020). However, both of these studies only analysed cells sorted with a specific transgenic TCR, precluding evaluation of the whole responding repertoire in combination with the

transcriptome. Similarly, Fontana *et al.* (2016) undertook single-cell TCR-seq of γδ T cells at day 21 post-infection, but only with separate bulk whole transcriptome RNA-seq analysis, and to date no *Plasmodium* studies have combined single cell RNA-seq of immune cell populations with combined targeted TCR analysis. Thus, single cell RNA-seq with TCR enrichment provides a novel route by which to explore the phenotypic landscape, in combination with clonality, of immune cells responding to a *Plasmodium* infection.

Here, to address the question of whether the previously seen conserved response is specific to a *P. chabaudi* infection, I firstly undertake analysis of publicly available TCRβ amplicon and bulk RNA-seq data. I then address the second key question of whether the conserved response seen in the acute phase of infection represents a distinct immune cell subtype, by profiling the response using single cell RNA-seq analysis combined with TCR capture.

4.3. Methods

4.3.1. TCR database search

The three principle curated databases of previously annotated TCRβ CDR3 sequences with known antigen specificities – VDJdb (Shugay *et al.*, 2018), iReceptor (Corrie *et al.*, 2018) and McPAS-TCR (Tickotsky *et al.*, 2017) - were searched for the CDR3 sequences present in OTU1 and OTU2, with and without masking with TRBV3 gene usage. The search was performed for the exact CDR3 amino acid sequence in each cluster, as well as searching within a levenshtein distance of 1 and 2 of the seed sequence for each cluster (function available in VDJdb only).

4.3.2. Publicly available TCRβ amplicon data

A literature search for publicly available TCRβ amplicon or RNA-seq data sets from mice challenged with a known non-specific immunogenic stimulus was performed. For comparative purposes, criteria for this was that the mice were wildtype C57Bl/6, and that time-points covered the peak post-challenge phase where TRBV3 expansion was

seen in our data. At time of writing, the only available data set fitting this criteria was from Madi *et al.* (2014), which was downloaded from the NCBI short read archive, accession number SRP042610. This data set consists of 28 TCRβ amplicon libraries, generated from $CD4^+$ splenocytes of C57Bl/6 mice. Of these, 12 samples are from spleens of unchallenged mice. Nine mice were injected IP with a 100 mg 1:1 ratio of Chicken Ovalbumin emulsified in CFA and seven mice with CFA only. Spleens from each challenged group were harvested on days 5 and 14 post-infection. TCRβ libraries in this data set were generated from total RNA extracted from magnetic-sorted $CD4^+$ splenocytes, which was reverse transcribed using a TCR-C specific primer, with cDNA amplified using a set of multiplexed TRBV-specific primers (Ndifon *et al.*, 2012; Madi *et al.*, 2014). This data set was aligned and TCRβ sequences extracted using MiXCR (Bolotin *et al.*, 2015) as per Methods 2.3.2. Only functional TCRβ sequences were included in analyses, and clonotype frequencies were corrected for potential TCRBV PCR amplification biases using the normalisation techniques presented in Ndifon *et al.* (2012). Differences in the mean TRBV3 gene usage of unchallenged and immunised mice were analysed by one-way ANOVA. The mean proportions of OTU1 and OTU2 in between unchallenged and immunised repertoires was also analysed by one-way ANOVA.

4.3.3. Publicly available RNA-seq data

A literature search for publicly available TCRβ amplicon or RNA-seq data sets from mice challenged with known pathogens was performed. For comparative purposes, criteria for use was that the mice were C57Bl/6, and that time-points covered the peak post-challenge phase where TRBV3 expansion was seen in our data. An RNA-seq data set from Singhania *et al.* (2019) was identified as matching this criteria. This data set consisted of RNA-seq data from both lung and whole-blood of wildtype and various knockout C57Bl/6 mice challenged with a variety of pathogens. For the comparative purposes of this analysis, only the whole-blood data from wildtype C57Bl/6 mice was analysed. This data set consisted of mice challenged with 10 separate pathogens examined at the peak of the murine response; *Toxoplasma gondii*, influenza A virus, murine cytomegalovirus, respiratory syncytial virus, *Candida albicans, Listeria monocytogenes*, *Burkholderia pseudomallei*, House dust mite allergen and *P.*

chabaudi. Each separate pathogen challenge included unchallenged wildtype C57Bl/6mice as controls. Raw fastq files of RNA-seq data obtained from whole blood of C57Bl/6 mice infected with a variety of pathogens, were downloaded from the NCBI short read archive (https://www.ncbi.nlm.nih.gov/sra), accession number GSE119856, and TCRβ sequences were aligned and extracted using MiXCR software as per Methods 2.3.2. Normalised trbv3 gene expression values were also downloaded for this data set from Singhania et al. (2019). For each pathogen, the percentage of TRBV3 gene usage between unchallenged controls and challenged mice was analysed using independent t-tests. As 10 comparisons were undertaken, resultant p values were adjusted for multiple testing using Holm-bonferroni correction.

4.3.4. Single cell RNA-seq

4.3.4.1. Mouse infections

Mouse infections were carried out according to Methods 2.1.1. Two C57BL/6 mice were infected with *P. chabaudi* (AS) by intra-peritoneal (IP) injection of ~10^5 parasitised erythrocytes from a mouse previously infected by mosquito transmission (Spence *et al*, 2012). Mice (n=2) were humanely euthanised on day 7 post- infection.

4.3.4.2. Cell sorting & processing

Splenocytes were isolated as per Methods 2.1.3. For this experiment, cells were only stained with CD3 and CD4 APC and CD3 PE/Cy7. Cells were sorted into 100μL 0.04% BSA in PBS, to generate the single-cell suspension required for 10x Genomics sequencing. These suspensions were then taken directly to the IGMM 10x Genomics sequencing facility on ice and processed within an hour of being harvested.

4.3.4.3. Single-cell sequencing, processing and data analysis

Two barcoded cDNA libraries were prepared from sorted samples using the Chromium Single Cell 5' Library Kit v2 (Zheng *et al.*, 2017). Full length V(D)J segments were enriched from amplified cDNA with primers specific to the TCR constant region using the Chromium Single Cell V(D)J Enrichment Kit – Mouse T-Cell. Sequencing was performed using the High-Output v2.5 Kit on a NextSeq 550 platform. Initial QC of the single-cell data was undertaken by Richard Clark (IGMM), and indicated that 90.2% of the data had a quality score >= Q30. All data has been deposited and made publicly available on ArrayExpress, accession E-MTAB-9691.

All further single-cell data processing was undertaken by Catherine Sutherland (University of Edinburgh). Initial processing of sequence files, including mapping of reads to the mouse reference genome (GRCm38), generation of count matrixes and assembly of TCR alpha and beta chains, was carried out using CellRanger 3.1.0. To exclude potential multiplets, poor quality cells or non T-cells, single T-cells were identified by the expression of a single productive beta chain. Barcodes lacking a beta chain or assigned to multiple were excluded, leaving data from 3333 single T-cells (1658 and 1675 from mouse 1 and 2 respectively). Downstream analyses were performed in R using Seurat 3.1.5 (Butler *et al.*, 2018). Genes expressed in fewer than 3 cells, as well as all *Trav/j* and *Trbv/d/j* genes were excluded. Cells expressing fewer than 200, or over 3000 genes and/or more than 5% mitochondrial genes were removed. The filtered matrix was normalised using Seurat's LogNormalize with default parameters and the top 2,000 variable genes were identified using the FindVariableFeatures 'vst' method, before centring and scaling of the matrix. Dimensionality reduction by PCA was carried out and the top 30 principal components were used as input for graph-based clustering. Clusters were visualised by UMAP. A small, poorly *Cd4*-expressing cluster was identified, and these cells were excluded as contaminants. The above normalisation and clustering steps were repeated with the remaining 2976 cells (1491 and 1485 from mouse 1 and 2 respectively). Differential gene expression analysis using the Wilcoxon rank sum test through FindAllMarkers was used to identify marker genes for each cluster.

4.4. Results

4.4.1. TCR database search results

To determine if the CDR3 sequences present in OTU1 and OTU2 had previously annotated antigen-specificities, three principle databases of TCRβ CDR3 sequences with known antigen specificities – VDJdb (Shugay *et al.*, 2018), iReceptor (Corrie *et al.*, 2018) and McPAS-TCR (Tickotsky *et al.*, 2017) - were searched for the CDR3 sequences present in OTU1 and OTU2, with and without masking with TRBV3 gene usage. The search was performed for the exact CDR3 amino acid sequence in each cluster, as well as searching within a levenshtein distance of 1 and 2 of the seed sequence for each cluster. No annotated specificities were found for either cluster as an exact match, or when searching within a levenshtein distance of <=2. An extensive literature search for the sequences in these clusters has also to date not revealed any previously annotated specificities.

4.4.2. Publicly available TCRβ amplicon data

To determine if either TRBV3 or the public CDR3 sequences previously found to be enriched in the acute phase of a *P. chabaudi (AS)* infection have been shown to expand in response to other immunogenic stimuli in C57Bl/6 mice, I analysed publicly available splenic $CD4^+$ TCRβ repertoire data from unchallenged mice and mice immunised with either OVA in CFA or CFA alone. (Madi *et al.*, 2014). CFA is known to be highly immune stimulating (Billiau and Matthys, 2001) and has been shown to induce non-specific by-stander activation of $CD4^+$ T-cells (Van Aalst *et al.*, 2017). Therefore if the TRBV3 enrichment previously identified was a non-specific innate response to inflammation in C57/Bl6 mice, it is expected this would be induced by CFA challenge. As the libraries in this data-set were generated using multiplex-PCR rather than 5'RACE or 5'RACE-UMI, they are quantitatively not directly comparable to my previously generated TCR repertoires in Chapter 3, as method-dependent bias in V gene usage has previously been observed due to use of different primers (Rosati *et al.*, 2017). However, if enriched TRBV gene or clonal usage were present within this data-set between treatment groups, such changes would be detectable.

On average, 59514.08 (±26587, 95% CI) unchallenged and 74693 (±17862, 95% CI) immunised functional TCRβ sequences were extracted from the dataset per sample respectively, representing 59000 unique unchallenged clones and 67765 unique challenged clones. At the V-gene level, TRBV3 was not found to be increased between unchallenged and immunised mice (one-way ANOVA, p=0.32) making up on average 8.83% (±1.37, 95% CI) of unchallenged repertoires, 8.10% (± 2.51, 95% CI) of repertoires challenged with CFA and 7.21% (±1.84, 95% CI) of repertoires challenged with CFA+OVA (Figure 4.1). In agreement with this, although detectable, the proportion of CDR3 sequences from OTU1 and OTU2 present in repertoires did not differ between unchallenged and immunised mice (one-way ANOVA, OTU1; p=0.39, OTU2;p=0.37) indicating the conserved expansion previously detected is not simply an innate-like response to inflammation (Figure 4.2).

Figure 4.1

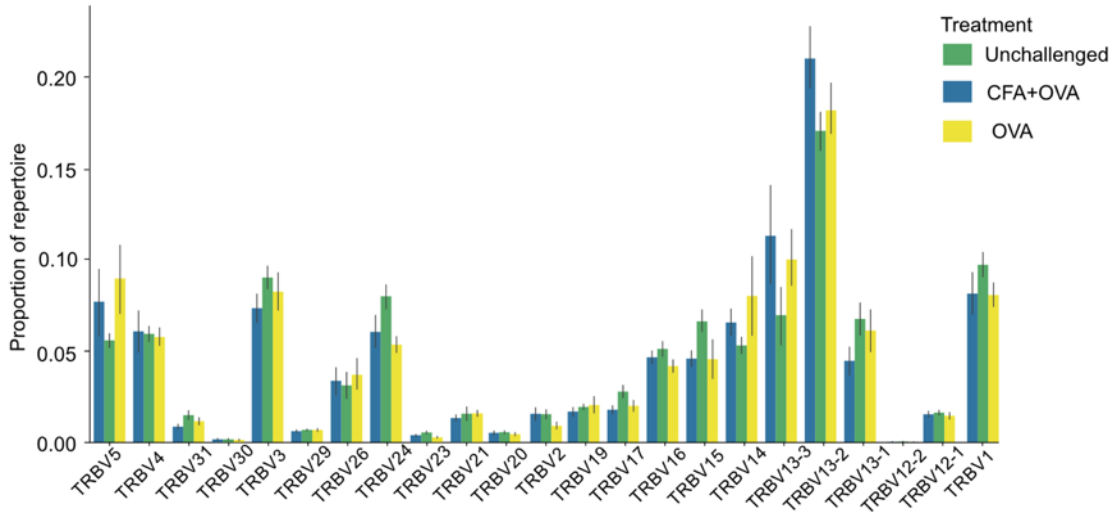


Figure 4.1:TRBV gene usage in unchallenged mice (green), mice immunised with CFA+OVA (blue) and mice immunised with OVA (yellow). Bars represent mean (±SEM) proportion of each TCR repertoire.

Figure 4.2

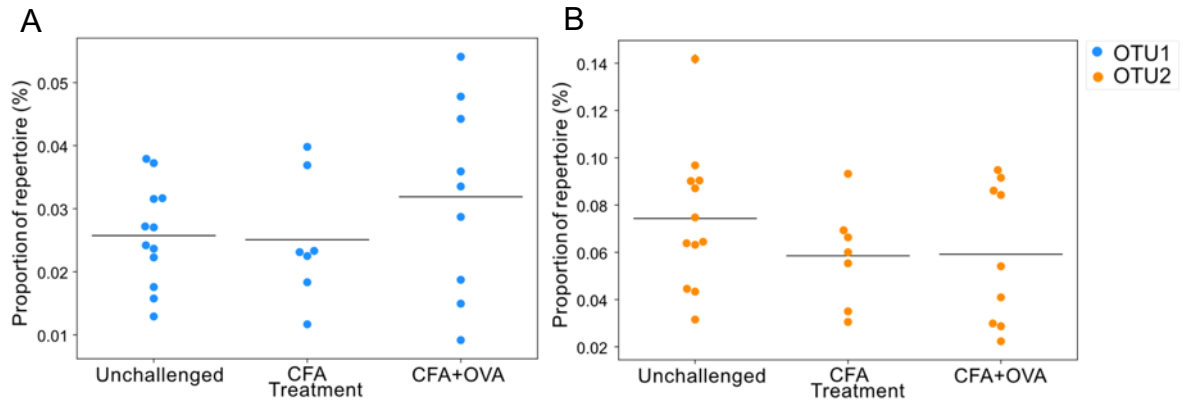


Figure 4.2: A) Proportion of (A) OTU1 and (B) OTU2 sequences present in the CD4+ splenic repertoires of unchallenged mice, and mice immunised with either CFA and OVA or CFA only.

4.4.3. Publicly available RNA-seq data

To determine if either TRBV3 or the public CDR3 sequences previously found to be enriched in the acute phase of a *P. chabaudi (AS)* infection have been shown to

expand in response to other pathogens, I next analysed publicly available RNA-seq data obtained from the whole blood of wildtype C57Bl/6 mice at the peak murine response following separate challenge with 9 different pathogens: *Toxoplasma gondii*, influenza A virus, murine cytomegalovirus, respiratory syncytial virus, *Candida albicans, Listeria monocytogenes*, *Burkholderia pseudomallei* (acute and chronic), House dust mite allergen and *P. chabaudi (AS).* Although it is preferable to mine RNA-seq data from T-cell enriched samples to increase coverage of the TCR loci (Bolotin *et al.*, 2017), meaningful repertoire profiling has been demonstrated in non-enriched tissues (Mandric *et al.*, 2020), particularly from spleen and whole blood samples within which T-cells already make up a significant proportion of the cell population. As the resultant repertoires are however superficially profiled, only enriched clones are likely to be detected and accurately quantified (Bai *et al.*, 2018). However, given the dominance of TRBV3 and clones detected in OTU1 and OTU2 clusters, if a similar response occurred following infection with other pathogens, this proliferation would be expected to be detected in RNA-seq data. As this particular data set also includes a set of mice challenged with *P. chabaudi* (AS), this is able to serve as an in-built positive control for the detection of TRBV3 expansion.

Altogether, this data set consisted of 103 unchallenged mice, and 96 challenged mice. On average, 450.13 (±75.73, 95% CI) and 407.59 (± 78.56, 95% CI) functional TCRβ sequences were extracted from each unchallenged sample and challenged sample respectively, which resulted in a total of 8852 unique unchallenged clones and 8051 unique challenged clones extracted altogether. Comparing each set of challenged repertoires compared to their unchallenged controls, TRBV3 was only found to be significantly increased in mice challenged with *P. chabaudi* (Figure 4.3, Table 4.1), a pattern that was in agreement with the whole-genome normalised Trbv3 expression values as determined by Singhania *et al*. (2019) (Figure 4.4). Searching the challenged data sets for CDR3 sequences from OTU1 and OTU2 , matching clones were found in 5 out of the 12 mice challenged with *P. chabaudi*, and in 1/10 mice challenged with RSV. No sequences present in OTU1 and OTU2 were found in any other challenged repertoires. Of the 103 control mice included in this data set, sequences from OTU1 were found in only 3 of the repertoires, with no sequences from OTU2 found in any unchallenged mice.

Figure 4.3

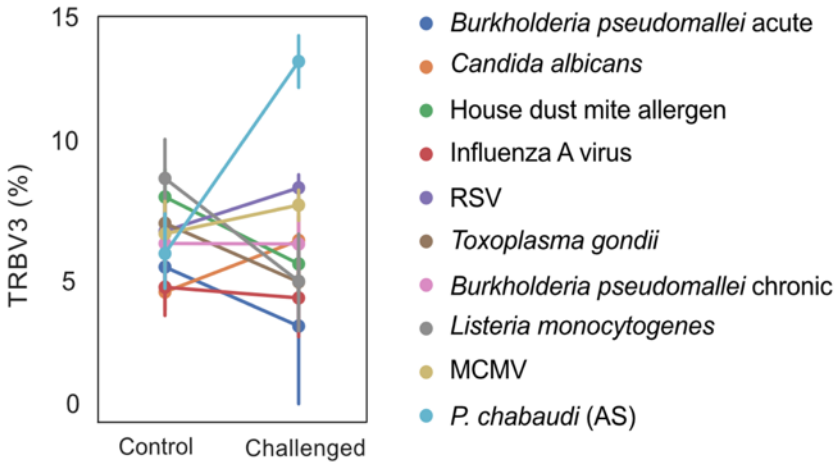


Figure 4.3: Mean TRBV3 gene usage from whole blood repertoires from unchallenged mice and mice challenged with a variety of pathogens, sampled at the peak of the murine response. Minimum n=6 mice for each point. Error bars represent SEM. This data is also presented in Table 4.1.

Table 4.1

Pathogen	Mean TRBV3 (%) Control (± SEM)	Mean TRBV3 (%) Challenged (± SEM)	p. value	p. value adj.
Candida albicans	4.77 (±0.67)	6.55 (±1.53)	0.321	1
B. pseudomallei acute	5.63 (±0.84)	3.59 (±2.62)	0.479	1
B. pseudomallei chronic	6.44 (±1.3)	6.42(±1.0)	0.991	1
House Dust Mite allergen	8.04 (±1.01)	5.74 (±1.03)	0.128	0.771
Influenza A virus	4.93 (±1.05)	4.56 (±1.42)	0.837	1
RSV	6.87 (±0.7)	8.36 (±0.49)	0.101	0.71
Toxoplasma gondii	7.14 (±0.84)	5.12 (±0.49)	0.056	0.505
Listeria monocytogenes	8.68 (±1.44)	5.14 (±1.77)	0.14	0.898
MCMV	6.77 (±1.13)	7.76 (±0.53)	0.442	1
P. chabaudi (AS)	6.09 (±1.37)	12.72 (±0.94)	0.0008	0.008

Table 4.1: Mean (± SEM) TRBV3 gene usage of whole blood repertoires from unchallenged control mice and mice challenged with a variety of pathogens, sampled at the peak of the murine response. Minimum n=6 mice for each group. Differences in mean TRBV3 gene usage between control and challenged groups were analysed using independent t-tests, with adjusted p values calculated using Holm-bonferroni correction for multiple comparisons.

Figure 4.4

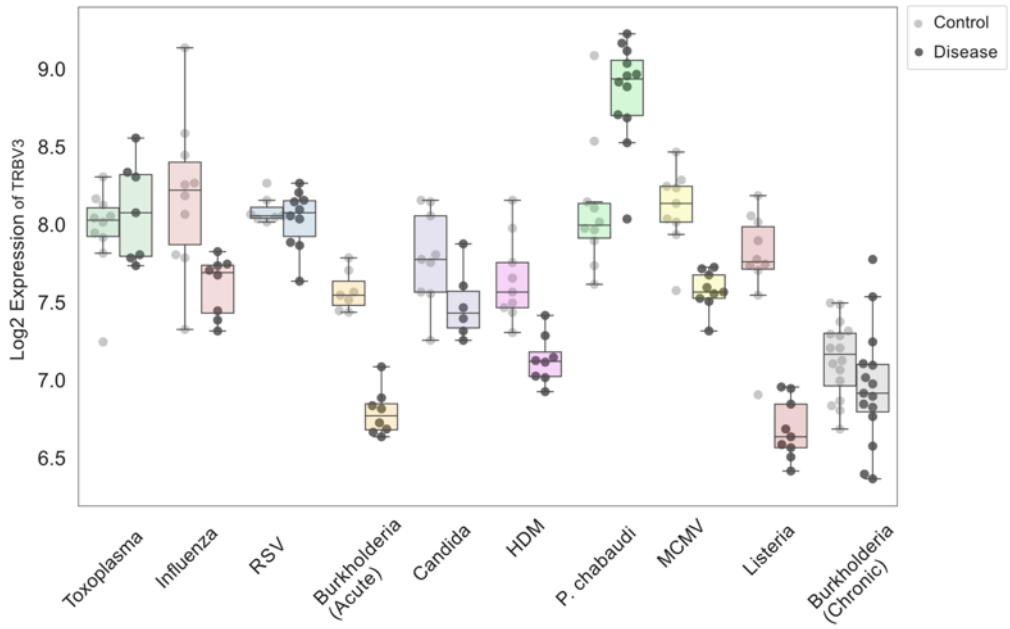


Figure 4.4: Log2 expression values (normalised RNA-seq counts) of TRBV3 from publicly available whole blood RNA-seq data from C57Bl/6 mice, taken at the peak of the murine response for each disease. Data is paired as unchallenged controls (light grey circles) and challenged (black circles). These values for Trbv3 expression were taken directly from Singhania *et al*. (2019).

4.4.4. Single cell analysis

Having found TRBV3 to be repeatedly expanded in the acute phase of a *P. chabaudi* infection, but not evident against other pathogens, I hypothesised that either a specific TRBV3-encoded innate cell population was being expanded in response to the parasite, or the response is part of an antigen-driven response against the parasite. To investigate this further and to determine the phenotype of the conserved TRBV3 response observed during the acute phase of a *P. chabaudi* infection, single-cell RNA sequencing of FACS sorted $CD3^+$ $CD4^+$ splenocytes from two mice at day 7 post-infection with MT parasites was undertaken. These cells were sequenced using the 10x Genomics Chromium platform, using the V(D)J enrichment protocol to obtain paired α/β TCR data for each cell. After quality control steps, expression profiles were obtained for 1491 and 1485 $CD4^+$ single T cells from each mouse respectively (2976 total). Following dimensionality reduction by principal component analysis (PCA), graph-based clustering was undertaken (Butler *et al.*, 2018) and resulting populations visualised using uniform manifold approximation and projection (UMAP) (Figure 4.5A). Seven discrete transcriptional clusters were identified, with cells from both mice evenly

distributed across all dominant clusters (Figure 4.5C). Four of these clusters, denoted as clusters 1, 2, 3 and 4, were classified overall as naïve on the basis of canonical markers (*Sell*, *Il7r*) (Figure 4.5B, Figure 4.6).

Figure 4.5

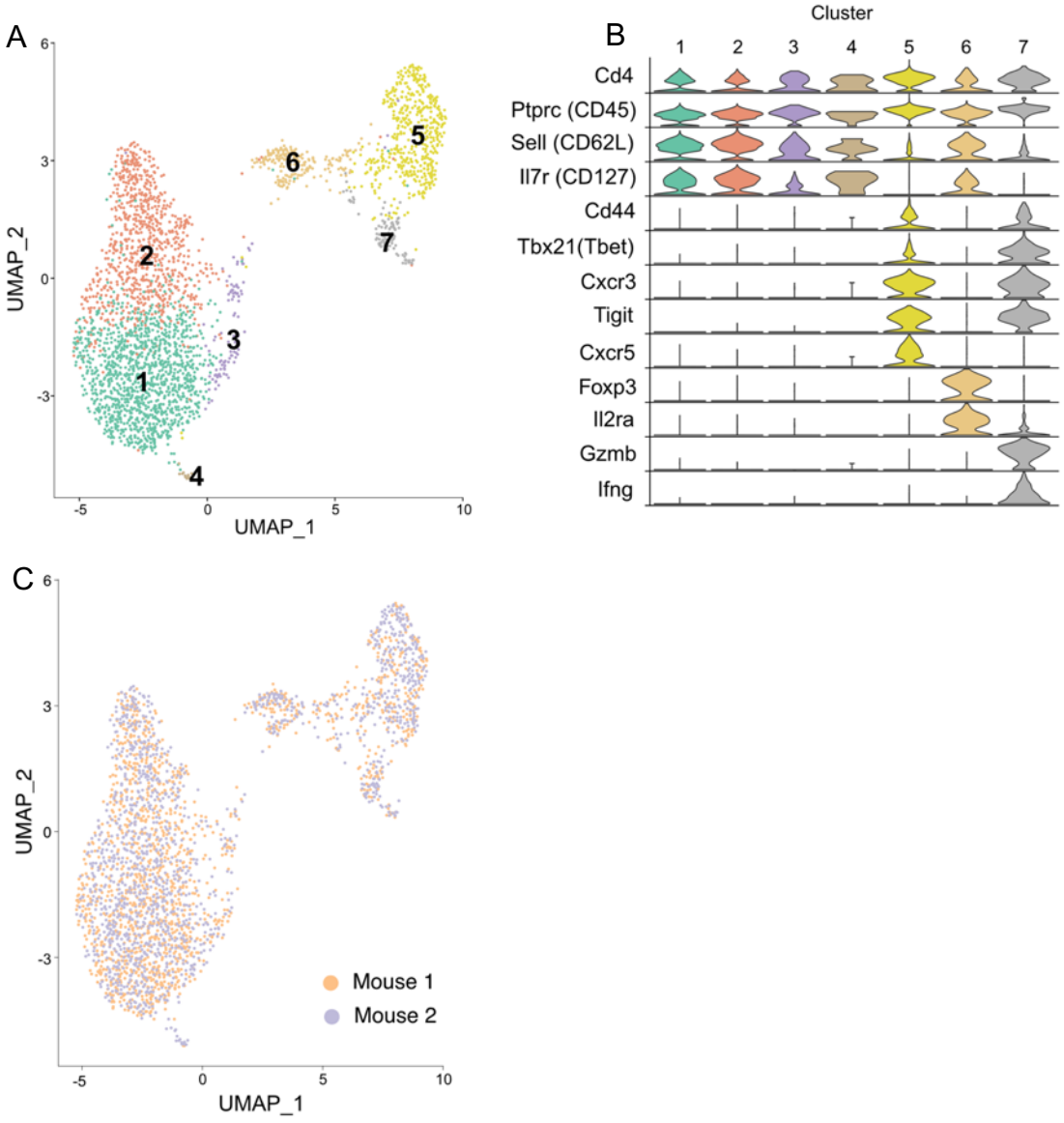


Figure 4.5: Cellular transcriptional phenotypes and paired TCRα and TCRβ sequences were captured using single-cell transcriptomics. A) Clustering of CD4⁺ splenocytes from 2 challenged mice, at day 7 post-infection, according to transcriptional profile similarity. Plot shows individual cell transcriptomes represented in two dimensions by uniform manifold approximation and projection (UMAP). Each point depicts a cell, coloured by Seurat cluster assignment as determined from gene expression profiles. B) Distributions of normalised expression level of major canonical T-cell markers for each cell cluster, displayed as violin plots. C) Two-dimensional UMAP visualisation of CD4+ splenocytes from 2 challenged mice, at day 7 post-infection, as per Seurat cluster assignment, with cells coloured according to each individual mouse.

Figure 4.6

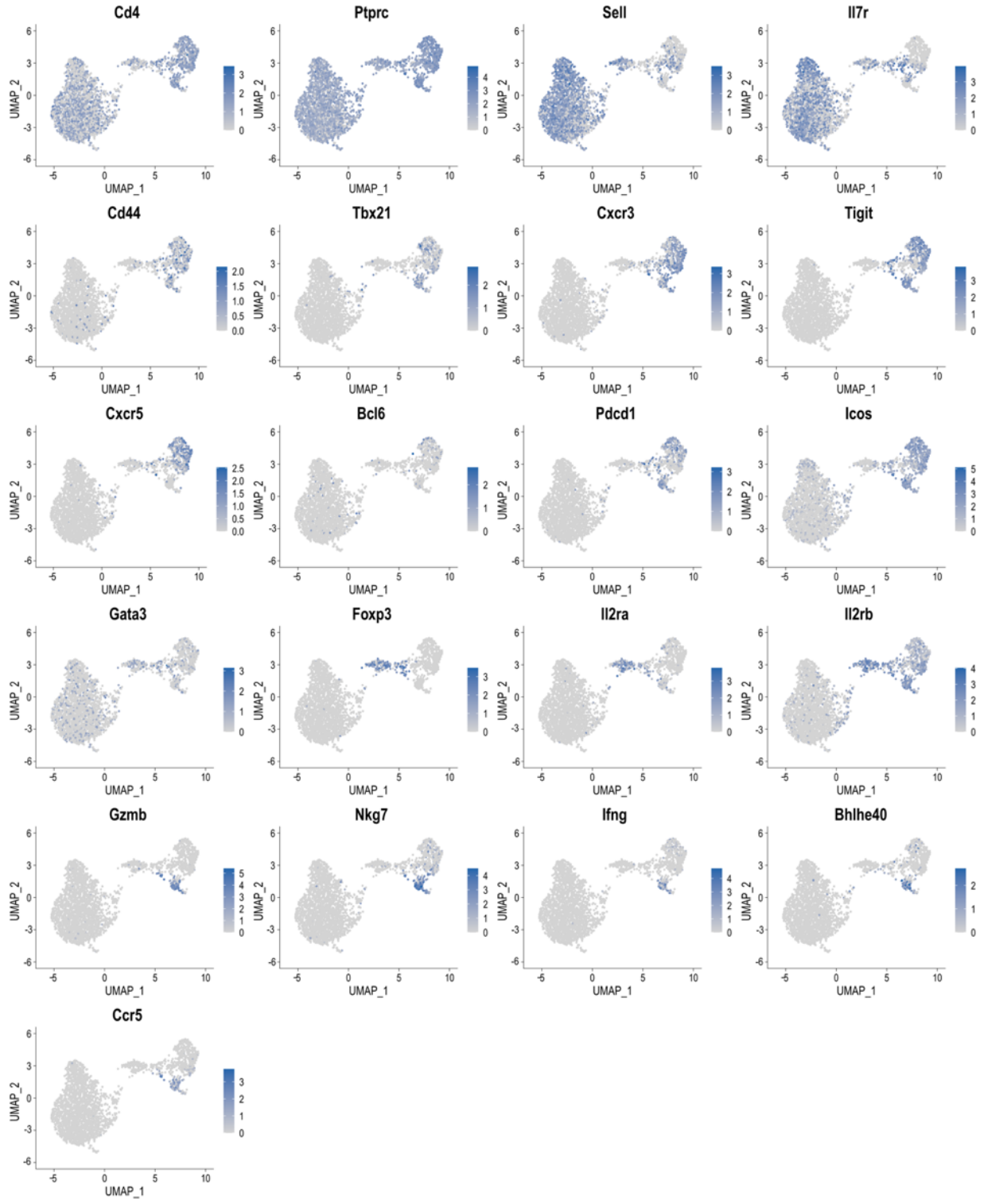


Figure 4.6: Expression of T-cell marker genes of interest from CD4^{+} splenocytes of 2 challenged mice at day 7 post *P. chabaudi* infection; cell positions are as per UMAP in Figure 4.5A.

Differential expression between all seven clusters indicated that clusters 2 and 3 were distinguishable by markers indicative of early T-cell activation, including *cd69*, *Itm2a*, *Zfp36* and *Ier2* for cluster 3, and type I interferon (IFN) response genes (*Gbp2*, *Ifit1*, *Ifit3*, *Isg15*, and *Socs1*) (Elyahu *et al.*, 2019) for cluster 2, whilst cluster 1 expressed a more definitive naïve phenotype (*Ief1*, *Ccr7*) (Figure 4.7). I also identified three clusters discrete from the naïve group: cluster 6 showed a typical Treg transcriptional phenotype with expression of *Foxp3* and *Il2ra*, whilst clusters 5 and 7 were both identified as activated effector populations (*Sell-, Il7r-*). Cluster 7 expressed genes associated with a pro-inflammatory T_H1 signature, including *Tbx21, Ifng, Gzmb, Cxcr6, Ccl5, NKg7* and *Bhlhe40* (Lönnberg *et al.*, 2017). In contrast, cluster 5 was *Ifng-* and appeared to be a more diverse helper population, which predominantly included cells expressing genes associated with a T_{FH} phenotype (*Cxcr5*, *Icos, Bcl6, Il21, Pdcd1)* as well as those with a T_H2 phenotype (*Gata3, Ccr4*) (Figure 4.8).

Figure 4.7

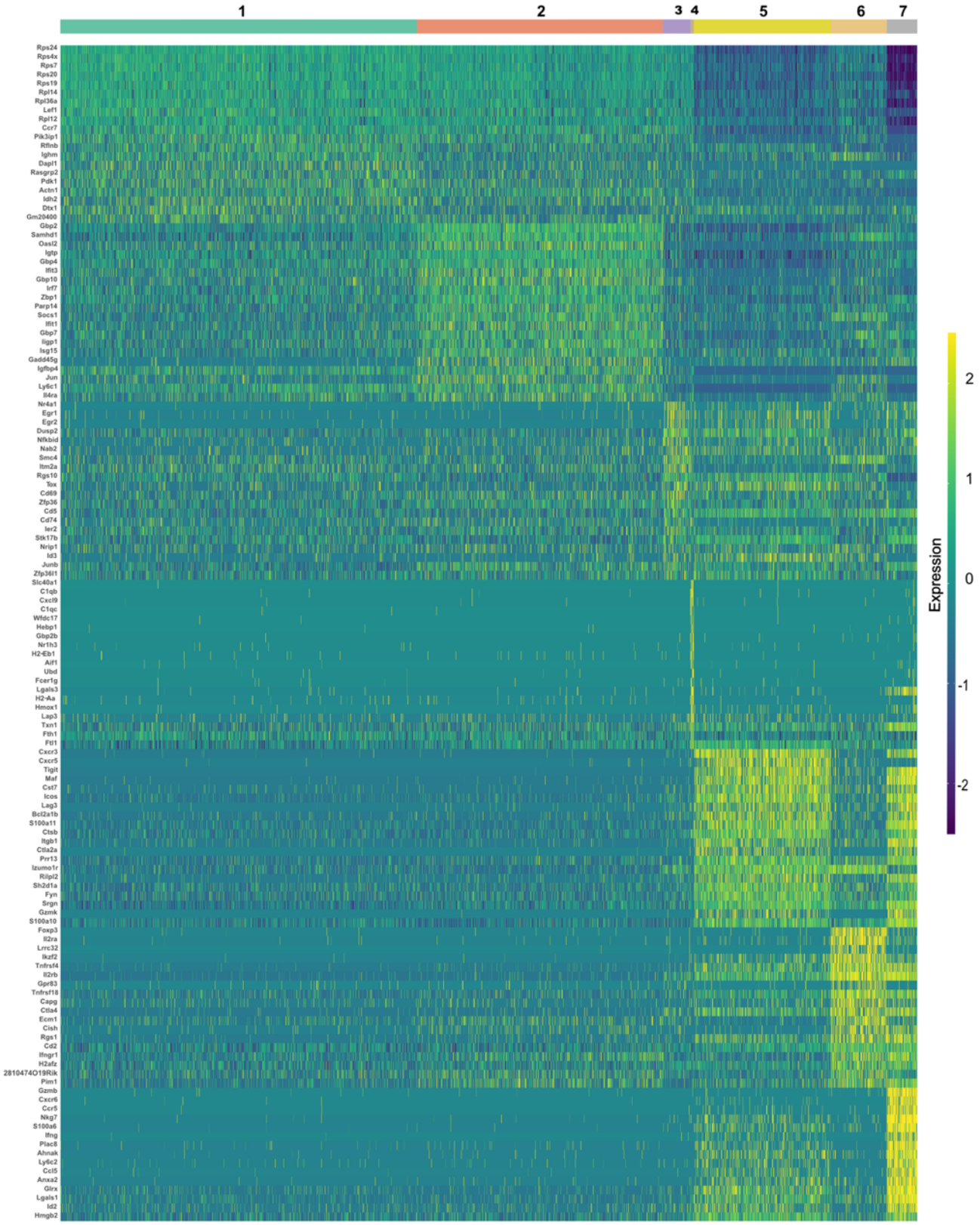


Figure 4.7: Heatmap depicts the top 20 differentially expressed genes and their expression values between all 7 distinct clusters of $CD4^+$ splenocytes of 2 challenged mice at day 7 post *P. chabaudi* infection. Clusters are as defined by Seurat cluster assignment according to transcriptional similarity, and are as described in Figure 4.5.

Figure 4.8

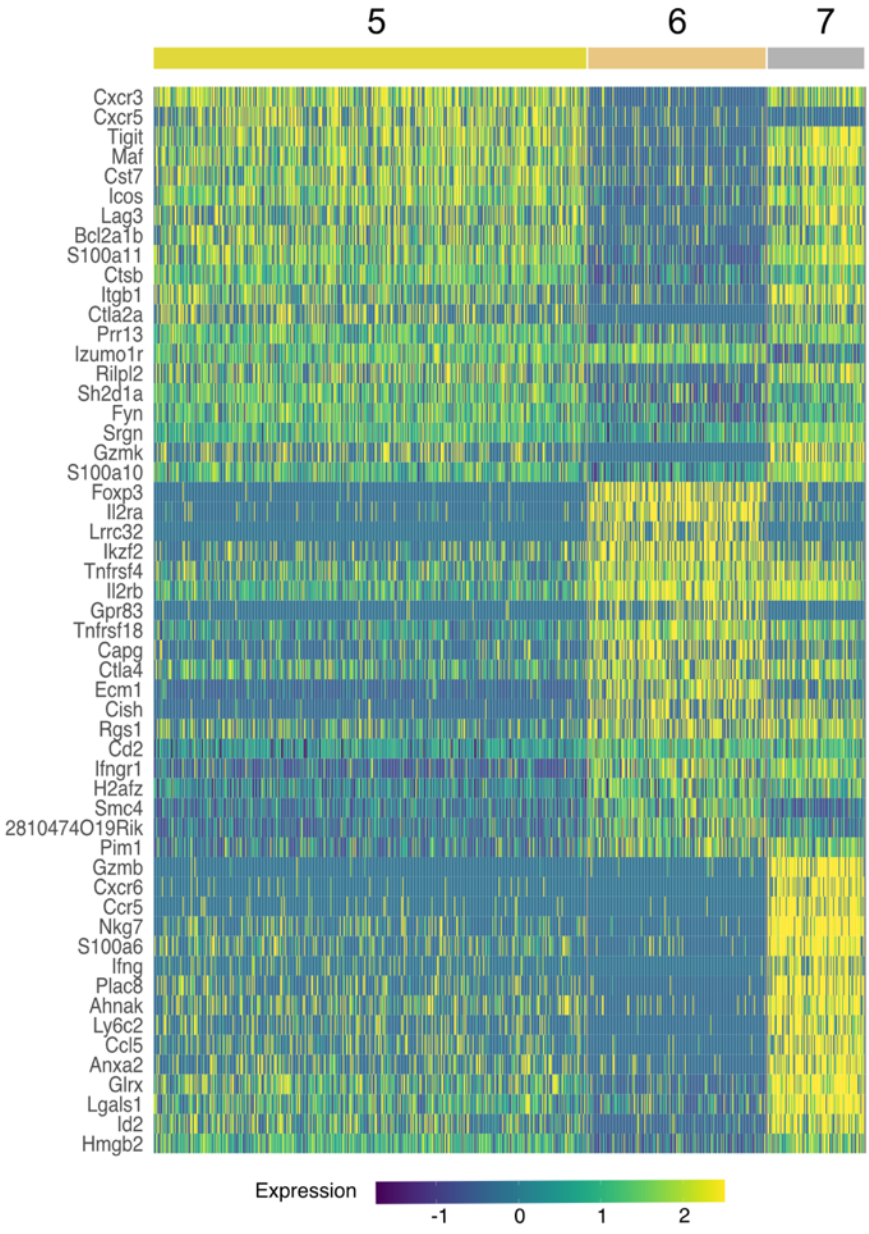


Figure 4.8: Heatmap showing the expression levels of the top 20 differentially expressed genes that define the activated clusters 5, 6 and 7.

Utilising the TCR data, we confirmed that the effector response is highly polyclonal: of the 581 effector cells present in clusters 5 and 7, only 3.6% (21/581) are clonally expanded (>1 cell with identical paired TCRα and TCRβ aa sequence) (Figure 4.9A) and no clone had more than three copies present. We again confirmed that TRBV3 is the most dominant TRBV gene encoding the TCRβ chain of activated effector populations, and show that out of all the clusters, cluster 7 had the highest proportion of TRBV3 (Figure 4.9B). Cells that were TRBV3^{+} did not however form a distinct cluster and instead displayed diverse phenotypes across all clusters. No TRAV gene was over-represented in a particular cluster (Figure 4.10). Of the clonotypes previously identified in the TCRβ bulk data in OTU1 and OTU2, two were present in the single-cell data set, both within the predominantly T_{FH} cluster 5. TCRβ nucleotide sequences displayed diverse probabilities of generation across all clusters (Figure 4.11).

Figure 4.9

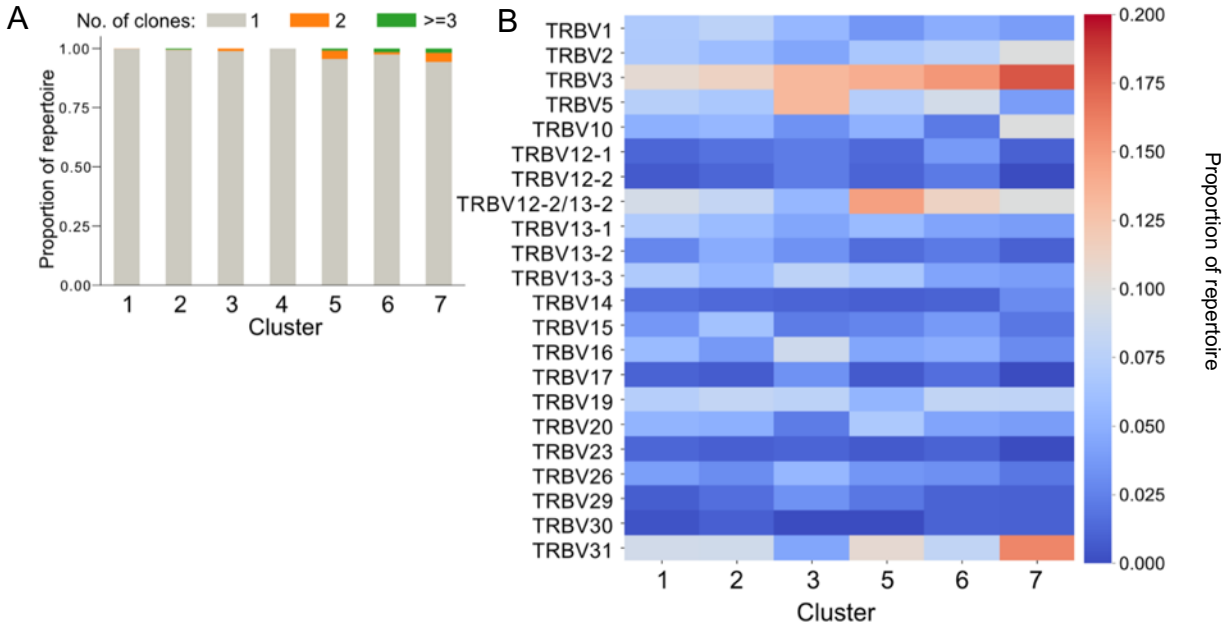


Figure 4.9: A) Stacked bar plot showing the proportion of each cluster that is clonally expanded (>1 cell with identical paired TCRα and TCRβ aa sequences). Repertoires of all clusters were highly polyclonal, with a minor increase in frequency of expanded clones in activated clusters. B) Heatmap displays proportional TRBV gene usage per cluster TCR repertoire.

Figure 4.10

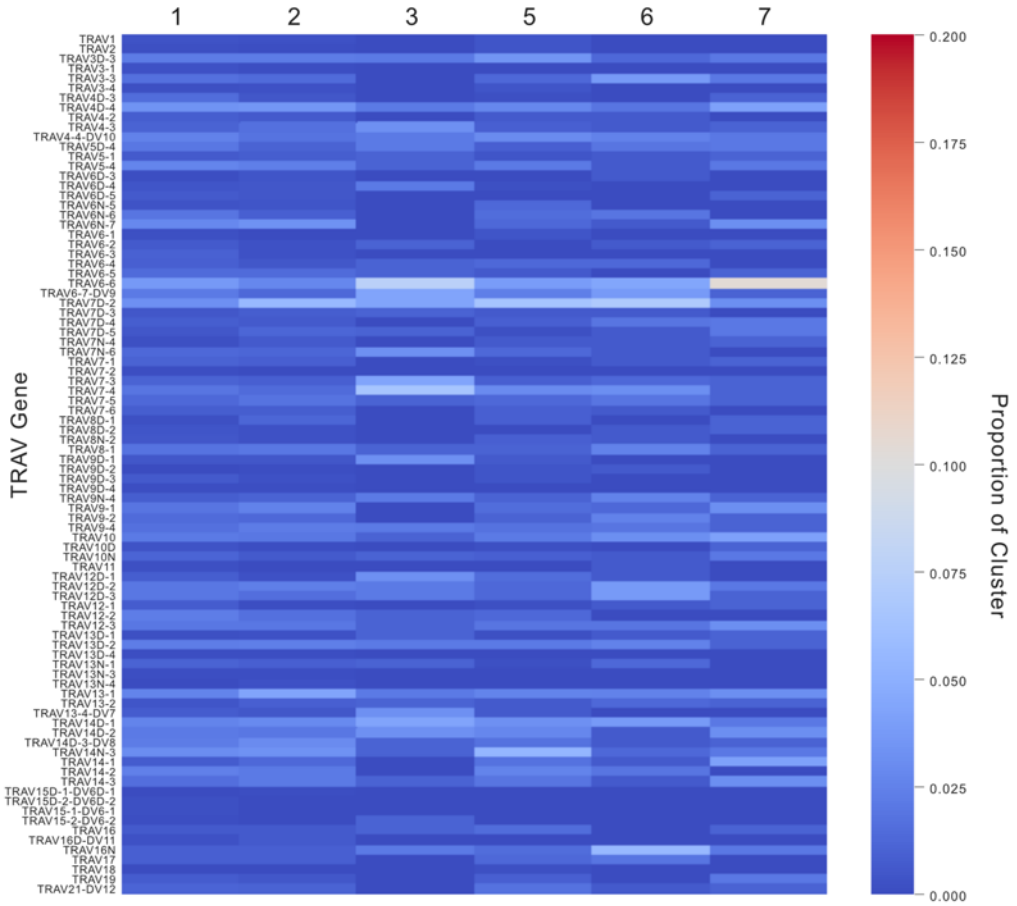


Figure 4.10: Heatmap displays the mean proportion of TRAV gene (rows) usage of each cluster 1 to 7 (columns) from the repertoires of both mice used in the single-cell experiment.

Figure 4.11

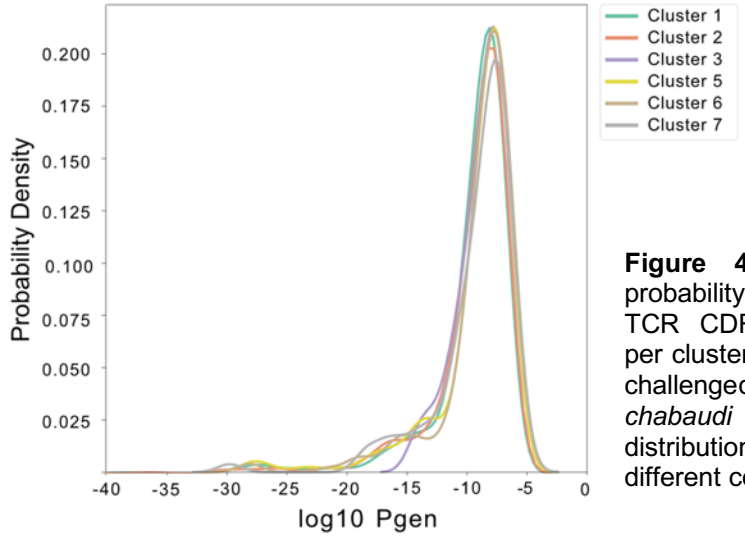


Figure 4.11: KDE plot of the probability of generation (log10) of TCR CDR3 nucleotide sequences per cluster, of CD4⁺ splenocytes of 2 challenged mice at day 7 post *P. chabaudi* infection. Each cluster distribution is represented by a different coloured line.

4.5. Discussion

Here, I have analysed publicly available TCRβ amplicon and RNA-seq data, and also undertaken single cell RNA-seq to deepen our understanding of the role of TRBV3 in the acute phase of a *P. chabaudi* infection. Firstly, to determine if the TRBV3 expansion seen was a non-specific response to the inflammation induced during a malaria infection, I analysed the TCRβ repertoires of $CD4^+$ splenocytes from C57Bl/6 mice immunised with CFA or OVA in CFA, with CFA being a known inflammatory trigger. However, in agreement with the authors original analysis (Madi *et al.*, 2014), no increased sharing between unchallenged and immunised mice was found, and neither TRBV3 gene usage, or proportion of clones present in OTU1 or OTU2 were found to be altered between treatment groups.

Secondly, to begin addressing the question of whether the conserved response seen is specific to a *P. chabaudi* infection, firstly the three principal TCR sequence databases (Tickotsky *et al.*, 2017; Corrie *et al.*, 2018; Shugay *et al.*, 2018) were searched for previous annotations, but to date no known epitope specificities to the sequences in OTU1 and OTU2 have been recorded in these databases. I next examined RNA-seq data generated from whole blood of C57Bl/6 mice challenged with a variety of pathogens. TRBV3 was not found to be expanded at the peak of infection for any of the pathogens examined from Singhania *et al.* (2019), except for the mice challenged with *P. chabaudi.* A major caveat of mining untargeted RNA-seq data for TCRβ sequences, is that the repertoire is only superficially profiled due to low coverage of the TCR loci (Bolotin *et al.*, 2017; Bai *et al.*, 2018). This is reflected in the much lower number of TCRβ sequences that were extracted from each sample compared to amplicon data, however as this data-set included mice challenged with *P. chabaudi (AS),* this was able to serve as an in-built positive control for detection of the expansion of TRBV3. For *P. chabaudi* challenged mice in this data set, TRBV3 was found to encode 12.72% of challenged repertoires, compared to 6.09% of unchallenged repertoires. A value of 12.72% at day 6 post-infection is lower than what was detected in the bulk TCRβ repertoires generated in Chapter 3, however the parasites used in this challenge were serially blood-passaged rather than recently

mosquito-transmitted, and in my previous analyses of SBP infections, TRBV3 was not found to peak until day 10 post-infection: a value of 12.72% is in agreement with the mean of 12.22% TRBV3 gene usage found at day 6 post-infection in SBP splenic repertoires in experiment PC01 from Chapter 3. Similarly, OTU1 was only found in 5/12 (41.67%) of the mice challenged with *P. chabaudi,* however in experiment PC01, OTU1 was not found to peak in SBP infections until day 10 post-infection, and this coupled with the repertoire only being superficially profiled, likely accounts for the reduced level of conservation of OTU1 between mice in this data-set. That OTU1 was only present in 3/103 (2.9%) of unchallenged mice, further confirms that although these sequences have a high probability of generation, they are not dominant in naïve repertoires.

Mining publicly available data is not an exhaustive approach by which to address the question of specificity of the conserved response to *P. chabaudi*. It is limited by what data-sets are publicly available, what pathogens these include, and further narrowed by requiring mice with the same syngeneic background in order to be comparable (Izraelson *et al.*, 2018). Similarly, curated epitope-specific TCR databases are still in their relative infancy, and all three principal databases examined are currently dominated by human TCR sequences. However, having found the conserved response repeatedly in the acute phase of a *P. chabaudi* infection, including in data generated by a variety of different methods, and to not have found this response following other challenges examined, strongly suggests the response is specific to a first infection with the parasite. TCRβ amplicon and RNA-seq data sets following infection with other *Plasmodium* species are not currently available, but as these are generated, the specificity of this response to the parasite and whether a similar response is elicited in a first infection by other *Plasmodium* species including *P. falciparum* will become evident. Eventually, as antigen-discovery techniques become more widely available, a distinct parasite peptide driving the response may ultimately be uncovered.

The ability to undertake such TCR-directed antigen-discovery investigations for *P. chabaudi* in the future is greatly strengthened by having undertaken the single cell

RNA-seq study, as this has provided us with key TCRα/β chain pairings. Knowledge of these is required for cell-based display antigen discovery methods, which currently represent the preferred method of choice for T-cell deorphanisation of complex pathogens. The single cell RNA-seq data also confirms two of the major findings from Chapter 3: firstly, that the $CD4^+$ effector response is highly polyclonal, and secondly that TRBV3 is the most dominant V-gene used in activated effector populations. It also reveals that $TRBV3^+$ cells in the acute phase of a first infection have diverse phenotypes. I can therefore conclude that the conserved TRBV3 response seen does not represent a discrete innate cell population but is instead part of an adaptive immune process against the parasite. The transcriptional phenotypes of activated effector cells are in agreement with those previously observed in the acute phase of a *P. chabaudi* infection (Lönnberg *et al.*, 2017) and indicate a large cluster of predominantly follicular helper cells as well as a distinct *Ifng*$^+$ T_H1 population. Both types of response have been shown to arise simultaneously during the acute phase of an infection (Lönnberg *et al.*, 2017) and to be essential in controlling blood-stage *P. chabaudi*; T_H1 responses are required for initial control of acute parasitaemia (Su and Stevenson, 2000; Soon and Haque, 2018) and T_{FH} are crucial for generating antibody-mediated immunity and controlling chronic infection (Pérez-Mazliah *et al.*, 2015, 2017). The presence of OTU1 and OTU2 in a predominantly T_{FH} cluster, suggests that they may play a role in guiding the developing humoral response against the parasite. A single naïve $CD4^+$ T-cell has been demonstrated to be able to give rise to clones with different cell fates (Lönnberg *et al.*, 2017) so, as only one of each clonotype was captured with the single cell sequencing depth used, I cannot firmly conclude that the transcriptional phenotypes identified here would broadly reflect the conserved TCRs identified in the bulk data. Nevertheless, the TCR of a cell has been shown to impart a strong preference for either a T_H1 or a T_{FH} fate, with longer dwell time between peptide-MHC:TCR biasing towards T_{FH} and GC-T_{FH} responses (Tubo *et al.*, 2013). Therefore, if driven by the same epitope, one would expect the conserved response to have a similar phenotype (Schattgen *et al.*, 2020).

The single cell RNA-seq data also indicates that PGen does not determine T-cell fate, in agreement with Sethna *et al* (2018) who demonstrated that the ability of a TCR to respond to a particular epitope, was not strongly correlated with its generation

probability. This adds further credence to the finding in Chapter 3; that although sequences in OTU1 and OTU2 have a high PGen, they are not expanded in effector repertoires simply because of this.

In summary, analyses of publicly available data presented here, together with the single cell RNA-seq work undertaken, strongly suggest that the conserved T-cell responses demonstrated in the acute phase of a *P. chabaudi* infection, are specific to a first infection with the parasite. Determination of the precise antigenic driver of this response will require cell-based display libraries to cater for potential de novo peptide discovery, presenting a future line of discovery. Ultimately, the relevance of these findings to malaria therapeutics is dependent on either such a peptide being conserved between *Plasmodium* species, or a similar conservation of response being detected in humans following a *Plasmodium* infection. Both present exciting avenues for future investigation.

Chapter 5: Longitudinal sequencing of the peripheral TCRβ repertoires elicited during a primary and secondary *Plasmodium falciparum* controlled human malaria infection

5.1. Abstract

T-cells are known to play key roles in the developing immune response to a *Plasmodium falciparum* infection, with both $CD4^+$, cytotoxic $CD8^+$ and γδ cells attributed functional roles. Aberrant T-cell responses have conversely been implicated in the pathogenesis of severe falciparum malaria. During CHMI, T-cells are recruited to the spleen, the site of the developing immune response, before being released back in to circulation post drug-treatment. Whether these kinetics represent movement of pathogen-specific clonally expanded populations has not previously been described. Here, using longitudinal PBMC samples from both a primary and secondary controlled human malaria infection model, I analysed the TCRβ repertoires generated over the time-course of a blood-stage *P. falciparum* CHMI. Overall, no clonally expanded populations were evident in response to either a first or second infection. However, the non-specific recruitment of established T-cell clones from the periphery was evident, a dynamic repeated in homologous re-challenge infections. Understanding the consequences of this non-specific trafficking, and whether or not it shapes an individual's response to a *Plasmodium* infection, warrants further investigation.

5.2. Introduction

Plasmodium falciparum is the causative agent of falciparum malaria, a disease responsible for ~200 million cases of malaria, with ~386,000 deaths in 2019 (WHO, 2020). Control measures are currently reliant on a combined toolbox of vector control measures and anti-malarial drug treatment. The most advanced vaccine candidate, Mosquirix®, has low efficacy (RTSS Clinical Trials Partnership, 2015), and although

more promising candidates are on the horizon (Datoo *et al.*, 2021; Minassian *et al.*, 2021), major challenges to immune-mediated elimination of malaria still remain. This is in part due to gaps in knowledge regarding how both clinical and anti-parasite immunity develops. Unlike murine models of malaria, where sterilising immunity can often be induced after a single infection, for *P. falciparum* anti-parasite immunity takes years of repeat exposure to develop. Clinical immunity to severe disease can however be acquired after only a more limited number of infections in endemic settings, independent of control of parasitaemia (Snow *et al.*, 1998; Gonçalves *et al.*, 2014). Both $\alpha\beta$ and $\gamma\delta$ T-cells have been attributed protective functional roles in the response to a first *P. falciparum* infection (reviewed in (Kurup, Butler and Harty, 2019)), but collateral damage to the host from aberrant effector T-cells including excessive proinflammatory cytokine responses have also been reported in acute infection and implicated in severe disease (Crompton *et al.*, 2014; Mandala *et al.*, 2017; Ghazanfari, Mueller and Heath, 2018).

Using the murine *P. chabaudi* model, in previous chapters I have shown that the splenic $CD4^+$ expansion in response to a first infection is highly polyclonal, though a conserved pathogen-specific response could be detected (Smith *et al.*, 2020). For $\gamma\delta$ cells, single-cell TCR sequencing has also demonstrated specific $\gamma\delta$ clones responding to a first *P. chabaudi* infection (Mamedov *et al.*, 2018). In falciparum malaria, TCR sequencing studies are still very limited. Mpina *et al*., (2017) reconstructed MAIT cell $\alpha\beta$ TCR sequences mined from RNA-seq data from individuals undergoing CHMI with sterile *P. falciparum* sporozoites, and found no significant changes in repertoires through time despite proliferation, arguing that the expansion was thus a cytokine rather than antigen driven response. In agreement, the PfEMP1 binding domain, CIDR-1α, has been shown to stimulate $CD4^+$ T-cells non-specifically through TCR-independent pathways (Ndungu *et al.*, 2006), and Treg proliferation during an infection has been shown to be able to be induced in an antigen non-specific manner (Scholzen *et al.*, 2009). To date, however, no published studies that use TCR repertoire sequencing to examine the clonality of the major adaptive T-cell responses to a *P. falciparum* infection have been identified. To begin to address this knowledge gap, in this Chapter I make use of access to longitudinal PBMC

samples from both a primary and secondary homologous re-challenge blood-stage *P. falciparum* controlled human malaria infection (CHMI), to profile the longitudinal TCRβ repertoires elicited over the time-course of both a first and second infection.

The wider study of which the human PBMC samples used in this chapter were obtained, have used systems immunology tools to investigate how immunity to severe disease – proposed to be a form of disease tolerance – can develop after only the first few infections of life (Muñoz Sandoval et al., (In Prep)). To date, the T-cell mass cytometry and whole-blood RNA-seq analyses from this wider study has indicated cross-sectionally that T-cell activation is much reduced between a first and third infection (Muñoz Sandoval et al., (In Prep)). It is proposed that cross-reactive or bystander activated non-specific T-cells may be responsible for the aberrant activation of T-cell subsets seen in a first infection, and that the reduction in activation of these non-specific responses is responsible for the reduction in severe disease following repeat exposure (Muñoz Sandoval et al., (In Prep)). To more fully explore and confirm this hypothesis, an understanding of the dynamics and clonality of the TCR repertoire longitudinally through re-challenge infections is required, to demonstrate if movement of non-specific clones occurs and how this may evolve during a re-challenge infection. As with the *P. chabaudi* model, determining if shared specific TCR signatures are generated between individuals, will also provide information regarding whether or not a convergent pathogen-specific TCR-mediated response against the parasite is elicited, that may drive discovery of immunodominant epitopes.

Previous studies and data from our collaborators indicate, in agreement with field clinical data, that marked lymphopenia occurs during CHMI (De Mast et al., 2008; Milne et al., 2021; Muñoz Sandoval et al., 2021 (In Prep); Van Wolfswinkel et al., 2017). The transient nature of this lymphopenia with recovery to baseline levels soon after drug treatment supports the theory that T-cells are recruited out of peripheral circulation, presumably to the spleen which is the main site for the developing immune response and elimination of parasitised red blood cells (del Portillo *et al.*, 2012b). As longitudinal human splenic samples are not obtainable, measurable T-cells that

constitute a response to the parasite, are those clones that are recruited out of circulation, as well as those that re-appear and may be enriched in the periphery post drug-treatment, when activated populations are released back into circulation (Edwards *et al.*, 2018; Bach *et al.*, 2021). The clonality and TCR repertoire composition during these kinetics has not previously been explored. Therefore, in this chapter, longitudinal CHMI TCRβ sequencing was undertaken, with the following key objectives:

- To analyse changes in TCRβ repertoire diversity induced longitudinally over the time-course of a primary and secondary *P. falciparum* CHMI
    - specifically, to determine if a clonally expanded population is detectable at post-treatment timepoints in response to the parasite and if clonality differs between first and second infection
- To characterise the TCRβ repertoire of T-cells recruited from the circulation
    - specifically, to determine if clones recruited from the circulation bear hallmarks of a non-specific response and whether or not this varies between a first and second *P. falciparum* CHMI
- To ascertain if distinct pathogen-specific clone trajectories within an individual are evident, and if these trajectories are shared between individuals
- To determine if any convergent TCRβ responses are generated between individuals

5.3. Methods

5.3.1. Study background and ethical approvals

Human PBMC samples utilised in this thesis were obtained as part of a collaboration with Dr Phil Spence (University of Edinburgh), as part of a wider collaboration with Dr Simon Draper at the Jenner Institute, University of Oxford. Samples were obtained from the clinal trial: “A Phase I/IIa clinical trial to assess the safety, immunogenicity and efficacy of the blood-stage *P. falciparum* malaria vaccine candidate RH5.1/AS01”, ClinicalTrials.gov Identifier: NCT02927145, VAC063 study. This trial was designed

and conducted by teams led by Dr Simon Draper and Dr Angela Minassian of the Jenner Institute, University of Oxford, and samples were obtained between November 2017 and June 2018 at the Centre for Clinical Vaccinology and Tropical Medicine (CCVTM) in Oxford. Ethical approval for this study was granted by the United Kingdom National Health Service Research Ethics Service (Oxfordshire Research Ethics Committee A, reference 16/SC/0345) and was conducted in full concordance with the Declaration of Helsinki 2008.

5.3.2. Volunteers

The above clinical trial was designed to assess the vaccine candidate RH5.1/AS01. As part of this, a group of volunteers who underwent both a primary and secondary CHMI, but did not receive the vaccine, were included as a control group. These two CHMI trials were categorised as VAC63A (primary CHMI) and VAC63B (secondary CHMI) respectively. As the objective of this thesis was to examine TCRβ repertoires following a malaria infection in unvaccinated individuals, only samples that were available to us from a sub-group of individuals within this control group were utilised. This comprised of 7 individuals, who underwent a secondary homologous challenge 4 months following their primary infection, inside a framework to study boosting of the vaccine candidate. All recruited volunteers were between 20-50 years old and were healthy malaria naïve males or non-pregnant females. Exclusion criteria during volunteer recruitment included no history of any previous malaria infection, seropositivity for HBV, HIV or HCV, recent (within 30 days) malaria chemoprophylaxis, G6PD deficiency or any known haemoglobinopathy. Detailed volunteer information is presented in Appendix 5.7, with HLA type presented in Appendix 5.8.

5.3.3. Volunteer infections

Volunteers underwent a blood stage infection only. The inoculum used for this came from cryopreserved vials of 3D7 *Plasmodium falciparum* infected erythrocytes kept at the QIMR in Brisbane, Australia (Cheng *et al.*, 1997). All infections were undertaken with a homologous 3D7 *P. falciparum* recently transmitted line, which came from a

single volunteer previously infected by mosquito bite, cryopreserved after 3 asexual blood cycles following liver egress (Cheng *et al.*, 1997). The inoculum was administered to volunteers in 5mL 0.9% saline intravenously at the CCVTM in Oxford, at an estimated dose of 1,000 parasitised erythrocytes per volunteer. Following the initiations of infections for each trial, a limiting dilution assay was used to determine viability of each inoculum, indicating a viability of 452 ring stage infected erythrocytes in the first infection and 857 in the second infection. Blood was collected daily from individuals starting from the day following first challenge (c+1), to monitor parasitaemia. Volunteers were diagnosed and treated if they were positive by thick blood smear, and/or if they had a parasite density greater than 5000 parasites per mL as quantified by qPCR and/or 1000 parasites per mL and were symptomatic. Volunteers were then treated with either Riamet® or Malarone®, dependent on any drug use contraindications. All volunteers successfully cleared the infection following treatment.

5.3.4. PBMC sample availability

PBMC samples consisted of longitudinal samples, taken from 7 individuals, undergoing a primary (VAC63A) and secondary (VAC63B) *P. falciparum* CHMI. Sample time-points available from these individuals were day before infection (c-1), days three, five and seven post infection (c+3, c+5, c+7), day of diagnosis, 28 days post-challenge (c+28) and 90 days post challenge (c+90). These time-points were available for both primary and secondary infections.

5.3.5. Blood count monitoring

Full blood counts were performed at the Churchill or John Radcliffe Hospitals, Oxford, by NHS staff. For these, blood samples were taken at baseline (c-1), c+6, day of diagnosis, and at a post-treatment timepoint (either c+28, c+45 and/or c+90). This count data included cell counts per mL of blood for lymphocytes, eosinophils, monocytes and neutrophils.

5.3.6. FACS analysis for VAC63B

Based on observations in VAC63A, for the secondary challenge (VAC63B), a finer precision of T-cell subset compartment dynamics was undertaken using FACS analysis at timepoints c-1 and c+9, as per Methods 2.2.4.2.

5.3.7. TCR amplification protocol optimisation

As human TCR sequencing had not previously been undertaken in our lab, I firstly adapted and optimised the human TCRβ amplification protocol published by Mamedov *et al.*, (2013). As with the murine TCRβ protocol presented in Chapter 3, this protocol was a 5-RACE-UMI technique, utilising a 12bp UMI. Following initial PCR optimisation, to ensure the resulting amplicon was indeed a complete TCRβ product, the protocol was tested on Jurkat cells. The Jurkat cell line has a clonal TCR, originally derived from a patient with T-cell leukaemia, and is thus suitable for Sanger sequencing. A volunteer PBMC sample was also amplified and sanger sequenced for control purposes. To test our custom sequence primer strategy before undertaking a NovaSeq run, amplified products from all volunteers were sequenced on a MiSeq platform.

5.3.8. Bulk TCRβ repertoire library prep

Bulk TCRβ library preps were carried out according to methods presented in Methods 2.5.2. Given the complexity of the library prep protocol, samples were run in batches of 12, with sample IDs randomly assigned to each batch. In brief, following RNA extraction, a modified protocol from Mamedov *et al.* (2013) was used to synthesise cDNA and amplify the TCRβ region. All PCR amplifications were performed on the same day as RNA extractions and cDNA synthesis, according to the optimised protocol presented in 2.2.5, and amplified products were visualised on a gel, with expected band size ~570 bp. Sample 96 produced no visible band, and unfortunately all samples in Batch 7 did not produce the expected dominant band seen with other samples. For these 13 samples, cDNA synthesis and amplification were repeated. However, this did not improve band appearance, indicating samples on this day likely

had poor quality RNA-extraction and based on clonotype numbers obtained may need to be excluded. All amplified products were stored at -80°C before being pooled and purified by extraction from excised agarose gel bands.

5.3.9. Bulk TCRβ sequencing

Asymmetric 400bp+100bp sequencing was performed initially on an Illumina MiSeq nano run to assess primer design, before being sequenced on an Illumina NovaSeq platform to ensure adequate depth of sequencing. Custom read primers were used, presented in Methods table 2.4. A diagram to demonstrate final amplicon construct and sequencing read location is presented in Figure 5.1. Library quality and concentration was assessed by Nanodrop prior to submission to Edinburgh Genomics. Edinburgh Genomics provided further quality control and fragment size analysis using a Bioanalyzer, as well as assessing read quality post-sequencing using inbuilt NovaSeq software. Quality was further assessed after receiving the raw fastq files, using FastQC (Andrews, 2010).

Figure 5.1

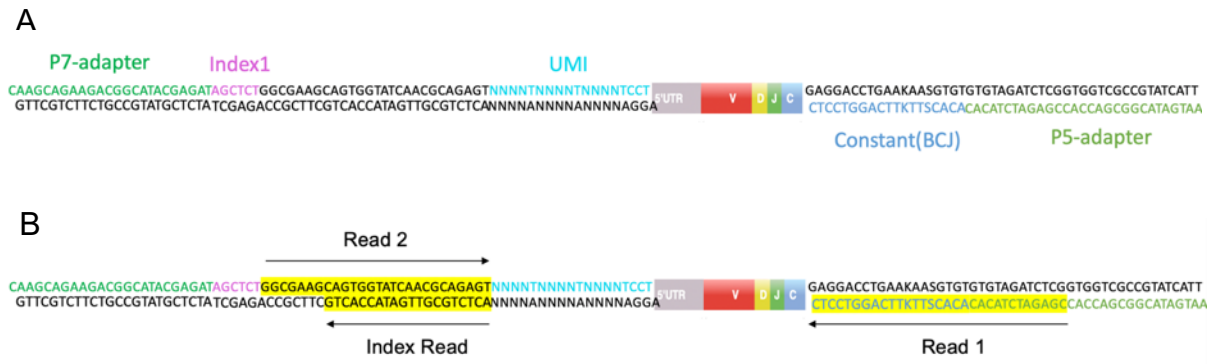


Figure 5.1: A) TCRβ amplicon construct with location of P7 and P5 adapter, constant region primer, 6bp index and 12bp UMI indicated by colour. B) TCR amplicon construct with location of Read 1, Read 2 and Index sequencing reads highlighted.

5.3.10. Raw data processing

Sample demultiplexing and adapter trimming was performed by Edinburgh Genomics. Raw repertoire sequencing data was further demultiplexed and collapsed to the UMI level using MiGEC (Shugay *et al.*, 2014). UMI-collapsed sequences were then aligned using MiXCR software (Bolotin *et al.*, 2015), utilising IMGT nomenclature (Lefranc, 1999), using the MiXCR `analyze amplicon` command. Only in-frame functional CDR3 sequences, as determined by MiXCR alignment, were included in analyses.

5.3.11. Bulk TCR repertoire analyses

PCA was performed on TRBV gene usage frequency using the *scikit-learn* python package (Pedregosa *et al.*, 2011), with proportional TRBV gene usage transformed using the *scikit-bio* central log ratio transformation for compositional data for this. Generalised Renyi entropy was used to create a diversity profile for each repertoire (Renyi, 1961; Rempala and Seweryn, 2013; Greiff *et al.*, 2015). Renyi's entropy is defined as:

$$H_\alpha = \frac{\log \sum_{i=1}^{s} p_i^\alpha}{1-\alpha}$$

Where p_i is the frequency of each clone, and s the total number of unique clones in the repertoire. The parameter α can be varied from 0 to ∞ to generate a diversity profile. At $\alpha = 0$, the entropy will be $\log s$ – simply a count of the number of clones present, corresponding to species richness of the repertoire. α = 1 cannot be defined for Renyi entropy, and instead is calculated as Shannon's entropy (Greiff *et al.*, 2015; Simnica *et al.*, 2019). α = 2 corresponds to the inverse of Simpson's diversity index. As α is increased, greater weight is applied to more frequent clones, and as $\alpha \to \infty$, the measure tends to only consider the proportional number of sequences in the largest clonotype. Diversity of repertoires through time was analysed using the Renyi entropy at various weightings of α, from 0 to 10, as it was noted that, diversity had

levelled off by α = 10, as observed in Greiff *et al.* (2015). Using the *lme4* package (Bates *et al.*, 2015) in R, linear mixed effects models were fit using the log-transformed Renyi entropy as the response variable, for α = 0, α = 1, α = 2 or α = 10, with timepoint as a fixed categorical effect and individual as a random categorical effect. Separate models were fit for a first and second infections, and key timepoints of c-1, diagnosis, c+28 and c+90 were included for each infection. Post-hoc pairwise analyses was carried out using the *emmeans* package (Lenth, 2021) in R adjusted for multiple testing using the Benjamini-Hochberg correction. An FDR < 0.05 was considered significant.

Probability of generation (PGen) of TCRβ NT sequences was calculated using OLGA software (Sethna *et al.*, 2018), and differences in log10 transformed PGen between persistent and non-persistent clones was analysed using Wilcoxon rank test, adjusted for multiple-testing using Benjamini-Hochberg correction, with FDR <0.05 considered significant.

To query if persistent clones had previously annotated epitope specificities, the VDJmatch software was used to query the VDJdb (Shugay *et al.*, 2018), with matching HLA-type and V/J gene usage, allowing for 1 mismatch at the CDR3 aa sequence level.

For GLIPH2 analysis, pooled c-1 repertoires from all individuals were used as the input reference. Repertoires from all individuals at day of diagnosis, c+28 and c+90 were analysed against pooled c-1 separately (by both timepoint and infection), to determine if there was enrichment of shared responses at specific timepoints, and then overlap analysis of significant clusters between timepoints was undertaken. As the algorithm's own reference set is made up of naïve T-cells, the OVE parameter (fold change) was altered from the default of 10 to 1, so small changes in already expanded cells could be detected.

The Jaccard index was used as a similarity metric between repertoires. This ranges from 0 to 1, with 1 representing complete overlap and 0 meaning no overlap at all. This was calculated on UMI-size matched data, unweighted by clonotype frequency. Changes in similarity through time were analysed using linear mixed effect models, with Jaccard index as the response variable, with timepoint as a fixed categorical effect and individual as a random categorical effect. Separate models were fitted for a first and second infection, and key timepoints of c-1, diagnosis, c+28 and c+90 were included for each infection. Post-hoc pairwise analyses was carried out using the *emmeans* package (Lenth, 2021) in R, adjusted for multiple testing using the Benjamini-Hochberg correction. An FDR < 0.05 was considered significant.

To identify clones with active longitudinal dynamics, trajectory analysis was undertaken, based on methods used in (Minervina et al., 2020a; Minervina et al., 2020b; Schultheiß et al., 2020). Proportions of the top 1000 clones from the key timepoints (c-1, diagnosis, c+28 and c+90) of both infections, were normalised by dividing by the maximum proportion of each clone longitudinally, and a euclidean distance matrix created between these normalised proportions. Unsupervised hierarchical clustering using the *SciPy* library (Virtanen *et al.*, 2020) was then undertaken, with average linkage using the euclidean distances between trajectories.

5.3.12. Whole blood RNA-seq analyses

As part of the wider CHMI trial, a tertiary re-challenge was also undertaken in some individuals. This tertiary trial was Vac63C, from which PBMC samples were not available. However, longitudinal whole blood RNA-seq analysis of all control volunteers from Vac63A, Vac63B and Vac63C had previously been undertaken (Muñoz Sandoval et al., (In Prep)). This RNA-seq data set was mined for TCR sequences using MiXCR software (Bolotin *et al.*, 2015), using the `analyze shotgun` command with default settings. Simpson's diversity and inverse Shannon's entropy was calculated using the `CalcDiversityStats` command from VDJTools (Shugay *et al.*,

2015), taking the mean value from recursively down-sampling to 200 unique clones in each repetition.

5.4. Results

5.4.1. TCR amplification protocol optimisation

Sanger sequencing of the amplified product from jurkat cells and from human PBMC sample was undertaken. The jurkat cell sample was split and run as 2 technical replicates. Sequences from both the jurkat cells and PBMCs were aligned using IMGT V-QUEST Both of the jurkat cell TCRβ amplicons showed 100% sequence identity for human TCRβ TRBV12-3*01 TRBD1*01 TRBJ1-2*01, with CDR3 amino acid sequence CASSFSTCSANYGYTF (Appendix 5.1A), consistent with known jurkat cell TCRβ sequence (Giudicelli *et al.*, 2006). The control PBMC sample yielded no alignment result on IMGT V-quest, as would be expected from sanger sequencing of a diverse amplicon product. Comparison of our anticipated amplicon construct with the resulting amplicon (Appendix 5.1B, C), also confirmed expected assembly of the amplified product.

5.4.2. MiSeq nano QC

To determine if the optimised TCRβ amplification protocol generated diverse TCRβ products, and to test the novel custom sequencing reads before undertaking an Illumina NovaSeq run, libraries generated from all volunteers and timepoints were run on a MiSeq nano platform. Pooling the libraries from all 96 samples, this run generated on average 2223.68 (±87.68 SEM) unique functional TCRβ clones from an average of 2977.51 (± 89.81 SEM) aligned reads per sample. This shallow sequencing depth is in line with what is expected from a MiSeq nano run. The low depth of a MiSeq nano run precludes use of UMIs (which require a high level of coverage for precision), but I was able to determine from this run that on average 90.62% (±0.29 SEM) of reads generated per library aligned to the human TCRβ locus with good quality coverage of the UMI and CDR3 region (Appendix 5.2), and that the libraries contained diverse TRBV and TRBJ genes as would be expected (Appendix 5.3). TRBV20-1, TRBV5-1 and TRBV7-9 were the most dominant TRBV genes present in these libraries,

consistent with other published human TCRβ studies utilising a range of sequencing techniques in both healthy and diseased individuals (Freeman *et al.*, 2009; Shugay *et al.*, 2015; Heather *et al.*, 2016; Farmanbar, Kneller and Firouzi, 2019; Servaas *et al.*, 2021). A strong positive correlation between read depth and number of unique clones was present, as would be expected for highly diverse libraries sequenced at a low depth, however the 12 samples from batch 7 identified as poor-quality RNA extractions were obvious outliers from this trend, suggesting the cDNA synthesis and amplification of these samples has indeed resulted in non-representative libraries (Appendix 5.4).

5.4.3. NovaSeq QC

As the libraries and sequencing protocol generated the expected diverse TCRβ repertoire data, these libraries were then sequenced on an Illumina NovaSeq platform. This was undertaken due to the much greater sequencing depth available in NovaSeq, allowing us to more deeply profile the repertoires given the large number of libraries (96 altogether). Quality scores for base calling in NovaSeq differ from other Illumina platforms, in that the quality scores are binned for more efficient storage. Thus, only 4 possible scores are available – 2, 12, 23 and 37 – corresponding to no-call, marginal quality (~Q15), medium quality (~Q20) and high-quality (>Q30) base calls respectively (https://emea.illumina.com/content/dam/illumina-marketing/documents/products/appnotes/novaseq-hiseq-q30-app-note-770-2017-010.pdf). FastQC analysis of Read 1 indicated high-quality coverage over the VDJ junction. For Read 2, however, poor quality scores were present at positions 5, 10, and from 15 to ~20. These positions correspond exactly to constant bases in the UMI region of the SMARTNNN primer (NNNNUNNNNUNNNNUCTTrGrGrG, constant bases highlighted), which do not contribute to the UMI itself, which is made up of the 12 'N' bases only. The bases which do cover the UMI had high-quality (>Q30) base calls. Poor quality only at constant base positions suggests there was a lack of diversity at these positions, indicating the level of PhiX spike-in (10%) used in the NovaSeq run was not high enough to produce enough diversity for high quality base calls at conserved base positions. Discussions with personnel from the Illumina development team confirmed this to be the most likely cause. However, as the UMI

region itself is extremely diverse and subsequently had high quality scores, this does not affect UMI quality or demultiplexing at the UMI level.

5.4.4. Clonotype counts and sample exclusions

On average, 17681.9 (±1212.6 SEM) unique functional TCRβ sequences were obtained per sample following stringent demultiplexing at the UMI level (Appendix 5.6). As expected, sample 96 and samples from batch 7 had low clonotype counts and were thus excluded (Appendix 5.6), resulting in 83 samples being included in all further analyses.

5.4.5. CHMI characteristics

The 7 volunteers used in the bulk TCR-seq study were all successfully infected in both their primary and secondary infection, and developed patent parasitaemia during the first 13 days following challenge, varying slightly in their day of diagnosis (Table 5.1). Parasite growth curves for all individuals in the control-arm of the overall study did not vary between first, second or third infection (Muñoz Sandoval et al., (In Prep)), with no decrease in maximum parasite density in an individual following homologous re-challenge.

Table 5.1

Volunteer	Day of diagnosis (first infection)	Day of diagnosis (second infection)
v1039_815	9	9
v1061_822	10	9.5
v1065_824	10.5	10.5
v1067_812	9	9.5
v1068_819	8.5	9
v1075_823	12.5	13.5
v6032_802	9	9

5.4.6. Blood counts and T-cell subset dynamics

In all 7 volunteers, marked lymphopenia was evident at day of diagnosis in both a primary and secondary infection. This was largely resolved by c+28 (Figure 5.2). The only exception to this was volunteer v1067_812 who during a first infection showed a mild lymphocytosis at day of diagnosis. This overall pattern of lymphopenia at day of diagnosis before resolution post-treatment, was present in all the control-arm participants of VAC63A, VAC63B and VAC63C (comprising all individuals, not just those in my subset of 7 volunteers). This included those undergoing a tertiary infection, with degree of lymphopenia similar at day of diagnosis independent of infection number (Appendix 5.9, Appendix 5.10). Lymphopenia is a hallmark of clinical malaria (Hviid *et al.*, 1997; Kotepui *et al.*, 2020) and has commonly been documented in previous CHMI volunteers at day of diagnosis (De Mast *et al.*, 2008; Van Wolfswinkel *et al.*, 2017; Milne *et al.*, 2021), though not all individuals undergoing primary CHMI with *P. falciparum* show this 'classical' response (Milne *et al.*, 2021). Lymphopenia is thought to be a result of recruitment away from circulation presumably to the spleen**.** More precise T-cell subset FACS data was available for the 7 volunteers in a secondary infection for baseline (c-1.2) and during the acute infection (c+9.2). These subsets were defined as $CD4^+$ or $CD4^-$ naïve T-cells ($CD45RA^+$ $CCR7^+$), $CD4^+$ or $CD4^-$ central memory T-cells ($CD45RA^-$ $CCR7^+$), $CD4^+$ or $CD4^-$ effector memory ($CD45RA^-$ $CCR7^-$) and $CD4^+$ or $CD4^-$ effector T cells ($CD45RA^+$ $CCR7^-$) (Figure 5.3). Timepoint c+9.2 is close to day of diagnosis for most of the volunteers (Table 5.1), except for volunteer v1075_823, who was not diagnosed until day 13.5 in their second infection. They are the only volunteer in a second infection to not have an absolute decrease in $CD3^+$ cells at c+9.2 (Figure 5.3, Figure 5.4), suggesting this volunteer may not have yet developed the lymphopenia seen at day of diagnosis. Overall, this subset dynamics data firstly indicates a large amount of variability between volunteers in terms of what subset made up the largest component of the peripheral T-cell compartment, with some volunteer's having large naïve cell populations, and others dominated by memory populations (Figure 5.4). Secondly, it demonstrates that the observed lymphopenia is not the result of 1 particular T-cell subset being lost from the peripheral circulation, but rather global recruitment of the adaptive T-cell subsets, as recently observed in CHMI with *P. vivax* (Bach *et al.*, 2021). Given the heterogeneity in subset dominance along with pan-recruitment of T-cell compartments, as only 1 vial of PBMCs was available per volunteer at each timepoint, it was decided that the whole

peripheral TCRβ repertoire would be sequenced, rather than sorting individual T-cell subsets. These T-cell dynamics confirm it is clones that are recruited out of circulation at day of diagnosis, as well as those that re-appear and may be enriched in the periphery post drug-treatment, that are the populations of interest.

Figure 5.2

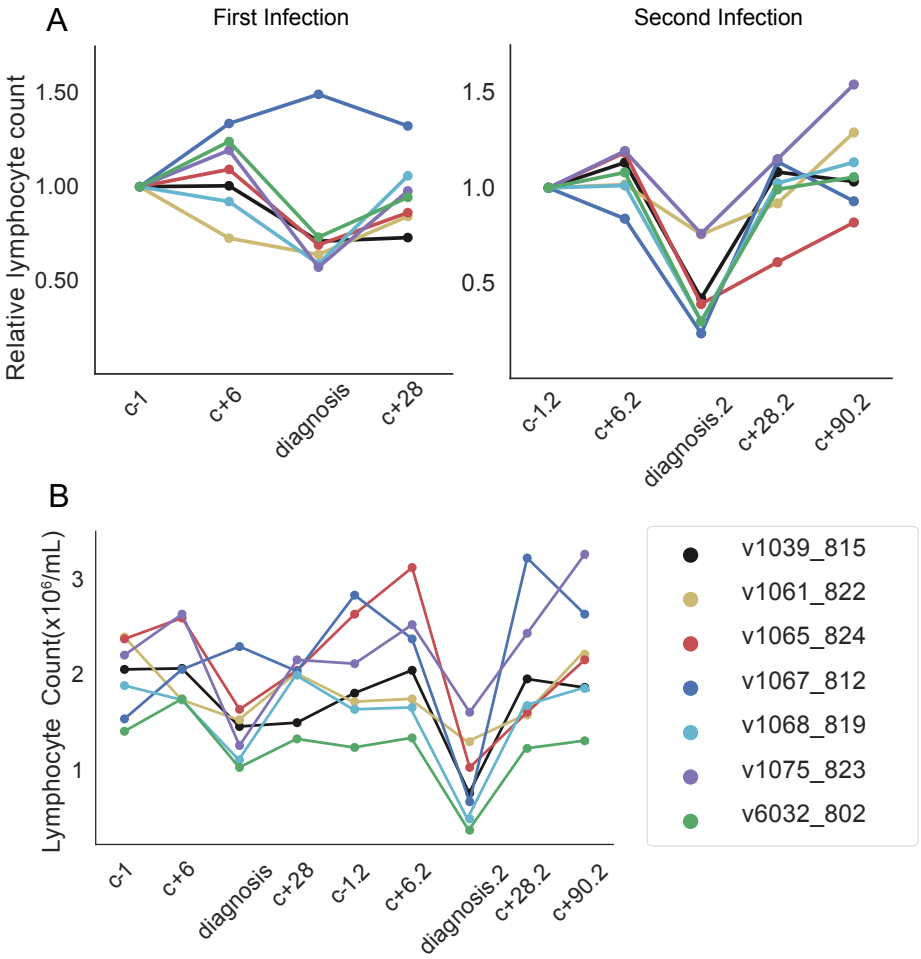


Figure 5.2: Lymphocyte counts (A) relative to baseline (c-1 for each infection), and (B) absolute counts ($x10^6$/mL) for the 7 volunteers undergoing a blood-stage *P. falciparum* CHMI. Data is coloured according to individual volunteer and depicted for a first and second (.2) infections, from baseline (c-1).

Figure 5.3

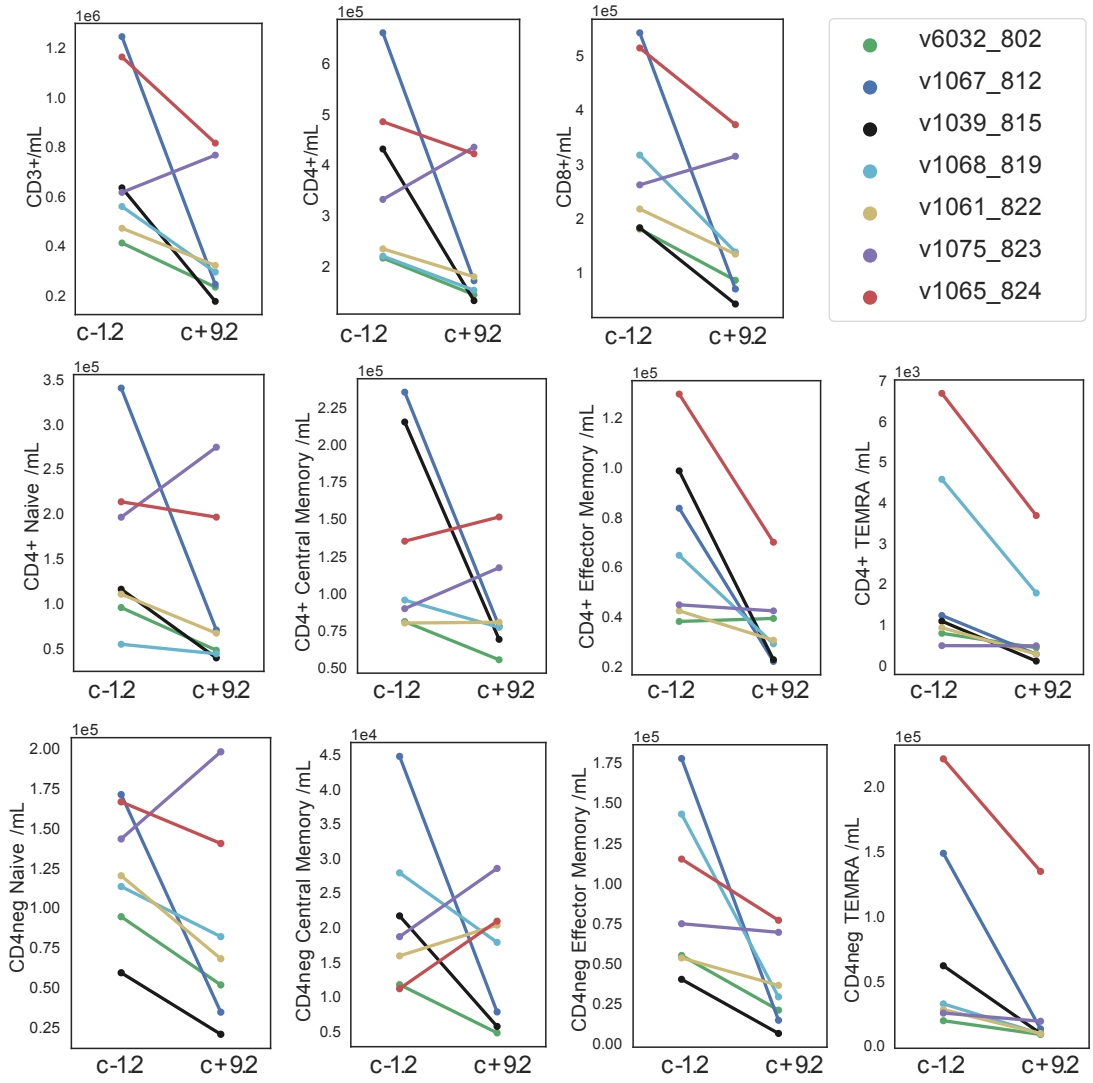


Figure 5.3: T-cell subset dynamics of the 7 volunteers during a second infection. Data represents the absolute values at baseline (c-1.2) and during infection (c+9.2). Data is coloured according to individual volunteer. Count per mL of blood ($x10^n$ as indicated) of each described T-cell subset is depicted on the y axis.

Figure 5.4

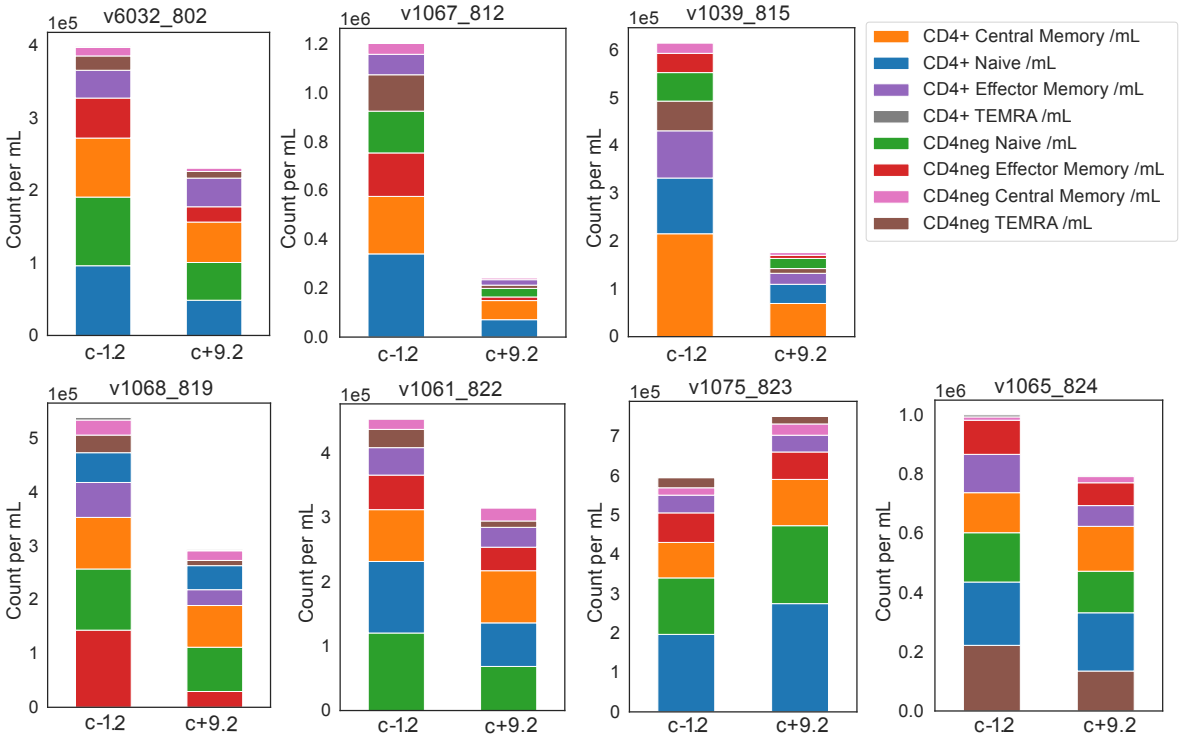


Figure 5.4: T cell subset dynamics in a second infection: Stacked bar plots are presented for each individual at baseline (c-1) and c+9. Data was only available for individuals undergoing their second infection. Each colour represents a different T-cell subset, with count per mL of blood (either $x10^5$ or $x10^6$ as indicated) depicted on the y axis.

5.4.7. TRBV and TRBJ gene usage

Despite the pronounced lymphopenia at day of diagnosis, throughout both infections, there was marked intra-individual stability in the TRBV gene usage of all volunteers (Figure 5.5). Between key time-points, the minimum correlation coefficient between TRBV gene usage within an individual was 0.94 (Figure 5.6), and Principal Components Analysis of TRBV and TRBV/J gene usage (Figure 5.8), clearly separated repertoires by individual rather than time-point. To quantify similarity in TRBV gene usage across all samples, the Manhattan distance between each repertoire was calculated. This again clearly indicated clustering by individual rather than timepoint, with marked stability of TRBV repertoires through time. There was also marked intra-individual stability in the TRBV/J gene repertoire throughout key timepoints of both infections (Figure 5.9).

Figure 5.5

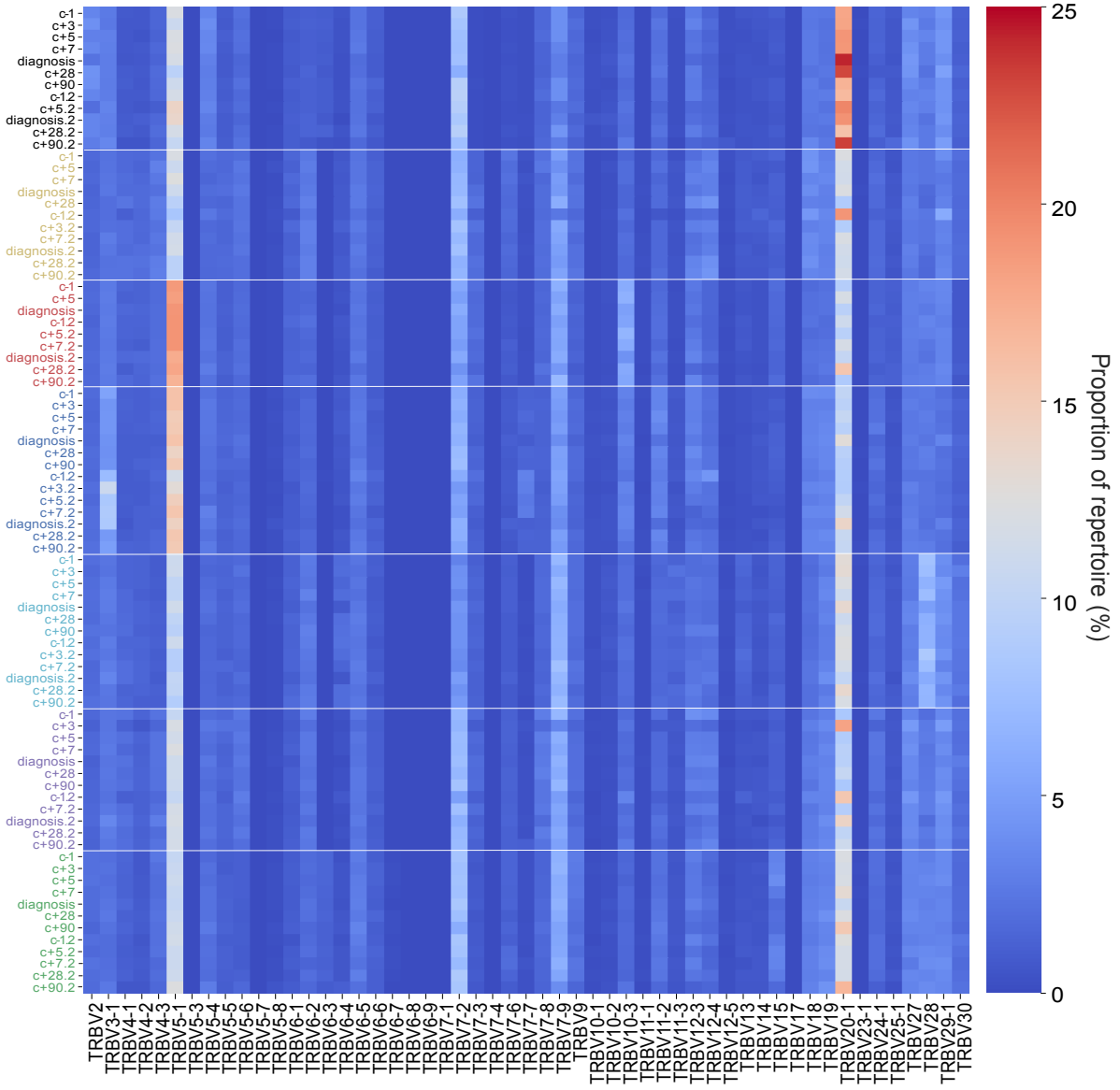


Figure 5.5: Proportion of TRBV gene usage. Data is presented for all 83 samples, where each row is an individual's repertoire at a particular timepoint, and each column is a unique TRBV gene. For each volunteer, indicated by colour of y-axis labels, repertoires are presented in longitudinal order, with (.2) indicating second infection.

Figure 5.6

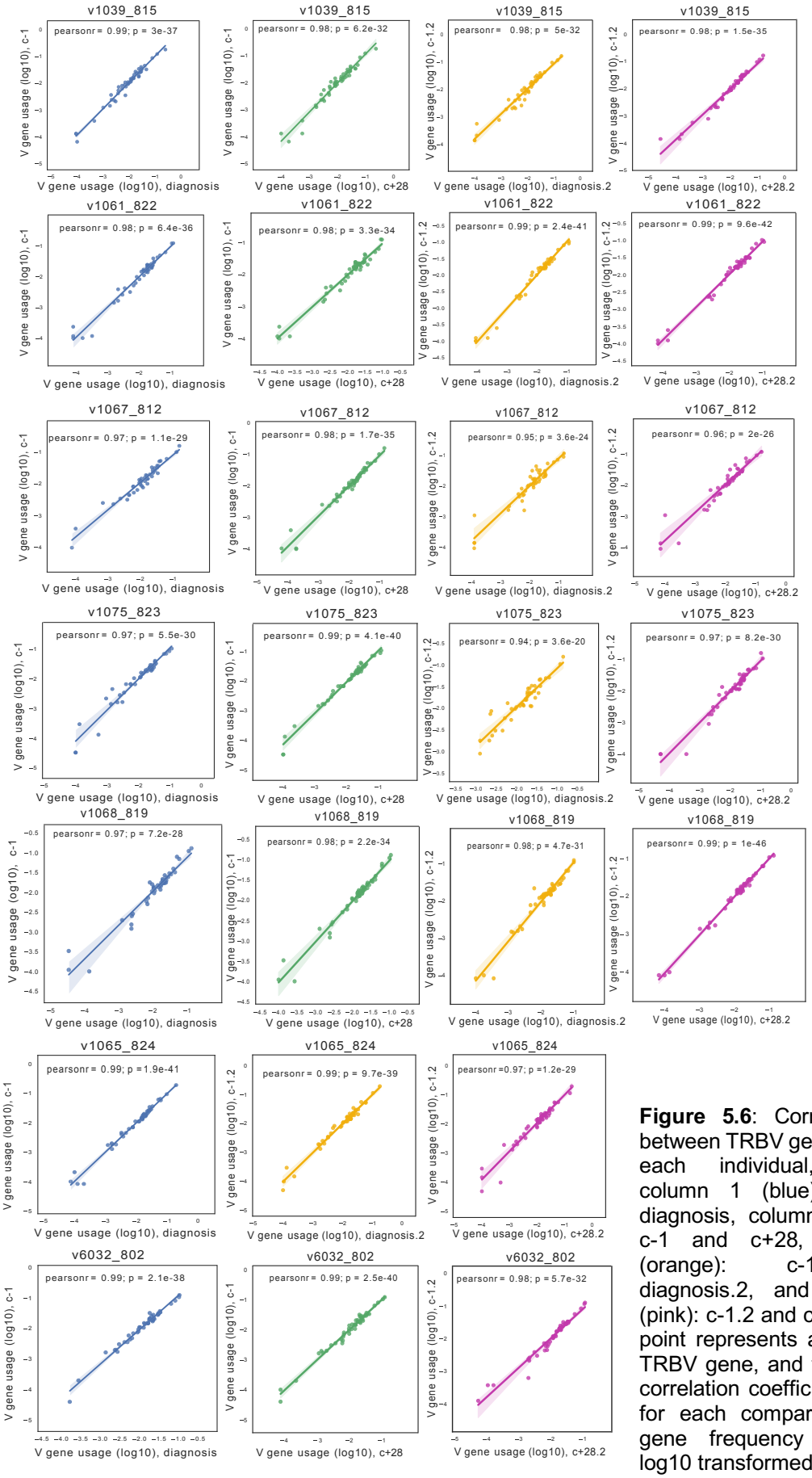


Figure 5.6: Correlation plot between TRBV gene usage for each individual, between column 1 (blue): c-1 and diagnosis, column 2 (green): c-1 and c+28, column 3 (orange): c-1.2 and diagnosis.2, and column 4 (pink): c-1.2 and c+28.2. Each point represents an individual TRBV gene, and the Pearson correlation coefficient is given for each comparison. TRBV gene frequency has been log10 transformed.

Figure 5.7

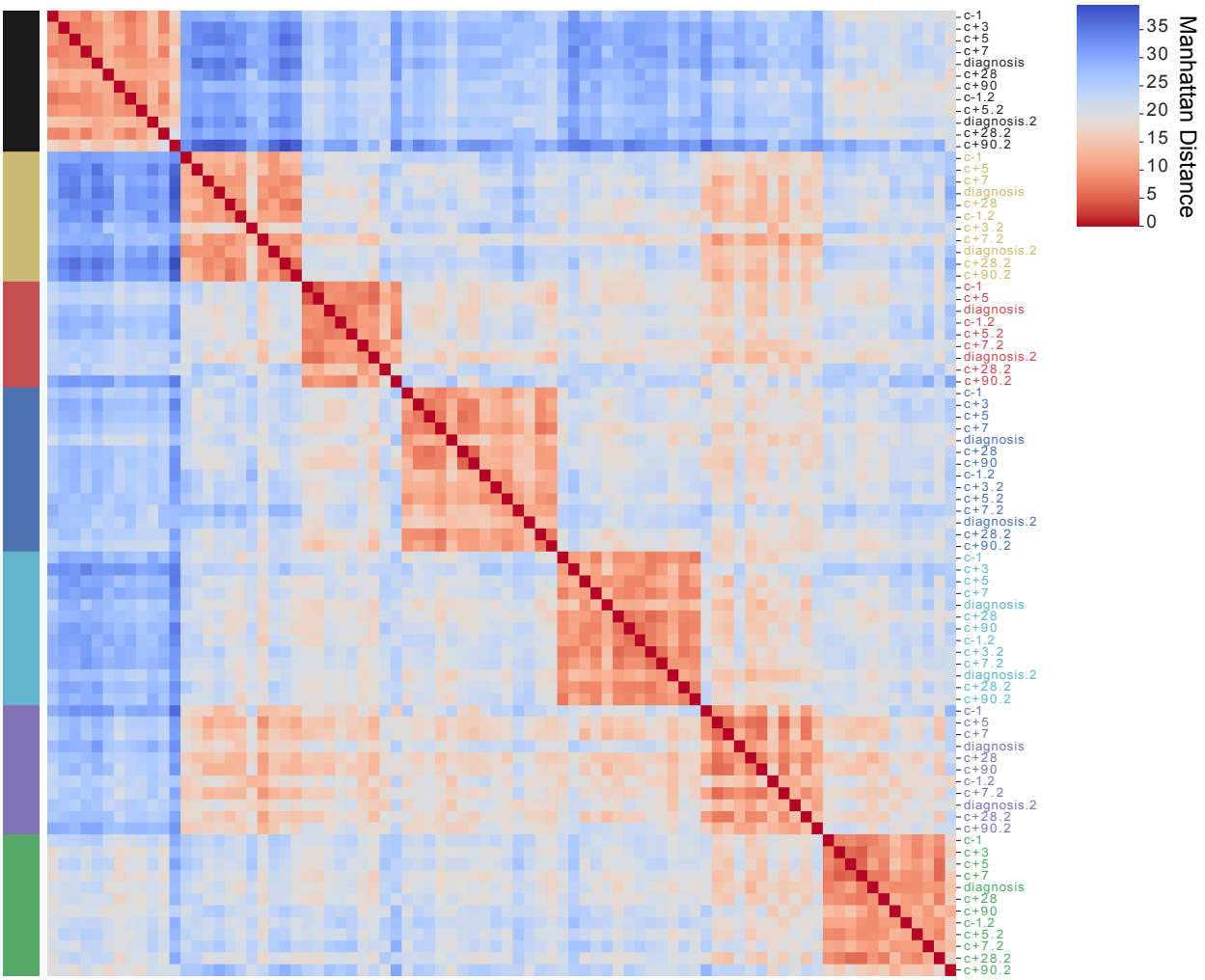


Figure 5.7: Manhattan distance matrix of TRBV gene usage. For each individual, indicated by colour of y-axis, repertoires are presented in longitudinal order, with (.2) indicating second infection. A smaller distance (red) indicates greater similarity between TRBV gene usage, with a distance of 0 indicating identical TRBV gene usage.

Figure 5.8

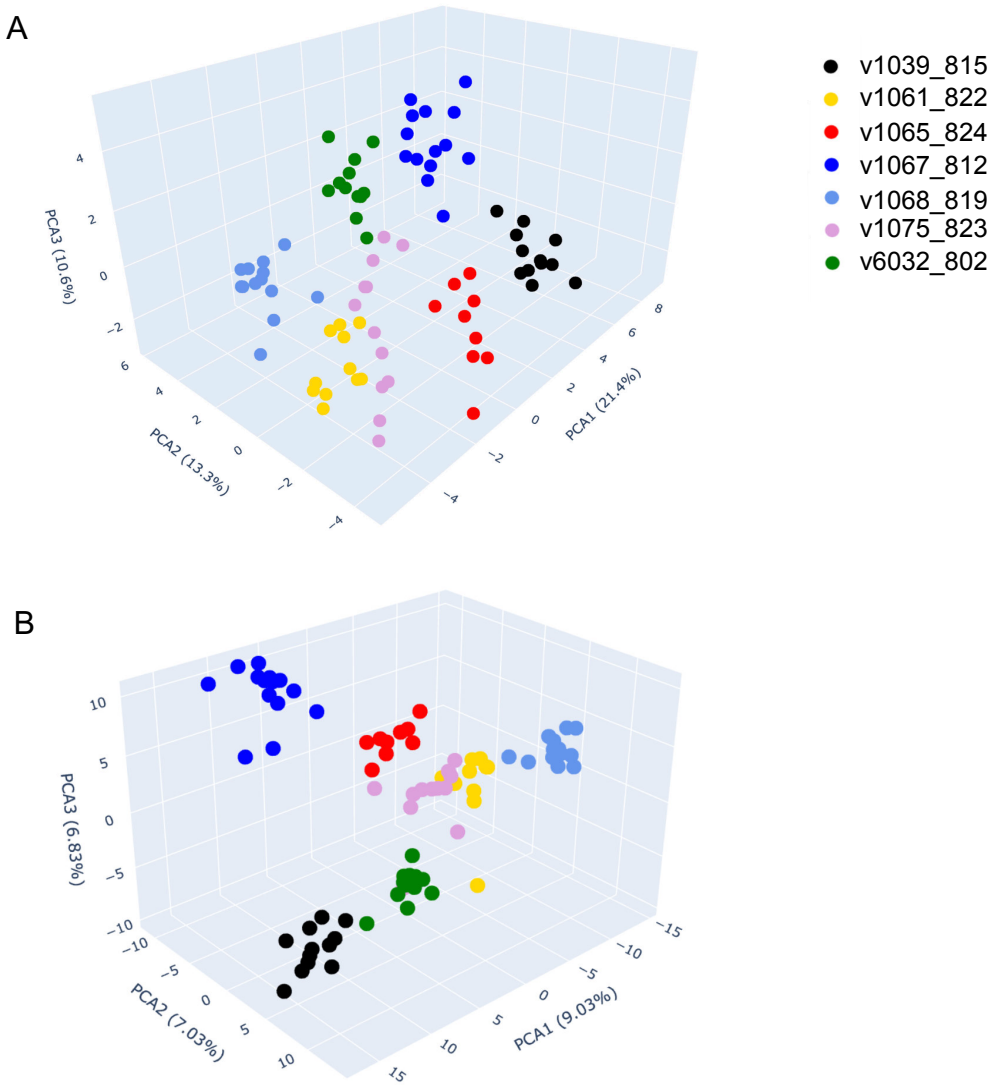


Figure 5.8: Principal Components analysis of (A) TRBV gene usage frequency and (B) TRBV/J gene usage frequency, for all 83 repertoires comprising all time-points. The variation encoded by each principal component (1, 2 and 3) is given for each axis, as a percentage of total variation in the dataset.

Figure 5.9

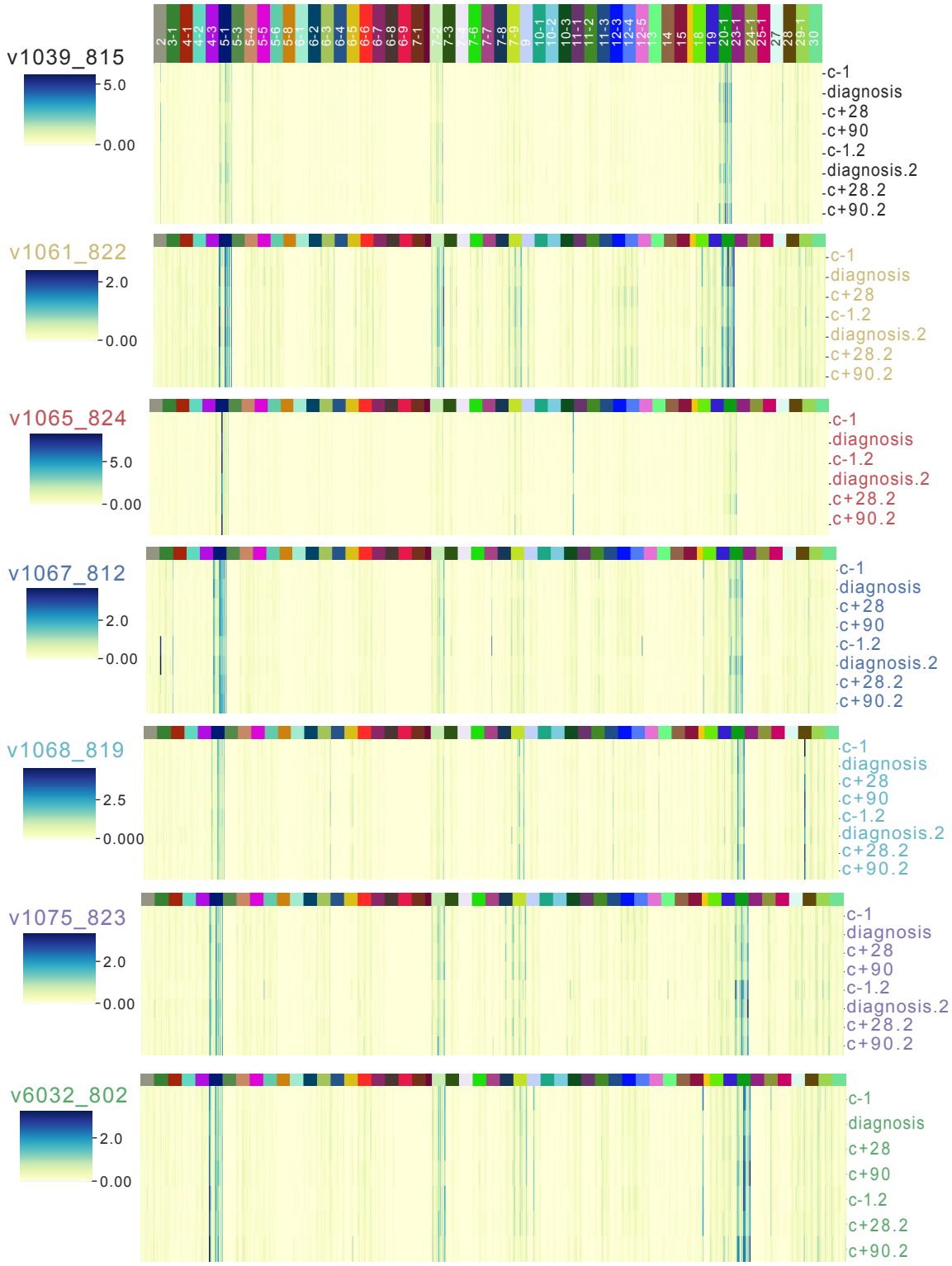


Figure 5.9: Heatmaps the TRBV/J proportion of each repertoire, for key time points (c-1, diagnosis, c+28, c+90, c-1.2, diagnosis.2, c+28.2, c+90.2). Data is presented longitudinally for each individual, and colour indicates proportion (%) of repertoire using a particular TRBV/J combination. Colour scale is per individual. Horizontal colour bar indicates TRBV gene used in the TRBV/J combination.

5.4.8. Repertoire clonality and architecture

I next examined the clonal architecture of the repertoires, observing firstly that background repertoire diversity and overall clonal composition was very specific to each volunteer. Individuals v1039_815, v1065_824, v1068_819 and v6032_802 had highly expanded (> 1% of repertoire) and persistent clones that dominated their repertoires throughout both infections (Figure 5.10). In contrast, individuals v1061_822 and v1075_823 lacked these highly expanded, persistent clones (Figure 5.10). For v1067_812, expansion of a dominant clone was evident from baseline of their second infection, through to c+28.2, and therefore likely represents a subclinical clonal response to an antigenic stimulus unrelated to CHMI (Figure 5.10B(iv)). Given the heterogeneity in background clonality between individuals, repertoire diversity was calculated using generalised Renyi entropy (Renyi, 1961; Rempala and Seweryn, 2013; Greiff *et al.*, 2015; Simnica *et al.*, 2019), rather than using a specific diversity metric. This was used as single diversity indices have been shown in some instances to give qualitatively different results when comparing heterogenous repertoires (Leinster and Cobbold, 2012). Using generalised Renyi entropy allows us to focus on different aspects of the repertoire structure. It is parameterised by alpha, which represents a weighting, and as alpha increases more weight is given to higher frequency clones. Several key diversity metrics are captured by changes in this weighting: alpha=0 is species richness i.e. how many unique clones are present, alpha=1 represents Shannon's entropy and alpha=2 is the inverse Simpson's diversity index. As alpha tends to infinity, more weight is placed on expanded clones. To characterise the clonal structure of the repertoires through time, a Renyi's entropy curve was constructed for each sample, with alpha varying from 0 to 10. As expected, these diversity profiles clearly separated volunteers v1061_822 and v1075_823 from volunteers with hyperexpanded clones in their repertoire (Figure 5.11). Given the kinetics of the T-cell response, if a pathogen-specific response was clonally expanded this would only be detectable at post-treatment timepoints when activated T-cells have re-entered circulation. Therefore, to determine if the diversity of repertoires changed through time, Renyi entropy only at key timepoints (c-1, diagnosis, c+28 and c+90) was analysed using linear mixed models. At alpha = 0, alpha = 1, alpha = 2 and alpha = 10, no difference in diversity metrics between c-1 and returning clones at c+28 and c+90 was detected in either a first or second infection. Rather, consistently across

variations of alpha, for both infections, the overall trend was for repertoires to become more diverse at day of diagnosis (Figure 5.12, Appendix 5.11). In a first infection, significant post-hoc pairwise comparisons of diversity were evident between c-1 and diagnosis at alpha=1 (FDR=0.039), alpha=2 (FDR=0.004) and alpha=10 (FDR=0.0024), with effect size increasing as alpha weighting is increased. In a second infection, this pattern of increased diversity at day of diagnoses was still present, though significant differences in diversity between key timepoints was not detected for any weightings of alpha. Lack of significance here is likely due to the dominant clonal expansion seen at baseline (c-1.2) in individual v1067_812 (Figure 5.10iv), skewing group diversity, as well as a lack of power to detect differences in diversity given the small number of individuals in the cohort. In sum, no clonal expansion at post-treatment timepoints was evident, but there was an overall trend for repertoires to be more diverse at day of diagnosis, present in both a primary and secondary infection.

Figure 5.10

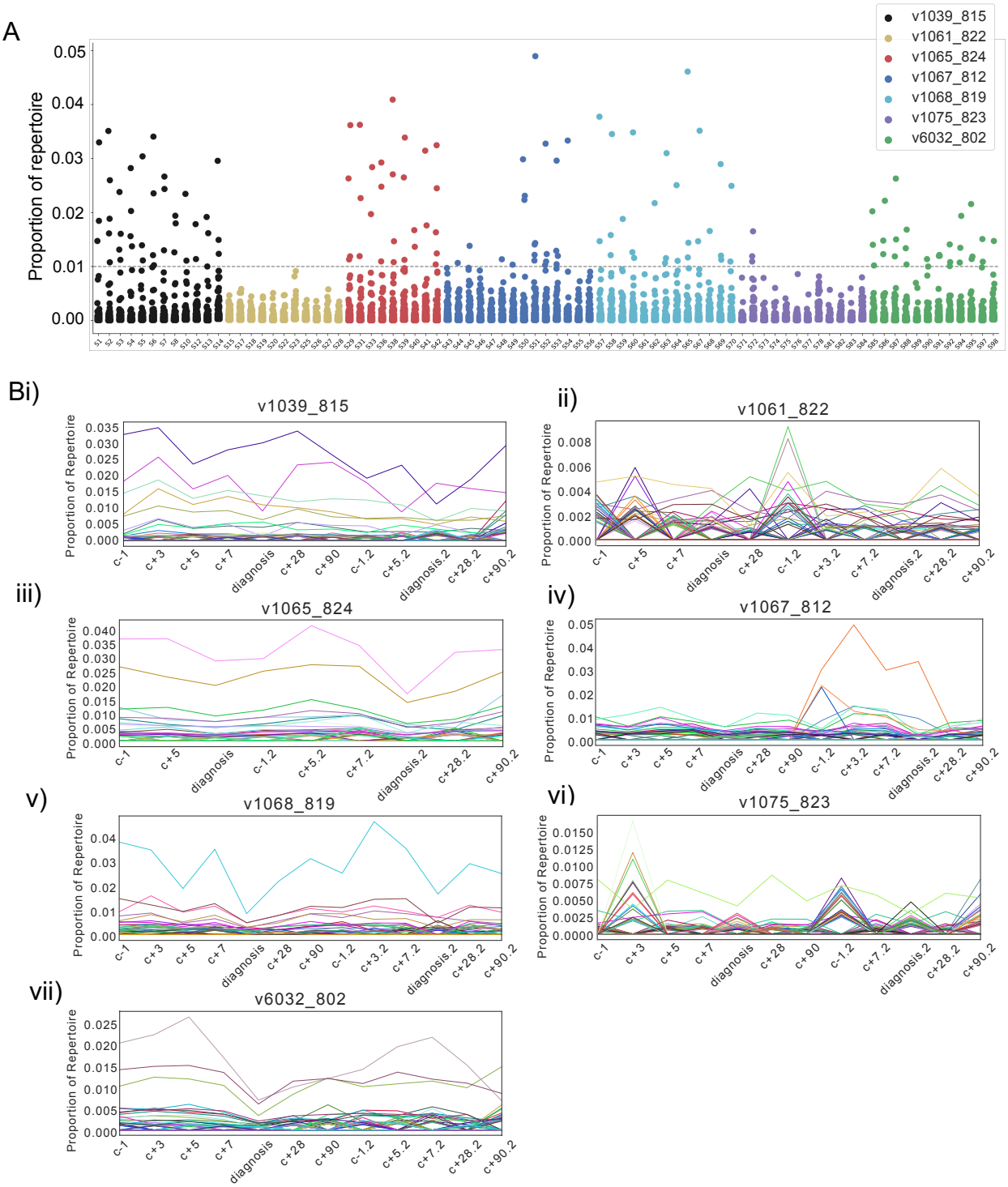


Figure 5.10: A) Strip plot displays the proportion of repertoire being taken up by each individual clone for each individual. Each point is an individual clone, and they are coloured according to individual. B) Parallel plots display the proportion of repertoire taken up by the top 25 clones in each individual at each available timepoint. Lines are coloured according to clone, and timepoints are displayed in longitudinal order for each individual.

Figure 5.11

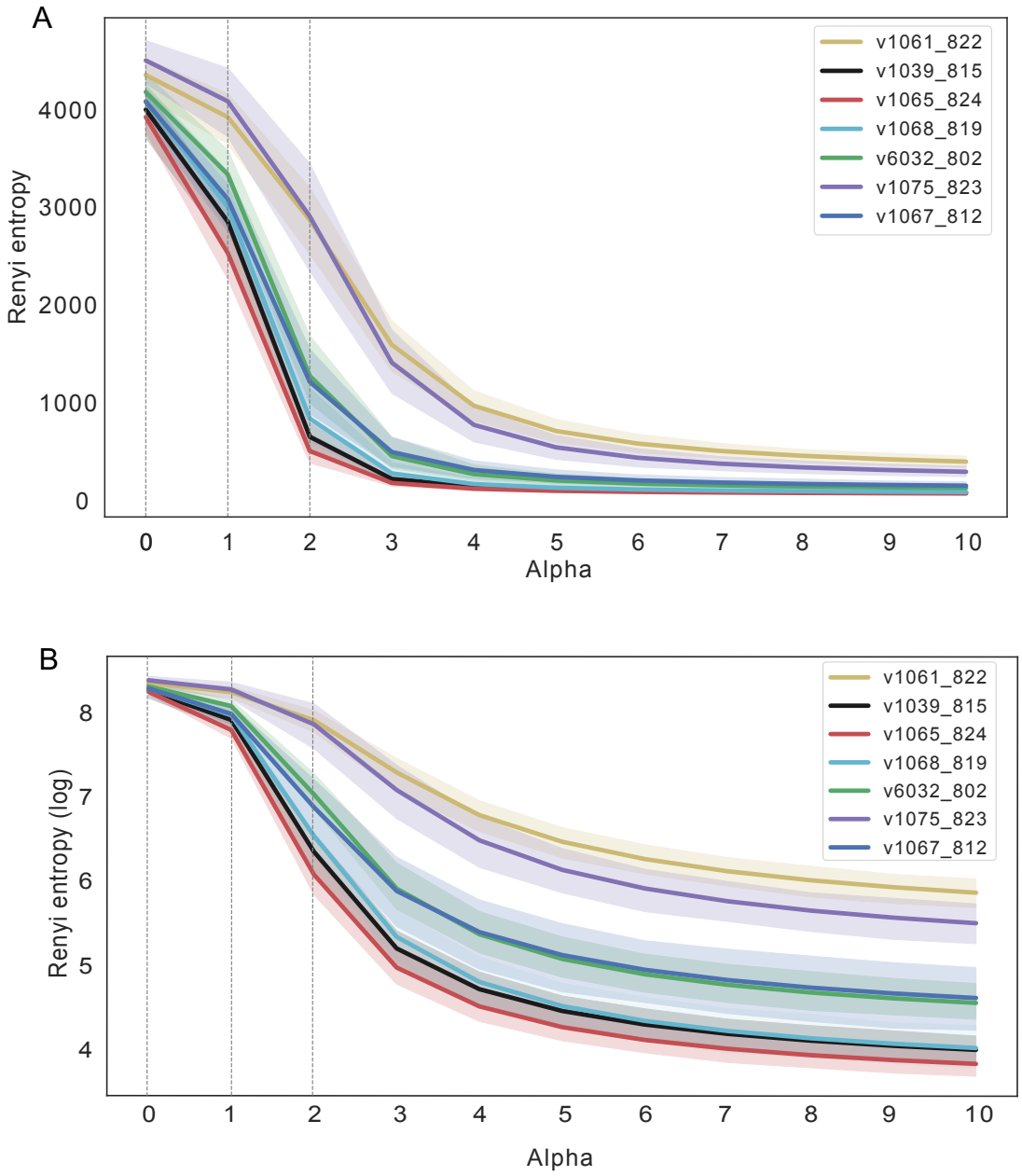


Figure 5.11: A) The mean Renyi entropy across all available timepoints is displayed for each individual, with alpha varying from 0 to 10. Shaded areas show 95% CI for each individual. Dashed vertical grey lines show Renyi entropy at alpha= 0, alpha=1 or alpha=2. These correspond to species richness, Shannon entropy and inverse Simpson's diversity index respectively. B) Same data as (A) but shown as log Renyi entropy. For both (A) and (B), repertoires have been normalised by down-sampling to 5000 UMI.

Figure 5.12

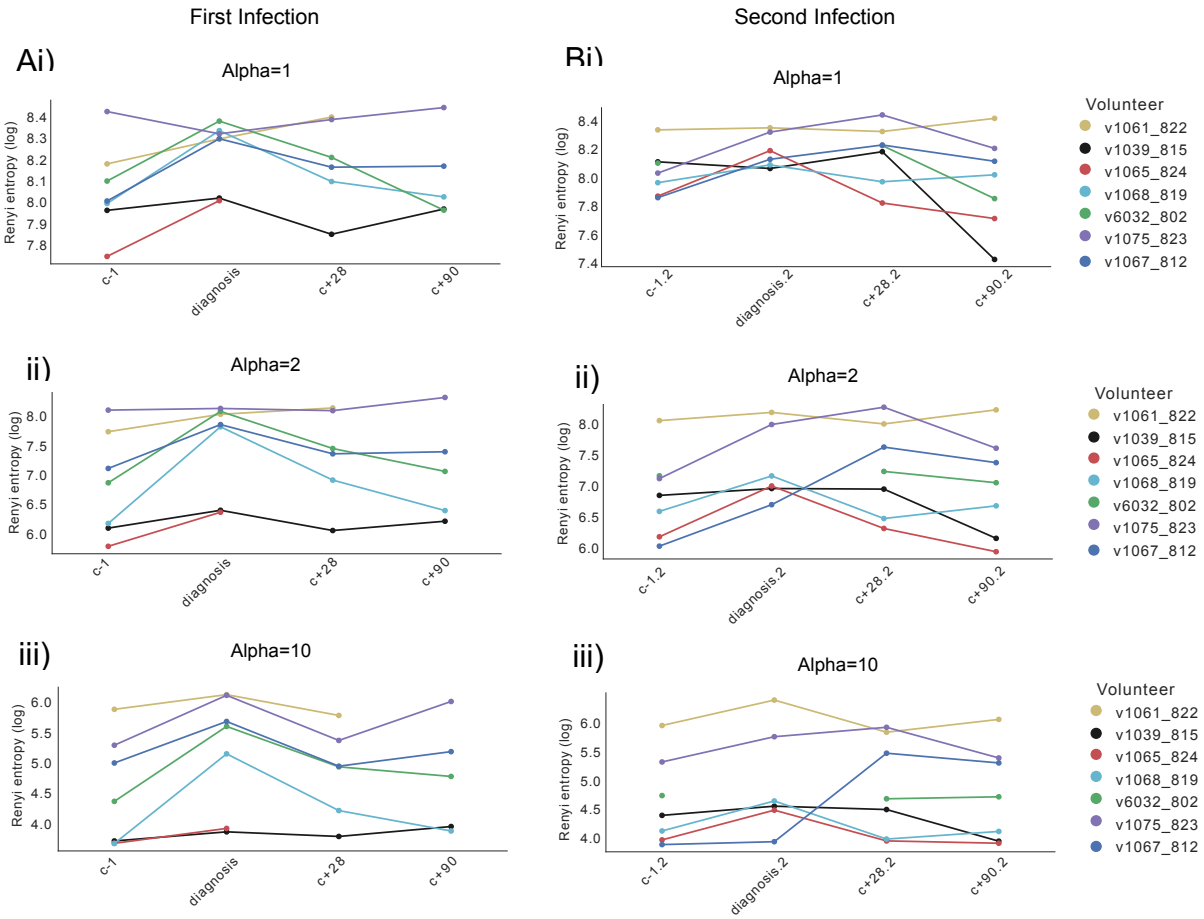


Figure 5.12: Parallel plots depict the Renyi entropy (log-transformed) for timepoints c-1, day of diagnosis, c+28 and c+90 for (A) individuals undergoing a first infection and (B) individuals undergoing a second infection. The Renyi entropy is depicted for (i) alpha=1, (ii) alpha=2 and (iii) alpha=10. Points are coloured by individual.

This trend of increased diversity at day of diagnosis, occurring at the trough of lymphopenia, suggests that established expanded clones are either being non-specifically recruited from the peripheral circulation, or that they are potentially being diluted by a more diverse cell population entering the periphery. This latter explanation is not supported by available FACS data or lymphocyte count data. Therefore, to investigate non-specific recruitment further, the proportion of each repertoire taken up by an individual's persistent clones (those present at every available timepoint for that individual), was analysed longitudinally across key timepoints using linear mixed effects models. Persistent clones were found to be significantly reduced in proportion

at day of diagnosis in both a primary and secondary infection (FDR<0.001), returning to approximate baseline levels by c+28 (Figure 5.13, Appendix 5.12). Persistent expanded clones have previously been demonstrated in longitudinal TCRβ sequencing in both healthy (Chu *et al.*, 2019) and diseased individuals (Servaas *et al.*, 2021), and likely represent a heterogenous pool of mostly antigen-experienced memory clones from previous or chronic antigenic exposure (Chu *et al.*, 2019), as well as clones with higher generational probabilities. To rule out that the volunteer's persistent clones are not consistently present simply due to biases in VDJ recombination (making them more likely to be present in the naïve TCR pool), the probability of generation (PGen) of persistent vs non-persistent clones was compared for each individual. This was calculated unweighted by clone frequency to prevent distribution skew due to expanded clones. The PGen of persistent versus non-persistent clones did not significantly differ for volunteers v1039_815, v1065_824, v1067_812, v1068_819 and v6032_802 (Figure 5.14A), indicating their persistent clones include previously antigen experienced populations. Individuals v1061_822 and v1075_823 did have slightly increased PGen of their persistent clones, as expected given their greater relative frequency of naïve CD4+ and CD8+ cells at baseline (Figure 5.4) and lack of expanded clones (Figure 5.10). However, the Pgen of persistent clones did not affect whether or not these clones were reduced at day of diagnosis compared to baseline (Figure 5.13B), indicating it isn't simply persistent clones that have a high Pgen (likely naïve) that are being recruited, in agreement with dynamics from the FACS data (Figure 5.4).

Figure 5.13

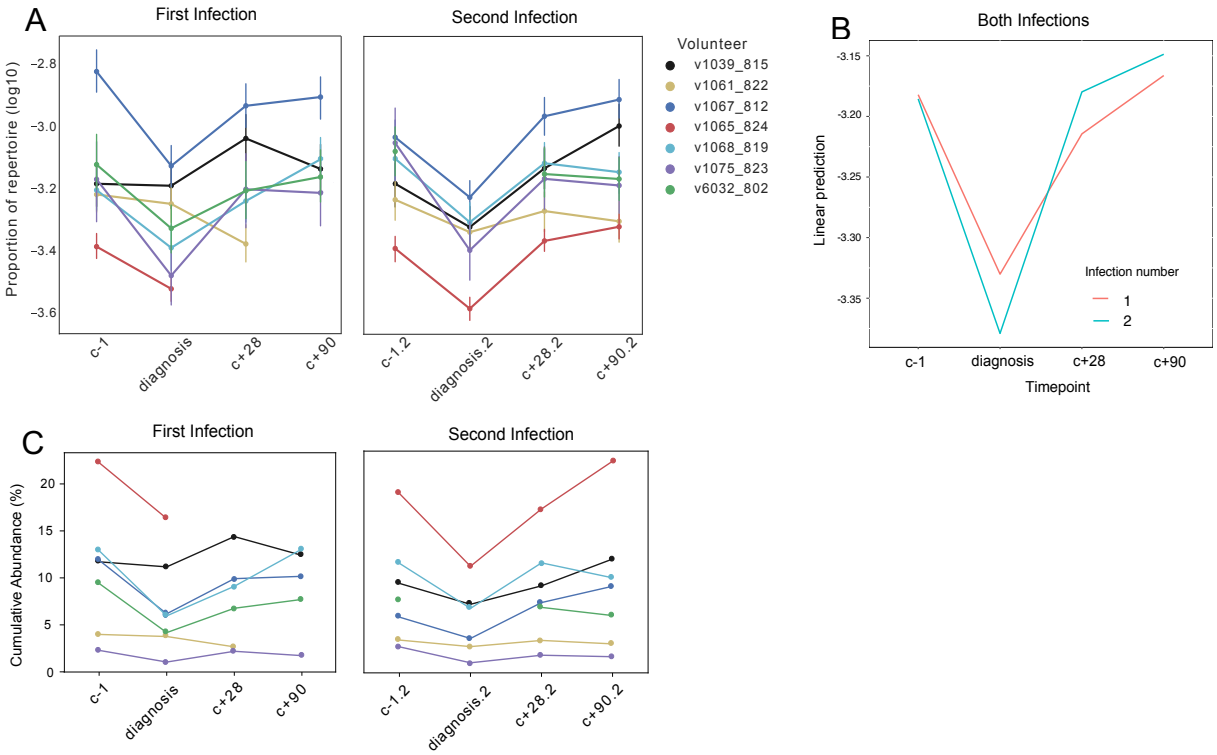


Figure 5.13: A) Mean proportion of persistent clones (log10-transformed) longitudinally displayed through key timepoints in both a primary and secondary CHMI infection. Error bars represent SEM and individuals are indicated by colour. B) Mean proportion through time was analysed using linear mixed models, with model predictive values for both a first and second infection depicted at key timepoints. C) Cumulative abundance of persistent clones per individual.

Figure 5.14

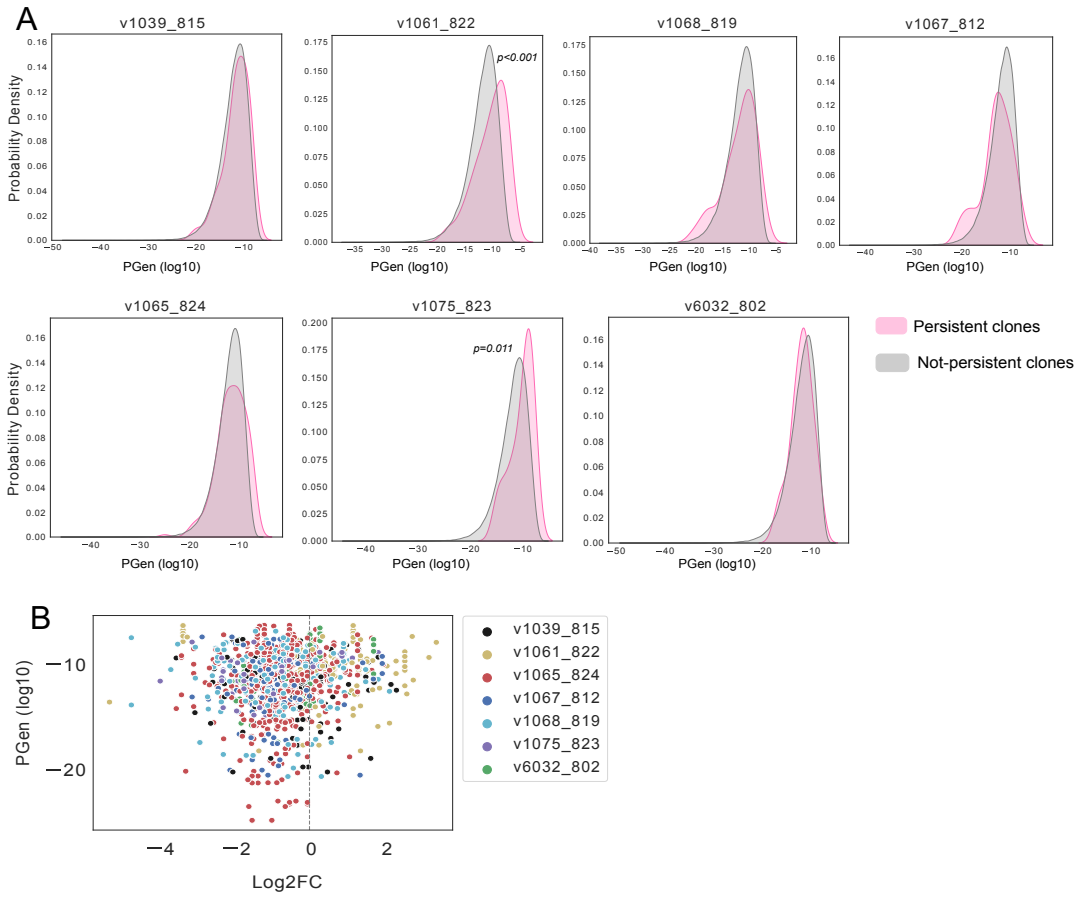


Figure 5.14: A) KDE plots of probability of generation (PGen) of TCR CDR3 nucleotide sequences for persistent (pink) and not-persistent clones (grey) for each individual. Persistent clones were defined as those being present at every available timepoint for that individual. Differences in Pgen (log 10) between persistent and not-persistent clones for each individual were analysed using Wilcoxon rank test. B) Log2 fold change in proportion of persistent clones between c-1 and diagnosis for both infections.

To investigate non-specific recruitment further, persistent clone sequences were queried against curated sequences in VDJdb (Shugay *et al.*, 2018), a database of previously identified TCR epitope specificities matched to an individual's HLA-type. Previously annotated epitope-specific clones to EBV, CMV and Influenza A were identified, the majority of which had a negative fold-change between baseline and day of diagnosis in both a primary and secondary infection (Table 5.2, Figure 5.15). Together, this supports the explanation that diversity is increased at day of diagnosis, due to the non-specific recruitment of previously established expanded clones out of peripheral circulation.

Table 5.2

Volunteer	CDR3	MHC A	MHC B	Epitope	Species	Log2FC 1	Log2FC 2
v1065_824	CASSPTRNTEAFF	HLA-B*07:02	B2M	pp65	CMV	-0.324	-3.194
v1065_824	CASSLEGDQPQHF	HLA-B*07:03	B2M	NP	Influenza A	-0.197	-0.49
v1065_824	CASSRTgDTGELFF	HLA-B*08:01	B2M	BZLf1	EBV	-1.43	-0.85
v6032_802	CSARARGLNTGELFF	HLA-A*11:01	B2M	EBNA4	EBV	-1.25	NA
v1068_819	CASSLNGGHYEQYF	HLA-A*11:01	B2M	EBNA5	EBV	1.09	-0.487
v1068_819	CASSLGThYGYTF	HLA-B*08:01	B2M	EBNA3A	EBV	-1.68	1.1
v1068_819	CASLGlAYNEQFF	HLA-A*11:01	B2M	EBNA4	EBV	-0.69	-0.12
v1068_819	CATSRqNSGANVLTF	HLA-A*03:01	B2M	IE1	CMV	-0.69	-1.9
v1068_819	CASSQDRGsQETQYF	HLA-A*11:01	B2M	EBNA4	EBV	-0.11	-0.49
v1068_819	CASSrRNTEAFF	HLA-A*03:01	B2M	IE1	CMV	-0.7	-1.22
v1068_819	CASSEeAGGPETQYF	HLA-A*03:01	B2M	IE1	CMV	0.11	-0.49
v1068_819	CASSElAGGPETQYF	HLA-A*03:01	B2M	EBNA3A	EBV	0.11	-0.49
v1075_823	CASSIGQAYEQYF	HLA-B*08:01	B2M	EBNA3A	EBV	-2.89	-1.27
v1061_822	CASSLAhTGELFF	HLA-A*01:01	B2M	pp50	CMV	-0.551	-0.877

Table 5.2: Epitope specificity results from querying an individual's persistent clone sequences against VDJdb, a curated database of known antigen specificities. Sequences were queried with identical V/Jgene usage and within 1 mismatch in aa CDR3 sequence, with matching HLA type to the individual. CDR3 sequences with 1 aa mismatch are indicated in lowercase. The log2 foldchange in proportion is given between c-1 and diagnosis in a first infection (Log2FC 1) and between c-1.2 and diagnosis.2 in a second infection (Log2FC 2).

Figure 5.15

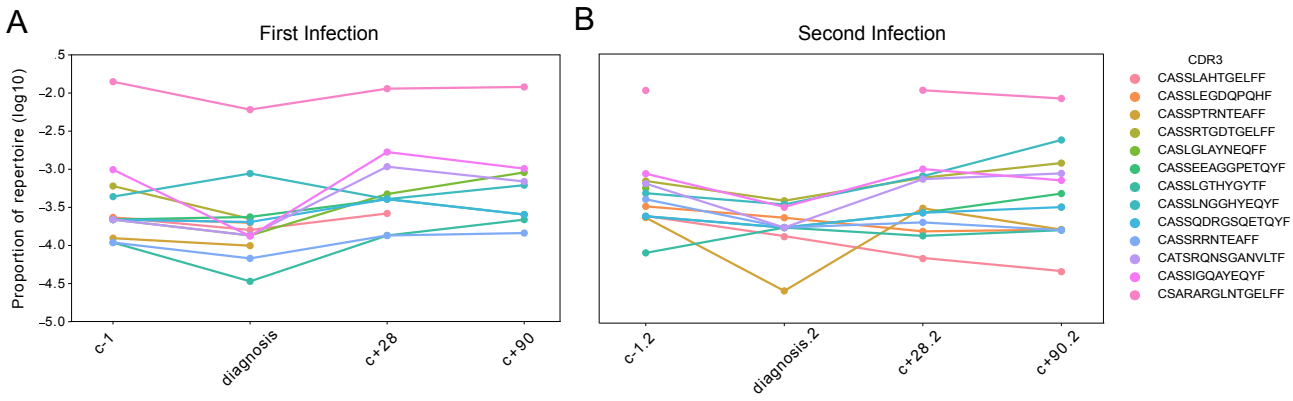


Figure 5.15: Proportion of the repertoire taken up by the CDR3 sequences that are listed in Table 5.2, across key timepoints of a first (A) and (B) second infection. Each CDR3 sequence represents a persistent clone with known epitope-specificity, as determined by querying the VDJdb database. Sequences were queried with identical V/J gene usage and within 1 mismatch in the aa CDR3 sequence, with matching HLA type to the individual.

5.4.9. Trajectory analysis of active clonal dynamics

Although clonality was unchanged at post-treatment timepoints compared to baseline, a highly heterogenous yet specific response to the parasite could potentially mask subtle changes in clonality and clonal dynamics. Trajectory analysis of clone frequencies through time has previously been successfully used to track and detect responding clonotypes in longitudinal repertoire data (Minervina et al., 2020a; Minervina et al., 2020b; Schultheiß et al., 2020). Such an approach has been shown to be effective when replicate repertoire data is not available (Minervina et al., 2020), therefore this technique was chosen to determine if particular responding clonal populations within an individual could be distinguished over time. Given expected TCR dynamics, if an epitope-specific response was present we would expect this to be enriched at c+28 and then again at c+28.2, and perhaps also still enriched at c+90 and c+90.2, when clones have reappeared in peripheral circulation. The available FACS data and movement of persistent clones suggests established clones are non-specifically recruited from the circulation. To further explore active clonal dynamics and determine if other groups of T-cell clones for example with opposing dynamics to persistent clones or kinetics to suggest they constitute a specific response to the parasite, trajectory analysis of clones through time was undertaken. Briefly, the

normalised frequencies of the top 1000 most abundant clones through key timepoints in both infections, were clustered using unsupervised hierarchical clustering as per (Minervina *et al.*, 2020b; Schultheiß *et al.*, 2020). For each individual, dendrogram visualisation indicated 2 main clusters, with multiple sub-branches (dendrogram for volunteer v1068_819 shown in Figure 5.16A). Setting cluster number to 2 revealed a main trajectory matching the pattern of previously seen dynamics of persistent clones (Figure 5.16B). However, no opposing trajectories or kinetics that represented a biologically relevant distinct responding cell population were identified; as cluster number was increased, over-fitting occurred, and individual trajectories simply represented background 'noise' of low frequency clones sequenced only at one timepoint, but not others, with the trajectory of persistent clones omnipresent. Varying the number of clones used for the trajectories, including using all clonal sequences, or restricting the analysis to clones present in at least 2 timepoints, did not alter this result. To determine the detection limit of this approach, clonal expansion at timepoints c+28 and c+28.2 was simulated, using the data from individual v1068_819. Simulation parameters were initially chosen based on replication of the trajectory analysis from publicly available repertoire data generated from longitudinal COVID-19 infections by Minervina *et al.*, (2020b), for which a mean fold-change of ~2.6 in proportion for a group of clones of absolute frequency ~70-80 out of 1000 unique clones in the trajectory was readily detectable. By varying degree of expansion and number of cells, I determined that 25 clones with a 2-fold-change, or 15 clones with a 4-fold-change in proportion at timepoints c+28 and c+28.2, was enough to be detected as a unique trajectory using this technique (Figure 5.16C). When simulating <20 clones with a 2-fold-change at these timepoints, no corresponding trajectory was detected without over-fitting, and this therefore represents the limit at which I could be confident in detecting a true trajectory using this technique. Overall, no specific responding clonal dynamics were detectable in any individual, except for the reduction in persistent clones at day of diagnosis in both infections. This is in agreement with the clonality analysis which did not detect expanded clones at c+28 compared to c-1 and corresponds with the mass cytometry data presented for CHMI volunteers in Muñoz Sandoval *et al* (In preparation), all of which together suggest a highly polyclonal and heterogenous T-cell response.

Figure 5.16

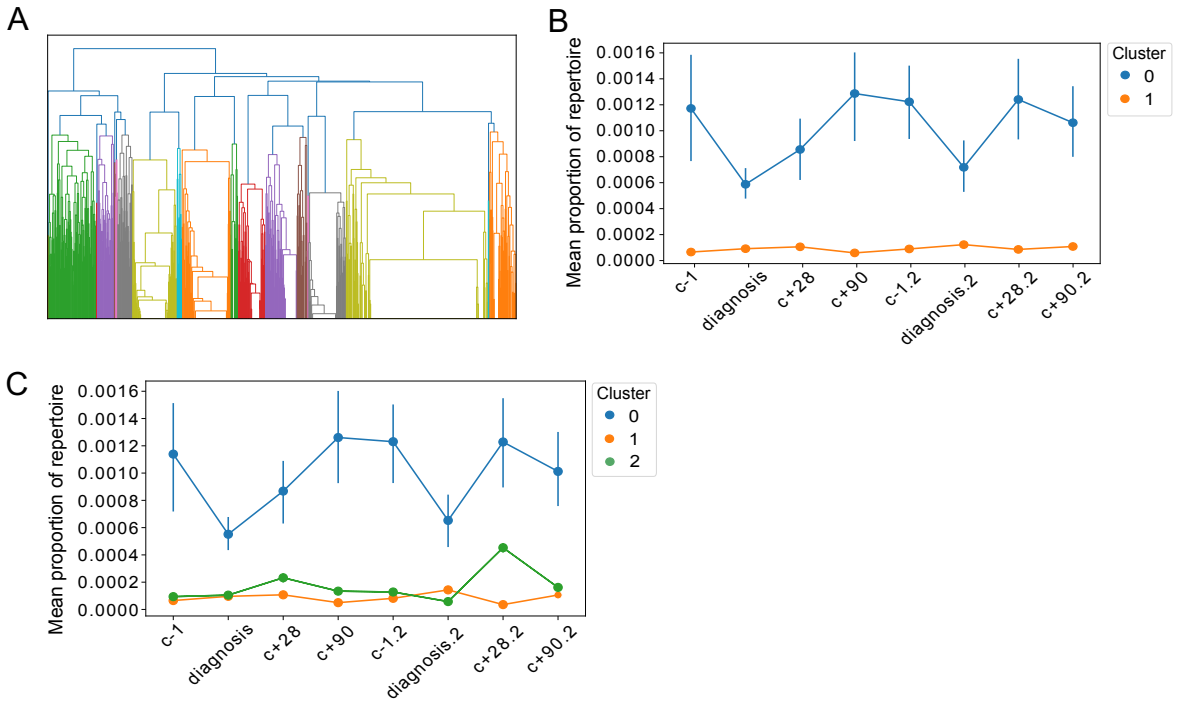


Figure 5.16: Dendrogram represents the unsupervised hierarchical clustering of the trajectories of the top 1000 clones for volunteer v1068_819, across key timepoints of both infections. The left-hand side branch is cluster 0 in (B) and (C) and represents persistent clones. B) Mean proportion of clusters when cluster number is set to 2. C) Mean proportion of clusters when clonal expansion of 2-fold has been simulated for 25 clones at c+28 and c+28.2, with simulated clones evident as cluster 3.

5.4.10. No conserved response detectable either within or between individuals

To further investigate this lack of a specific response using alternative methods, and to examine any shared responses between individuals, repertoire overlap analysis both within and between individuals was undertaken. The Jaccard index was used as a normalised sharing metric both within and between individuals. This indicated that samples taken from the same individual shared a high number of sequences, whilst there was very limited overlap between different individuals (Figure 5.17A, B). Given the expected T-cell dynamics, if a pathogen-specific response was enriched within an individual at post-treatment timepoints, we would expect a higher level of sharing between c+28 and c+90 than other time-points. However, in congruence with the trajectory analysis, no increased sharing between post-treatment timepoints was found compared to between baseline and post-treatment timepoints, for either infection (Figure 5.17C, Appendix 5.13). There was less overlap between key timepoints and day of diagnosis, in agreement with the drop in proportion in persistent clones at this timepoint (Figure 5.17C, Appendix 5.13). Clustering repertoires within 1 amino-acid mismatch using the modified Swarm algorithm used in Chapter 3, and then examining overlap in clusters, did not change this pattern. Similarly, no increase in network connectivity within post-treatment timepoints was evident when networks were created between CDR3 sequences using the Levenshtein distance (Figure 5.17E).

If a shared response between individuals was present at the clonal level, we would again expect to see an increase in sharing at either of the post-treatment timepoints. However, no differences in clonal sharing post-infection between individuals was found compared to other timepoints, and no increased sharing was found between individuals in a secondary infection (Figure 5.17D). A lack of sharing at the clonal level between volunteers is not unexpected, given volunteers are not matched by HLA-type. However, individuals with diverse HLA-type have been shown to share epitope-specific motifs in their CDR3 sequence in response to both mycobacterial (Huang *et al.*, 2020) and viral (Schultheiß *et al.*, 2020) infections. Therefore in order to further explore shared pathogen-specific responses between volunteers, analysis using GLIPH2 (“Grouping of Lymphocyte Interactions by Paratope Hotspots”) (Huang *et al.*,

2020) was undertaken. GLIPH2 uses sequence similarity and motif analyses to group together enriched TCR sequences that may be responding to the same epitope. For this analysis, pooled c-1 repertoires from all individuals were used as the input reference for GLIPH2. Repertoires of all individuals at day of diagnosis, c+28 and c+90 were analysed against pooled c-1 separately (by both timepoint and infection), to determine if there was enrichment of shared responses at specific timepoints, and then overlap analysis of significant clusters between timepoints was undertaken. This approach was used, as if true epitope-specific responses are being detected, given the dynamics of T-cells returning to peripheral circulation post-treatment, one would expect the same motifs to be found enriched at either both c+28 and c+90, or enriched again in post-treatment timepoints in a second infection, rather than enriched at only 1 timepoint. To exclude false positive results, GLIPH2 clusters were considered significant if they had a Fisher's exact p-value <=0.05, contained at least 3 unique CDR3 sequences, appeared in at least 2 individuals at that timepoint, and had similar V gene usage (Vβ score <= 0.05). This approach is similar - if not less conservative - to what has been used by other groups including the GLIPH2 authors (Glanville *et al.*, 2017; Huang *et al.*, 2020; Servaas *et al.*, 2021). Although significantly enriched motifs were detected at each timepoint relative to baseline, there was little to no overlap of these significant motifs between timepoints, with post-treatment timepoints sharing similar low counts of motifs between them as with day of diagnosis, when recruited T-cells are not expected to be enriched in the peripheral circulation. Motifs that were shared between key timepoints only overlapped with one other key timepoint and accounted for very low frequency clones (max count of 3, mostly singletons). These motifs therefore likely represent background 'noise' of motifs simply not sequenced at baseline, rather than true enriched convergent responses between individuals. To explore this observation further, the 5 (out of 402) overlapping motifs between c+28 and c+28.2, were mapped back on to key timepoints (Figure 5.18C), where they were indeed found to have similar counts at day of diagnosis and c+90, in agreement with these likely being false positive results when considering expected T-cell dynamics. The sensitivity of GLIPH2 to identify enriched motifs compared to reference input is evident here, in that clones with only a very small increase in frequency compared to c-1 are detected. However, the lack of temporal conservation and expected dynamics of these responses through time means, overall, no support for a conserved

convergent response between individuals was evident. Using the in-built GLIPH2 naïve reference files as a baseline (and excluding clusters that were enriched at c-1 in an individual), did not change this result.

Figure 5.17

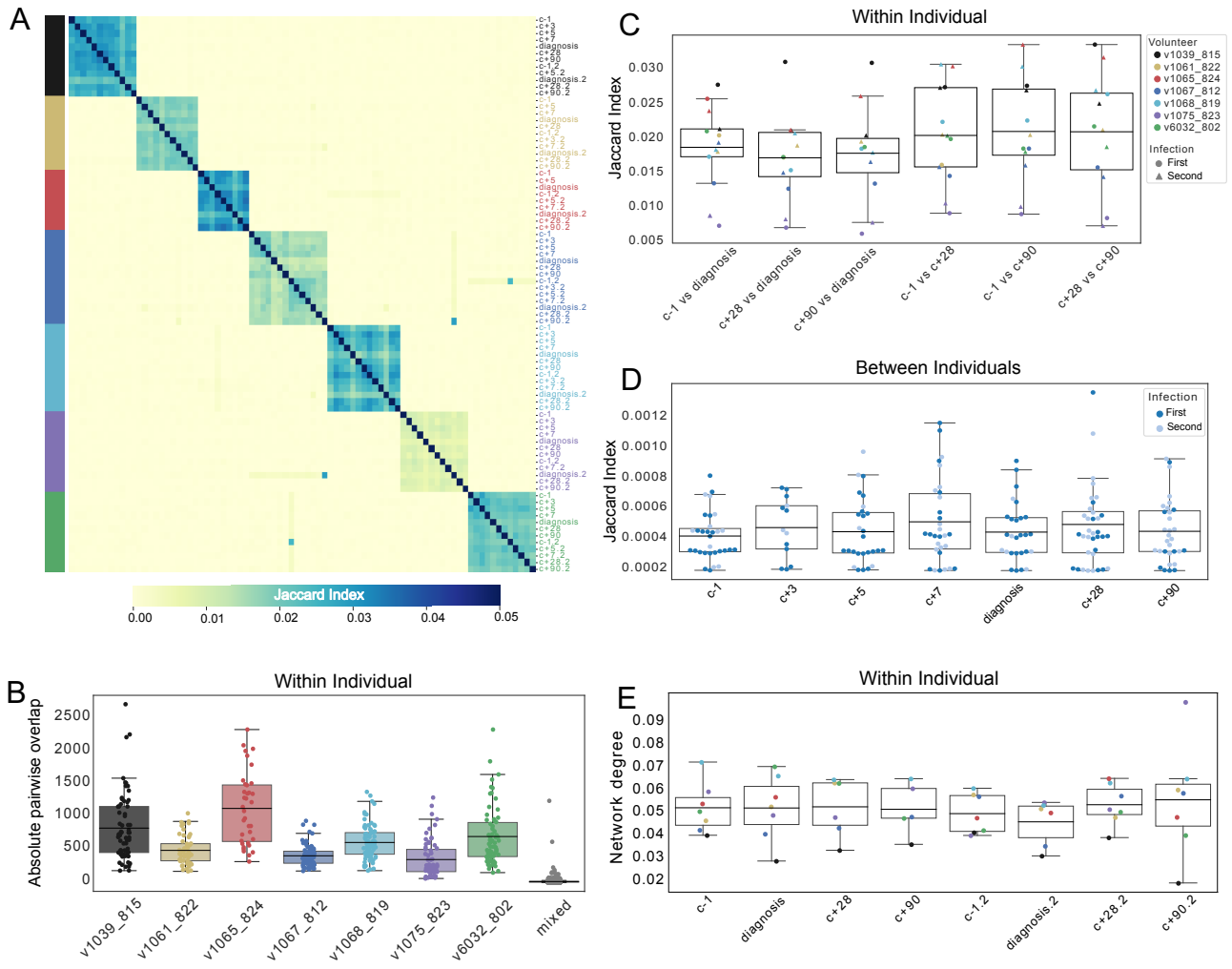


Figure 5.17: A) Heatmap depicts the pairwise Jaccard index between each of the 83 repertoires. The data has been ordered by individual (y axis colour bar) longitudinally through the time course of both infections. Jaccard index was calculated on UMI size-matched data. B) The absolute pairwise overlap between repertoires within an individual, unweighted by clonotype frequency. Each point represents the absolute number of TCR sequences shared between 2 timepoints within an individual. C) Jaccard index between key timepoints within an individual. Comparison is given on the x axis. D) Jaccard index between individuals at timepoints given on the x-axis. E) Network degree – edges per 100 unique TCR sequences. Edges were created if they were within a levenshtein distance of 1 of each other and normalised by unique number of clones in the network. Data is presented longitudinally through the key time points of both infections, and points are coloured by individual.

Figure 5.18

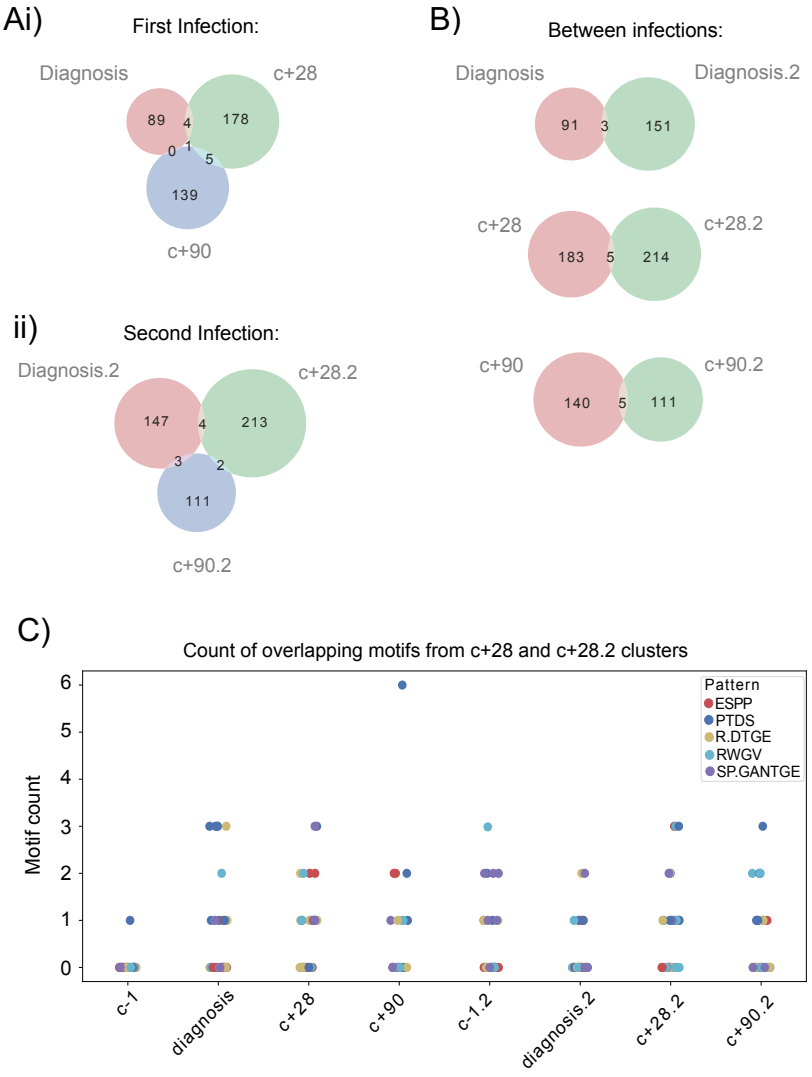


Figure 5.18: A) Venn diagrams show number of significant CDR3 amino acid motifs enriched compared to baseline (c-1) as determined by GLIPH2, that overlap with other key timepoints in a first (i) and second (ii) infection, and (B) between key timepoints of a first and second infection. C) The count of the overlapping motifs between c+28 and c+28.2 (5 motifs altogether), when mapped back on to key timepoints of each individual.

5.4.11. Mining whole-blood RNA-seq data

As part of the wider clinical trial, longitudinal whole blood RNA-seq of all the control-arm volunteers from Vac63A, Vac63B and Vac63C had previously been undertaken. Based on data from a concurrently running *P. vivax* CHMI trial which indicated that an activated T-cell population was evident soon after treatment, as well as data from Edwards *et al*., (2018) showing transcriptional changes in $CD4^+$ cells 6-7 days post-treatment, the timepoint of 6 days post-treatment (T+6) was included in Vac63C by collaborators. Vac63C consisted of 3 volunteers undergoing a primary infection, 2 undergoing a secondary and 6 volunteers undergoing a tertiary infection. Of the 6 individuals undergoing a third infection, 4 of them were volunteers from my initial data set (v1039, v1061, v1068, v6032, v1075). This RNA-seq data set was mined for TCR sequences, firstly to corroborate findings from the TCR-seq analysis, and secondly to determine if including data from the new T+6 timepoint and from tertiary infections would provide any additional insights in to repertoire dynamics.

Table 5.3

Volunteer	c-1	c+6	c+7	c+8	c+9	c+10	c+11	c+12	c+14	c+15	c+18	diagnosis	t+6	memory
v313	✓	✓	✓	✓	✓							✓	✓	✓
v315	✓	✓	✓	✓	✓							✓	✓	✓
v320	✓	✓	✓	✓	✓							✓	✓	✓
v806	✓✓	✓✓		✓	✓✓	✓✓	✓					✓✓	✓	✓
v818	✓✓	✓✓	✓✓	✓✓								✓✓	✓	✓
v1065	✓✓	✓✓		✓✓	✓✓	✓✓						✓✓		✓
v1067	✓✓	✓✓	✓✓	✓✓	✓							✓✓		✓
v1039	✓✓✓	✓✓✓	✓✓✓	✓✓✓	✓							✓✓✓	✓	✓
v1061	✓✓✓	✓✓✓		✓✓✓	✓✓✓	✓						✓✓✓	✓	✓

v1068	✓✓✓	✓✓✓	✓✓✓	✓✓✓								✓✓✓	✓	✓
v6032	✓✓✓	✓✓✓	✓✓✓	✓✓✓	✓✓							✓✓✓	✓	✓
v1075	✓✓✓	✓✓✓		✓✓✓		✓✓✓		✓✓	✓			✓✓✓	✓	✓
v1040	✓✓✓	✓✓✓		✓	✓✓✓	✓✓	✓	✓		✓	✓	✓✓✓	✓	✓

Table 5.3: Table displays what RNA-seq samples were available for each volunteer. Red tick indicates sample from a first infection, blue tick indicates sample form a second infection, black tick indicates sample form a third infection. Timepoint is presented in the table headings.

Altogether, 188 samples were available from 13 individuals (Table 5.3). Each sample on average yielded 369.2 (±9.2 SEM) unique TCRβ clones from an average of 581.7 (±14.9 SEM) functional TCRβ sequences (Appendix 5.14). This lower number of clones is expected given TCR transcript availability in non-targeted RNA-seq data, with short base pair reads (50bp x 50bp) (Bolotin *et al.*, 2017). Sample v1040_7 (c-1.2 for this volunteer), only yielded 1 clone and was excluded from further analysis. Comparing TRBV gene assignment between the baseline repertoires from the TCR-seq with their corresponding RNA-seq repertoires, showed good agreement between the 2 techniques, with an overall Pearson correlation coefficient of 0.81 despite the shorter read length of the RNA-seq data set. To confirm the dynamics of persistent clones from the bulk TCR-seq data, the frequency of these previously identified clones in the RNA-seq data from the 7 original volunteers was analysed. This showed a remarkably similar pattern of reduction in proportion at day of diagnosis of these persistent clones (Figure 5.19). Tracking these in a third infection showed similar dynamics to a first and second infection and indicated that at the novel timepoint of T+6, in 4 of the 5 volunteers with this timepoint, persistent clones had returned to baseline levels by this timepoint (Figure 5.19).

Figure 5.19

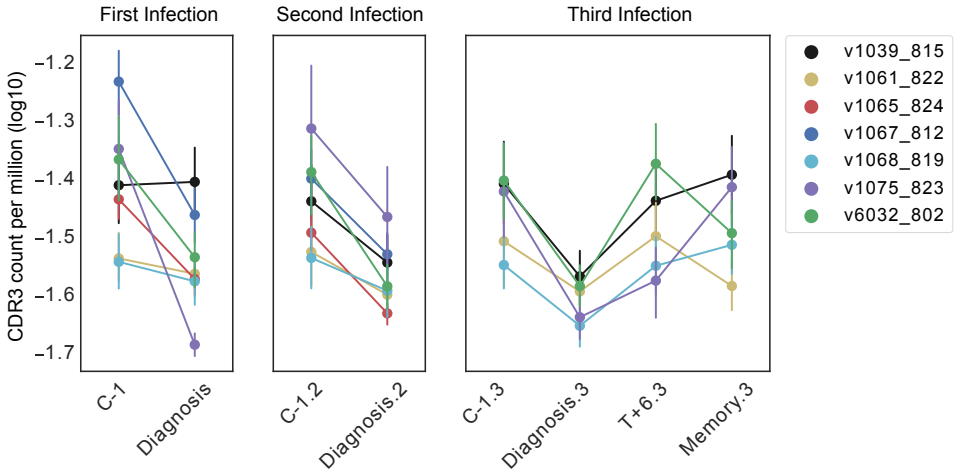


Figure 5.19: Mean clone count per million aligned RNA-seq reads (log10 transformed), of persistent clones as identified form the bulk TCR-seq data. Volunteers v1065_824 and v1067_812 did not undergo a third infection.

The pattern of increased sequence sharing within an individual and minimal sharing between individuals was again evident, with similar trends to that seen with the bulk-TCR for specific individuals (Figure 5.20A). The lower absolute numbers of clones shared across an individual's repertoires (Figure 5.20Bi), is reflective of the shallow depth when mining RNA-seq data for TCR sequences, and this longitudinal sparsity of clones precluded the use of trajectory analysis for this data set. Similarly, no increase in sharing of clonal sequences, or sequences clustered within 1 amino acid mismatch, was evident between volunteers at the novel timepoint of T+6 (Figure 5.20Bii), when analysing data only from Vac63C.

Figure 5.20

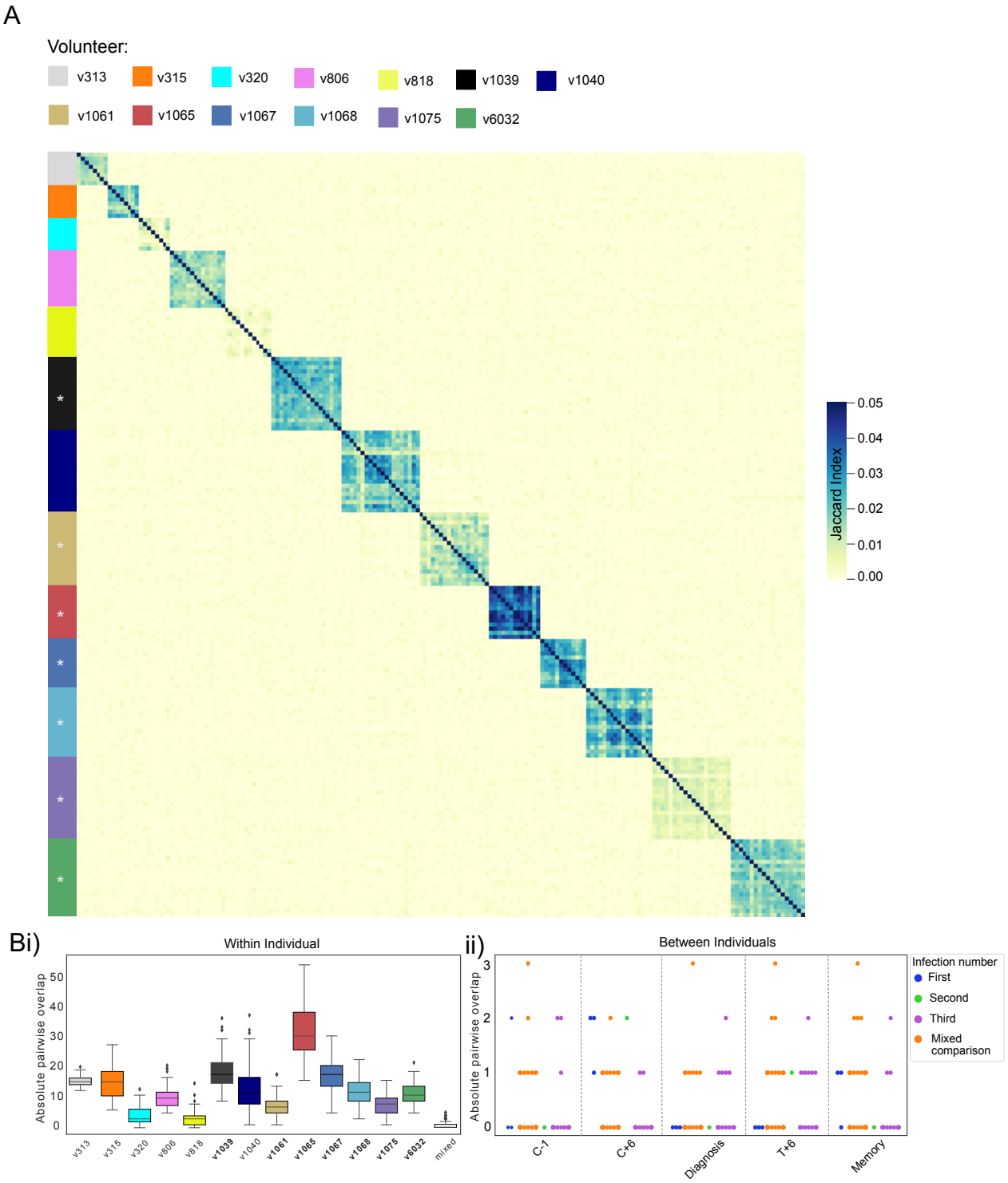


Figure 5.20: A) Heatmap depicts the pairwise Jaccard index between each of the 187 repertoires mined from whole blood RNA-seq data. Data has been ordered by individual (y axis colour bar) longitudinally, with individuals who were included in the bulk TCR-seq data indicated with a star symbol. Bi) The absolute pairwise overlap between repertoires within an individual, unweighted by clonotype frequency. Box plots show the absolute number of TCR sequences shared between 2 timepoints within an individual. Individuals included in the bulk TCR-seq data are indicated in bold, for comparison with Figure 5.17B. ii) Absolute number of shared TCR sequences between volunteers from Vac63C. Points are coloured by infection number, dependent on if the pairwise comparison was from matched infection number or mixed comparison (e.g. first infection vs third).

Given the novel timepoint of T+6 was only present in Vac63C, to determine if data from this timepoint could provide any unique insights, only data from this trial was analysed further. As the mass cytometry data presented in Muñoz Sandoval et al., (In Prep) indicated large amounts of T-cell activation at T+6 in a first infection, not having this timepoint available for the TCR-seq data may have precluded detection of a clonally enriched population expanding at this timepoint but contracted by c+28. Due to the lower availability of TCR sequences in untargeted RNA-seq data, the lymphopenia at day of diagnosis was evident in this data set, with lower numbers of clones detected at this timepoint (Figure 5.21A). This was observed in a first, second, and third infection (Figure 5.21B), consistent with the lymphopenia count data which showed similar levels of lymphopenia at day of diagnosis regardless of infection number (Appendix 5.9). Whilst most volunteers, despite infection number, had clone counts return to baseline by T+6, volunteer v315 had a four-fold increase in clone counts at this timepoint. Mass cytometry data presented in Muñoz Sandoval et al. (In prep), indicated this volunteer had a very large increase in CD4+ activation at T+6 compared to other volunteers and the increase clone counts here may reflect that activation and proliferation. To determine if the increased count represents a clonally expanded population, diversity of repertoires at key timepoints was calculated. Given the low counts at day of diagnosis resulting in uneven repertoires, for normalisation, the mean diversity of each repertoire was calculated by recursively down sampling each repertoire to 200 clones using VDJTools software (Shugay et al., 2015), rather than using the entropy curves. An overall general trend of repertoires being more diverse at day of diagnosis was again evident, with return to near baseline levels at T+6 for most volunteers (Figure 5.22, Figure 5.23A, B). Volunteer v315, however, was a clear outlier at T+6, with a highly expanded repertoire at this timepoint. Tracking the top 10 clones at baseline and T+6 in this volunteer, indicated that at T+6 this individual had a highly expanded repertoire consisting of clones mostly not detected at any other timepoint, with overall return of baseline clones by memory timepoint (Figure 5.22, Figure 5.23Ci). This could indicate a specific responding population at T+6, but as this is only evident in one volunteer, it is hard to draw firm conclusions. The other 2 volunteers undergoing a first infection in Vac63C, v313 and v320, had too little clonal sharing within their repertoires to be able to track clones through time, although volunteer v320 did have a more clonal repertoire at T+6, though to a much lesser extent than v315. No expanded populations at T+6 were evident in individuals

undergoing a second or third infection, where for example with volunteer v1039, baseline clones predominated at T+6 (Figure 5.22, Figure 5.23Cii). It is tempting to suggest that this could indicate a secondary and tertiary response of lower magnitude, as seen in the P. chabaudi work presented in Chapter 3 of this thesis, and as seen for example following secondary vaccination (Minervina et al., 2020c). However, comparing differences in clonality at this timepoint between individuals undergoing a first and third infection cannot be undertaken with this data set, as data for T+6 is not longitudinal, and differences in an individual's background clonality make cross-sectional comparisons redundant with so few volunteers. Overall, analysing this timepoint suggests in some individuals a responding population may be detectable at T+6, and future longitudinal studies should therefore incorporate this timepoint to determine if the degree of expansion observed in volunteer v315 is a more general finding in a larger population. Including this timepoint in longitudinal studies will also allow the question of whether a responding population at T+6 is reduced in clonality/magnitude in re-challenge compared to a first infection within an individual.

Figure 5.21

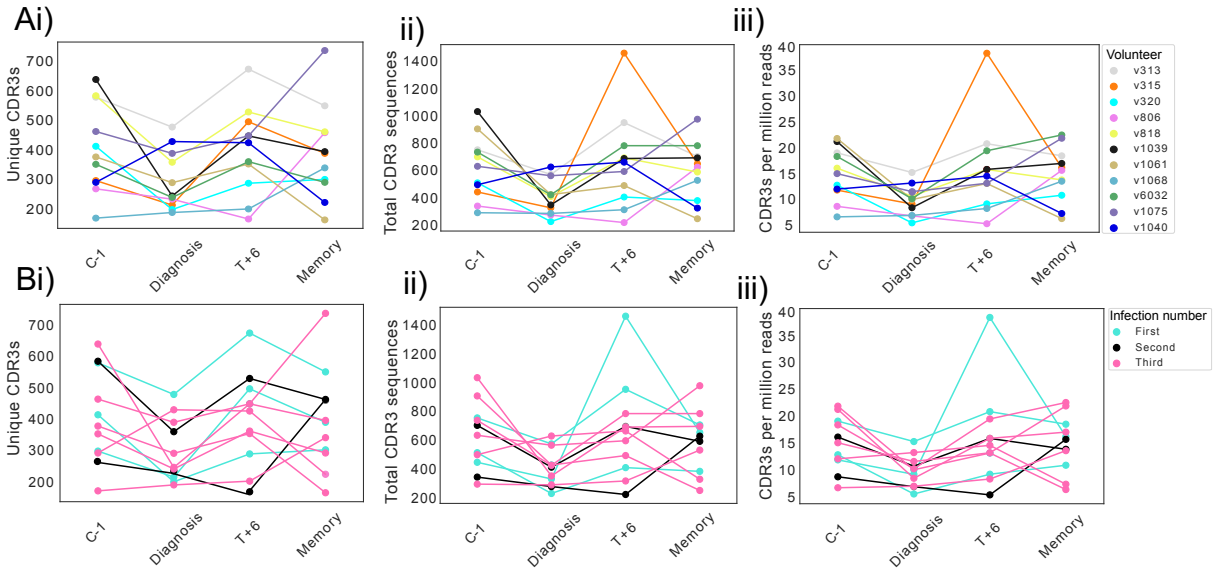


Figure 5.21: Clone counts from the key timepoints from volunteers in Vac63C: Ai) Unique functional CDR3 sequences, ii) total number of functional CDR3 sequences and iii) number of functional CDR3 sequences per million RNA-seq reads, with each colour representing an individual volunteer. B) Same data as for (A) but coloured by infection number. For Vac63C, 3 individuals (light blue) were undergoing a first infection, 2 individuals (black) were undergoing a second CHMI infection and 6 individuals (pink) were undergoing a third infection.

Figure 5.22

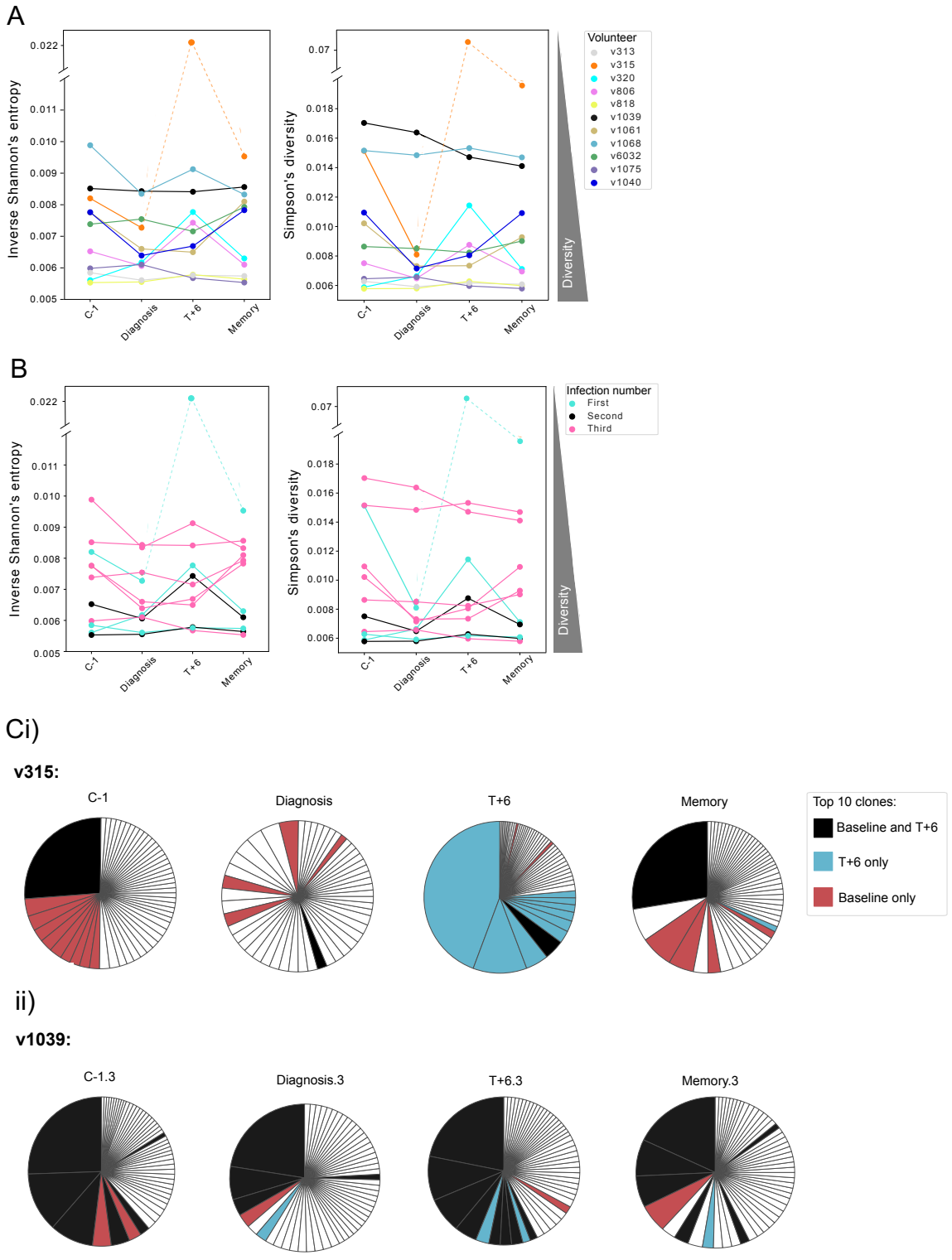


Figure 5.22: Inverse Shannon's entropy and Simpson's diversity index from key timepoints of the Vac63C RNA-seq data, coloured by (A) individual, and (B) infection number. Points represent the mean diversity metric, as determined by recursively down sampling each repertoire to 200 unique clones using VDJTools. C) Top 50 clones from the RNA-seq data set at key timepoints, for volunteers (i) v315 and (ii) v1039. Each segment represents an individual clone. The top 10 clones at baseline are coloured red, top 10 at T+6 are coloured blue, and clones are coloured black if they are present in the top 10 at both baseline and T+6 for that volunteer. For v1039, the top 10 clones from baseline of a first infection were used to colour clone segments.

5.5. Discussion

T-cells are known to play a critical role in the immune response against the pathological blood-stage of malaria (reviewed in (Nlinwe *et al.*, 2018; Kurup, Butler and Harty, 2019)), though they have also been associated with severe disease. However, much still remains unknown regarding the acquisition of T-cell mediated immunity against *Plasmodium,* including the precise clonal dynamics of the response to a first infection in which they are proposed to also play a role in severe disease (Crompton et al., 2014; Mandala et al., 2017; Muñoz Sandoval et al., (In Prep)). In this study, I have used high-throughput TCRβ sequencing to document the clonal T-cell kinetics in both a primary and secondary blood-stage *P. falciparum* CHMI. At time of writing, this is the first study to examine longitudinal CHMI samples using this powerful technique. I firstly show that despite the marked lymphopenia induced in both infections, the TRBV and TRBV/J gene repertoires of individuals remained remarkably stable over time, indicating that no particular TRBV gene is specifically recruited out of circulation over others, nor is it expanded post-treatment. Consequently, no specific response is detectable at the V-gene level, and there is no evidence to support superantigen activity despite the breadth and magnitude of the T-cell responses induced. Furthermore, an individual's long-standing peripheral TRBV repertoire appears to be unaltered by a primary or secondary homologous *P. falciparum* CHMI. This latter finding is perhaps not unexpected given these individuals have been drug-treated and there is therefore no ongoing antigen exposure, which has previously been shown to drive hyper-expansion of memory clones in other chronic infections (Chang *et al.*, 2020; Schober *et al.*, 2020). Whether this picture would be the same in an endemic setting where chronic *Plasmodium* exposure does occur, remains an intriguing question for future studies.

One of the main advantages of truly longitudinal TCR repertoire data is that it allows tracking of private responses within an individual as well as detection of shared public responses between individuals challenged with the same pathogen (Minervina et al., 2020a; Pogorelyy et al., 2018). Mass cytometry analysis carried out in *P. vivax* CHMI (Bach *et al.*, 2021) and in the third *P. falciparum* CHMI trial Vac63C (Muñoz Sandoval et al., (In Prep)), indicated T-cell activation was only evident in the periphery several

days after drug treatment, and it is therefore post-treatment timepoints where I hypothesised I would expect to see specific responding clones either expanded or established as memory populations. However, within an individual, no specific population that was enriched post-infection relative to baseline in a first or second infection was detected, nor was clonality increased. This could be due to several factors. Firstly, given the broad recruitment of T-cell populations, the repertoires released back in to circulation post-treatment may be too polyclonal for a specific response to be detected. Longitudinal TCRβ repertoire studies that have previously detected specific responding populations using trajectory analysis have only been undertaken for viral infections or post-vaccination, where specific kinetics of the response are well documented and the pathogen much less antigenically complex compared to the *Plasmodium* parasite (Minervina et al., 2020a; Minervina et al., 2020b; Pogorelyy et al., 2018; Schultheiß et al., 2020). Similarly, the highly heterogenous response of all T-cell lineages being recruited may preclude detection of specific clonal dynamics without pre-sorting of T-cell subsets. Unsorted PBMCs have previously been sufficient to detect specific clonal dynamics in other diseases, however. Secondly, it cannot be ruled out that by not having a timepoint closer to drug-treatment than c+28, a specific expanded population of T-cells as they return to the periphery has been missed. The timepoint T+6 was only included in the last clinical trial Vac63C. Thus, this timepoint was not available for the TCR-seq work. Nevertheless, if specific responses had become established as memory responses, these should be detectable through both similarity and trajectory analysis. That they weren't, indicates that a dominant clonal memory response is not established post-CHMI. More subtle changes in repertoire architecture may require deeper sequencing depth, and/or the sorting of memory sub-populations, in order to be detectable.

No specific 'public' pathogen-specific response between individuals was evident, either at the clonal or motif level. At the clonal level, this was not unexpected given individuals were not matched by HLA type. At the motif level, public responses have been detected between unmatched individuals including to more complex pathogens such as mycobacteria (Glanville *et al.*, 2017; Huang *et al.*, 2020). However, these studies have had larger cohorts of individuals (n=22 and n=58 respectively) and are

detecting public responses driven by chronic antigen exposure. It could be our sample size is too small to detect shared responses from an antigenically complex pathogen. Alternatively, the shared responses detected in the *P.chabaudi* work in this thesis were only evident in the acute phase of infection and not enhanced upon re-challenge. Not having T+6 or a similar timepoint may have therefore precluded detection of a temporally similar acute public response. However, again if such a response had remained prevalent in the repertoire at later timepoints, this should have been detectable through both similarity and GLIPH2 analysis between individuals.

Interestingly, persistent clones - which include expanded likely memory cells - were reduced at day of diagnosis. These included sequences that have previously been identified as epitope-specific to common viral antigens. Movement of these previously expanded clones out of circulation is reflected in the increased diversity of repertoires at day of diagnosis, in both infections. These findings were confirmed in the RNA-seq data set and indicate that these clones are being recruited from the periphery in a non-specific manner, regardless of infection number. This is in agreement with available cell-subset count data for these volunteers, as well as mass cytometry data with finer precision of subsets from related *P. vivax* CHMI data presented in Bach *et al.* (2021), which showed indiscriminate recruitment of T-cells from circulation independent of type or function. Such non-specific recruitment has previously been hypothesised to disrupt the developing immune response and prevent sterile immunity being induced in mycobacteria TB granulomas (Millar *et al.*, 2021). Recruitment is primarily driven by CXCL9 and CXCL10 (Groom and Luster, 2011), which have been shown to peak at day of diagnosis in *P. falciparum* CHMI, as the secreted effectors of an emergency innate myeloid response induced by the parasite (Milne *et al.*, 2021). This innate myeloid response has been described as ‘hardwired’, occurring across all three infections of the volunteers examined in this thesis (Muñoz Sandoval et al., (In Prep)), and likely accounts for the similar levels of non-specific recruitment seen in the TCR-seq data independent of infection number. Assuming these clones are trafficked to the spleen, the major site of the developing immune response in a malaria infection (del Portillo *et al.*, 2012b), what happens to them there if they are not being recirculated to the periphery? One hypothesis is that they may undergo bystander activation. This

phenomenon - where T-cells are activated in an inflammation-dependent manner without cognate antigen, predominantly affecting memory T-cells – has been reported previously in infections that produce systemic inflammation (Whiteside, Snook, Williams, *et al.*, 2018), and was suggested to contribute to the highly polyclonal splenic *P. chabaudi* $CD4^+$ effector response presented in Chapter 3 of this thesis (Smith *et al.*, 2020). Similarly, the magnitude and breadth of activation across all major T-cell subsets at T+6 following a first infection in both *P. vivax* and *P. falciparum* CHMI, has been hypothesised to be driven by bystander activation (Bach et al., 2021; Muñoz Sandoval et al., (In prep)). The most common driver is dysregulated inflammation, with elevated cytokine levels including IL-6, IL-7, IL-12 and IL-15, IL-18 and IL-21 all implicated (Kim *et al.*, 2018; Whiteside, Snook, Williams, *et al.*, 2018). Analysis of both plasma proteins and whole-blood RNA-seq data from the volunteers in Vac63C showed IFN driven systemic inflammation was induced by CHMI, with levels peaking at day of diagnosis. The level of systemic inflammation induced was not attenuated by re-challenge and remained comparable across infections (Muñoz Sandoval *et al.,* (In prep)), making bystander activation plausible. If these clones do not become activated following recruitment, either due to lack of cognate p:MHC or they are not bystander activated, they could still potentially disrupt the developing immune response due to spatial overcrowding limiting interactions between specific TCR's and their cognate p:MHC as suggested to be the cause for non-sterilising immunity in mycobacterial infections by Millar *et al.* (2021).

If bystander activation of these non-specific persistent clones trafficked to the spleen was occurring, how would this manifest in the TCR repertoire? If they were proliferating, it would be expected that tracking these clones through time would show their expansion at post-treatment timepoints when they return to peripheral circulation. In the bulk TCR data, at c+28 this was not evident, with persistent clones returned to baseline levels in both a first and second infection by this timepoint. This does not preclude bystander activation - it may have been that expansion of these populations was missed by not having the earlier post-treatment timepoint of T+6 for this data. However, expansion of these clones at T+6 in the RNA-seq data set, in which these volunteers are undergoing their third infection, was also not evident. This does not rule

out that such an observation could have been made following a first infection. Of the 3 volunteers undergoing a first infection in the RNA-seq data set, 2 out of the 3 (v313 and v320) do not have expanded persistent clones detectable in their repertoire, and as such their data is too sparse to follow persistent clones longitudinally with any certainty. The data for volunteer v315 is also sparse given the depth of RNA-seq data, but as this volunteer has a more expanded repertoire at baseline, this allowed tracking of the most dominant clone at baseline, which had returned to baseline levels at T+6 without being expanded further, although it had been superseded by new more dominant clones at T+6, which weren't present at other timepoints. The data from this volunteer hints at a responding expanded population at this timepoint, warranting its inclusion in future studies, but ultimately no novel firm conclusions regarding a first infection can be drawn from the RNA-seq data, given low sequencing depth and small cohort numbers.

Another explanation for persistent clones not being clonally expanded post-treatment if bystander activation is occurring, is that non-specific recruitment of T-cells does not necessarily result in their proliferation. For example, non-specific memory $CD8^+$ cells in a mouse model of *Listeria monocytogenes* were trafficked to the spleen in the acute phase of infection where they were activated - staining positive for cytotoxic molecules including granzyme B - but not staining for Ki-67, a marker of cellular proliferation (Maurice *et al.*, 2019). Bystander activation could be a potential cause for the variation in T-cell activation between individuals seen during a first infection in the accompanying whole transcriptome RNA-seq data (Muñoz Sandoval et al., (In Prep)). Volunteer v315 had a less diverse, more expanded repertoire at baseline compared to v313 and v320, indicating a greater repertoire proportion of hyperexpanded likely memory clones, and in turn this volunteer showed by far the largest amount of activation at T+6 in a first infection, predominantly of the $CD4^+$ effector memory population (Muñoz Sandoval et al., (In Prep)). It is tempting therefore to speculate that if bystander activation does occur, the landscape of an individual's pre-existing memory repertoire may influence their T-cell response to a *Plasmodium* infection, and this would be an interesting avenue to explore in future studies. Further experimental proof to confirm bystander activation for example through the use of MHC-tetramer

staining of activated populations as seen with the use of EBV-specific tetramers in Port *et al.* (2020), is warranted. Similarly, combining longitudinal TCR-seq with single-cell RNA-seq would confirm the non-specific activation of pre-existing memory populations.

In summary, examining the longitudinal TCRβ repertoire dynamics during a primary and secondary *P. falciparum* CHMI, provides evidence for non-specific recruitment of established T-cell clones from the peripheral circulation. This dynamic is replicated in a primary, secondary and tertiary infection, and is not attenuated by re-challenge. Understanding the consequences of this non-specific trafficking and whether or not it shapes an individual's response to a *Plasmodium* infection warrants further investigation. That no pathogen-specific clonal responses were detected, may reflect limitations in study design rather than true absence of such responses. Future CHMI TCR studies that aim to discover public or private specific responses, should ideally include a finer precision of timepoints immediately following treatment at day of diagnosis, such as T+6, and sort T-cell sub-populations of interest to reduce background repertoire noise.

5.6. Appendix

Appendix 5.1

A

Sequence: 1 Jurkat2_P5

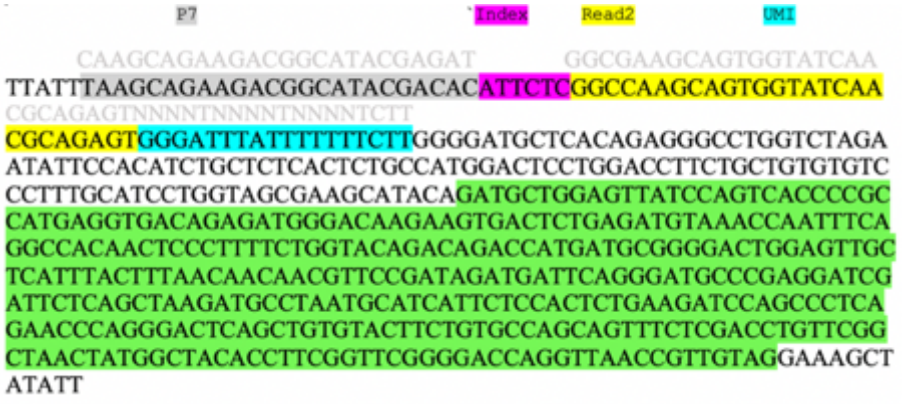
Analysed sequence length: 554.
Sequence analysis category: 3 (complementay reverse_no indel search).
Complementary reverse sequence compared with the *Homo sapiens* (human) TR set from the IMGT reference directory

```
>Jurkat2_P5
ttatttaagcagaagacggcatacgacacattctcggccaagcagtggtatcaacgcaga
gtgggatttatttttttcttggggatgctcacagagggcctggtctagaatattccacat
ctgctctcactctgccatggactcctggaccttctgctgtgtgtccctttgcatcctggt
agcgaagcatacagatgctggagttatccagtcaccccgccatgaggtgacagagatggg
acaagaagtgactctgagatgtaaaccaatttcaggccacaactccctttttctggtacag
acagaccatgatgcggggactggagttgctcatttactttaacaacaacgttccgataga
tgattcagggatgcccgaggatcgattctcagctaagatgcctaatgcatcattctccac
tctgaagatccagccctcagaacccagggactcagctgtgtacttctgtgccagcagttt
ctcgacctgttcggctaactatggctacaccttcggttcggggaccaggttaaccgttgt
aggaaagctatatt
```

Result summary:	Productive TRB rearranged sequence (no stop codon and in-frame junction)		
V-GENE and allele	Homsap TRBV12-3*01 F	score = 1375	identity = 100.00% (276/276 nt)
J-GENE and allele	Homsap TRBJ1-2*01 F	score = 240	identity = 100.00% (48/48 nt)
D-GENE and allele by IMGT/JunctionAnalysis	Homsap TRBD1*01 F	D-REGION is in reading frame 1	
FR-IMGT lengths, CDR-IMGT lengths and AA JUNCTION	[26.17.38.10]	[5.6.14]	CASSFSTCSANYGYTF

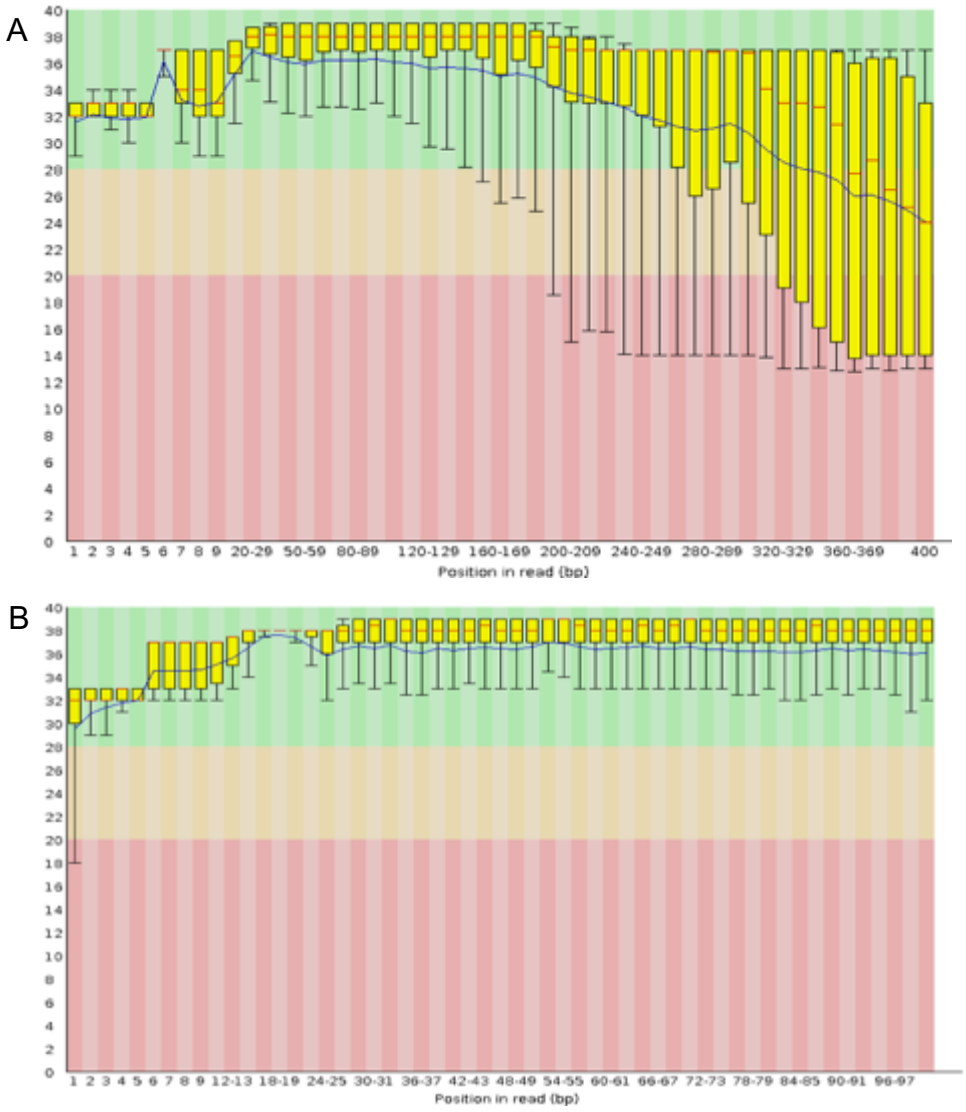
B

C



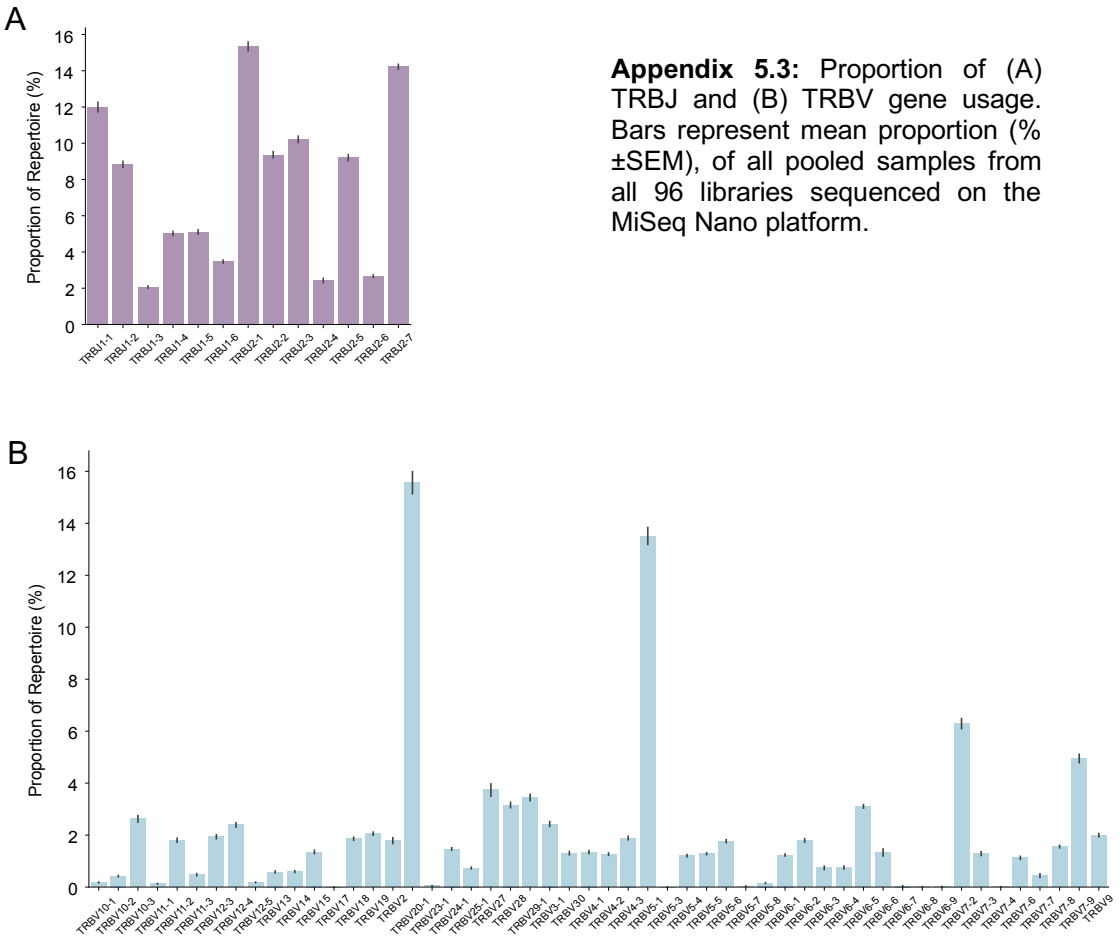
Appendix 5.1: A) Summary alignment results from a BLAST alignment of Sanger sequencing results of the TCRβ chain of jurkat cells, aligned using IMGT V-QUEST. B) Expected primer amplicon orientation. C) Orientation of primers overlayed on the Jurkat sanger sequence.

Appendix 5.2



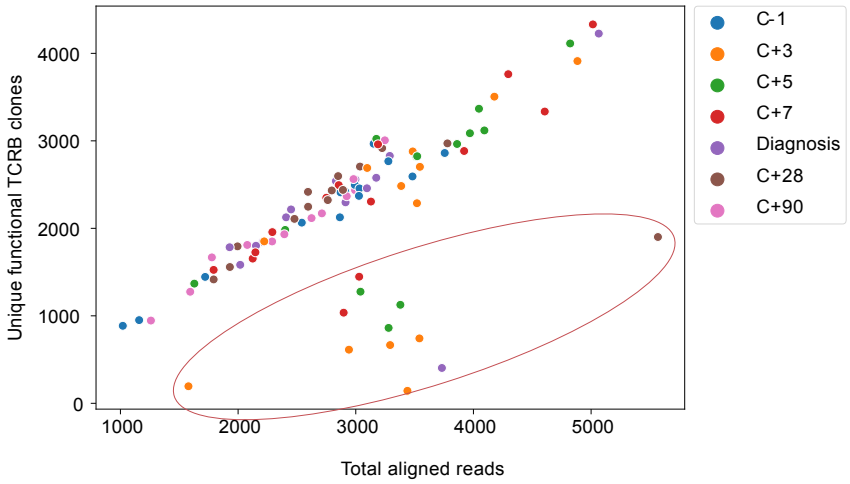
Appendix 5.2: FastQC plots from the MiSeq Nano run for (A) Read 1 and (B) Read 2 from representative samples

Appendix 5.3



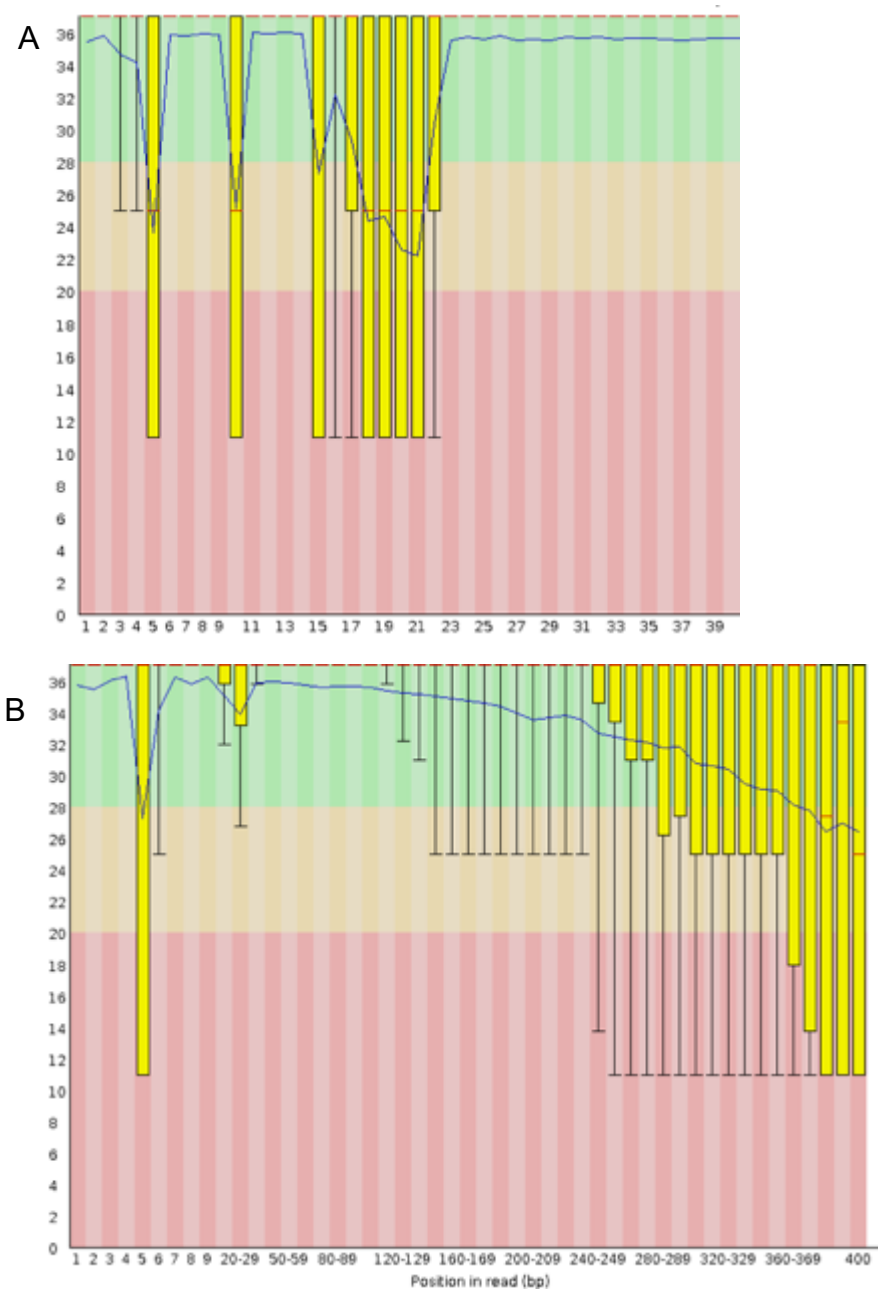
Appendix 5.3: Proportion of (A) TRBJ and (B) TRBV gene usage. Bars represent mean proportion (% ±SEM), of all pooled samples from all 96 libraries sequenced on the MiSeq Nano platform.

Appendix 5.4



Appendix 5.4: Scatterplot depicts the number of unique functional TCR clones and their sequencing depth form the MiSeq Nano run. Points are coloured according to timepoint, and samples suspected to have failed library prep from batch 7 are circled in red.

Appendix 5.5



Appendix 5.5: FastQC scores from the NovaSeq run for (A) Read 2 and (B) Read 1. For (A), base positions have been ungrouped to show quality at each individual base, and only the first 40 bases are shown.

Appendix 5.6

Sample_ID	Cell_type	Timepoint	Infection_number	Filtered	Unique_CDR3	Unique_UMIS	Functional TCR
S1	v1039_815	c-1	First	N	11533	15555	2745679
S2	v1039_815	c+3	First	N	25135	36461	3697762
S3	v1039_815	c+5	First	N	8048	10592	2294259
S4	v1039_815	c+7	First	N	19228	25697	2516613
S5	v1039_815	diagnosis	First	N	8590	11057	2013272
S6	v1039_815	c+28	First	N	7944	10748	1660032
S7	v1039_815	c+90	First	N	10498	13923	2761017
S8	v1039_815	c-1	Second	N	11007	14178	2645269
S9	v1039_815	c+3	Second	Y	472	1143	3191786
S10	v1039_815	c+5	Second	N	5321	7038	1537597
S11	v1039_815	c+7	Second	Y	899	1403	2576156
S12	v1039_815	diagnosis	Second	N	6644	8353	2587138
S13	v1039_815	c+28	Second	N	28398	36994	2777479
S14	v1039_815	c+90	Second	N	4507	5484	1293060
S15	v1061_822	c-1	First	N	6829	8588	2758513
S16	v1061_822	c+3	First	Y	401	1168	2822706
S17	v1061_822	c+5	First	N	8117	9939	3810756
S18	v1061_822	c+7	First	N	7592	9190	2154760
S19	v1061_822	diagnosis	First	N	10230	12582	2688856
S20	v1061_822	c+28	First	N	23279	26740	2145930
S22	v1061_822	c-1	Second	N	13323	16333	1217102
S23	v1061_822	c+3	Second	N	3405	4044	3056904
S24	v1061_822	c+5	Second	Y	924	1416	2879110
S25	v1061_822	c+7	Second	N	6011	7624	1597815
S26	v1061_822	diagnosis	Second	N	15732	22404	2599390
S27	v1061_822	c+28	Second	N	11988	14395	2091243
S28	v1061_822	c+90	Second	N	18587	21335	1568380
S29	v1065_824	c-1	First	N	33348	48363	3056124
S30	v1065_824	c+3	First	Y	113	742	1389881
S31	v1065_824	c+5	First	N	14305	19729	2209002
S32	v1065_824	c+7	First	Y	1549	2366	2812939
S33	v1065_824	diagnosis	First	N	31095	40463	3123764
S34	v1065_824	c+28	First	Y	1521	2734	4563614
S36	v1065_824	c-1	Second	N	14918	21027	2636739
S37	v1065_824	c+3	Second	Y	60	1798	3341306
S38	v1065_824	c+5	Second	N	14790	23248	4302461
S39	v1065_824	c+7	Second	N	6936	10420	2822214
S40	v1065_824	diagnosis	Second	N	30644	38021	2641952

S41	v1065_824	c+28	Second	N	4623	6361	1667767
S42	v1065_824	c+90	Second	N	31409	48316	2195707
S43	v1067_812	c-1	First	N	8022	10593	2456132
S44	v1067_812	c+3	First	N	13895	17825	3294643
S45	v1067_812	c+5	First	N	14294	19170	3712064
S46	v1067_812	c+7	First	N	24062	31104	2545793
S47	v1067_812	diagnosis	First	N	11224	13140	2125054
S48	v1067_812	c+28	First	N	12473	15963	2403556
S49	v1067_812	c+90	First	N	16251	20316	3234522
S50	v1067_812	c-1	Second	N	16506	22346	1553056
S51	v1067_812	c+3	Second	N	7356	11439	3162897
S52	v1067_812	c+5	Second	N	10260	17957	3559953
S53	v1067_812	c+7	Second	N	4961	8446	1755551
S54	v1067_812	diagnosis	Second	N	8210	10293	2685430
S55	v1067_812	c+28	Second	N	11616	14305	2249586
S56	v1067_812	c+90	Second	N	6440	8084	2541857
S57	v1068_819	c-1	First	N	7426	9193	955355
S58	v1068_819	c+3	First	N	11352	22066	2390330
S59	v1068_819	c+5	First	N	16155	21223	3566778
S60	v1068_819	c+7	First	N	8601	12343	3917706
S61	v1068_819	diagnosis	First	N	24024	29685	4346777
S62	v1068_819	c+28	First	N	11365	14862	2934430
S63	v1068_819	c+90	First	N	20370	27628	1414181
S64	v1068_819	c-1	Second	N	8536	12133	3417394
S65	v1068_819	c+3	Second	N	7301	9935	3077467
S66	v1068_819	c+5	Second	N	1164	1616	2411736
S67	v1068_819	c+7	Second	N	6635	9473	3689299
S68	v1068_819	diagnosis	Second	N	4567	5673	1925653
S69	v1068_819	c+28	Second	N	10397	14574	3085533
S70	v1068_819	c+90	Second	N	9199	12277	2796591
S71	v1075_823	c-1	First	N	27514	30510	2979764
S72	v1075_823	c+3	First	N	3389	4119	3291704
S73	v1075_823	c+5	First	N	44020	49891	3039279
S74	v1075_823	c+7	First	N	18037	20904	4466780
S75	v1075_823	diagnosis	First	N	6332	7576	2743219
S76	v1075_823	c+28	First	N	15556	17911	3090556
S77	v1075_823	c+90	First	N	21474	23649	2833321
S78	v1075_823	c-1	Second	N	4546	5628	2525842
S79	v1075_823	c+3	Second	Y	1	2	292
S80	v1075_823	c+5	Second	Y	603	1208	2606582
S81	v1075_823	c+7	Second	N	26648	29553	2965561

S82	v1075_823	diagnosis	Second	N	7036	8120	1888495
S83	v1075_823	c+28	Second	N	17571	19621	2616782
S84	v1075_823	c+90	Second	N	4656	5506	2130265
S85	v6032_802	c-1	First	N	18798	25321	3283389
S86	v6032_802	c+3	First	N	15109	19940	4158643
S87	v6032_802	c+5	First	N	33813	44825	4288243
S88	v6032_802	c+7	First	N	27750	35783	3824492
S89	v6032_802	diagnosis	First	N	15731	18219	1878033
S90	v6032_802	c+28	First	N	10528	13110	2746372
S91	v6032_802	c+90	First	N	7196	12117	2359077
S92	v6032_802	c-1	Second	N	6328	8236	3289083
S93	v6032_802	c+3	Second	Y	2414	4952	3213338
S94	v6032_802	c+5	Second	N	8347	10995	3495455
S95	v6032_802	c+7	Second	N	5231	6678	2015650
S96	v6032_802	diagnosis	Second	Y	223	1124	3365133
S97	v6032_802	c+28	Second	N	15342	18835	2590083
S98	v6032_802	c+90	Second	N	3529	5299	1543036

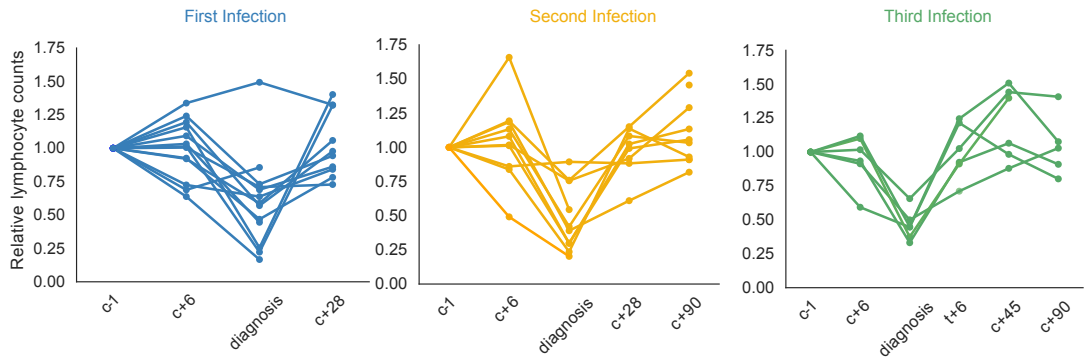
Appendix 5.7: Volunteer Attributes

Volunter_id_63A	Volunteer_id_63B	Volunteer_id	Gender	Age	BMI	Smoking	CMV	EBV	Ethnicity	PMR	Treatment
v1039	815	v1039_815	F	32	26.2	N	Positive	Positive	White Portuguese	8.6	Riamet
v1061	822	v1061_822	M	24	25.2	Y	Negative	Positive	White British	10.87	Riamet
v1065	824	v1065_824	M	20	20.8	N	Positive	Not tested	White British	11.29	Riamet
v1067	812	v1067_812	M	21	26.7	N	Negative	Not tested	White British	9.01	Malarone
v1068	819	v1068_819	M	33	25.4	N	Positive	Positive	South Korean	13.15	Riamet
v1075	823	v1075_823	M	22	33.4	N	Negative	Positive	White British	6.1	Riamet
v6032	802	v6032_802	M	22	19.5	N	Negative	Positive	White British	9.61	Riamet

Appendix 5.8: Volunteer HLA Type

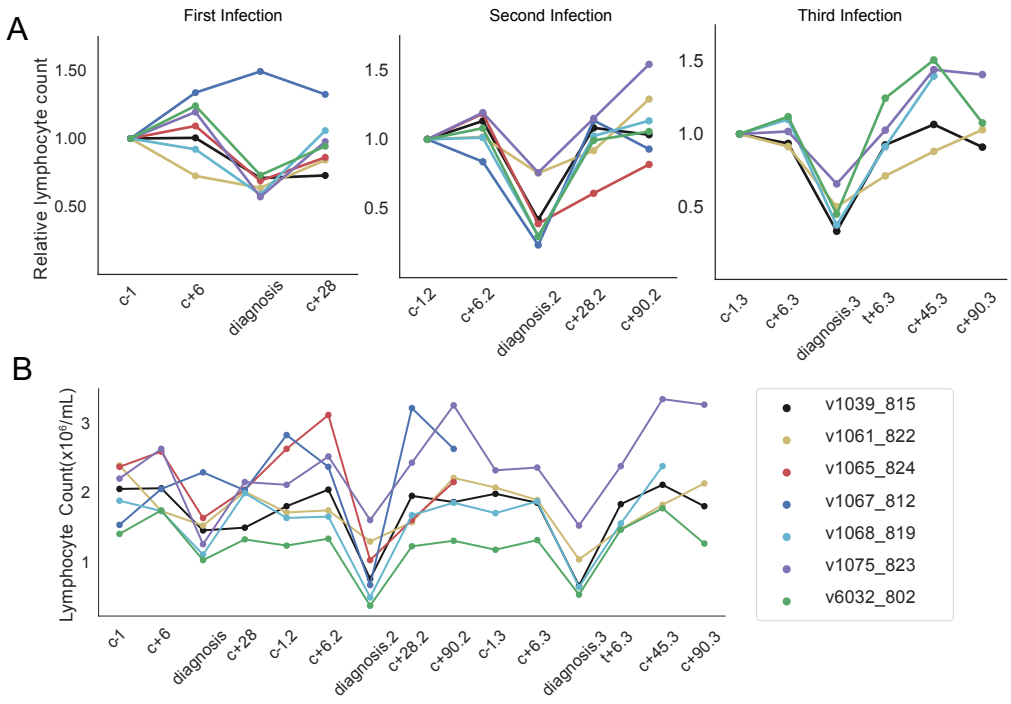
Volunteer_id	HLA_A	HLA_B	HLA_C	HLA_DRB1	HLA_DRB3	HLA_DRB4	HLA_DRB5	HLA_DQB1
v1039_815	*69 *68	*38 *15	*7 *12	*13 *13	*2 *1	---	---	*6
v1061_822	*68 *1	*57 *27	*6 *2	*7	---	*1 *1	---	*3 *2
v1065_824	*1 *26	*8 *7	*7 *7	*3	*1	---	---	*2
v1067_812	*32 *25	*58 *40	*2 *7	*4 *13	*2	*1	---	*3 *6
v1068_819	*11 *3	*51 *44	*5 *14	*9 *13	*1	*1	---	*6 *3
v1075_823	*29 *1	*8 *7	*7 *7	*7 *14	*2	*1	---	*2 *5
v6032_802	*11 *26	*44 *44	*4 *5	*7 *13	*2	*1	---	*2 *6

Appendix 5.9



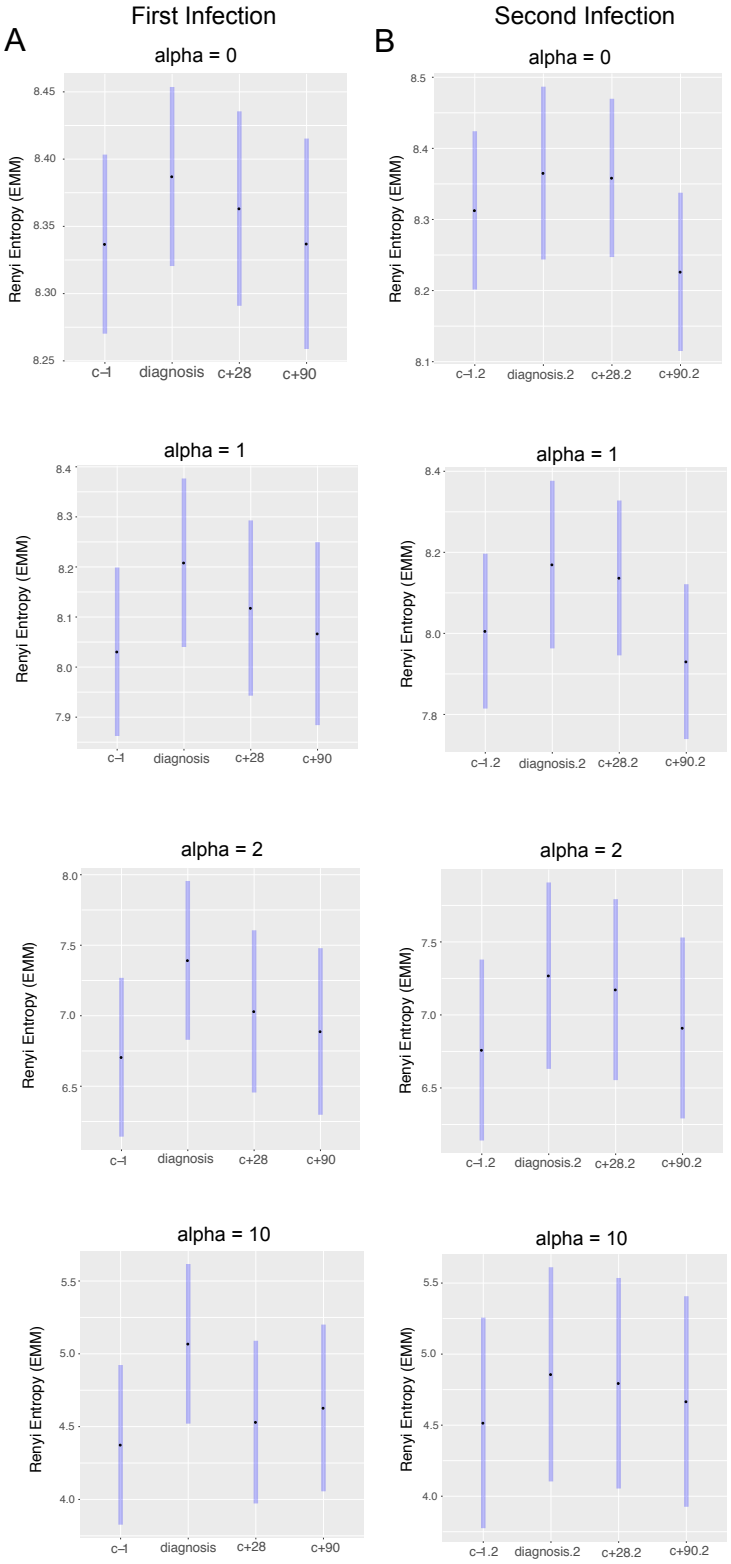
Appendix 5.9: Lymphocyte counts relative to baseline (C-1 for each infection), for all volunteers who underwent a *P. falciparum* blood-stage CHMI infection as part of the control arm for Vac063A, Vac063B and Vac063C.

Appendix 5.10

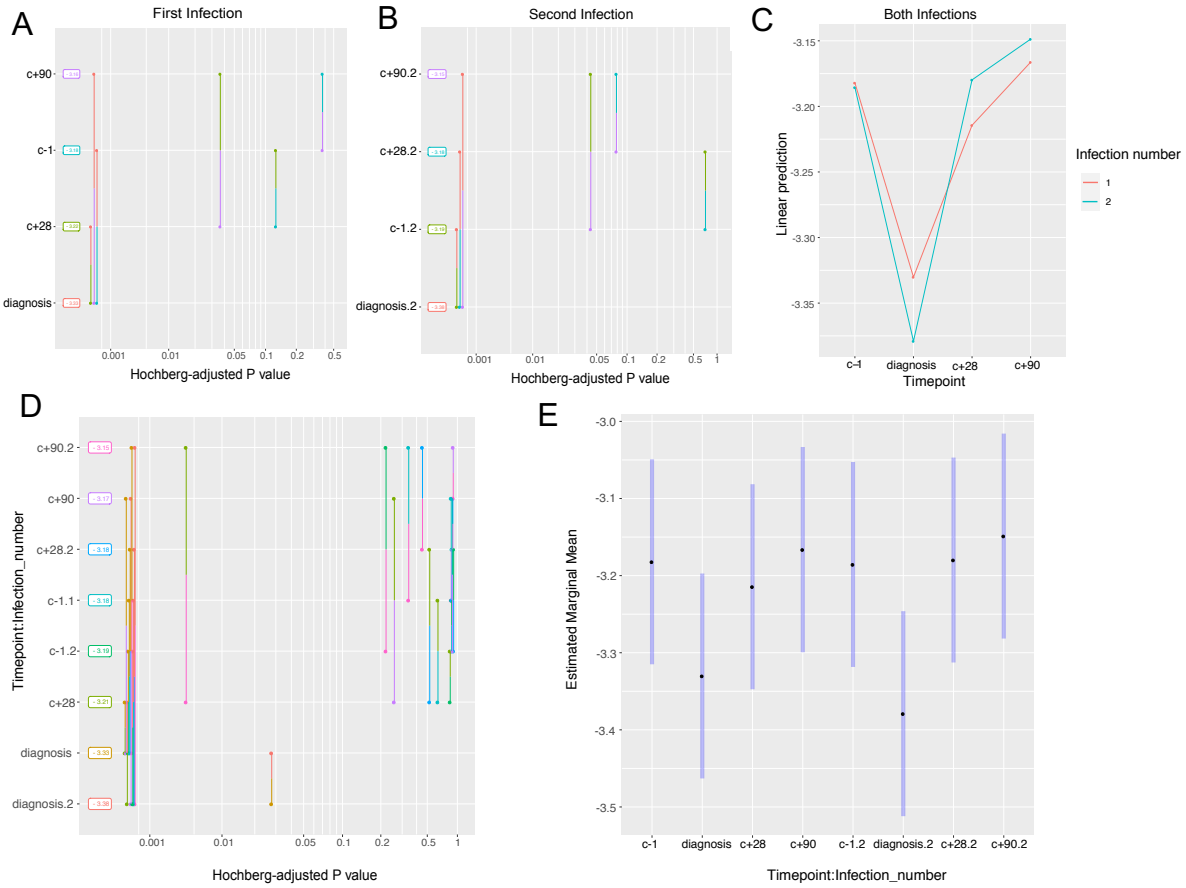


Appendix 5.10: Lymphocyte counts (A) relative to baseline (C-1 for each infection), and (B) absolute counts (x10^6/mL) for volunteers undergoing a blood-stage *P. falciparum* CHMI. Data is coloured according to individual volunteer. Data is depicted for first, second (.2) and third (.3) infections. Volunteers v1065_824 and v1067_812 did not undergo a third infection.

Appendix 5.11

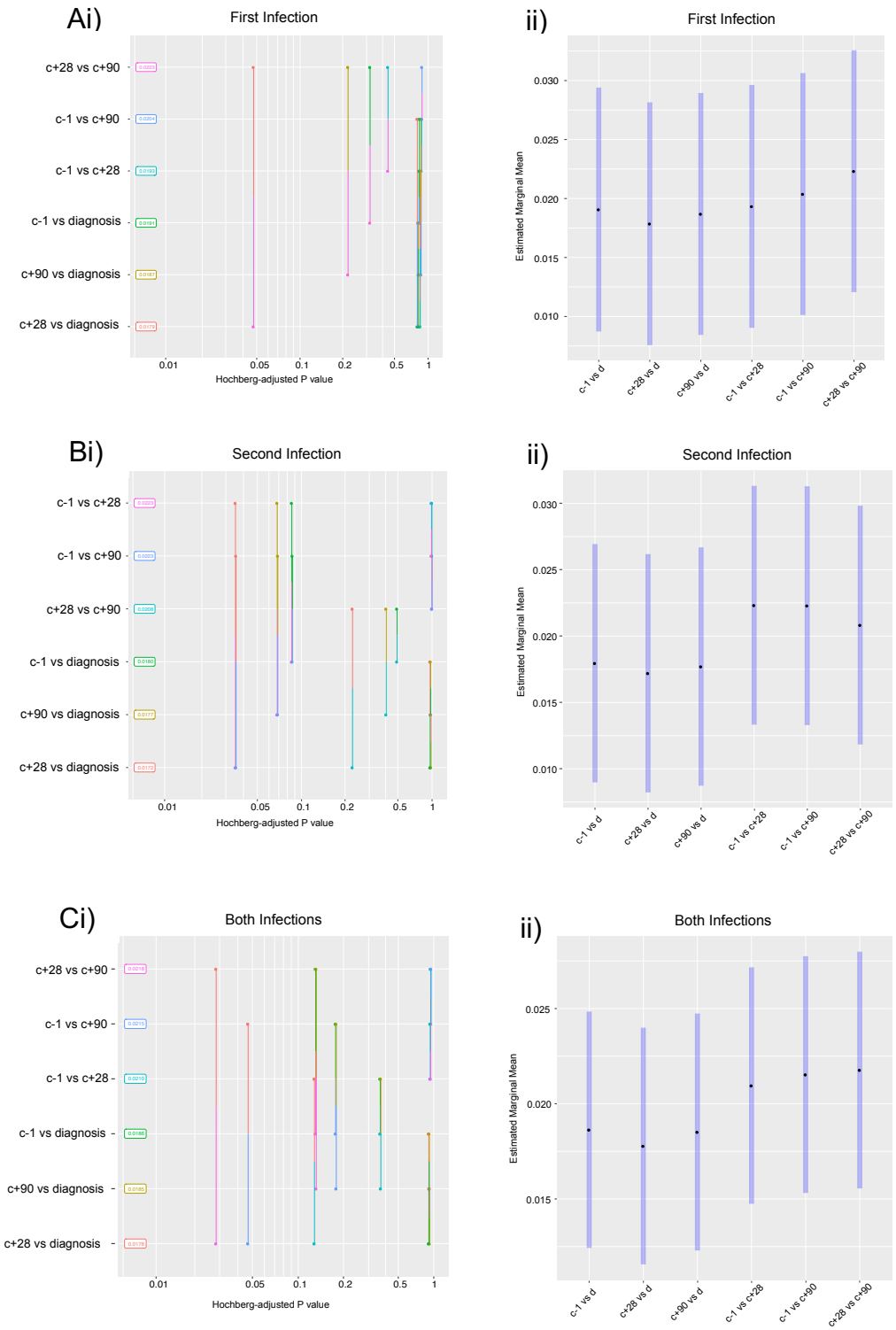


Appendix 5.12



Appendix 5.12: Mean proportion of persistent clones through time was analysed using linear mixed effect models, with timepoint as a fixed effect and both clone and individual as random fixed effects. A) and B) show Benjamini-Hochberg adjusted p-values for pairwise comparisons between key timepoints in a first and second infection respectively. For C) and D), to compare infections, timepoint*infection_number was included as a fixed effect. E) Shows estimated marginal means and 95% CI.

Appendix 5.13



Appendix 5.13: Differences in the pairwise Jaccard index between key timepoints within an individual, were analysed using linear mixed effects models, with pairwise comparison as the response variable, and Volunteer as a random fixed effect for A) first infection B) second infection and C) both infections. For both infections, infection number was considered as a fixed effect. For A, B and C, (i) show the Benjamini-Hochberg adjusted p-values for pairwise comparisons, with (ii) displaying estimated marginal means for each group, with 95% CI.

Appendix 5.14: clone counts from whole blood RNA-seq data

Sample_ID	Volunteer	Timepoint	Vac	Unique_CDR3s	Functional TCR sequences
v313_1	v313	C-1	C	582	768
v313_2	v313	C+6	C	474	630
v313_3	v313	C+7	C	488	646
v313_4	v313	C+8	C	497	686
v313_5	v313	C+9	C	541	690
v313_6	v313	Diagnosis	C	481	587
v313_7	v313	T+6	C	677	967
v313_8	v313	Memory	C	553	719
v315_1	v315	C-1	C	300	459
v315_2	v315	C+6	C	426	623
v315_3	v315	C+7	C	579	873
v315_4	v315	C+8	C	431	628
v315_5	v315	C+9	C	301	400
v315_6	v315	Diagnosis	C	217	343
v315_7	v315	T+6	C	499	1476
v315_8	v315	Memory	C	392	666
v320_1	v320	C-1	C	416	526
v320_2	v320	C+6	C	298	382
v320_3	v320	C+7	C	367	498
v320_4	v320	C+8	C	170	213
v320_5	v320	C+9	C	294	382
v320_6	v320	Diagnosis	C	202	242
v320_7	v320	T+6	C	291	422
v320_8	v320	Memory	C	304	396
v1065_1	v1065	C-1	A	584	1043
v1065_2	v1065	C+6	A	520	799
v1065_3	v1065	C+8	A	529	898
v1065_4	v1065	C+9	A	535	852
v1065_5	v1065	C+10	A	322	527
v1065_6	v1065	Diagnosis	A	436	673
v1065_7	v1065	C-1	B	556	891
v1065_8	v1065	C+6	B	575	1003
v1065_9	v1065	C+8	B	401	675
v1065_10	v1065	C+9	B	372	545
v1065_11	v1065	C+10	B	384	676
v1065_12	v1065	Diagnosis	B	362	485
v1065_13	v1065	Memory	B	333	519
v1067_1	v1067	C-1	A	430	637
v1067_2	v1067	C+6	A	398	586
v1067_3	v1067	C+7	A	381	557

v1067_4	v1067	C+8	A	207	291
v1067_5	v1067	Diagnosis	A	338	462
v1067_6	v1067	C-1	B	434	798
v1067_7	v1067	C+6	B	369	612
v1067_8	v1067	C+7	B	304	520
v1067_9	v1067	C+8	B	271	425
v1067_10	v1067	C+9	B	271	392
v1067_11	v1067	Diagnosis	B	295	484
v1067_12	v1067	Memory	B	476	659
v806_1	v806	C-1	B	365	474
v806_2	v806	C+6	B	485	673
v806_3	v806	C+8	B	401	543
v806_4	v806	C+9	B	234	293
v806_5	v806	C+10	B	245	309
v806_6	v806	Diagnosis	B	215	265
v806_7	v806	C-1	C	272	356
v806_8	v806	C+6	C	427	591
v806_9	v806	C+9	C	380	492
v806_10	v806	C+10	C	279	346
v806_11	v806	C+11	C	344	455
v806_12	v806	Diagnosis	C	237	290
v806_13	v806	T+6	C	170	235
v806_14	v806	Memory	C	462	641
v818_1	v818	C-1	B	410	502
v818_2	v818	C+6	B	450	571
v818_3	v818	C+7	B	366	453
v818_4	v818	C+8	B	206	272
v818_5	v818	Diagnosis	B	173	196
v818_6	v818	C-1	C	587	716
v818_7	v818	C+6	C	395	475
v818_8	v818	C+7	C	303	363
v818_9	v818	C+8	C	435	520
v818_10	v818	Diagnosis	C	362	425
v818_11	v818	T+6	C	532	709
v818_12	v818	Memory	C	465	605
v1039_1	v1039	C-1	A	665	1008
v1039_2	v1039	C+6	A	473	876
v1039_3	v1039	C+7	A	435	716
v1039_4	v1039	C+8	A	314	482
v1039_5	v1039	Diagnosis	A	455	721
v1039_6	v1039	C-1	B	548	830

v1039_7	v1039	C+6	B	540	810
v1039_8	v1039	C+7	B	410	637
v1039_9	v1039	C+8	B	404	663
v1039_10	v1039	Diagnosis	B	306	456
v1039_11	v1039	C-1	C	642	1048
v1039_12	v1039	C+6	C	385	571
v1039_13	v1039	C+7	C	559	898
v1039_14	v1039	C+8	C	406	636
v1039_15	v1039	C+9	C	299	515
v1039_16	v1039	Diagnosis	C	248	364
v1039_17	v1039	T+6	C	451	704
v1039_18	v1039	Memory	C	398	709
v1061_1	v1061	C-1	A	285	481
v1061_2	v1061	C+6	A	274	480
v1061_3	v1061	C+8	A	141	224
v1061_4	v1061	C+9	A	179	304
v1061_5	v1061	Diagnosis	A	252	403
v1061_6	v1061	C-1	B	242	418
v1061_7	v1061	C+6	B	318	579
v1061_8	v1061	C+8	B	317	497
v1061_9	v1061	C+9	B	193	297
v1061_10	v1061	Diagnosis	B	285	403
v1061_11	v1061	C-1	C	380	921
v1061_12	v1061	C+6	C	300	577
v1061_13	v1061	C+8	C	342	602
v1061_14	v1061	C+9	C	376	641
v1061_15	v1061	C+10	C	215	331
v1061_16	v1061	Diagnosis	C	293	440
v1061_17	v1061	T+6	C	356	506
v1061_18	v1061	Memory	C	167	263
v1068_1	v1068	C-1	A	289	465
v1068_2	v1068	C+6	A	392	685
v1068_3	v1068	C+7	A	324	585
v1068_4	v1068	C+8	A	172	309
v1068_5	v1068	Diagnosis	A	336	515
v1068_6	v1068	C-1	B	286	488
v1068_7	v1068	C+6	B	295	672
v1068_8	v1068	C+7	B	210	609
v1068_9	v1068	C+8	B	312	592
v1068_10	v1068	Diagnosis	B	307	482
v1068_11	v1068	C-1	C	173	307

v1068_12	v1068	C+6	C	319	674
v1068_13	v1068	C+7	C	290	699
v1068_14	v1068	C+8	C	261	440
v1068_15	v1068	Diagnosis	C	192	302
v1068_16	v1068	T+6	C	204	329
v1068_17	v1068	Memory	C	343	544
v6032_1	v6032	C-1	A	427	881
v6032_2	v6032	C+6	A	411	856
v6032_3	v6032	C+7	A	326	659
v6032_4	v6032	C+8	A	343	880
v6032_5	v6032	C+9	A	297	615
v6032_6	v6032	Diagnosis	A	378	645
v6032_7	v6032	C-1	B	430	753
v6032_8	v6032	C+6	B	417	768
v6032_9	v6032	C+7	B	333	655
v6032_10	v6032	C+8	B	223	508
v6032_11	v6032	Diagnosis	B	153	307
v6032_12	v6032	C-1	C	355	751
v6032_13	v6032	C+6	C	330	709
v6032_14	v6032	C+7	C	274	731
v6032_15	v6032	C+8	C	403	748
v6032_16	v6032	C+9	C	267	431
v6032_17	v6032	Diagnosis	C	243	438
v6032_18	v6032	T+6	C	364	798
v6032_19	v6032	Memory	C	294	798
v1075_1	v1075	C-1	A	568	768
v1075_2	v1075	C+6	A	654	897
v1075_3	v1075	C+8	A	569	778
v1075_4	v1075	C+10	A	469	623
v1075_5	v1075	Diagnosis	A	238	282
v1075_6	v1075	C-1	B	631	910
v1075_7	v1075	C+6	B	402	557
v1075_8	v1075	C+8	B	628	806
v1075_9	v1075	C+10	B	594	829
v1075_10	v1075	C+12	B	498	714
v1075_11	v1075	Diagnosis	B	486	655
v1075_12	v1075	C-1	C	466	647
v1075_13	v1075	C+6	C	464	648
v1075_14	v1075	C+8	C	445	653
v1075_15	v1075	C+10	C	375	497
v1075_16	v1075	C+12	C	238	308

v1075_17	v1075	C+14	C	310	413
v1075_18	v1075	Diagnosis	C	392	578
v1075_19	v1075	T+6	C	451	609
v1075_20	v1075	Memory	C	740	992
v1040_1	v1040	C-1	A	457	746
v1040_2	v1040	C+6	A	358	657
v1040_3	v1040	C+9	A	215	335
v1040_4	v1040	C+10	A	162	252
v1040_5	v1040	C+11	A	175	248
v1040_6	v1040	Diagnosis	A	90	310
v1040_7	v1040	C-1	B	1	1
v1040_8	v1040	C+6	B	568	844
v1040_9	v1040	C+9	B	381	635
v1040_10	v1040	C+12	B	465	748
v1040_11	v1040	C+15	B	316	746
v1040_12	v1040	C+18	B	451	809
v1040_13	v1040	Diagnosis	B	470	699
v1040_14	v1040	C-1	C	294	512
v1040_15	v1040	C+6	C	101	235
v1040_16	v1040	C+8	C	355	528
v1040_17	v1040	C+9	C	281	533
v1040_18	v1040	C+10	C	305	482
v1040_19	v1040	Diagnosis	C	432	642
v1040_20	v1040	T+6	C	428	677
v1040_21	v1040	Memory	C	226	341

Chapter 6: General Discussion

This thesis used T-cell receptor repertoire sequencing as a novel tool to further our understanding and reveal previously unknown insights into the complex immune responses triggered during a first *Plasmodium* challenge and following re-infection. I sought to demonstrate the utility of this powerful technique, by documenting the clonal dynamics and molecular architecture of the T-cell repertoires that are elicited following exposure to the parasite, and determine if pathogen-specific signatures are generated. Below, I summarise the key findings from both the *P. chabaudi* project and the *P. falciparum* CHMI study, and outline future directions for this research.

6.1. Overview of key findings from *P. chabaudi* model

6.1.1. A conserved TCRβ response is detectable in a first *P. chabaudi* infection

Results from the bulk splenic TCR-seq data indicate a conserved TCRβ response is repeatedly elicited in the acute phase of a first murine *P. chabaudi* infection, and that these are dominated by TRBV3 gene usage. These findings were confirmed with analyses of independent RNA-seq data. TCR repertoire studies of murine *Plasmodium* spp infections have only previously used more rudimentary techniques such as CDR3 length analysis (spectratyping), to indicate that a specific T-cell response may be elicited by the parasite. By using unbiased TCR-seq methods, this study goes into much finer detail than previous studies, by revealing the exact molecular sequences that constitute this specific response, opening up exciting new avenues for further research. This conserved response that was repeatedly elicited in the syngeneic mice, has the classic hallmarks of a public response including convergent recombination. Public responses following pathogen challenge have increasingly garnered attention as it is thought they may be driven by immunodominant peptides. Although public responses can simply be present due to biases in V(D)J recombination, the conserved signature detected was not evident in other pathogens examined, though current lack of publicly available data sets meant examining this was by no means an exhaustive process. As TCR repertoire data is increasingly generated, the question of specificity of this response can be more fully addressed. The response was further characterised using single-cell RNA-seq combined with TCR capture, which indicated clones of

interest were present in a mostly Tfh cluster, indicating they are part of the adaptive process against the parasite rather than representing an innate cell population. As only a few clones were captured in the single-cell analyses and given marked T-cell plasticity, this phenotype cannot however be extrapolated definitively to the bulk TCR-seq data. Acute TCR effector repertoires were also examined in a re-challenge infection, where the conserved signature was not re-expanded nor was TRBV3. Given the near-sterile immunity that is induced after only one infection in a *P. chabaudi*, this finding cannot be uncoupled from antigen load. Recent methods that overcome this limitation by using the more virulent *P. chabaudi AJ* genotype to generate matched parasite burdens in re-challenge (Nahrendorf, Ivens and Spence, 2021), would therefore be of interest to undertake. This would allow better comparison with the findings of the human CHMI studies that indicated reduced T-cell activation in re-challenge despite identical pathogen load (Muñoz Sandoval *et al.*, (In prep)). Observations in the re-challenge study were however in agreement with TCR-seq of the elicited memory populations which appeared to be private responses to the individual, an observation requiring truly longitudinal analyses for confirmation.

6.1.2. Splenic CD4+ responses elicited by *P. chabaudi* are highly polyclonal

Despite the conserved response detected, analysis from the bulk TCR-seq data as well as the single-cell TCR-seq data revealed that the splenic $CD4^+$ repertoires elicited during a first *P. chabaudi* infection are highly polyclonal, with the heterogeneity of TRBV/J gene usage and extent of polyclonality indicative of bystander activation and/or activation of cross-reactive T-cells. Although this cannot be conclusively proven in this study, it is hard to envisage how such a highly heterogenous response could solely be made up of clonally expanded populations to multiple antigens, and findings presented here provide supporting evidence for non-specific activation of $CD4^+$ T-cells in a first infection. Future experiments that make use of tetramer staining or peptide stimulation assays would, together with the repertoire analyses presented here, aid confirmation of non-specific activation. Results from the single-cell analysis indicated that activated $CD4^+$ populations clustered in previously observed subsets of T_H1 and predominantly T_{FH}, in line with there currently being no definitive evidence for a specific murine marker that differentiates bystander activated $CD4^+$ populations from those

that are activated in an antigen-specific manner. The implications of non-specific activation of T-cells in a *P. chabaudi* infection remain unknown, as it has been implicated in both protective and pathogenic immune responses (Lee, Cho and Choi, 2020). Benefits to the host in other disease models have included increased IFN-γ and granzyme B production from bystander activated $CD8^+$ memory T-cells, limiting pathogen replication prior to the onset of de novo antigen-specific responses in a mouse model of *Listeria monocytogenes* (Chu *et al.*, 2013). Similarly, cross-protection against Influenza virus from bystander activated CMV-specific T-cells (Furman *et al.*, 2015) and enhanced immunity in other well-studied viral models of disease including hepatitis A, B and C has also been demonstrated (Kim and Shin, 2019). Bystander activation of $CD4^+$ T-cells is less well characterised (Lee, Cho and Choi, 2020), but benefits to the host have been reported for example in clearance of helminth infections (Guo *et al.*, 2015) and providing nonspecific host resistance against *Toxoplasma gondii,* again through production of IFN-γ (Kawabe *et al.*, 2017). Conversely, nonspecific activation has also been repeatedly implicated in host pathology, for example in Lyme disease-induced arthritis (Whiteside, Snook, Ma, *et al.*, 2018), hepatic injury in acute viral hepatitis (Kim *et al.*, 2018) and in autoimmune encephalitis (Lee *et al.*, 2019). Although clonality was not compared between SBP and MT infections, SBP infections are known to induce higher levels of pro-inflammatory cytokines (Spence *et al.*, 2013), which have been implicated in triggering bystander activation. If this translated in to a greater degree of non-specific T-cell activation in this more virulent model, this could provide further explanation for the higher levels of immunopathology induced in the SBP model and why humoral responses appear to be reduced in such infections (Spence *et al.*, 2013). Given the findings of non-specific recruitment of T-cells from the periphery in the *P. falciparum* CHMI study, future studies dissecting the precise role non-specific T-cells play in malaria immunity are warranted.

6.1.3. Future directions for *P. chabaudi* project

Several outstanding questions from the *P. chabaudi* work undertaken in this thesis still remain. Firstly, does the conserved response have a protective or pathogenic role, or what is its relevance to the response? Public responses elicited following pathogen challenge have increasingly garnered interest as they are thought to be driven by immunodominant peptides, that could represent shared therapeutic targets. As yet however, there is no definitive evidence that generating public responses affects for example vaccine efficacy, and this remains a theoretical hypothesis whilst the field remains in its relative infancy. Conversely, observations in the recent CHMI study (Muñoz Sandoval *et al.*, (In Prep)) suggest that distinct T-cell responses in a first infection may be responsible for the more severe disease seen in a first falciparum infection. Therefore, determining if the response has a protective role or not warrants further research. In future, this could be addressed for example through adoptive transfer experiments of TRBV3^{+} T-cells cells in to nude mice, work which is currently hindered by a lack of monoclonal agents against TRBV3. Use of anti-TRBV3 therapy to explore the question of pathogenicity is also not yet available, though this would also be revealing. Secondly, if the conserved response is against an immunodominant epitope, how can the identity of this be determined? As discussed in Chapter 4, determination of the precise antigenic driver of this response will ideally utilise cell-based display libraries to cater for peptide discovery on a large scale. An example of this novel methodology used cell-based predicted p:MHC libraries and transgenic Jurkat T-cells expressing the TCR of interest with GFP expression of the former linked to intracellular signalling following successful TCR:pMHC interaction. The utility of this method has been shown in neoantigen discovery for a tumour-reactive melanoma TCR (Joglekar *et al.*, 2019). Determining if a conserved immunodominant epitope is driving this conserved response could potentially lead to discovery of a new therapeutic target for malaria if the epitope were to be conserved across species. Ultimately, the *P. chabaudi* model is used as a model of falciparum malaria in humans, and future research should consider the potential translational impact to human disease given the differences in parasite biology between *P. chabaudi* and *P. falciparum*. An immunodominant epitope may represent a conserved ortholog between species, making such research worthy of pursuit, but future studies should

where possible work in conjunction with and/or utilise the CHMI model or ideally endemic country field studies to maximise translatability of findings.

6.2. Overview of key findings from *P. falciparum* CHMI study

6.2.1. Non-specific recruitment of clones from periphery

The CHMI results presented in this thesis are one avenue of exploration from a wider study utilising systems immunology to understand how individuals exposed to *P. falciparum* can develop immunity to severe disease after only their first infection. The aim of this wider study is to determine how tolerance to the parasite can evolve through the first few infections of life, protecting the individual from collateral damage inflicted by the host immune response, independent of controlling parasitaemia. Recent results from this wider study from which the human PBMC samples and RNA-seq data presented in this thesis came from, indicate that $CD4^+$ T-cell activation is much reduced in a third infection compared to a first, despite no differences in either parasitaemia or levels of systemic inflammation across the first three infections of life (Muñoz Sandoval *et al.*, (In Prep)). The precise mechanism by which this is orchestrated is not defined, but it is suggested that bystander activation may account for the broad activation seen in a first infection, leading to host collateral damage, but that only more specific responses are activated in a second infection, similar to observations from the *P. chabaudi* study. Some support for this hypothesis is evident in the TCR-seq data, but it is not fully conclusive. Despite the marked lymphopenia evident in individuals at day of diagnosis, the TRBV/J gene repertoires of individuals remained remarkably stable across the time-course of both a first and second infection, indicating broad recruitment with no specific TRBV gene or V/J combination recruited out of circulation nor found to be expanded post drug-treatment. The most striking finding in the longitudinal TCR-seq, was the movement of previously established and persistent clones out of peripheral circulation, including clones with previously defined epitope specificity for common chronic viral diseases including CMV and EBV. Persistent clones troughed at day of diagnosis, before returning to baseline levels post drug-treatment. This finding was also evident in the TCR data mined from the whole-blood RNA-seq, and indicates that clones are being recruited non-specifically from the periphery, confirming the broad recruitment suggested by the

available FACS data. Given the systemic inflammation shown evident in individuals, it is highly plausible that these non-specific T-cells become activated in the inflamed splenic environment, but further work as discussed below is required to prove this. The TCR-seq results do suggest that the same level of non-specific recruitment occurs regardless of infection. Combining this with the wider study findings that activation of T-cells returning to the periphery is much reduced by a third infection, insinuates that these non-specific cells although still appear to be recruited, are not activated by the inflammatory splenic milieu in a third infection, despite levels of systemic inflammation being the same across all infections (Muñoz Sandoval *et al.*, (In Prep)). How this could occur remains unknown, but it is proposed that given these non-specific cells are activated either independently of TCR engagement or low affinity reactions, mechanisms may exist whereby only specific T-cells are re-activated (Muñoz Sandoval *et al.*, (In Prep)). Ongoing studies that combine single cell transcriptomics with TCR-capture may provide more clarity. Ultimately, how the non-specific recruitment evident here may affect the developing immune response in the spleen, either through aberrant cytokine production from bystander activation, spatially limiting specific p:MHC interactions or poor effector function due to low-affinity cross-reactivity remains unresolved and an exciting avenue for future research.

No clonally expanded populations relative to baseline were detected in the periphery at 28 days post challenge in the TCR-seq data in any individual, in either infection. This does not rule out a clonal response closer to post-drug treatment timepoints, as noted in the RNA-seq data, at 6 days post-treatment in 1 individual; a finer precision with closer spacing of timepoints should be used to address this in future studies. This finding does show however that repertoire architecture remains strikingly unchanged following a CHMI infection, with no expanded clones becoming established relative to baseline. This may be due to drug-treatment early in infection removing the antigen exposure that would otherwise be chronic in endemic settings, with expanded clones in the periphery generally driven by chronic exposure. If memory responses were primed from a first infection, this might have been expected to be reflected in recall responses in re-challenge infection, but these were not evident. Again, this cannot be ruled out with timepoints available, and similarly given the dynamic and ‘noise’ of the

movement of established clones, the heterogeneity of the response may be masking a specific recall response.

6.2.2. Future directions for *P. falciparum* CHMI study

Given the evidence for non-specific recruitment of T-cells out of peripheral circulation, a key avenue for future research arising from this study is what affect this has on the developing immune response. Bystander activation should be confirmed or ruled out with future studies that combine TCR-seq with single-cell RNA-seq to determine the activation and phenotype of non-specific recruited T-cell populations, including those with previously annotated specificities, as well as peptide-stimulation assays. Combining this with longitudinal studies through the first few infections of life, may be illuminating regarding how levels of T-cell activation can be much reduced in a third infection compared to a first, despite levels of systemic inflammation and non-specific recruitment being similar. The mechanism by which this could occur remains an active area of research. If bystander activation is occurring, what are the consequences of this? As discussed for *P. chabaudi,* the phenomenon has been associated with both protective and pathogenic consequences, enhancing immunity in some responses, but enhancing immunopathology in others, outcomes which appear to be context rather than pathogen dependent (Lee, Cho and Choi, 2020). For example, bystander activated T-cells are implicated in severe cases of Lassa fever, where increased activation of non-specific cells with homing capacity to inflamed tissues is associated with fatal cases compared to those who survive infection (Port *et al.*, 2020). If bystander activation does occur, the landscape of an individual's pre-existing memory repertoire could influence their response to a first falciparum infection, and further research in to this phenomenon is warranted.

This is the first longitudinal TCRβ repertoire study to examine the repertoires elicited in both a first *P. falciparum* infection and following re-challenge. No specific TCR signatures of disease were found either longitudinally within an individual or shared between individuals. It is plausible that no specific response is elicited given the

complex heterogeneity of responses demonstrated and non-specific recruitment. However, this finding is not conclusive due to potential limitations of the study design, which was not conceived specifically for repertoire analyses. The study has been revealing in terms of unveiling the non-specific recruitment of clones out of peripheral circulation, but has also highlighted and helped refine requirements for future studies that aim to definitively determine if clonal signatures of disease are generated. A finer precision of timepoints post drug-treatment is needed, to ascertain the exact kinetics of when T-cells are released back into periphery, and sorting cell populations of interest would also be beneficial given the dynamics of persistent and established clones shown here which may be masking plasmodium-specific T-cell trajectories. Similarly, increased sequencing depth may be required to detect TCR signatures of disease, given no populations appeared to be highly expanded relative to baseline in data presented here. A limitation of the CHMI model is that volunteers have to be drug-treated early in infection for ethical reasons. This may prevent adequate memory formation that could alter the repertoire architecture, and repertoire studies that aim to investigate such responses should focus on endemic setting field studies where chronic antigen exposure may make such responses easier to detect. Larger cohorts of volunteers are also required to definitively rule in or out convergent responses between individuals.

Overall, application of TCR-seq in the *P. chabaudi* model and *P. falciparum* CHMI in this thesis demonstrates the utility of the technology in understanding the dynamics and clonality of the complex T-cell responses elicited during a malaria infection. The data presented complements and elaborates on existing hypotheses, provides novel insights in to public TCRβ signatures in the murine model, and reveals previously unknown dynamics of non-specific recruitment in the CHMI model. Findings from both studies provide a platform for a plethora of future research ideas that will continue to contribute to our understanding of this deadly disease.

Bibliography

Van Aalst, S. *et al.* (2017) 'Bystander activation of irrelevant CD4+ T cells following antigen-specific vaccination occurs in the presence and absence of adjuvant', *PLoS ONE*, 12(5), pp. 1–13. doi: 10.1371/journal.pone.0177365.

Achtman, A. H. *et al.* (2007) 'Malaria-specific antibody responses and parasite persistence after infection of mice with Plasmodium chabaudi chabaudi', *Parasite Immunology*, 29(9), pp. 435–444. doi: 10.1111/j.1365-3024.2007.00960.x.

Almanzar, G. *et al.* (2013) 'Assessment of IgG avidity against pertussis toxin and filamentous hemagglutinin via an adapted enzyme-linked immunosorbent assay (ELISA) using ammonium thiocyanate', *Journal of Immunological Methods*. Elsevier B.V., 387(1–2), pp. 36–42. doi: 10.1016/j.jim.2012.09.008.

Alves Sousa, A. de P. *et al.* (2019) 'Comprehensive Analysis of TCR-β Repertoire in Patients with Neurological Immune-mediated Disorders', *Scientific Reports*, 9(1), pp. 1–10. doi: 10.1038/s41598-018-36274-7.

Andrade, C. M. *et al.* (2020) 'Increased circulation time of Plasmodium falciparum underlies persistent asymptomatic infection in the dry season', *Nature Medicine*. Springer US, 26(12), pp. 1929–1940. doi: 10.1038/s41591-020-1084-0.

Andrews, S. (2010) 'FastQC: A quality control tool for high throughput sequence data.'

Awandare, G. A. *et al.* (2006) 'Increased levels of inflammatory mediators in children with severe Plasmodium falciparum malaria with respiratory distress', *Journal of Infectious Diseases*, 194(10), pp. 1438–1446. doi: 10.1086/508547.

Bach, F. *et al.* (2021) 'Parasite species regulates T cell activation in human malaria', (1).

Bai, Y. *et al.* (2018) 'Evaluation of the capacities of mouse TCR profiling from short read RNA-seq data', *PLoS ONE*, 13(11), pp. 1–19. doi: 10.1371/journal.pone.0207020.

Barennes, P. *et al.* (2020) 'Benchmarking of T cell receptor repertoire profiling methods reveals large systematic biases', *Nature Biotechnology*. Springer US. doi: 10.1038/s41587-020-0656-3.

Barrera, V. *et al.* (2019) 'Comparison of CD8+ T Cell Accumulation in the Brain During Human and Murine Cerebral Malaria', *Frontiers in immunology*, 10(July), p. 1747. doi: 10.3389/fimmu.2019.01747.

Bastian, M., Heymann, S. and Jacomy, M. (2009) 'Gephi : An Open Source Software for Exploring and Manipulating Networks', in.

Bates, D. *et al.* (2015) 'Fitting linear mixed-effects models using lme4', *Journal of Statistical Software*, 67(1). doi: 10.18637/jss.v067.i01.

Beeson, J. G. *et al.* (2019) 'Challenges and strategies for developing efficacious and long-lasting malaria vaccines', *Science Translational Medicine*, 11(474). doi: 10.1126/scitranslmed.aau1458.

Belnoue, E. *et al.* (2002) ' On the Pathogenic Role of Brain-Sequestered αβ CD8 + T Cells in Experimental Cerebral Malaria ', *The Journal of Immunology*, 169(11), pp. 6369–6375. doi: 10.4049/jimmunol.169.11.6369.

Benati, D. *et al.* (2016) 'Public T cell receptors confer high-avidity CD4 responses to HIV controllers', *Journal of Clinical Investigation*, 126(6), pp. 2093–2108. doi: 10.1172/JCI83792DS1.

Billiau, A. and Matthys, P. (2001) 'Modes of action of Freund's adjuvants in experimental models of autoimmune diseases.', *Journal of leukocyte biology*, 70(6), pp. 849–60. doi: 10.1189/jlb.70.6.849.

Bolotin, D. A. *et al.* (2015) 'MiXCR : software for comprehensive adaptive immunity profiling', *Nature methods*. Nature Publishing Group, 12(5), pp. 380–381. doi: 10.1038/nmeth.3364.

Bolotin, D. A. *et al.* (2017) 'Antigen receptor repertoire profiling from RNA-seq data', *Proc. Natl. Acad. Sci. USA*, 35(10), pp. 908–911.

Bradley, P. and Thomas, P. G. (2019) 'Using T Cell Receptor Repertoires to Understand the Principles of Adaptive Immune Recognition', *Annual Review of Immunology*, 37(1), pp. 547–570. doi: 10.1146/annurev-immunol-042718-041757.

Britanova, O. V. *et al.* (2014) 'Age-Related Decrease in TCR Repertoire Diversity Measured with Deep and Normalized Sequence Profiling', *The Journal of Immunology*, 192(6), pp. 2689–2698. doi: 10.4049/jimmunol.1302064.

Butler, A. *et al.* (2018) 'Integrating single-cell transcriptomic data across different conditions, technologies, and species', *Nature Biotechnology*, 36(5), pp. 411–420. doi: 10.1038/nbt.4096.

Cavacini, L. A., Long, C. A. and Weidanz, W. P. (1986) 'T-cell immunity in murine malaria: Adoptive transfer of resistance to Plasmodium chabaudi adami in nude mice with splenic T cells', *Infection and Immunity*, 52(3), pp. 637–643. doi: 10.1128/iai.52.3.637-643.1986.

Chan, J. A. *et al.* (2019) 'Antibody targets on the surface of plasmodium falciparum- infected erythrocytes that are associated with immunity to severe malaria in young children', *Journal of Infectious Diseases*, 219(5), pp. 819–828. doi: 10.1093/infdis/jiy580.

Chang, Y. M. *et al.* (2020) 'T Cell Receptor Diversity and Lineage Relationship between Virus-Specific CD8 T Cell Subsets during Chronic Lymphocytic Choriomeningitis Virus Infection', *Journal of Virology*, 94(20), pp. 1–13. doi: 10.1128/jvi.00935-20.

Cheesman, S. *et al.* (2009) 'Strain-specific immunity may drive adaptive polymorphism in the merozoite surface protein 1 of the rodent malaria parasite Plasmodium chabaudi', *Infection, Genetics and Evolution*, 9(2), pp. 248–255. doi: 10.1016/j.meegid.2008.12.002.

Cheng, Q. *et al.* (1997) 'Measurement of Plasmodium falciparum growth rates in vivo: a test of malaria vaccines', *American Journal of Tropical Medicine and Hygiene*, 57(4), pp. 495–500.

Chu, N. D. *et al.* (2019) 'Longitudinal immunosequencing in healthy people reveals persistent T cell receptors rich in highly public receptors', *BMC Immunology*. BMC Immunology, 20(1), pp. 1–12. doi: 10.1186/s12865-019-0300-5.

Chu, T. *et al.* (2013) 'Bystander-Activated Memory CD8 T Cells Control Early Pathogen Load in an Innate-like, NKG2D-Dependent Manner', *Cell Reports*. The Authors, 3(3), pp. 701–708. doi: 10.1016/j.celrep.2013.02.020.

Cohen, S., McGregor, I. and Carrington, S. (1961) 'Gamma-Globulin and Aquired Immunity to Human Malaria', *Nature*, 192(4804), pp. 733–737. doi: 10.1038/192733a0.

Corrie, B. D. *et al.* (2018) 'iReceptor: A platform for querying and analyzing antibody/B-cell and T-cell receptor repertoire data across federated repositories', *Immunological Reviews*, 284(1), pp. 24–41. doi: 10.1111/imr.12666.

Couper, K. N., Blount, D. G. and Riley, E. M. (2008) 'IL-10: The Master Regulator of Immunity to Infection', *The Journal of Immunology*, 180(9), pp. 5771–5777. doi: 10.4049/jimmunol.180.9.5771.

Covacu, R. *et al.* (2016) 'System-wide Analysis of the T Cell Response', *Cell Reports*. The Authors, 14(11), pp. 2733–2744. doi: 10.1016/j.celrep.2016.02.056.

Crompton, P. D. *et al.* (2014) 'Malaria Immunity in Man and Mosquito: Insights into Unsolved Mysteries of a Deadly Infectious Disease', *Annual Review of Immunology*, 32(1), pp. 157–187. doi: 10.1146/annurev-immunol-032713-120220.

Crotty, S. (2014) 'T Follicular Helper Cell Differentiation, Function, and Roles in Disease', *Immunity*. Elsevier Inc., 41(4), pp. 529–542. doi: 10.1016/j.immuni.2014.10.004.

Dantzler, K. W. and Jagannathan, P. (2018) 'γδ T cells in antimalarial immunity: New insights into their diverse functions in protection and tolerance', *Frontiers in Immunology*, 9(OCT), pp. 1–14. doi: 10.3389/fimmu.2018.02445.

Dash, P. *et al.* (2011) 'Paired analysis of TCRa and TCRb chains at the single-cell level in mice', *Journal of Clinical Investigation*, 121(1), pp. 288–295. doi: 10.1172/JCI44752.288.

Dash, P. *et al.* (2017a) 'Quantifiable predictive features define epitope-specific T cell receptor repertoires', *Nature*, 547(7661), pp. 89–93. doi: 10.1038/nature22383.

Dash, P. *et al.* (2017b) 'Quantifiable predictive features define epitope-specific T cell receptor repertoires', *Nature*. Nature Publishing Group, 547(7661), pp. 89–93. doi: 10.1038/nature22383.

Datoo, M. S. *et al.* (2021) 'Efficacy of a low-dose candidate malaria vaccine, R21 in adjuvant Matrix-M, with seasonal administration to children in Burkina Faso: a randomised controlled trial.', *Lancet (London, England)*, 397, pp. 1809–1818. doi: 10.1016/S0140-6736(21)00943-0.

Deitsch, K. W. and Dzikowski, R. (2017) 'Variant Gene Expression and Antigenic Variation by Malaria Parasites', *Annual Review of Microbiology*, 71, pp. 625–641. doi: 10.1146/annurev-micro-090816-093841.

Deseke, M. and Prinz, I. (2020) 'Ligand recognition by the γδ TCR and discrimination between homeostasis and stress conditions', *Cellular and Molecular Immunology*. Springer US, 17(9), pp. 914–924. doi: 10.1038/s41423-020-0503-y.

Doolan, D. L., Dobaño, C. and Baird, J. K. (2009) 'Acquired immunity to Malaria', *Clinical Microbiology Reviews*, 22(1), pp. 13–36. doi: 10.1128/CMR.00025-08.

Doumbo, O. K. *et al.* (2009) 'High levels of Plasmodium falciparum rosetting in all clinical forms of severe malaria in African children', *American Journal of Tropical Medicine and Hygiene*, 81(6), pp. 987–993. doi: 10.4269/ajtmh.2009.09-0406.

Duffy, M. F. *et al.* (2016) 'Differences in PfEMP1s recognized by antibodies from patients with uncomplicated or severe malaria', *Malaria Journal*. BioMed Central, 15(1), pp. 1–12. doi: 10.1186/s12936-016-1296-4.

Dupic, T. *et al.* (2019) 'Genesis of the αβ T-cell receptor', *PLoS computational biology*, 15(3), p. e1006874. doi: 10.1371/journal.pcbi.1006874.

Dziubianau, M. *et al.* (2013) 'TCR repertoire analysis by next generation sequencing allows complex differential diagnosis of T cell-related pathology', *American Journal of Transplantation*, 13(11), pp. 2842–2854. doi: 10.1111/ajt.12431.

Edwards, C. L. *et al.* (2018) 'Early changes in CD4+ t-cell activation during blood-stage plasmodium falciparum infection', *Journal of Infectious Diseases*, 218(7), pp. 1119–1129. doi: 10.1093/infdis/jiy281.

Egorov, E. S. *et al.* (2015) 'Quantitative Profiling of Immune Repertoires for Minor Lymphocyte Counts

Using Unique Molecular Identifiers', *The Journal of Immunology*, 194(12), pp. 6155–6163. doi: 10.4049/jimmunol.1500215.

Elyahu, Y. *et al.* (2019) 'Aging promotes reorganization of the CD4 T cell landscape toward extreme regulatory and effector phenotypes', *Science Advances*, 5(8), p. eaaw8330. doi: 10.1126/sciadv.aaw8330.

Emerson, R. O. *et al.* (2017) 'Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire', *Nature Genetics*. Nature Publishing Group, 49(5), pp. 659–665. doi: 10.1038/ng.3822.

Ewer, K. J. *et al.* (2013) 'Protective CD8 + T-cell immunity to human malaria induced by chimpanzee adenovirus-MVA immunisation', *Nature Communications*, 4, pp. 4–13. doi: 10.1038/ncomms3836.

Farmanbar, A., Kneller, R. and Firouzi, S. (2019) 'RNA sequencing identifies clonal structure of T-cell repertoires in patients with adult T-cell leukemia/lymphoma', *npj Genomic Medicine*. Springer US, 4(1), pp. 1–9. doi: 10.1038/s41525-019-0084-9.

Ferreira, M. U. and Katzin, A. M. (1995) 'The assessment of antibody affinity distribution by thiocyanate elution: a simple dose-response approach', *Journal of Immunological Methods*, 187(2), pp. 297–305. doi: 10.1016/0022-1759(95)00186-4.

Figueiredo, M. M. *et al.* (2017) 'T follicular helper cells regulate the activation of B lymphocytes and antibody production during Plasmodium vivax infection', *PLoS Pathogens*, 13(7), pp. 1–23. doi: 10.1371/journal.ppat.1006484.

Fink, K. (2019) 'Can we improve vaccine efficacy by targeting T and B cell repertoire convergence?', *Frontiers in Immunology*, 10(FEB), pp. 1–10. doi: 10.3389/fimmu.2019.00110.

Fontana, M. F. *et al.* (2016) 'Macrophage Colony Stimulating Factor Derived from CD4 + T Cells Contributes to Control of a Blood-Borne Infection', *PLoS Pathogens*, 12(12), pp. 1–26. doi: 10.1371/journal.ppat.1006046.

Freeman, J. D. *et al.* (2009) 'Profiling the T-cell receptor beta-chain repertoire by massively parallel sequencing', *Genome Research*, 19(10), pp. 1817–1824. doi: 10.1101/gr.092924.109.

Frimpong, A. *et al.* (2018) 'Novel strategies for malaria vaccine design', *Frontiers in Immunology*, 9(NOV), pp. 1–14. doi: 10.3389/fimmu.2018.02769.

Furman, D. *et al.* (2015) 'Cytomegalovirus infection enhances the immune response to influenza', *Science Translational Medicine*, 7(281). doi: 10.1126/scitranslmed.aaa2293.

Galatas, B., Bassat, Q. and Mayor, A. (2016) 'Malaria Parasites in the Asymptomatic: Looking for the Hay in the Haystack', *Trends in Parasitology*. Elsevier Ltd, 32(4), pp. 296–308. doi: 10.1016/j.pt.2015.11.015.

Gasper, D. J., Tejera, M. M. and Suresh, M. (2014) 'CD4 T-cell memory generation and maintenance.', *Critical reviews in immunology*, 34(2), pp. 121–46. doi: 10.1615/critrevimmunol.2014010373.

Gee, M. H. *et al.* (2018) 'Antigen Identification for Orphan T Cell Receptors Expressed on Tumor-Infiltrating Lymphocytes Antigen Identification for Orphan T Cell Receptors Expressed on Tumor-Infiltrating Lymphocytes', *Cell*. Elsevier Inc., 172(3), pp. 549-556.e16. doi: 10.1016/j.cell.2017.11.043.

Ghazanfari, N., Mueller, S. N. and Heath, W. R. (2018) 'Cerebral malaria in mouse and man', *Frontiers in Immunology*, 9(SEP), pp. 1–11. doi: 10.3389/fimmu.2018.02016.

Giudicelli, V. *et al.* (2006) 'IMGT/LIGM-DB, the IMGT comprehensive database of immunoglobulin and

T cell receptor nucleotide sequences.', *Nucleic acids research*, 34(Database issue), pp. 781–784. doi: 10.1093/nar/gkj088.

Glanville, J. *et al.* (2017) 'Identifying specificity groups in the T cell receptor repertoire', *Nature*. Nature Publishing Group, 547(7661), pp. 94–98. doi: 10.1038/nature22976.

Gonçalves, B. P. *et al.* (2014) 'Parasite Burden and Severity of Malaria in Tanzanian Children', *New England Journal of Medicine*, 370, pp. 1799–1808. doi: 10.1056/NEJMoa1303944.

Gonzales, S. J. *et al.* (2020) 'Naturally Acquired Humoral Immunity Against Plasmodium falciparum Malaria', *Frontiers in Immunology*, 11(October), pp. 1–15. doi: 10.3389/fimmu.2020.594653.

Gorski, J. *et al.* (1994) 'Circulating T cell repertoire complexity in normal individuals and bone marrow recipients analyzed by CDR3 size spectratyping. Correlation with immune status.', *The Journal of Immunology*, 152, pp. 5109–5119.

Grau, G. E. *et al.* (1989) 'Monoclonal antibody against interferon γ can prevent experimental cerebral malaria and its associated overproduction of tumor necrosis factor', *Proceedings of the National Academy of Sciences of the United States of America*, 86(14), pp. 5572–5574. doi: 10.1073/pnas.86.14.5572.

Greiff, V. *et al.* (2015) 'A bioinformatic framework for immune repertoire diversity profiling enables detection of immunological status.', *Genome medicine*. Genome Medicine, 7(1), p. 49. doi: 10.1186/s13073-015-0169-8.

Groom, J. R. and Luster, A. D. (2011) 'CXCR3 ligands: Redundant, collaborative and antagonistic functions', *Immunology and Cell Biology*, 89(2), pp. 207–215. doi: 10.1038/icb.2010.158.

Grun, J. L. and Weidanz, W. P. (1981) 'Immunity to Plasmodium chabaudi adami in the B-cell-deficient mouse', *Nature*, 290(5802), pp. 143–145. doi: 10.1038/290143a0.

Guo, L. *et al.* (2015) 'Innate immunological function of T H2 cells in vivo', *Nature Immunology*, 16(10), pp. 1051–1059. doi: 10.1038/ni.3244.

Gupta, S. *et al.* (1999) 'Immunity to non-cerebral severe malaria is acquired after one or two infections.', *Nature medicine*, 5(3), pp. 340–343. doi: 10.1038/6560.

Hashimoto, K. *et al.* (2019) 'Single-cell transcriptomics reveals expansion of cytotoxic CD4 T cells in supercentenarians', *Proceedings of the National Academy of Sciences of the United States of America*, 116(48), pp. 24242–24251. doi: 10.1073/pnas.1907883116.

Heather, J. M. *et al.* (2016) 'Dynamic perturbations of the T-Cell receptor repertoire in chronic HIV infection and following antiretroviral therapy', *Frontiers in Immunology*, 6(JAN), pp. 1–15. doi: 10.3389/fimmu.2015.00644.

Hill, D. L. *et al.* (2019) 'The adjuvant GLA-SE promotes human Tfh cell expansion and emergence of public TCRβ clonotypes', *The Journal of Experimental Medicine*, 216, pp. 1857–1873. doi: 10.1084/jem.20190301.

Holla, P. *et al.* (2021) 'Shared transcriptional profiles of atypical B cells suggest common drivers of expansion and function in malaria , HIV , and autoimmunity', (May).

Hozumi, N. and Tonegawa, S. (1976) 'Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions', *Proceedings of the National Academy of Sciences of the United States of America*, 73(10), pp. 3628–3632. doi: 10.1073/pnas.73.10.3628.

Huang, H. *et al.* (2020) 'Analyzing the Mycobacterium tuberculosis immune response by T-cell receptor

clustering with GLIPH2 and genome-wide antigen screening', *Nature Biotechnology*. Springer US, (38), pp. 1194–1202. doi: 10.1038/s41587-020-0505-4.

Hunt, N. H. *et al.* (2014) 'Cerebral malaria: Gamma-interferon redux', *Frontiers in Cellular and Infection Microbiology*, 4(AUG), pp. 1–12. doi: 10.3389/fcimb.2014.00113.

Huth, A. *et al.* (2019) 'Antigen-Specific TCR Signatures of Cytomegalovirus Infection', *The Journal of Immunology*, 202(3), pp. 979–990. doi: 10.4049/jimmunol.1801401.

Hviid, L. *et al.* (1997) 'Rapid reemergence of T cells into peripheral circulation following treatment of severe and uncomplicated Plasmodium falciparum malaria', *Infection and Immunity*, 65(10), pp. 4090–4093.

Illingworth, J. *et al.* (2013) ' Chronic Exposure to Plasmodium falciparum Is Associated with Phenotypic Evidence of B and T Cell Exhaustion ', *The Journal of Immunology*, 190(3), pp. 1038–1047. doi: 10.4049/jimmunol.1202438.

Izraelson, M. *et al.* (2018) 'Comparative analysis of murine T-cell receptor repertoires', *Immunology*, 153(2), pp. 133–144. doi: 10.1111/imm.12857.

Joglekar, A. V. *et al.* (2019) 'T cell antigen discovery via signaling and antigen-presenting bifunctional receptors', *Nature Methods*. Springer US, 16(2), pp. 191–198. doi: 10.1038/s41592-018-0304-8.

Joglekar, A. V. and Li, G. (2020) 'T cell antigen discovery', *Nature Methods*. Springer US. doi: 10.1038/s41592-020-0867-z.

Jones, E., Oliphant, T. E. and Peterson, P. (2001) *SciPy: Open Source Scientific Tools for Python*. Available at: http://www.scipy.org.

Junqueira, C. *et al.* (2021) 'γδ T cells suppress Plasmodium falciparum blood-stage infection by direct killing and phagocytosis', *Nature Immunology*, 22(3), pp. 347–357. doi: 10.1038/s41590-020-00847-4.

Kawabe, T. *et al.* (2017) 'Memory-phenotype CD4+ T cells spontaneously generated under steady-state conditions exert innate TH1-like effector function', *Science Immunology*, 2(12). doi: 10.1126/sciimmunol.aam9304.

Kester, K. E. *et al.* (2009) 'Randomized, double-blind, phase 2a trial of falciparum malaria vaccines RTS,S/AS01B and RTS,S/AS02A in malaria-naive adults: Safety, efficacy, and immunologic associates of protection', *Journal of Infectious Diseases*, 200(3), pp. 337–346. doi: 10.1086/600120.

Khosravi-Maharlooei, M. *et al.* (2019) 'Crossreactive public TCR sequences undergo positive selection in the human thymic repertoire', *The Journal of clinical investigation*, 130, pp. 2446–2462. doi: 10.1172/JCI124358.

Kim, J. *et al.* (2018) 'Innate-like Cytotoxic Function of Bystander-Activated CD8 + T Cells Is Associated with Liver Injury in Acute Hepatitis A', *Immunity*. Elsevier Inc., 48(1), pp. 161-173.e5. doi: 10.1016/j.immuni.2017.11.025.

Kim, T. S. and Shin, E. C. (2019) 'The activation of bystander CD8+ T cells and their roles in viral infection', *Experimental and Molecular Medicine*. Springer US, 51(12), pp. 1–9. doi: 10.1038/s12276-019-0316-1.

King, C. L. *et al.* (2015) 'Biosignatures of Exposure/Transmission and Immunity', *The American journal of tropical medicine and hygiene*, 93(3), pp. 16–27. doi: 10.4269/ajtmh.15-0037.

King, T. and Lamb, T. (2015) 'Interferon-γ: The Jekyll and Hyde of Malaria', *PLoS Pathogens*, 11(10), pp. 8–13. doi: 10.1371/journal.ppat.1005118.

Kivioja, T. *et al.* (2012) 'Counting absolute numbers of molecules using unique molecular identifiers', *Nature Methods*, 9(1), pp. 72–74. doi: 10.1038/nmeth.1778.

Klarenbeek, P. L. *et al.* (2012) 'Deep Sequencing of Antiviral T-Cell Responses to HCMV and EBV in Humans Reveals a Stable Repertoire That Is Maintained for Many Years', *PLoS Pathogens*, 8(9). doi: 10.1371/journal.ppat.1002889.

Kotepui, M. *et al.* (2020) 'Reduction in total leukocytes in malaria patients compared to febrile controls: A systematic review and meta-analysis', *PLoS ONE*, 15(6), pp. 1–28. doi: 10.1371/journal.pone.0233913.

Krishna, C. *et al.* (2020) 'Genetic and environmental determinants of human TCR repertoire diversity', *Immunity and Ageing*. Immunity & Ageing, 17(1), pp. 1–7. doi: 10.1186/s12979-020-00195-9.

Kumar, R. *et al.* (2019) 'The regulation of CD4+ T cells during malaria', *Immunological Reviews*, 293(1), pp. 70–87. doi: 10.1111/imr.12804.

Kumar, R., Ng, S. and Engwerda, C. (2019) 'The role of IL-10 in malaria: A double edged sword', *Frontiers in Immunology*, 10(FEB), pp. 1–10. doi: 10.3389/fimmu.2019.00229.

Kurup, S. P., Butler, N. S. and Harty, J. T. (2019) 'T cell-mediated immunity to malaria', *Nature Reviews Immunology*. Springer US, (19), pp. 457–471. doi: 10.1038/s41577-019-0158-z.

Langhorne, J. *et al.* (2008) 'Immunity to malaria: more questions than answers.', *Nature immunology*, 9(7), pp. 725–732. doi: 10.1038/ni.f.205.

Lee, H. G. *et al.* (2019) 'Pathogenic function of bystander-activated memory-like CD4 + T cells in autoimmune encephalomyelitis', *Nature Communications*. Springer US, 10(1). doi: 10.1038/s41467-019-08482-w.

Lee, H. G., Cho, M. Z. and Choi, J. M. (2020) 'Bystander CD4+ T cells: crossroads between innate and adaptive immunity', *Experimental and Molecular Medicine*. Springer US, 52(8), pp. 1255–1263. doi: 10.1038/s12276-020-00486-7.

Lefranc, M. (1999) 'IMGT, the international ImMunoGeneTics database', *Nucleic Acids Research*, 27(1), pp. 219–221. doi: 10.1093/nar/28.1.219.

Leinster, T. and Cobbold, C. A. (2012) 'Measuring diversity: The importance of species similarity', *Ecology*, 93(3), pp. 477–489. doi: 10.1890/10-2402.1.

Lenth, R. V (2021) 'emmeans: Estimated Marginal Means, aka Least-Squares Means'.

Lewis, M., Pfeil, J. and Mueller, A. K. (2011) 'Continuous oral chloroquine as a novel route for Plasmodium prophylaxis and cure in experimental murine models', *BMC Research Notes*. BioMed Central Ltd, 4(1), p. 262. doi: 10.1186/1756-0500-4-262.

Li, B. *et al.* (2017) 'Ultrasensitive detection of TCR hypervariable-region sequences in solid-tissue RNA-seq data', *Nature Genetics*, 49(4), pp. 483–484. doi: 10.1038/ng.3820.

Li, N. *et al.* (2020) 'T-cell receptor repertoire analysis for the diagnosis and treatment of solid tumor: A methodology and clinical applications', *Cancer Communications*, 40(10), pp. 473–483. doi: 10.1002/cac2.12074.

Liao, M. *et al.* (2020) 'Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19', *Nature Medicine*. Springer US, 26(6), pp. 842–844. doi: 10.1038/s41591-020-0901-9.

Link, C. S. *et al.* (2016) 'Abundant cytomegalovirus (CMV) reactive clonotypes in the CD8+ T cell receptor alpha repertoire following allogeneic transplantation', *Clinical and Experimental Immunology*, 184(3), pp. 389–402. doi: 10.1111/cei.12770.

Liu, L. *et al.* (2020) 'Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike', *Nature*. Springer US, 584(7821), pp. 450–456. doi: 10.1038/s41586-020-2571-7.

Liu, X. *et al.* (2016) 'Systematic comparative evaluation of methods for investigating the TCRβ repertoire', *PLoS ONE*, 11(3), pp. 1–18. doi: 10.1371/journal.pone.0152464.

Lönnberg, T. *et al.* (2017) 'Single-cell RNA-seq and computational analysis using temporal mixture modeling resolves TH1/TFH fate bifurcation in malaria', *Science Immunology*, 2(9), pp. 1–12. doi: 10.1126/sciimmunol.aal2192.

Luty, A. J. F. *et al.* (1999) 'Interferon-γ responses are associated with resistance to reinfection with Plasmodium falciparum in young African children', *Journal of Infectious Diseases*, 179(4), pp. 980–988. doi: 10.1086/314689.

Ly, A. and Hansen, D. S. (2019) 'Development of B cell memory in malaria', *Frontiers in Immunology*, 10(APR), pp. 1–11. doi: 10.3389/fimmu.2019.00559.

Lyke, K. E. *et al.* (2004) 'Serum levels of the proinflammatory cytokines interleukin-1 beta (IL-1β), IL-6, IL-8, IL-10, tumor necrosis factor alpha, and IL-12(p70) in Malian children with severe Plasmodium falciparum malaria and matched uncomplicated malaria or healthy controls', *Infection and Immunity*, 72(10), pp. 5630–5637. doi: 10.1128/IAI.72.10.5630-5637.2004.

Lythe, G. *et al.* (2016) 'How many TCR clonotypes does a body maintain?', *Journal of Theoretical Biology*. Elsevier, 389, pp. 214–224. doi: 10.1016/j.jtbi.2015.10.016.

MacDonald, R. A., Hosking, C. S. and Jones, C. L. (1988) 'The measurement of relative antibody affinity by ELISA using thiocyanate elution', *Journal of Immunological Methods*, 106(2), pp. 191–194. doi: 10.1016/0022-1759(88)90196-2.

Mackinnon, M. J., Walker, P. R. and Rowe, J. A. (2002) 'Plasmodium chabaudi : rosetting in a rodent malaria model', *Experimental Parasitology*, 101, pp. 121–128.

Mackroth, M. S. *et al.* (2016) 'Acute Malaria Induces PD1+CTLA4+ Effector T Cells with Cell-Extrinsic Suppressor Function', *PLoS Pathogens*, 12(11), pp. 1–24. doi: 10.1371/journal.ppat.1005909.

Madi, A. *et al.* (2014) 'T-cell receptor repertoires share a restricted set of public and abundant CDR3 sequences that are associated with self-related immunity', *Genome Research*, 24, pp. 1603–1612. doi: 10.1101/gr.170753.113.24.

Madi, A. *et al.* (2017) 'T cell receptor repertoires of mice and humans are clustered in similarity networks around conserved public CDR3 sequences', *eLife*, 3, pp. 1–17. doi: 10.7554/eLife.22057.

Mahanta, A. *et al.* (2015) 'Heightened inflammation in severe malaria is associated with decreased IL-10 expression levels and neutrophils', *Innate Immunity*, 21(5), pp. 546–552. doi: 10.1177/1753425914561277.

Mahe, F. *et al.* (2014) 'Swarm: robust and fast clustering method for amplicon-based studies', *PeerJ*, 2(593), pp. 1–13. doi: 10.7717/peerj.593.

Mamedov, I. Z. *et al.* (2013) 'Preparing unbiased T-cell receptor and antibody cDNA libraries for the deep next generation sequencing profiling', *Frontiers in Immunology*, 4(DEC), pp. 1–10. doi: 10.3389/fimmu.2013.00456.

Mamedov, M. R. *et al.* (2018) 'A Macrophage Colony-Stimulating-Factor-Producing γδ T Cell Subset Prevents Malarial Parasitemic Recurrence', *Immunity*. Elsevier Inc., 48(2), pp. 350-363.e7. doi: 10.1016/j.immuni.2018.01.009.

Mandala, W. L. *et al.* (2017) 'Cytokine Profiles in Malawian Children Presenting with Uncomplicated Malaria ', 24(4), pp. 1–11. doi: org/10.3389/fimmu.2018.02016.

Mandric, I. *et al.* (2020) 'Profiling immunoglobulin repertoires across multiple human tissues using RNA sequencing', *Nature Communications*, 11(1), pp. 1–14. doi: 10.1038/s41467-020-16857-7.

Marcou, Q., Mora, T. and Walczak, A. M. (2018) 'High-throughput immune repertoire analysis with IGoR', *Nature Communications*. Springer US, 9(1). doi: 10.1038/s41467-018-02832-w.

Mariotti-Ferrandiz, E. *et al.* (2016) 'A TCRB Repertoire Signature Can Predict Experimental Cerebral Malaria', *PLoS ONE*, 11(2), pp. 1–17. doi: 10.1371/journal.pone.0147871.

De Mast, Q. *et al.* (2008) 'A decrease of plasma macrophage migration inhibitory factor concentration is associated with lower numbers of circulating lymphocytes in experimental Plasmodium falciparum malaria', *Parasite Immunology*, 30(3), pp. 133–138. doi: 10.1111/j.1365-3024.2007.01008.x.

Maurice, N. J. *et al.* (2019) 'CXCR3 enables recruitment and site-specific bystander activation of memory CD8+ T cells', *Nature Communications*. Springer US, 10(1), pp. 1–15. doi: 10.1038/s41467-019-12980-2.

McHeyzer-Williams, L. J. *et al.* (1999) 'Evolution of antigen-specific T cell receptors in vivo: Preimmune and antigen-driven selection of preferred complementarity-determining region 3 (CDR3) motifs', *Journal of Experimental Medicine*, 189(11), pp. 1823–1837. doi: 10.1084/jem.189.11.1823.

Miles, J. J. *et al.* (2010) 'Genetic and Structural Basis for Selection of a Ubiquitous T Cell Receptor Deployed in Epstein-Barr Virus Infection', *PLoS Pathogens*, 6(11), pp. 1–15. doi: 10.1371/journal.ppat.1001198.

Miles, J. J., Douek, D. C. and Price, D. A. (2011) 'Bias in the αB T-cell repertoire: Implications for disease pathogenesis and vaccination', *Immunology and Cell Biology*, 89(3), pp. 375–387. doi: 10.1038/icb.2010.139.

Millar, J. A. *et al.* (2021) 'Spatial Organization and Recruitment of Non-Specific T Cells May Limit T Cell-Macrophage Interactions Within Mycobacterium tuberculosis Granulomas', *Frontiers in Immunology*, 11(January), pp. 1–16. doi: 10.3389/fimmu.2020.613638.

Milne, K. *et al.* (2021) 'Mapping immune variation and var gene switching in naive hosts infected with Plasmodium falciparum.', *eLife*, 10, pp. 1–31. doi: 10.7554/eLife.62800.

Minassian, A. M. *et al.* (2021) 'Reduced blood-stage malaria growth and immune correlates in humans following RH5 vaccination Reduced blood-stage malaria growth and immune', *Med*. Elsevier Inc., pp. 1–19. doi: 10.1016/j.medj.2021.03.014.

Minervina, A. *et al.* (2020) 'Primary and secondary anti-viral response captured by the dynamics and phenotype of individual T cell clones.', *eLife*, 9, pp. 1–21. doi: 10.7554/eLife.53704.

Minervina, A. A. *et al.* (2020) 'Longitudinal high-throughput TCR repertoire profiling reveals the dynamics of T cell memory formation after mild COVID-19 infection', *eLife*, (e63502). doi: https://doi.org/10.7554/eLife.63502.

Miyasaka, A. *et al.* (2019) 'Next-generation sequencing analysis of the human T-cell and B-cell receptor repertoire diversity before and after hepatitis B vaccination', *Human Vaccines and Immunotherapeutics*. Taylor & Francis, 15(11), pp. 2738–2753. doi: 10.1080/21645515.2019.1600987.

Mpina, M. *et al.* (2017) 'Controlled Human Malaria Infection Leads to Long-Lasting Changes in Innate and Innate-like Lymphocyte Populations', *The Journal of Immunology*, 199(1), pp. 107–118. doi: 10.4049/jimmunol.1601989.

Muñoz Sandoval, D. *et al.* (2021) 'Adaptive T cells regulate disease tolerance in human malaria', *In Prep*.

Murugan, A. *et al.* (2012) 'Statistical inference of the generation probability of T-cell receptors from sequence repertoires', *Proceedings of the National Academy of Sciences of the United States of America*, 109(40), pp. 16161–16166. doi: 10.1073/pnas.1212755109.

Nahrendorf, W., Ivens, A. and Spence, P. J. (2021) 'Inducible mechanisms of disease tolerance provide an alternative strategy of acquired immunity to malaria', *eLife*, 10, pp. 1–33. doi: 10.7554/eLife.63838.

Ndifon, W. *et al.* (2012) 'Chromatin conformation governs T-cell receptor Jβgene segment usage', *Proceedings of the National Academy of Sciences of the United States of America*, 109(39), pp. 15865–15870. doi: 10.1073/pnas.1203916109.

Ndungu, F. M. *et al.* (2006) 'CD4 T Cells from Malaria-Nonexposed Individuals Respond to the CD36-Binding Domain of Plasmodium falciparum Erythrocyte Membrane Protein-1 via an MHC Class II-TCR-Independent Pathway', *The Journal of Immunology*, 176(9), pp. 5504–5512. doi: 10.4049/jimmunol.176.9.5504.

Nguyen, T. N. *et al.* (2018) 'The persistence and oscillations of submicroscopic Plasmodium falciparum and Plasmodium vivax infections over time in Vietnam: an open cohort study', *The Lancet Infectious Diseases*. The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY 4.0 license, 18(5), pp. 565–572. doi: 10.1016/S1473-3099(18)30046-X.

Niu, X. *et al.* (2020) 'Longitudinal Analysis of T and B Cell Receptor Repertoire Transcripts Reveal Dynamic Immune Response in COVID-19 Patients', *Frontiers in Immunology*, 11(September), pp. 1–9. doi: 10.3389/fimmu.2020.582010.

Nlinwe, O. N. *et al.* (2018) 'T-cell responses against Malaria: Effect of parasite antigen diversity and relevance for vaccine development', *Vaccine*. Elsevier Ltd, 36(17), pp. 2237–2242. doi: 10.1016/j.vaccine.2018.03.023.

Oakes, T. *et al.* (2017) 'Quantitative characterization of the T cell receptor repertoire of naïve and memory subsets using an integrated experimental and computational pipeline which is robust, economical, and versatile', *Frontiers in Immunology*, 8(OCT), pp. 1–17. doi: 10.3389/fimmu.2017.01267.

Obeng-Adjei, N. *et al.* (2015) 'Circulating Th1-Cell-type Tfh Cells that Exhibit Impaired B Cell Help Are Preferentially Activated during Acute Malaria in Children', *Cell Reports*. The Authors, 13(2), pp. 425–439. doi: 10.1016/j.celrep.2015.09.004.

Oettinger, M. A. *et al.* (1990) 'RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination', *Science*, 248(4962), pp. 1517–1523. doi: 10.1126/science.2360047.

Ofori, M. F. *et al.* (2002) 'Malaria-induced acquisition of antibodies to Plasmodium falciparum variant surface antigens', *Infection and Immunity*, 70(6), pp. 2982–2988. doi: 10.1128/IAI.70.6.2982-2988.2002.

Oftedal, B. E. *et al.* (2017) 'T cell receptor assessment in autoimmune disease requires access to the most adjacent immunologically active organ', *Journal of Autoimmunity*, 81, pp. 24–33. doi: 10.1016/j.jaut.2017.03.002.

Opata, M. M. *et al.* (2015) 'Early Effector Cells Survive the Contraction Phase in Malaria Infection and

Generate Both Central and Effector Memory T Cells', *The Journal of Immunology*, 194(11), pp. 5346–5354. doi: 10.4049/jimmunol.1403216.

Opata, M. M. *et al.* (2018) 'Protection by and maintenance of CD4 effector memory and effector T cell subsets in persistent malaria infection', *PLoS pathogens*, 14(4), p. e1006960. doi: 10.1371/journal.ppat.1006960.

Opata, M. M. and Stephens, R. (2017) 'Chronic Plasmodium chabaudi infection generates CD4 memory T cells with increased T cell receptor sensitivity but poor secondary expansion and increased apoptosis', *Infection and Immunity*, 85(3), pp. 1–16. doi: 10.1128/IAI.00774-16.

Pedregosa, F. *et al.* (2011) 'Scikit-learn: Machine learning in Python', *Journal of Machine Learning Research*, 12, pp. 2825–2830. doi: 10.1007/s13398-014-0173-7.2.

Penkava, F. *et al.* (2020) 'Single-cell sequencing reveals clonal expansions of pro-inflammatory synovial CD8 T cells expressing tissue-homing receptors in psoriatic arthritis', *Nature Communications*. Springer US, 11(1). doi: 10.1038/s41467-020-18513-6.

Pérez-Mazliah, D. *et al.* (2015) 'Disruption of IL-21 Signaling Affects T Cell-B Cell Interactions and Abrogates Protective Humoral Immunity to Malaria', *PLoS Pathogens*, 11(3), pp. 1–24. doi: 10.1371/journal.ppat.1004715.

Pérez-Mazliah, D. *et al.* (2017) 'Follicular Helper T Cells are Essential for the Elimination of Plasmodium Infection', *EBioMedicine*. The Authors, 24, pp. 216–230. doi: 10.1016/j.ebiom.2017.08.030.

Pogorelyy, M. V., Minervina, A. A., Chudakov, D. M., *et al.* (2018) 'Method for identification of condition-associated public antigen receptor sequences', *eLife*, 7(D), p. e33050. doi: 10.7554/eLife.33050.

Pogorelyy, M. V., Minervina, A. A., Touzel, M. P., *et al.* (2018) 'Precise tracking of vaccine-responding T-cell clones reveals convergent and personalized response in identical twins', *PNAS*. doi: 10.1073/pnas.1809642115.

Pogorelyy, M. V. *et al.* (2019) 'Detecting T cell receptors involved in immune responses from single repertoire snapshots', *PLoS Biology*, 17(6), pp. 1–13. doi: 10.1371/journal.pbio.3000314.

Port, J. R. *et al.* (2020) 'Severe Human Lassa Fever Is Characterized by Nonspecific T-Cell Activation and Lymphocyte Homing to Inflamed Tissues', *Journal of Virology*, 94(21), pp. 1–18. doi: 10.1128/jvi.01367-20.

del Portillo, H. A. *et al.* (2012a) 'The role of the spleen in malaria', *Cellular Microbiology*, 14(3), pp. 343–355. doi: 10.1111/j.1462-5822.2011.01741.x.

del Portillo, H. A. *et al.* (2012b) 'The role of the spleen in malaria', *Cellular Microbiology*, 14(3), pp. 343–355. doi: 10.1111/j.1462-5822.2011.01741.x.

Portugal, S. *et al.* (2015) 'Malaria-associated atypical memory B cells exhibit markedly reduced B cell receptor signaling and effector function', *eLife*, 4(MAY), pp. 1–21. doi: 10.7554/eLife.07218.

Portugal, S. *et al.* (2017) 'Atypical memory B cells in human chronic infectious diseases: An interim report', *Cellular Immunology*. Elsevier, 321(June), pp. 18–25. doi: 10.1016/j.cellimm.2017.07.003.

Qi, Q. *et al.* (2014) 'Diversity and clonal selection in the human T-cell repertoire', *Proceedings of the National Academy of Sciences of the United States of America*, 111(36), pp. 13139–13144. doi: 10.1073/pnas.1409155111.

Reddy, S. B. *et al.* (2012) 'High affinity antibodies to plasmodium falciparum merozoite antigens are associated with protection from malaria', *PLoS ONE*, 7(2), pp. 1–9. doi: 10.1371/journal.pone.0032242.

Reece, W. H. H. *et al.* (2004) 'A CD4+ T-cell immune response to a conserved epitope in the circumsporozoite protein correlates with protection from natural Plasmodium falciparum infection and disease', *Nature Medicine*, 10(4), pp. 406–410. doi: 10.1038/nm1009.

Rempala, G. A. and Seweryn, M. (2013) 'Methods for diversity and overlap analysis in T-cell receptor populations', *Journal of Mathematical Biology*, 67(6–7), pp. 1339–1368. doi: 10.1007/s00285-012-0589-7.

Renyi, A. (1961) *On measures of entropy and information*, *Fourth Berkeley Symposium on Mathematical Statistics and Probability.* University of California Press; USA.

Ribot, J. C., Lopes, N. and Silva-Santos, B. (2021) 'γδ T cells in tissue physiology and surveillance', *Nature Reviews Immunology*. Springer US, 21(4), pp. 221–232. doi: 10.1038/s41577-020-00452-4.

Riggle, B. A. *et al.* (2020) 'CD8+ T cells target cerebrovasculature in children with cerebral malaria', *Journal of Clinical Investigation*, 130(3), pp. 1128–1138. doi: 10.1172/JCI133474.

Ritvo, P.-G. *et al.* (2018) 'High-resolution repertoire analysis reveals a major bystander activation of Tfh and Tfr cells', *Proceedings of the National Academy of Sciences*, 115(38), pp. 9604–9609. doi: 10.1073/pnas.1808594115.

Robins, H. S. *et al.* (2009) 'Comprehensive assessment of T-cell receptor β-chain diversity in αβ T cells', *Blood*, 114(19), pp. 4099–4107. doi: 10.1182/blood-2009-04-217604.

Robinson, M. W. *et al.* (2016) 'Tracking TCRβ sequence clonotype expansions during antiviral therapy using high-throughput sequencing of the hypervariable region', *Frontiers in Immunology*, 7(APR), pp. 1–9. doi: 10.3389/fimmu.2016.00131.

Romero, P. *et al.* (1989) 'Cloned cytotoxic T cells recognize an epitope in the circumsporozoite protein and protect against malaria', *Nature*, 341(6240), pp. 323–326. doi: 10.1038/341323a0.

Rosati, E. *et al.* (2017) 'Overview of methodologies for T-cell receptor repertoire analysis', *BMC Biotechnology*. BMC Biotechnology, 17(1), pp. 1–16. doi: 10.1186/s12896-017-0379-9.

RTSS Clinical Trials Partnership (2015) 'Effi cacy and safety of RTS , S / AS01 malaria vaccine with or without a booster dose in infants and children in Africa : fi nal results of a phase 3 , individually randomised , controlled trial', *The Lancet*. Elsevier Ltd, 6736(15), pp. 31–45. doi: 10.1016/S0140-6736(15)60721-8.

Ryg-Cornejo, V., Ly, A. and Hansen, D. S. (2016) 'Immunological processes underlying the slow acquisition of humoral immunity to malaria.', *Parasitology*, 143(02), pp. 1–9. doi: 10.1017/S0031182015001705.

Sanni, L. A. *et al.* (2004) 'Cerebral Edema and Cerebral Hemorrhages in Interleukin-10-Deficient Mice Infected with Plasmodium chabaudi', *Infection and Immunity*, 72(5), pp. 3054–3058. doi: 10.1128/IAI.72.5.3054-3058.2004.

Schattgen, S. A. *et al.* (2020) 'Linking T cell receptor sequence to transcriptional profiles with clonotype neighbor graph analysis (CoNGA)', *bioRxiv*. doi: 10.1038/s41587-020-0505-4.

Schatz, D. G., Oettinger, M. A. and Baltimore, D. (1989) 'The V(D)J recombination activating gene, RAG-1', *Cell*, 59(6), pp. 1035–1048. doi: 10.1016/0092-8674(89)90760-5.

Schmidt, N. W. *et al.* (2010) 'Extreme CD8 t cell requirements for anti-malarial liver-stage immunity following immunization with radiation attenuated sporozoites', *PLoS Pathogens*, 6(7), pp. 1–15. doi: 10.1371/journal.ppat.1000998.

Schober, K. *et al.* (2020) 'Reverse TCR repertoire evolution toward dominant low-affinity clones during chronic CMV infection', *Nature Immunology*. Springer US, 21(4), pp. 434–441. doi: 10.1038/s41590-020-0628-2.

Scholzen, A. *et al.* (2009) 'Plasmodium falciparum – Mediated Induction of Human CD25 hi Foxp3 hi CD4 T Cells Is Independent of Direct TCR Stimulation and Requires IL-2 , IL-10 and TGF b', *PLoS Pathogens*, 5(8), pp. 26–32. doi: 10.1371/journal.ppat.1000543.

Schultheiß, C. *et al.* (2020) 'Next-Generation Sequencing of T and B Cell Receptor Repertoires from COVID-19 Patients Showed Signatures Associated with Severity of Disease', *Immunity*. Elsevier Inc., 53(2), pp. 442-455.e4. doi: 10.1016/j.immuni.2020.06.024.

Servaas, N. H. *et al.* (2021) 'Longitudinal analysis of T-cell receptor repertoires reveals persistence of antigen-driven CD4+ and CD8+ T-cell clusters in systemic sclerosis', *Journal of Autoimmunity*. Elsevier Ltd, 117(September 2020), p. 102574. doi: 10.1016/j.jaut.2020.102574.

Sethna, Z. *et al.* (2018) 'OLGA: fast computation of generation probabilities of B- and T-cell receptor amino acid sequences and motifs', *Bioinformatics*, 35(January), pp. 2974–2981. doi: 10.1093/bioinformatics/btz035.

Sewell, A. K. (2012) 'Why must T cells be cross-reactive?', *Nature Reviews Immunology*. Nature Publishing Group, 12(9), pp. 669–677. doi: 10.1038/nri3279.

Sharma, G., Rive, C. M. and Holt, R. A. (2019) 'Rapid selection and identification of functional CD8+ T cell epitopes from large peptide-coding libraries', *Nature Communications*. Springer US, 10(1), pp. 1–13. doi: 10.1038/s41467-019-12444-7.

Sheehy, S. H. *et al.* (2012) 'ChAd63-MVA-vectored blood-stage Malaria vaccines targeting MSP1 and AMA1: Assessment of efficacy against mosquito bite challenge in humans', *Molecular Therapy*. The American Society of Gene & Cell Therapy, 20(12), pp. 2355–2368. doi: 10.1038/mt.2012.223.

Shi, W. *et al.* (2015) 'Transcriptional profiling of mouse B cell terminal differentiation defines a signature for antibody-secreting plasma cells', *Nature Immunology*, 16(6), pp. 663–673. doi: 10.1038/ni.3154.

Shomuradova, A. S. *et al.* (2020) 'SARS-CoV-2 Epitopes Are Recognized by a Public and Diverse Repertoire of Human T Cell Receptors', *Immunity*. Elsevier Inc., 53(6), pp. 1245-1257.e5. doi: 10.1016/j.immuni.2020.11.004.

Shretta, R., Avanceña, A. L. V. and Hatefi, A. (2016) 'The economics of malaria control and elimination: a systematic review', *Malaria Journal*. BioMed Central, 15(1), pp. 1–14. doi: 10.1186/s12936-016-1635-5.

Shugay, M. *et al.* (2014) 'Towards error-free profiling of immune repertoires', 11(6), pp. 6–10. doi: 10.1038/nmeth.2960.

Shugay, M. *et al.* (2015) 'VDJtools: Unifying Post-analysis of T Cell Receptor Repertoires', *PLoS Computational Biology*, 11(11), pp. 1–16. doi: 10.1371/journal.pcbi.1004503.

Shugay, M. *et al.* (2018) 'VDJdb : a curated database of T-cell receptor sequences with known antigen specificity', *Nucleic Acids Research*. Oxford University Press, 46(September 2017), pp. 419–427. doi: 10.1093/nar/gkx760.

Simnica, D. *et al.* (2019) 'T cell receptor next-generation sequencing reveals cancer-associated repertoire metrics and reconstitution after chemotherapy in patients with hematological and solid tumors', *OncoImmunology*. Taylor & Francis, 8(11). doi: 10.1080/2162402X.2019.1644110.

De Simone, M., Rossetti, G. and Pagani, M. (2018) 'Single cell T cell receptor sequencing: Techniques

and future challenges', *Frontiers in Immunology*, 9(JUL), pp. 1–7. doi: 10.3389/fimmu.2018.01638.

Singh, M. *et al.* (2019) 'High-throughput targeted long-read single cell sequencing reveals the clonal and transcriptional landscape of lymphocytes', *Nature Communications*. Springer US, 10(1). doi: 10.1038/s41467-019-11049-4.

Singhania, A. *et al.* (2019) 'Transcriptional profiling unveils type I and II interferon networks in blood and tissues across diseases', *Nature Communications*, 10(1), pp. 1–21. doi: 10.1038/s41467-019-10601-6.

Smith, J. D. *et al.* (1995) 'Switches in expression of plasmodium falciparum var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes', *Cell*, 82(1), pp. 101–110. doi: 10.1016/0092-8674(95)90056-X.

Smith, N. L. *et al.* (2020) 'A Conserved TCRβ signature dominates a highly polyclonal T-cell expansion during the acute phase of a murine malaria infection', *Frontiers in Immunology*, 11(November), p. 2020.07.27.222570. doi: 10.3389/fimmu.2020.587756.

Snow, R. W. *et al.* (1998) 'Risk of severe malaria among African infants: Direct evidence of clinical protection during early infancy', *Journal of Infectious Diseases*, 177(3), pp. 819–822. doi: 10.1086/517818.

Sood, S. *et al.* (2016) 'Loss of the Protein Tyrosine Phosphatase PTPN22 Reduces Mannan-Induced Autoimmune Arthritis in SKG Mice', *J Immunol*, (197), pp. 1–11. doi: 10.4049/jimmunol.1502656.

Soon, M. S. F. *et al.* (2020) 'Transcriptome dynamics of CD4+ T cells during malaria maps gradual transit from effector to memory', *Nature Immunology*. Springer US, 21(12), pp. 1597–1610. doi: 10.1038/s41590-020-0800-8.

Soon, M. S. F. and Haque, A. (2018) ' Recent Insights into CD4 + Th Cell Differentiation in Malaria ', *The Journal of Immunology*, 200(6), pp. 1965–1975. doi: 10.4049/jimmunol.1701316.

Spence, P. J. *et al.* (2013) 'Vector transmission regulates immune control of Plasmodium virulence.', *Nature*, 498, pp. 228–31. doi: 10.1038/nature12231.

Spence, P. J., Brugat, T. and Langhorne, J. (2015) 'Mosquitoes Reset Malaria Parasites', *PLoS Pathogens*, 11(7), pp. 10–15. doi: 10.1371/journal.ppat.1004987.

Spencer, A. J. *et al.* (2017) ' The Threshold of Protection from Liver-Stage Malaria Relies on a Fine Balance between the Number of Infected Hepatocytes and Effector CD8 + T Cells Present in the Liver ', *The Journal of Immunology*, 198(5), pp. 2006–2016. doi: 10.4049/jimmunol.1601209.

Ssewanyana, I. *et al.* (2017) 'Avidity of anti-malarial antibodies inversely related to transmission intensity at three sites in Uganda', *Malaria Journal*. BioMed Central, 16(1), p. 67. doi: 10.1186/s12936-017-1721-3.

Stanisic, D. I. *et al.* (2015) 'Acquisition of antibodies against Plasmodium falciparum merozoites and malaria immunity in young children and the influence of age, force of infection, and magnitude of response', *Infection and Immunity*, 83(2), pp. 646–660. doi: 10.1128/IAI.02398-14.

Stephens, R., Culleton, R. L. and Lamb, T. J. (2012) 'The contribution of Plasmodium chabaudi to our understanding of malaria', *Trends in Parasitology*. Elsevier Ltd, 28(2), pp. 74–83. doi: 10.1016/j.pt.2011.10.006.

Stephens, R. and Langhorne, J. (2010) 'Effector memory Th1 CD4 T cells are maintained in a mouse model of chronic malaria', *PLoS Pathogens*, 6(11), p. e1001208. doi: 10.1371/journal.ppat.1001208.

Su, Z. and Stevenson, M. M. (2000) 'Central role of endogenous gamma interferon in protective

immunity against blood-stage Plasmodium chabaudi AS infection', *Infection and Immunity*, 68(8), pp. 4399–4406. doi: 10.1128/IAI.68.8.4399-4406.2000.

Sun, P. *et al.* (2003) ' Protective Immunity Induced with Malaria Vaccine, RTS,S, Is Linked to Plasmodium falciparum Circumsporozoite Protein-Specific CD4 + and CD8 + T Cells Producing IFN-γ ', *The Journal of Immunology*, 171(12), pp. 6961–6967. doi: 10.4049/jimmunol.171.12.6961.

Sun, Y. *et al.* (2017) 'Specificity, privacy, and degeneracy in the CD4 T cell receptor repertoire following immunization', *Frontiers in Immunology*, 8(APR), pp. 1–12. doi: 10.3389/fimmu.2017.00430.

Surette, F. A. *et al.* (2021) 'Extrafollicular CD4 T cell-derived IL-10 functions rapidly and transiently to support anti-Plasmodium humoral immunity', *PLoS Pathogens*, 17(2), pp. 1–21. doi: 10.1371/JOURNAL.PPAT.1009288.

Swanson, P. A. *et al.* (2016) 'CD8+T Cells Induce Fatal Brainstem Pathology during Cerebral Malaria via Luminal Antigen-Specific Engagement of Brain Vasculature', *PLoS Pathogens*, 12(12), pp. 1–34. doi: 10.1371/journal.ppat.1006022.

Thera, M. A. *et al.* (2011) 'A Field Trial to Assess a Blood-Stage Malaria Vaccine', *New England Journal of Medicine*, 365(11), pp. 1004–1013. doi: 10.1056/nejmoa1008115.

Thomas, N. *et al.* (2014) 'Tracking global changes induced in the CD4 T-cell receptor repertoire by immunization with a complex antigen using short stretches of CDR3 protein sequence', *Bioinformatics*, 30(22), pp. 3181–3188. doi: 10.1093/bioinformatics/btu523.

Tickotsky, N. *et al.* (2017) 'McPAS-TCR: A manually curated catalogue of pathology-associated T cell receptor sequences', *Bioinformatics*, 33(18), pp. 2924–2929. doi: 10.1093/bioinformatics/btx286.

Tubo, N. J. *et al.* (2013) 'Single naive CD4+ T cells from a diverse repertoire produce different effector cell types during infection', *Cell*, 153(4), pp. 785–796. doi: 10.1016/j.cell.2013.04.007.

Turchaninova, M. A. *et al.* (2016) 'High-quality full-length immunoglobulin profiling with unique molecular barcoding', *Nature Protocols*, 11(9), pp. 1599–1616. doi: 10.1038/nprot.2016.093.

Turner, S. J. *et al.* (2004) 'Characterization of CD8+ T cell repertoire diversity and persistence in the influenza a virus model of localized, transient infection', *Seminars in Immunology*, 16(3), pp. 179–184. doi: 10.1016/j.smim.2004.02.005.

Vavassori, S. *et al.* (2013) 'Butyrophilin 3A1 binds phosphorylated antigens and stimulates human γδ T cells', *Nature Immunology*, 14(9), pp. 908–916. doi: 10.1038/ni.2665.

Venturi, V. *et al.* (2006) 'Sharing of T cell receptors in antigen-specific responses is driven by convergent recombination', *PNAS*, 103(49), pp. 18691–18696. doi: 10.1073/pnas.0608907103.

Venturi, V. *et al.* (2007) 'Methods for comparing the diversity of samples of the T cell receptor repertoire', *Journal of Immunological Methods*, 321(1–2), pp. 182–195. doi: 10.1016/j.jim.2007.01.019.

Virtanen, P. *et al.* (2020) 'SciPy 1.0: fundamental algorithms for scientific computing in Python', *Nature Methods*, 17(3), pp. 261–272. doi: 10.1038/s41592-019-0686-2.

Waickman, A. T. *et al.* (2019) 'Dissecting the heterogeneity of DENV vaccine-elicited cellular immunity using single-cell RNA sequencing and metabolic profiling', *Nature Communications*. Springer US, 10(1), pp. 1–16. doi: 10.1038/s41467-019-11634-7.

Walther, M. *et al.* (2009) 'Distinct roles for FOXP3+ and FOXP32 CD4+ T cells in regulating cellular immunity to uncomplicated and severe plasmodium falciparum malaria', *PLoS Pathogens*, 5(4). doi: 10.1371/journal.ppat.1000364.

Wang, X. *et al.* (2021) 'Host-derived lipids orchestrate pulmonary γδ T cell response to provide early protection against influenza virus infection', *Nature Communications*. Springer US, 12(1), pp. 1–19. doi: 10.1038/s41467-021-22242-9.

Weiss, D. J. *et al.* (2019) 'Mapping the global prevalence, incidence, and mortality of Plasmodium falciparum, 2000–17: a spatial and temporal modelling study', *The Lancet*. The Authors. Published by Elsevier Ltd. This is an Open Access article under the CC BY 4.0 license, 394(10195), pp. 322–331. doi: 10.1016/S0140-6736(19)31097-9.

Weiss, G. E. *et al.* (2010) 'The Plasmodium falciparum-Specific Human Memory B Cell Compartment Expands Gradually with Repeated Malaria Infections', *PLoS Pathogens*, 6(5), p. e1000912. doi: 10.1371/journal.ppat.1000912.

Weiss, W. R. *et al.* (1988) 'CD8+ T cells (cytotoxic/suppressors) are required for protection in mice immunized with malaria sporozoites', *Proceedings of the National Academy of Sciences of the United States of America*, 85(2), pp. 573–576. doi: 10.1073/pnas.85.2.573.

Whiteside, S. K., Snook, J. P., Williams, M. A., *et al.* (2018) 'Bystander T Cells: A Balancing Act of Friends and Foes', *Trends in Immunology*. Elsevier Ltd, 39(12), pp. 1021–1035. doi: 10.1016/j.it.2018.10.003.

Whiteside, S. K., Snook, J. P., Ma, Y., *et al.* (2018) 'IL-10 Deficiency Reveals a Role for TLR2-Dependent Bystander Activation of T Cells in Lyme Arthritis', *The Journal of Immunology*, 200(4), pp. 1457–1470. doi: 10.4049/jimmunol.1701248.

WHO (2020) *World Malaria Report 2020. World Health Organization.*, *World Health*.

WHO Strategic Advisory Group on Malaria Eradication (2019) *Malaria eradication: benefits, future scenarios and feasibility. A report of the Strategic Advisory Group on Malaria Eradication*, *Geneva: World Health Organization*.

Willcox, B. E. and Willcox, C. R. (2019) 'γδ TCR ligands: the quest to solve a 500-million-year-old mystery', *Nature Immunology*. Springer US, 20(2), pp. 121–128. doi: 10.1038/s41590-018-0304-y.

Wipasa, J. *et al.* (2011) 'Short-lived IFN-γ effector responses, but long-lived IL-10 memory responses, to malaria in an area of low malaria endemicity', *PLoS Pathogens*, 7(2), p. e1001281. doi: 10.1371/journal.ppat.1001281.

Wolf, K., Hether, T., Gilchuk, P., Kumar, A., *et al.* (2018) 'Identifying and Tracking Low-Frequency Virus-Specific TCR Clonotypes Using High-Throughput Sequencing', *Cell Reports*. ElsevierCompany., 25(9), pp. 2369-2378.e4. doi: 10.1016/j.celrep.2018.11.009.

Wolf, K., Hether, T., Gilchuk, P., Ahn, T., *et al.* (2018) 'Identifying and Tracking Low-Frequency Virus-Specific TCR Clonotypes Using High-Throughput Sequencing', *Cell Reports*. ElsevierCompany., 25(9), pp. 2369-2378.e4. doi: 10.1016/j.celrep.2018.11.009.

Van Wolfswinkel, M. E. *et al.* (2017) 'Changes in total and differential leukocyte counts during the clinically silent liver phase in a controlled human malaria infection in malaria-naïve Dutch volunteers', *Malaria Journal*. BioMed Central, 16(1), pp. 1–8. doi: 10.1186/s12936-017-2108-1.

Wong, G. K. *et al.* (2017) 'Immune dysregulation in immunodeficiency disorders: The role of T-cell receptor sequencing', *Journal of Autoimmunity*. Elsevier Ltd, 80, pp. 1–9. doi: 10.1016/j.jaut.2017.04.002.

Xu, J. L. and Davis, M. M. (2000) 'Diversity in the CDR3 Region of V H Is Sufficient for Most Antibody Specificities', *Immunity*, 13, pp. 37–45.

Yañez, D. M. *et al.* (1996) 'Participation of lymphocyte subpopulations in the pathogenesis of experimental murine cerebral malaria.', *Journal of immunology (Baltimore, Md.: 1950)*, 157(4), pp. 1620–4.

Zarnitsyna, V. I. *et al.* (2013) 'Estimating the diversity, completeness, and cross-reactivity of the T cell repertoire', *Frontiers in Immunology*, 4(DEC), pp. 1–11. doi: 10.3389/fimmu.2013.00485.

Zhang, H. *et al.* (2020) 'Investigation of antigen-specific T-cell receptor clusters in human cancers', *Clinical Cancer Research*, 26(6), pp. 1359–1371. doi: 10.1158/1078-0432.CCR-19-3249.

Zhang, J. Y. *et al.* (2020) 'Single-cell landscape of immunological responses in patients with COVID-19', *Nature Immunology*. Springer US, 21(9), pp. 1107–1118. doi: 10.1038/s41590-020-0762-x.

Zhang, N. and Bevan, M. J. (2011) 'CD8+ T Cells: Foot Soldiers of the Immune System', *Immunity*. Elsevier Inc., 35(2), pp. 161–168. doi: 10.1016/j.immuni.2011.07.010.

Zheng, G. X. Y. *et al.* (2017) 'Massively parallel digital transcriptional profiling of single cells', *Nature Communications*. Nature Publishing Group, 8, p. 14049. doi: 10.1038/ncomms14049.

frontiers in Immunology

ORIGINAL RESEARCH
published: 23 November 2020
doi: 10.3389/fimmu.2020.587756

A Conserved TCRβ Signature Dominates a Highly Polyclonal T-Cell Expansion During the Acute Phase of a Murine Malaria Infection

Natasha L. Smith, Wiebke Nahrendorf, Catherine Sutherland, Jason P. Mooney, Joanne Thompson, Philip J. Spence and Graeme J. M. Cowan*

Institute of Immunology and Infection Research, School of Biological Sciences, University of Edinburgh, Edinburgh, United Kingdom

CD4$^+$ αβ T-cells are key mediators of the immune response to a first *Plasmodium* infection, undergoing extensive activation and splenic expansion during the acute phase of an infection. However, the clonality and clonal composition of this expansion has not previously been described. Using a comparative infection model, we sequenced the splenic CD4$^+$ T-cell receptor repertoires generated over the time-course of a *Plasmodium chabaudi* infection. We show through repeat replicate experiments, single-cell RNA-seq, and analyses of independent RNA-seq data, that following a first infection - within a highly polyclonal expansion - T-effector repertoires are consistently dominated by TRBV3 gene usage. Clustering by sequence similarity, we find the same dominant clonal signature is expanded across replicates in the acute phase of an infection, revealing a conserved pathogen-specific T-cell response that is consistently a hallmark of a first infection, but not expanded upon re-challenge. Determining the host or parasite factors driving this conserved response may uncover novel immune targets for malaria therapeutic purposes.

Keywords: T-cell, TCR, immune-repertoire, single-cell, *Plasmodium chabaudi*, malaria

OPEN ACCESS

Edited by:
Juarez Antonio Simões Quaresma, Evandro Chagas Institute, Brazil

Reviewed by:
Henrique Borges da Silva, University of Minnesota Twin Cities, United States
Gregoire S. Lauvau, Albert Einstein College of Medicine, United States

****Correspondence:***
Natasha L. Smith
n.l.smith-2@sms.ed.ac.uk

Specialty section:
This article was submitted to Microbial Immunology, a section of the journal Frontiers in Immunology

Received: *27 July 2020*
Accepted: *27 October 2020*
Published: *23 November 2020*

Citation:
Smith NL, Nahrendorf W, Sutherland C, Mooney JP, Thompson J, Spence PJ and Cowan GJM (2020) A Conserved TCRβ Signature Dominates a Highly Polyclonal T-Cell Expansion During the Acute Phase of a Murine Malaria Infection. Front. Immunol. 11:587756. doi: 10.3389/fimmu.2020.587756

INTRODUCTION

Although protective natural immunity against clinical malaria is slow to develop and requires years of repeated exposure (1), protection against severe disease is obtained after a more limited number of symptomatic infections (2, 3). The acquisition of this naturally acquired immunity is mediated by both antibody [reviewed in (4, 5)] and T-cell responses (6); the latter being crucial for B-cell class switching and affinity maturation. As well as guiding the humoral response, CD4$^+$ T-cells play a key role in restricting the growth and pathogenesis of blood-stage *Plasmodium* through cytokine secretion and macrophage activation [reviewed in (7)]. However, the antigenic drivers and developmental dynamics underlying this naturally acquired immunity remain poorly understood, presenting major challenges for effective vaccine design.

In animal models of malaria, a *Plasmodium* infection in previously unexposed individuals initially produces a massive expansion of CD4$^+$ T-cells in the spleen (8, 9), a major site of the developing immune response (10). The size of this response, together with the generation of a highly diverse range of cellular responses, suggests that the splenic expansion of CD4$^+$ populations is

highly polyclonal, as opposed to the expansion of a minor (oligoclonal) subset of the repertoire. However, it is not known whether this expansion is primarily a non-specific response, such as a result of cytokine-driven bystander activation, or whether it is dominated by antigen-specific responses generated through classic TCR-engagement mediated clonal expansion. Spectratyping (CDR3 length analysis) of T-cell receptor (TCR) β chain repertoires induced by the rodent malaria *Plasmodium berghei* has previously detected a unique TCRβ CDR3 length signature enhanced over the course of infection, suggesting that there may be a clonal response to specific antigenic peptides (11). In agreement with this, an expanded fraction of $CD4^+$ T-cells and fast-responding cytokine secretors that respond to a secondary challenge has been observed following a *Plasmodium chabaudi (AJ)* infection in mice, indicating initial priming by the parasite, and the presence of pathogen-specific T-cells within the $CD4^+$ T-cell population (9). Alternatively, there is evidence from *P. falciparum*, that the PfEMP1 binding domain, CIDR-1α, stimulates $CD4^+$ T-cells non-specifically through TCR-independent pathways (12), and that regulatory T-cell (Treg) proliferation during an infection can be induced in an antigen non-specific manner (13). Non-specific proliferation of T-cells due to cross-reactivity in response to *P. falciparum* antigens has also been reported (14).

Overall, proliferation is likely to be a combination of activation dynamics. However, whether a detectable clonal malaria-specific $CD4^+$ T cell response that is conserved between individuals, and thus a potential focal target for therapeutics, is induced, has not previously been demonstrated.

Advances in high-throughput TCR repertoire sequencing techniques now allow deep profiling of immune responses. This approach has been used to ascertain clonality of T-cell responses, identify expanded T-cell clones and determine if conserved or 'public' responses between individuals are generated following antigenic stimulation [reviewed in (15)]. Repertoire sequencing thus provides a novel, immune-focused approach to delineate the clonality of the developing immune response to a malaria infection.

Here, using bulk TCRβ repertoire sequencing we examine the dynamics and clonal structure of the splenic $CD4^+$ T-cell repertoires generated during infection with the well-established mouse malaria model *Plasmodium chabaudi (AS)*. By comparing serially blood passaged (SBP) and recently mosquito-transmitted (MT) *P. chabaudi* infections, Spence et al. (8) demonstrated that vector transmission of *P. chabaudi* intrinsically modified parasite gene expression in asexual blood-stage parasites, eliciting an altered host immune response that in turn regulates parasite virulence. In this model, infection with SBP parasites leads to hyperparasitaemia with more severe disease during the acute phase of infection. In contrast, mosquito transmission (MT) attenuates parasite growth and virulence, through a mechanism associated with epigenetic reprogramming of the expression of the subtelomeric multigene families, including the variant surface antigen (VSA) family. We have used this comparative model to compare TCR repertoires integral to an immune response that rapidly controls parasite growth, against a less effective response that fails to control parasite replication and induces immunopathology (8, 16). We have sequenced the T-naive (T_N), T-effector (T_E), T-effector memory (T_{EM}) and T-central memory (T_{CM}) $CD4^+$ splenic TCRβ repertoires elicited in mice over the time-course of both MT and SBP *P. chabaudi* infections. We report, for both infection types, that the T_E expansion seen during the acute phase of a *P. chabaudi* infection is highly polyclonal. However, within this diverse expansion, a conserved pathogen-specific response characterized by TRBV3 gene usage consistently dominates the effector repertoire following a first infection, and we further profile this response using single-cell RNA-seq.

MATERIALS AND METHODS

Mice Infections

All procedures were carried out in accordance with UK Home Office regulations (Animals Scientific Procedures Act, 1986; project license number 70/8,546 and P04ABDCAA) and were approved by the Ethical Review Body of the University of Edinburgh. C57Bl/6 mice were bred and housed under specific pathogen free conditions at the University of Edinburgh and subjected to regular pathogen monitoring by sentinel screening. Mice were housed with at least one companion in individually ventilated cages furnished with autoclaved woodchip, fun tunnel and tissue paper at 21°C ± 2°C under a reverse light-dark cycle (light, 19.00 – 07.00; dark, 07.00 – 19.00) at a relative humidity of 55% ± 10%. *P. chabaudi (AS)* parasites were obtained from the European Malaria Reagent Repository at the University of Edinburgh. Eight- to 10-week-old C57Bl/6 female mice were infected with *P. chabaudi (AS)* by intra-peritoneal injection of $1x10^5$ parasitized erythrocytes that had either been maintained by serial blood-passage over a high number of generations (SBP) or undergone a single passage following mosquito transmission (MT) as per Spence et al. (17). Each transmission group consisted of five cages of five mice, with five unchallenged mice from the same cohort used as experimental controls. Mice (n=5 per transmission group) were euthanized on days 6, 10, 20, 40, and 60 post-infection. In a repeat experiment, mice (n=4 per time-point) were infected with MT *P. chabaudi (AS)-GFP* (18) by intra-peritoneal injection of $1x10^5$ parasitized erythrocytes, and were euthanized at days 4, 7, 11, and 14 post-infection. For the re-challenge experiment, mice were infected with MT *P. chabaudi (AS)-GFP* by intra-peritoneal injection of $1x10^5$ parasitized erythrocytes. Mice (n=4) underwent a homologous re-challenge at day 60 post infection and were euthanized 7 days post re-challenge (day 67). To eliminate chronic infection before re-challenge, 0.288 mg/ml of chloroquine diphosphate salt (Sigma), supplemented with glucose for palatability, was dissolved in drinking water daily for 10 days, from day 30 to day 40 post-infection (19). For the single-cell RNA-seq experiment, mice (n=2) were infected with MT *P. chabaudi (AS)-GFP* by intra-peritoneal injection of $1x10^5$ parasitized erythrocytes, and euthanized at day 7 post-infection.

Cell Sorting

$CD4^+$ splenic T-cell populations of interest were isolated by FACS using a BD FACSAria III instrument, according to gates described by Spence et al. (8): T_N ($CD62L^+$ $CD127^+$), T_E ($CD62L^-$ $CD127^-$), $T_{EM,}$ (CD44HI $CD127^+$ $CD62L^-$) (20), and T_{CM} (CD44HI $CD127^+$ $CD62L^+$) (**Supplementary Figure 1**). For the repeat and re-challenge experiments, only T_E and T_N populations were isolated. Cells were sorted into 50 µl FACS buffer and stored at -80°C until processing. For the single-cell experiment, $CD3^+$ $CD4^+$ splenocytes were sorted in to 100 µl 0.04% BSA in PBS, to generate the single-cell suspension required for 10x Genomics sequencing.

Unbiased Bulk TCR Amplification and Sequencing

RNA was extracted from $CD4^+$ splenic T-cell populations of interest using Dynabeads mRNA purification kit (ThermoFisher Scientific). cDNA was synthesized from each RNA preparation by adding the following to each sample: 4 µl First-strand Buffer, 2 µl 10mM dNTP mix, 2 µl 20mM DTT and 2 µl SMART-PTO2 oligo (5' AAGCAGTGGTATCAACGGAGAGTACATrGrGr 3'), 0.5 µl RNase inhibitor (Clontech 2313A), and 2 µl (100 U/µl) SMARTScribe reverse transcriptase (Clontech). For the repeat experiments, unique molecular identifiers (UMIs) (21) were incorporated during cDNA synthesis by replacing the template-switch oligo with 2 µl SMARTNNN oligo (5'AAGCAGUGGTAUCAACGCAGAGUNNNNUNNNNUN NNNUCTTrGrGrG 3'). Samples were then incubated at 42°C for 70 min, before the reaction was terminated by heating at 70°C for 10 min. cDNA synthesized with SMARTNNN oligos were treated with 1 µl of Uracil DNA glycosylase (5 U/µl, New England Biolabs) and incubated for 15 min at 37°C. PCR was then used to generate TCRβ V-region amplicons, using indexed forward primers composed of the SMART synthesis oligo sequence fused to a P7 Illumina tag, and a reverse primer within the TCR-C region fused to a P5 Illumina tag (P5-mTCRBrev3: 5' AATGATACGGCGACCACCGAGATCT ACACCTTGGGTGGAGTCACATTTCT 3'). Amplified products were purified by extraction from excised agarose gel bands. Single-end 1x400bp and asymmetric 100bp+400bp (to incorporate UMIs) sequencing was performed on an Illumina MiSeq platform.

Bulk TCR Repertoire Analyses

Bulk TCR sequence data was initially processed using MiGEC (22) and MiXCR (23) software with default settings. Samples were excluded from further analyses if the repertoire contained fewer than 10,000 total reads after processing, as this was indicative of poor sample preservation or preparation. A combination of custom pipelines of Python scripts and VDJtools software (24) was used to analyze and plot the MiXCR output. Statistical analyses were performed using SciPy Python software (25). A TCR clone was defined by 100% amino acid sequence identity of the CDR3 region, and IMGT nomenclature used for gene usage. Only in-frame (functional) CDR3s were analyzed. A modified version of the Swarm algorithm (26) was used to cluster highly homologous CDR3 amino sequences, with identical V-gene usage, within 1 amino-acid mismatch of each other. The Gliph2 package (27) was also used to identify enriched amino acid motifs within the contact region of CDR3 sequences; the unchallenged T-naïve repertoires were used to make custom murine reference files for this. Network analyses was undertaken using Gephi (28) software (v0.9.2). Generation probability of TCRs (Pgen) was calculated using OLGA (29).

Publicly Available RNA-Seq Data Analyses

Raw FASTQ files from RNA-seq data obtained from the spleens of C57Bl/6 mice infected with blood-stage *Plasmodium chabaudi (AS)* and *Plasmodium chabaudi (CB)* were downloaded from the ArrayExpress archive: ENA - ERP004042 and ENA - ERP005730 respectively. For comparison of infection with other pathogens, raw FASTQ files of RNA-seq data obtained from whole blood of C57Bl/6 mice infected with a variety of pathogens, were downloaded from the NCBI short read archive, under accession SRR7821557. Normalized Trbv3 gene expression values were also obtained from Singhania et al. (30). RNA-seq data was aligned using MiXCR (23) software, and a combination of custom pipelines of Python scripts and VDJtools software (24) was used to analyze and plot the MiXCR output.

Single-Cell Sequencing, Data Processing, and Analyses

Two barcoded cDNA libraries were prepared from sorted samples using the Chromium Single Cell 5' Library Kit v2 (31). Full length V(D)J segments were enriched from amplified cDNA with primers specific to the TCR constant region using the Chromium Single Cell V(D)J Enrichment Kit – Mouse T-Cell. Sequencing was performed using the High-Output v2.5 Kit on a NextSeq 550 platform. Initial processing of sequence files, including mapping of reads to the mouse reference genome (GRCm38), generation of count matrixes and assembly of TCR alpha and beta chains, was carried out using CellRanger 3.1.0. To exclude potential multiplets, poor quality cells or non T-cells, single T-cells were identified by the expression of a single productive beta chain. Barcodes lacking a beta chain or assigned to multiple were excluded, leaving data from 3,333 single T-cells (1,658 and 1,675 from mouse 1 and 2 respectively). Downstream analyses were performed in R using Seurat 3.1.5 (32). Genes expressed in fewer than three cells, as well as all *Trav/j* and *Trbv/d/j* genes were excluded. Cells expressing fewer than 200, or over 3,000 genes and/or more than 5% mitochondrial genes were removed. The filtered matrix was normalized using Seurat's LogNormalize with default parameters and the top 2,000 variable genes were identified using the FindVariableFeatures 'vst' method, before centering and scaling of the matrix. Dimensionality reduction by PCA was carried out and the top 30 principal components were used as input for graph-based clustering. Clusters were visualized by UMAP. A small, poorly *Cd4*-expressing cluster was identified, and these cells were excluded as contaminants. The above normalization and clustering steps were repeated with the remaining 2,976 cells

(1,491 and 1,485 from mouse 1 and 2 respectively). Differential gene expression analysis using the Wilcoxon rank sum test through FindAllMarkers was used to identify marker genes for each cluster.

RESULTS

Mice were infected with *P. chabaudi (AS)* parasitized erythrocytes from donor mice infected with either recently MT or SBP parasites, and followed for 60 days of infection. For repeat experiments, only MT parasites were used, as these represent a less artificial experimental model. At each time point, spleens were harvested and $CD4^+$ splenic T-cells populations of interest were isolated by FACS before RNA was extracted, reverse transcribed, and TCRβ chains amplified and sequenced.

TRBV3 Gene Usage Dominates a Highly Polyclonal T-Effector Expansion

Consistent with previously published data for *P. chabaudi* (8, 33), $CD4^+$ splenic T-effectors (T_E) reached maximum levels of expansion in the acute phase of infection, increasing by up to 10-fold. Expansion coincided with the peak of parasitemia and contracted back to pre-challenged levels between days 20 and 40 post-infection (**Figures 1A, B**). We first hypothesized that if the T_E expansion in the acute phase of infection was solely the result of non-specific activation, V gene usage and V/J allele usage within the T_E repertoire would mirror that of T_N repertoires, despite the vast cellular proliferation. Thus, there would be no change in the distribution of V or V/J allele usage post parasite-challenge. However, a distinct increase in TRBV3 gene usage was observed during the acute phase of infection, differentiating challenged T_E repertoires from both unchallenged T_N and T_E repertoires, and from challenged T_E repertoires at later time-points (**Figures 2A, B**, **Supplementary Figure 2**). For MT infections during the acute phase, TRBV3 encoded on average 23.7% (± 2.03 95% CI) of the effector repertoire at day 6 and 21.6% (± 2.21 95% CI) at day 10 post-infection, compared to only 7.6% (± 0.47, 95% CI) of the unchallenged naïve repertoire. This finding was repeated in a second independent experiment (**Supplementary Figure 3**), and through analysis of publicly available RNA-seq data for *P. chabaudi (AS)* and *P. chabaudi (CB)* (accession E-ERAD-221 and E-ERAD-289 respectively, **Supplementary Figure 4**). This increase in TRBV3 usage was more delayed in SBP challenged repertoires, and not apparent until day 10 post-challenge, where at its peak it encoded 17.2% (± 2.54, 95% CI) of the T_E repertoire.

For V/J allele combination usage, an overall strong positive correlation between challenged T_E and unchallenged T_N repertoires was evident during the effector expansion, indicating a highly polyclonal response with broad expansion of the naïve precursor pool. However, three specific TRBV3/J allelic combinations, TRBV3-TRBJ1-1, TRBV3-TRBJ2-4 and TRBV3-TRBJ2-7, were disproportionately increased in challenged T_E repertoires of mice infected with MT parasites at both days 6 and day 10, and for SBP infections by day 10 post-infection (**Figure 2C**). This specific V/J usage was conserved across all individual replicate mice infected with MT parasites during the acute phase of infection (**Figure 2D**) and was evident in the repeat experiment (**Supplementary Figures 5A, Biii**). During the late phase of infection, as the T_E population contracted, this conserved V/J signature was lost in both infection types (**Supplementary Figures 5Bi, Bii**).

Changes in diversity of a TCR repertoire following pathogen exposure are indicative of the extent to which clonal expansion within a repertoire has occurred (34). To assess the diversity of the acute $CD4^+$ T_E expansion response at the clonal level, repertoire diversity was calculated using Simpson's diversity index on age-matched (35) and size-matched repertoires from the repeat experiment. This allowed us to sample an equal number of UMI-labeled cDNA molecules for the precise normalization required for comparing diversity metrics (36). Although unchallenged T_N repertoires were, as expected, significantly more diverse than challenged T_E repertoires

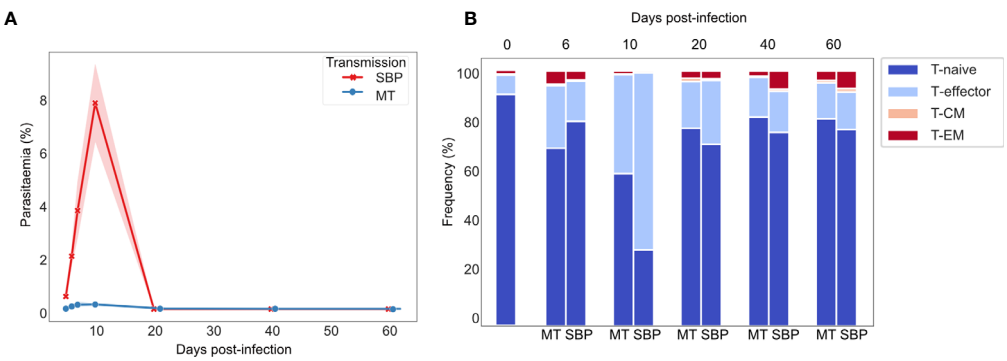


FIGURE 1 | Dynamics of *P. chabaudi* infection: T-effector expansion coincides with peak parasitemia. **(A)** Parasitaemia of C57Bl/6 mice infected with either serially blood passaged (SBP) (red) or recently MT (blue) 5 x10^5 *P. chabaudi* parasitized erythrocytes, n=5 mice per infection type per time point, shaded area depicts 95% CI. **(B)** Phenotypic profiling of $CD4^+$ T cells as determined by FACS. Representative frequencies over the time course of infection of T-naïve $CD4^+$ T cells ($CD62L^+$ $CD127^+$), T-effector $CD4^+$ T cells (T_E) ($CD62L^-$ $CD127^-$), effector memory (T_{EM}) (CD44HI $CD127^+$ $CD62L^-$) and central memory (T_{CM}) (CD44HI $CD127^+$ $CD62L^+$) $CD4^+$T cells. Unchallenged control mice are also represented.

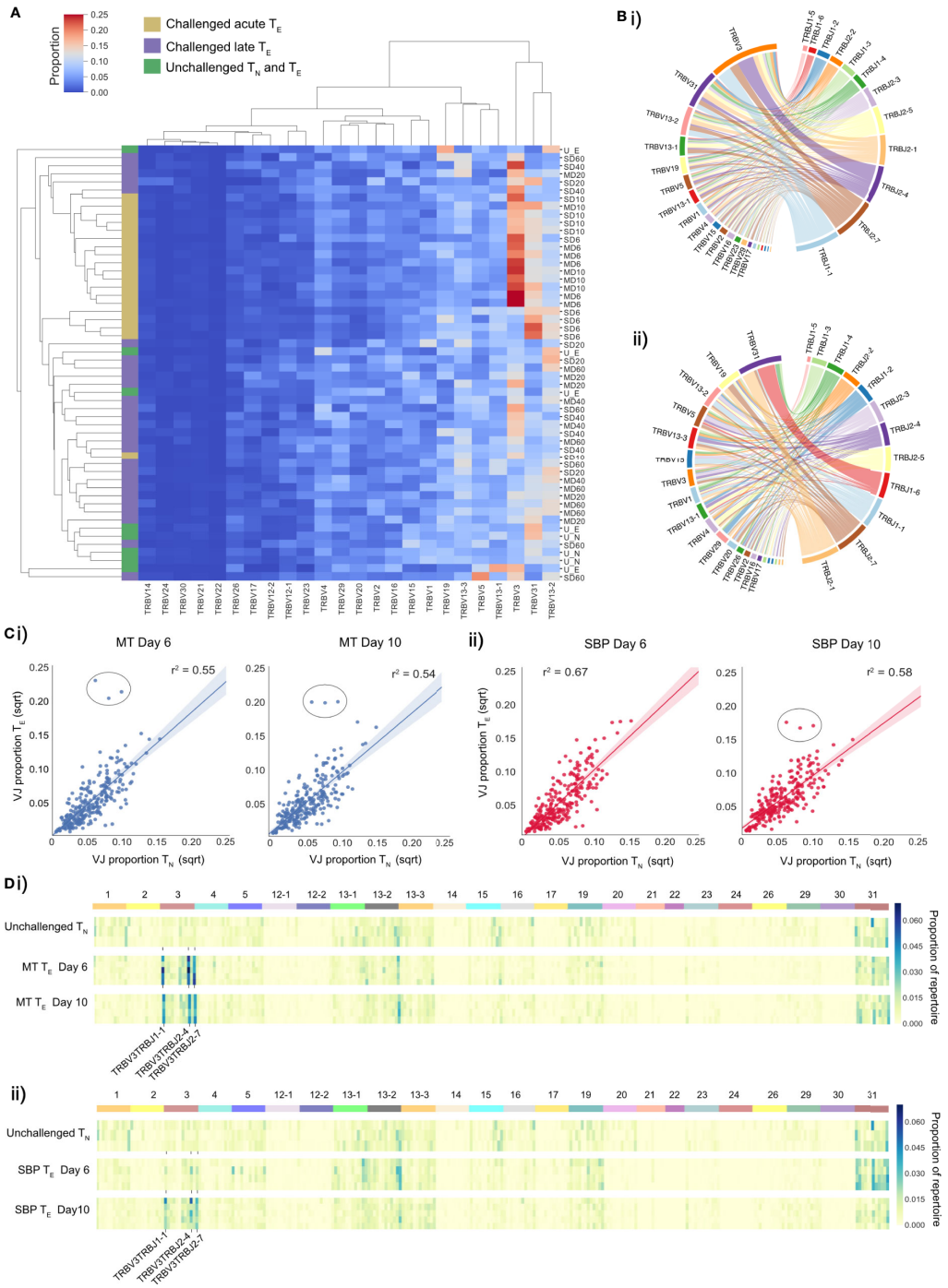


FIGURE 2 | T_E repertoires have dominant TRBV3 gene usage during the T_E expansion in acute phase of a *P. chabaudi* infection. **(A)** Clustermap displays TRBV gene proportional usage for individual splenic CD4$^+$ T-cell receptor (TCR) repertoires from challenged acute T_E repertoires (gold), challenged late phase T_E repertoires (purple) and unchallenged T_N and T_E repertoires (green). Each column is a unique TRBV gene and each row an individual repertoire. **(B)** Circos plots show V/J gene usage from a representative repertoire of i) a challenged MT T_E repertoire at day 6 post-infection and ii) an unchallenged T_N repertoire. Band width is proportional to usage frequency. **(C)** Mean proportion of each V/J allelic combination in unchallenged T_N repertoires versus challenged T_E repertoires at days 6 and 10 post-infection for mice infected with (i) MT parasites (blue) and (ii) SBP parasites (red). Each point represents a unique V/J combination. **(D)** Heatmaps depict proportional usage of each V/J allelic combination (columns) for individual replicate mice (rows) for (i) unchallenged T_N repertoires and acute T_E MT repertoires and (ii) unchallenged T_N repertoires and acute T_E SBP repertoires. Data for both days 6 and 10 post-infection are displayed. Horizontal color bar indicates TRBV gene used in the V/J combination.

(**Figure 3A**, p<0.01), the T_E repertoires after challenge were still highly polyclonal – with the most abundant clone taking up on average only 0.72% (± 0.11%) of the repertoire compared to 0.112% (± 0.01%) in T_N unchallenged repertoires (**Figure 3B**).

T-Effector Repertoires Have Greater Similarity During Acute Infection

If a pathogen-specific response is elicited at the clonal level, we would expect challenged T_E repertoires of replicate mice to contain an increased number of shared clones, and to therefore be more similar to each other than unchallenged repertoires. To examine the degree of repertoire sharing between replicate mice, the Jaccard index, a similarity or 'overlap' metric was used, matching at the CDR3 amino acid sequence level. Over the course of infection, for each infection type, similarity between replicates was significantly altered (one-way ANOVA, MT: p<0.001, SBP: p<0.001) (**Figure 3C**), with replicates being more similar to each other in the acute phase of infection than at later time-points. MT repertoires were also more similar to each other during the acute phase than SBP infections (day 6: t=2.6, p=0.016, day 10: t=7.2, p<0.001). For both infection types, during the acute phase, replicate repertoires were more similar to each other than to unchallenged T_N (day 6: MT: t=15.13, p <0.001, SBP: t=11.7, p<0.001, day 10: MT: t=13.4, p<0.001, SBP: t=9.2, p<0.001) and unchallenged T_E repertoires (day 6: MT: t=15.1, p<0.001, SBP: t=11.7, p <0.001, day 10: MT: t=13.4, p<0.001, SBP: t=9.2, p<0.001). Randomly sampling the same number of sequences (10^4) to produce size-matched repertoires did not alter this pattern of results, nor did size-matching the repeat UMI data (**Supplementary Figure 6**). Further exploration using multi-dimensional scaling (MDS) of size-matched repertoires, also indicated clustering of acute T_E repertoires for both infection types according to Jaccard similarity, with MT repertoires at day 6 and 10 most tightly co-clustered (**Figure 3D**).

The TCR repertoire is a dynamic network, so examining similarity solely at the clonal level can fail to take in to account the degree of extended clonal networks that may be present. Despite not undergoing somatic hypermutation, T-cell repertoires have been shown to contain networks generated by sequence similarity, with CDR3 sequence similarity and thus network connectivity increased in antigen-experienced repertoires (37, 38). To explore connectivity between CDR3s in

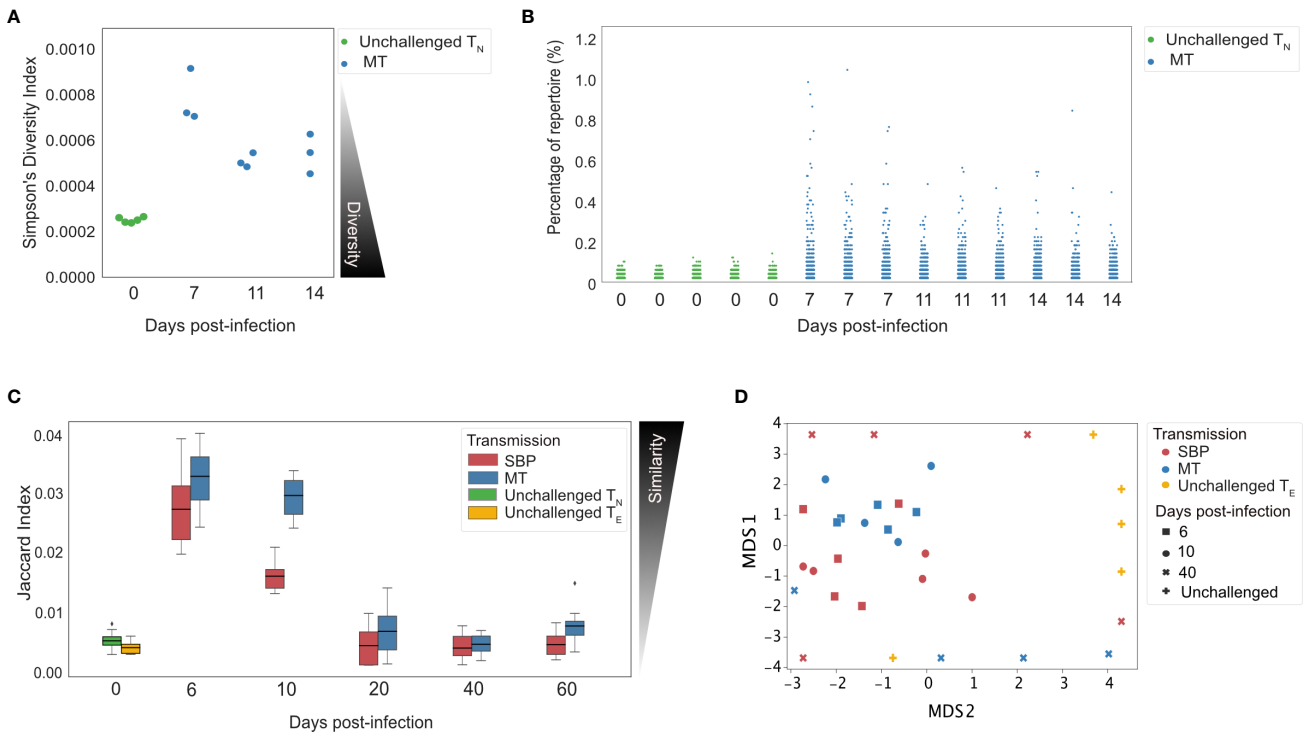


FIGURE 3 | Splenic T_E repertoires elicited by *Plasmodium chabaudi* are highly diverse. **(A)** Simpson's Diversity Index (SI) for UMI-size matched unchallenged T_N (green) and acute MT T_E repertoires (blue). SI varies from 0 to 1, and for T-cell receptor (TCR) repertoires represents the probability that identical TCRs (as determined by identical CDR3 amino acid sequence) will be drawn from the repertoire with two independent draws. SI of 0 therefore represents maximal diversity. **(B)** Each point in the strip plot represents a clone and the percentage of repertoire they occupy, for individual replicate mice for T_N unchallenged repertoires (green) and MT T_E repertoires (blue). Data for **(A, B)** was normalized by down-sampling to 5000 UMI. **(C)** Similarity between replicate repertoires is demonstrated using the Jaccard index, a normalized sharing metric which represents the degree of clonal overlap between repertoires. Jaccard index varies form 0 (no similarity) to 1 (identical repertoires). T_E replicate repertoires are more similar to each other in the acute phase of infection for both infection types. **(D)** Multi-dimensional-scaling (MDS) analysis using Jaccard similarity index of T_E repertoires for unchallenged (yellow), MT (blue) and SBP (red) repertoires. Data for **(C, D)** was normalized by down-sampling to 10^4 reads and calculated on weighted data to include clonotype frequency.

the T_E repertoires, network analysis was undertaken between the top 100 most abundant clones, using Levenshtein distance. Networks were constructed between replicate repertoires, by creating an edge between unique CDR3 sequences (nodes) if they were within a Levenshtein distance of one of each other (**Figure 4A**). Node degree (the average number of edges per node within network) for each individual indicated a higher degree of connectivity for both infection types in the acute stage of infection compared to unchallenged T_E repertoires (Mann-Whitney U, (day 6: MT: $p<0.01$, SBP: $p=0.022$, day10: MT: $p<0.01$, SBP: $p=0.045$) (**Figure 4Avi**)). MT repertoires also tended to have a higher node degree than SBP repertoires at days 6, 10 and 20 post-infection, although significance was not detected at day 10 (Mann-Whitney U, day 6: $p<0.01$, day 10: $p=0.088$, day 20: $p=0.039$). Within each individual, there was also an increased frequency of CDR3 pairwise comparisons that

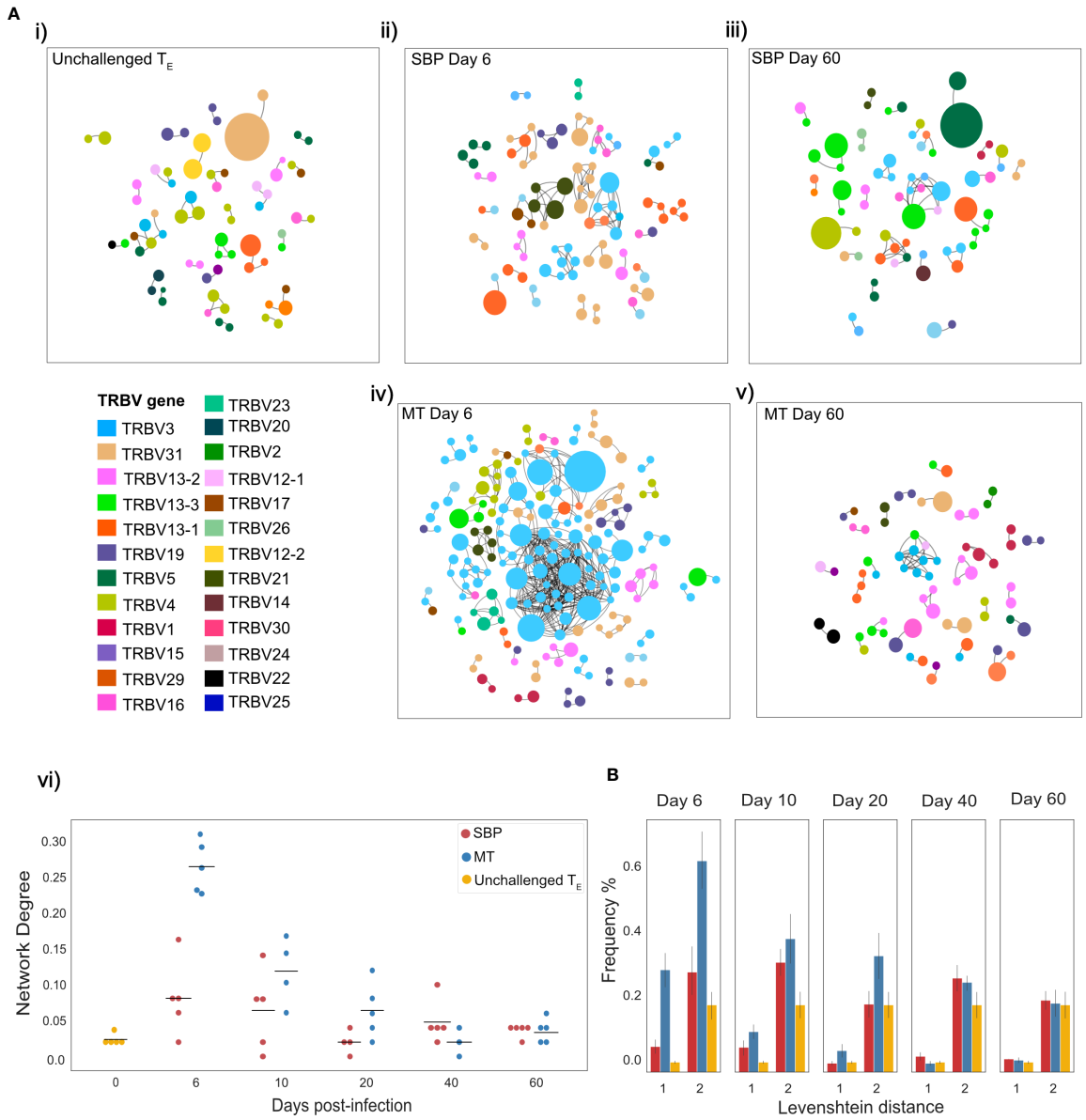


FIGURE 4 | Mosquito-transmitted (MT) acute repertoires have greater clonal network connectivity. **(A)** (i–v) Networks showing the top 100 most abundant CDR3 amino acid sequences in replicate repertoires within a Levenshtein distance of 1 of each other. For T-cell receptors (TCRs) a Levenshtein distance of 1 represents 1 amino acid mismatch (insertion, deletion or substitution) between CDR3 sequences. Each node represents a TCR clone as defined by CDR3 amino acid sequence, with node size indicating proportion of repertoire occupied by clone. Nodes are colored according to TRBV-gene usage. An edge is drawn between nodes if within a Levenshtein distance of 1, with unconnected nodes not depicted, (vi) network degree (mean number of edges per node) for each individual T_E repertoire network. **(B)** Frequency (%) of individual CDR3 sequence pairwise-comparisons, within the top 100 most abundant CDR3 sequences of each individual T_E repertoire, that are within a Levenshtein distance of 1 and 2.

differed by a distance of one or two for MT repertoires in the acute phase of infection compared to both SBP and unchallenged T_E repertoires, highlighting their greater repertoire similarity (**Figure 4B**).

The Same Public Cluster Is Dominant in the Majority of Acute T-Effector Repertoires

Given the increased connectivity in challenged repertoires, and the knowledge that TCRs recognizing the same antigen typically have a high global similarity to each other (38, 39), we clustered CDR3 sequences of individual repertoires within one amino acid mismatch of each other using a modified Swarm algorithm (26). This identified two clusters of highly similar CDR3 sequences, hereafter referred to as OTU1 and OTU2, that dominated T_E repertoires and were conserved across replicates in the acute phase of infection (**Figures 5A, B**). A near-identical cluster to OTU1 was also found to be expanded and dominant in a majority of challenged repertoires at day 7 and 11 post-infection (**Supplementary Figure 7B**), and both clusters were also observed in analyses of publicly available RNA-seq data sets for *P. chabaudi (AS)* and *P. chabaudi (CB)* (accession E-ERAD-221 and E-ERAD-289 respectively, **Supplementary Figure 7C**), showing a similar temporal pattern of expansion. We also applied the recently published Gliph2 algorithm (27) to our data. Gliph2 is designed to identify TCRs recognizing the same antigen by clustering sequences with enriched amino acid motifs in the high-contact-probability region of CDR3β (IMGT positions 107-116). It identified significant clusters that corresponded to both OTU1 and OTU2 in the acute phase of MT infections, as well several other clusters which were not as dominant nor as well-conserved in their response (**Supplementary Table 1**, **Supplementary Figure 7A**).

Given the conserved nature of OTU1 between individual mice, we hypothesized that the CDR3 sequences it contains would share similar properties with other known 'public' CDR3 sequences, defined simply as TCR clones shared by different individuals. Public TCR clones have been detected in numerous T-cell responses in multiple species, and although their functional significance remains unknown, they have been shown to be expanded in response to antigenic stimulation (40), viral infection (41, 42) and associated with regulatory self-immunity (43). In some previous studies, public sequences have been shown to have minimal alterations to germline V, D, and J gene sequences. In agreement with this, we found fewer recombination events in the CDR3 sequences in OTU1, with the mean number of randomly inserted nucleotides in the CDR3 sequences in these clusters significantly lower than that for CDR3 sequences in both challenged ($t=-61.5$, $p<0.001$) and unchallenged repertoires ($t=-58.7$, $p<0.001$) (**Figure 5Ci**). The mean number of nucleotides lying between the V and J gene segment sequences was also significantly lower (unchallenged: $t=-22.1$, $p=0.002$, challenged: $t=-83.7$, $p<0.001$) (**Figure 5Cii**). This cluster also showed a greater degree of convergent recombination – considered an important mechanism of public TCR generation (40, 44) - with a higher average number of unique CDR3 nucleotide sequences that code for the same CDR3 amino acid sequence (**Figure 5Ciii**) compared to CDR3 sequences in unchallenged ($t=4.4$, $p<0.001$) and challenged ($t=5.3$, $p<0.001$) repertoires. Consequently, CDR3 amino acid sequences in OTU1 had a higher probability of generation (Pgen) (29) than CDR3 sequences in unchallenged T_N repertoires and challenged repertoires (**Figure 5Di**). In a scenario of non-specific polyclonal expansion, a higher Pgen could indicate sharing and detection of this cluster incidentally due to higher abundance in the pre-selection pool, rather than as a result of common specificity (convergent selection) or function (45). However, the cluster was either not found or was present at a very low level in unchallenged T_N and T_E repertoires, and our use of UMI-corrected data for the repeat experiment confirmed that the CDR3s within this cluster are clonally expanded. Further, a large proportion of CDR3s with a similar high Pgen in unchallenged T_N populations, were either not present in T_E repertoires, or present but not expanded, demonstrating that the public cluster is likely to be pathogen-associated (**Figure 5Dii**).

To determine if the CDR3 sequences in OTU1 have been shown to expand in response to other antigenic stimuli in C57Bl/6 mice, we analyzed publicly available splenic $CD4^+$ TCRβ repertoire data (43) from unchallenged mice and mice immunized with either OVA or CFA and OVA. Although detectable, the proportion of OTU1 did not differ between unchallenged and immunized mice, indicating the expansion seen is not simply an innate response to inflammation. TCR repertoire data was also extracted from publicly available whole blood RNA-seq data from C57Bl/6 mice challenged with a variety of pathogens; *Toxoplasma gondii*, influenza A virus, murine cytomegalovirus, respiratory syncytial virus, *Candida albicans*, *Listeria monocytogenes*, *Burkholderia pseudomallei*, and House dust mite allergen (30). Of the 9 pathogens examined at the peak of the murine response, only *P. chabaudi* elicited an expansion in TRBV3 gene usage (**Supplementary Figure 8**). As such, OTU1 was not found to be expanded by any of the other pathogens examined, supporting the probable specificity of this response.

Splenic Memory and Re-Challenged Repertoires Are Predominantly Private Responses

Splenic $CD4^+$ T_{EM} and T_{CM} populations have been shown to expand following a *P. chabaudi* infection (8, 9, 20). To determine if a conserved splenic memory response was also generated, the TCRβ repertoires of splenic T_{EM} and T_{CM} populations from the first experiment were sequenced, for both infection types. All similarity and CDR3 overlap analyses were conducted unweighted, to avoid biases from effector clonal expansion. Little to no sharing was found between either the T_{EM} or T_{CM} replicate repertoires themselves for either infection type, nor between the T_{EM}, T_{CM} and T_E populations (**Figure 6A**). Clustering the memory repertoires within one amino acid mismatch using the same modified Swarm algorithm (26) did not improve the similarity between memory replicates, indicating these responses are mostly 'private' to each individual at this sequencing depth. As a conserved

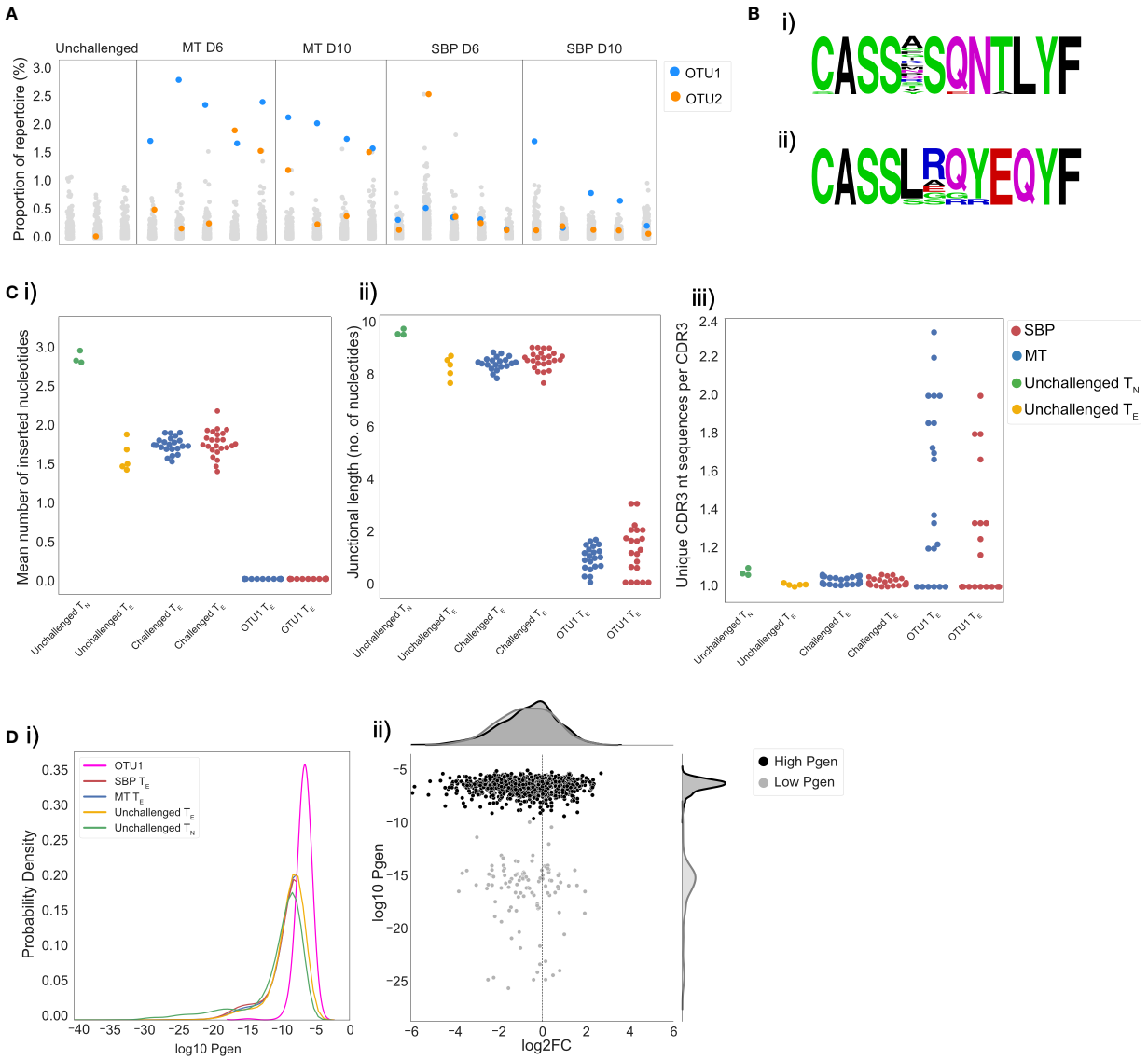


FIGURE 5 | Acute mosquito-transmitted (MT) T_E repertoires are dominated by the same cluster of clones. Repertoires were clustered using a modified Swarm algorithm, to cluster CDR3 sequences within 1 amino-acid mismatch of each other, with identical V-gene usage. **(A)** Strip plots display the proportion of repertoire occupied by each cluster in individual repertoires for unchallenged T_N and MT and SBP T_E repertoires at day 6 and 10 post-infection. Each column is an individual repertoire, with each point representing a unique cluster. Cluster OTU1 is displayed in blue and OTU2 in orange. **(B)** Representative amino acid sequence logos of clusters OTU1 and OTU2. **(C)** OTU1 CDR3 sequences have hallmarks of a public response: (i) mean number of inserted random nucleotides in CDR3 sequences, (ii) mean number of nucleotides lying between V and J gene segment sequences and (iii) convergence (mean number of unique nucleotide sequences that encode a particular CDR3 sequence), for unchallenged T_N (green), unchallenged T_E repertoires (yellow), all MT T_E repertoires (blue) and all SBP T_E repertoires (red) and for cluster OTU1 in all MT and SBP T_E repertoires. **(D)** (i) Distributions of the generation probabilities (log10) of CDR3 nucleotide sequences for unchallenged T_N (green) and T_E repertoires (yellow), SBP (red) and MT (blue) T_E repertoires and CDR3 nucleotide sequences of cluster OTU1 (pink). (ii) Log2 fold change of clones present in unchallenged T_N repertoires versus challenged T_E repertoires. Each point represents an individual clone, and Pgen is separated in to high (>−10) (black) and low (<-10) (gray). Log2FC was calculated on UMI normalized data.

response was not detected within memory populations, a homologous re-challenge infection was undertaken to determine if a similar expansion of TRBV3, as seen in primary infection, was observed upon secondary challenge. Mice were re-challenged at day 60 post-primary infection, and then sacrificed at day 67 (day 7 post re-challenge). Following re-challenge, mice did not develop a detectable parasitemia, but a splenic expansion of an activated effector population was evident, though not as marked as following primary challenge (**Figure 7A**). Although TRBV3 was again confirmed to be expanded 7 days after primary challenge, encoding 21.3% (± 3.6, 95% CI) compared to 6.0% (± 2.1, 95% CI) of unchallenged T_N, no re-expansion of TRBV3 was detected

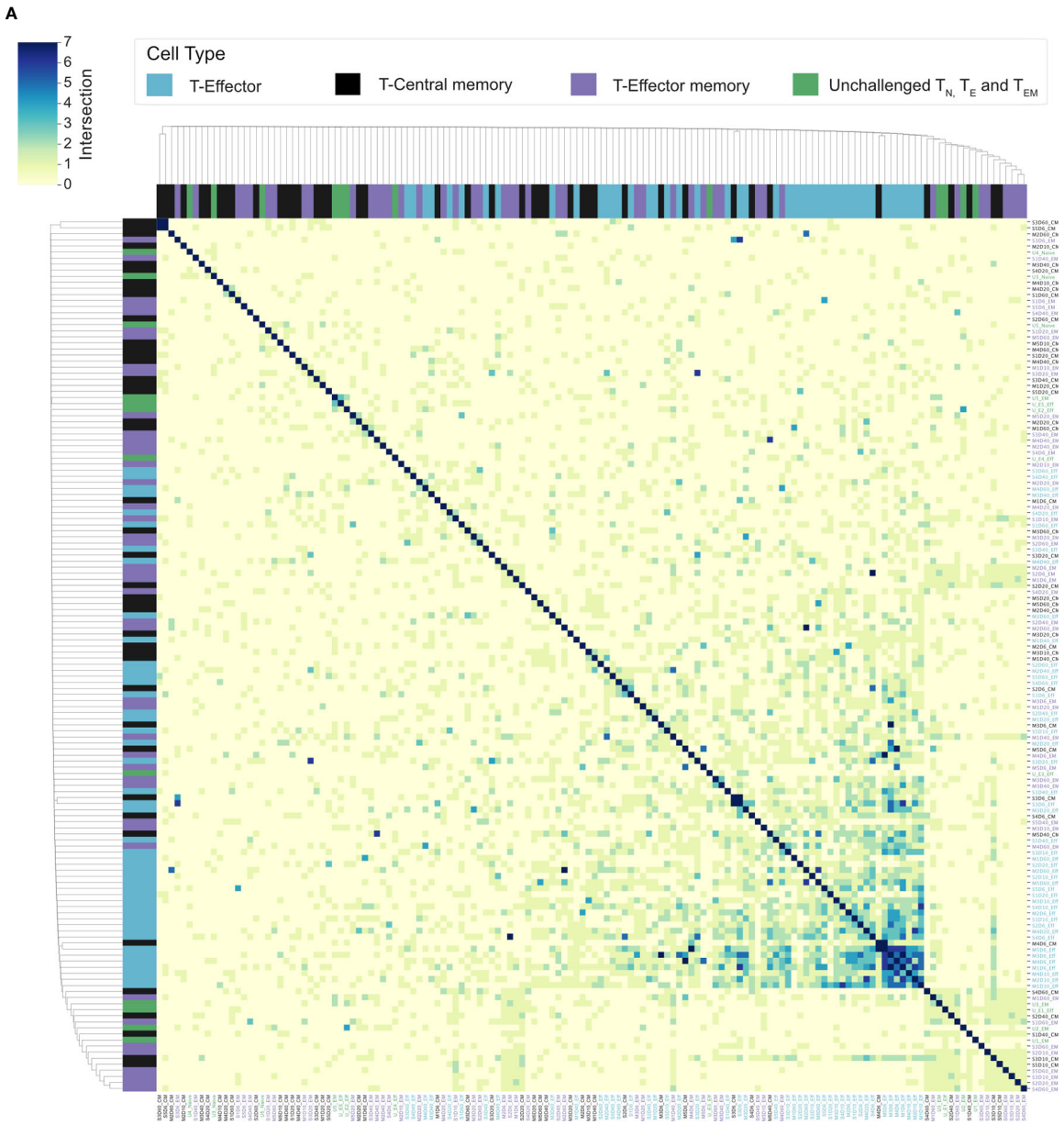


FIGURE 6 | Splenic memory responses have little to no overlap: **(A)** Cluster map shows the pairwise T-cell receptor (TCR) repertoire overlap of the top 100 CDR3 sequences of individual repertoires, of challenged T_E (blue), T_{CM} (black), and T_{EM} (purple) repertoires, and unchallenged T_N, T_E and T_{EM} repertoires (green). The size of the pairwise intersection between each repertoire (number of shared CDR3 amino acid sequences) is displayed, with greatest overlap evident between acute T_E repertoires.

following re-challenge. TRBV3 encoded 11.5% (± 0.4, 95% CI) of the re-challenged T_E repertoire, compared to 10.5% (± 1.7, 95% CI), for mice who had not undergone a secondary challenge, and re-challenged T_E TRBV gene repertoires clustered with later time-points rather than acute repertoires (**Figure 7D**). At the clonal level, re-challenged repertoires were found to be as dissimilar to each other as unchallenged repertoires (Mann Whitney U, p=0.29) (**Figure 7C**), with little to no sharing of the 100 most abundant CDR3 amino acid sequences (**Figure 7B**).

Activated TRBV3$^+$ Cells Have Diverse Transcriptional Phenotypes

To further our understanding of the phenotype of the conserved TRBV3 response observed during the acute phase of a *P. chabaudi* infection, we undertook single-cell RNA sequencing of FACS sorted CD3$^+$ CD4$^+$ splenocytes from two mice at day 7 post-infection with MT parasites. We sequenced these cells using the 10x Genomics Chromium platform, using the V(D)J enrichment protocol to obtain paired α/β TCR data for each

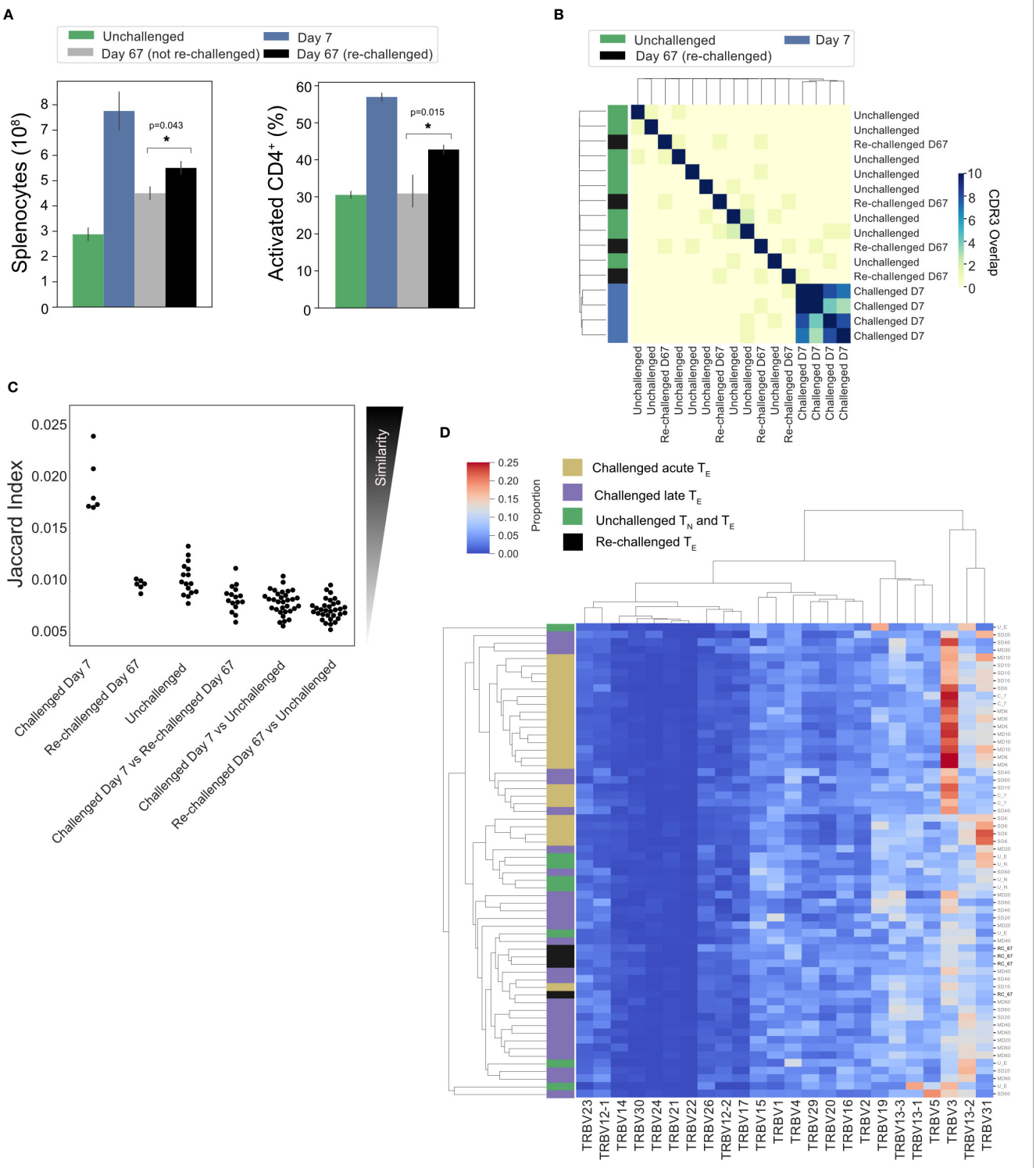


FIGURE 7 | No conservation of response following homologous re-challenge: **(A)** Splenocyte numbers and frequency of activated CD4$^+$T cells (CD44HI) of unchallenged mice (green), day 7 post-primary infection (blue), day 67 post primary infection (gray, not re-challenged) and 7 days following re-challenge (black, re-challenged day 67). **(B)** Clustermap depicts the pairwise overlap of the top 100 most abundant CDR3 amino acid sequences of repertoires from the re-challenge experiment. The size of the pairwise intersection between each repertoire (number of shared CDR3 amino acid sequences) is displayed. CDR3 overlap was calculated unweighted by clonotype frequency. **(C)** Jaccard (similarity) index of replicate repertoires (normalized by down-sampling to 5000 UMI) and **(D)** Clustermap displays the TRBV gene proportion of each repertoire for challenged acute T_E repertoires (gold), challenged late phase T_E repertoires (purple) and unchallenged T_N and T_E repertoires (green) from the first experiment, with re-challenged T_E repertoires (black) also included.

cell. After quality control steps, we obtained expression profiles for 1491 and 1485 $CD4^+$ T cells from each mouse respectively (2976 total). Following dimensionality reduction by principal component analysis (PCA), we undertook graph-based clustering (32) and visualized resulting populations using uniform manifold approximation and projection (UMAP) (**Figure 8A**). We identified seven discrete transcriptional clusters, with cells from both mice evenly distributed across all dominant clusters (**Supplementary Figure 9**). Four of these clusters, denoted as clusters 1, 2, 3, and 4, were classified overall as naïve on the basis of canonical markers (*Sell*, *Il7r*) (**Figure 8B**, **Supplementary Figure 10**). Differential expression between all seven clusters indicated that clusters 2 and 3 were distinguishable by markers indicative of early T-cell activation, including *cd69*, *ltm2a*, *Zfp36* and *ler2* for cluster 3, and type I interferon (IFN) response genes (*Gbp2*, *Ifit1*, *Ifit3*, *Isg15*, and *Socs1*) (46) for cluster 2 (**Supplementary Figure 11**), while cluster 1 expressed a more definitive naïve phenotype (*lef1*, *Ccr7*). We also identified three clusters discrete from the naïve group: cluster 6 showed a typical Treg transcriptional phenotype with expression of *Foxp3* and *Il2ra*, while clusters 5 and 7 were both identified as activated effector populations (*Sell*-, *Il7r*-). Cluster 7 expressed genes associated with a pro-inflammatory T_H1 signature, including *Tbx21, Ifng, Gzmb, Cxcr6, Ccl5, NKg7*, and *Bhlhe40* (47). In contrast, cluster 5 was *Ifng*- and appeared to be a more diverse helper population, which predominantly included cells expressing genes associated with a T_{FH} phenotype (*Cxcr5*, *Icos, Bcl6, Il21, Pdcd1)* as well as those with a T_H2 phenotype (*Gata3, Ccr4*) (**Figure 8C**). Utilizing the TCR data, we confirmed that the effector response is highly polyclonal: of the 581 effector cells present in clusters 5 and 7, only 3.6% (21/581) are clonally expanded (>1 cell with identical paired TCRα and TCRβ aa sequence) (**Figure 8D**), and no clone had more than three copies present. We again confirmed that TRBV3 is the most dominant TRBV gene encoding the TCRβ chain of activated effector populations, and show that out of all the clusters, cluster 7 had the highest proportion of TRBV3 (**Figure 8E**). Cells that were $TRBV3^+$ did not however form a distinct cluster and instead displayed diverse phenotypes across all clusters. No TRAV gene was over-represented in a particular cluster (**Supplementary Figure 9B**). Of the clonotypes previously identified in the TCRβ bulk data in OTU1 and OTU2, two were present in the single-cell data set, both within the predominantly T_{FH} cluster 5. TCRβ nucleotide sequences displayed diverse probabilities of generation across all clusters (**Supplementary Figure 12**).

DISCUSSION

$CD4^+$ T-cells play a critical role in the immune response against the pathological blood-stage of malaria [reviewed in (6, 7)]. However there is a lack of deep mechanistic understanding regarding the development of T-cell mediated immunity against *Plasmodium* (7). To our knowledge, this is the first study to use bulk TCR deep-sequencing to examine the composition of $CD4^+$ T-cell repertoires induced by a *Plasmodium* infection. We found that in both the MT and SBP infection models, which show differential control of parasite growth and degrees of immunopathology (8, 16), T_E repertoires elicited upon infection are highly diverse and polyclonal. Despite the 10-fold increase in splenic T_E cell numbers, the V/J gene usage frequencies of acute challenged T_E repertoires are strongly positively correlated with that of unchallenged T_N repertoires, indicating a highly polyclonal broad expansion of the naïve repertoire, consistent with the massive cellular expansion seen. This polyclonality is confirmed in our single-cell RNA-seq data. The degree of correlation between VJ usage frequencies likely indicates a degree of non-specific expansion, although a highly heterogeneous response to the vast number of potential antigens expressed by the parasite, argued as a potential cause for the preponderance seen in *P. falciparum* IgG responses (48), cannot be ruled out. However, within this highly polyclonal effector proliferation, a strong and oligo-clonal conserved response is observed in the bulk TCR data following a first infection. Here, repertoires are skewed to TRBV3 gene usage, have a higher degree of clonal sharing, and show increased amino acid sequence similarity of the CDR3 region between dominant clones. Thus, we demonstrate that despite the antigenic complexity of *P. chabaudi*, T_E repertoires bear the hallmarks of a specific response (38, 39), mirroring that observed with less antigenically diverse organisms (49–51). This conserved response is evident in both infection models, although it is delayed and less marked in mice infected with SBP parasites, and does not become evident until parasitemia has peaked. Adaptive immune responses of lower magnitude have previously been documented in SBP infections compared to MT (8) and while the precise reasons for this difference remain speculative and represent a complex scenario, the fact that the conserved response does eventually become evident suggests that the drivers of this response are likely to be common to both MT and SBP parasites despite apparent distinct kinetics in the TRBV3 response. Sequencing of total parasite RNA for both MT and SBP parasites undertaken by Spence et al. (8), indicated the parasites' *cir* gene family - believed to encode a large set of variable antigens displayed on parasitized erythrocytes -was differentially expressed between the two infection models. MT parasites upregulated expression of many genes within this family equally, re-setting broad expansion of the antigenic repertoire, while SBP selected for a limited but dominant *cir* expression. Our results are therefore inconsistent with proteins encoded by *cir* genes driving the detected conserved response, which would instead have been expected to result in a more dominant response in SBP infections. We hypothesize that the differences observed in the timing and magnitude of the TRBV3-restricted shared response seen between the two infection models are a result of the systemic inflammation induced by SBP parasites (8) disrupting or delaying the formation of an appropriate T-cell response. Future work to determine MHC-presentation pathway and ligands or peptides involved is ultimately required to determine what specific parasite epitope elicits the conserved response.

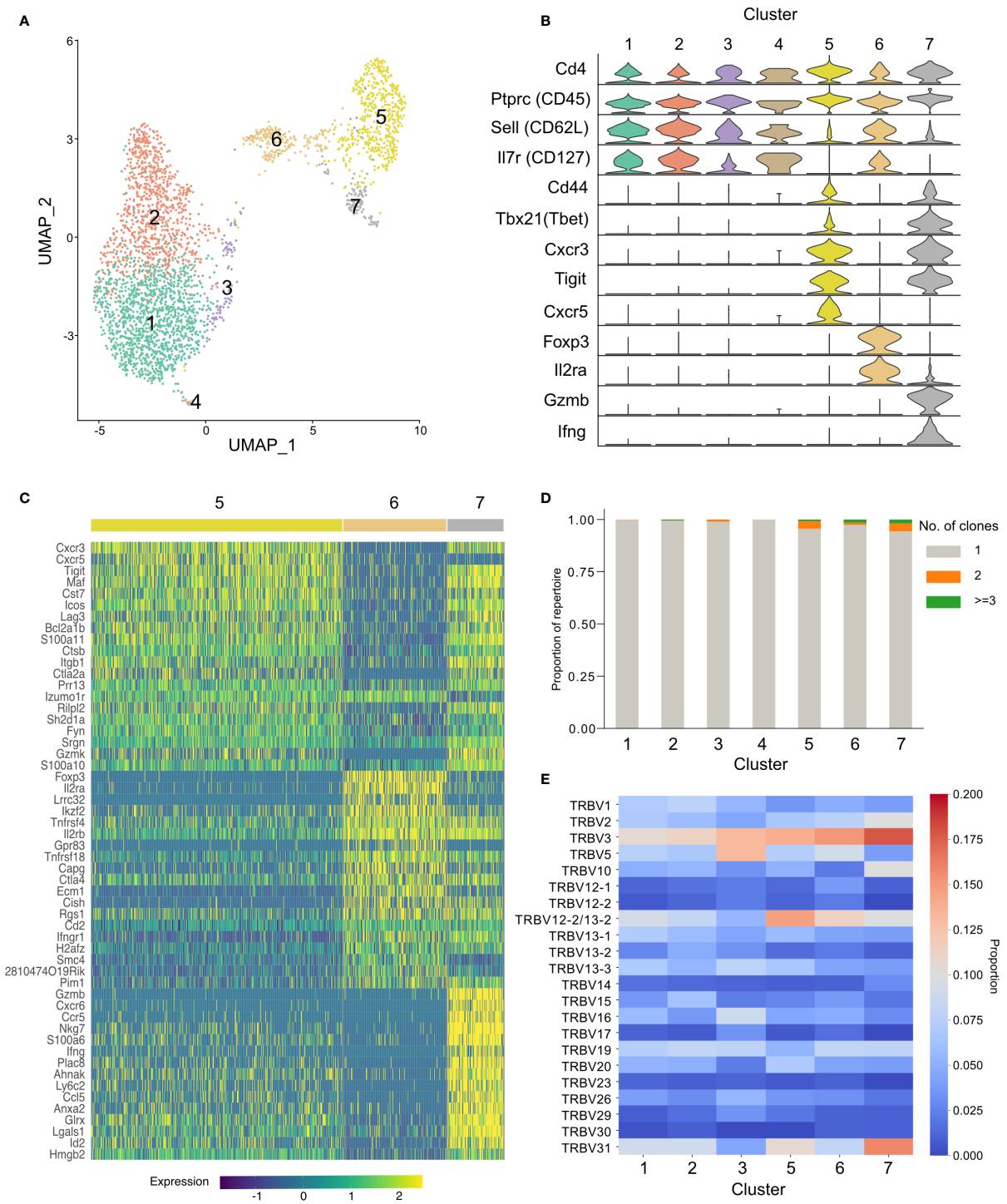


FIGURE 8 | Activated TRBV3^{+} cells have diverse transcriptional phenotypes. Cellular transcriptional phenotypes and paired T-cell receptor (TCR) α and TCRβ sequences were captured using single-cell transcriptomics. **(A)** Clustering of CD4^{+} splenocytes from two challenged mice, at day 7 post-infection, according to transcriptional profile similarity. Plot shows individual cell transcriptomes represented in two dimensions by uniform manifold approximation and projection (UMAP). Each point depicts a cell, colored by Seurat cluster assignment as determined from gene expression profiles. **(B)** Distributions of normalized expression level of major canonical T-cell markers for each cell cluster, displayed as violin plots. **(C)** Heatmap showing the expression levels of the top 20 differentially expressed genes that define clusters 5, 6, and 7. **(D)** Stacked bar plot showing the proportion of each cluster that is clonally expanded (>1 cell with identical paired TCRα and TCRβ aa sequences). Repertoires of all clusters were highly polyclonal, with a minor increase in frequency of expanded clones in activated clusters. **(E)** Heatmap displays proportional TRBV gene usage per cluster TCR repertoire.

The single cell RNA-seq data confirms the dominance of TRBV3 in activated effector populations and reveals that TRBV3$^+$ cells in the acute phase of a first infection have diverse phenotypes. We therefore show that the conserved response seen does not represent a discrete innate cell population but is instead part of an adaptive immune process against the parasite. The transcriptional phenotypes of activated effector cells are in agreement with those previously observed in the acute phase of a *P. chabaudi* infection (47) and indicate a large cluster of predominantly follicular helper cells as well as a distinct *Ifng*$^+$ T_H1 population. Both types of response have been shown to arise simultaneously during the acute phase (47) and to be essential in controlling blood-stage *P. chabaudi*; T_H1 responses are required for initial control of acute parasitemia (52, 53) and T_{FH} are crucial for generating antibody-mediated immunity and controlling chronic infection (54, 55). The presence of OTU1 and OTU2 in a predominantly T_{FH} cluster, suggests that they may play a role in guiding the developing humoral response against the parasite. A single naïve CD4$^+$ T-cell has been demonstrated to be able to give rise to clones with different cell fates (47) so, as only one of each clonotype was captured with the single cell sequencing depth used, we cannot firmly conclude that the transcriptional phenotypes identified here would broadly reflect the conserved TCRs identified in the bulk data. Nevertheless, the TCR of a cell has been shown to impart a strong preference for either a T_H1 or a T_{FH} fate, with longer dwell time between peptide-MHC : TCR biasing toward T_{FH} and GC-T_{FH} responses (56). Therefore, if driven by the same epitope, we would expect the conserved response to have a similar phenotype (57).

Public TCR sequences are shared between multiple individuals either due to biases in V(D)J recombination, and/or convergent selection by a common antigen (45, 49). The CDR3 amino acid sequences in the most dominant and conserved cluster detected (OTU1) had similar features to other previously observed public TCR responses. These include a reduced number of nucleotide addition and deletion events during VDJ recombination and a greater degree of recombinant convergence (different nucleotide sequences encoding the same amino acid CDR3 sequence). Such recombinational biases during T-cell development mean these TCRs will have a higher probability of generation (PGen) during VDJ recombination, and are therefore more likely to be present in a naïve pool. However, the degree of clonal expansion observed in OTU1 sequences was greater than many other sequences of equal or higher PGen, indicating that these clones were truly expanded and not simply found to be of high frequency as a result of recombinational biases. We also show that PGen does not determine T-cell fate, in agreement with Sethna et al. (29) who demonstrated that the ability of a TCR to respond to a particular epitope, was not strongly correlated with its generation probability. It has been hypothesized that public TCR responses may provide rapid cross-reactive immunity (58) to cope with diverse antigenic challenge, allowing time for more specific private responses to develop (40, 59). Thus, during a *P. chabaudi infection*, a public response that is mobilized rapidly due to high Pgen and/or a higher chance of positive selection if cross-reactive, may act as a first line of defense against the parasite before more specific responses become effective. In agreement with this, OTU1 appears to be temporally associated with enhanced control of parasitaemia. It arises earlier and is more dominant in MT infections compared with SBP, where rapid parasite growth is observed alongside a delayed and less marked conserved response. Mice infected with MT parasites also show reduced disease severity (8). Despite these positive associations, whether this conserved response is truly beneficial to the host, remains unknown. There are reports of public TCR responses being implicated in self-related immunity (43), and in *P. berghei*, the presence of conserved pathogenic CD8$^+$ T-cells has been used to predict cerebral malaria (11).

No expanded conserved response was evident in memory populations at the timepoints tested. It is known that precursors of memory T-cells derive from an earlier T_E population (CD127$^-$ CD62L^{hi}) that precedes the terminally differentiated CD62L$^-$ CD127$^-$ T_E population captured by our T_E gating strategy (60). While that precursor population successfully expands to dominate the T_E repertoire, since memory T populations preferentially favor TCRs with high avidity, it is possible that the public TCR clones identified may lack sufficient avidity required to enter the memory pool (61). A low level of T_M repertoire overlap indicates that splenic memory responses are diverse and mostly private to each individual. Similarly, in the re-challenged T_E compartment, we did not detect an increase in TRBV3 compared to later time points that had not undergone re-infection, and at the clonal level responses had as little overlap between replicates as unchallenged repertoires. This does not imply lack of a memory/primed response; mice are protected from re-infection and have previously been shown to have primed secondary responses (9). Heterologous responses could be due to T-cells more specific for the parasite than the TRBV3+ public response surviving the contraction phase/entering memory to be able to respond. Diverse or 'private' responses have previously been seen with other complex pathogens such as *mycobacteria* (62, 63) and would not be unexpected given the antigenic complexity of *Plasmodium*. However, primed private secondary responses cannot be confirmed without experiments that sample the same individuals longitudinally over the course of a primary and secondary infection, and this hypothesis therefore remains untested and an avenue for future investigation.

These results also do not exclude the possibility that a low-level re-expansion of TRBV3 does occur upon re-challenge, given only a small increase in the number of activated cells was observed. This may not have constituted a large enough proportion of the repertoire, given the number of animals used in re-infection, to detect subtle differences in re-challenged T_E V-gene usage. This is supported by work done by Opata et al. (9), who demonstrated that T-cells specific to *P. chabaudi* MSP-1 that survive the contraction phase, do not re-expand upon secondary challenge due to increased apoptosis. Such clones that are responding but do not accumulate/re-expand would be difficult to discern from baseline controls using bulk repertoire sequencing techniques. Future re-infection studies may benefit

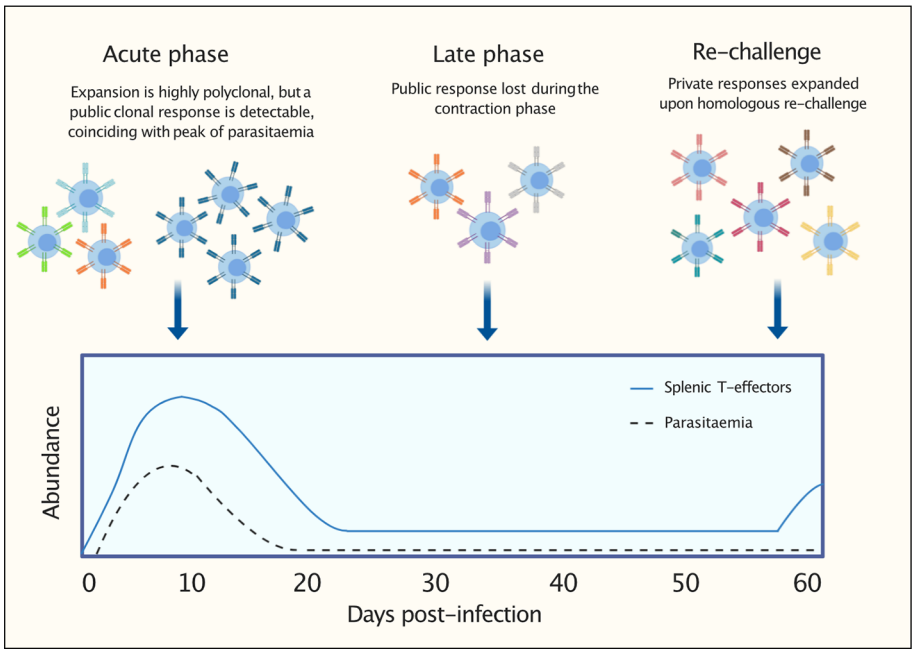


FIGURE 9 | Graphical representation to summarize findings: following a first *P. chabaudi* infection, within a highly polyclonal T-effector expansion, a public T-cell receptor (TCR) β clonal response encoded by TRBV3 is detectable. This conserved response contracts during the late phase of infection and is not re-expanded upon homologous re-challenge. This response is consistently found to be a hallmark of a first *P. chabaudi* infection.

from including single-cell transcriptomics to help differentiate such responding cells based on phenotype.

To have found a near-identical expanded T_E cluster in independent repeat experiments, and in publicly available *P. chabaudi* (AS and CB) RNA-seq data sets, demonstrates the public TCR signature is consistently expanded in multiple individuals in response to a first *Plasmodium chabaudi* infection. TRBV3 was not expanded at peak infection for any of the other eight pathogens examined and searches of annotated TCR sequence databases have not revealed any other known specificities for any of the CDR3 amino acid sequences in OTU1 or OTU2 (50, 64). TCR repertoire and RNA-seq data sets following infection with other *Plasmodium* species are currently unavailable, but as these are generated, the specificity of this response to the parasite and whether a similar response is elicited in a first infection by other *Plasmodium* species including *P. falciparum* will become evident.

Future studies are required to determine whether the conserved response plays a critical role in protection during a first infection. MT infections develop chronic recrudescing infection, so the response does not fully clear infection, although it is temporally associated with control of parasite growth and reduced disease severity (8). Whether entirely favorable to the host or not, we hypothesize the response allows time for mechanisms that govern the formation of more specific private responses and subsequent immunity to the parasite to develop. Mice are known to develop highly effective strain-specific anti-parasite immunity after a single malaria episode, while this takes years to develop in people. Therefore, if some degree of first line protection were shown, such a response could be a novel target to promote in malaria naïve individuals through vaccination, providing initial cover while more specific but slower private responses develop, conveying higher levels of protection. Even if only partially protective, the response is conserved between individuals and receptor sequences have a high probability of generation. Targeting these is predicted therefore to increase vaccine success rate (29, 65).

In summary, we have demonstrated that a conserved TCRβ signature encoded by TRBV3 is consistently expanded in response to a first *Plasmodium chabaudi* infection (**Figure 9**). In contrast, memory formation and re-challenged repertoires appear to be more private responses to the individual. Understanding the antigenic drivers and contribution to protection (or pathogenesis) of this conserved signature, that is consistently a hallmark of a first infection, is ultimately required to determine if it should be promoted or mitigated for malaria therapeutic purposes.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ebi.ac.uk/ena, PRJEB40867 https://www.ebi.ac.uk/arrayexpress/, E-MTAB-9691.

ETHICS STATEMENTS

The animal study was reviewed and approved by the Ethical Review Body of the University of Edinburgh.

AUTHOR CONTRIBUTIONS

NS, GC, and CS designed the project, analyzed the data, and wrote the manuscript. Experimental work was carried out by NS, GC, WN, JM, PS, and JT. All authors reviewed the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was made possible by a studentship from the Wellcome Trust to NS (204511/Z/16/A) and Wellcome Trust ISSF funding to GC.

ACKNOWLEDGMENTS

This work has made use of the resources provided by the Edinburgh Computer and Data Facility (ECDF) (http://www.ecdf.ed.ac.uk/). The authors thank Dr. Martin Waterfall of the IIIR Flow Cytometry Facility for help with cell sorting. Bulk TCR sequencing was carried out by Edinburgh Genomics at the University of Edinburgh. Edinburgh Genomics is partly supported through core grants from NERC (R8/H10/56), MRC (MR/K001744/1), and BBSRC (BB/J004243/1). Single-cell sequencing was carried out by the IGMM flow cytometry service and the Edinburgh Clinical Research Facility, at the University of Edinburgh. **Figure 9** was created using BioRender.com.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2020.587756/full#supplementary-material

REFERENCES

1. Ryg-Cornejo V, Ly A, Hansen DS. Immunological processes underlying the slow acquisition of humoral immunity to malaria. *Parasitology* (2016) 143:1–9. doi: 10.1017/S0031182015001705
2. Gupta S, Snow RW, Donnelly CA, Marsh K, Newbold C. Immunity to non-cerebral severe malaria is acquired after one or two infections. *Nat Med* (1999) 5:340–3. doi: 10.1038/6560
3. Gonçalves BP, Huang C-Y, Morrison R, Holte S, Kabyemela E, Prevots DR, et al. Parasite Burden and Severity of Malaria in Tanzanian Children. *N Engl J Med* (2014) 370:1799–808. doi: 10.1056/NEJMoa1303944
4. Ly A, Hansen DS. Development of B cell memory in malaria. *Front Immunol* (2019) 10:559. doi: 10.3389/fimmu.2019.00559
5. Doolan DL, Dobaño C, Baird JK. Acquired immunity to Malaria. *Clin Microbiol Rev* (2009) 22:13–36. doi: 10.1128/CMR.00025-08
6. Nlinwe ON, Kusi KA, Adu B, Sedegah M. T-cell responses against Malaria: Effect of parasite antigen diversity and relevance for vaccine development. *Vaccine* (2018) 36:2237–42. doi: 10.1016/j.vaccine.2018.03.023
7. Kurup SP, Butler NS, Harty JT. T cell-mediated immunity to malaria. *Nat Rev Immunol* (2019) 19:457–71. doi: 10.1038/s41577-019-0158-z
8. Spence PJ, Jarra W, Lévy P, Reid AJ, Chappell L, Brugat T, et al. Vector transmission regulates immune control of Plasmodium virulence. *Nature* (2013) 498:228–31. doi: 10.1038/nature12231
9. Opata MM, Stephens R. Chronic Plasmodium chabaudi infection generates CD4 memory T cells with increased T cell receptor sensitivity but poor secondary expansion and increased apoptosis. *Infect Immun* (2017) 85:1–16. doi: 10.1128/IAI.00774-16
10. del Portillo HA, Ferrer M, Brugat T, Martin-Jaular L, Langhorne J, Lacerda MVG. The role of the spleen in malaria. *Cell Microbiol* (2012) 14:343–55. doi: 10.1111/j.1462-5822.2011.01741.x
11. Mariotti-Ferrandiz E, Pham HP, Dulauroy S, Gorgette O, Klatzmann D, Cazenave PA, et al. A TCRB Repertoire Signature Can Predict Experimental Cerebral Malaria. *PloS One* (2016) 11:1–17. doi: 10.1371/journal.pone.0147871
12. Ndungu FM, Sanni L, Urban B, Stephens R, Newbold CI, Marsh K, et al. CD4 T Cells from Malaria-Nonexposed Individuals Respond to the CD36-Binding Domain of Plasmodium falciparum Erythrocyte Membrane Protein-1 via an MHC Class II-TCR-Independent Pathway. *J Immunol* (2006) 176:5504–12. doi: 10.4049/jimmunol.176.9.5504
13. Scholzen A, Mittag D, Rogerson SJ, Cooke BM, Plebanski M. Plasmodium falciparum – Mediated Induction of Human CD25 hi Foxp3 hi CD4 T Cells Is Independent of Direct TCR Stimulation and Requires IL-2 , IL-10 and TGF b. *PloS Pathog* (2009) 5:26–32. doi: 10.1371/journal.ppat.1000543
14. Wipasa J, Okell L, Sakkhachornphop S, Suphavilai C, Chawansuntati K, Liewsaree W, et al. Short-lived IFN-γ effector responses, but long-lived IL-10 memory responses, to malaria in an area of low malaria endemicity. *PloS Pathog* (2011) 7:e1001281. doi: 10.1371/journal.ppat.1001281
15. Bradley P, Thomas PG. Using T Cell Receptor Repertoires to Understand the Principles of Adaptive Immune Recognition. *Annu Rev Immunol* (2019) 37:547–70. doi: 10.1146/annurev-immunol-042718-041757
16. Spence PJ, Brugat T, Langhorne J. Mosquitoes Reset Malaria Parasites. *PloS Pathog* (2015) 11:10–5. doi: 10.1371/journal.ppat.1004987
17. Spence PJ, Jarra W, Levy P, Nahrendorf W, Langhorne J. Mosquito transmission of the rodent malaria parasite Plasmodium chabaudi. *Malar J* (2012) 11:407. doi: 10.1186/1475-2875-11-407
18. Marr EJ, Milne RM, Anar B, Girling G, Schwach F, Mooney JP, et al. An enhanced toolkit for the generation of knockout and marker-free fluorescent Plasmodium chabaudi. *Wellcome Open Res* (2020) 5:1–20. doi: 10.12688/wellcomeopenres.15587.1
19. Lewis M, Pfeil J, Mueller AK. Continuous oral chloroquine as a novel route for Plasmodium prophylaxis and cure in experimental murine models. *BMC Res Notes* (2011) 4:262. doi: 10.1186/1756-0500-4-262
20. Stephens R, Langhorne J. Effector memory Th1 CD4 T cells are maintained in a mouse model of chronic malaria. *PloS Pathog* (2010) 6:e1001208. doi: 10.1371/journal.ppat.1001208
21. Kivioja T, Vähärautio A, Karlsson K, Bonke M, Enge M, Linnarsson S, et al. Counting absolute numbers of molecules using unique molecular identifiers. *Nat Methods* (2012) 9:72–4. doi: 10.1038/nmeth.1778
22. Shugay M, Britanova OV, Merzlyak EM, Turchaninova MA, Mamedov IZ, Tuganbaev TR, et al. Towards error-free profiling of immune repertoires. *Nat Methods* (2014) 11:653–6. doi: 10.1038/nmeth.2960
23. Bolotin DA, Poslavsky S, Mitrophanov I, Shugay M, Mamedov IZ, Putintseva EV, et al. MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods* (2015) 12:380–1. doi: 10.1038/nmeth.3364
24. Shugay M, Bagaev DV, Turchaninova MA, Bolotin DA, Britanova OV, Putintseva EV, et al. VDJtools: Unifying Post-analysis of T Cell Receptor Repertoires. *PloS Comput Biol* (2015) 11:1–16. doi: 10.1371/journal.pcbi.1004503
25. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* (2020) 17:261–72. doi: 10.1038/s41592-019-0686-2
26. Mahe F, Rognes T, Quince C, De VC, Dunthorn M. Swarm: robust and fast clustering method for amplicon-based studies. *PeerJ* (2014) 2:1–13. doi: 10.7717/peerj.593

27. Huang H, Wang C, Rubelt F, Scriba TJ, Davis MM. Analyzing the Mycobacterium tuberculosis immune response by T-cell receptor clustering with GLIPH2 and genome-wide antigen screening. *Nat Biotechnol* (2020) 38:1194–202. doi: 10.1038/s41587-020-0505-4
28. Bastian M, Heymann S, Jacomy M. *Gephi: An Open Source Software for Exploring and Manipulating Networks*. International AAAI Conference on Weblogs and Social Media (2009).
29. Sethna Z, Elhanati Y, Callan CG, Walczak AM, Mora T. OLGA: fast computation of generation probabilities of B- and T-cell receptor amino acid sequences and motifs. *Bioinformatics* (2018) 35:2974–81. doi: 10.1093/bioinformatics/btz035
30. Singhania A, Graham CM, Gabryšová L, Moreira-Teixeira L, Stavropoulos E, Pitt JM, et al. Transcriptional profiling unveils type I and II interferon networks in blood and tissues across diseases. *Nat Commun* (2019) 10:1–21. doi: 10.1038/s41467-019-10601-6
31. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. *Nat Commun* (2017) 8:14049. doi: 10.1038/ncomms14049
32. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* (2018) 36:411–20. doi: 10.1038/nbt.4096
33. Mamedov MR, Scholzen A, Nair RV, Cumnock K, Kenkel JA, Oliveira JHM, et al. A Macrophage Colony-Stimulating-Factor-Producing γδ T Cell Subset Prevents Malarial Parasitemic Recurrence. *Immunity* (2018) 48:350–63.e7. doi: 10.1016/j.immuni.2018.01.009
34. Venturi V, Kedzierska K, Turner SJ, Doherty PC, Davenport MP. Methods for comparing the diversity of samples of the T cell receptor repertoire. *J Immunol Methods* (2007) 321:182–95. doi: 10.1016/j.jim.2007.01.019
35. Britanova OV, Putintseva EV, Shugay M, Merzlyak EM, Turchaninova MA, Staroverov DB, et al. Age-Related Decrease in TCR Repertoire Diversity Measured with Deep and Normalized Sequence Profiling. *J Immunol* (2014) 192:2689–98. doi: 10.4049/jimmunol.1302064
36. Izraelson M, Nakonechnaya TO, Moltedo B, Egorov ES, Kasatskaya SA, Putintseva EV, et al. Comparative analysis of murine T-cell receptor repertoires. *Immunology* (2018) 153:133–44. doi: 10.1111/imm.12857
37. Madi A, Poran A, Shifrut E, Reich-zeliger S, Greenstein E, Zaretsky I, et al. T cell receptor repertoires of mice and humans are clustered in similarity networks around conserved public CDR3 sequences. *Elife* (2017) 3:1–17. doi: 10.7554/eLife.22057
38. 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) 547:94–8. doi: 10.1038/nature22976
39. 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) 547:89–93. doi: 10.1038/nature22383
40. Covacu R, Philip H, Jaronen M, Douek DC, Efroni S, Quintana FJ, et al. System-wide Analysis of the T Cell Response. *Cell Rep* (2016) 14:2733–44. doi: 10.1016/j.celrep.2016.02.056
41. Miles JJ, Bulek AM, Cole DK, Gostick E, Schauenburg AJA, Venturi V, et al. Genetic and Structural Basis for Selection of a Ubiquitous T Cell Receptor Deployed in Epstein-Barr Virus Infection. *PloS Pathog* (2010) 6:1–15. doi: 10.1371/journal.ppat.1001198
42. Benati D, Galperin M, Lambotte O, Gras S, Lim A, Mukhopadhyay M, et al. Public T cell receptors confer high-avidity CD4 responses to HIV controllers. *J Clin Invest* (2016) 126:2093–108. doi: 10.1172/JCI83792DS1
43. Madi A, Shifrut E, Reich-zeliger S, Gal H, Best K, Ndifon W, et al. T-cell receptor repertoires share a restricted set of public and abundant CDR3 sequences that are associated with self-related immunity. *Genome Res* (2014) 24:1603–12. doi: 10.1101/gr.170753.113.24
44. 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. *PNAS* (2006) 103:18691–6. doi: 10.1073/pnas.0608907103
45. Pogorelyy MV, Minervina AA, Chudakov DM, Mamedov IZ, Lebedev YB, Mora T, et al. Method for identification of condition-associated public antigen receptor sequences. *Elife* (2018) 7:e33050. doi: 10.7554/eLife.33050
46. Elyahu Y, Hekselman I, Eizenberg-Magar I, Berner O, Strominger I, Schiller M, et al. Aging promotes reorganization of the CD4 T cell landscape toward extreme regulatory and effector phenotypes. *Sci Adv* (2019) 5:eaaw8330. doi: 10.1126/sciadv.aaw8330
47. Lönnberg T, Svensson V, James KR, Fernandez-Ruiz D, Sebina I, Montandon R, et al. Single-cell RNA-seq and computational analysis using temporal mixture modeling resolves TH1/TFH fate bifurcation in malaria. *Sci Immunol* (2017) 2:1–12. doi: 10.1126/sciimmunol.aal2192
48. Weiss GE, Traore B, Kayentao K, Ongoiba A, Doumbo S, Kone Y, et al. The Plasmodium falciparum-Specific Human Memory B Cell Compartment Expands Gradually with Repeated Malaria Infections. *PloS Pathog* (2010) 6:e1000912. doi: 10.1371/journal.ppat.1000912
49. 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) 49:659–65. doi: 10.1038/ng.3822
50. Shugay M, Bagaev DV, Zvyagin IV, Vroomans RM, Crawford JC, Dolton G, et al. VDJdb: a curated database of T-cell receptor sequences with known antigen specificity. *Nucleic Acids Res* (2018) 46:419–27. doi: 10.1093/nar/gkx760
51. Wolf K, Hether T, Gilchuk P, Ahn T, Joyce S, Dipaolo RJ, et al. Identifying and Tracking Low-Frequency Virus-Specific TCR Clonotypes Using High-Throughput Sequencing. *Cell Rep* (2018) 25:2369–78.e4. doi: 10.1016/j.celrep.2018.11.009
52. Su Z, Stevenson MM. Central role of endogenous gamma interferon in protective immunity against blood-stage Plasmodium chabaudi AS infection. *Infect Immun* (2000) 68:4399–406. doi: 10.1128/IAI.68.8.4399-4406.2000
53. Soon MSF, Haque A. Recent Insights into CD4 + Th Cell Differentiation in Malaria. *J Immunol* (2018) 200:1965–75. doi: 10.4049/jimmunol.1701316
54. Pérez-Mazliah D, Ng DHL, Freitas do Rosário AP, McLaughlin S, Mastelic-Gavillet B, Sodenkamp J, et al. Disruption of IL-21 Signaling Affects T Cell-B Cell Interactions and Abrogates Protective Humoral Immunity to Malaria. *PloS Pathog* (2015) 11:1–24. doi: 10.1371/journal.ppat.1004715
55. Pérez-Mazliah D, Nguyen MP, Hosking C, McLaughlin S, Lewis MD, Tumwine I, et al. Follicular Helper T Cells are Essential for the Elimination of Plasmodium Infection. *EBioMedicine* (2017) 24:216–30. doi: 10.1016/j.ebiom.2017.08.030
56. Tubo NJ, Pagán AJ, Taylor JJ, Nelson RW, Linehan JL, Ertelt JM, et al. Single naive CD4+ T cells from a diverse repertoire produce different effector cell types during infection. *Cell* (2013) 153:785–96. doi: 10.1016/j.cell.2013.04.007
57. Schattgen SA, Guion K, Crawford JC, Souquette A, Barrio AM, Stubbington MJT, et al. Linking T cell receptor sequence to transcriptional profiles with clonotype neighbor graph analysis (CoNGA). *bioRxiv* (2020). doi: 10.1101/2020.06.04.134536. 2020.06.04.134536.
58. Khosravi-Maharlooei M, Obradovic A, Misra A, Motwani K, Holzl M, Seay HR, et al. Crossreactive public TCR sequences undergo positive selection in the human thymic repertoire. *J Clin Invest* (2019) 129:2446–62. doi: 10.1172/JCI124358
59. Miles JJ, Douek DC, Price DA. Bias in the αβ T-cell repertoire: Implications for disease pathogenesis and vaccination. *Immunol Cell Biol* (2011) 89:375–87. doi: 10.1038/icb.2010.139
60. Opata MM, Carpio VH, Ibitokou SA, Dillon BE, Obiero JM, Stephens R. Early Effector Cells Survive the Contraction Phase in Malaria Infection and Generate Both Central and Effector Memory T Cells. *J Immunol* (2015) 194:5346–54. doi: 10.4049/jimmunol.1403216
61. Gasper DJ, Tejera MM, Suresh M. CD4 T-cell memory generation and maintenance. *Crit Rev Immunol* (2014) 34:121–46. doi: 10.1615/critrevimmunol.2014010373
62. Thomas N, Best K, Cinelli M, Reich-Zeliger S, Gal H, Shifrut E, et al. Tracking global changes induced in the CD4 T-cell receptor repertoire by immunization with a complex antigen using short stretches of CDR3 protein sequence. *Bioinformatics* (2014) 30:3181–8. doi: 10.1093/bioinformatics/btu523
63. Sun Y, Best K, Cinelli M, Heather JM, Reich-Zeliger S, Shifrut E, et al. Specificity, privacy, and degeneracy in the CD4 T cell receptor repertoire following immunization. *Front Immunol* (2017) 8:430. doi: 10.3389/fimmu.2017.00430
64. Tickotsky N, Sagiv T, Prilusky J, Shifrut E, Friedman N. McPAS-TCR: A manually curated catalogue of pathology-associated T cell receptor sequences. *Bioinformatics* (2017) 33:2924–9. doi: 10.1093/bioinformatics/btx286

65. Hill DL, Pierson W, Bolland DJ, Mkindi C, Carr EJ, Wang J, et al. The adjuvant GLA-SE promotes human Tfh cell expansion and emergence of public TCRβ clonotypes. *J Exp Med* (2019) 216:1857–73. doi: 10.1084/jem.20190301

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.