Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

High-resolution analysis of human cytomegalovirus-specific T-cell receptor repertoires

Alina Huth

aus

Aachen, Deutschland

2019

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Herrn PD Dr. Andreas Moosmann betreut und von Herrn Prof. Dr. Horst Domdey vor der Fakultät für Chemie und Pharmazie vertreten.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, den 31.07.2019

_____

Alina Huth

Dissertation eingereicht am 06.08.2019

1. Gutachter: Prof. Dr. Horst Domdey

2. Gutachter: PD Dr. Andreas Moosmann

Mündliche Prüfung am 28.10.2019

Teile dieser Arbeit wurden wie folgt veröffentlicht:

Huth, A., Liang, X., Krebs, S., Blum, H., and Moosmann, A. (2019). Antigen-Specific TCR Signatures of Cytomegalovirus Infection. *The Journal of Immunology* 202(3): 979–990.

Teile dieser Arbeit wurden im Rahmen der Dissertation wie folgt präsentiert:

02.09.–05.09.2018 Amsterdam, NL	**Konferenz ECI 2018 – 5th European Congress of Immunology** Poster: “T cell receptor signatures in cytomegalovirus carriers”
18.10.–20.10.2017 Teisendorf, DE	**PhD Retreat SFB-TR36 "Principles and applications of adoptive T cell therapy"** Vortrag: “Analysis of T cells targeting human cytomegalovirus by next-generation sequencing”
06.05.–13.05.2017 Porto Cervo, IT	**12th ENII EFIS EJI Summer school in advanced immunology** Poster: “Analysis of T cells targeting human cytomegalovirus by next-generation sequencing”
30.04.–04.05.2017 Noordwijkerhout, NL	**Konferenz CMV 2017** Vortrag und Poster: “Analysis of T cells targeting human cytomegalovirus by next-generation sequencing”
24.11–27.11.2016 Berlin, DE	**PhD Retreat SFB-TR36 "Principles and applications of adoptive T cell therapy"** Poster: “High throughput analysis of the T cell repertoire against human cytomegalovirus”
27.09.–30.09.2016 Hamburg, DE	**46th Annual Meeting of the DGfI** Poster: “High throughput analysis of the T cell repertoire against human cytomegalovirus”

TABLE OF CONTENTS

SUMMARY

The human herpesvirus cytomegalovirus (CMV) is a prevalent pathogen and infects a person for life. After primary infection, CMV latently resides in certain body cells to avoid clearance by the immune system. Primary infection and reactivations from latency usually go unnoticed in persons with an intact immune system. But in immunocompromised persons, such as transplant recipients, CMV can cause severe diseases, because these persons cannot keep the virus in check. Virus-specific $CD8^+$ T cells are the key component of CMV-directed immunity and their presence is associated with protection from overt disease. Several studies have already explored the human T-cell response against the virus. However, the CMV-specific T-cell repertoire is highly complex, and it remains an enigma which $CD8^+$ T cells with which CMV antigen specificity are the most protective and best control the virus.

$CD8^+$ T cells recognise antigens using a specialised surface protein, the heterodimeric $\alpha\beta$ T-cell receptor (TCR). This receptor is highly variable between T cells and billions of different TCRs can be found on the T cells of a single person. In this PhD project, the $CD8^+$ T-cell repertoire specific for CMV was assessed at high resolution: Virus-specific $CD8^+$ T cells from healthy virus carriers were enriched through in vitro stimulation with short peptides and analysed by TCR mRNA sequencing, primarily of the β-chain. CMV-specific TCRβ sequences were identified by computational analysis via multi-sample comparisons. Subsequently, the magnitude and diversity of interindividual and intraindividual CMV-specific TCRβ repertoires was analysed. The focus was on the TCR repertoire specific for HLA-C–restricted CMV peptides, because presentation of such peptides is least affected by viral immunomodulation.

In total, 1809 CMV-specific TCRβ amino acid sequences were identified in this project. The CMV-specific TCR repertoires were generally highly clonal, but repertoire diversity differed between peptides. CMV-specific TCRβ clonotypes, particularly those specific for HLA-C–restricted peptides, were highly frequent in peripheral blood of virus carriers; in 3 of 9 donors, the TCR with the highest frequency in the entire CD8+ T cell repertoire was specific for one CMV peptide. Several TCRβ clonotypes with the same CMV peptide specificity were identical or strikingly similar at the amino acid level and shared by multiple donors. The cumulative frequency of these 162 shared TCRβ clonotypes was significantly higher in CMV-positive than CMV-negative donors in this cohort and

a large independent validation cohort. Consequently, CMV infection leaves a specific TCRβ signature in the T-cell repertoires of its human hosts.

Signature TCRβ sequences will be valuable in disease monitoring, for instance as markers for the presence of a CMV-specific T-cell response. In addition, such TCRs hold great potential for adoptive T-cell transfer, since they are tolerant to a wide range of HLA-self peptide complexes and are therefore less likely to cause toxicity in the recipient.

ABBREVIATIONS

APC	antigen-presenting cell
B blast	CD40-stimulated B cell
bp	base pairs
C	constant gene segment
CAR	chimeric antigen receptor
CD	cluster of differentiation
CDR	complementarity-determining region
cDNA	complementary DNA
CMV	human cytomegalovirus
CsA	Cyclosporin A
d	day
D	diversity gene segment
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
EBV	Epstein-Barr virus
ELISA	enzyme-linked immunosorbent assay
ER	endoplasmic reticulum
ERAP1	endoplasmic reticulum aminopeptidase 1
FACS	fluorescence-activated cell sorting
FCS	fetal calf serum
FSC-A	forward scatter area
FSC-H	forward scatter height
$\times$g	times gravitational force equivalent
h	hour
HIV	human immunodeficiency virus
HLA	human leukocyte antigen
HSCT	hematopoietic stem cell transplantation
HTS	high-throughput sequencing
IFN	interferon
Ig	immunoglobulin
IL	interleukin
IQR	interquartile range

J	joining gene segment
KIR	killer-cell immunoglobulin-like receptor
MACS	magnetic-activated cell sorting
MHC	major histocompatibility complex
mini-LCL	mini-lymphoblastoid cell line
mRNA	messenger RNA
N	degenerated nucleotide
n.d.	not done
NK cells	natural killer cells
nt	nucleotide
ORF	open reading frame
PBMC	peripheral blood mononuclear cell
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
RNA	ribonucleic acid
rt	room temperature
SSC-A	side scatter area
TAP	transporter associated with antigen processing
TCR	T-cell receptor
TNF	tumour necrosis factor
TRB	gene locus encoding the TCRβ chain
TRA	gene locus encoding the TCRα chain
UL	unique long
US	unique short
V	variable gene segment

1 INTRODUCTION

1.1 Human cytomegalovirus, a large viral pathogen

1.1.1 Virology

Human cytomegalovirus (CMV) is the prototypic representative of the β-herpesvirus subfamily (Crough and Khanna 2009) and is also referred to as human herpesvirus 5. CMV hast the largest genome of all known human viruses (Boeckh and Geballe 2011): Its double-stranded deoxyribonucleic acid (DNA) genome comprises approximately 235 kilo base pairs (bp) (Gardner and Tortorella 2016). The genome consists of a unique long (UL) and a unique short (US) region, which are separated by an internal repeat region and flanked by terminal repeats (Murphy et al. 2003). It contains more than 200 open reading frames (ORFs), of which 151 were, so far, shown to encode immunogenic proteins (Sylwester et al. 2005). In the virion, the genome is enclosed by the icosahedral capsid, the tegument and the viral envelope (Figure 1.1).

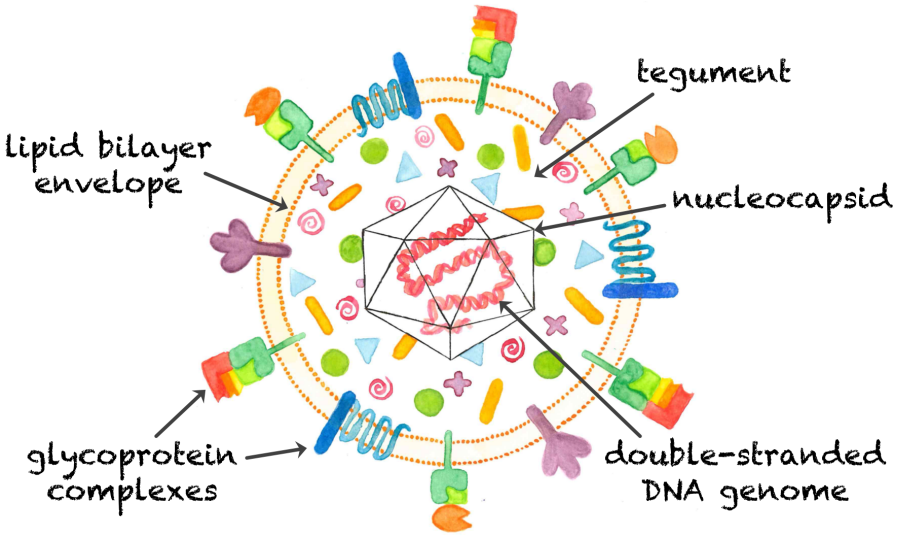


Figure 1.1: Schema of the CMV virion. The double-stranded DNA genome is encapsulated by the icosahedral nucleocapsid, the tegument, and the lipid bilayer envelope. Most phosphoproteins are contained in the tegument, while glycoproteins are located in the outer membrane, where they form protein complexes.

The tegument contains many phosphoproteins, among them the lower matrix phosphoprotein 65 (pp65), the virion transactivator pp71, and the core virion maturation protein pp150 (Crough and Khanna 2009). The viral envelope consists of a lipid bilayer membrane sprinkled with glycoprotein complexes (Gardner and Tortorella 2016). Ap-

proximately 20 different glycoproteins with functions in cell attachment and cell entry were found to be located in the envelope, including glycoprotein B (gB), gH, gL, and gO (Crough and Khanna 2009; Gardner and Tortorella 2016). Spherical mature virions measure 200–300 nm in diameter (Crough and Khanna 2009; Yu et al. 2017).

1.1.2 Life cycle

Human CMV has a broad tropism. It infects and subsequently replicates in various cell types, including fibroblasts and epithelial cells, but also macrophages, neurons, and muscle cells (Revello and Gerna 2010; Beltran and Cristea 2014; Acquaye-Seedah et al. 2015). Infection of host cells is initiated by adsorption of CMV virions onto the cell surface (Figure 1.2) followed by different entry mechanisms depending on the cell type (Compton et al. 1993). Entry into fibroblasts occurs via direct membrane fusion and is mediated by protein gB in cooperation with the gH/gL/gO trimeric complex (Huber and Compton 1998; Isaacson and Compton 2009). Entry into endothelial and

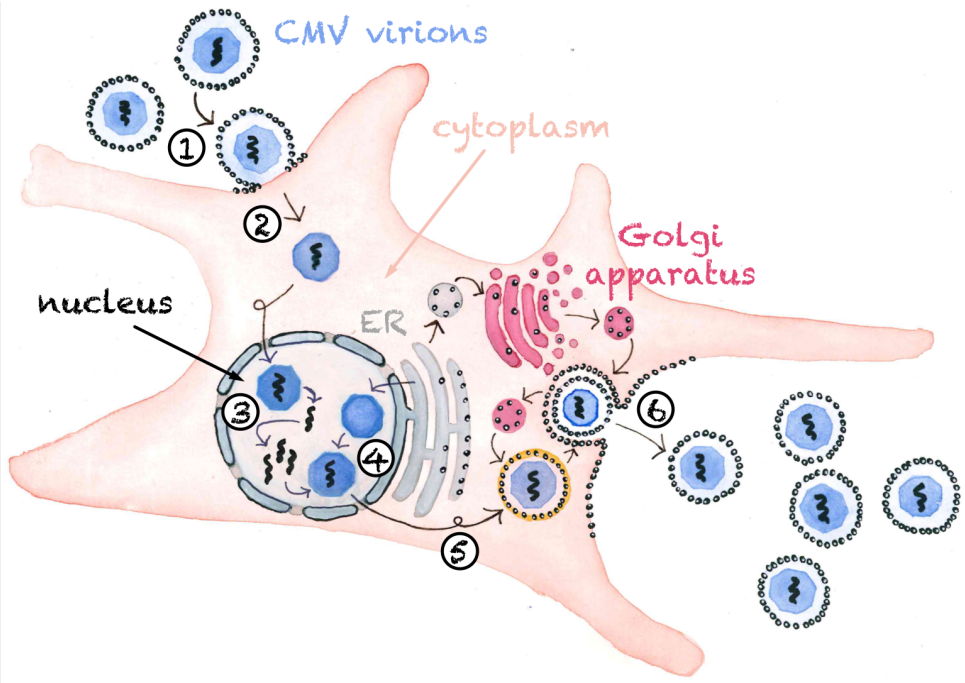


Figure 1.2: CMV life cycle. The virion attaches to its host cell by adsorption and enters the cell through direct fusion with the cell membrane or endocytosis (1). Next, the capsid containing the CMV genome is released into the cytoplasm (2). The capsid enters the nucleus, where the viral DNA is released and replicated (3). Viral proteins are translated, and the viral DNA is encapsulated in newly assembled capsids (4). After nuclear egress (5), the capsids are enveloped (5) and the mature virions are either released by cell lysis or transmitted from cell to cell (6). Figure based on Crough and Khanna 2009.

epithelial cells, as well as into macrophages, occurs via endocytosis combined with low-pH fusion (Ryckman et al. 2006; Acquaye-Seedah et al. 2015). It is mediated by the gH/gL/UL128–131 pentameric complex (Hahn et al. 2004; Wang and Shenk 2005; Ryckman et al. 2008). During viral entry, the capsid enclosing the viral DNA is released into the cytoplasm and translocates to the nucleus (Crough and Khanna 2009; Acquaye-Seedah et al. 2015). Next, the viral DNA is replicated and translated. Viral DNA and freshly produced proteins are assembled to virions and enveloped. New virions are either released into the extracellular space (cell-free transmission) or directly into a neighbouring cell (cell-to-cell transmission) (Acquaye-Seedah et al. 2015; Murrell et al. 2017).

Apart from infecting cells productively, CMV can also establish latency in cells of the myeloid lineage, particularly in monocytes, and hematopoietic progenitor cells (Dupont and Reeves 2016). Other cell types that may be latent CMV reservoirs are endothelial progenitor cells and derived aortic endothelial cells, and neuronal progenitor cells (Fish et al. 1998; Belzile et al. 2014; Dupont and Reeves 2016), but more data is required to validate these putative sites of latency. Latency is defined as a state in which the viral genome is present in the cell as an extrachromosomal plasmid (episome) and no infectious virus is produced (Noriega et al. 2012). However, the virus does not remain completely silent during latency. It was shown for other herpesviruses that non-coding ribonucleic acids (RNAs) exert various latency-promoting functions, such as epigenetic modifications, anti-apoptotic effects, and regulation of viral and cellular gene expression (Perng et al. 2000; Pfeffer et al. 2004; Cai et al. 2005; Umbach et al. 2008; Dupont and Reeves 2016). Non-coding RNAs were also found in cells that were latently infected with CMV. For instance, several long non-coding RNAs were found in latently infected cells, of which one, RNA4.9, is likely involved in epigenetic silencing of immediate-early genes (Rossetto et al. 2013). Furthermore, CMV-encoded microRNAs that suppress immediate-early gene expression were identified (Grey et al. 2007; Murphy et al. 2008). Consequently, CMV, like its herpesvirus siblings, also uses non-coding RNAs to establish and maintain latency. In addition to these RNAs, a discrete set of CMV proteins is expressed during latency. These proteins are predominantly involved in immunoevasion and modulation of the cellular environment, but may also exert other functions, such as maintenance of the viral genome (Dupont and Reeves 2016). Reac-

tivation of CMV from latency is thought to be induced by cellular differentiation: It was shown that CMV reactivates upon differentiation of monocytes to macrophages or dendritic cells (Taylor-Wiedeman et al. 1994; Reeves et al. 2005). Presumably, reactivation of CMV is promoted by histone modifications, which, in turn, lead to activation of the major immediate-early promoter that is repressed during latency (Dupont and Reeves 2016). Despite recent advances in the field, the precise mechanisms leading to establishment and maintenance of latency and to viral reactivation from latency remain poorly understood (Crough and Khanna 2009; Dupont and Reeves 2016; Goodrum 2016).

Based on their sequence of expression, CMV proteins are divided into 3 categories: immediate-early, early, and late proteins (Stinski 1978; Stenberg et al. 1990; Crough and Khanna 2009). Immediate-early proteins are expressed directly after infection (0h–2h) and require no prior synthesis of viral proteins (Stinski 1978; Stenberg et al. 1990; Crough and Khanna 2009; Goodrum 2016). The most prominent immediate-early proteins are IE1 and IE2. Expression of early proteins requires the presence of immediate-early proteins and takes place up to 24h after infection. Late proteins are expressed more than 24 hours after infection. The most intensively studied late protein is structural matrix protein pp65.

1.1.3 Pathology

CMV is a widespread human pathogen. Between 30%–90% of persons carry the virus and its prevalence increases with age (Crough and Khanna 2009). Upon initial contact, CMV establishes a persistent infection in its host and, in doing so, switches to a quiescent latent state with occasional reactivations. The virus can be transmitted in various ways: sexually, via the placenta, via saliva, via transplants, or via blood transfusion (Crough and Khanna 2009). Primary infection with CMV and its reactivation are usually asymptomatic in healthy persons, although CMV mononucleosis is infrequently observed during seroconversion (Sissons and Carmichael 2002). In immunocompromised persons, however, CMV can cause severe clinical complications (Crough and Khanna 2009). Groups at risk of developing CMV disease are congenitally infected fetuses, human immunodeficiency virus (HIV) carriers, and patients receiving solid organ transplantation or hematopoietic stem cell transplantation (HSCT).

The highest-risk constellation in the setting of HSCT is when a CMV-positive patients receives a graft from a CMV-negative donor, because the donor graft does not contain antigen-experienced, protective T cells (thymus-derived lymphocytes) against the virus. Additional risk factors for CMV infection and disease after HSCT include mismatched human leukocyte antigen (HLA) alleles between donor and recipient, and the use of T-cell–depleted or cord-blood–derived grafts (Chen et al. 2018). Owing to the advancement of CMV prevention strategies, less than 5% of patients today develop CMV disease after receiving HSCT (Camara 2016). Nevertheless, morbidity and mortality are severe in transplant recipients who do present with CMV disease. Common clinical manifestations of early CMV disease, which sets on up to 100 days post transplantation, are pneumonitis and gastrointestinal disease. Both conditions are associated with poor overall survival at mortality rates of ca. 70% each (Camara 2016; Baroco and Oldfield 2008). Late CMV disease ($>$100 days post transplantation) is considered a major threat to long-term survival of patients and mostly manifests as pneumonitis, although retinitis and encephalitis were also observed (Crough and Khanna 2009).

1.1.4 Treatment

Management of CMV after HSCT can be prophylactic, preemptive, or therapeutic (Camara 2016). In prophylactic treatment, patients receive antivirals, such as intravenous ganciclovir, in the absence of detectable virus to prevent CMV infection or reactivation. Preemptive treatment means that measures are taken when active CMV infection is diagnosed in a patient, but CMV disease has not yet clinically manifested. Measures of preemptive treatment include intravenous injection of ganciclovir, treatment with foscarnet, or oral administration of valganciclovir in low-risk patients (Ljungman et al. 2008). A randomised double-blind study (Boeckh et al. 1996) showed that overall survival after HSCT was similar for patients treated prophylactically or preemptively with ganciclovir. Since prophylactic treatment was associated with increased incidence of fungal and bacterial infections, preemptive treatment is currently the gold standard of CMV treatment strategies. When the worst cannot be averted and an HSCT recipient is diagnosed with CMV disease, therapeutic measures are put to action. These measures include the administration of antivirals gangciclovir or foscarnet, often in combination with CMV immunoglobulin (Ig) (Camara 2016).

1.1.4.1 Antivirals

Different antivirals have been developed to prevent or treat CMV disease. Common practice is the administration of the established antivirals ganciclovir and its oral analogue valganciclovir, foscarnet or cidofovir (Boeckh and Geballe 2011). All three antivirals inhibit the viral DNA polymerase and were shown to be effictive in clinical trials. Unfortunately, treatment with these antivirals has severe adverse effects, such as organ-directed toxicity, and it frequently leads to drug resistance (Boeckh and Geballe 2011). Recently, three new antivirals have been evaluated in clinical trials: maribavir, brincidofovir, which is a prodrug of cidofovir, and letermovir. Maribavir targets CMV protein kinase UL97, whereas brincidofovir acts on the viral DNA polymerase in the same way as cidofovir. Both antivirals failed clinical trial phase III, since their effectiveness against CMV could not be demonstrated at the tested doses. A second phase III trial with higher doses has been started for maribavir, since it was well tolerated and had virtually no side effects in the first round of clinical trials (Camara 2016; Chen et al. 2018). The third novel antiviral, letermovir, uses yet a different mechanism of action and targets the CMV terminase complex UL56. Letermovir effectively reduced CMV reactivation and mortality in two randomised studies with tolerable adverse effects (Chemaly et al. 2014; Marty et al. 2017). Hence, it was approved by the US Food and Drug Administration and the European Medicines Agency in 2017 (Chen et al. 2018). Despite the success of letermovir, a number of cases of breakthrough CMV viremia and CMV pneumonia were already reported (Knoll et al. 2019; Lischka et al. 2015). They were likely caused by different mutations in CMV UL56 in patients on letermovir prophylaxis. It was shown that letermovir rapidly induces diverse mutations which confer drug resistance to letermovir in vitro (Chou 2015). These recent findings imply that letermovir induces drug-resistant CMV strains and is therefore not ideal for prevention and treatment of CMV disease. To treat CMV disease, it is currently recommended to combine antiviral treatment with intravenous administration of purified anti-CMV IgG from pooled human sera. However, a large study showed that the addition of anti-CMV IgG had no beneficial effects on survival of patients with CMV disease (Camara 2016).

1.1.4.2 Experimental vaccines

Great efforts are made to try and develop new and innovative CMV prevention and treatment stategies. A major milestone in the prevention of CMV infection and reacti-

vation would be the development of a CMV vaccine that is able to boost potent adaptive immune responses. Experimental CMV vaccines have been developed for several decades, but without resounding success. Vaccine candidates range from live attenuated virus to viral-vectored and plasmid DNA vaccines, and subunit vaccines (McVoy 2013). A live attenuated virus vaccine derived from the CMV strain Towne was safe and induced cellular and humoral responses, but failed to protect mothers from contracting primary CMV infection from their children in a blind randomised study. This was likely due to lower neutralising titers with the vaccine than after wild type infection (Adler et al. 1995; McVoy 2013). To increase immunogenicity of the Towne vaccine, 25% of the Towne vaccine genome were replaced with analogous sequences from CMV strain Toledo (McVoy 2013). The four resulting Towne/Toledo chimeric vaccines were safe, but did not increase existing immunity of CMV-positive subjects in a phase I study (Heineman et al. 2006). In an additional phase I trial with 36 healthy CMV-negative men, vaccination with the Towne/Toledo chimeras lead to detectable anti-CMV antibody responses in 11 subjects (Adler et al. 2016). An additional $CD4^+$ T-cell response was detected in 2 of the responsive subjects and an additional $CD8^+$ T-cell response was detected in 8 of the responsive subjects. $CD8^+$ T-cell responses were always directed against IE1, but $CD8^+$ T cells against pp65, UL32, UL36, UL48, and UL55 were also infrequently detected. Towne/Toledo chimeras 2 and 4 were most immunogenic and will thus be tested in additional clinical trials (Adler et al. 2016). ASP 0113, also known as TransVax, is a plasmid vaccine comprised of adjuvant and a DNA plasmid encoding CMV proteins gB and pp65. The outcome of a phase II trial of ASP 0113 in HSCT recipients was encouraging (Kharfan-Dabaja et al. 2012), but it was discontinued in January 2018 due to lack of efficacy in the phase III trial (Astellas Pharma 2018). Besides, it was not effective in the prevention of CMV viremia in kidney transplant recipients in another phase II trial (Vincenti et al. 2018). VCL 6365 or CyMVectin, which consists of the same plasmid DNA as ASP 0113, but an improved adjuvant, was withdrawn from a phase I clinical trial in 2016 and discontinued shortly after in March 2018 (AdisInsight database 2018). RNA virus-vectored vaccine AVX 601 encodes gB and a pp65/IE1 fusion protein. It was safe in a phase I clinical trial and induced polyfunctional $CD4^+$ and $CD8^+$ T cell responses (Bernstein et al. 2009; AlphaVax a). It is still in the AlphaVax pipeline (AlphaVax b). The subunit vaccine gB/MF59 combines a soluble

form of CMV gB with adjuvant MF59. The vaccine reduced primary infection in women by 50% (Pass et al. 2009; McVoy 2013) and was marginally protective against CMV infection of teenage girls in an additional phase II clinical trial (Bernstein et al. 2016). Due to the overall insufficient efficacy, the gB/MF59 vaccine was not continued as a standalone vaccine for CMV infection (Bernstein et al. 2016). Additional approaches to CMV vaccination of transplant patients are currently tested in clinical trials. These include a recombinant Modified Vaccinia Ankara virus encoding pp65, IE1, and IE2, and a vaccine with a pp65 peptide plus adjuvant (Diamond et al. 2018).

1.1.4.3 Adoptive T-cell transfer

Adoptive transfer of CMV-specific T cells is an alternative approach to prevent or treat CMV disease in HSCT recipients. CMV-specific T cells can be obtained by selective enrichment from CMV-positive bone marrow donors or by genetic modification of T cells from CMV-negative bone marrow donors (Peggs 2009; Roddie and Peggs 2017; Kaeuferle et al. 2019). The first proof of principle study of adoptive T-cell therapy used CMV-specific $CD8^+$ T-cell clones, which were generated by ex vivo stimulation of donor-derived peripheral blood mononuclear cells (PBMCs). Adoptive transfer of these $CD8^+$ CMV-specific T-cell clones led to reconstitution of CMV immunity in HSCT patients (Walter et al. 1995). Polyclonal CMV-specific T-cell lines, which were produced by stimulation of PBMCs from bone marrow donors with CMV lysate or CMV-derived peptide, were also potent in reconstituting CMV immunity (Einsele et al. 2002; Peggs et al. 2003; Micklethwaite et al. 2007). In another clinical study, multi-specific T-cell lines for adoptive T-cell transfer were produced by stimulation of donor-derived PBMCs with Epstein-Barr virus (EBV)-transformed, adenovirus-transduced B cells as antigen presenting cells (Leen et al. 2006). The transferred multi-specific T cells had antiviral activity against all three viruses (CMV, EBV, and adenovirus) in vivo. In a follow-up study, multi-specific T-cell lines generated by in vitro stimulation with overlapping peptide pools derived from 5 viruses, including CMV, were protective in 94% of the tested donors (Papadopoulou et al. 2014). Direct selection of CMV-specific T cells from donor samples by peptide-MHC multimer-based enrichment (Cobbold et al. 2005; Neuenhahn et al. 2017) or enrichment of T cells secreting interferon (IFN)γ in response to CMV protein or peptide (Peggs 2009) can also be used to obtain CMV-specific T cells for adoptive T-cell transfer. Peptide-MHC multimers, for short multimers, are oligomers of peptide-

loaded major histocompatibility complex (MHC)-encoded molecules that can be coupled to a fluorophore. They are used to stain T cells specific for a given peptide-MHC complex. A major drawback of all aforementioned methods to enrich CMV-specific T cells is that they require a CMV-positive bone marrow donor. Furthermore, all methods are costly, since T-cell preparations have to be generated for each patient individually as a highly personalised medication. Lastly, chances are the transferred T cells cause alloreactivity or graft-versus-host disease, especially in less well-defined preparations, such as polyclonal and multi-specific T-cell lines. Genetically modified T-cell products are a promising alternative to circumvent these problems. CMV-pp65–specific T-cell receptor (TCR)-transgenic T cells (Schub et al. 2009) and CMV-gB–specific chimeric antigen-receptor (CAR) T cells (Full et al. 2010) showed relevant effector functions in vitro. Consequently, adoptive transfer of genetically modified T cells holds great potential for prevention of CMV infection, reactivation, and disease in HSCT recipients.
Despite the availability of various treatment options, CMV remains a major threat to the health of immunocompromised patients and newborns (Navarro 2016). Thus, there is a great need for new safe and effective CMV therapeutics.

1.1.5 Diagnosis

Viral infection can be diagnosed in various ways. Firstly, the presence of virus particles in blood (viremia) or urine (viruria), or in other patient samples, can be diagnosed by cell culture. Traditionally, CMV was isolated in fibroblast tissue cultures, which can be produced from various patient samples. An existing CMV infection will cause a cytopathic effect in the fibroblast cultures, which can be confirmed by fluorescent antibody stainings. A variation to standard fibroblast cell culture is the shell vial method, in which samples are centrifuged onto fibroblast monolayers and antibody-stained for early CMV antigens after a brief incubation period. The shell vial method dramatically reduces the time from sample collection to diagnosis compared to the classical cell culture protocol, but also reduces the sensitivity by up to 32% (Hoz et al. 2002). Secondly, an active CMV infection can be confirmed by detecting CMV pp65 antigen with pp65-specific monoclonal antibodies (antigenemia). Compared to viral culture, CMV antigenemia was equally specific and more sensitive, but it was less reliable when specimens other than blood or cerebrospinal fluid were tested (Hoz et al. 2002). Thirdly, CMV serology

can indirectly diagnose CMV infection by detecting the presence of a humoral response to CMV in serum from patients. Most commonly, anti-CMV IgG or IgM antibody titres are determined with an enzyme-linked immunosorbent assay (ELISA) (Kotton 2013). However, CMV serology can sometimes be ambigous with absorption values in the range of the cutoff controls. Last but not least, the presence of CMV DNA (DNAemia) can be detected to diagnose infection. In this method, CMV DNA is amplified by polymerase chain reaction (PCR) from clinical samples or after in vitro proliferation in culture. PCR is a highly sensitive method and can detect CMV appearance earlier than cell culture or antibodies (Storch et al. 1994). It is currently the diagnostic standard to detect active CMV infection.

1.2 The cellular adaptive immune system

Every living creature is permanently surrounded and attacked by various pathogens and other damage-causing agents that are a potential threat to its well-being. The human body defends itself against pathogens with a complex network of specialised cells, proteins, and messenger molecules: the immune system. The immune system is divided into two main components: the innate immune system and the adaptive immune system. While the innate immune system reacts immediately with non-specific defense mechanisms, the adaptive immune system is based on acquired immunity and highly pathogen-specific. After an initial response to a pathogen is elicited, the adaptive immune system forms an immunological memory. This immunological memory enables the body to mount a rapid immune response in case a pathogen is re-encountered. The adaptive immune system is subdivided into humoral immunity, which is based on macromolecules, such as antibodies secreted by B lymphocytes (B cells), and cellular immunity, which is largely mediated by pathogen-specific immune cells, the T lymphocytes (T cells). T cells do not directly remove pathogens, but recognise processed antigen presented on body cells. Based on their distinct functions, T cells are divided into two subsets: cytotoxic T cells and T helper cells. When cytotoxic T cells are activated, they release cytotoxins and thereby induce apoptosis of infected cells. T helper cells regulate innate and adaptive immune responses, but mostly do not exert direct target killing function.

1.2.1 T-cell development

T cells derive from hematopoietic stem cells in the bone marrow and are named after their site of maturation, the thymus (Murphy and Weaver 2016). When progenitor T cells enter the thymus, they receive a signal via their Notch1 receptor, which activates particular genes, thus inducing commitment to the T-cell lineage. Initially, progenitor T cells in the thymus, the thymocytes, do not express the prototypic T-cell surface molecules required for recognition of infected cells. These surface molecules are the pathogen-specific T-cell receptor (TCR) and its co-receptors cluster of differentiation (CD)3, and CD4 or CD8. Because neither CD4 or CD8, the co-receptors defining T-cell function later on, are expressed in the early stage of T-cell development, the immature T cells are called double-negative thymocytes. Most double-negative thymocytes differentiate to $\alpha\beta$ T cells, which express a heterodimeric TCR consisting of an α-chain and a β-chain. In the double-negative stage of $\alpha\beta$ T-cell development, the TCRβ chain locus is rearranged. If gene rearrangement was productive, the expressed TCRβ chain and a surrogate pre-TCRα chain pair to form the pre-TCR. The pre-TCR is expressed on the thymocyte surface in complex with CD3. Pre-TCR–signalling induces thymocyte proliferation and the simultaneous expression of surface co-receptors CD4 and CD8. Thus, the thymocyte enters the double-positive stage of T-cell development, in which the TCRα chain locus is rearranged. Successfully recombined $\alpha\beta$ TCRs are positively selected for their ability to recognise self-peptide/self-MHC complexes. T cells expressing a functional TCR pass positive selection and mature to single-positive thymocytes, which lose expression of either CD4 or CD8. During and after the double-positive stage, developing T cells undergo a second round of quality control, the negative selection. In the negative selection stage, autoreactive T cells that respond to self-peptides are eliminated. After passing negative selection, single-positive mature T cells exit the thymus. The antigen-inexperienced naïve T cells circulate in the blood and peripheral lymphoid organs, where they may encounter their cognate antigen and become activated.

1.2.2 Antigen presentation, processing, and recognition

T cells recognise pathogen-derived foreign peptides, which are processed from antigenic proteins and presented by body cells on specialised molecules encoded by the

MHC gene set (Murphy and Weaver 2016). MHC molecules come in two distinct classes. MHC class I molecules are heterodimers and consist of a long and variable α-chain paired with a smaller and invariant β2-microglobulin chain. They are expressed on all nucleated body cells and primarily present peptides derived from intracellular antigen. Intracellular antigenic proteins in the cytosol are mainly degraded via the proteasomal pathway and give rise to many short peptides. The resulting peptides are continually transported to the endoplasmic reticulum (ER) lumen by a protein complex called "transporter associated with antigen processing" (TAP). MHC class I α-chains and β2-microglobulin chains are also transported to the ER, where they are assembled to heterodimeric MHC class I molecules. Binding of a matching short peptide to the MHC binding groove stabilises the MHC molecule. The stable peptide-MHC complex is then translocated to the surface, where it is exposed to T cells. Peptides presented by MHC class I molecules are recognised by cytotoxic $CD8^+$ T cells, which are activated, secrete cytotoxins and induce apoptosis of the presenting body cell. MHC class II molecules are also heterodimers, but consist of two homogenous peptide chains, the α- and β-chain. They are only expressed by professional antigen-presenting cells (APCs), such as dendritic cells or macrophages, and they primarily present peptides derived from extracellular antigen. Extracellular proteins from exogenous sources are internalised by endocytosis or phagocytosis and are subsequently degraded by lysosomal proteolysis under acidic conditions (Blum et al. 2013). Lysosomes containing degraded peptide fuse with vesicles containing MHC class II molecules. Subsequently, peptide-MHC complexes are formed and transported to the cell surface for antigen presentation to T cells. Peptides presented by MHC class II molecules are recognised by $CD4^+$ T cells, which subsequently orchestrate further immune responses. In some cases, $CD4^+$ T cells can also directly kill the peptide-presenting cells. This was, for instance, shown in the context of antiviral T-cell responses, particular those against CMV (Suni et al. 2001; Hegde et al. 2005; Marshall and Swain 2011; Pachnio et al. 2016). In humans, MHC molecules are referred to as human leukocyte antigens (HLA) and are encoded by several genes of the *MHC* locus on chromosome 6. There are three classical HLA class I heavy chains (HLA-A, HLA-B, and HLA-C), and three gene loci encoding the major HLA class II chains (HLA-DP, HLA-DQ, and HLA-DR).

1.2.3 TCR gene rearrangement

T cells recognise infected cells with their heterodimeric TCR, which typically consists of an α-chain and a β-chain (Murphy and Weaver 2016). Less frequently, T cells express an alternative TCR consisting of a γ-chain paired with a δ-chain. Such $\gamma\delta$ T cells include δ2 T cells that are responsible for the recognition of phosphoantigens, but they will not be addressed in this thesis. TCRs are highly diverse molecules, but each T cell only expresses a single functional version of the TCRβ chain and 1–2 TCRα chains, one of which does not pair with the β-chain and is therefore non-functional. TCR diversity is created during T-cell maturation through rearrangement of gene segments in the TCRα and TCRβ gene loci, *TRA* and *TRB* (Figure 1.3). This process was termed V(D)J recombination. The human *TRA* locus is located on chromosome 14 and

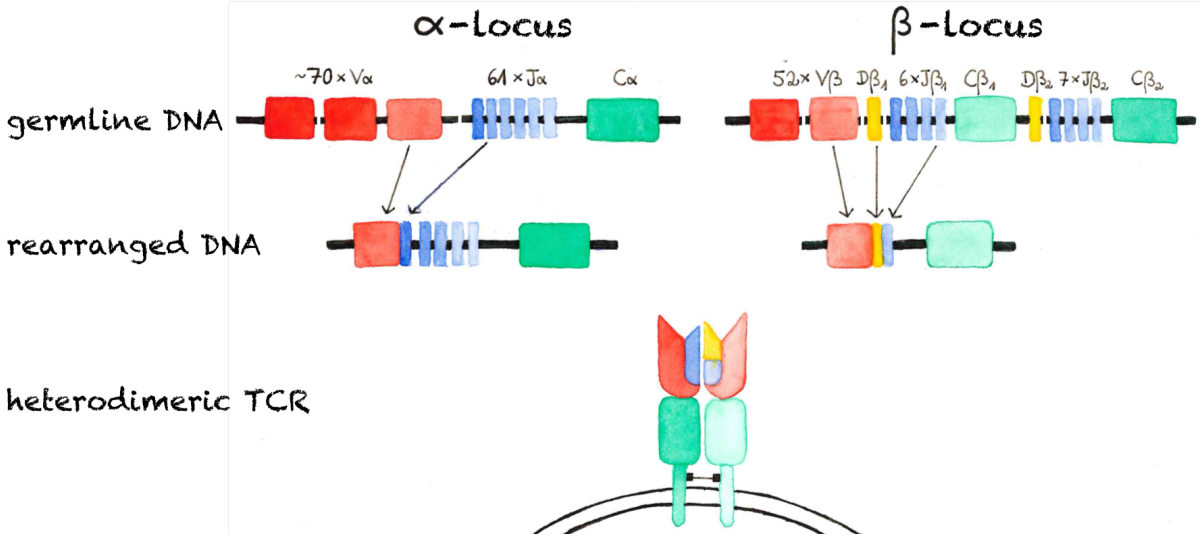


Figure 1.3: V(D)J recombination of the TCRα (*TRA*) locus and the TCRβ (*TRB*) locus during T-cell development. Functional VJα exons are generated by pasting one of $\approx$70 Vα to one of 61 Jα gene segments. Functional VDJβ exons are generated by combining one of 52 Vβ with one of 2 Dβ segments and one of 13 Jβ gene segments, whereby only Dβ and Jβ gene segments of the same gene segment cluster can be combined (there are two DJC clusters). Random nucleotides are inserted or deleted at the fusion sites of the gene segments. Successfully rearranged TCR chains are transcribed and the V(D)J exon is spliced to the C exon during messenger RNA (mRNA) processing. The spliced mRNA is translated and the α-chain and β-chain pair to form a functional heterodimeric TCR.

contains $\approx$70 different variable region (V) gene segments, 61 joining region (J) gene segments, and 1 constant region (C) gene segment. During T-cell maturation, one of the Vα and Jα gene segments each are pasted together to obtain a functional VJ exon. Random nucleotides are inserted or deleted at the fusion site between V gene segment

and J gene segment, which further increases TCRα chain diversity. The human *TRB* locus is located on chromosome 7 and consists of 52 V gene segments, 2 diversity region (D) gene segments, 13 J gene segments, and 2 C gene segments. The Dβ, Jβ, and Cβ gene segments form 2 clusters, and only genes of the same cluster are rearranged together to form a productive VDJ exon. Because of the additional D gene segments and the insertion or deletion of non-template nucleotides to both sides of the D gene segment during V(D)J recombination, the TCRβ chain is even more diverse than the TCRα chain. After successful rearrangement of the *TRB* and subsequently the *TRA* gene locus, the two TCR chains are transcribed and translated independently. Newly synthesised TCRα and TCRβ protein chains pair and are transported to the T-cell surface.

1.2.4 TCR:peptide-MHC interaction

The TCR generally contacts the peptide-MHC complex with 3 loops of each chain. These loops are termed complementarity-determining regions (CDRs, Figure 1.4). CDR1 and CDR2 are completely encoded by the V gene segments and mainly contact the MHC molecule. CDR3 is highly variable, because it is located at the site where V, (D,) and J gene segments are joined in an imprecise way with template-independent insertions and deletions of nucleotides, and binds directly to the presented peptide (Garcia et al. 1996). Between 1–400 interactions of TCRs with peptide-MHC com-

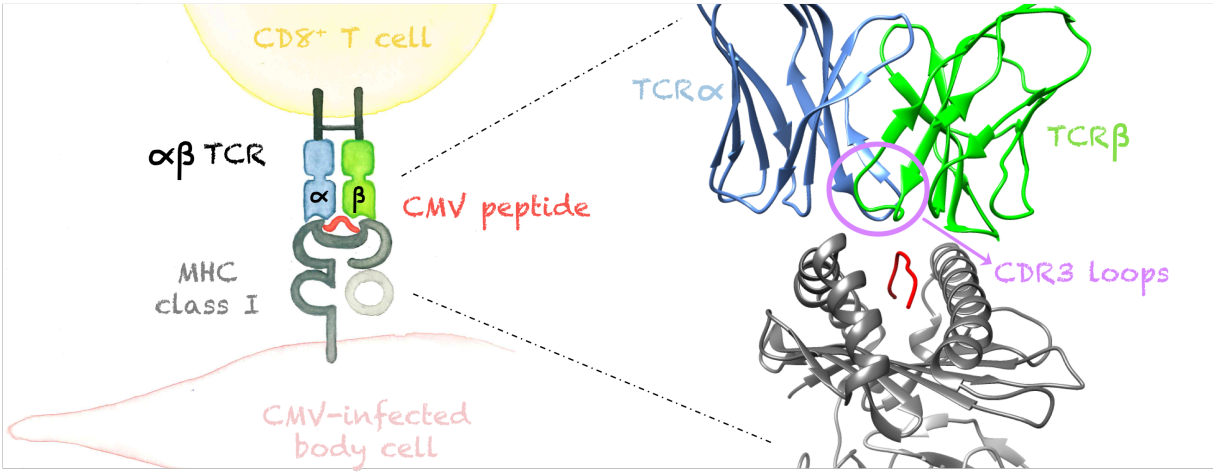


Figure 1.4: Interaction of the TCR with a peptide-MHC complex. The CDR1 and CDR2 loops of each TCR chain are in close contact with the MHC α-helices, while the CDR3 loops make direct contact to the peptide. Crystal structure modified from Ishizuka et al. (PDB ID: 2VLJ) (Ishizuka et al. 2008)

plexes are estimated to be required for activation of a T cell (Huppa and Davis 2003). Apart from the TCR binding to its cognate peptide-MHC complex, signalling through T-cell co-receptor CD3, and often stabilisation by co-receptor CD4 or CD8, is required for T-cell activation. Typically, peptides presented on MHC class I molecules are 8–10 amino acids long (Wieczorek et al. 2017). The ends of the peptide are anchored in the confined MHC class I binding groove. Peptides presented on MHC class II molecules are generally longer and measure 13–25 amino acids; their termini protrude at the sides from the open binding groove. Immunogenic peptides are presented by their restricting MHC in a particular spatial conformation. The portion of the peptide-MHC complex that is bound by TCRs is referred to as T-cell epitope.

1.3 CMV-directed immunity and viral immune evasion

Infection with CMV initiates a massive immune response in the human host, which involves both humoral and cellular immunity. Although the human body mobilises a strong humoral response against CMV antigens, in particular against the late phase envelope glycoprotein gB (Britt et al. 1990), it remains controversial whether or not the CMV-directed humoral immunity is protective against viral disease. The effectiveness of the humoral immunity in virus control appears to be limited, because CMV is mostly located within body cells, where it is protected from antibody neutralisation (Hoz et al. 2002). Nonetheless, it is believed that CMV-specific antibodies block cell-free transmission of the virus and thereby reduce the severity of CMV disease (Jonjić et al. 1994; Boppana and Britt 1995; Crough and Khanna 2009).

The CMV-specific T-cell response is, to date, the largest known human T-cell response against a single pathogen. CMV-specific T cells, in particular effector memory T cells, accumulate in virus carriers after initial infection. This process was designated ”memory inflation” and has been described in numerous studies in mice (Holtappels et al. 2000; Karrer et al. 2003) and humans (Gillespie et al. 2000; Sylwester et al. 2005; Hosie et al. 2017). Accordingly, it was observed that CMV-specific $CD8^+$ T cell are more frequent in the elderly than in young persons (Klenerman and Oxenius 2016). Memory inflation is presumably a result of recurring virus reactivation and subsequent mobilisation of specific T cells over time (Smith and Khanna 2013). $CD8^+$ T cells are key players in virus control and their presence is associated with protection from overt

CMV disease (Quinnan et al. 1982; Cwynarski et al. 2001; Bunde et al. 2005; Sacre et al. 2008). Indeed, adoptive transfer of virus-specific $CD8^+$ T cells was sufficient to restore CMV-protective immunity in HSCT recipients (Walter et al. 1995; Cobbold et al. 2005; Neuenhahn et al. 2017). The role of CMV-specific $CD4^+$ T cells in virus control is still unclear. Several studies indicate that the presence of $CD4^+$ T cells after transplantation contributes to protection from the virus, likely because they help maintaining an effective $CD8^+$ T cell response (Walter et al. 1995; Gerna et al. 2006).

1.3.1 Immune evasion strategies of CMV

Considering that CMV elicits such an enormous immune response, it seems surprising that the virus manages to persist after initial infection. During co-evolution with its human host, the virus has developed a multitude of immune evasion strategies to modulate and mitigate attacks by the immune system (Tortorella et al. 2000). Firstly, CMV is able to hide from the immune system by entering a latent state in which no infectious virus is produced and very few viral transcripts are present in the cell. This reduces immunogenicity of the virus to a minimum (Noriega et al. 2012). In addition, CMV has dedicated a substantial fraction of its genome to encode immunoevasins which interfere with immune control on multiple levels. For instance, CMV is thought to evade humoral immunity by upregulating host complement control proteins and incorporating them into the virion (Hengel et al. 1998; Alcami and Koszinowski 2000). Moreover, CMV evades innate immunity and inflammation. For this purpose, the viral genome encodes homologues of the immunoregulatory cytokine interleukin (IL)-10 and chemokine receptors (Smith and Khanna 2013).

Most importantly, CMV tampers with antigen presentation so that infected cells cannot be recognised by T cells. CMV encodes various immunoevasins to downregulate antigen presentation mediated by MHC class I. These immunoevasins interfere with many steps of antigen processing and presentation (Noriega et al. 2012). Some immunoevasins block the supply of antigenic peptide. They inhibit proteasomal antigen degradation that produces antigenic peptide, TAP-dependent transport of antigenic peptide to the ER, or trimming of antigenic peptide mediated by endoplasmic reticulum aminopeptidase 1 (ERAP1) in the ER. Other immunoevasins induce degradation of MHC class I or decelerate or stop translocation of the peptide-MHC complex to the

cell surface. Notably, CMV immunoevasins most strongly block antigen presentation of peptides presented on HLA-A and HLA-B, but less so on HLA-C (Schust et al. 1998). It remains controversial whether CMV also interferes with antigen presentation on MHC class II molecules. A study reported that the immunoevasive gene product of US2 lead to the degradation of certain MHC class II-encoded proteins (Tomazin et al. 1999). This finding was later refuted by Rehm et al. who showed that gene products of US2 and US11 only affected the surface expression of MHC class I molecules, whereas MHC class II molecules were spared from immunomodulation (Rehm et al. 2002).

CMV has also evolved mechanisms to block downstream signalling of the antiviral interferons IFNα and IFNβ, thereby tampering with activation of natural killer (NK) cells (García-Sastre and Biron 2006; Loewendorf and Benedict 2010). NK cells are traditionally considered part of the innate immune system, because they have invariant receptors to detect infected cells. They have both direct cytotoxic activity and secrete inflammatory cytokines, such as IFNγ, to promote the function of other immune cells (Murphy and Weaver 2016). NK cells express activating and inhibitory receptors to distinguish healthy from infected or otherwise abnormal body cells. Some inhibitory receptors bind to certain MHC class I molecules and their signal prevents NK-cell–mediated cytotoxicity. In case of low MHC class I expression, for instance due to reduction by viral immunoevasins, the inhibitory signal is abolished and the MHC class I–deficient cell is killed by NK cells. This mechanism of action was termed ”missing self” theory and was proposed several decades ago (Ljunggren and Kärre 1990). HLA-C molecules, as a group, are particularly important to protect cells from NK-cell lysis (Colonna et al. 1993), because for every HLA-C molecule there is a matched inhibitory NK receptor from the killer-cell immunoglobulin-like receptor (KIR)2DL group. Members of inhibitory KIR groups are expressed on a large subset of NK cells across donors with diverse ethnic backgrounds (Middleton et al. 2008). Every KIR haplotype contains such inhibitory receptors for any type of HLA-C molecule, and therefore HLA-C is the universal molecule for NK-cell inhibition, although expression levels and ability to induce self-tolerance can vary for different inhibitory KIR2DL and HLA-C alleles (Gardiner 2008; Charoudeh et al. 2012). CMV immunoevasins may spare HLA-C from downregulation to prevent NK cells from killing the host cell. In line with this hypothesis, recent work of our group showed that an IE1-derived T-cell epitope restricted by HLA-C*07:02 was

constitutively presented by CMV-infected cells and capable of inducing a specific cytotoxic T-cell response (Ameres et al. 2013). By contrast, other IE1 epitopes restricted by different HLA-A or HLA-B alleles were no longer presented and thus not recognised by cytotoxic T cells. Deletion of four immunoevasins restored surface expression of HLA-A and HLA-B alleles and T-cell mediated killing of CMV-infected target cells. CMV immunoevasins also modulate NK-cell activity by downregulation of ligands required for NK-cell activation (Loewendorf and Benedict 2010). In addition, CMV expresses an HLA class I homologue, the UL18 gene product (Beck and Barrell 1988), which prevents killing of CMV-infected cells by a subset of NK cells, whereas another subset is activated by UL18 (Prod'homme et al. 2007). However, UL18 is also a target of $CD8^+$ T cells and its expression leads to lysis of cells infected with CMV (Saverino et al. 2004). It has been shown that absence of NK cells leads to severe CMV disease, indicating that NK cells are protective against the virus (Biron et al. 1989).

1.3.2 Protectivity of specific T cells

Although the presence of a virus-specific T-cell response is associated with protection from CMV, it is unknown which T cells (of which antigen and epitope specificities) have protective effector functions, and which T cells are only innocent bystanders and do not contribute to virus control. The $CD8^+$ T-cell response appears to be particularly important for virus control (Quinnan et al. 1982; Cwynarski et al. 2001). Despite the large number of immunogenic ORFs, only few antigens elicit $CD8^+$ T-cell responses in a substantial number of donors. The most immunogenic antigens include immediate-early proteins IE1 and IE2, envelope glycoprotein gB, phosphoproteins pp65 and pp150, which both have late kinetics and are located in the virion tegument, and the UL29/28 gene product, which promotes accumulation of IE mRNA and regulates transcription factor p53 to prevent cellular stress responses (Sylwester et al. 2005; Terhune et al. 2010; Savaryn et al. 2013). Especially CMV antigens IE1 and pp65 give rise to many different epitopes and induce stong $CD8^+$ immune responses (Elkington et al. 2003; Khan 2007). CMV antigen pp65 has been shown to elicit a strong and diverse functional $CD8^+$ T-cell response in vitro and ex vivo (Manley et al. 2004; Lacey et al. 2005), and clinical responses were observed after transfer of pp65-specific T cells (Einsele et al. 2002; Neuenhahn et al. 2017). However, when T cells of different specificities (IE1,

pp65) and their association with viral reactivation was analysed in transplant patients, IE1-specific $CD8^+$ T cells appeared more protective than pp65-specific T cells (Bunde et al. 2005; Sacre et al. 2008; Gratama et al. 2008). Data from our group implies that not only the nature of the antigen, but also the restricting HLA determines functionality of the T-cell response. HLA-C–restricted epitopes may be superior to epitopes restricted by HLA-A or HLA-B in eliciting a protective $CD8^+$ T-cell response, because they are less affected by CMV immunoevasion and are thus continuously presented on infected cells (Ameres et al. 2013; Ameres et al. 2014); this enables clearance of the infected cells by $CD8^+$ T cells. Still, the identity of protective T cells is far from being understood, and further research is required to expand our knowledge of the functional roles of the different elements that constitute a CMV-specific T-cell response.

1.3.3 Factors contributing to diversity of the CMV-specific T-cell response

Understanding the overall human T-cell response to CMV is difficult, because it is greatly diverse. Factors contributing to this immense diversity are (i) CMV antigen diversity, (ii) CMV epitope diversity, (iii) HLA diversity, and (iv) TCR diversity. Once CMV infects a body cell, viral proteins are degraded to peptides, which are in turn presented to T cells on HLA molecules (Figure 1.5). CMV harbours the largest genome of all presently known human viruses. One group studied $CD4^+$ and $CD8^+$ T-cell responses to 213 known ORFs in CMV, and they identified T-cell responses against 151 of the encoded CMV proteins (Sylwester et al. 2005). Each of the many CMV antigens is degraded and can potentially give rise to multiple different immunogenic peptides that are presented on HLA molecules. HLA molecules are highly polymorphic and many different alleles exist in the human gene pool. Three classical HLA class I heavy chains are encoded by the major genes HLA-A, -B, and -C. There are also three major gene loci encoding HLA class II α-chains and β-chains; they are HLA-DP, -DQ, and -DR. The combination of HLA alleles found on a single chromosome are called an HLA haplotype. Due to the close vicinity of the HLA-B and HLA-C loci, there is a particularly strong linkage disequilibrium between alleles of these two HLA genes. The frequency distribution of HLA alleles differs between ethnic groups (Cao et al. 2001; Schmidt et al. 2009; Gragert et al. 2013). For instance, HLA-A*02:01 is the most frequent HLA-A allele in people of European descent (allele frequency 27.6%), but less common in

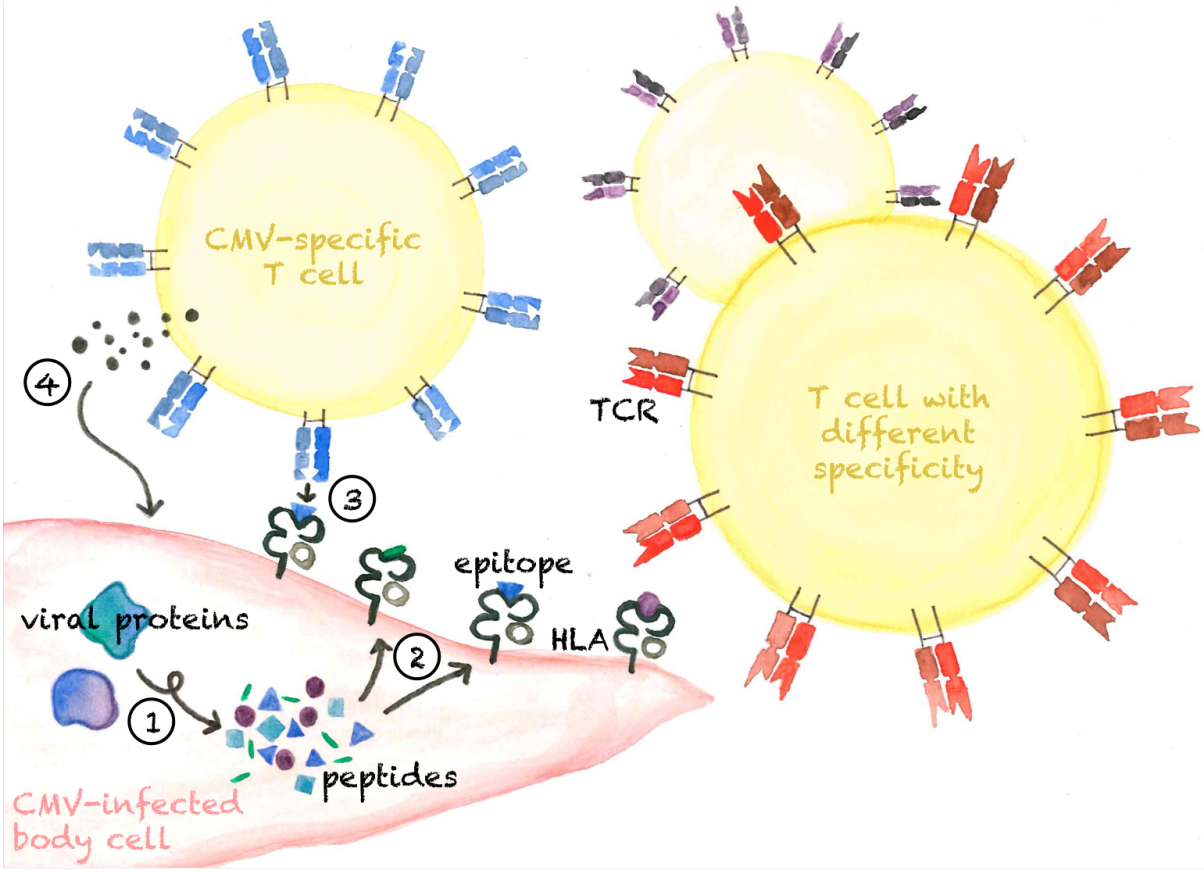


Figure 1.5: The CMV-specific $CD8^+$ T-cell response. After CMV infects a human cell, viral proteins are degraded to peptides (1). Some of these peptides are loaded onto matching HLA class I molecules and transported to the cell surface (2). T cells can recognise immunogenic CMV peptides in complex with their restricting HLA. The part of the peptide-MHC complex that is bound by the TCR is referred to as the T-cell epitope. CMV-specific T cells bind to matching epitopes presented by infected cells (3) and are activated. Activated $CD8^+$ T cells secrete cytotoxins (4), which in turn induce apoptosis in the infected cell.

Southeast Asians (5.8%); the latter most frequently express HLA-A*11:01 (17.5%) and HLA-A*24:02 (13.8%) (Gragert et al. 2013). Furthermore, the individual HLA alleles are joined to different haplotypes in different ethnic groups (Cao et al. 2001). For example, HLA-C*07:02 is the most frequent HLA-C allele in Southeast Asians (12.9%) and the second most frequent HLA-C allele in Europeans (14.1%) (Gragert et al. 2013); in Germans, HLA-C*07:02 is even the most frequent HLA-C allele (15.2%) (Schmidt et al. 2009). However, HLA-C*07:02 is most commonly associated with HLA-A*26:01 and HLA-B*08:01 in Southeast Asians, whereas it is most commonly associated with HLA-A*03:01 and HLA-B*07:02 in Europeans (Gragert et al. 2013). Different HLAs bind to and present different antigenic peptides. Hence, two people expressing dif-

ferent HLA alleles will present different CMV epitopes and therefore mobilise T cells with different TCRs. This makes it difficult to get an overall picture of the human T-cell respone against CMV. Lastly, the TCR comes in various different compositions. It is estimated that between 10^{15} and 10^{20} distinct TCR clonotypes can theoretically be produced by V(D)J recombination (Laydon et al. 2015). By definition, a TCR clonotype is the entirety of T cells expressing the same TCR. However, the estimated total number of T cells in the human body is only 10^{13} (Bianconi et al. 2013), and some antigen-experienced TCR clonotypes are highly abundant. Consequently, the actual number of different TCR clonotypes in a person is much lower than the theoretically possible number of different TCR clonotypes. A study by theoretical biologists estimated the amount of distinct TCR clonotypes per person to be in the range of 10^{10}, which was one order of magnitude lower than the estimated number of naïve T cells in the human body (Lythe et al. 2016). Still, there are countless different TCR clonotypes among which the CMV-specific ones need to be identified.

1.4 Studies of the CMV-specific TCR repertoire

Several studies have previously investigated the magnitude and diversity of the CMV-specific T-cell response against epitopes, antigens, and full-length virus using phenotypical peptide-MHC multimer staining or functional immunoassays (Gillespie et al. 2000; Hosie et al. 2017; Elkington et al. 2003; Jackson et al. 2014; Sylwester et al. 2005), but the T-cell response to CMV is complex and many questions remain unanswered. In order to fully understand the CMV-specific human T-cell response, it is inevitable to go into more detail and also assess the composition, frequency, and binding properties of the TCRs that bind CMV-derived epitopes. Although some virus-specific TCRs have been described to date, our knowledge of the CMV-directed TCR repertoire is still very limited. Early studies have identified some CMV-specific TCRβ chains targeting pp65-derived epitopes, sometimes in combination with matching TCRα chains. For instance, a group sequenced TCRα and TCRβ chains of epitope-specific T-cell clones that were generated by stimulation with short immunogenic CMV peptides in order to investigate the clonal composition of the pp65-specific $CD8^+$ T-cell response (Weekes et al. 1999). Both TCR chains of the clones were amplified in multiple separate PCR reactions using a panel of V region-specific primers followed by Sanger se-

quencing. In doing so, the authors discovered clonally expanded CMV peptide-specific $CD8^+$ TCR clonotypes in all tested virus carriers. TCRβ sequences specific for pp65 epitopes were also identified by other groups, either from ex vivo multimer-enriched populations or from short-term peptide stimulated and subsequently multimer-enriched cells (Price et al. 2005; Venturi et al. 2008; Schwanninger et al. 2008; Miconnet et al. 2011; Brennan et al. 2012; Koning et al. 2013). In these studies, bulk RNA was isolated, amplified, and transformed into competent *Escherichia coli* cells, from which selected colonies were Sanger-sequenced.

The advent of high-throughput immunosequencing a decade ago (Robins et al. 2009) enabled the simultaneous readout of thousands of $\alpha\beta$ TCRs from single cells or hundreds of millions of TCRα or TCRβ chains from bulk cell samples. This led to a massive increase in sequencing depth, since not only the most expanded specific TCR clonotypess, but also less expanded ones, can be found, and large TCR data sets have become available. Several groups have already exploited high-throughput sequencing to analyse CMV-specific TCR repertoires. In multiple studies, CMV-specific T cells were enriched by sorting with pp65-derived peptide-MHC multimers, or, less frequently, IE1-derived or pp50-derived multimers (Klarenbeek et al. 2012; Dash et al. 2017; Glanville et al. 2017). Multimer-enriched cells were high-throughput–sequenced either from bulk cells to obtain high-resolution, but separate, TCRα and/or TCRβ repertoires, or from single cells for paired TCR$\alpha\beta$ data at a lower resolution. These studies identified a number of CMV multimer-specific TCR sequences and found that the specific TCRβ repertoires were stably maintained after primary infection (Klarenbeek et al. 2012) and that TCR repertoires of the same specificity frequently used common CDR3 amino acid sequence motifs (Dash et al. 2017; Glanville et al. 2017). An additional study combined short-term peptide stimulation with multimer sorting and found that the TCR repertoire specific for a CMV epitope was diverse, but contained several short CDR3 sequence motifs (Chen et al. 2017). One group sequenced PBMCs that were short-term–stimulated with an overlapping peptide pool of CMV late protein pp65 and used a frequency cutoff of $>$1% of total TCRβ reads to identify antigen-specific TCRβ sequences in their samples (Dziubianau et al. 2013). They showed that pp65-specific TCRβ repertoires were oligoclonal and largely stable over time. Except for the last study, which did not investigate the precise epitope specificity and HLA restrictions

of their specific TCRβ sequences, all previous studies only analysed TCR repertoires specific for CMV epitopes restricted by HLA-A and HLA-B alles. The findings presented in this PhD thesis expand on these previous results by exploring TCRβ repertoires not only against CMV epitopes restricted by HLA-A and HLA-B alles, but also include HLA-C–restricted CMV epitopes.

1.5 Aims of this PhD project

Previous studies that investigated the CMV-specific TCR repertoire greatly relied on multimer sorting to isolate virus-specific T cells. However, multimers can be difficult to produce, are quite expensive, and they are only available for selected CMV-derived epitopes. In this project, the CMV-specific TCR repertoire is explored by combining in vitro stimulation with short immunogenic peptides with high-throughput Illumina sequencing. Short peptides are comparatively inexpensive and easy to produce. They can be externally loaded onto HLA molecules, where they are presented to T cells. With the peptide stimulation approach, TCR clonotypes specific for any CMV-derived peptide matching the donors HLA can be identified, even if no multimer is available. Such an assay will be simple and ready to be scaled up to test different CMV peptide specificities simultaneously. Apart from investigation of TCR repertoires against HLA-A–restricted and HLA-B–restricted CMV epitopes, which have been studied before to different levels of detail, the CMV-specific HLA-C–restricted TCR repertoire is investigated here for the first time. The TCR repertoires of an unstimulated sample and sample(s) stimulated with control peptide(s) were sequenced in addition to the peptide-stimulated sample (multi-sample comparison) with the aim to distinguish CMV peptide-specific TCRβ clonotypes from those with different specificities with precision and efficacy.

The CMV-specific TCR repertoire data obtained by this technique were to be analysed with regard to the following questions:

1. Is it possible to obtain a measure of the frequency of epitope-specific T cells directly from TCRβ sequencing? How well does this measure correlate with the frequency of specific T cells determined by peptide-MHC multimer staining?

2. To what extent is TCRβ chain sequencing sufficient to characterise the CMV epitope-specific T-cell repertoire? Is knowledge of the pairing TCRα chain indispensable to gain meaningful insights in the CMV-specific T-cell response?

3. How frequent are circulating CMV-specific TCRs against selected epitopes in virus carriers? What portion of T cells and expressed TCRs is estimated to be specific for CMV in total?

4. Are there differences in magnitude or clonality of CMV-specific TCR repertoires between epitopes derived from different antigens or presented by different HLAs?

5. How similar is the interindividual and intraindividual TCR repertoire against particular CMV epitopes? Why do different donors mobilise different TCRs against the same epitopes?

6. Are there common features or patterns between TCRs of the same specificity, such as usage of particular gene segments or preferred CDR3 lengths? Can amino acid sequence motifs be identified that are indicative the epitope specificity and may be used to predict the specificity of a given TCR?

7. Are there CMV epitope-specific TCRs that are frequently shared between donors? Can we identify a minimum set of shared specific TCRs against selected CMV epitopes that can be used to derive a predictive TCR signature that can, for instance, be used in CMV disease monitoring? This question is of great clinical interest because extensively shared CMV-specific TCRs also hold great potential for adoptive T-cell therapy, since they are tolerant to many HLA backgrounds and therefore less likely to cause graft-versus-host disease.

2 MATERIALS

2.1 Laboratory equipment and devices

Table 2.1: Laboratory equipment and devices.

Item	Model	Manufacturer
Autom. electrophoresis system	2100 Bioanalyzer	Agilent Technologies, Santa Clara, US
Beaker	Polypropylene, 1l	Brand, Wertheim, DE
Centrifuge	Rotana 46 RSC	Hettich, Tuttlingen, DE
Centrifuge	Rotana 460 R	Hettich, Tuttlingen, DE
Electrophoresis power supply	E835	Consort, Turnhout, BE
Electrophoresis system	PerfectBlue Mini S	PEQLAB Biotechnologie, Erlangen, DE
Electrophoresis system	PerfectBlue Mini L	PEQLAB Biotechnologie, Erlangen, DE
Flow cytometer	BD LSRFortessa	BD Biosciences, Heidelberg, DE
Freezer, -20°C	KG39NVW30	Siemens, Berlin & Munich, DE
Freezer, -80°C	Igloo 830l	Telstar, Terrassa, ES
Fridge, 4°C	Bosch economic	Bosch, Gerlingen-Schillerhöhe, DE
Gel documentation system	Quantum ST4	Vilber Lourmat, Eberhardzell, DE
Glass bottles	Duran 50 ml–1 l	SCHOTT, Mainz, DE
Ice machine	AF206	Scotsman, Vernon Hills, US
Incubator	CB150	Binder, Tuttlingen, DE
Irradiation device (Cs137)	Gammacell 40	Best Theratronics, Ottawa, CA
Laminar flow hood	S 1800	BDK, Sonnenbühl-Genkingen, DE
Liquid nitrogen tank	CS 540 B	Cryo Anlagenbau, Wilnsdorf, DE
Magnetic separator	MidiMACS Separator	Miltenyi, Bergisch Gladbach, DE
Magnetic separator	OctoMACS Separator	Miltenyi, Bergisch Gladbach, DE
Measuring cylinder	Polypropylene, 1l	Kartell LABWARE, Noviglio, IT
Microcentrifuge	Centrifuge 5415 R	Eppendorf, Hamburg, DE
Microcentrifuge	Pico 21	Heraeus, Hanau, DE
Microscope	Axiovert 25	Zeiss, Jena, DE
Microwave	R-202	Sharp, Sakai, JP
Mini centrifuge	SPROUT	Diversified Biotech, Dedham, US
Multichannel pipette, 20–200 µl	Transferpette-12	Eppendorf, Hamburg, DE
PCR Thermal Cycler	TGradient	Biometra, Göttingen, DE
Pipet controller	Pipetboy acu 2	Integra Biosciences, Hudson, US
Pipette 2 µl	Pipetman G P2G	Gilson, Middleton, US
Pipette 10 µl	Pipetman G P10G	Gilson, Middleton, US
Pipette 20 µl	Pipetman G P20G	Gilson, Middleton, US
Pipette 200 µl	Pipetman G P200G	Gilson, Middleton, US
Pipette 1000 µl	Pipetman G P1000G	Gilson, Middleton, US
Plate reader	EL800	Bio-Tek Instruments, Winooski, US
Roller mixer	CAT RM5-30V	neoLab, Heidelberg, DE
Thermomixer	Thermostat plus	Eppendorf, Hamburg, DE
UV-Vis spectrophotometer	Nanodrop ND-1000	PEQLAB Biotechnologie, Erlangen, DE
Vacuum pump	N86KN.18	KNF Neuberger, Freiburg, DE
Vortex mixer	7-2020	neoLab, Heidelberg, DE
Vortex mixer	Vortex-Genie 2	Scientific Industries, New York, US
Water bath	1003	GFL, Burgwedel, DE

2.2 Consumables

Table 2.2: Consumables.

Item	Manufacturer
Butterfly needle, Vacutainer safety-lok green	BD Biosciences, Heidelberg, DE
Cell culture flask, Falcon 25cm^2, vented cap	Corning, New York, US
Cell culture flask, EasYFlask 75cm^2, filter cap	Nunc, Roskilde, DK
Cell strainer, Falcon 100 μm	Corning, New York, US
Conical centrifuge tube, Falcon 15 ml	Corning, New York, US
Conical centrifuge tube, Falcon 50 ml	Corning, New York, US
Cryotube, 1.8 ml	Nunc, Roskilde, DK
Disposable Pasteur pipette, 3.2 ml	Carl Roth, Karlsruhe, DE
FACS Test tube, Falcon 5 ml, round-bottom	Corning, New York, US
Gloves, NextGen Nitrile S	Meditrade, Kiefersfelden, DE
Gloves, Kimtech Purple Nitrile S	Kimberly-Clark Professional, Roswell, US
MACS column, LS	Miltenyi, Bergisch Gladbach, DE
MACS column, MS	Miltenyi, Bergisch Gladbach, DE
Multiwell plate, 6 wells, flat bottom	Corning, New York, US
Multiwell plate, 12 wells, flat bottom	Corning, New York, US
Multiwell plate, 24 wells, flat bottom	Corning, New York, US
Multiwell plate, 48 wells, flat bottom	Corning, New York, US
Multiwell plate, 96 wells, flat bottom	Corning, New York, US
Multiwell plate, 96 wells, V-bottom	Nunc, Roskilde, DK
Hemocytometer, C-Chip	NanoEnTek, Seoul, KR
PCR strip, 200 μl, 8 tubes	Brand, Wertheim, DE
Pipette tips, 10 μl, Diamond D10	Gilson, Middleton, US
Pipette tips, 200 μl, Diamond D200	Gilson, Middleton, US
Pipette tips, 1000 μl, Diamond D1000	Gilson, Middleton, US
Reaction tubes, 1.5 ml	Eppendorf, Hamburg, DE
Reaction tubes, 2 ml	Eppendorf, Hamburg, DE
Serological pipette, 5 ml	Greiner Bio-One, Kremsmünster, AT
Serological pipette, 10 ml	Greiner Bio-One, Kremsmünster, AT
Serological pipette, 25 ml	Greiner Bio-One, Kremsmünster, AT
Serum tube, Primavette S 10 ml	KABE Labortechnik, Nümbrecht, DE
Syringe, INFUJECT 50/60 ml	Dispomed Witt, Gelnhausen, DE
Syringe filter, Minisart $\varnothing$0.2 μm	Sartorius, Göttingen, DE

2.3 Commercial kits

Table 2.3: Commercial kits.

Name	Manufacturer
DNA 1000 Kit	Agilent Technologies, Santa Clara, US
NEBNext Multiplex Oligos for Illumina (Dual Index Primers Set 1)	New England Biolabs, Ipswich, US
Human Anti-Cytomegalovirus IgG ELISA Kit	abcam, Cambridge, UK
Human IFN-γ ELISA development kit (ALP)	Mabtech AB, Nacka Strand, SE
QuantiTect Reverse Transcription Kit	Qiagen, Hilden, DE
RNeasy Mini Kit	Qiagen, Hilden, DE

2.4 Chemicals and buffers

Table 2.4: Chemicals and buffers.

Name	Manufacturer
Agarose	Biozym, Hessisch Oldendorf, DE
Agencourt AMPure XP Beads	Beckman Coulter, Brea, US
Aqua ad iniectabilia, 50 ml	B Braun, Melsungen, DE
Dimethylsulfoxide (DMSO)	Carl Roth, Karlsruhe, DE
dNTP mix	ThermoFisher scientific, Waltham, US
Ethanol, ROTIPURAN 2.5l	Carl Roth, Karlsruhe, DE
Ethidium bromide, 0.07%, 15 ml dropper bottle	AppliChem, Darmstadt, DE
Ficoll, Pancoll 500 ml	PAN-Biotech, Aidenbach, DE
Formaldehyde, 37% solution	Carl Roth, Karlsruhe, DE
GeneRuler DNA ladder, 100 bp	ThermoFisher scientific, Waltham, US
Heparin, Heparin-Natrium-25000	ratiopharm, Ulm, DE
DNA Gel Loading Dye (6×)	ThermoFisher scientific, Waltham, US
$MgSO_4$	ThermoFisher scientific, Waltham, US
NaOH, pellets	Carl Roth, Karlsruhe, DE
Dulbecco's PBS, 500 ml	Sigma-Aldrich, St. Louis, US
RNase-free H_2O	Qiagen, Hilden, DE
TAE buffer, Rotiphorese 50x	Carl Roth, Karlsruhe, DE
TRIS hydrochloride	Carl Roth, Karlsruhe, DE
Trypan blue	Merck, Darmstadt, DE

2.5 Oligonucleotides

2.5.1 Multiplex PCR primers for TCRα library preparation

Table 2.5: List of multiplex PCR primers for TCRα libraries that were designed in this thesis. 51 forward primers and 2 reverse primer candicates were ordered from metabion (Planegg, DE). The non-pairing primer overhangs are equivalent to the Illumina Read 1 and Illumina Read 2 primer sites. In this table, "ACACTCTTTCCCTACACGACGCTCTTCCGATCT" (Illumina Read 1 primer sequence) is abbreviated to "ACA[...]TCT" and "5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3'" (Illumina Read 2 primer sequence) is abbreviated to "GTG[...]TCT". The target-binding portions of the primers are highlighted in bold font.

Name	Sequence (5'→ 3')
multiTRAV1-1	ACA[...]TCT**CGCTCTGATAGTTATGGTTACCTCCTTCTAC**
multiTRAV1-2	ACA[...]TCT**TCTAAAGGGTACAGTTACCTCCTTTTGAAGG**
multiTRAV2	ACA[...]TCT**GTTCTCTTCATCGCTGCTCATCCTCC**
multiTRAV3	ACA[...]TCT**CAAACCTCCTTCCACCTGAAGAAACCATC**
multiTRAV4	ACA[...]TCT**AGTCCAGCACTCTGAGCCTGCC**
multiTRAV5	ACA[...]TCT**GGATAAACATCTGTCTCTGCGCATTGCAG**
multiTRAV6	ACA[...]TCT**CCCTTAAACAGAGTTTGTTTCATATCACAGCC**
multiTRAV6*04	ACA[...]TCT**CCCTTAAACAGAGTTTGTTTCATGTCACAGC**
multiTRAV7	ACA[...]TCT**TGGAAGCAGCTTGTACATTACAGCCG**
multiTRAV8-1	ACA[...]TCT**AATTCTCCTTTAATCTGAGGAAACCCTCTG**
multiTRAV8-2+4	ACA[...]TCT**CCTCCTTCCACCTGACGAAACCCTC**

Table 2.5: (continued)

Name	Sequence (5'→ 3')
multiTRAV8-3	ACA[...]TCT**TCAATCTTCCTTCAATCTGAGGAAACCCTC**
multiTRAV8-3*02	ACA[...]TCT**CAATCTTCCTTCAACCTGAGGAAACCCTC**
multiTRAV8-4*02	ACA[...]TCT**CCTCCTTCCACCTGACAAAACCCTC**
multiTRAV8-4*06	ACA[...]TCT**TCCTTCCACCTGACGAAACCCG**
multiTRAV8-6	ACA[...]TCT**ACTTCCTTCCACTTGAGGAAACCCTC**
multiTRAV8-7	ACA[...]TCT**GAAACCTCCTTCTACCTGAGGAAACCATC**
multiTRAV9-1	ACA[...]TCT**GAAACCACTTCTTTCCACTTGGAGAAAGAC**
multiTRAV9-2	ACA[...]TCT**CCACTTCTTTCCACTTGGAGAAAGGC**
multiTRAV10	ACA[...]TCT**AGCAAAGCTCTCTGCACATCACAGC**
multiTRAV12-1	ACA[...]TCT**GCCAGTATATTTCCCTGCTCATCAGAGAC**
multiTRAV12-2	ACA[...]TCT**CCAGCCAGTATGTTTCTCTGCTCATCAG**
multiTRAV12-3	ACA[...]TCT**TCCAGCAAGTATATCTCCTTGTTCATCAGAG**
multiTRAV13-1	ACA[...]TCT**CCAAACATTTCTCCCTGCACATCACAG**
multiTRAV13-2	ACA[...]TCT**AGACAGTGAAACATCTCTCTCTGCAAATTG**
multiTRAV14/TRDV4	ACA[...]TCT**CGCCAACCTTGTCATCTCCGC**
multiTRAV16	ACA[...]TCT**GCGAGACATCTTTCCACCTGAAGAAACC**
multiTRAV17	ACA[...]TCT**GAAAAGCAGTTCCTTGTTGATCACGGC**
multiTRAV18	ACA[...]TCT**GTTCCTTCCACCTGGAGAAGCCCTC**
multiTRAV19	ACA[...]TCT**GTTCCTTCAACTTCACCATCACAGCC**
multiTRAV20	ACA[...]TCT**AAGCTTTCTGCACATCACAGCCC**
multiTRAV21	ACA[...]TCT**CATCAGGACGTAGTACTTTATACATTGCAGC**
multiTRAV22	ACA[...]TCT**CGGAACGCTACAGCTTATTGTACATTTCCTC**
multiTRAV23/TRDV6	ACA[...]TCT**GCCAAGCAGTTCTCATTGCATATCATGG**
multiTRAV24	ACA[...]TCT**GGAGGGTTACAGCTATTTGTACATCAAAGG**
multiTRAV25+41	ACA[...]TCT**ACAGCTCCCTGCACATCACAGC**
multiTRAV26-1	ACA[...]TCT**AGTCCAGCACCTTGATCCTGCC**
multiTRAV26-2	ACA[...]TCT**AGTCCAGTACCTTGATCCTGCACCG**
multiTRAV27	ACA[...]TCT**AAAGGACAGTTCTCTCCACATCACTGC**
multiTRAV29/TRDV5	ACA[...]TCT**GCCAAGCACCTCTCTCTGCACATTG**
multiTRAV29/TRDV5*02	ACA[...]TCT**GCCAAGCACCTCTCTCTCGACATTG**
multiTRAV30	ACA[...]TCT**GCAAAGCTCCCTGTACCTTACGGC**
multiTRAV34	ACA[...]TCT**CAGCAAAGTTCCCTGCATATCACAGC**
multiTRAV35	ACA[...]TCT**AGAAAGGACAGCTTCCTGAATATCTCAGC**
multiTRAV36/TRDV7	ACA[...]TCT**CAGCATCCTGAACATCACAGCCAC**
multiTRAV38-1+2/TRDV8	ACA[...]TCT**GCAGCCAAATCCTTCAGTCTCAAGATCTC**
multiTRAV39	ACA[...]TCT**TCTCAGCACCCTCCACATCACAGC**
multiTRAV40	ACA[...]TCT**GACAAAAACTCCCCCATTGTGAAATATTCAG**
multiTRDV1	ACA[...]TCT**AATCCGTCGCCTTAACCATTTCAGCC**
multiTRDV2	ACA[...]TCT**AGAACCTGGCTGTACTTAAGATACTTGCAC**
multiTRDV3	ACA[...]TCT**AAGCCTTTCACTTGGTGATCTCTCCAG**
CαA	GTG[...]TCT**CGGCAGGGTCAGGGTTCTGG**
CαB	GTG[...]TCT**GGATTTAGAGTCTCTCAGCTGGTACACG**

2.5.2 Multiplex PCR primers for TCRβ library preparation

Table 2.6: Multiplex PCR primers for TCRβ libraries. 45 forward primers and 3 reverse primers were ordered from metabion (Planegg, DE). All target-binding regions of the forward primers, except for ss-BV15, were published by Robins et al. (Robins et al. 2009). Forward primer ss-BV-15 and the reverse primers were designed by Xiaoling Liang. In this table, "ACACTCTTTCCCTACACGACGCTCTTCCGATCT" (Illumina Read 1 primer sequence) is abbreviated to "ACA[...]TCT" and "5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-3'" (Illumina Read 2 primer sequence) is abbreviated to "GTG[...]TCT". The binding portions of the primers are printed in bold font.

Name	Sequence (5'→ 3')
SS-BV2	GTG[...]TCT**TCAAATTTCACTCTGAAGATCCGGTCCACAA**
SS-BV3-1	GTG[...]TCT**GCTCACTTAAATCTTCACATCAATTCCCTGG**
SS-BV4-1	GTG[...]TCT**CTTAAACCTTCACCTACACGCCCTGC**
SS-BV(4-2, 4-3)	GTG[...]TCT**CTTATTCCTTCACCTACACACCCTGC**
SS-BV5-1	GTG[...]TCT**GCTCTGAGATGAATGTGAGCACCTTG**
SS-BV5-3	GTG[...]TCT**GCTCTGAGATGAATGTGAGTGCCTTG**
SS-BV(5-4, 5-5, 5-6, 5-7, 5-8)	GTG[...]TCT**GCTCTGAGCTGAATGTGAACGCCTTG**
SS-BV6-1	GTG[...]TCT**TCGCTCAGGCTGGAGTCGGCTG**
SS-BV(6-2, 6-3)	GTG[...]TCT**GCTGGGGTTGGAGTCGGCTG**
SS-BV6-4	GTG[...]TCT**CCCTCACGTTGGCGTCTGCTG**
SS-BV6-6	GTG[...]TCT**CGCTCAGGCTGGAGTTGGCTG**
SS-BV6-5	GTG[...]TCT**GCTCAGGCTGCTGTCGGCTG**
SS-BV6-7	GTG[...]TCT**CCCCTCAAGCTGGAGTCAGCTG**
SS-BV6-8	GTG[...]TCT**CACTCAGGCTGGTGTCGGCTG**
SS-BV6-9	GTG[...]TCT**CGCTCAGGCTGGAGTCAGCTG**
SS-BV7-1	GTG[...]TCT**CCACTCTGAAGTTCCAGCGCACAC**
SS-BV7-2	GTG[...]TCT**CACTCTGACGATCCAGCGCACAC**
SS-BV7-3	GTG[...]TCT**CTCTACTCTGAAGATCCAGCGCACAG**
SS-BV7-4	GTG[...]TCT**CCACTCTGAAGATCCAGCGCACAG**
SS-BV7-6	GTG[...]TCT**CACTCTGACGATCCAGCGCACAG**
SS-BV7-7	GTG[...]TCT**CCACTCTGACGATTCAGCGCACAG**
SS-BV7-8	GTG[...]TCT**CCACTCTGAAGATCCAGCGCACAC**
SS-BV7-9	GTG[...]TCT**CACCTTGGAGATCCAGCGCACAG**
SS-BV9	GTG[...]TCT**GCACTCTGAACTAAACCTGAGCTCTCTG**
SS-BV10-1	GTG[...]TCT**CCCCTCACTCTGGAGTCTGCTG**
SS-BV10-2	GTG[...]TCT**CCCCCTCACTCTGGAGTCAGCTA**
SS-BV10-3	GTG[...]TCT**CCTCCTCACTCTGGAGTCCGCTA**
SS-BV(11-1, 11-3)	GTG[...]TCT**CCACTCTCAAGATCCAGCCTGCAG**
SS-BV11-2	GTG[...]TCT**CTCCACTCTCAAGATCCAGCCTGCAA**
SS-BV(12-3, 12-4, 12-5)	GTG[...]TCT**CCACTCTGAAGATCCAGCCCTCAG**
SS-BV13	GTG[...]TCT**CATTCTGAACTGAACATGAGCTCCTTGG**
SS-BV14	GTG[...]TCT**CTACTCTGAAGGTGCAGCCTGCAG**
SS-BV15	GTG[...]TCT**CTGCTTTCTTGACATCCGCTCACCAG**
SS-BV16	GTG[...]TCT**CTGTAGCCTTGAGATCCAGGCTACGA**
SS-BV17	GTG[...]TCT**CTTCCACGCTGAAGATCCATCCCG**
SS-BV18	GTG[...]TCT**GCATCCTGAGGATCCAGCAGGTAG**
SS-BV19	GTG[...]TCT**CCTCTCACTGTGACATCGGCCC**
SS-BV20-1	GTG[...]TCT**CTTGTCCACTCTGACAGTGACCAGTG**
SS-BV23-1	GTG[...]TCT**CAGCCTGGCAATCCTGTCCTCAG**
SS-BV24-1	GTG[...]TCT**CTCCCTGTCCCTAGAGTCTGCCAT**

Table 2.6: (continued)

Name	Sequence (5'→ 3')
SS-BV25-1	GTG[...]TCT**CCCTGACCCTGGAGTCTGCCA**
SS-BV27	GTG[...]TCT**CCCTGATCCTGGAGTCGCCCA**
SS-BV28	GTG[...]TCT**CTCCCTGATTCTGGAGTCCGCCA**
SS-BV29-1	GTG[...]TCT**CTAACATTCTCAACTCTGACTGTGAGCAACA**
SS-BV30	GTG[...]TCT**CGGCAGTTCATCCTGAGTTCTAAGAAGC**
R-seq-BC18	TACA[...]TCT**CACAGCGACCTCGGGTGGGA**
R-seq-N1-BC18	TACA[...]TCT**NCACAGCGACCTCGGGTGGGA**
R-seq-N2-BC18	TACA[...]TCT**NNCACAGCGACCTCGGGTGGGA**

2.5.3 Miscellaneous primers

Table 2.7: Miscellaneous primers. These primers are required for reverse transcription of TCR mRNA or for quality control of mini-LCLs.

Name	Sequence (5'→ 3')	Use
CP1*	GCACCTCCTTCCCATTCAC	TCRβ reverse transcription primer
RT TRAC A	CGGTGAATAGGCAGACAGAC	TCRα reverse transcription primer candidate
RT TRAC B	GCTCTTGAAGTCCATAGACCTC	TCRα reverse transcription primer candidate
Nextera1	AATGATACGGCGACCACCGA	Rescue PCR primer
Nextera2	CAAGCAGAAGACGGCATACGA	
cam-up†	TTCTGCCGACATGGAAGCCATC	Target chloramphenicol acetyltransferase as a
cam-down†	GGAGTGAATACCACGACGATTTCC	marker for mini-EBV–specific sequences
gp85c†	TGGTCAGCAGCAGATAGTGAACG	Target the BXLF1 gene, which encodes
gp85d†	TGTGGATGGGTTTCTTGGGC	glycoprotein 65, as marker for wild-type EBV

* Published in Zhou et al. 2006
† Published in Moosmann et al. 2002

2.6 Enzymes

All enzymes were delivered with their appropriate 10$\times$ standard buffer.

Table 2.8: Enzymes.

Name	Manufacturer
Taq DNA polymerase, 5U/μl,	New England Biolabs, Ipswich, US
Phusion High-Fidelity DNA Polymerase, 2U/μl	ThermoFisher scientific, Waltham, US

2.7 Antibodies and multimers

Throughout this project, anti-human antibody conjugates were used to phenotypically analyse human cells by flow cytometry or enrich cells by magnetic-activated cell sorting (MACS). In addition, multimers, which consist of multiple peptide-MHC complexes linked to a fluorophore, were used to visualise epitope-specific T cells.

2.7.1 Fluorescence-activated cell sorting (FACS) antibodies

Over the course of this thesis, the antibody panels used to stain human PBMCs for flow cytometry were optimised. All antibodies and their combinations in the different panels are listed in Table 2.9. The compensation control for PE, which was used to tag multimers, was anti-CD4-PE. FITC-labelled and BV605-labelled KIR2DL2/3 antibodies were compensated with antibodies against CD4 and CD19, respectively, because KIR2DL2/3-positive cells are not frequent enough to get a large positive population.

Table 2.9: Flow cytometry antibodies. All antibodies are monoclonal mouse anti-human antibodies. The antibody targets are named according to the cluster of differentiation (CD) nomenclature. In this thesis, however, CD158b will be referred to as KIR2DL2/3, and CD197 will be referred to as CCR7.

Target	Clone	Isotype	Fluorophore	Panel	Manufacturer
CD3	HIT3a	IgG2a,κ	Alexa700	I,III,IV	Biolegend, San Diego, US
CD3	HIT3a	IgG2a,κ	APC/Cy7	II	Biolegend, San Diego, US
CD4	RPA-T4	IgG1,κ	APC	I,IV	Biolegend, San Diego, US
CD4	RPA-T4	IgG1,κ	APC/Cy7	III	Biolegend, San Diego, US
CD4	RPA-T4	IgG1,κ	FITC	I*	Biolegend, San Diego, US
CD4	RPA-T4	IgG1,κ	PE	I*,II,III*,IV*	Biolegend, San Diego, US
CD8	RPA-T8	IgG1,κ	Pacific Blue	I,II,III,IV	Biolegend, San Diego, US
CD19	HIB19	IgG1,κ	BV605	III*	Biolegend, San Diego, US
CD45RA	HI100	IgG2b,κ	APC	II,III	Biolegend, San Diego, US
CD62L	DREG-56	IgG1,κ	Alexa488	III†	Biolegend, San Diego, US
CD158b‡	CH-L	IgG2b,κ	BV605	III	BD Biosciences, Heidelberg, DE
CD158b‡	CH-L	IgG2b,κ	FITC	I,II,IV	BD Biosciences, Heidelberg, DE
CD197§	G043H7	IgG2a,κ	Alexa488	III†	Biolegend, San Diego, US
CD197§	G043H7	IgG2a,κ	PE/Cy7	II	Biolegend, San Diego, US

* Compensation control
† Either anti-CD197 or anti-CD62L was used as differentiation marker in panel III
‡ Here referred to as KIR2DL2/3
§ Here referred to as CCR7

2.7.2 Magnetic-activated cell sorting (MACS) microbeads

Table 2.10: Magnetic-activated cell sorting microbeads.

Target	Isotype	Product name	Manufacturer
CD4	mouse IgG1	CD4 MicroBeads	Miltenyi, Bergisch Gladbach, DE
CD8	mouse IgG2a	CD8 MicroBeads	Miltenyi, Bergisch Gladbach, DE

2.7.3 Multimers

Table 2.11: Pentamers and streptamers for staining of epitope-specific T cells in flow cytometry. Unlabelled pentamers were purchased from ProImmune ("Pro5 MHC Class I Pentamers" series, ProImmune, Oxford, UK). Peptide/HLA-C*07:02 monomers for streptamer assembly were synthesised and kindly provided by Fabian Schlott, Michael Neuenhahn, and Dirk Busch (Technical University Munich, Munich, DE).

Peptide	HLA molecule	Type	Antigen	Virus
CRVLCCYVL	HLA-C*07:02	monomer	IE1	CMV
CRVLCCYIL	HLA-C*07:02	monomer	IE1	CMV
ELKRKMIYM	HLA-B*08:01	pentamer	IE1	CMV
ELRRKMMYM	HLA-B*08:01	pentamer	IE1	CMV
FRCPRRFCF	HLA-C*07:02	monomer	UL29/28-encoded protein	CMV
HERNGFTVL	HLA-B*40:01	pentamer	pp65	CMV
NLVPMVATV	HLA-A*02:01	pentamer	pp65	CMV
QIKVRVDMV	HLA-B*08:01	pentamer	IE1	CMV
RPHERNGFTVL	HLA-B*07:02	pentamer	pp65	CMV
RPPIFIRRL	HLA-B*07:02	pentamer	EBNA-3A	EBV
TPRVTGGGAM	HLA-B*07:02	pentamer	pp65	CMV
VLEETSVML	HLA-A*02:01	pentamer	IE1	CMV
VTEHDTLLY	HLA-A*01:01	pentamer	pp50	CMV
YSEHPTFTSQY	HLA-A*01:01	pentamer	pp65	CMV

Table 2.12: Fluorophores for multimer labelling.

Name	Fluorophore	Target multimer	Manufacturer
StrepTactin PE	PE	streptamer	IBA Lifesciences, Göttingen, DE
Pro5 Fluorotag R-PE	PE	pentamer	ProImmune, Oxford, UK

2.8 Peptides

All peptides were produced by JPT Peptide Technologies (Berlin, DE) with at least 70% purity. Lyophilized peptides were solved in DMSO to a stock concentration of 10 mg/ml and stored in a freezer at -20°C. Table 2.13 lists all peptides used in this project. A new batch of peptides was used for the peptide pool stimulations.

Table 2.13: List of immunogenic CMV peptides tested in this thesis. Each peptide is listed with its full-length amino acid sequence in one-letter code, its abbreviation (**Abbr.**, usually 3 letters), the peptide length in amino acids (**Len.**), the antigen it originates from, its HLA restriction (**HLA res.**), and the reference(s) to its first description with the stated HLA restriction

Name	Abbr.	Len.	Antigen	HLA res.	Reference(s)
ATVQGQNLK	ATV	9	pp65	A*11:01	Kondo et al. 2004
AYAQKIFKI	AYA	9	IE1	A*24:02	Kirchner et al. 2008
AYAQKIFKIL	AYA-10	10	IE1	A*24:02*	Kuzushima et al. 2001
CEDVPSGKL	CED	9	pp65	B*40:01	Kondo et al. 2004
CRVLCCYIL	CRV-I	9	IE1	C*07:02	Ameres et al. 2013
CRVLCCYVL	CRV	9	IE1	C*07:02	Ameres et al. 2013
DALPGPCI	DAL	8	pp65	B*51:01	Kondo et al. 2004
DEEDAIAAY	DEE-D	9	IE1	B*18:01	Retière et al. 2000
DEEEAIVAY	DEE	9	IE1	B*18:01†	Gavin et al. 1993; Ameres 2013
EDAIAAYTL	EDA	9	IE1	B*40:01	Smith et al. 2016
EEAIVAYTL	EEA	9	IE1	B*40:01‡	Braendstrup et al. 2014
EFFWDANDIY	EFF	10	pp65	B*44:02	Wills et al. 1996; Weekes et al. 1999
ELKRKMIYM	ELK-I	9	IE1	B*08:01§	Saulquin et al. 2000
ELKRKMMYM	ELK	9	IE1	B*08:01	Elkington et al. 2003
ELRRKMIYM	ELR-I	9	IE1	B*08:01	Weißbrich 2016
ELRRKMMYM	ELR	9	IE1	B*08:01	Elkington et al. 2003
FPKTTNGCSQA	FPK	11	IE1	B*55:01	Khan et al. 2007; Gibson et al. 2007
FPTKDVAL	FPT	8	pp65	B*35:02‖	Longmate et al. 2001; Wiesner et al. 2005
FRCPRRFCF	FRC	9	UL28/29	C*07:02	Kim et al. 2011
GPISGHVLK	GPI	9	pp65	A*11:01	Hebart et al. 2002
HERNGFTVL	HER	9	pp65	B*40:01/02	Kondo et al. 2004
ILEETSVML	ILE	9	IE1	A*02:01	Link et al. 2017
IPSINVHHY	IPS	9	pp65	B*35:01	Gavin et al. 1993
KEVNSQLSL	KEV	9	IE1	B*40:01	Nastke et al. 2005
KLGGALQAK	KLG	9	IE1	A*03:01	Elkington et al. 2003
KLGGALKAK¶	KLG-K	9	IE1	A*03:01	unpublished
KQIKVRVDM	KQI	9	IE1	C*06:02	Slezak et al. 2007
NLVPMVATV	NLV	9	pp65	A*02:01	Diamond et al. 1997
NQWKEPDVY	NQW	9	pp65	B*35:01	Wiesner et al. 2005
QEFFWDANDIY	QEF	11	pp65	B*44:02	Nastke et al. 2005; Gibson et al. 2007
QIKVRVTMV**	QIK	9	IE1	B*08:01	unpublished
QYDPVAALF	QYD	9	pp65	A*24:02	Kuzushima et al. 2001
RCPEMISVL	RCP	9	pp65	C*01:02	Kondo et al. 2004
RIKEHMLKK	RIK	9	IE1	A*03:01	Ameres et al. 2013; Ameres et al. 2014
RPHERNGFTV	RPH-10	10	pp65	B*07:02	Weekes et al. 1999
RPHERNGFTVL	RPH	11	pp65	B*07:02	Longmate et al. 2001
TPRVTGGGAM	TPR	10	pp65	B*07:02	Wills et al. 1996; Weekes et al. 1999
TRATKMQVI	TRA	9	pp65	C*06:02††	Rist et al. 2009
TSDACMMTMY	TSD	10	IE1	A*01:01	Braendstrup et al. 2014
VLEETSVML	VLE	9	IE1	A*02:01	Khan et al. 2002a
VTEHDTLLY	VTE	9	pp50	A*01:01	Elkington et al. 2003
VYALPLKML	VYA	9	pp65	A*24:02	Masuoka et al. 2001
YSEHPTFTSQY	YSE	11	pp65	A*01:01	Longmate et al. 2001

* A*23:01 discussed (Burrows et al. 2003)
† A*01:01 discussed (Elkington et al. 2003)
‡ B*44:02 discussed (Khan et al. 2007)
§ B*18:01 discussed (Retière et al. 2000)
‖ Often imprecisely published as B35-restricted (Elkington et al. 2003)
¶ CMV strain variant
** The correct CMV epitope is QIKVRVDMV (Elkington et al. 2003). Because of a typing error, it was ordered with a D→T exchange.
†† Also published as B13 (Nastke et al. 2005); B57 (Crough et al. 2005); B57 and B58 (Walker et al. 2007); B57, B58, and Cw6 (Zhong et al. 2008)

2.9 Cell lines and cell culture materials

2.9.1 Primary human cells

PBMCs were isolated from blood of healthy adult volunteers with their informed consent or obtained from buffy coats of anonymous blood donors (Institut für Transfusionsmedizin, Ulm, DE).

Table 2.14: Donor cohort of this thesis project.The table lists alias name, CMV status, and HLA class I alleles of the 21 CMV-positive and 10 CMV-negative donors. In addition, the table shows information on TRBV4-3 deletions and use of the TRBJ2-7*02 allele variant. No donor was homozygous for TRBJ2-7*02.

Donor	CMV	HLA-A	HLA-B	HLA-C	△BV4-3	BJ2-7*02
P01	+	A*03:01,A*30:03	B*07:02,B*58:01	C*07:02,C*07:18	yes	no
P02	+	A*02:01,A*02:01	B*07:02,B*35:03	C*04:01,C*07:02	no	no
P03	+	A*02:01,A*02:01	B*07:02,B*40:02	C*02:02,C*07:02	no	yes
P04	+	A*02:01,A*11:01	B*07:02,B*07:02	C*07:02,C*07:02	no	no
P05	+	A*11:01,A*24:02	B*07:02,B*35:01	C*04:01,C*07:02	no	no
P06	+	A*03:01,A*03:01	B*07:02,B*50:01	C*06:02,C*07:02	no	no
P07	+	A*02:01,A*31:01	B*07:02,B*44:02	C*05:01,C*07:02	yes	no
P08	+	A*02:01,A*03:01	B*07:02,B*57:01	C*06:02,C*07:02	no	no
P09	+	A*02:01,A*02:01	B*07:02,B*08:01	C*07:01,C*07:02	no	no
P10	+	A*01:01,A*01:01	B*07:02,B*08:01	C*07:01,C*07:02	no	no
P11	+	A*11:01,A*11:01	B*07:02,B*55:01	C*01:02,C*07:02	no	no
P12	+	A*02:01,A*02:01	B*07:02,B*27:05	C*02:02,C*07:02	no	no
P13	+	A*02:01,A*03:01	B*07:02,B*44:02	C*05:01,C*07:02	no	no
P14	+	A*03:01,A*32:01	B*07:02,B*38:01	C*07:02,C*12:03	no	no
P15	+	A*24:02,A*68:01	B*07:02,B*15:01	C*03:03,C*07:02	no	no
P16	+	A*02:01,A*24:02	B*07:02,B*39:01	C*07:02,C*12:03	no	no
P17	+	A*01:01,A*02:01	B*08:01,B*18:01	C*07:01,C*07:02	no	no
P18	+	A*02:03,A*02:06	B*40:01,B*55:02	C*01:02,C*07:02	yes	no
P19	+	A*01:01,A*03:01	B*35:01,B*57:01	C*04:01,C*06:02	yes	no
P20	+	A*03:01,A*24:03	B*18:01,B*37:01	C*06:02,C*12:03	yes	no
P21	+	A*01:01,A*02:01	B*08:01,B*40:01	C*03:04,C*07:01	no	no
N01	–	A*01:01,A*02:01	B*07:02,B*57:01	C*06:02,C*07:02	no	no
N02	–	A*02:01,A*02:05	B*07:02,B*50:01	C*06:02,C*07:02	no	no
N03	–	A*02:01,A*03:01	B*07:02,B*27:05	C*02:02,C*07:02	yes	no
N04	–	A*03:01,A*11:01	B*07:02,B*49:01	C*07:01,C*07:02	no	no
N05	–	A*03:01,A*03:01	B*07:02,B*35:01	C*04:01,C*07:02	no	yes
N06	–	A*02:01,A*03:01	B*07:02,B*35:01	C*04:01,C*07:02	no	yes
N07	–	A*03:01,A*24:02	B*07:02,B*35:03	C*04:01,C*07:02	no	no
N08	–	A*02:01,A*24:02	B*07:02,B*07:02	C*07:02,C*07:02	no	no
N09	–	A*02:05,A*03:01	B*07:02,B*49:01	C*07:01,C*07:02	yes	no
N10	–	A*01:01,A*26:01	B*35:01,B*57:01	C*04:01,C*06:02	no	yes

2.9.2 Cell lines

Table 2.15: Cell lines.

Name	Description	Reference(s)
B95.8	EBV-transformed cotton-top tamarin (*Saiguinus oedipus*) cell line B95.8	Miller et al. 1972; Miller and Lipman 1973
LL8	CD40-ligand–expressing LL8 mouse fibroblasts	Rancan et al. 2015

2.9.3 T-cell clones

Table 2.16: T-cell clones used in IFNγ ELISAs. They were obtained by limiting dilution of stimulated cell lines made from CMV-positive donors.

Name	Specificity	Description
F46M #160 CRV	CRV	Made by Stefanie Ameres from IE1 mini-LCL–stimulated cell line
XLBC13 #8 FRC	FRC	Made by Xiaoling Liang from FRC peptide-stimulated cell line

2.9.4 Cell culture media and additives

Table 2.17: Cell culture media and additives for cell culture of human lymphocyte-derived primary cells and cell lines and of LL8 feeder cells.

Name	Manufacturer
Cyclosporin A	Novartis, Nürnberg, DE
FCS, Fetal bovine serum 500 ml	Bio&SELL, Feucht bei Nürnberg, DE
Interleukin-2, Proleukin S	Novartis, Nürnberg, DE
Interleukin-4	R&D Systems, Minneapolis, US
Penicillin-Streptomycin, gibco 10^5U/mL	ThermoFisher scientific, Waltham, US
RPMI1640 liquid medium, gibco 500 ml, with L-Glutamine and Phenol Red	ThermoFisher scientific, Waltham, US
Sodium selenite, $\sim$98%	MP Biomedicals, Santa Ana, US

2.10 Hardware, software, and bioinformatics tools

2.10.1 Hardware

Table 2.18: Hardware.

Item	Manufacturer
Hard Drive, Canvio Basics, 3 TB	Toshiba, Tokyo, JP
iMac computer (21.5-inch, late 2013) **Operating system:** macOS High Sierra (v10.13.6) **Processor:** 3.1 GHz Intel Core i7 **Memory:** 16 GB 1600 MHz DDR3 **Startup Disk:** Macintosh HD **Graphics:** NVIDIA GeForce GT 750M 1024 MB **Serial Number:** DGKQF00YF8J8	Apple, Cupertino, US

2.10.2 Software and programming languages

Table 2.19: Software and programming languages.

Name	Version	Developer
Cisco AnyConnect Secure Mobility Client	4.5.05030	Cisco Systems, San Jose, US
Cytoscape	3.7.0	Cytoscape Developer Team, National Resource for Network Biology
FlowJo	10.4.1	FlowJo LLC, Ashland, US
GraphPad Prism	7.0d	GraphPad, San Diego, US
Java	1.8.0_60	Oracle, Redwood City, US
MacVector	14.5.2 (24)	MacVector, Apex, US
Microsoft Office Microsoft Excel Microsoft Powerpoint Microsoft Word	16.16.5	Microsoft, Redmond, US
MiXCR	1.8.1	MiLaboratory, Moscow, RU (Bolotin et al. 2015)
Python	2.7.10	Python Software Foundation, Delaware, US
R	3.5.1	R Core Team, GNU project
packages:		
beeswarm	0.2.3	
circlize	0.4.5	
colorspace	1.3-2	
data.table	1.11.8	
directlabels	2018.5.22	
ggplot2	3.1.0	
plyr	1.8.4	
reshape2	1.4.3	
scales	1.1.0	
stringdist	0.9.5.1	
stringr	1.3.1	
RStudio	1.1.463	RStudio, Boston, US
Terminal	2.8.3 (404.1)	Apple, Cupertino, US

2.10.3 Online bioinformatics tools

Table 2.20: Online bioinformatics tools.

Name	URL	Developer
Galaxy	https://galaxyproject.org/, private repository	Galaxy Community, The Galaxy Project
Multiple Primer Analyzer	https://www.thermofisher.com/	ThermoFisher scientific, Waltham, US
WebLogo	https://weblogo.berkeley.edu/	University of California, Berkeley, US

2.11 Services and cooperations

High-resolution HLA class I typing was done in the Center for Human Genetics and Laboratory Diagnostics (MVZ Martinsried, Munich, DE). High-throughput sequencing of TCR libraries was performed by Dr. Stefan Krebs and Dr. Helmut Blum at the Laboratory for Functional Genome Analysis (Gene Center, Ludwig Maximilian University of Munich, Munich, DE).

3 METHODS

3.1 Molecular biology

3.1.1 Preparation of TCRβ libraries

3.1.1.1 RNA isolation

Cells were lysed and total RNA was extracted with the RNeasy Mini Kit as per manufacturer's instructions. The quality and concentration of RNA in the eluate was measured with the Nanodrop ND-1000 spectrophotometer using 1 µl eluate. The RNA samples were either stored at -80°C in a freezer or, preferably, directly transcribed to complementary DNA (cDNA).

3.1.1.2 Reverse transcription

Per sample, 1 µg RNA or, if low RNA yields were obtained, a maximum of 12 µl RNA eluate, were used to make TCR libraries. The multimer-sorted samples of donors P01 and P04 had the lowest amounts of input RNA (0.14 µg–0.73 µg). RNA was treated with genomic DNA wipeout buffer and reversely transcribed to cDNA using the QuantiTect Reverse Transcription Kit. In the reverse transcription, a TCRβ-gene–specific primer (CP1 (Zhou et al. 2006)), which was designed to target both Cβ1 and Cβ2, was used at c_{end} = 0.2 µM and elongation was carried out for 30 min at 42°C. Then, the TCRβ cDNA was amplified in two subsequent PCRs.

3.1.1.3 Multiplex PCR

First, a gene-specific multiplex PCR was run (Table 3.1). In this first PCR step, equimolar amounts of 45 different forward primers covering all possible human *TRBV* gene segments (ss-BV-mix) and a mix of *TRBC*-specific reverse primers (R-seq-BC18-mix) with identical priming site in the *TRBC* region were used. In addition to the cDNA binding sites, each primer had an overhang sequence complementary to the Illumina Read 2 and Illumina Read 1 priming sequence, respectively. Thus, the Illumina read priming sites were added to the PCR products in the multiplex PCR. Three different versions of the R-seq-BC18 primer with 0, 1, or 2 degenerated (N) nucleotides inserted between the *TRBC* priming site and the Illumina Read 1 sequence were mixed at equimolar concentrations. Using primers with different N nucleotide insertions in this

position is required, because it considerably increases the nucleotide (nt) diversity of TCRβ reads in the first 5 base positions. Thereby cluster recognition during sequencing is facilitated and sequencing artefacts are reduced. To minimise amplification bias, forward and reverse primers were used in excess and only 10 cycles were run in the multiplex PCR. Each PCR was split into 2 PCR tubes each containing 50 µl per tube to ensure that the target temperatures are reached quickly during the short PCR cycles.

Table 3.1: Assembly (left) and thermocycler settings (right) of the first PCR (= multiplex PCR) of the TCRβ library preparation protocol.

Ingredient	**V**$_{stock}$	**c**$_{end}$
TCRβ cDNA	20 µl	-
dNTP mix	10 µl	0.2 mM
$MgSO_4$ mix	16 µl	4 mM
ss-BV-mix	3 µl	Σ0.3 µM
R-seq-BC18-mix	3 µl	Σ0.3 µM
Pfu polymerase	2 µl	0.05 U/µl
10× Pfu buffer	10 µl	1×
H_2O	36 µl	-
V$_{total}$	**100 µl** (2 tubes)	

PCR step	**T [°C]**	**t [s]**	
Initial denat.	95	120	
Denaturation	95	30	10×
Annealing	59	30	
Elongation	72	60	
Final elong.	72	600	

3.1.1.4 Purification of the intermediate PCR product

After the first PCR step, the two PCR reactions of the same sample were combined (V_{total} = 100µl) and the PCR product was purified with Agencourt AMPure XP Beads. The principle behind this system, solid-phase reversible immobilisation, uses paramagentic beads coated with carboxyl groups and a buffer containing polyethylene glycol and NaCl (DeAngelis et al. 1995). In the presence of the buffer, DNA molecules can bind to the coated beads. When proportionally more buffer is added, smaller DNA fragments can bind to the beads. PCR product purification was carried out according to the manufacturer's protocol using 0.8 µl beads suspension per 1 µl PCR product. At this concentration, DNA fragments that are at least 150 bp long, such as the TCRβ amplicons, can bind to the beads, while the shorter primers and self-annealed primer by-products are washed off. The purified PCR product was eluted in 20 µl H_2O.

3.1.1.5 Barcode PCR

Next, the PCR product was amplified in a second PCR using commercially available NEBNext Multiplex Oligos for Illumina (Dual Index Primers Set 1). These primers bind to the Illumina Read 1 and Illumina Read 2 priming sites that were introduced in the first

PCR. In addition, they contain 8-nt–long barcodes, which are required for correct assignment of sequence reads to the original sample, and the Illumina i5 and i7 adapters, which are needed for binding of the TCRβ amplicons to the Illumina flow cell. Primers i501-i508 were combined with primers i701-i712 so that each of the samples that will be combined on one sequencing lane has a unique combination of an i5 and i7 barcode. Table 3.2 shows the reagents and thermocycler settings used in this barcode PCR step. Generally, 15 cycles were sufficient to obtain as much DNA as needed to sequence the TCRβ library. However, in case the yield was very low, which affected approximately a third of all samples, the second PCR was repeated with 16-20 cycles to obtain satisfactory amounts of TCRβ PCR product.

Table 3.2: Assembly (left) and thermocycler settings (right) of the second PCR (=barcode PCR) of the TCRβ library preparation protocol.

Ingredient	**V_{stock}**	**c_{end}**
1^{st} PCR product	10 µl	-
dNTP mix	5 µl	0.2 mM
$MgSO_4$ mix	8 µl	4 mM
i5 primer	2 µl	Σ0.3 µM
i7 primer	2 µl	Σ0.3 µM
Pfu polymerase	1 µl	0.05 U/µl
10× Pfu buffer	5 µl	1×
H_2O	17 µl	-
V_{total}	**50 µl**	

PCR step	**T [°C]**	**t [s]**	
Initial denat.	95	120	
Denaturation	95	30	15×
Annealing	59	30	
Elongation	72	60	
Final elong.	72	600	

3.1.1.6 Purification of the final TCRβ PCR product

After the second PCR step, the PCR product was again purified using the Agencourt AMPure XP Beads as described above, with the alteration that the purified DNA was eluted in 10 µl H_2O.

3.1.1.7 Automated chip electrophoresis

To determine the TCRβ amplicon concentration in the samples and for the sake of quality control, the DNA was measured in an automated chip electrophoresis system: Samples were placed in a microchannel chip using the Agilent DNA 1000 Kit and measured with the Agilent 2100 Bioanalyzer automated electrophoresis system as per manufacturer's instructions. The TCRβ amplicons are approximately 300 bp in length and a concentration of no less than 5 nM, but ideally higher than 10 nM, was required for high-throughput sequencing. The presence of residual primers in the sample (Figure 3.1A) does not negatively impact the sequencing procedure, since these frag-

ments are too short to contain both the i5 and i7 adapter sequence; hence they cannot be amplified on the Illumina flow cell. For samples lacking the peak at 300 bp, the second PCR step was repeated with an increased number of cycles. Samples that had a very small peak at 300 bp were amplified in an additional rescue PCR using primers Nextera 1 and Nextera 2, which are complementary to the Illumina i5 and i7 adapter sequences. The rescue PCR was also done to rescue samples with overamplification peaks (Figure 3.1B). These overamplification peaks are caused by insufficient amounts of primer, which prevents correct elongation and leads annealing of mismatched DNA strands, which in turn generates bulky DNA structures. Overamplification can severely decrease sequencing quality, since it leads to underestimation of the amount of total TCRβ DNA in the sample, which in turn results in an overload of the sequencing flow cell. One cycle of denaturing, annealing, and elongation with suitable primers restores the complementary strands of the TCRβ amplicons and removes bulky secondary DNA structures. After the rescue PCR, samples were purified with Agencourt AMPure XP Beads and remeasured on the Agilent 2100 Bioanalyzer.

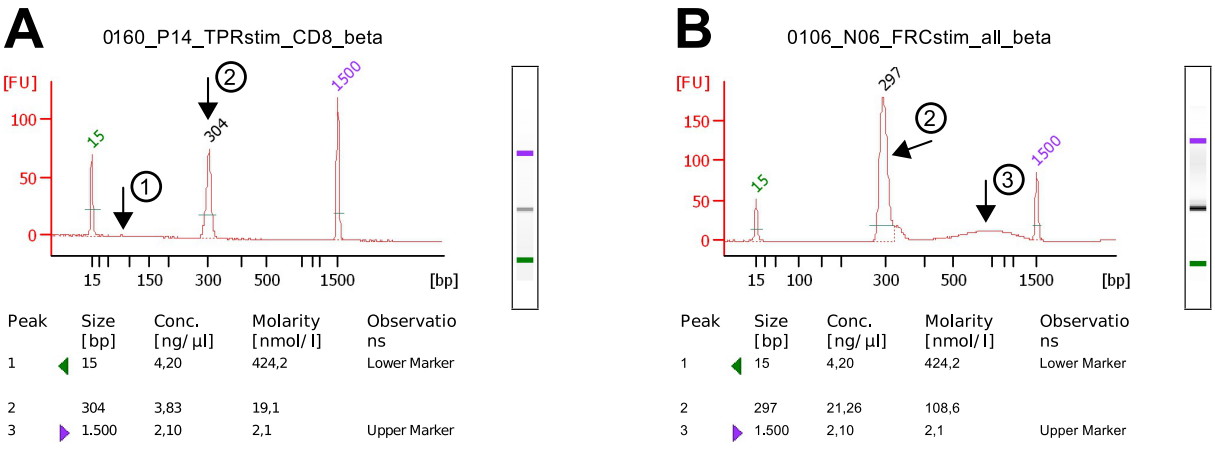


Figure 3.1: Exemplary electropherograms of two TCRβ libraries. The electropherograms were generated using the Agilent 2100 Bioanalyzer. The lower DNA marker (green arrow) is 15 bp long and the upper DNA marker (purple arrow) is 1500 bp long. (**A**) This TCRβ sequencing sample of T cells from donor P14 which were stimulated with peptide TPR and subsequently CD8-enriched is sufficiently pure and contains enough DNA for high-throughput sequencing. The TCRβ amplicon peak is at 304 bp (2) and its signal strength corresponds to a concentration of 19.1 nM. A second peak suggestively shows at approximately 80 bp (1), which indicates the presence of residual primers that were not removed by paramagnetic bead purification. (**B**) This sample was derived from cells of donor N06 stimulated with peptide FRC. Apart from the TCRβ amplicon peak at 297 bp (2), there is a broad DNA peak between 500 bp and 1500 bp (3) indicating overamplification of the sample.

3.1.2 Preparation of TCRα libraries

Libraries for sequencing of TCRα repertoires were prepared analogously to the TCRβ libraries (Section 3.1.1) with minor alterations. To reversely transcribe the total RNA, TCRα-gene–specific primer RT-TRAC-A was used. For the first PCR step, a new set of TCRα multiplex primers (TRAV/TRDV-mix) was designed. Importantly, the overhangs of the TCRα primers were switched: The multiplex primers binding to the variable region were complementary to the Illumina Read 1 priming site, and the reverse primer binding to the constant region was complementary to the Illumina Read 2 priming site. Because of this, no N nucleotide insertions were required to increase read diversity in the first 5 bases and therefore a single reverse primer (CαB) was sufficient. However, a drawback is that TCRα libraries alone may be not diverse enough for sequencing without TCRβ libraries on the same lane, especially if strongly expanded clonotypes are present in the sample. The full-length PCR products for high-throughput sequencing are slightly larger than the TCRβ products and measure around 320 bp.

3.1.3 High-throughput Illumina sequencing

TCR libraries that passed the Bioanalyzer quality check were handed over to our cooperation partners Dr. Stefan Krebs and Dr. Helmut Blum (Laboratory for Functional Genome Analysis, Gene Center, Ludwig Maximilian University of Munich, Munich, DE). All TCR libraries that were to be sequenced together on the same flow cell lane were mixed at equimolar concentrations to obtain a concentation of 10 nM DNA in total. Next, 10 μl TCR DNA mix was denatured by adding 10 μl 1M sodium hydroxide. The reaction was neutralised by addition of 980 μl Illumina hybridisation buffer. Next, 60 μl of the neutralised denatured DNA was combined with 440 μl Illumina hybridisation buffer to obtain a 12 pM DNA concentration for sequencing. Our TCR libraries are of lower variability than i.e. whole transcriptome libraries, since we only sequence one type of transcript, namely the TCRα or TCRβ cDNA. Hence, a high amount of phiX control library (10%) was added to the sample before sequencing. The template hybridisation and first extension was done using the cBot 2 System (Illumina, San Diego, US). Subsequently, the TCR libraries were paired-end sequenced in rapid mode. Sequence read length was 150 bp or 175 bp in each direction, and index read length was 8 nt per direction.

3.1.4 Isolation and PCR amplification of mini-EBV–specific DNA

Infection of B cells with mini-EBV leads to expression of genes delivered by the virus. In order to confirm the successful transformation of the B cells through mini-EBV, DNA was isolated and samples were tested for the presence of mini-EBV–specific DNA sequences. First, 1 ml cell suspension was transferred to an 1.5 ml reaction tube and spun down in a desktop centrifuge at 300×g for 10 min at room temperature (rt). The supernatant was removed and cells were lysed by addition of 25 µl 50 mM sodium hydroxide and subsequent incubation at 95°C for 5 min. Samples were then cooled down to rt and 4 µl 1 M TRIS hydrochloride (pH=7.0) was added to neutralise the reaction. To remove cell debris, samples were centrifuged at 14,000×g for 1 min at rt and 25 µl supernatant was transferred to a new 1.5 ml reaction tube. Next, the PCRs were assembled and run as shown in Table 3.3.

Table 3.3: Assembly (left) and thermocycler settings (right) of the PCR for amplification of EBV DNA amplicons.

Ingredient	**V_{stock}**	**c_{end}**
mini-LCL DNA	2 µl	-
dNTP mix	5 µl	0.2 mM
cam-up primer	1.5 µl	0.3 µM
cam-down primer	1.5 µl	0.3 µM
gp85c primer	1.5 µl	0.3 µM
gp85d primer	1.5 µl	0.3 µM
Taq polymerase	1 µl	0.1 U/µl
10× Taq buffer	5 µl	1×
H_2O	31 µl	-
V_{total}	**50 µl**	

PCR step	**T [°C]**	**t [s]**	
Initial denat.	95	300	
Denaturation	95	45	30×
Annealing	59	45	
Elongation	72	45	
Final elong.	72	600	

3.1.5 Agarose gel electrophoresis

Argarose gel electrophoresis was used to validate functionality of the newly designed TCRα multplex PCR primers for sequencing library preparation and to confirm the identity of freshly transformed (mini-)LCLs. Depending on the number of PCR samples, 1.5% agarose gels were made by mixing 1.05 g (small gel) or 1.8 g (large gel) agarose with 70 ml or 120 ml 1×TAE buffer, respectively. To fully dissolve the agarose, the mixture was brought to boil using the microwave. Right before pouring the gels, 2 or 3 drops ethidium bromide were added to stain the DNA amplicons on the gel. 1.5 mm combs were used to make 10 or 20 pockets in the gel. 25 µl of a mixture consisting of 25 µl sample and 5 µl 6x loading dye were loaded in each sample pocket. As a marker,

1 µl 100 bp GeneRuler ladder, diluted with 19 µl H_2O, was mixed with 5 µl loading dye and 20 µl were applied. The gel electrophoresis was run with 1×TAE buffer as running buffer at a voltage of approximately 9 V/cm for 1 hour (h). DNA bands on the gel were visualised using the Vilber Lourmat gel documentation system.

3.2 Cell culture

In this thesis, most experiments were performed with human PBMCs or derived cell lines. Moreover, CD40-ligand–expressing LL8 mouse fibroblasts were used to produce and maintain CD40-stimulated human B cell lines (B blasts) (Garrone et al. 1995; Wiesner et al. 2008; Rancan et al. 2015). Cell culture medium was always RPMI1640 liquid medium supplemented with 8% or 10% fetal calf serum (FCS), 100 U/ml penicillin, 100 µg/ml streptomycin, and 100 nM sodium selenite. Unless stated otherwise, cells in 15 ml or 50 ml Falcon tubes were spun down in a large bench centrifuge at speed 300×g and rt for 10 min. Cells were handled under sterile conditions in a laminar flow hood. They were counted and phenotypically observed under a microscope at 10× and 50× magnification.

3.2.1 Isolation of human peripheral blood mononuclear cells

Blood was drawn from healthy adults with their informed consent or obtained from buffy coats of anonymous blood donors (Institut für Transfusionsmedizin, Ulm, DE). Clotting of freshly drawn blood was prevented by adding 0.5 ml heparin per 60 ml syringe. The blood was transferred to 50 ml Falcon tubes at up to 20 ml (buffy coats) or up to 30 ml (fresh blood) per tube. Next, the blood was diluted with phosphate-buffered saline (PBS) to 35 ml. For PBMC separation by density gradient centrifugation, approximately 13 ml Ficoll with a density of 1.077 g/ml was carefully layered beneath the diluted blood. Next, the blood samples were centrifuged at 1000×g at rt for 25 min with slow acceleration and deceleration in order to separate the blood samples into layers (Figure 3.2). Optionally, 10 ml plasma were removed and stored in a 15 ml Falcon tube at -20°C for serological tests. The PBMC layer was transferred to a new 50 ml Falcon tube and diluted to 50 ml with PBS to lower the cell density. After standard centrifugation, the supernatant was discarded and the pellet was resuspended in PBS. If a sample was split into multiple tubes, cells from the same sample were combined in this step. Again,

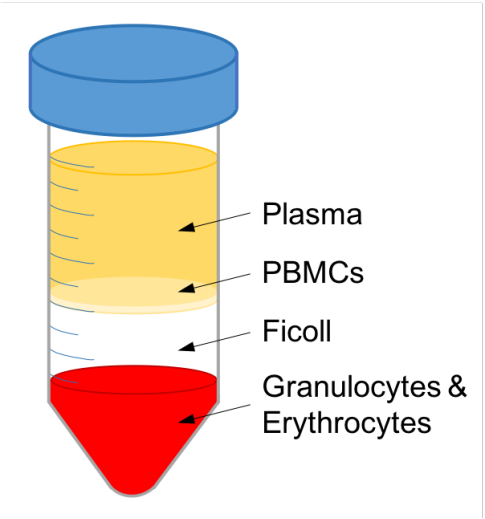


Figure 3.2: Layers of blood samples after density gradient centrifugation.

the suspension was topped up to 50 ml, cells were spun down by centrifugation and the supernatant was removed. This washing procedure with PBS was repeated two more times before the PBMCs were counted. The freshly isolated PBMCs were either used in experiments or were divided into aliquots and cryopreserved until use.

3.2.2 Cell counting

Per sample, 10 µl cell suspension was mixed with 10 µl 0.4% trypan blue solution. Then, 10 µl stained cell suspension was transferred to a hemocytometer and counted. If cell concentrations were too high and exceeded 4 million cells/ml, cell suspensions were diluted 1:5 or 1:10 with PBS prior to staining with trypan blue solution.

3.2.3 Cryopreservation of cells

Cryomedium for long-term storage of human cells was RPMI1640 liquid cell culture medium mixed with FCS and dimethyl sulfoxide (DMSO) at a ratio of 5:4:1. Cells were spun down by centrifugation (300×g, 10 min, rt), resuspended in 1 ml cryomedium, and transferred to 1.8 ml Nunc CryoTubes. The cryotubes were kept on ice and stored at -80°C as soon as possible. After few weeks to months, the cryopreserved cells were relocated to liquid nitrogen tank, where they were stored in the gas phase until use. Frozen cells were quickly thawed with and taken up in 10–20 ml pre-warmed standard cell culture medium. If the thawed cells were isolated from buffy coats, they were passed through a cell strainer to remove aggregates. Then, cells were centrifuged, the

DMSO-containing supernatant was removed, and cells were resuspended in fresh cell culture medium or PBS.

3.2.4 Establishment of mini-lymphoblastoid cell lines

In order to analyse the human TCR repertoire against full-length CMV antigens, mini-lymphoblastoid cell lines (mini-LCLs) were established by infecting PBMCs with a mini-EBV containing encoding only EBV proteins, or EBV proteins together with one additional CMV antigen, which was either pp65 or IE1. The mini-EBV particles for B-cell transformation were produced in our lab as previously described (Moosmann et al. 2002; Ameres et al. 2015). As a positive control for infection, full-length EBV produced by the cotton-top tamarin cell line B95.8 (Miller et al. 1972; Miller and Lipman 1973) was used. Stable mini-LCL cell lines were generated with existing virus supernatants as described by Moosmann et al. (Moosmann et al. 2002). Firstly, PBMCs were thawed according to Section 3.2.3 and seeded at 400,000 cells per well on a 96-well flat-bottom plate in 100 µl cell culture medium supplemented with 1 µg/ml Cyclosporin A (CsA). CsA is a T-cell inhibitor and prevents outgrowth of T cells in mini-LCL cultures. Per mini-LCL, as well as for the positive and negative controls, six replicates were seeded. Then, 100 µl freshly thawed mini-EBV–containing cell culture supernatant was added to each well for mini-LCL generation. For the positive control, 10 µl EBV-containing B95.8 supernatant was diluted with 90 µl cell culture medium and 100 µl diluted B95.8 supernatant was added per well. For the negative control, 100 µl cell culture medium was added to each well. The cells were co-incubated with mini-EBV supernatants or controls at 37°C for 3 h or overnight for infection. Next, 150 µl medium was carefully replaced by 150 µl fresh CsA medium without touching the infected cells on the bottom of each well. Cells were then incubated until they proliferated strongly and formed spherical aggregates. Every week, 100 µl supernatant was exchanged with 100 µl fresh CsA medium. When the cell culture medium started to turn acidic and mini-LCL outgrowth was observed, mini-LCLs were expanded 1:2 or 1:3 once or twice a week depending on the growth rates. The mini-LCLs grow non-adhesively and were first expanded on 48-well plates, then 24-well plates, 12-well plates, 6-well plates, and, finally, in cell culture flasks stored upright in the incubator. Approximately 4 weeks past infection, addition of CsA to the medium was discontinued. The identities of the best-

growing mini-LCLs were confirmed as described in Section 3.1.4 and further passaged or cryopreserved in 1 ml aliquots until use.

3.2.5 Generation and maintenance of CD40-stimulated B cell lines

CD40-stimulated B cells (B blasts) were produced from human PBMCs by co-culture with LL8 feeder cells (Garrone et al. 1995; Wiesner et al. 2008; Rancan et al. 2015). For this purpose, the LL8 cells were harvested and irradiated at $\geq$180 Gy for 5 h in a ^{137}Cs device before they were plated at 1 million cells per plate on a 96-well flat-bottom plate. The irradiated LL8 cells were rested for 1 day in an incubator which allows them to adhere to the bottom. On the next day, PBMCs were thawed as described in Section 3.2.3. All supernatant was removed from the irradiated LL8 feeder cell wells and the PBMCs were seeded at different concentrations in 200 µl cell culture medium containing 1 µg/ml CsA and 2 ng/ml IL-4. The cell numbers seeded per well were 2×10^5, 10^5, 5×10^4, 2.5×10^4, and 1.25×10^4; 6 wells were prepared for each cell number. Every week, cells were transferred to new LL8 feeder plates and half of the medium was exchanged by freshly prepared cell culture medium with CsA and IL-4. When B blasts proliferated strongly, they were split 1:2 to a larger plate, i.e. to a 48-well plate at first, and in later passages to 24-well and 12-well plates. Of all cultures that proliferated, the cultures with the lowest seeded cell number were kept and further expanded. After 3 months in culture, CsA was omitted in B blast cell culture medium.

3.3 T-cell stimulation assays

3.3.1 Peptide stimulation assay

CMV-epitope–specific T cells were expanded by in vitro stimulation of PBMCs with a single HLA class I-restricted peptide per culture. 25 million PBMCs were suspended in 2 ml standard cell culture medium containing 5 mg/ml peptide and were incubated at 37°C for 1 h. Subsequently, cells were washed thrice with PBS to remove excess peptide. Next, the cells were resuspended in 12.5 ml cell culture medium supplemented with 50 U/ml IL-2 and distributed at 2.5 ml/well to a 12-well plate. The plate was placed in an incubator. After 6$\pm$1 days, the cells of each well were resuspended, distributed to two wells, and 1 ml of fresh culture medium supplemented with IL-2 was added to each

well. Cells were harvested on day 10 of culture and cryopreserved, used to isolate RNA, or stained for flow cytometry.

3.3.2 Stimulation with mini-LCLs

Firstly, mini-LCLs were γ-irradiated with 50 Gy in a ^{137}Cs device. Then, 1.5×10^5 mini-LCL cells were combined with 6×10^6 PBMCs in 3 ml standard cell culture medium per replicate in a 12-well plate. Four replicates were made per culture. After 9 days, and then every 7 days, the T cells were restimulated. For this purpose, cultures were harvested, washed, and counted. 3×10^6 T cells per well were seeded on a 12-well plate and coincubated with 10^6 irradiated mini-LCLs in 3 ml medium with 50 U/ml IL-2. On day 30, cultures were harvested and T cells were further analysed.

3.3.3 Stimulation with peptide pools

PBMCs were stimulated as described in Section 3.3.1, but with multiple epitopes simultaneously. For this pupose, 32 CMV-derived peptides were mixed in 8 different pools as listed in Table 3.4. These CMV peptides were selected, because they were previously published as CD8$^+$ T-cell epitopes with known HLA restrictions. The sole exception was epitope KLG-K, which was added because it is a CMV strain variant of epitope KLG. Per pool, 12 or 13 different epitopes were combined, and each peptide was present in a unique subset of pools, which we termed the ”address” of the peptide. Epitope variants were also included. These variants were of different types: (i) epitope variants from different CMV strains, (ii) shorter or longer versions of epitopes, and (iii) shifted versions of epitopes. The variant peptides were added to the same pools as the corresponding reference peptide. To make the pools, each peptide was dissolved in 500 µl DMSO (c_{stock}=10 mg/ml) and 20 µl of each peptide stock was added to the respective pools. The pools were topped up with DMSO to 400 µl, so that the final concentration per peptide was 0.5 mg/ml. Up to 200 million PBMCs from healthy donors were resuspended in 16 ml cell culture medium and split to 8 different 15 ml Falcon tubes. Then, 4 µl of one of the eight peptide pools was added to each sample (c_{end}=1 µg/ml per peptide). Each sample was incubated, washed, and processed as described for the single peptide stimulation in Section 3.3.1. The pool stimulation experiments were performed by Fenja Gerpott within the framework of her Master's thesis and supervised by me.

Table 3.4: Composition of 8 peptide pools containing 12 or 13 defined CMV-derived peptide sequences each. Every peptide was present in exactly 3 pools, as indicated by ”x”, and the combination of pools was unique for each peptide. Peptide variants were distributed to the same pools as the reference peptide; they are listed in parentheses. The number of the unique pool combinations are designated ”address” of a peptide. See Table 2.13 for full sequences and further information on all peptides.

Peptide	Pool								Address
	1	2	3	4	5	6	7	8	
ATV	x	x		x					124
AYA (AYA-10)		x	x					x	238
CED	x					x	x		167
CRV (CRV-I)	x				x			x	158
DAL	x					x		x	168
DEE (DEE-D)	x	x	x						123
EEA (EDA)		x					x	x	278
EFF (QEF)		x	x	x					234
ELK (ELR,ELK-I,ELR-I)				x		x	x		467
FPK	x		x	x					134
FPT			x		x	x			356
FRC		x	x				x		237
GPI		x		x			x		247
HER	x		x				x		137
IPS					x		x	x	578
KEV						x	x	x	678
KLG (KLG-K)	x			x		x			146
KQI			x		x			x	358
NLV				x		x		x	468
NQW		x		x	x				245
QIK	x						x	x	178
QYD					x	x		x	568
RCP	x	x						x	128
RIK			x	x	x				345
RPH (RPH-10)					x	x	x		567
TPR				x	x	x			456
TRA	x			x	x				145
TSD			x		x		x		357
VLE (ILE)	x		x		x				135
VTE			x	x		x			346
VYA		x				x	x		267
YSE		x		x		x			246

3.3.4 Stimulation of T-cell clones for the IFNγ ELISA

Since the results from peptide pool stimulation assays raised the concern that some cross-contamination of peptides or peptide pools may have occurred, the purity of CRV and FRC peptides in the old peptide batch, the new peptide batch, and the assembled pools was indirectly assessed by quantifying the response of antigen-specific Tcell clones to these peptide preparations. To do so, T-cell clones with known specificity for CRV or FRC were short-term–stimulated with peptide-loaded B blasts to measure the

resulting IFNγ-secretion. B blasts expressing the required HLA for presentation of CRV and FRC, HLA-C*07:02, were counted, spun down and resuspended at 20,000 cells/ml in standard cell culture medium. Per condition, 1 ml cell suspension was transferred to a 15 ml Falcon tube. Peptides or peptide pools were added to the B blasts at a final concentration of 1 μg/ml and cells were incubated at 37°C for 1 h. Subsequently, B blasts were washed with PBS to remove excess peptide and resuspended in 1 ml medium. 10,000 B blasts (in 100 μl medium) per well were seeded in triplicates for each condition on a 96-well V-bottom plate. Then, 10,000 cells of the CRV-specific or FRC-specific T-cell clone in 100 μl standard cell culture medium per well were added to the peptide-loaded B blasts. The plate was incubated overnight at 37°C. The following day, the supernatant was tested for IFNγ as described in Section 3.4.3.

3.4 Immunological methods

3.4.1 Serum isolation

Human serum was isolated from fresh blood samples. Approximately 5 ml blood were drawn from healthy donors into a serum tube without anti-coagulant. It was left for blood clotting at rt for approximately 30 min. The supernatant, which is the serum, was then transferred to 1.5 ml reaction tubes and stored at -20°C until use.

3.4.2 CMV serology

CMV status was determined by ELISA using the Human Anti-Cytomegalovirus IgG ELISA Kit (abcam). The ELISA was conducted according to the manufacturer's protocol using the supplied reagents and consumables. Absorption was measured at 450 nm with the automated plate reader EL800.

3.4.3 IFNγ ELISA

An IFNγ ELISA with supernatants from overnight peptide-stimulated T-cell clones (Section 3.3.4) was performed using the Human IFN-γ ELISA development kit (Mabtech) as per manufacturer's instructions, but using half of the recommended volumes of supernatant samples and diluted reagents (anti-IFNγ capture antibody, biotinylated anti-IFNγ-antibody, and streptavidin-alkaline phosphatase). Plates were then incubated with 100 μl of alkaline phosphatase substrate solution (final concentration: 1 mg/ml

para-nitrophenylphosphate, 0.8 mM MgSO4, 10% diethanolamine in water, pH 9.5) for 20min. Alkaline phosphatase activity was measured via absorption at 450 nm with the automated plate reader EL800.

3.4.4 Magnetic bead separation

Before or after stimulation with single peptides (Section 3.3.1), PBMCs or peptide-stimulated cells were CD8-enriched by magnetic-activated cell sorting (MACS) as per CD8 MicroBeads instructions with few modifications described as follows. During the whole procedure, all cells and reagents were kept on ice. The columns and the magnet were pre-cooled in a fridge. In summary, cells were harvested or thawed, spun down by standard centrifugation (300×g, 10 min) and subsequently taken up in ice-cold MACS buffer, which was PBS supplemented with 2% FCS; 80 µl MACS buffer were used per 10^7 cells. Cells were then incubated with CD8 MicroBeads using 20 µl beads per 10^7 cells. After thorough washing, the magnetically labelled cells were transferred to an MS or LS column fixed in a MACS Separator magnet and labelled cells were immobilised on the column. The flow-through was collected and further analysed or cryopreserved as CD8-depleted reference sample. Columns were washed twice with ice-cold PBS buffer to remove residual unlabelled cells, before CD8-positive cells were flushed out and further analysed or cryopreserved. Analogously, CD4-positive lymphocytes were enriched from PBMCs by magnetic sorting with CD4 MicroBeads.

3.4.5 Flow cytometry

Phenotypes of PBMCs and sorted and/or stimulated cell samples were analysed by flow cytometry. In doing so, information about T-cell subsets within samples was obtained and the poportions of epitope-specific T cells were determined. Fluorescence-activated cell sorting (FACS), in this thesis synonymously used for flow cytometry measurement without cell sorting unless explicitly stated otherwise, was also used to check the purity of MACS-enriched T-cell entities. FACS buffer was always PBS supplemented with 2% FCS. All cell samples, buffers, and antibodies were kept on ice at all times and centrifugation was carried out at 4°C.

3.4.5.1 Antibody and multimer titration

In order to determine the minimum amount of antibody or multimer required to sufficiently stained all marker-positive cells in a sample, 3×10^5 cells each were co-incubated with different concentrations of fluorochrome-labelled antibodies or peptide-MHC multimers on ice for 20 min and then washed twice with 1 ml FACS buffer. The fluorescence was then measured with the BD LSRFortessa flow cytometer. For each antibody or multimer, the optimum amount for staining was determined as the lowest amount that did not lead to a major downward shift of the positive population (Figure 3.3).

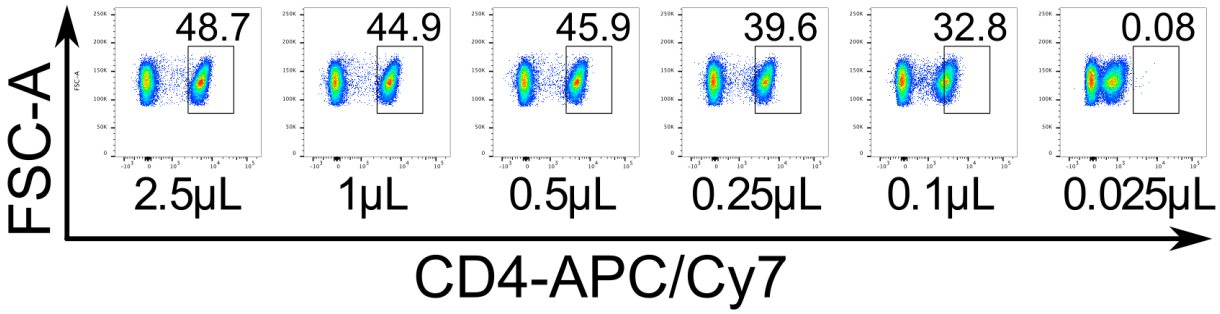


Figure 3.3: Titration of fluorescent-labelled antibodies and multimers, exemplarily shown here for PBMCs from donor P01 stained with αCD4-APCCy7. Samples shown here were gated on lymphocytes in the forward scatter vs. side scatter view. The lowest concentration that does not shift the positive population out of the gate, here 0.5 µl αCD4-APCCy7 antibody per 3×10^5 cells, was later used in the staining panels.

3.4.5.2 Surface marker staining

In order to assess the different cell populations within samples, cell surface markers were stained with different fluorescence-labelled, commercially available antibodies. Over the course of this thesis, the panel was changed for improved staining results. Generally, the surface markers CD3, CD4, CD8, and KIR2DL2/3 (CD158b) were always stained. CD3 is a ubiquitous receptor expressed on T cells, CD4 is a marker for T helper cells, CD8 is a maker for cytotoxic T cells, and KIR2DL2/3 is a receptor expressed on NK cells. KIR2DL2/3-positive cells were excluded from quantification of HLA-C–restricted T-cell responses in the flow cytometry data analysis step, since HLA-C streptamers may also stain NK cells or T cells of irrelevant specificity that express KIR2DL2/3. In addition to these markers, the differentiation markers CD45RA and CCR7 (CD197) were occasionally stained to investigate the differentiation state of specific T cells in the sample. The panels used throughout the thesis can be found in Table 2.9. Per sample, up to 3×10^5 cells were stained. A master mix containing

all panel antibodies was prepared by mixing the amounts of antibody determined by titration and adding FACS buffer to 20 µl per sample. Next, cells were washed with FACS buffer once and distributed to 1.5 ml reaction tubes or V-bottom 96-well plates at up to 3×10^5 cells per tube or well. After centrifugation at 4°C, the supernatant was removed and cells were thoroughly resuspended in 20 µl master mix per sample. Then, the cells were incubated with the antibodies on ice in the dark for ≈20 min. Samples were washed twice. Subsequently, cells were taken up in 100 µl FACS buffer for direct measurement, or they were fixed with 100 µl FACS buffer containing 1% formaldehyde and stored at 4°C in the dark until the next day.

3.4.5.3 Pentamer staining with commercially available multimers

All of the HLA-A–restricted and HLA-B–restricted multimers used in this project were ready-to-use pre-assembled pentamers purchased from ProImmune (Table 2.11). Per sample, up to 3×10^5 cells were stained with 0.5 µl unlabelled pentamer in 9.5 µl FACS buffer. Staining took place at rt for ≈15 min. For the negative controls, cells were incubated in 10 µl with no pentamer added. After an intermediate washing step, cells were stained with a surface marker panel as described above, to which 1 µl Fluorotag-PE was added per sample. Fluorotag-PE binds to the unlabelled pentamers on the cell surface and is required to visualise the multimer-specific cells in the flow cytometer.

3.4.5.4 Streptamer staining

For HLA-C–restricted CMV peptides, no off-the-shelf multimers were available until very recently. In this project, peptide-MHC streptamers were therefore assembled in vitro from peptide-MHC monomers, which were produced and kindly provided by Fabian Schlott, Michael Neuenhahn, and Dirk Busch (Technical University Munich, Munich, DE). To assemble the multimers, 0.4 µl peptide-MHC monomer was mixed with 0.5 µl Strep-Tactin PE and 9.1 µl FACS buffer and incubated on ice shielded from light for ≈30 min. For the negative controls, 0.5 µl Strep-Tactin PE was combined with 9.5 µl FACS buffer and incubated under the same conditions. Per up to 3×10^5 cells, 10 µl incubated streptamer or control was mixed thoroughly with the cells and samples were incubated on ice in the dark for ≈30 min. Subsequently, cells were spun down and stained with a surface marker panel as described above.

3.4.5.5 Compensation controls

Compensation controls were included in each flow cytometry experiment. For each fluorochrome, 5×10^4–10^5 cells were stained with one of the fluorescence-labelled antibodies used in the surface marker panel alone. Since KIR2DL2/3 was not expressed on a high proportion of cells, anti-CD4 antibody labelled with the same fluorophore was used as compensation control for the KIR2DL2/3 antibody. Cells for the compensation controls were resuspended in a mix of 1 µl fluorescence-labelled antibody and 19 µl FACS buffer and incubated on ice in the dark for ≈20 min. Excess antibody was removed by washing twice, and cells were resuspended in 100 µl FACS buffer for measurement or fixed with formaldehyde and stored until use.

3.4.5.6 Flow cytometry measurements

The stained cell samples were transferred to FACS test tubes and left on ice in the dark until BD LSRFortessa flow cytometer measurement. Fluorescence was compensated by measuring 10^4 events of unstained cells and single fluorochrome-stained samples. After compensation, up to 10^5 events were recorded per sample using the built-in BD FACSDiva software. Recorded FACS data was later analysed as described in Section 3.5.2.

3.5 Bioinformatics and data analysis

3.5.1 Primer design for TCRα library generation

The binding portions of the primers for TCRα sequencing were designed to resemble the properties of the TCRβ sequencing primers. With the exception of primer ss-BV15, which was designed by Xiaoling Liang, the binding portions of the TCRβ sequencing primers were originally published by Robins et al. (Robins et al. 2009). From this set of TCRβ primers, the criteria listed in Table 3.5 were derived and applied to TCRα primer design. In addition to these requirements, the risk of TCRα primer dimerisation, duplex formation, and hairpin structure formation was minimised using the ”Quicktest Primer” option of the software MacVector. Finally, the Illumina Read 1 priming site was added to the 5’ end of each TRAV-binding forward primer sequence as a non-pairing overhang. In addition, two different reverse primers (CαA and CαB), which bind to

different parts of the TRAC region, were designed and the Illumina Read 2 priming site was added to the 5' end of both primers.

Table 3.5: Criteria for the design of TCRα multiplex primers as derived from the previously established TCRβ primers (Robins et al. 2009).

Parameter	Required value
Length	20 – 31 nt
Upstream position*	31 – 38
Last base	C or G
GC content	40 – 60%
T_m(Santa Lucia)†,‡	63.2 – 68.2°C
T_m(Multiple Primer Analyzer)‡,§	70.0 – 75.0°C

* Position upstream of CDR3, which here starts after the canonic cysteine codon
† Determined using MacVector
‡ T_m = melting temperature
§ Determined with the online tool "Multiple Primer Analyzer"

3.5.2 Flow cytometry data analysis

The flow cytometry data were analysed using the software FlowJo. Figure 3.4 shows the gating strategy that was followed throughout this project. The events were first gated on lymphocytes in a plot of the forward scatter area (FSC-A) and the side scatter area (SSC-A). Next, doublets were excluded in an FSC-A versus forward scatter height (FSC-H) plot. Finally, events were gated on CD3-expressing T cells. Since NK cell receptor KIR2DL2/3 may be bound by HLA-C–streptamers, these cells are excluded in streptamer-stained samples by gating on KIR2DL2/3-negative events.

3.5.3 Sequencing data analysis

Sequencing data was uploaded by Stefan Krebs to the group's Galaxy repository (`http://blum-galaxy.genzentrum.lmu.de/galaxy/`), where raw data was processed, downloaded and further analysed using various software.

3.5.3.1 Raw data processing

Raw data deposited on the Galaxy platform were accessed through the virtual private network of Ludwig-Maximilians-University Munich. Using the built-in online tools of the Galaxy platform, raw data were first demultiplexed. During demultiplexing, all reads obtained from a sequencing lane were grouped by barcode. Two separate files con-

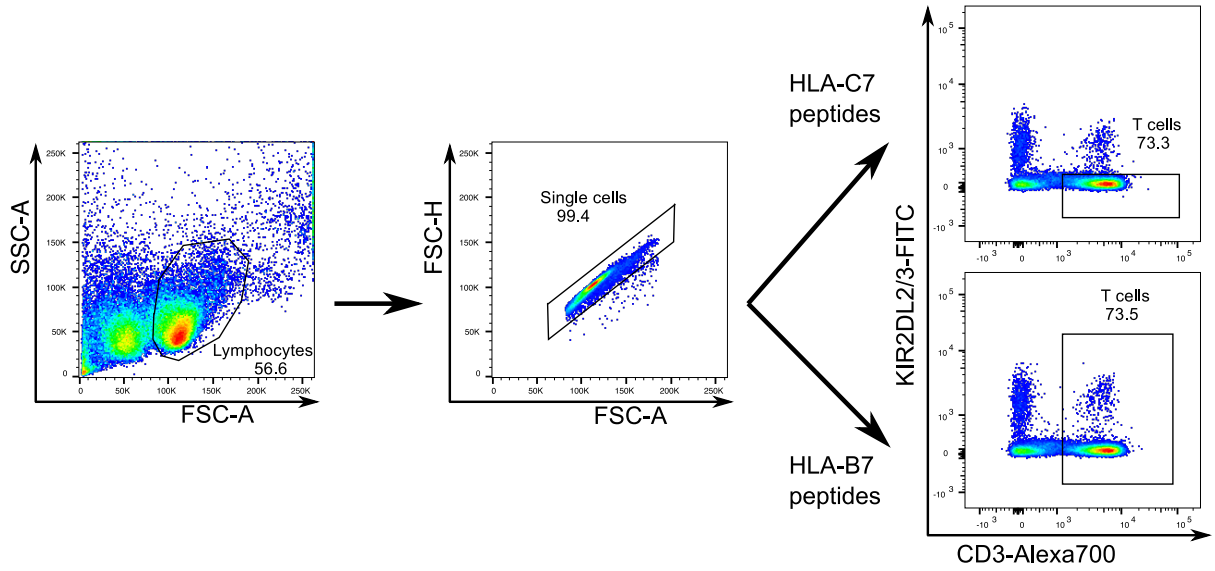


Figure 3.4: Gating strategy for T cells, here exemplarily shown for PBMCs isolated from donor P07. All events were gated on lymphocytes and subsequently on single cells. Next, T cells were isolated by gating on CD3-expressing cells. For multimer staining of T cells targeting HLA-C*07:02–restricted epitopes, KIR2DL2/3-positive cells were excluded to avoid false positive events due to the presence of NK cells or irrelevant KIR2DL2/3-positive T cells in the samples.

taining the forward and reverse reads, respectively, were generated for each barcode. When demultiplexing reads from single-barcoded libraries produced by Xiaoling Liang, no barcode mismatch was tolerated in the 6-nt barcodes. For the dual-barcoded libraries generated in this thesis, a single mismatch per 8-nt barcode was permitted. Demultiplexed sequencing data was then quality-filtered using the built-in FastqFilter tool in paired-end mode. FastqFilter trims low-quality bases from each side and removes reads that are shorter than a user-defined length. If a forward read is removed due to poor quality in this process, the paired read in the reverse read file is also removed and vice versa. Here, 100 nt was used as the minimum tolerated length in FastqFilter. Demultiplexed and quality-filtered forward and reverse read files were then downloaded from Galaxy and long-term stored on an external hard drive.

3.5.3.2 Alignment of reads and clonotype clustering

Forward and reverse read files from Galaxy were further processed with the java-based software MiXCR (Bolotin et al. 2015). MiXCR was accessed via Terminal. Reads were aligned to TCRα or TCRβ template gene sequences, clustered to clonotypes of identical reads, and exported in tab-delimited format with the following command lines:

```
// 1. Align reads:
mixcr align --loci TRB --species hsa --report ALIGN_REPORT_samplename.txt
    -OminSumScore=258
    -OvParameters.geneFeatureToAlign={FR3End\(-26\):FR3End\(+3\)}
    -OcParameters.geneFeatureToAlign={CBegin:CBegin\(+20\)}
    -OjParameters.geneFeatureToAlign={FR4Begin\(-3\):FR4End}
    -OjParameters.parameters.floatingRightBound=false
    -OvParameters.parameters.minAlignmentLength=25
    -OjParameters.parameters.minAlignmentLength=25
    -OvParameters.parameters.absoluteMinScore=115
    -OjParameters.parameters.absoluteMinScore=125
    -OvParameters.parameters.relativeMinScore=0.8
    -OjParameters.parameters.relativeMinScore=0.8
    -OcParameters.parameters.relativeMinScore=0.8
    -OvParameters.parameters.maxHits=5 -OjParameters.parameters.maxHits=5
    -OvParameters.parameters.scoring.gapPenalty=-28
    -OjParameters.parameters.scoring.gapPenalty=-28 input_fwd_read.fastqsanger
    input_rev_read.fastqsanger align_samplename.vdjca

// 2. Cluster clonotypes:
mixcr assemble --report ASSEBMBLE_REPORT_samplename.txt
    -OassemblingFeatures=[CDR3] -OminimalClonalSequenceLength=6
    -OqualityAggregationType=Average -ObadQualityThreshold=20
    -OmaxBadPointsPercent=0 -OaddReadsCountOnClustering=true
    -OcloneClusteringParameters.searchDepth=2
    -OcloneClusteringParameters.searchParameters=oneMismatch
    -OcloneClusteringParameters.clusteringFilter.specificMutationProbability=1E-4
    align_samplename.vdjca clones_samplename.clns

// 3. Export clonotype tables:
mixcr exportClones --filter-out-of-frames --filter-stops -count -fraction
    -sequence -aaFeature CDR3 -vHit -vHitsWithScore -jHit -jHitsWithScore
    -dHit -cHit -lengthOf CDR3 -quality -minFeatureQuality CDR3
    clones_samplename.clns clones_samplename.txt
```

Forward and reverse read pairs were first aligned to the best fitting V-region, J-region and C-region of TCR template genes. Secondly, core clonotypes were assembled by grouping identical reads of good quality. Reads containing few low-quality nucleotides were then mapped to core clonotypes with similar sequences if possible. Furthermore, high-frequency parent clonotypes and highly similar clonotypes with significantly lower read counts were clustered, whereby only the information of the parent clonotype is kept. Thirdly, relevant clonotype information, such as read count, proportion of reads in sample, V and J gene use, CDR3 nt and amino acid sequence, and average CDR3 read quality, is extracted and exported to a tab-delimited .txt output file.

3.5.3.3 In-depth TCR repertoire analyses

All analyses described in the following sections were performed in R via the integrated development environment RStudio. The R packages used in this project are listed in Table 2.19. Custom R scipts were written for each analysis step.

3.5.3.4 Specific TCRβ clonotypes from single peptide stimulations

Specific TCRβ clonotypes for each CMV peptide were identified by comparing clonotype frequencies in three samples from the same donor: a sample *S* that was stimulated with the specific peptide of interest, a sample *C* that was stimulated with a control peptide, and a sample *U* derived from unstimulated PBMCs. To count as specific, TCRβ clonotypes were demanded to be enriched in *S* over *C* and in *S* over *U*, and to exceed a sample-specific read count (Figure 3.5). Let s_i, c_i and u_i be the proportion of reads (relative read frequency) of clonotype *i* in the three samples. TCRβ clonotypes must exceed two distinct enrichment cutoffs to count as specific (Figure 3.5B). The first enrichment cutoff was determined as a local minimum of a weighted density distribution of $\log_{10}(s_i/c_i)$ of all clonotypes *i* that met the requirement $s_i c_i > 10^{-6}$, i.e., of all medium- to high-frequency clonotypes. Analogously, the second enrichment cutoff was determined as a local minimum of a weighted density distribution of $\log_{10}(s_i/u_i)$ of all clonotypes *i* fulfilling the condition $s_i u_i > 10^{-7}$. In order to eliminate low-fidelity background signals, specific clonotypes were also required to exceed a specific sample read count cutoff (Figure 3.5C). This cutoff was identified by analysing the two density distributions of $\log_{10}(s_i)$ for all clonotypes *i* that were of low frequency in control samples (i.e., an absolute frequency of 1–10 reads in sample *C* or sample *U*, respectively).

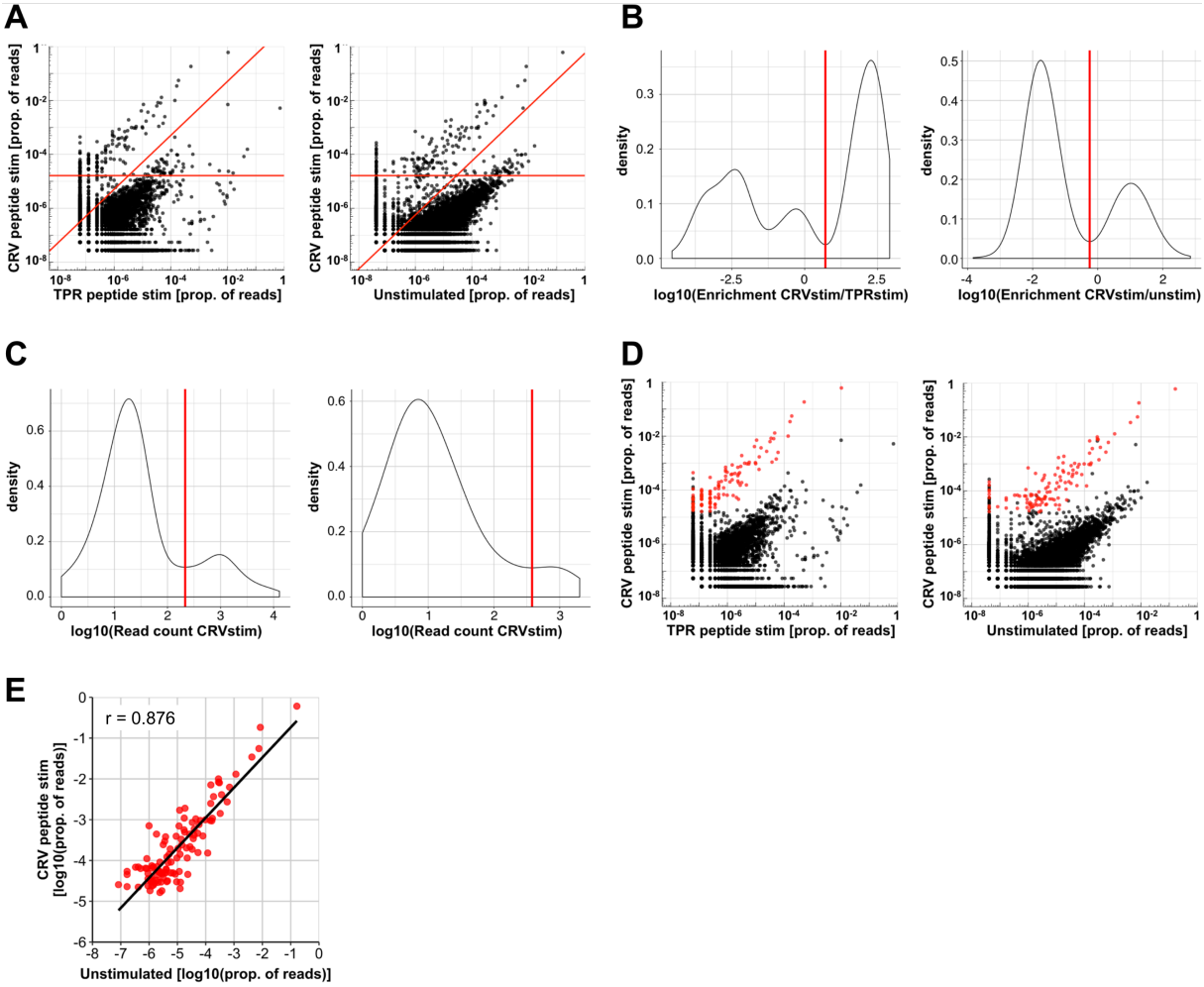


Figure 3.5: Criteria for identification of peptide-specific TCRβ clonotypes. As an example, representative data from donor P01 for identification of CRV-specific T cells are shown. Diagrams on the left relate to the binary comparison of the CRV-stimulated sample and the TPR-stimulated sample (control peptide). Diagrams on the right relate to the comparison of the CRV-stimulated sample and the unstimulated PBMC sample. Only TCRβs that fulfilled all criteria were categorised as CRV epitope-specific. In dot plots, a frequency of zero is plotted as a pseudofrequency of 0.5 reads. **(A)** Relative frequencies of TCRβ clonotypes. Red lines indicate the enrichment cut-off (diagonal line) and the specific sample read count cut-off (horizontal line). **(B)** The enrichment cut-off (red line) is defined as the local minimum of the weighted density distribution of the logarithm of clonotype enrichment. Clonotype enrichment is the ratio of clonotype frequencies in the CRV-stimulated sample and either control. **(C)** Read count cut-offs (red line) were defined as local minima of the distribution of the logarithms of clonotype counts, determined for clonotypes with a read count of 1 to 10 in the TPR-stimulated (left) or unstimulated (right) sample. The mean of these two cut-offs was used as the final specific sample read count cut-off. **(D)** Relative frequencies of TCRβ clonotypes as in A, but TCRβ clonotypes that fulfil all inclusion criteria for being CRV epitope-specific are shown in red. **(E)** Correlation of the log-transformed read frequencies of CRV-specific clonotypes in the unstimulated and CRV-stimulated sample, with Pearson correlation coefficient *r*. Figure and legend taken from Huth et al. 2019.

The read count at a local minimum of both two distributions was determined, and the mean of these two individual read counts served as the read cutoff value. If any of these density distribution did not have a local minimum, the cutoff was positioned at the global maximum of the density distribution $\times$ 100 (i.e., at $\geq$100 reads in sample *S*).

3.5.3.5 Antigen-specific TCRβ clonotypes from mini-LCL stimulations

Identification of antigen-specific TCRβ sequences from mini-LCL stimulations was achieved in a similar manner as described in Section 3.5.3.4: The frequency of each clonotype in the CMV antigen-stimulated sample ($\hat{=}s_i$) was compared with its *ex vivo* frequency ($\hat{=}u_i$) and with its frequency in the samples obtained by stimulation with an empty-vector mini-LCL control ($\hat{=}c_i$). Unlike for the single-peptide stimulation, a fixed enrichment value of 5-fold enrichment was demanded in the comparison of the CMV antigen-stimulated sample and the unstimulated sample: $s_i \geq 5 \times u_i$. Setting a fixed enrichment cutoff compared to the unstimulated sample was necessary, since populations of not-enriched and enriched TCRβ clonotypes were not as sharply separated as in the single peptide stimulation assay. This was likely due to the fact that a large number of T-cells, both CMV-specific and EBV-specific, were enriched by mini-LCL stimulation. Consequently, the minimum of the density distibutions in the mini-LCL–stimulated versus unstimulated samples could not be identified with the R function for analysis of single-peptide–stimulated samples. Another difference between the analyses was that the read cutoff was determined by comparing only read frequencies in the CMV antigen-stimulated and empty control mini-LCL–stimulated samples. Here, the read cutoff was set at 2$\times$ the absolute maximum of the logarithmic density distribution of clonotypes with 1–10 reads in the control sample. Only TCRβ sequences that were enriched compared with both control samples and exceeded the computed read count cutoff were considered specific for epitopes from the tested CMV antigen.

3.5.3.6 Specific TCRβ clonotypes from peptide pool stimulations

Epitope-specific TCRβ sequences from peptide pool stimulations were identified with a simplified analysis of TCRβ frequencies. For every peptide, the 5 pools that did not contain the peptide served as controls for the 3 pools in which the peptide was contained. All clonotypes that had at least 10-fold higher proportion of reads in any 3 pools compared to the remaining 5 pools were extracted from the MiXCR-processed

sequencing data. Of these, all clonotypes whose pool combinations matched one of the peptide addresses (Table 3.4) were categorised as specific for the given peptide. Here is an example: A TCRβ clonotype had ≥10× higher proportional read frequency in pools 4, 5, and 6 compared to pools 1, 2, 3, 7, and 8. This clonotype was assigned the specificity TPR, because TPR is contained in pools 4, 5, and 6. Peptide-specific TCRβ sequences identified in this manner were removed when they had <10 reads in any of the 3 pools matching the peptide address to exclude sequencing artefacts.

3.5.3.7 Shared or related epitope-specific TCRβ sequences

TCRβ sequences specific for a particular epitope were compared to find identical or highly similar TCRβ sequences shared between donors. For this purpose, all TCRβ sequences that were identical or differed in 1 amino acid from at least one other TCRβ sequences with the same CMV epitope specificity were isolated. This resulted in a set of shared and similar TCRβ sequences, which was further divided into TCRβ clusters which are present in multiple donors, and TCRβ cluster candidates, which are TCRβ chain variants within the same donor.

3.5.4 Data visualisation

VJ use of epitope-specific TCRs was visualised with chord diagrams using the circlize package in R. Scatter plots, heatmaps, and density plots comparing clonotype frequencies in different samples were plotted in R with ggplot2. Formation probabilities of shared or related and private epitope-specific TCRβ sequences were determined using the python-based OLGA algorithm (Sethna et al. 2019) and visualised in R. Bar plots and overlays were made with Microsoft Excel combined with GraphPad Prism or R. Multi-panel figures were assembled with Inkscape and Microsoft Powerpoint. Amino acid logos were made from CDR3 amino acid sequences of TCRβ families using Berkeley's WebLogo tool. Clusters of specific TCRβ sequences were made with R and Cytoscape.

4 RESULTS

4.1 High-resolution TCRβ profiling by short-term CMV peptide stimulation

In this PhD project, the CMV-specific TCR repertoires of individuals and cohorts were analysed in order to deepen our understanding of the human T-cell response against the virus. The focus of the project was on analysis of the TCRβ repertoire, because the TCRβ repertoire is generally more diverse than the TCRα repertoire (Murphy and Weaver 2016); thus the β-chain seemed more likely to provide precise markers of TCR specificity than the α-chain. Moreover, the TCRβ chain is a better determinant for T-cell clonotypes, since a single β-chain is expressed per T cell, whereas appoximately 25% of T-cells express 2 distinct functional α-chain sequences (Malissen et al. 1992; Murphy and Weaver 2016). To assess the CMV-specific TCR repertoire, a former post-doc in our group, Xiaoling Liang, established a protocol combining short term in vitro stimulation of PBMCs with synthetic peptide and high-throughput Illumina sequencing (Figure 4.1). First, virus-specific T cells were enriched by incubation of peptide-loaded PBMCs for 10 days in the presence of IL-2. Next, RNA was isolated from bulk peptide-stimulated cells and unstimulated PBMCs of the same donor and reversely transcribed to cDNA with a TCRβ gene-specific primer. In two subsequent PCR steps, a TCRβ gene-specific multiplex PCR and an adapter PCR, TCRβ DNA was amplified and nucleotide indices (barcodes) as well as the required primer binding sites and adapters for Illumina sequencing were appended to the amplicons. Raw sequencing data was quality-filtered and TCRβ clonotypes were built using the software MiXCR (Bolotin et al. 2015). TCRβ sequences that were enriched in the peptide-stimulated sample compared to both the unstimulated control sample and a sample stimulated with a different peptide (three-sample comparison) were identified to be specific for the tested CMV epitope (Section 3.5.3.4).

Initially, the single peptide stimulation assay was tested with CMV-positive donors that co-expressed HLA-B*07:02 and HLA-C*07:02, because these HLA alleles are highly frequent in persons of European descent (Gragert et al. 2013) and present some of the most immunogenic CMV epitopes known to date (Ameres et al. 2013; Kim et al. 2011; Weekes et al. 1999). In donors that additionally expressed HLA-A*02:01 or HLA-

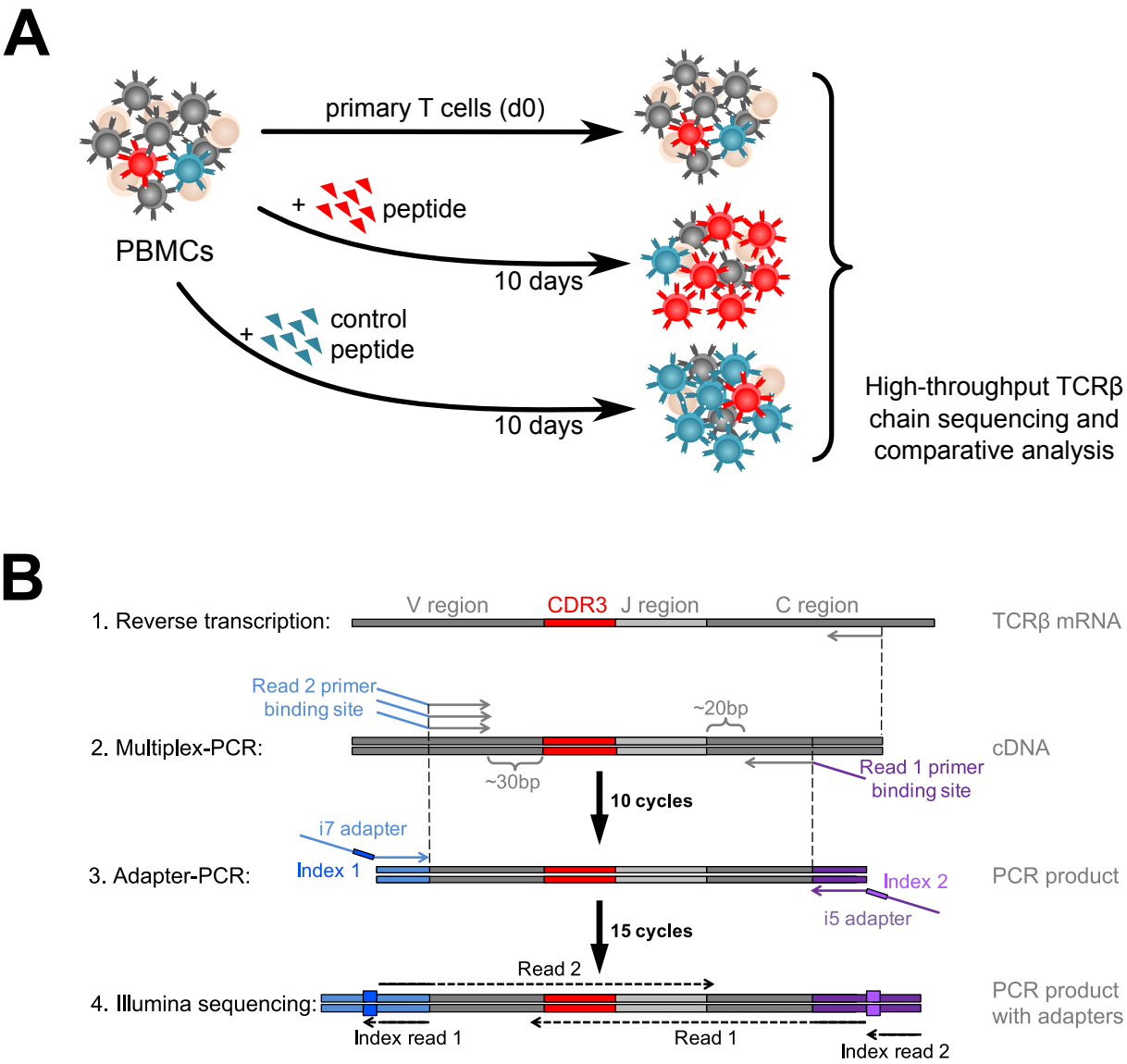


Figure 4.1: Experimental setup. (**A**) Schema of the peptide stimulation assay and three-sample comparison for expansion and analysis of CMV-specific T cells. PBMCs were isolated from peripheral blood of healthy donors, loaded with single immunogenic CMV peptides, and cultured for 10 days with IL-2. Cells before and after stimulation were lysed, and TCRβ libraries were prepared from bulk RNA and high-throughput–sequenced. Specific TCRβ sequences for each epitope were identified by comparing TCRβ clonotype frequencies in three samples. These samples were 1) stimulated with specific peptide, 2) stimulated with control peptide, and 3) not stimulated. Clonotypes that were enriched by stimulation with the specific peptide, but not in controls, were considered specific. (**B**) Preparation of TCRβ libraries for paired-end sequencing of the CDR3. After total RNA isolation, TCRβ RNA was reversely transcribed using a Cβ gene-specific primer (1). In the first PCR step (2), cDNA was amplified by semi-multiplexed PCR with a mixture of 45 forward primers that covered all Vβ genes. All forward primers had identical overhangs that served to append the Illumina sequencing read 2 priming site to the product. The reverse primer was complementary to both Cβ genes and appended the Illumina sequencing read 1 priming site to the amplicon. In the second PCR step (3), a single primer was used on each side to add Illumina i5 and i7 adapters and sample indices (barcodes) for multiplexed Illumina sequencing (4). Figure taken from Huth et al. 2019.

B*35:01, CMV epitopes restricted by these HLA alleles (Diamond et al. 1997; Khan et al. 2002a; Gavin et al. 1993) were also included. T cells from 9 CMV-positive donors (P01–P08 and P14) were stimulated with 1 of 7 HLA-matched CMV peptides. Of the 7 peptides, 2 were HLA-A*02:01–restricted (NLV and VLE), 2 were HLA-B*07:02–restricted (RPH and TPR), 1 was HLA-B*35:01–restricted (IPS), and 2 were HLA-C*07:02–restricted (CRV and FRC). Comprehensive lists of all tested donors with their HLA type and of all CMV-derived peptides used in this project can be found in Table 2.14 and Table 2.13, respectively.

4.2 Peptide stimulation increases the RNA content per T cell

Cell were counted in samples before and after stimulation with a CMV-derived peptide. After 10 days of co-culture with CMV peptide and IL-2, absolute cell numbers were often lower than the amount of initially seeded cells in each stimulation (Figure 4.2A). Cells from 5 of 9 donors reacted strongly to stimulation with FRC, as demonstrated by the 1.25-fold to 2.1-fold increase in cell number at harvest. Stimulation with IPS and TPR also led to an increase in cell number in 1 of 1 donor and 1 of 9 donors,

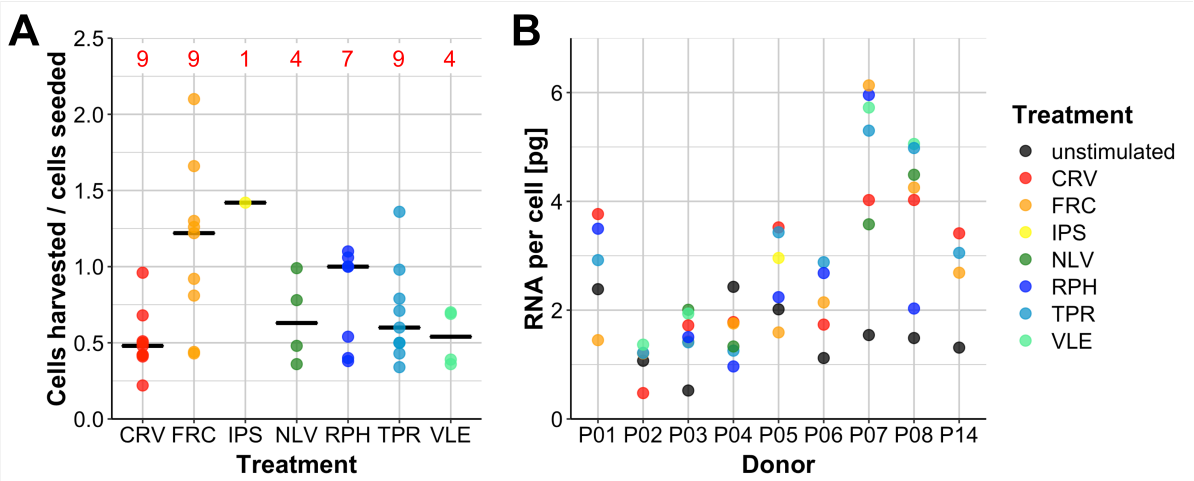


Figure 4.2: Change in cell number and RNA content per cell in PBMCs from 9 donors stimulated with 1 of 7 HLA-matched CMV peptides. (**A**) Ratios of the number of cells harvested after stimulation divided by the cell number seeded. Data were grouped by the CMV-derived peptide with which the cells were stimulated. Red digits indicate the number of donors that were treated with each peptide and black lines show the median of each data subset. (**B**) Average RNA content per cell before and after stimulation grouped by donor. The average amount of RNA obtained per cell is shown for all peptides with which the donor's cells were treated.

respectively. RPH-stimulation marginally increased the cell number in samples of 2 of 7 donors. Cells stimulated with CRV, NLV, or VLE always decreased in number upon peptide stimulation. To this point, it is unclear why changes in cell numbers due to 10-day stimulation were different between donors and peptides. Possible reasons include different strengths of the initiated CMV-specific T-cell responses, different proportion of epitope-specific T cells in PBMCs, and different proliferative potential of activated T cells.

Total RNA was isolated from unstimulated and peptide-stimulated PBMCs. The average RNA content per lysed cell was between 0.5 pg and 6 pg (Figure 4.2B). Unstimulated samples had RNA yields below 2.5 pg per cell. In 8 of 9 donors, the RNA content in the unstimulated sample was the lowest or second lowest of all samples. By contrast, the highest RNA yield per cell was obtained from the unstimulated sample in donor P04. When comparing the RNA contents per cell after stimulation with different peptides, TPR–stimulated and VLE–stimulated cells were often among the samples with the highest RNA yields of a donor. Interestingly, CRV–stimulated cells harboured either one of the highest or lowest amounts of RNA compared to differently treated samples of a donor. In general, there was a positive linear correlation between the input cell number and output RNA yield of unstimulated samples (Pearson correlation coefficient r=0.66; Figure 4.3) and peptide-stimulated samples (r=0.51). CMV

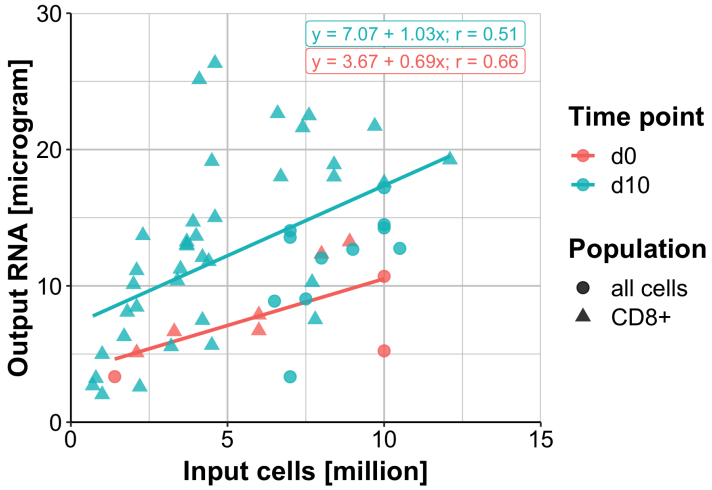


Figure 4.3: Comparison of RNA content in samples before and after stimulation with CMV peptide. The plot shows total RNA yields of unstimulated and peptide-stimulated samples as a function of the input cell numbers. CD8-enriched samples are represented by triangles and unsorted samples by circles. Pearson correlation coefficients r were calculated for both time points.

peptide-stimulated cells contained on average 1.75-fold more RNA than unstimulated PBMCs and were slightly larger in size as observed under the miscroscope. Total RNA yields were lower when cell samples were not CD8-enriched by MACS prior to lysis and RNA isolation. A measurable increase in average RNA content per cell was anticipated, because CMV-specific T cells are highly frequent in the blood of virus carriers (Gillespie et al. 2000; Sylwester et al. 2005; Hosie et al. 2017) and because activated T cells upregulate TCR expression approximately a day after activation (Paillard et al. 1990). Since only $CD8^+$ T-cell epitopes were used to stimulate CMV-specific T cells, it was expected that RNA content in CD8-enriched samples was higher than in unsorted samples which contain higher proportions of unactivated cells.

4.3 Peptide stimulation enriches CMV multimer-positive T cells

Unstimulated PBMCs and peptide-stimulated cells were stained with peptide-MHC multimers as described in Section 3.4.5 in order to determine the proportions of epitope-specific T cells. Total events were first gated on lymphocytes, then single cells, and finally $CD3^+$ T cells as explained in Figure 3.4. KIR2DL2/3-positive events were excluded in stainings with HLA-C multimers, because the NK-cell receptor KIR2DL2/3 can also bind to HLA-C, which may lead to false-positive events. The proportions of multimer-positive and thus peptide-MHC–specific T cells were calculated based on the number of $CD3^+$ lymphocytes. CMV-positive donors had substantial amounts of CMV-specific T cells in their blood, while such T cells were rare or absent in CMV-negative donors (Figure 4.4). Stimulation with HLA-matched peptides increased the percentage of epitope-specific T cells in samples derived from CMV-positive donors, but not in samples from CMV-negative donors. In the flow cytometry data analyses presented here, background signals determined by control stainings without multimers were subtracted from the proportion of multimer-positive T cells. Table 4.1 shows the proportions of multimer-positive T cells in CMV-positive donors P01–P08 and P14 ex vivo (d0) and after 10-day stimulation with the corresponding CMV peptide (d10). Proportions of T cells specific for the tested peptides in ex vivo PBMCs varied between donors. On average, FRC-specific T cells were most frequent in the ex vivo repertoire of CMV-positive donors followed by CRV-specific T cells, with geometric means of 0.5%

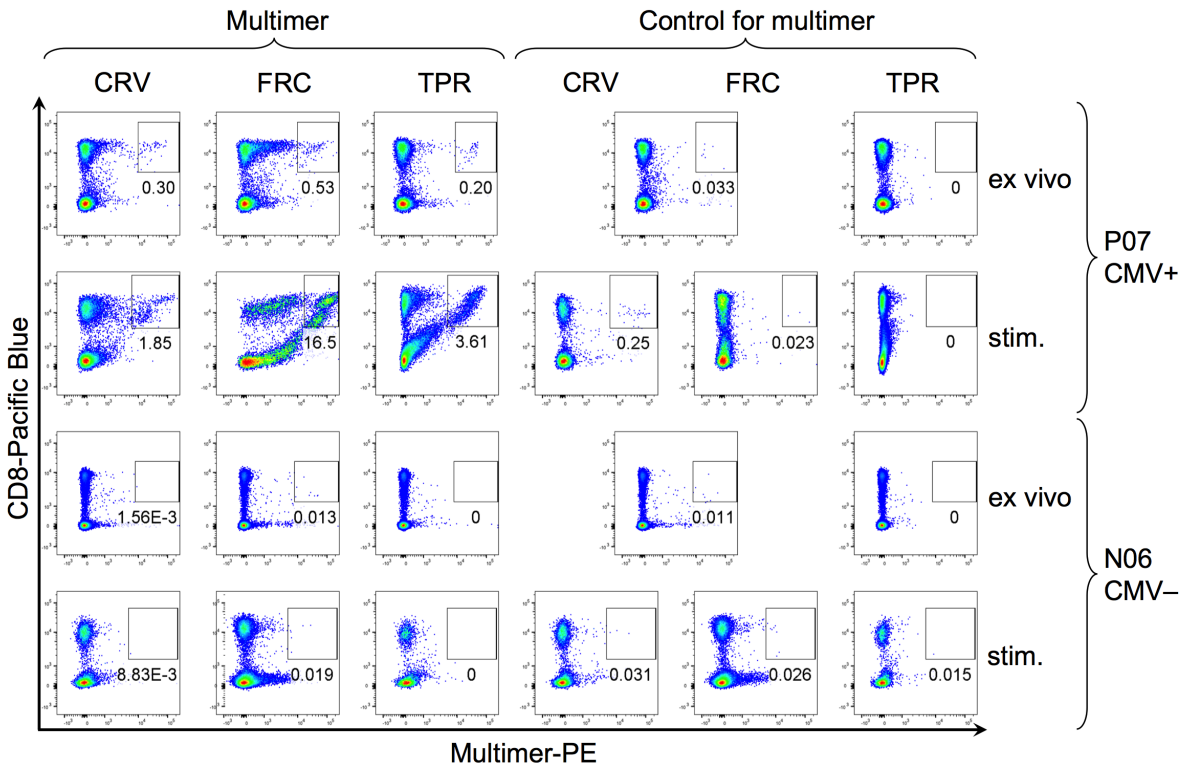


Figure 4.4: Frequency of multimer-positive T cells in PBMCs (ex vivo) and samples stimulated with single CMV peptides from a CMV-positive donor (P07) and a CMV-negative donor (N06). The figure shows representative flow cytometry plots of samples stained with streptamers CRV/HLA-C*07:02 or FRC/HLA-C*07:02, or pentamer TPR/HLA-B*07:02. The controls for streptamer- or pentamer-treated samples were treated with StrepTactin-PE or Fluorotag-PE, respectively, in the absence of peptide-MHC monomer or multimer. Since the ex vivo samples were all derived from the same starting material, one control was sufficient for both streptamer stainings (CRV and FRC).

and 0.23%. Ex vivo frequencies of RPH- and TPR-specific T cells varied the most and ranged from 0.0059% to 0.6% and from 0.0016% to 1.6%, respectively. Surprisingly, levels of VLE-specific T cells were very similar in ex vivo samples from 3 of 4 donors. The differences in ex vivo frequencies of CMV-specific T cells between donors were even more apparent in a plot showing the proportions of multimer-positive T cells per epitope for each donor (Figure 4.5A). While CRV-specific and FRC-specific T cells were frequent across donors, proportions of RPH-specific and TPR-specific T cells were high in some donors, but very low in others. Next, the relative enrichment of CMV epitope-specific T cells by peptide stimulation was calculated by dividing the proportion of specific T cells in the stimulated sample by their proportion in the ex vivo sample (d10 ÷ d0). Average enrichment of epitope-specific T cells was strikingly sim-

Table 4.1: Proportions of specific T cells in unstimulated (d0) and peptide-stimulated (d10) samples of donors P01–P08 and P14 in percent [%] of $CD3^+$ T cells, as measured by flow cytometry, and calculated T-cell enrichment. Approximately 60% of stainings were performed by Xiaoling Liang and reanalysed in this thesis. Some stainings are missing because the FRC/HLA-C*07:02-multimer was not available at the time of staining, or because the stimulation was not done; they are marked by n.d. (”not done”). Blank cells indicate that no staining was done because the donor does not express the relevant HLA for the peptide. Geometric means of the T-cell proportions and the enrichment (d10÷d0) are noted in the bottom row. Mean values of IPS stainings are not representative, since only one donor was analysed. Values <100 were rounded to 2 significant digits; values ≥ 100 were rounded to 0 decimal places. Enrichment was calculated from non-rounded frequency values.

Donor	CRV			FRC			IPS			NLV		
	d0 [%]	d10 [%]	Enrichment	d0 [%]	d10 [%]	Enrichment	d0 [%]	d10 [%]	Enrichment	d0 [%]	d10 [%]	Enrichment
P01	0.20	30	150	0.36	6.8	19						
P02	2.4	3.7	1.5	2.8	32	11				0.76	11	14
P03	0.039	3.6	93	0.13	4.8	37				0.11	2.6	23
P04	0.83	24	29	n.d.	n.d.	-				0.0031	0.17	55
P05	0.27	4.4	16	n.d.	n.d.	-	0.19	36	192			
P06	0.061	0.58	9.4	n.d.	n.d.	-						
P07	0.27	1.6	6.0	0.50	16	33				0.24	3.7	15
P08	0.92	5.0	5.5	n.d.	n.d.	-				0	0.82	-
P14	0.030	0.027	0.90	n.d.	n.d.	-						
Geometric mean	**0.23**	**2.6**	**11**	**0.50**	**11**	**23**	**0.19**	**36**	**192**	**0.090**	**1.7**	**23**

Donor	RPH			TPR			VLE		
	d0 [%]	d10 [%]	Enrichment	d0 [%]	d10 [%]	Enrichment	d0 [%]	d10 [%]	Enrichment
P01	0.011	4.6	419	0.11	26	238			
P02	0.16	0.42	2.6	0.18	16	91	0.36	3.3	9.1
P03	0.084	3.3	39	0.025	1.0	40	0.28	2.3	8.1
P04	0.011	0.66	59	0.0062	5.7	912	0.0029	n.d.	-
P05	0.60	5.4	9.0	0.0016	0.000060	0.040			
P06	0.0059	0.22	38	0.18	4.2	24			
P07	0.10	1.6	16	0.20	3.6	18	0.33	6.2	19
P08	0.19	4.8	25	0.29	4.0	14	0	0.040	-
P14	0.50	1.9	3.7	1.6	1.3	0.79			
Geometric mean	**0.073**	**1.6**	**22**	**0.076**	**1.3**	**17**	**0.099**	**1.2**	**11**

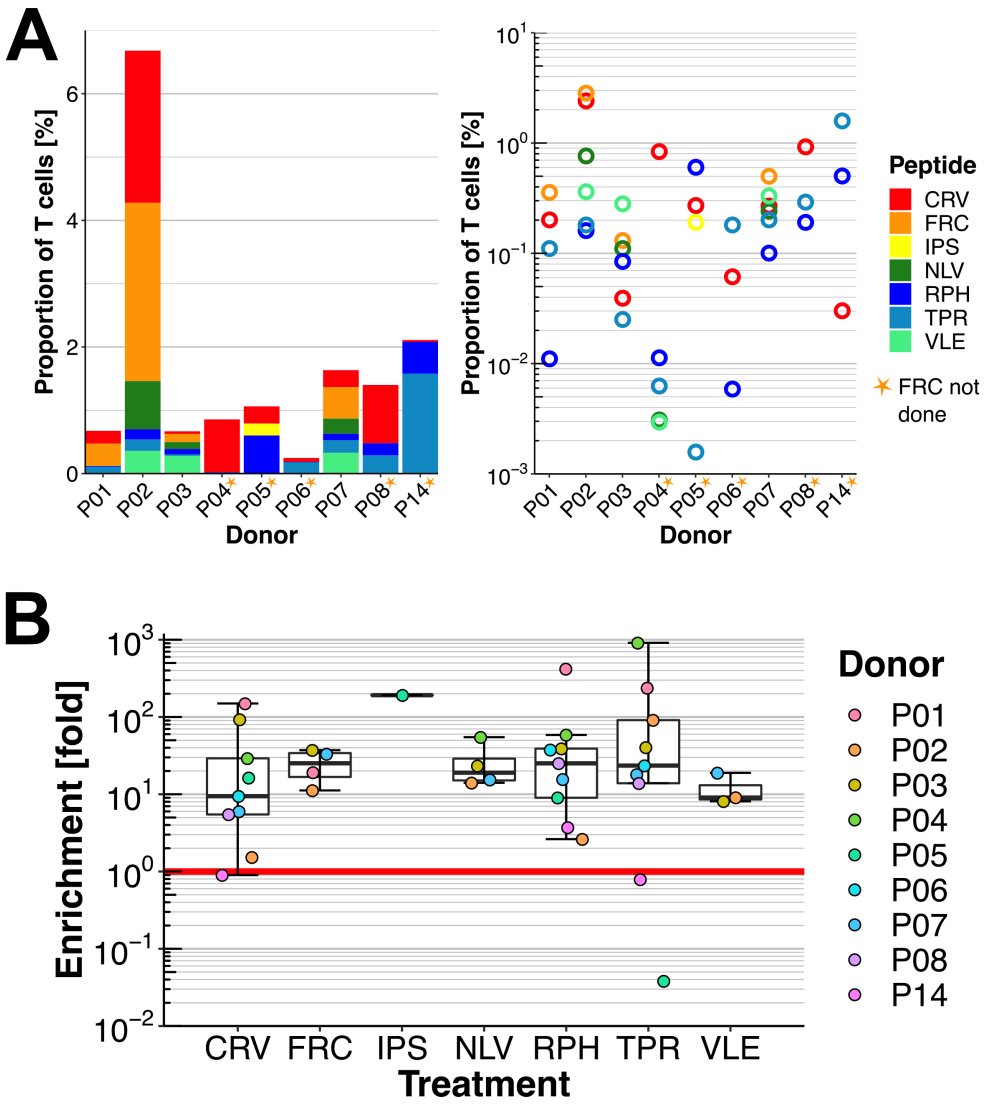


Figure 4.5: Ex vivo frequencies and enrichment of CMV-specific T cells in samples of donors P01–P08 and P14 as determined by flow cytometry. Background signal measured in the negative controls was subtracted from the proportions of positive T cells in each sample. Samples derived from donors P04, P05, P06, P08, and P14 were not stained with FRC streptamer, since it was not available at the time. (**A**) Proportion of T cells that are specific for the tested epitopes in ex vivo samples. The proportions of positive T cells were plotted on a linear y-axis (bar chart) and a logarithmic y-axis (dot plot). (**B**) Relative enrichment of peptide-specific T cells by 10-day stimulation. Enrichment was calculated by dividing the proportion of positive T cells after stimulation by the ex vivo proportion of peptide-specific T cells. The red line marks no change in proportion of positive T cells by peptide stimulation. Bottom and top box ends represent the lower quartile (Q1) and upper quartile (Q3), the horizontal line is the median (Q2), and the whiskers extend to highest value and the lowest value of each data subset. Values that are $\leq$Q1–1.5$\times$the interquartile range (IQR; IQR=Q3-Q1) and values that are $\geq$Q3+1.5$\times$IQR are considered to be outliers. Outliers are ignored when determining whisker endpoints. NLV- and VLE-specific T cells were absent in the ex vivo sample of donor P08 and are not depicted in the plots. VLE-staining was not done for VLE-stimulated cells in donor P04 and is therefore not depicted.

ilar for all epitopes except IPS, which is not representative, because only one donor was tested (Figure 4.5B). Enrichment by stimulation with FRC, NLV, or VLE was similiar across donors, while stimulation with CRV, RPH, or TPR induced more variable enrichment depending on the donor. The geometric means of enrichment by peptide stimulation were between 11 (CRV & VLE) and 23 (FRC & NLV). Despite the similar average enrichment of epitope-specific T cells, the enrichments observed for individual donors were rather diverse. For instance, enrichment of TPR-specific T cells by peptide stimulation varied between 0.04-fold and 912-fold. Donor P05 had extremely low initial frequency of TPR-specific T cells, and the frequency of such T cells was even lower after stimulation, indicating that this donor may not mobilise T cells against CMV epitope TPR at all. In conclusion, CMV-specific T cells were abundant in PBMCs and their proportion was enriched by stimulation with most CMV-derived peptides in all donors.

4.4 Peptide stimulation identifies CMV peptide-specific TCRβ clonotypes

After TCRβ library preparation and high-throughput sequencing, raw data were processed as described in Section 3.5.3. The software MiXCR returned TCRβ repertoire files with aligned and clustered clonotypes. In the MiXCR output files, the read count and the proportion of reads on the total number of reads for each clonotype were listed alongside their V and J gene usage and CDR3 nucleotide and amino acid sequences. The TCRβ clonotype files were further analysed in R using custom scripts that I wrote for this project. Table A1 in the appendix lists all single peptide stimulation samples that were used to identify CMV-specific TCRβ sequences. The median productive read count per sample was 5.1×10^6 reads, and the median number of TCRβ clonotypes was 128115.

4.4.1 Clonotype distribution in TCRβ repertoires is influenced by stochastic effects

To begin with, the frequency distribution of TCRβ clonotypes in unstimulated and CMV peptide-stimulated samples was examined. Low-frequency clonotypes were far more abundant than clonotypes with high read counts. The vast majority of TCRβ clonotypes identified in unstimulated or stimulated samples were detected at absolute frequencies

of less than 10 reads. For example, a total of 2745103 clonotypes were identified in unstimulated and peptide-stimulated samples of donor P03 (Figure 4.6A), whereby clonotypes that were found in multiple samples were counted multiple times. Of these 2745103 clonotypes, 2139 or 0.078% of clonotypes were found with at least 1000 reads in their corresponding sample. Conversely, 2358163 clonotypes, or 85.9% of clonotypes, were found with less than 10 reads in their corresponding sample. Low-

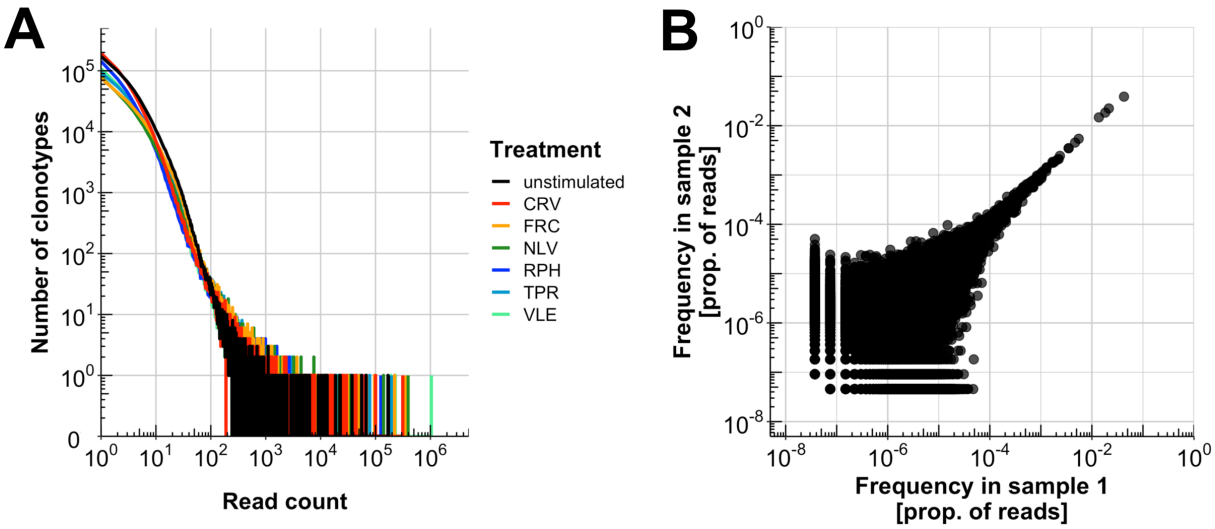


Figure 4.6: Clonotype distribution in TCRβ repertoires. (**A**) Frequency of clonotypes with a particular read count in unstimulated and stimulated samples of CMV-positive donor P03. The higher the number of reads was, the fewer clonotypes had this read count. (**B**) Biological replicates of ex vivo PBMCs derived from donor P03. Each TCRβ clonotype is defined as the entirety of identical reads on the nucleotide level and represented by a black dot. Clonotypes that were undetectable in one sample were assigned a pseudo-frequency corresponding to 0.5 reads to enable their display on a logarithmic axis. Two separate TCRβ library preparations were carried out starting from the same PBMC sample. Frequency variation of a TCRβ clonotype increased with decreasing relative frequency of a clonotype in either sample.

frequency clonotypes were not only highly abundant in number, but their relative frequency values were less stable between biological replicates than the relative frequencies of high-frequency clonotypes. Figure 4.6B compares the relative frequencies of TCRβ clonotypes in two biological replicates derived from the same PBMC sample of donor P03. Before cell lysis, the PBMC sample was split into two distinct samples. RNA isolation, TCRβ library preparation, and high-throughput sequencing were carried out independently for both replicates. While the relative frequency of the most frequent TCRβ clonotypes was similar in the replicates, the relative frequency varied more strongly the lower the TCRβ clonotype frequency was in one of the replicates.

This led to a broader clonotype distribution at the lower frequency end and impacted all TCRβ repertoire comparisons between samples dealt with in this project. The distortion of high-throughput sequencing data by PCR has been investigated both empirically and theoretically in a recent study (Kebschull and Zador 2015). The researchers analysed the impact of PCR bias, stochasticity, template switching, and polymerase errors on the representation of DNA templates in the sequenced samples. They found that stochastic effects contribute the most to skewed high-throughput sequencing data and that this skewing was most significant for low copy numbers. Since PCR is not a perfect process, not all templates are amplified in each PCR cycle. This affects low-frequency TCRβ sequences more severely than high-frequency TCRβ sequences because even if some cDNA copies from high-frequency TCRβ clonotypes are not amplified in a cycle, the majority of them still is. By contrast, if a TCRβ clonotype with only one initial cDNA copy is not amplified in a cycle, its relative frequency decreases by approximately 50% and it is further disadvantaged in the next cycle of amplification. In addition to the stochastic effects on low-frequency sequences, they are also more prone to sampling effects during Illumina sequencing. The stochastic effects on TCRβ repertoire data were considered in all analyses presented in this thesis. For example, read frequency cutoffs were introduced when samples were compared to each other to reduce indicences of false-positive peptide-specific TCRβ clonotypes (e.g. Figure 3.5).

4.4.2 Clusters of enriched TCRβ clonotypes emerge after stimulation of PBMCs from CMV-positive donors

When plotting TCRβ clonotype frequencies before and after 10-day peptide stimulation, distinct clusters of specifically enriched clonotypes are present in CMV-positive, but not in CMV-negative donors. Figure 4.7 exemplarily shows such scatter plots for CMV-positive donor P01 and CMV-negative donor N05 treated with peptides CRV, FRC, RPH, or TPR. The stochastic variation and consequently wider distribution of low-frequency clonotypes explained in the previous section is visible in all scatter plots. Apart from this, a proportional effect influenced the TCRβ clonotype frequency distribution. This effect became apparent when one or more highly frequent TCRβ clonotypes dominated the sample. In the CRV-stimulated and TPR-stimulated samples of donor P01 (Figure 4.7), the dominant TCRβ clonotype takes up 60.5% and 71.6% of all reads. This leaves less reads for the remaining TCRβ clonotypes of the sample,

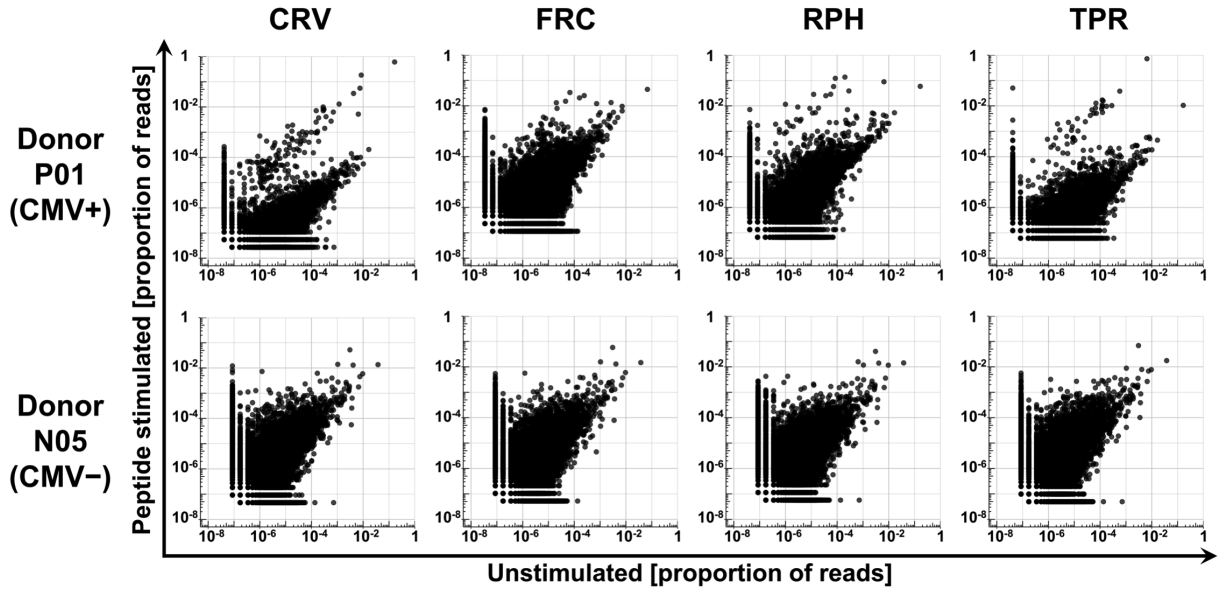


Figure 4.7: Relative frequencies (proportion of reads) of TCRβ clonotypes before (x-axis) and after (y-axis) stimulation with one of four CMV peptides in CMV-positive donor P01 (upper panel) and CMV-negative donor N05 (lower panel). Each TCRβ clonotype is defined as the entirety of identical reads on the nucleotide level and represented by a black dot. Clonotypes that were undetectable in one condition were assigned a pseudo-frequency corresponding to 0.5 reads to enable their display on a logarithmic axis.

which in turn leads to a seemingly lower enrichment of all expanded TCRβ clonotypes compared to the ex vivo sample. As a result, the enrichment, which was defined as the ratio of the frequency after stimulation divided by the frequency before stimulation, becomes smaller when the frequency of responding clonotypes is relatively high ex vivo. However, this does not impair the discriminatory power of the experiment, since the corresponding ratio of frequencies of non-responding clonotypes becomes smaller as well. Because of the strong variation in enrichment values of enriched TCRβ clonotype clusters obtained for different donors or peptides, no uniform enrichment cutoff, such as at least 10-fold enrichment, could be applied to all samples in order to identify epitope-specific TCRβ clonotype. Hence, a dynamic discriminator was required to identify CMV epitope-specific TCRβ clonotypes for each three-sample comparison individually. This dynamic discriminator was implemented as a custom R script that automatedly searched for the minima of TCRβ clonotype distributions between the responder and non-responder clonotypes (see Section 3.5.3.4 and Figure 3.5). Like this, enriched clusters of TCRβ sequences were identified and the corresponding specific TCRβ sequences were extracted from the data.

4.4.3 High-throughput sequencing identifies TCRβ clonotypes of various CMV epitope specificities

Epitope-specific TCRβ clonotypes were identified as described in Section 3.5.3.4 and Figure 3.5. The computed cutoff values that were required for TCRβ clonotypes of a particular sample to qualify as CMV epitope-specific are listed in Table A2 in the appendix. All in all, 1440 specifically enriched TCRβ sequences were identified from 9 CMV-positive donors P01–P08 and P14. Importantly, some of these 1440 specific TCRβ sequences were identical on the amino acid level within a donor's repertoire or even identical on the nucleotide level between donors (donor-nucleotide-unique). In this section, the specific TCRβ repertoires of the 9 donors are analysed individually and no comparisons between TCRβ sequences of different donors are made. 90 TCRβ sequences were assigned to more than one epitope specificity in the same donor. They were excluded from the following analyses, leaving 1350 monospecific TCRβ sequences specific for one of the 7 tested CMV-derived epitopes (Table 4.2). The table shows how many nucleotide-unique specific TCRβ clonotypes were found

Table 4.2: Number of nucleotide-unique monospecific TCRβ clonotypes per CMV epitope and donor.

Donor	CRV	FRC	IPS	NLV	RPH	TPR	VLE	Sum	Mean
P01	119	24			27	39		**209**	**52.3**
P02	31	25		n.d.	n.d.	38	5	**99**	**24.8**
P03	39	16		8	15	22	15	**115**	**19.2**
P04	45	13		12	16	20	n.d.	**106**	**21.2**
P05	118	99	26		63	13		**319**	**63.8**
P06	49	36			21	21		**127**	**31.8**
P07	33	39		20	30	9	10	**141**	**23.5**
P08	25	18		26	18	12	2	**101**	**16.8**
P14	26	52			n.d.	55		**133**	**44.3**
Sum	**485**	**322**	**26**	**66**	**190**	**229**	**32**	**Total TCRβs = 1350**	
Mean	**53.9**	**35.8**	**26**	**16.5**	**27.1**	**25.4**	**8**	**Overall mean = 31.4**	

in a donor for each epitope. On average, each single-peptide stimulation lead to the identification of 31.4 CMV-specific TCRβ sequences. Most specific TCRβ sequences were found after stimulation with CRV, namely between 25 and 119 TCRβ sequences per donor. The lowest number of specific TCRβ sequences was found after stimulation with VLE, where 2–15 TCRβ sequences were specifically enriched. Donor P05 had a particularly diverse T cell response against the 5 tested CMV epitopes; 319 specific

TCRβ sequences were discovered in samples derived from this donor. The remaining donors mobilised between 99 and 209 TCRβ clonotypes against 3–6 tested epitopes.

4.4.4 Peptide stimulation enriches CMV-specific TCRβ clonotypes

In Section 4.3, frequency and enrichment of CMV-specific T cells was analysed on the cellular level using flow cytometry. Next, this analysis was carried to the molecular level: The cumulative proportions and the enrichment values of specific TCRβ sequences in the ex vivo and peptide-stimulated samples were computed. For this purpose, the read frequencies of all TCRβ sequences that were monospecific for a given CMV epitope in a donor were added together, and the percentages of these reads among total reads in ex vivo and peptide-stimulated samples of this donor was obtained (Table 4.3). Overall enrichment of specific TCRβ sequences was calculated by dividing the frequency of all specific TCRβ sequences in the relevant peptide-stimulated sample by their frequency in the unstimulated sample. Excluding IPS, which was analysed in only one donor, TCRβ sequences against HLA-C–restricted epitopes CRV and FRC had, on average, the highest mean frequencies, both ex vivo and in stimulated samples. CRV-specific TCRβ sequences accounted for an average 2% of reads in the unstimulated and 47% in the peptide-stimulated samples. FRC-specific TCRβ sequences had mean frequencies of 0.89% in the unstimulated and 50% in the peptide-stimulated sample. Ex vivo frequencies of CRV-specific TCRβ sequences varied strongly between donors, and ranged from 0.068% in donor P03 to 19% in donor P01. In 4 out of 9 donors, the percentages of specific TCRβ sequences in the CRV-stimulated samples exceeded 90%. Importantly, the proportions of specific TCRβ sequences presented in this section for donors P02 and P03 stem from unsorted samples, whereas the samples from all other donors were CD8-enriched before sequencing. This means that the frequency of CMV epitope-specific TCRβ sequences will be higher in CD8-enriched samples from donors P02 and P03. Nevertheless, the unsorted samples were analysed here because they were used to identify specific TCRβ sequences. A comparison of ex vivo frequencies of CMV epitope-specific TCRβ sequences in unsorted ex vivo samples of all donors can be found later on (Section 4.4.9). TCRβ sequences targeting HLA-A*02:01-restricted epitopes NLV and VLE had the lowest median frequencies in ex vivo samples (0.081% and 0.066%, respectively). Among HLA-matched donors, the ex vivo

Table 4.3: Proportions of reads that CMV-specific TCRβ sequences hold in unstimulated (d0) and peptide-stimulated (d10) samples of donors P01–P08 and P14 in percent [%] of total reads and overall enrichment of specific TCRβ sequences. Approximately 66% of stimulations were performed by Xiaoling Liang; all samples were reanalysed in this thesis. Geometric means of the read proportions and the enrichment (d10÷d0) are noted in the bottom row. Mean values of IPS-specific TCRβ sequences are not representative since only one donor was analysed. Values <100 were rounded to 2 significant digits; values ≥100 were rounded to 0 decimal places. Enrichment was calculated from non-rounded frequency values.

Donor	CRV			FRC			IPS			NLV		
	d0 [%]	d10 [%]	Enrichment	d0 [%]	d10 [%]	Enrichment	d0 [%]	d10 [%]	Enrichment	d0 [%]	d10 [%]	Enrichment
P01	19	97	5.1	0.073	11*	147						
P02	1.2*	16*	13	0.44*	46*	105				n.d.	n.d.	-
P03	0.068*	14*	211	0.13*	12*	95				0.059*	13*	225
P04	4.0	87	22	0.19	70	363				0.020	9.7	483
P05	27	95	3.6	5.6	91	16	0.70	63	89			
P06	1.5	91	63	2.4	86	36						
P07	6.0	81	13	0.83	71	86				0.50	69	139
P08	0.17	12	70	9.5	94	9.9				0.075	10	135
P14	1.7	67	40	4.3	89	21						
Geometric mean	**2.0**	**47**	**24**	**0.89**	**50**	**56**	**0.70**	**63**	**89**	**0.081**	**17**	**213**

Donor	RPH			TPR			VLE		
	d0 [%]	d10 [%]	Enrichment	d0 [%]	d10 [%]	Enrichment	d0 [%]	d10 [%]	Enrichment
P01	0.054	47	877	0.85	94	111			
P02	n.d.	n.d.	-	0.22*	50*	226	0.043*	16*	379
P03	0.016*	7.7*	477	0.064*	11*	165	1.7*	34*	20
P04	0.047	9.3	196	0.21	42	206	n.d.	n.d.	-
P05	1.2	47	41	0.0041	0.63	153			
P06	0.22	17	81	0.33	63	193			
P07	0.38	48	128	1.6	69	43	0.42	44	103
P08	0.26	34	128	0.96	45	47	0.00060	0.092	154
P14	n.d.	n.d.	-	0.013	5.2	399			
Geometric mean	**0.14**	**24**	**172**	**0.16**	**22**	**140**	**0.066**	**6.9**	**105**

* Samples were not CD8-enriched before RNA isolation and TCRβ sequencing library preparation.

frequencies of VLE-specific TCRβ sequences differed by more than 3 orders of magnitude, whereas the lowest and highest cumulative frequencies of NLV-specific TCRβ sequences, 0.059% and 0.5%, were within the same order of magnitude. Compared to the highly frequent HLA-C–restricted and the lower frequency HLA-A–restricted TCRβ sequences, HLA-B*07:02–restricted TCRβ sequences were of intermediate ex vivo frequency: RPH-specific TCRβ sequences had a mean frequency of 0.14% and TPR-specific TCRβ sequences had a mean frequency of 0.16%.

Cumulative ex vivo frequencies of CMV-specific TCRβ sequences were computed by adding together the relative read frequencies of CMV epitope-specific TCRβ of each individual donor. Cumulative frequencies of TCRβ clonotypes specific for any tested CMV epitope ranged from 2% in the unsorted samples of donors P02 and P03 and 4.4%–4.5% in the CD8-enriched samples of donors P04 and P06 to maximum 34% in donor P05 (Figure 4.8A). In 8 out of 9 donors, HLA-C*07:02–restricted CRV-specific or FRC-specific TCRβ sequences were the most abundant specificity in ex vivo samples. Only the ex vivo sample of donor P03 was dominated by VLE-specific TCRβ sequences. Enrichment of epitope-specific TCRβ sequences was the highest through NLV-stimulation and varied between 135-fold and 483-fold in the HLA-A*02:01–expressing donors; the mean enrichment was approximately 200-fold. The lowest enrichment values were observed for CRV-stimulated samples with a mean enrichment of 24-fold across all donors. Median TCRβ enrichment values were in a similar range for all epitopes but CRV and FRC (Figure 4.8B). The enrichment values for epitopes CRV and FRC were spread the most between donors. When comparing enrichment values of specific TCRβ clonotypes from different donors or peptide stimulations, two aspects need to be considered. Firstly, unlike the T-cell enrichment values measured by flow cytometry, enrichment values from TCRβ sequencing presented here are not independent from the method to identify CMV-specific TCRs, since enrichment compared to the unstimulated sample was a criterion for peptide specificity of TCRβ clonotypes. Secondly, as discussed above, the presence of high frequency CMV-specific TCRβ clonotypes reduces the reads leads to lower relative read frequencies of the non-responding TCRβ clonotypes. Consequently, the calculated enrichment of TCRβ clonotypes compared to their ex vivo frequency may be lower, but at the same time the enrichment of the CMV-specific TCRβ clonotypes compared to the non-responding

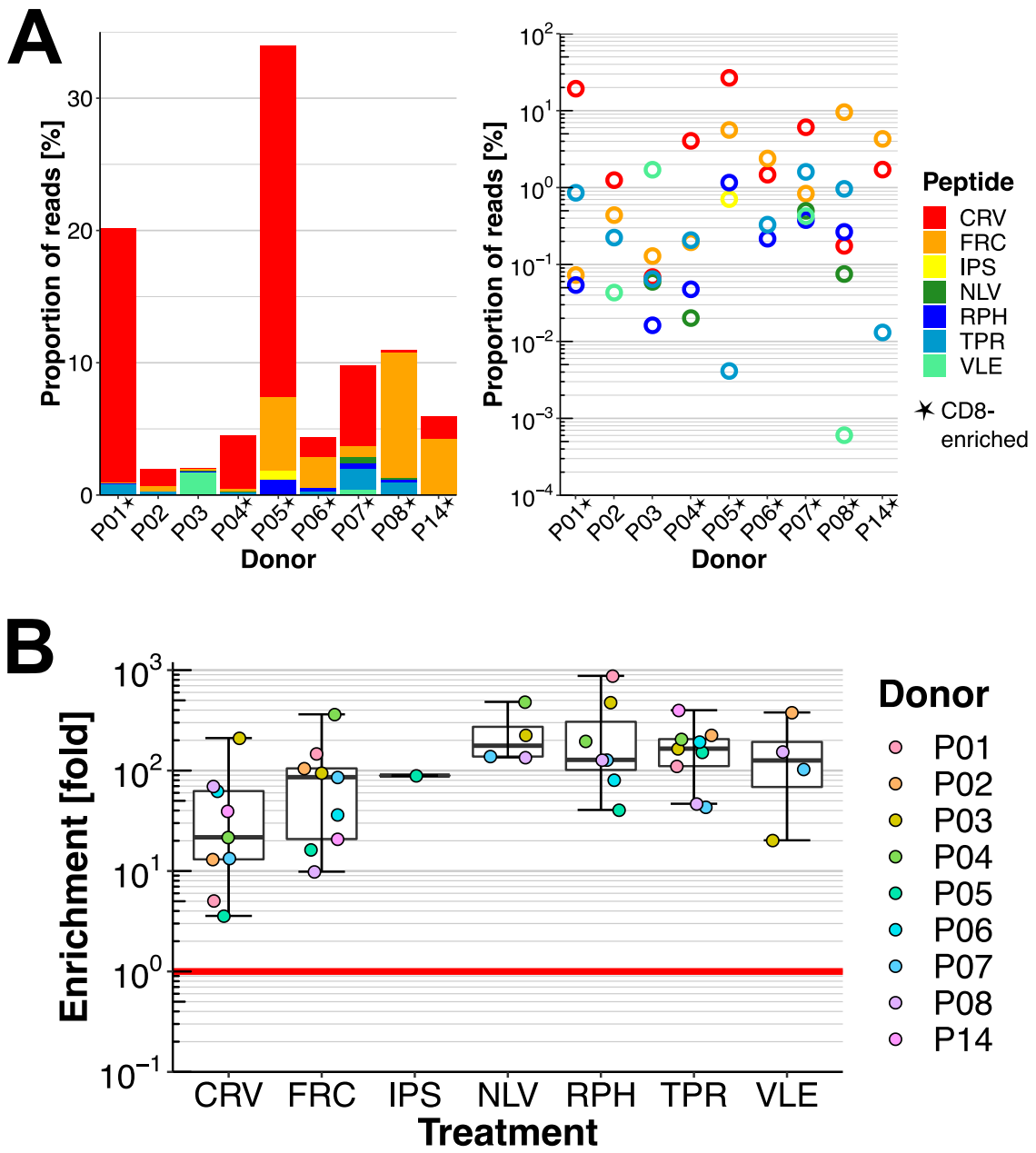


Figure 4.8: Ex vivo frequencies and overall enrichment of CMV-specific TCRβ sequences in samples of donors P01–P08 and P14. Samples derived from donors P01 (except the FRC-stimulated sample), P04, P05, P06, P07, P08, and P14 were CD8-enriched by MACS before sequencing. (**A**) Proportion of reads of CMV epitope-specific TCRβ sequences in ex vivo samples. The percentages were plotted on a linear y-axis (bar chart) and a logarithmic y-axis (dot plot). (**B**) Relative enrichment of epitope-specific T cells by 10-day stimulation. Enrichment was calculated by dividing the total proportion of epitope-specific TCRβ sequence reads in the sample after stimulation by their ex vivo frequency. The red line marks no change in proportion of positive T cells by single peptide stimulation. Bottom and top box ends represent the lower quartile (Q1) and upper quartile (Q3), the horizontal line is the median (Q2), and the whiskers extend to highest value and the lowest value of each data subset. Values that are $\leq$Q1–1.5$\times$IQR and values that are $\geq$Q3+1.5$\times$IQR are considered to be outliers. Outliers are ignored when determining whisker endpoints.

TCRβ clonotypes may be similar to samples without high frequency specific clonotypes. For example, TCRβ clonotypes in the CRV-enriched cluster from donor P01 were, on average, approximately 10-fold more frequent after peptide stimulation (enrichment=10), but they were approximately 1000-fold enriched compared to the TCRβ clonotypes not responding to CRV peptide (see Figure 4.7). Donor P01 responded strongly to peptide CRV and a CRV-specific TCRβ clonotype was already highly frequent ex vivo (>10% of total reads). No such high-frequency TCRβ clonotype was found after stimulation with RPH in this donor. The frequency of RPH-specific TCRβ clonotypes after peptide stimulation compared to their ex vivo frequency was, on average, 1000-fold increased (enrichment=1000). However, RPH-specific clonotypes were also 1000-fold enriched compared to the not RPH-responsive TCRβ clonotype population. Although TCRβ clonotypes of both specificities were similarly enriched by short-term peptide stimulation compared to the TCRβ clonotypes with different specificities, the computed enrichment factors presented in Table 4.3 for donor P01 were 2 orders of magnitude higher for RPH than CRV. In conclusion, the informative power of the calculated enrichment values in Table 4.3 is limited, since it does not take into account the degree of reduction in relative frequency of non-responding TCRβ clonotypes.

4.4.5 Correlation of flow cytometry and sequencing data

To answer the question whether there was a correlation between the proportion of multimer-positive T cells and read frequencies of TCRβ sequences specific for the corresponding epitope, the data from flow cytometry and high-throughput sequencing were compared (Figure 4.9). A moderate positive correlation between relative frequencies of specific T cells and specific TCRβ sequences was observed in ex vivo samples and peptide-stimulated samples from donors P01–P08 and P14 with Pearson correlation coefficients of r=0.434 and r=0.642, respectively. A major problem with these comparisons is that most samples were CD8-enriched prior to high-throughput sequencing, whereas flow cytometry was always performed on unsorted samples. CD8-enriched samples tended to contain higher proportions of CMV-specific TCRβ reads than unsorted samples (Figure 4.9), which may have negatively influenced the correlation between flow cytometry and sequencing data. This phenomenon may also explain why the data from stimulated samples correlated slightly better: Since only

CD8$^+$ T cell epitopes were investigated, CD8$^+$ T cells were enriched by CMV-peptide stimulation and the overall proportion of CD8$^+$ T cells in the stimulated samples was higher than in the ex vivo samples. MACS enrichment of CD8$^+$ T cells changed the ratio of CD8$^+$:CD4$^+$ T cells in the sequenced samples after peptide stimulation less than in the ex vivo samples, which contained a higher proportion of CD4$^+$ T cells. Hence, the sequenced peptide-stimulated samples will be more similar in their composition to the not CD8-enriched samples analysed by flow cytometry than the unstimulated samples, which may have led to the better positive correlation.

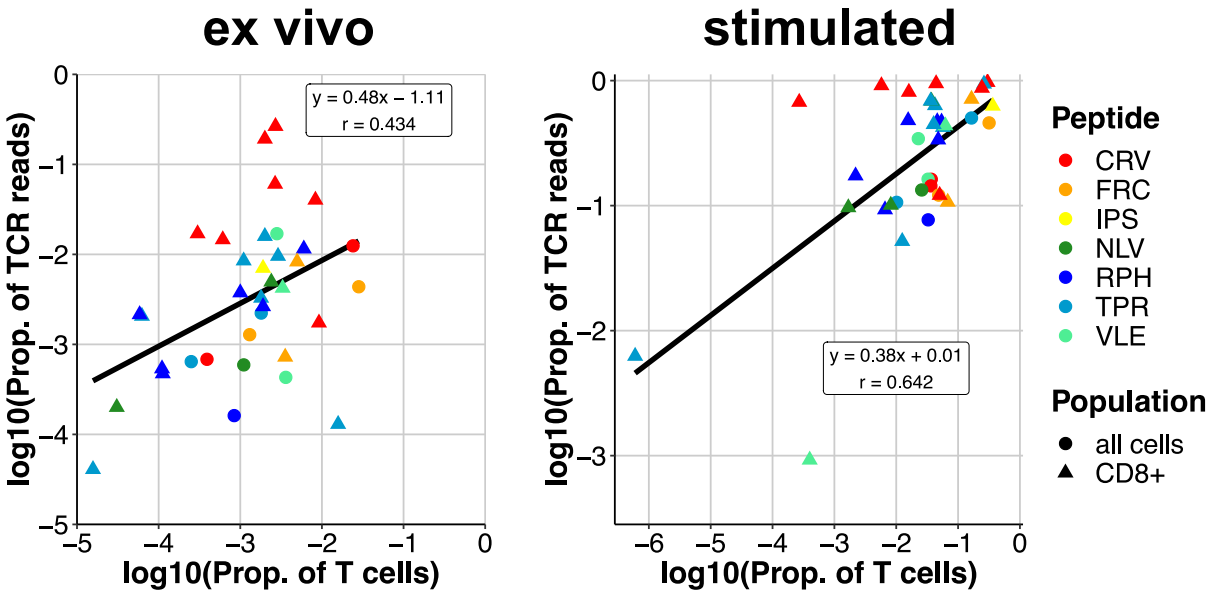


Figure 4.9: Comparison of frequencies of specific T cells and cumulative specific TCRβ sequences before and after peptide stimulation. The equation and Pearson correlation coefficient *r* for the linear regression are noted in the framed boxes. The proportions of multimer-positive T cells on total CD3$^+$ T cells are plotted on the x-axes, while the proportions of reads of specific TCRβ sequences are plotted on the y-axes. Samples marked with a triangle were CD8-enriched prior to sequencing.

4.4.6 Clonality of the CMV-specific TCRβ repertoire

When comparing the proportional read frequencies of CMV epitope-specifc TCRβ sequences in the corresponding single peptide-stimulated samples, it becomes apparent that most epitope-specific TCRβ repertoires are dominated by one or few highly frequent clonotypes (Figure 4.10). This observation was particularly striking in samples from donors P07 and P08, where the most frequent specific TCRβ clonotype(s) were up to 500-fold (P07, VLE) more frequent than the remaining ones. In 29 of 43 peptide-stimulated samples, the top most frequent TCRβ clonotype(s) amounted to

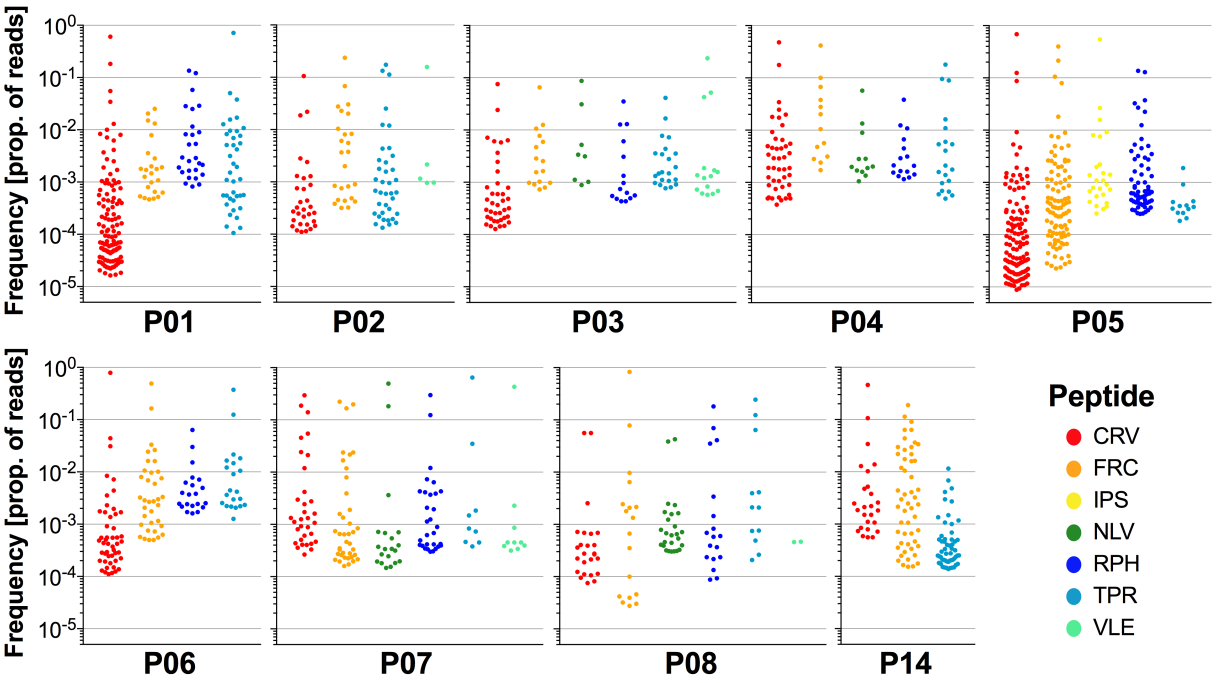


Figure 4.10: Clonality of CMV epitope-specific TCRβ repertoires. The plots show the proportional read frequencies of CMV-specific TCRβ clonotypes in the peptide-stimulated samples of CMV-positive donors P01–P08 and P14. Samples from donors P01 (except FRC), P04–P08, and P14 were CD8-enriched before sequencing.

at least 10% of all reads each, while the frequency of the majority of epitope-specific TCRβ clonotypes was well below 0.5%. Importantly, all samples from donors P02 and P03, and the FRC-stimulated sample from donor P01 were not CD8-enriched before sequencing and thus the frequency of epitope-specific TCRβ clonotypes in these samples is underestimated here. The most frequent TCRβ clonotype found in a stimulated sample was an FRC-specific clonotype in donor P08 with a relative read frequency of 83.1%.

4.4.7 VJ gene segment usage in TCRβ repertoires

There is a concern that multiplex PCR-based sequencing library preparation introduces an amplificaton bias, because not all primers bind equally well to their template. Hence, the obtained TCRβ repertoire data was initially compared to TCRβ repertoires of an independent control cohort from a recently published study by Ruggiero et al. (Ruggiero et al. 2015). The V and J (VJ) gene segment usage in ex vivo TCRβ repertoires of an extended donor cohort of 21 CMV-positive donors P01–P21 and 10 CMV-negative donors N01–N10 was compared to the VJ gene segment usage in ex vivo TCRβ repertoires of 6 healthy donors from the control cohort. The TCRβ libraries of

the independent control cohort were produced using a non gene-specific amplification system: TCRβ mRNA was reversely transcribed using a C-region–specific, biotinylated primer and the resulting TCRβ cDNA was captured on streptavidin beads. Then, a single-stranded linker cassette was appended to the free cDNA ends and TCRβ cDNA was amplified in two PCR steps using primers complementary to the C region and the linker cassette. Consequently, no multiplex primer-based PCR bias was introduced in the TCRβ repertoire data of the control cohort. Figure 4.11 shows the relative frequencies of the various Vβ and Jβ genes used in ex vivo TCRβ repertoires of healthy donors from this cohort and the cohort published by Ruggiero et al. It is important to note that some Vβ segments could not be distinguished from each other in our TCRβ repertoire data, since their nucleotide sequence was identical in the amplified and sequenced portion of the TCRβ chain. This affects Vβ genes TRBV3-1 and 3-2, TRBV6-1, 6-5 and 6-6, TRBV6-2 and 6-3, and TRBV12-3 and 12-4. Sequencing reads for these Vβ gene groups were added together, which likely contributed to the observation that these groups were among the most frequent Vβ genes used. VJ gene usage was remarkably similar in both cohorts which demonstrates there was no mea-

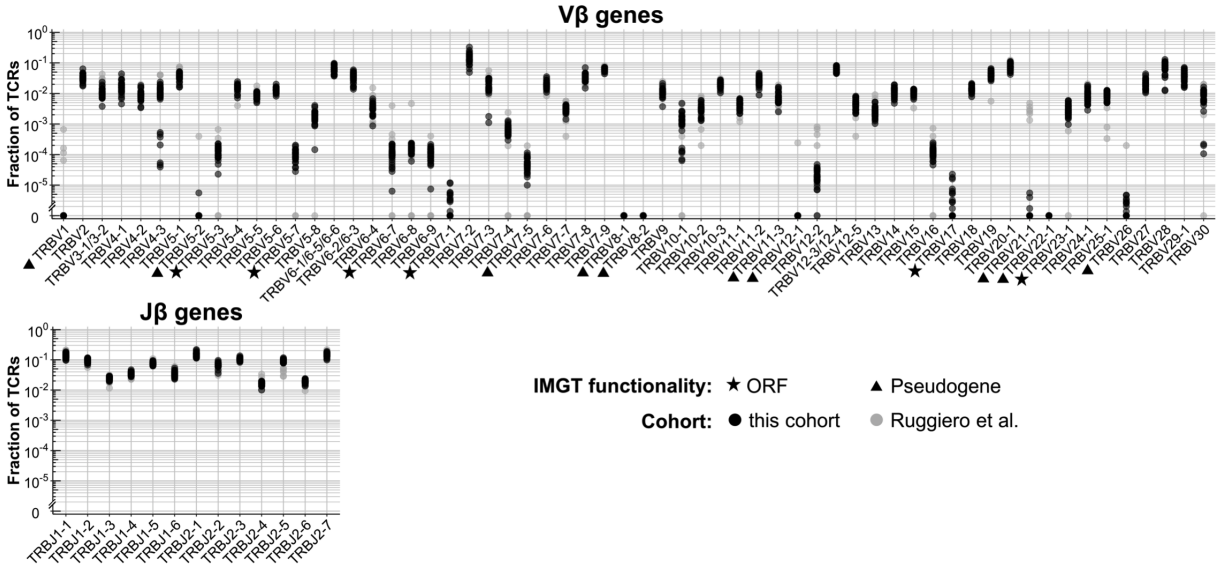


Figure 4.11: Proportion of TCRβ clonotypes using a particular TRBV gene (upper panel) or TRBJ gene (lower panel) in ex vivo PBMCs of donors P01–P08 and P14 and an independent cohort of 6 healthy donors (Ruggiero et al. 2015). Grey dots correspond to donors from the cohort published by Ruggiero et al. and black dots correspond to donors of this cohort. TRAV and TRAJ genes that are categorised as open reading frames or pseudogenes according to IMGT (Lefranc 2001) are marked with a star and a triangle, respectively.

surable PCR bias introduced by the multiplexed Vβ gene-specific forward primers used in this project. Functional Vβ genes were used with different relative frequencies that approximately ranged from 10^{-3} to 10^{-1}, with the exception of TRBV7-2, which was the most frequently Vβ gene used in ex vivo TCR repertoires across donors from both cohorts. Vβ genes derived from ORFs or pseudogenes were of low frequency in our cohort, since no primers targetting these Vβ genes were included in the set of multiplex primers. In the control cohort, however, some Vβ ORFs or pseudogenes were found at substantial frequencies in some or all donors, for instance TRBV21-1. The distribution of Jβ genes used in ex vivo TCR repertoires was similar in this cohort and the control cohort, although TRBJ2-5 was slightly overrepresented in this cohort.

The overall VJ gene segment usage in ex vivo TCRβ repertoires was comparable for donors in this cohort. Stimulation with CMV peptides led to strong shifts in VJ usage of TCRβ repertoires from CMV-positive donors, whereas VJ usage in TCRβ repertoires from CMV-negative donors was not influenced by stimulation with CMV peptides. For instance, single peptide stimulation of PBMCs derived from CMV-positive donor P07 led to dramatic changes in VJ gene usage in these samples (Figure 4.12). By contrast, relative frequencies of VJ gene segments used only marginally changed in peptide-stimulated samples of CMV-negative donor N05 compared to the unstimulated control sample. VJ usage in the ex vivo repertoires of donor P07 and donor N05 was similar: Both donors frequently use TRBV7-2, TRBV7-8, TRBV7-9, and TRBV28, and TRBJ1-1, TRBJ2-1, TRBVJ2-3, TRBJ2-5, and TRBJ2-7 with similar porportional read frequencies. TRBV7-2 was by far the most frequently used V gene in TCRβ sequences of ex vivo repertoires across CMV-positive donors P01–P08 and P14, and CMV-negative donors N01–N08. Upon single peptide stimulation, the proportions of VJ genes used in TCRβ repertoires of donor P07 shifted strongly to few highly frequent VJ combinations. Notably, the most frequent VJ combinations were completely different in 4 samples stimulated of donor P07 stimulated with different peptides. This implies that the underlying individual CMV-specific TCRβ clonotypes were very diverse for different epitope specificities.

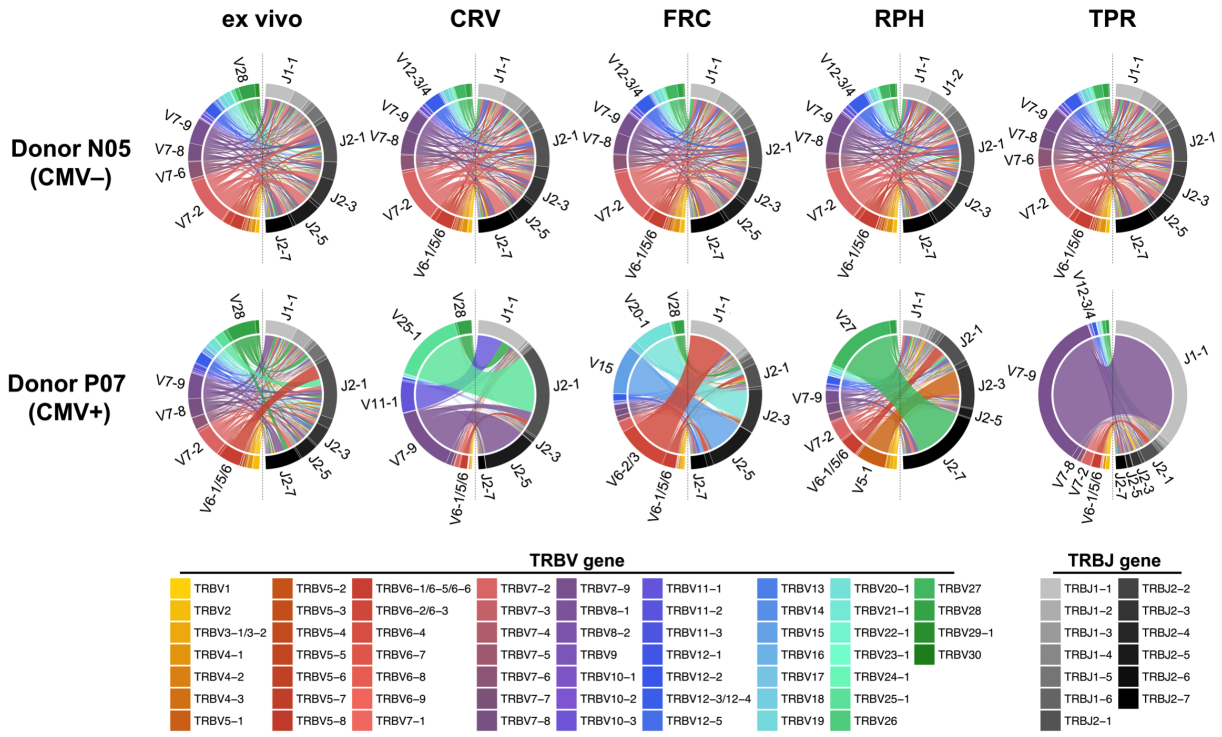


Figure 4.12: Usage of Vβ genes (TRBV) and Jβ genes (TRBJ) in samples before and after single peptide stimulation. Chord diagrams show the V gene usage (left semi-circle) and J gene usage (right semi-circle) in ex vivo or CMV peptide-stimulated TCRβ repertoires of CMV-negative donor N05 and CMV-positive donor P07. All samples were CD8-enriched before sequencing. The sector widths correspond to the relative proportion of reads using a particular V or J gene. The chord thickness is proportional to the read frequencies of particular combinations of V and J genes in the sample. The top 5 most frequently used gene segments are labelled in each chord diagram.

4.4.8 Epitope-reactive TCRβ clonotypes in CMV-negative donors

In addition to donors P01–P08 and P14, eight CMV-negative donors (N01–N08) were subjected to the single peptide stimulation assay. PBMCs derived from donors N01–N08 were stimulated with CMV peptides CRV, FRC, and TPR. PBMCs from donors N03 and N05 were also stimulated with HLA-matched CMV peptide(s) NLV, or IPS and RPH. Single peptide stimulation of CMV-negative donors with CMV peptides followed by high-throughput sequencing and data analysis using the same data analysis pipeline as described for the CMV-positive donors (Section 3.5.3.4) led to the identification of 399 epitope-reactive TCRβ sequences (Table A3, Table A4), of which 375 were monoreactive within samples of a donor. An additional 12 TCRβ sequences were found twice with different specificity in the same donor. As already implied by the flow cytometry data shown in Figure 4.4, there were hardly any multimer-positive T cells in stimulated samples derived from CMV-negative donors. In addition, no dis-

tinct clusters of enriched TCRβ clonotypes were seen in samples of CMV-negative donors when comparing ex vivo TCRβ read frequencies and TCRβ read frequencies in peptide-stimulated samples (Figure 4.7). Epitope-reactive TCRβ sequences identified in CMV-negative donors were absent or of low relative frequency ex vivo (0%–0.059% or reads) and contributed up to 7.2% of reads in the stimulated samples (Table A5). V gene usage among the 375 epitope-reactive TCRβ sequences was similar to the ex vivo distribution: 38% of TCRβ sequences used TRBV7-2, and gene segments TRBV6-1/6-5/6-6, TRBV7-6, TRBV7-8, TRBV7-9, TRBV12-3/12-4, TRBV28 were also frequently used (see Figure 4.12, leftmost chord diagrams). Similarly, the most frequently used J genes were TRBJ2-1, followed by TRBJ2-3, TRBJ2-7, and TRBJ1-1, which were also the most abundant J gene segments used ex vivo. When comparing the epitope-reactive TCRβ sequences on the amino acid level between donors, there was no sharing observed between donors. All these observations imply that stochastic effects, which were introduced by library preparation and sequencing, rather than epitope-specific in vitro enrichment led to the identification of specific TCRβ sequences in CMV-negative donors (see Section 4.4.1 for an explanation of this phenomenon). However, it cannot be excluded that, sporadically, naïve T cells or cross-reactive T cells of a CMV-negative donor were indeed specifically activated and expanded in the single peptide stimulation assay.

4.4.9 CMV-specific TCRβ sequences are highly abundant in the ex vivo repertoires of virus carriers

In a next step, the impact of CMV-infection on the ex vivo TCRβ repertoire were investigated. For this purpose, the top 100 most frequent TCRβ clonotypes in ex vivo PBMCs or CD8-sorted samples were extracted and compared to the CMV epitope-specific TCRβ sequences identified by single peptide stimulation of PBMCs from the same donor (Figure 4.13). The top 100 most frequent TCRβ clonotypes in ex vivo samples were similarly distributed in CMV-positive and CMV-negative donors, with some highly frequent clonotypes accumulating above 1% of reads and the majority of clonotypes with read frequencies between 0.01%–1%. In the CD8-enriched samples of donors P01, P03, P14, N02, N03, N04, and N07, 1–2 TCRβ clonotypes had frequencies $>$10% of the total acquired TCRβ reads. 2–15 of the 100 most frequent TCRβ clonotypes in PBMCs and 5–17 in CD8-enriched samples of CMV-positive donors were

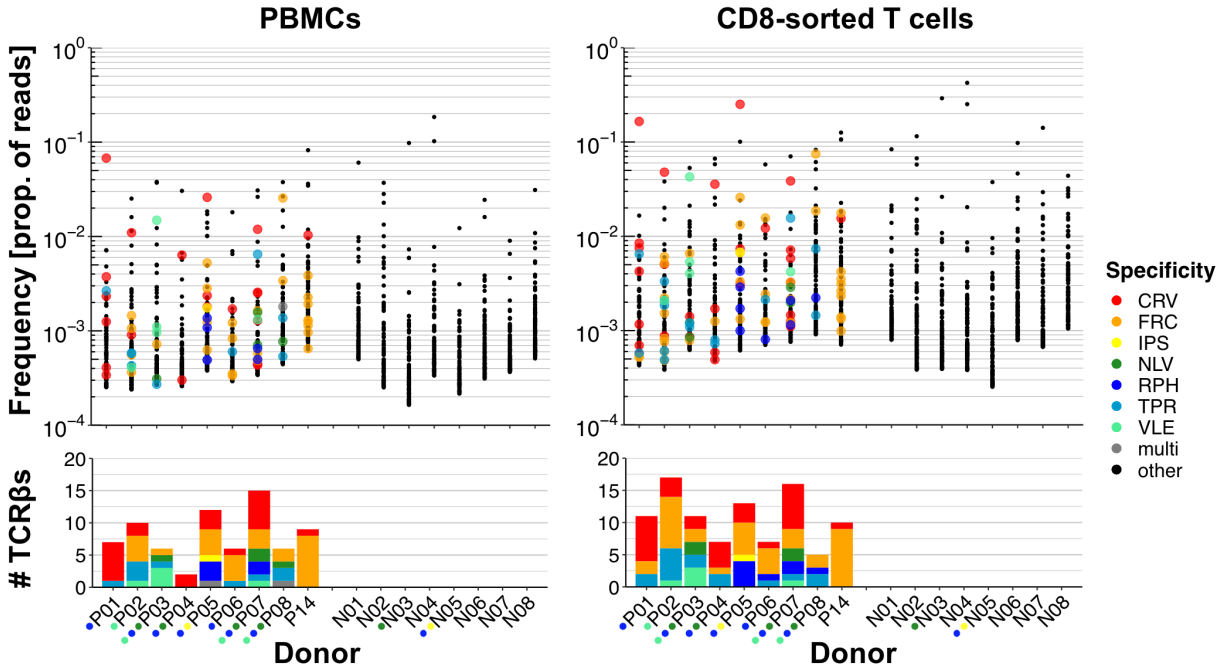


Figure 4.13: Frequency of CMV epitope-specific TCRβ sequences in ex vivo repertoires of virus carriers and uninfected persons. The ex vivo frequencies of the top 100 most frequent TCRβ sequences in PBMCs (left) or CD8-enriched cells (right) are plotted in the top panel for CMV-positive donors P01–P08 and P14 as well as CMV-negative donors N01–N08. In the lower panel, the number of different CMV-specific TCRβ sequences in the top 100 most frequent clonotypes are plotted by epitope. Specific TCRβ sequences were only searched for in the ex vivo repertoire of the donor in which they were identified. All donors were single peptide-stimulated with CRV, FRC, and TPR. Some donors were additionally stimulated with IPS, NLV, RPH, or VLE if they expressed the matching HLA; these donors are marked with circles in the colour of the corresponding epitope.

specific for one of the tested epitopes, whereas no epitope-reactive TCRβ sequences identified in CMV-negative donors were among the top 100 most frequent TCRβ clonotypes of these donors. In 8 of 9 virus carriers, at least one CMV-specific TCR clonotype was among the top 5 most frequent TCRβ clonotypes in the entire repertoire, and all virus carriers had CMV-specific TCRβ clonotypes among the top 5 most frequent TCRβ sequences in the CD8-sorted samples. The most frequent CMV-specific TCRβ clonotype ex vivo was specific for CRV in 7 of 9 CMV-positive donors, and in donors P03 and P08, the top CMV-specific clonotype was specific for VLE and FRC, respectively. Overall, the most abundant CMV-specific TCRβ clonotypes in the ex vivo PBMCs and CD8-enriched samples of donors P01–P08 and P14 were specific for the HLA-C*07:02–restricted epitopes CRV and FRC, with the exception of P03, in which VLE-specific TCRβ clonotypes were highly frequent ex vivo. Taken together, these

findings show that CMV infection strongly shapes the TCRβ repertoire of its carriers, as CMV-specific TCRβ sequences are highly frequent in the blood of infected donors. T cells specific for HLA-C*07:02–restricted CMV epitopes were particularly frequent. This observation is in line with the findings of a recent study (Hosie et al. 2017) in which researchers showed that $CD8^+$ T cells specific for HLA-C*07:02-restricted CMV epitopes dominate the immune repertoire of CMV-positive elderly persons.

4.5 Peptide stimulation discerns peptide-specific TCRβ clonotypes better than MACS enrichment

The majority of recently published CMV-specific TCRβ sequences were identified by combining peptide-MHC multimer-based sorting of peripheral blood T cells and high-throughput sequencing of the sorted cell population (Klarenbeek et al. 2012; Glanville et al. 2017; Emerson et al. 2017; Chen et al. 2017). To test whether this approach yields comparable results to the single peptide stimulation approach, PBMCs from two CMV-positive donors whose T-cell repertoires were also explored with the peptide stimulation assay (P01 and P04) were labelled with CRV/HLA-C*07:02 streptamers or TPR/HLA-B*07:02 pentamers. Multimer-positive cells were isolated by MACS using ”Anti-PE MicroBeads” (Miltenyi, Bergisch Gladbach, DE), and subsequently their TCRβ repertoires were sequenced by Xiaoling Liang, a former post-doc of the group. A reanalysis of these data was performed by me for this project.

Between 0.14% and 0.68% of input cells were obtained as the multimer-positive fraction after MACS enrichment (Table 4.4). Relative enrichment of multimer-positive T cells by MACS, as determined by flow cytometry, was between 30-fold and 1258-fold, leading to a purity between 15.1% and 67.4% in the positive fractions that were sequenced. The absolute number of sequenced CMV-specific T cells, which was computed by multiplication of the cell count in the eluate with the proportion of multimer-positive cells measured by flow cytometry, was between 1.3×10^4 and 1.5×10^5. Relative frequencies and enrichment values of CMV-specific T cells from flow cytometry measurements were in the same range as those measured for peptide-stimulated samples (see Table 4.1), but the absolute cell count after peptide stimulation (7×10^6–84×10^6, median = 23×10^6) was considerably higher than the cell count after MACS enrichment. Therefore, the number of CMV-specific T cells in the peptide-stimulated samples substan-

Table 4.4: Enrichment of multimer-labelled T cells by MACS. Cells were counted before (input) and after (eluate) MACS-enrichment of T cells specific for CMV epitopes CRV or TPR from PBMCs of donor P01 or P04. Multimer-staining and flow cytometry were used to determine the relative frequency of multimer-positive T cells in the input PBMCs and eluate (positive) fractions. Yield = Percentage of multimer-labelled cells eluted from the MACS column, Comp. count = computed number of multimer-positive cells in the eluted fraction, Rel. enrichment = relative enrichment of multimer-positive T cells by MACS.

		Input		Eluate	
Donor	**Multimer**	**Cells [10^6]**	**Multimer+ [%]**	**Cells [10^6]**	**Multimer+ [%]**
P01	CRV	25	0.67	0.09	20.1
P01	TPR	25	0.16	0.17	27.1
P04	CRV	60	0.96	0.22	67.4
P04	TPR	60	0.012	0.086	15.1

		Multimer+ Cells		
Donor	**Multimer**	**Yield [%]**	**Comp. count [10^6]**	**Rel. enrichment [-fold]**
P01	CRV	0.36	0.018	30
P01	TPR	0.68	0.046	169
P04	CRV	0.37	0.148	70
P04	TPR	0.14	0.013	1258

tially exceeded the number of CMV-specific T cells in the multimer-sorted samples, although similar or higher input cell numbers were used for the latter samples. After sequencing, the relative frequencies of TCRβ clonotypes in single peptide-stimulated samples or multimer-enriched samples were compared to their frequencies in the corresponding ex vivo control samples (Figure 4.14). The clusters of enriched TCRβ clonotpes after single peptide stimulation were larger and more clearly separated in the frequency scatter plots than the clusters of enriched TCRβ clonotypes after multimer sorting. Especially low frequency epitope-specific TCRβ clonotypes were not visibly enriched in the multimer-sorted samples. Between 7947 and 33353 distinct TCRβ clonotypes were found in each of the multimer-sorted samples (Table A6), suggesting that the TCRβ clonotypes obtained from multimer sorting were far from being exclusively epitope-specific. Computational identification of epitope-specific TCRβ sequences from multimer-sorted samples using the data analysis pipeline developed in this thesis would therefore result in a significantly lower amount of epitope-specific TCRβ clonotypes than with the single peptide stimulation assay. Only the most abundant epitope-specific TCRβ clonotypes would be found by multimer sorting, whereas the low-frequency epitope-specific TCRβ clonotypes would go unrecognised. This limited set of sequencing data obtained from CMV multimer-enriched samples suggests

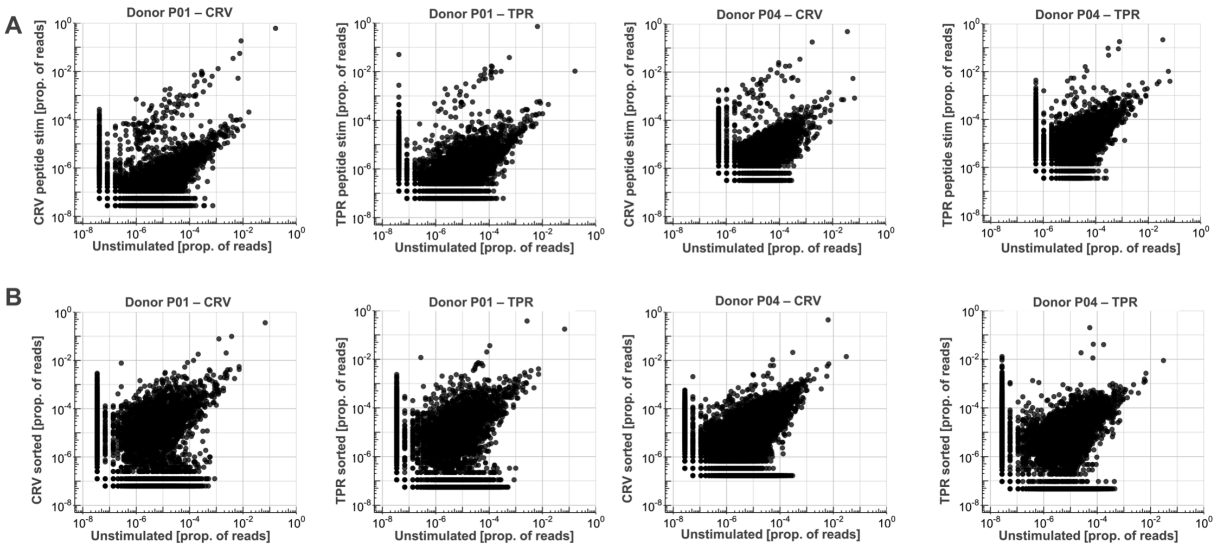


Figure 4.14: Comparison of TCRβ clonotype frequencies in single peptide-stimulated and multimer-MACS samples of donors P01 and P04. Clonotypes that were undetectable in one sample were assigned a pseudo-frequency corresponding to 0.5 reads to enable their display on a logarithmic axis. (**A**) TCRβ clonotype frequencies in CRV-stimulated and TPR-stimulated samples of donors P01 and P04 (y-axes) compared to their ex vivo frequencies (x-axes). All samples were CD8-enriched prior to RNA isolation. (**B**) TCRβ clonotype frequencies in CRV:HLA-C*07:02 or TPR:HLA-B*07:02 multimer-sorted samples of donors P01 and P04 (y-axes) compared to their frequencies in PBMCs (x-axes).

that single peptide stimulation combined with high-throughput sequencing is superior in identification of CMV epitope-specific TCRβ sequences compared to TCRβ sequencing of multimer-sorted samples. A limitation of this comparison is that the purity of multimer-positive T cells after MACS-based isolation may be lower than the purity that could potentially be achieved by flow cytometry sorting.

4.6 Co-culture with CMV antigen-presenting cells enriches epitope-specific T cells

In vitro peptide stimulation to enrich CMV-specific T cells has two major limitations. Firstly, synthetic peptides do not carry post-translational modifications. Therefore, doubts have been raised over whether T cells responding to synthetic peptides would also recognise endogenously expressed and processed viral antigens, and whether genuinely virus-specific T cells may be missed because they do not react to unmodified synthetic peptides (Yewdell 2005). Secondly, CMV antigens are processed and presented in different quantities and the physologically relevant amount of presented peptide required for T-cell activation is unknown. In the peptide stimulation approach,

synthetic peptide is loaded in excess to ensure that specific T cells are activated. However, physiological levels of presented antigen can be expected to be much lower and it is not clear whether physiological levels of CMV peptide activate the same T cells as those identified by peptide stimulation. Because of the excess of synthetic peptide, there is a risk of activating low-affinity cross-reactive T cells that recognise similar epitopes from completely unrelated antigens in the artificial in vitro peptide stimulation assay. This may also explain, in part, the background of epitope-reactive TCRβ found in CMV-negative donors (see Section 4.4.8).

In order to address these questions, PBMCs derived from 3 of 9 CMV-positive donors who were previously tested with the single peptide stimulation assay (P01, P02, and P03) and a CMV-negative donor (N09) were long-term stimulated with autologous B cell lines transformed by infection with a minimal EBV genome, so-called mini-LCLs (see Section 3.2.4). All 4 donors were positive for EBV. The mini-LCLs expressed one additional full-length CMV antigen (pp65 or IE1) together with the EBV antigens encoded by the mini-EBV genome, or no additional CMV antigen (empty mini-LCL). For two donors, autologous CD40-activated B cells (B blasts) were used as negative controls. PBMCs were co-cultivated with one of the mini-LCLs or a B blast line for 30 days as described in Section 3.3.2). At each time point of restimulation, cells were counted and samples were taken for flow cytometry analysis. Figure 4.15 shows the growth rates and extrapolated cell numbers of mini-LCL–stimulated PBMCs from donors P01–P03 and N09. Cell counts of mini-LCL–stimulated PBMCs declined over the first 9 days, but then increased until day 23 post start of culture. The observed growth rates were highest at day 16 and day 23 of co-culture, and often declined towards the endpoint on day 30. Co-stimulation of PBMCs with autologous B blasts from donor P01 was terminated on d16, because the stimulated cells were barely viable. In donor P01, stimulation with IE1-expressing mini-LCL led to higher growth rates between day 9 and day 16 than with the other mini-LCLs or B blasts, but was comparable to growth rates of the other co-cultures from day 23 onwards. In CMV-negative donor N09, growth rates were similar for all stimulations except PBMCs stimulated with autologous B blasts, which only proliferated around day 23.

Next, the mini-LCL–stimulated cells were phenotypically analysed for their specificity by flow cytometry. Samples were stained with CMV multimers and EBV antigen EBNA3A-

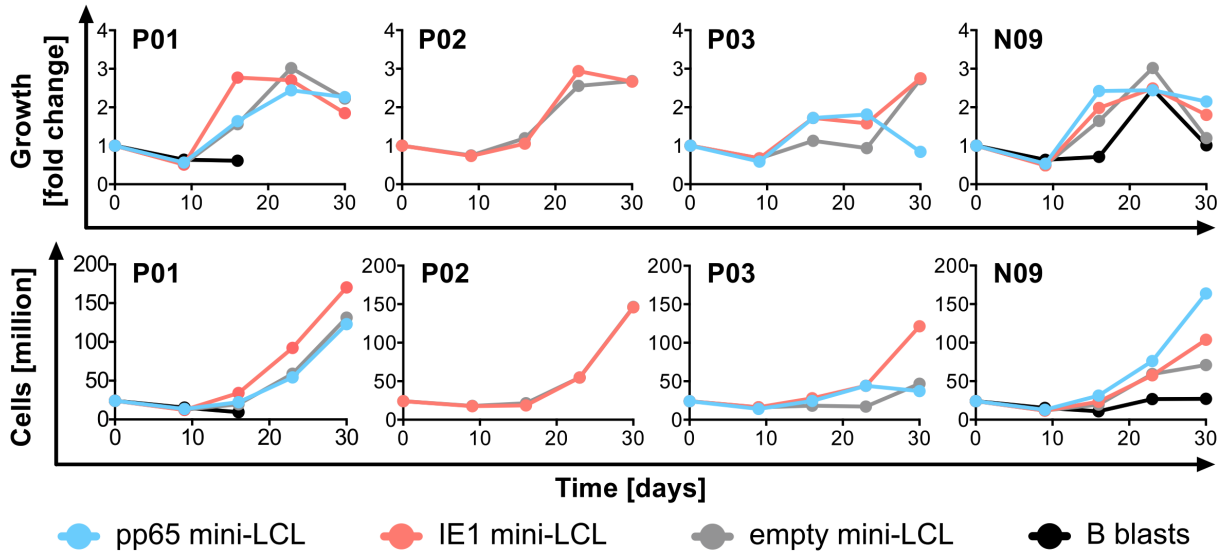


Figure 4.15: Cell growth from each time point to the next (upper panel) and extrapolated cell counts (lower panel) during 30-day stimulations of CMV-positive donors P01–P03 and CMV-negative donor N09 with autologous mini-LCLs or CD40-activated B cells (B blasts). Growth rates were calculated by dividing the number of harvested cells at each time point by the number of cells seeded on the previous time point. Extrapolated cell numbers were calculated by successively multiplying the initially seeded number of PBMCs with the growth rates determined on day 9, 16, 23, and 30 of coculture. They represent the cell numbers that would have been obtained had no samples been taken at previous time points. Donors P02 and P03 were not stimulated with autologous B blasts, and donor P02 was neither stimulated with autologous pp65 mini-LCL, since the pp65 mini-LCL of this donor was not viable.

derived RPP multimer, all matching the donors' HLA types, and data were gated on $CD3^+$ lymphocytes. Data analysis showed that, over time, CMV epitope-specific T cells were enriched by stimulation with a mini-LCL expressing the matching antigen in virus carriers. Figure 4.16 shows representative FACS plots of IE1 mini-LCL–stimulated T-cell cultures derived from donors P01 and N09 stained with CRV or RPP multimer. CRV-specific T cells strongly increased throughout mini-LCL stimulation in CMV-positive donor P09, but not in CMV-negative donor N09. EBV-targeting RPP-specific T cells increased over time in both donors, which meets the expectations since both donors are EBV-positive. The flow cytometry data of all donors and mini-LCL–stimulations show that enrichment of epitope-specific T cells over time was exclusively observed in samples derived from donors with the correct virus status and stimulated with the corresponding antigen (Figure 4.17). For example, CRV-positive T cells were solely enriched in IE1 mini-LCL–stimulated T cells of donor P01, and TPR- and RPH-positive T cells were solely found in pp65 mini-LCL–stimulated T cells of donor P01.

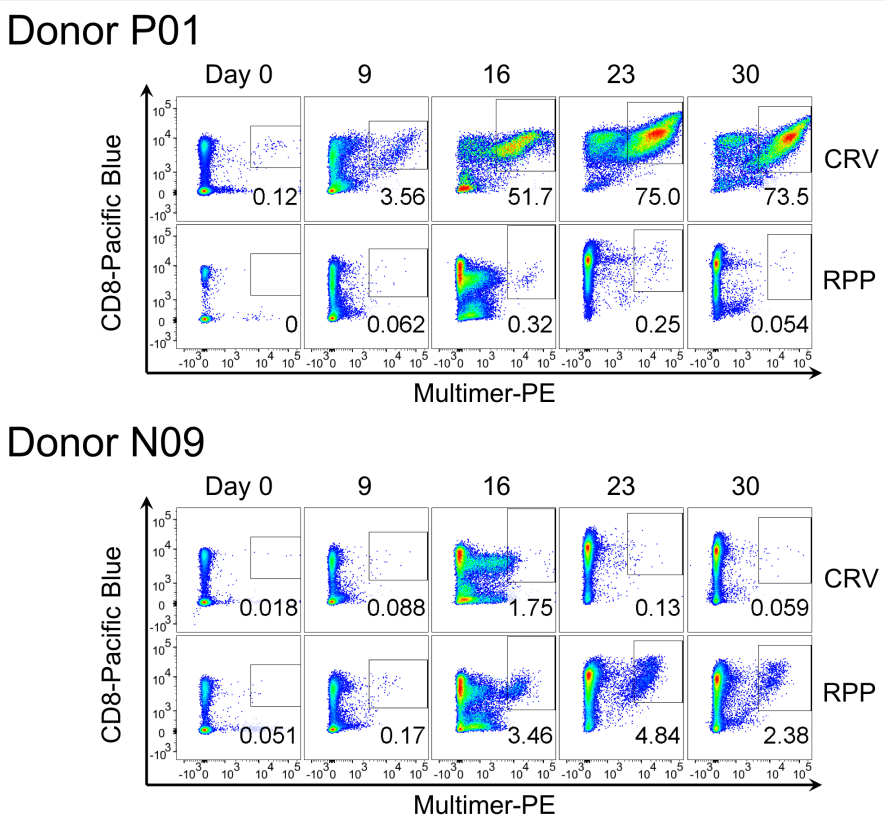


Figure 4.16: Specific enrichment of multimer-positive T cells by mini-LCL stimulation over time. Flow cytometry data of IE1 mini-LCL–stimulated PBMCs from donors P01 and N09 stained with PE-labelled CRV streptamer or RPP multimer. Only donor P01 is CMV-positive, but both donors carry EBV.

Neither CRV-positive, nor RPH- or TPR-positive T cells emerged after stimulation with empty control mini-LCLs. No CMV-multimer–positive T cells were enriched in donor N09 throughout the mini-LCL stimulations, except on day 16, where multimer staining resulted in ambiguous low-positive populations (compare Figure 4.16). RPP-specific T cells against EBV were enriched in all four donors and peaked around day 16 and day 23. Multimer-positive T cells started to accumulate around day 9 and either continued to outgrow until day 30, or peaked on day 23 and declined towards the end of co-culture. CRV-specific T cells were the most abundant with 50%–75% in IE1-stimulated samples of all 3 CMV-positive donors. In donor P03, an additional 20% of T cells were specific for VLE, accounting for a total of more than 70% of CMV-specific T cells with known epitope specificity in this sample. T cells against pp65-derived epitopes were less frequent and accounted for up to 20% of T cells. Weak epitope-specific responses upon full-length antigen stimulation were observed in donor P02 for VLE and in donor P03 for TPR.

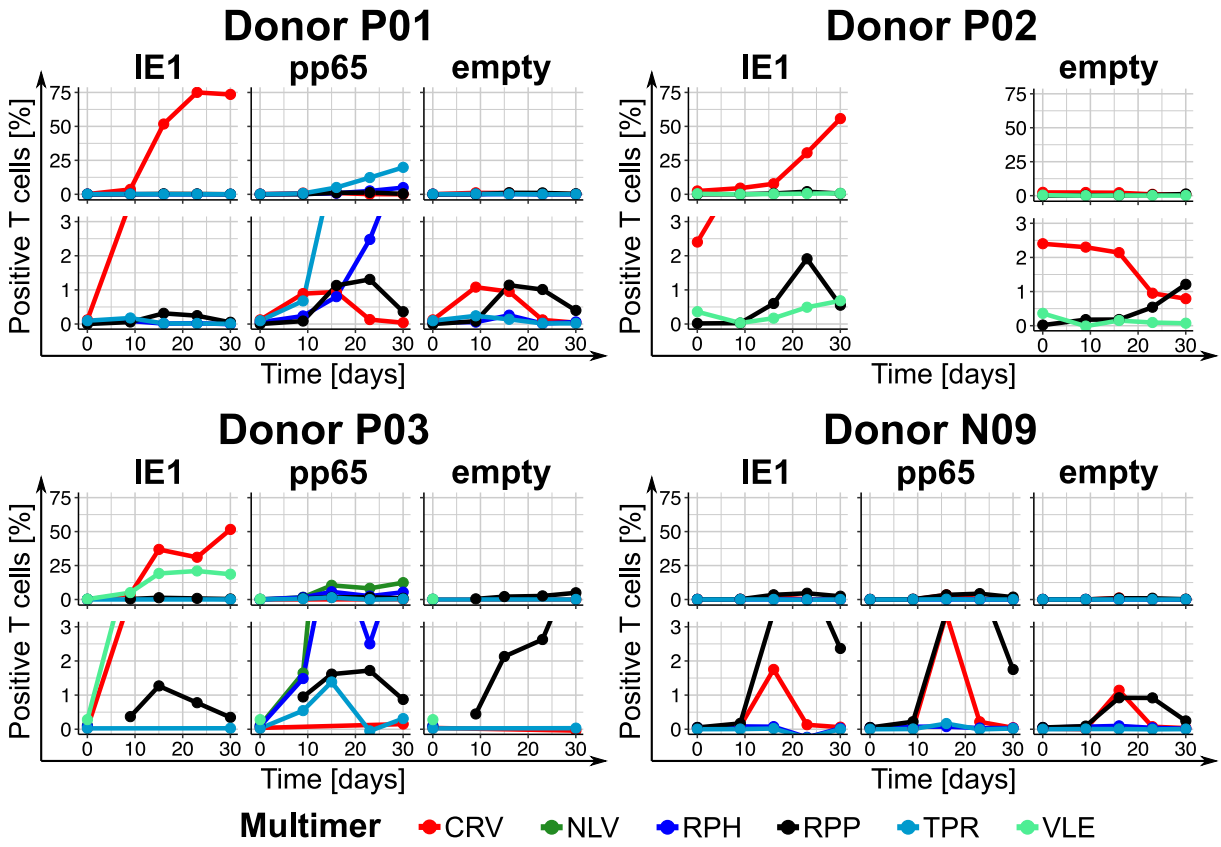


Figure 4.17: Relative frequencies of CMV epitope-specific T cells in mini-LCL–stimulated samples as measured by flow cytometry. Cells were gated on lymphocytes, then single cells and finally $CD3^+$ T cells. Percentages of multimer-positive T cells refer to the number of $CD3^+$ T cells in the samples. For each donor and mini-LCL, frequencies of multimer-positive T cells are shown on a large scale (0%–80%) and as a close up on the y-axis (0%–3%). Multimer stainings in which the multimer did not match the antigen were not done for all samples from donors P02 and P03. As a marker for EBV-specific T cells, all samples were stained with RPP-multimer derived from the EBV antigen EBNA-3A.

4.7 Peptide stimulation expands TCRβ clonotypes that recognise endogenously processed antigen

Samples from day 23 or day 30 of mini-LCL–stimulation were chosen for TCRβ library preparation and high-throughput sequencing. When comparing TCRβ clonotype frequencies in the IE1 mini-LCL–stimulated or pp65 mini-LCL–stimulated samples to their frequencies in the corresponding empty mini-LCL control sample, clusters of specifically enriched T cells were solely found in CMV-positive donors P01–P03, while CMV-negative donor N09 lacked enriched TCRβ sequences (Figure 4.18). CMV antigen-specific TCRβ sequences were identified only from CMV-positive donors P01–P03 by comparing the clonotype frequencies in the CMV antigen-stimulated sample (pp65 or

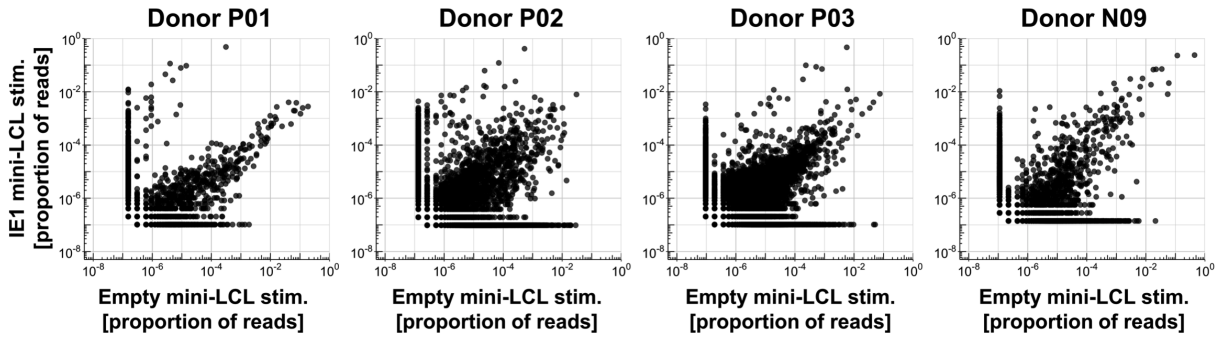


Figure 4.18: Frequencies of TCRβ clonotypes in PBMCs and mini-LCL–stimulated samples. Each dot represents one TCRβ clonotype. Proportional TCRβ clonotype frequencies in the empty mini-LCL–stimulated samples of donors P01, P02, P03, and N09 are plotted on the x-axes; relative frequencies after stimulation with IE1-expressing mini-LCLs are plotted on the y-axes. TCRβ clonotypes that were only found in one of the two samples shown in each plot were assigned a pseudofrequency of 0.5 reads in the sample in which they were absent in order to depict them on a logarithmic scale.

IE1), in the empty mini-LCL–stimulated sample, and in the unstimulated PBMCs as explained in Section 3.5.3.5. The median productive read count per sample was 4.5 million reads, and the median number of TCRβ clonotypes was 71045 (Table A7). In total, 1572 nucleotide-unique TCRβ clonotypes were found in donors P01, P02, and P03 (Table A8). 11 of these TCRβ clonotypes were found to be specific for both IE1 and pp65 in donor P03, leaving 1550 monospecific TCRβ sequences. Of these, 915 were specific for IE1 and 635 were specific for pp65. Next, the antigen-specific TCRβ sequences identified by mini-LCL stimulation were compared to the epitope-specific TCRβ sequences found by single peptide stimulation of the same donors P01–P03. Of the 915 IE1-specific TCRβ sequences, 138 were specific for CRV, 6 for VLE, and no epitope could be assigned to 771 TCRβ sequences. Of the 635 pp65-specific TCRβ sequences, 4 were specific for NLV, 26 for RPH, 38 for TPR, and 566 were of unknown epitope specificity. One pp65-enriched TCRβ sequence was enriched with both RPH and TPR in the single peptide stimulation assay and therefore of ambiguous specificity. Although the precise epitope specificity of the majority of different TCRβ clonotypes could not determined, the majority of pp65-specific and IE1-specific TCRβ reads could be assigned to epitope specificities in 4 out of 5 samples (Figure 4.19). For example, 95.6% of TCRβ reads belonged to CRV-specific clonotypes in the IE1 mini-LCL–stimulated sample of P01. In the pp65 mini-LCL–stimulated sample of this donor, 56.1% of reads belonged to TPR-specific clonotypes and another 7.6% to RPH-specific clonotypes. Nonetheless, 32.2% of IE1-stimulated TCRβ reads in donor P02

and 21.6% of pp65-stimulated TCRβ reads in donor P03 could not be assigned to an epitope specificity based on the data obtained by single peptide stimulation of the same donor. Hence, it may be interesting to test these donor for additional IE1-derived and pp65-derived epitopes, respectively.

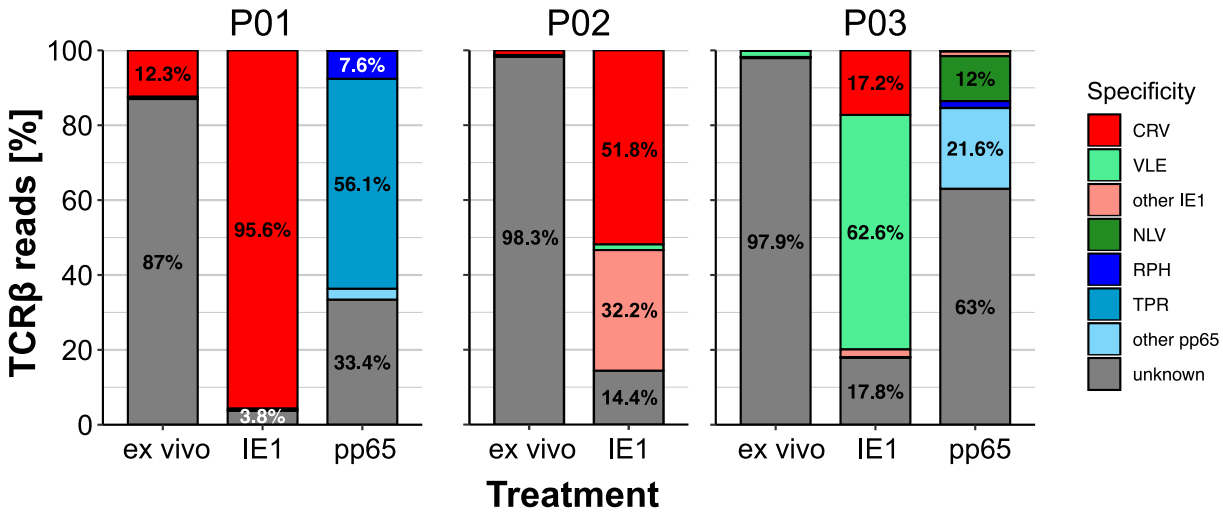


Figure 4.19: The majority of full-length antigen-specific TCRβ reads identified by mini-LCL–stimulation belongs to clonotypes with known epitope specificity. Cumulative read frequencies of TCRβ sequences of each epitope specificity in the ex vivo and mini-LCL–stimulated samples of donors P01, P02, and P03 are shown on the y-axes.

To find out whether the CMV epitope-specific TCRβ clonotypes identified by single peptide stimulation also recognise endogenously processed and presented epitopes from full-length antigens, specific TCRβ sequences identified using the single peptide stimulation assay were compared to enriched TCRβ sequences from mini-LCL stimulation (Figure 4.20). With the exception of TPR-specific TCRβ clonotypes identified from donor P03, all top frequency epitope-specific TCRβ clonotypes activated and proliferated by co-culture with an autologous mini-LCL presenting the corresponding antigen. Moreover, many less frequent epitope-specific TCRβ clonotypes were also identified as antigen-specific in the mini-LCL stimulation assay. Especially CRV-specific TCRβ sequences were abundant in IE1 mini-LCL–stimulated samples of donors P01–P03. The data show that majority of peptide-specific TCRβ clonotypes was specificially enriched with physiological levels of antigen exogenously processed and presented by mini-LCL. These findings corroborate that the single peptide stimulation assay is a valid approach to identify genuine CMV epitope-specific TCRβ sequences. The assay is also suitable to identify rare CMV-specific TCRβ clonotypes that are of low frequency ex vivo, but selectively enrich upon peptide stimulation. In addition, the data suggest that a majo-

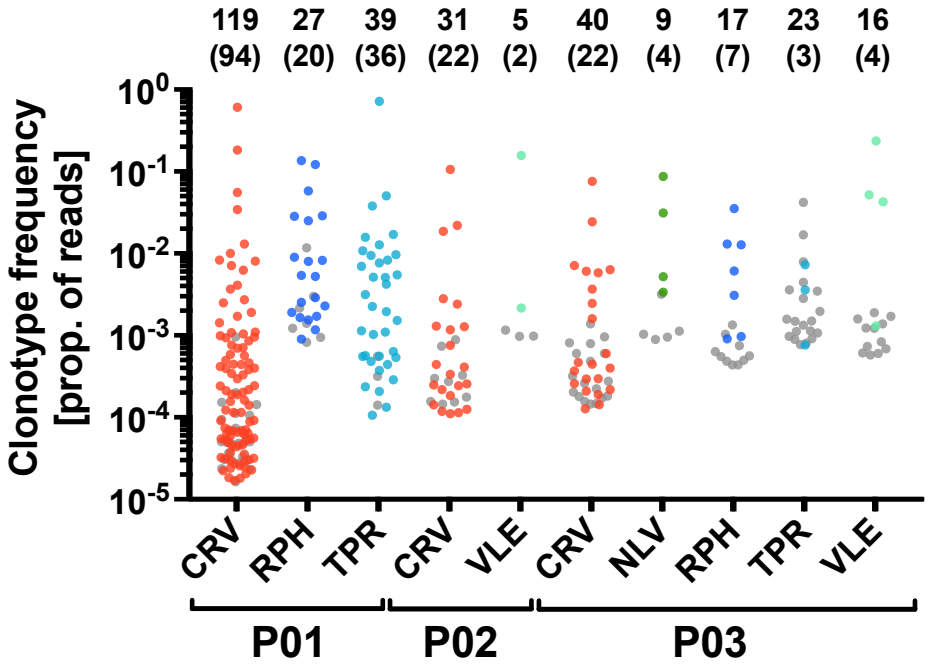


Figure 4.20: Specific TCRβ clonotypes identified by peptide stimulation recognise endogenously processed CMV antigen. The analysis was performed for antigens IE1 (epitopes CRV and VLE) in CMV-positive donors P01–P03 and pp65 (epitopes NLV, RPH and TPR) in donors P01 and P03. The plot depicts all TCRβ clonotypes that were identified as epitope-specific in the peptide stimulation assay. The y-axis shows the frequency of each clonotype after peptide stimulation. Coloured dots represent clonotypes that were specifically enriched by stimulation with an autologous mini-LCL expressing the corresponding CMV antigen; grey dots indicate clonotypes for which this was not the case. The numbers on top indicate the total number of epitope-specific TCRβ clonotypes and, in parentheses, the number of clonotypes responding to antigen endogenously processed by mini-LCLs. Samples of donor P01 were CD8-enriched before sequencing.

rity of IE1-specific and pp65-specific T cells in donors with HLA backgrounds similar to P01–P03 are directed against a small number of already known immunodominant epitopes. However, this may not hold true for donors with less well-studied HLA alleles. Comparative TCRβ sequencing can serve to identify donors with antigen-specific T-cell responses that are unaccounted for on the epitope level, in order to identify presently unknown immunodominant epitopes.

4.8 Stimulation with multiplexed CMV peptides to characterise T cells with various specificities

Since the identification of CMV-specific TCRβ clonotypes by single peptide stimulation is quite laborious and one sample is required per specificity, the assay was scaled up

to test multiple specificities simultaneously with as few samples as possible in donors with various HLA types. For her master's thesis in molecular biotechnology, Fenja Gerpott conducted a set of experiments under my supervision in which she distributed 32 immunogenic CMV peptides, all of which were previously described $CD8^+$ T-cell targets, to 8 different pools. The peptides were distributed in a way that each peptide was present in a unique combination of 3 pools and absent in the remaining 5 pools, resulting in a unique "address" for each peptide. To optimise coverage of CMV-specific T cells, homologous peptides whose sequence shows some variation in different CMV strains were combined and treated as one epitope. The full-length peptide sequences are listed in Table 2.13 and the pool distributions are noted in Table 3.4. As in the single peptide stimulation assay, PBMCs were loaded with pooled peptides, excess peptide was washed off, and cells were cultured in the presence of IL-2 for 10 days. Cells were counted before and after the 10-day stimulation period, and expansion rates for each stimulation were calculated by dividing the cell number harvested by the cell number seeded (Figure 4.21). For the peptide pool stimulation experiments, CMV-positive

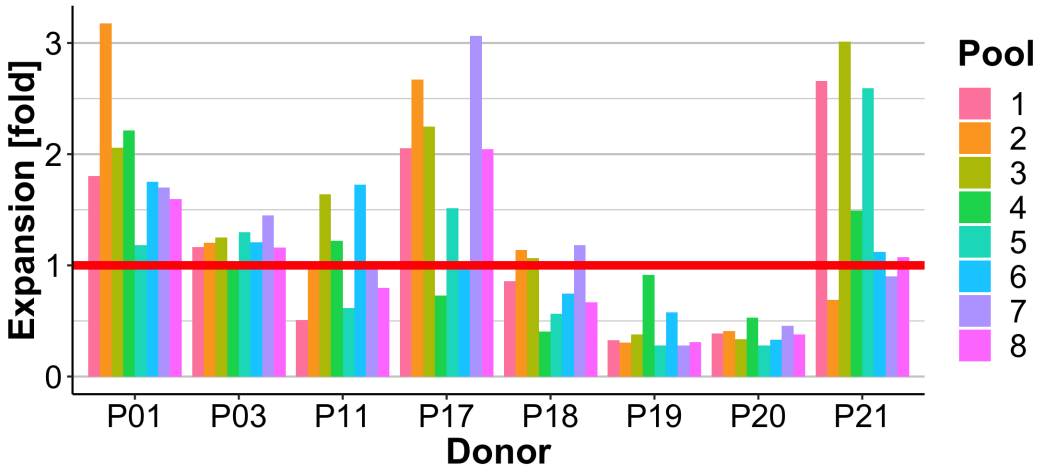


Figure 4.21: Expansion rates of CMV peptide pool-stimulated PBMCs derived from 8 CMV-positive donors. The red line indicates identical cell counts before and after stimulation (expansion rate=1). Experiments were performed jointly by Fenja Gerpott and me.

donors with a wide variety of different HLA types matching to multiple of the 32 CMV peptides were chosen (P01, P03, P11, P17–P21). Consistent with this donor choice, the observed expansion in culture was strongly dependent on the donor and the pep-

tide pool. While absolute cell numbers were up to 3-fold increased after stimulation of PBMCs derived from donors P01, P17, and P21 with some peptide pools, cell numbers were decreased in donors P19 and P20. Variations in cell expansion between peptide pool stimulations of the same donor were smallest in donors P03 and P20, where harvested cell numbers were in the same range for all pools. Peptide pool-dependent expansion rates varied most strongly in donors P17 and P21. No CMV peptide pool led to an increase in cell numbers in all donors, but pool 3 led to growth rates >1 in 6 out of 8 donors. In donor P21, cell numbers after stimulation with CMV pools 1, 3, and 5 were more than twice as high than after stimulation with the remaining pools. This suggested that this donor reacted particularly strongly to the HLA-A*02:01–restricted epitope VLE and/or its strain variant ILE, which were contained in exactly these pools (1,3,5) and matched the donor's HLA type.

4.8.1 Multimer-positive T cells are enriched by stimulation with CMV peptide pools

Samples from before and after CMV peptide pool stimulation derived from donor P21 were stained with HLA-matched multimers for epitope-specific T cells and analysed by flow cytometry (Figure 4.22). CMV-specific T cells targeting various epitopes presented on different HLAs, namely ELK, NLV, QIK, VLE, VTE, YSE, were already frequent in PBMCs. The measured proportions in ex vivo $CD3^+$ T cells ranged from 0.13% (ELK) to 0.81% (VTE). Multimer-positive T cells against epitopes ELK, NLV, VLE, and VTE were enriched with all 3 pools containing the respective peptide and were not enriched with the remaining 5 pools. T cells against YSE were enriched in the 3 pools containing the peptide (pools 2,4, and 6), but also in pools 1 and 5, which should not contain YSE. No enrichment of multimer-positive T cells was observed in pools containing ELR, HER, or QIK. While QIK-specific cells were nicely stained in the ex vivo sample, no specific cells were observed when staining with ELR or HER multimers. The lack of enrichment of QIK-specific T cells may be related to the fact that the QIK amino acid sequence that was used in the peptide pools contained an error (see Table 2.13). This could mean that donor P21 did not react to the erroneous QIK variant at all, or that T cells that were enriched with the erroneous QIK variant could not be stained with the QIK-multimer with the correct CMV epitope sequence. ELK-specific T cells, but not

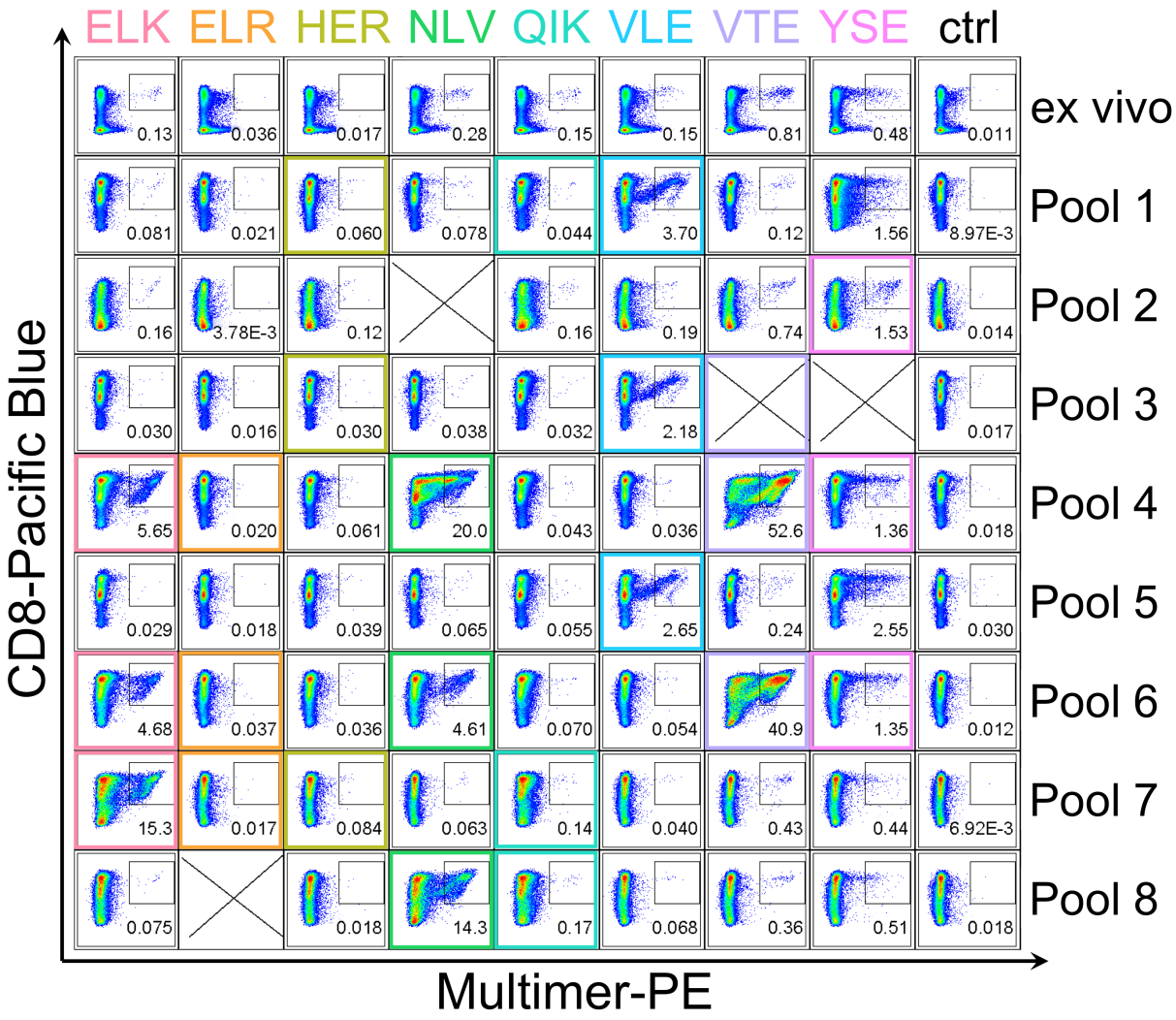


Figure 4.22: Multimer stainings of unstimulated and peptide pool-stimulated samples derived from donor P21 with multimers matching the donor's HLA type. Cells were gated on $CD3^+$ lymphocytes. The multimer stainings of samples stimulated with a CMV peptide pool containing the corresponding peptide are framed by boxes in the colour of the peptide. Some samples could not be analysed since handling errors occurred during staining; such samples are marked wit an "x". All stainings were performed by Fenja Gerpott for her master's thesis under my supervision.

ELR-specific T cells, were found ex vivo and were enriched by stimulation with ELK and/or its variant epitopes ELR, ELK-I, and ELR-I. It has been described before that ELK-specific T cells did not respond to stimulation with ELR peptide (Elkington et al. 2003). Furthermore, a study showed that cross-reactivity of ELR-specific and ELK-specific T cells was donor-dependent (Smith et al. 2016). In that study, ELR-specific T cells were absent in a CMV-infected donor who carried CMV strains that mostly contained the ELR epitope variant, whereas a T-cell response against the subdominant epitope variant ELK existed in that donor. In two other CMV-infected donors of that

study, however, T cells cross-reactive towards ELK and ELR were found. As the flow cytometry data implies, donor P21 belongs to the class of donors with T cells targeting exclusively the epitope variant ELK. Moreover, donor P21 had no HER-specific T cells in their repertoire, which is in agreement with previous reports in which HER-specific T cells were only found in a subset of donors expressing the restricting HLA-B*40:01 and/or HLA-B*40:02 (Kondo et al. 2004; Trivedi et al. 2005). In summary, T cells specific for 6 of 8 HLA-matched CMV peptides were present in the ex vivo repertoire of donor P21, and T cells specific for 5 of these 6 CMV peptides (all except QIK) were enriched by stimulation with the relevant peptide pools.

When looking at the cumulative proportions of CMV multimer-positive T cells in samples of donor P21 (Figure 4.23A), most multimer-positive T cells were found after stimulation with pool 4 (80%) followed by pool 6 (51%). In these two samples, VTE-specific T cells were most abundant, followed by NLV-specific T cells and ELK-specific T cells. VLE-specific T cells were only enriched in pools 1, 3 and 5, but their proportions in the stimulated samples were $<$5%. Since T cells specific for the VLE/ILE pool combination (Table 3.4) appear to be dominant in this donor (Section 4.8), this donor's CMV-specific $CD8^+$ T-cell repertoire seems dominated by strain-specific T cells that recognise ILE, but not VLE. An ILE/HLA-A*02:01 multimer to verify this hypothesis was not available at the time. There appears to be no reduction of relative frequencies of epitope-specific T cells due to competition in pools where the donor reacted to multiple epitopes. For example, the proportion of NLV-specific T cells in the sample stimulated with pool 4 was not smaller than its proportion in pool 8, although donor P21 had a strong T cell response against VTE in pool 4 and reacted to no other specificity stained for in pool 8. By multiplication of the number of cells harvested by the proportion of multimer-positive cells, the absolute number of multimer-positive T cells in the samples can be estimated (Figure 4.23B). Numbers of multimer-positive T cells in the relevant pools were comparable for specificities ELK (1.31–3.44 million cells) and VLE (1.63–2.45 million cells), but varied more for specificities NLV (1.29–7.45 million cells) and VTE (11.47–19.61 million cells). YSE-multimer–positive T cells were most abundant with pools 1 and 5, which did not contain YSE peptide. This multimer showed a relatively high staining in all samples, and the staining pattern took the form of an extended "smear" between negative and positive cells (Figure 4.22); the quality of the multimer

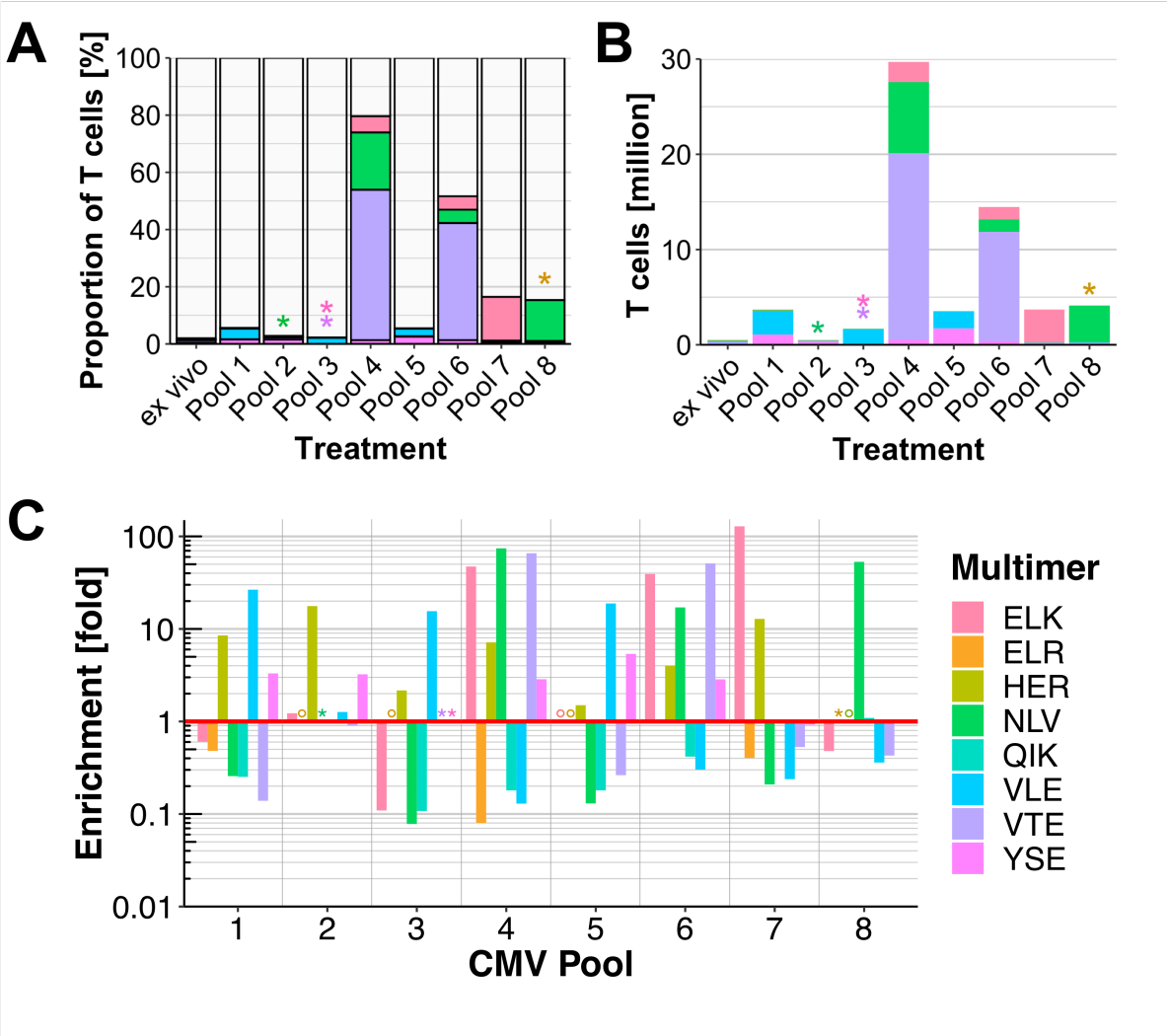


Figure 4.23: Flow cytometry confirms selective enrichment of multimer-positive T cells after pool stimulation. Stainings of unstimulated and peptide pool-stimulated samples derived from donor P21 and stained with HLA-matched multimers. Cells were gated on $CD3^+$ lymphocytes and the background from samples stained without peptide-MHC monomer or multimer was subtracted. All stainings were performed by Fenja Gerpott for her master's thesis under my supervision. (**A**) Relative frequencies of multimer-positive T cells in ex vivo PBMCs and after stimulation with CMV peptide pools 1–8. (**B**) Number of multimer-positive T cells before and after peptide pool stimulation. Absolute numbers of multimer-positive T cells were calculated by multiplying the proportion of multimer-positive T cells determined by flow cytometry with the harvested amount of cells in a sample. (**C**) Enrichments of multimer-positive T cells by CMV peptide pool stimulations were calculated by dividing the relative frequencies of multimer-positive cells after stimulation by their frequencies ex vivo. The red line marks no enrichment through peptide pool stimulation. $*$ Unusable for analysis due to handling errors during staining; $\circ$ Background-corrected frequency in stimulated sample ≤ 0.

reagent may have suffered during its long-term storage. Since the relative frequency of CMV epitope-specific T cells after stimulation is influenced by the ex vivo frequency of multimer-positive T cells, the enrichment of multimer-positive T cells by peptide pool stimulation was also analysed (Figure 4.23C). T cells specific for ELK, NLV, VLE, and VTE were enriched exclusively by stimulation with 3 peptide pools containing the respective peptides and depleted in the remaining 5 pools. Enrichment of T cells with one of these specificities ranged from 16-fold (VLE, pool 3) to 129-fold (ELK, pool 7) when stimulated with matching peptide pools and from 0.08-fold (NLV, pool 3) to 1.3-fold (VLE, pool 2) when stimulated with mismatched peptide pools. Consequently, there was a more than 10-fold difference in enrichment of T cells specific for ELK, NLV, VLE, and VTE after stimulation with matched peptide pools than after stimulation with mitmatched peptide pools. Taken together, the flow cytometry results demonstrate that T cells with different specificities can be specifically enriched simultaneously by short-term CMV peptide pool stimulation.

4.8.2 Peptide pool stimulation identifies TCRβ clonotypes of various specificities

High-throughput sequencing of TCRβ chains before and after CMV peptide pool stimulations (median reads=2.1×10^6, Table A9) and subsequent bioinformatic identification of specific TCRβ clonotypes as described in Section 3.5.3.6 led to the identification of 944 CMV epitope-specific TCRβ sequences in 8 seropositive donors (Table 4.5). These TCRβ sequences were nucleotide-unique in each of the donors, but were counted multiple times if found with identical nucleotide sequence in different donors. The 944 TCRβ sequences included 92 TCRβ sequences that were enriched in every pool that contained at least one of two strongly overlapping peptides: DEE+EEA, KQI+QIK, and RPH+HER. Both peptides of each pair are independent CMV epitopes with different HLA restrictions. However, because of their large sequence overlap it cannot be excluded that under the conditions of our experiment one peptide may bind to the other HLA, although probably less efficiently. Therefore, the 3 pairs of strongly overlapping CMV peptides were included in this analysis. Of the 944 CMV-specific TCRβ sequences, 625 (66.2%) were specific for epitopes known to be restricted by HLAs matching the donor's HLA type. For all CMV epitopes tested, at least one specific TCRβ sequence was found in HLA-matched donors, with the exception of epitopes

Table 4.5: Number of specific TCRβ sequences found by peptide pool stimulation per peptide and donor. TCRβ sequences that can be presented by the donor because they have the matching HLA allele are shown in black font and HLA-mismatched peptides are shown in red font. Three peptides strongly overlapped and may be presented on 2 distinct HLA alleles; if only one of 2 alleles matched the donor's HLA type, numbers are printed in gray font. The numbers in parentheses indicate the numbers of HLA-matched TCRβ sequences for each donor and epitope.

Epitope	P01	P03	P11	P17	P18	P19	P20	P21	Sum Epitope
ATV	0	0	1	0	0	4	1	1	**7** (1)
AYA	0	0	0	2	0	7	0	0	**9** (0)
CED	0	0	0	0	1	1	3	2	**7** (3)
CRV	9	8	12	14	9	1	0	0	**53** (52)
DAL	0	0	0	0	0	3	1	0	**4** (0)
DEE	0	1	0	0	0	3	5	1	**10** (5)
DEE+EEA	0	0	82	0	0	1	1	0	**84** (0)
EEA	0	0	0	0	1	2	4	8	**15** (9)
EFF	0	0	1	0	0	45	0	0	**46** (0)
ELK	0	0	1	10	0	1	3	14	**29** (24)
FPK	0	0	7	1	0	2	1	0	**11** (7)
FPT	0	0	0	0	0	1	0	0	**1** (1)
FRC	24	61	30	60	63	1	3	1	**243** (238)
GPI	0	2	0	0	4	1	3	1	**11** (0)
HER	0	1	1	0	0	4	3	0	**9** (1)
IPS	0	0	0	0	0	15	1	0	**16** (15)
KEV	0	0	0	1	10	4	2	2	**19** (12)
KLG	0	0	0	0	2	2	1	0	**5** (3)
KQI	0	0	0	0	0	2	2	0	**4** (4)
KQI+QIK	0	1	1	0	0	2	1	0	**5** (0)
NLV	0	10	0	11	22	1	6	31	**81** (52)
NQW	0	1	0	0	0	2	4	0	**7** (2)
QIK	0	2	11	0	1	9	3	1	**27** (1)
QYD	0	1	0	0	0	0	0	0	**1** (0)
RCP	0	1	5	0	1	2	3	3	**15** (6)
RIK	0	0	0	0	0	3	5	0	**8** (8)
RPH	1	4	3	0	0	1	1	1	**11** (8)
RPH+HER	0	0	0	0	0	0	2	1	**3** (0)
TPR	1	3	4	1	0	2	2	0	**13** (9)
TRA	0	0	0	0	0	5	20	0	**25** (25)
TSD	0	0	0	0	0	5	1	1	**7** (6)
VLE	0	1	0	0	0	4	0	10	**15** (11)
VTE	3	0	2	11	0	19	1	47	**83** (77)
VYA	0	0	0	3	1	1	3	2	**10** (0)
YSE	0	1	0	5	0	22	4	18	**50** (45)
Sum Donor	**38** (35)	**98** (88)	**161** (62)	**119** (112)	**115** (85)	**178** (76)	**90** (33)	**145** (134)	**944** (625)

HER, RIK, TSD, and VLE. Specific TCRβ sequences against the latter epitopes were found in 1/3 HLA-matched donors (HER), or 2/3 HLA-matched donors (RIK, TSD, VLE). This is not surprising, since CMV epitopes are not equally immunogenic or, as it is commonly termed, immunodominant. For example, VLE-specific T cells were found in 6/18

donors in a study by Khan et al. (Khan et al. 2002a), in which the epitope was identified, and in 2/6 donors and 14/26 donors in other studies (Nastke et al. 2005; Khan et al. 2004). HER appears to be even less immunodominant, as specific T cells have been found in 0/10 CMV-positive donors by multimer staining ((Nastke et al. 2005)) and in 2/4 CMV-positive donors by IFNγ screening (Trivedi et al. 2005). Consistently, no HER/HLA-B*40:01 multimer-positive T cells were found in donor P21 in this project (Figure 4.22). A low number of specific TCRβ sequences from donors lacking the expected HLA type was observed for specificities CRV, ELK, FRC, IPS, KQI, RIK, TRA, TSD, VTE, and YSE. For these specificities, $\geq$80% of identified TCRβ sequences were found in donors expressing the HLA molecule that was expected to present the peptide. Most TCRβ sequences in an HLA-mismatched donor were found in donor P19 for EFF followed by donor P18 for NLV. For the overlapping specificities DEE+EEA, KQI+QIK, and RPH+HER, 0–2 specific TCRβ sequences were found per donor expressing one of 2 matching HLAs. The low number of TCRβ sequences with overlapping specificity implies that the two overlapping peptides of each pair are unlikely presented on the same HLA, and that such peptides are not recognised by the same TCRβ clonotypes. The only donor expressing both HLAs for one of the overlapping epitope pairs, namely donor P03 for RPH+HER, did not mobilise any TCRβ sequences recognising both RPH and HER on their respective HLAs, although 4 and 1 TCRβ sequences were found for each individual epitope. In donor P11, 82 specific TCRβ sequences were found for the overlapping epitope pair DEE+EEA, although the donor expressed none of the required HLAs to present either peptide. The DEE+EEA double-specific TCRβ sequences were of low frequency in the ex vivo sample (0 or 1 reads) and at the low-frequency end of specific TCRβ sequences. Hence it is likely that these TCRβ sequences are, in fact, unspecific background. By far the largest number of unique HLA-matched specific TCRβ sequences was found for FRC; for this epitope, between 24–63 specific TCRβ sequences were identified per donor. Other epitopes eliciting a diverse specific T-cell response on the TCRβ clonotype level in HLA-matched donors were epitopes CRV (8–14), ELK (10–14), NLV (10–31), TRA (5–20), VTE (11–47), and YSE (5–22).

In conclusion, peptide pool stimulation specifically enriched TCRβ sequences against epitopes presented on HLA-molecules matching the donor's HLA type. Nevertheless, there was occasional enrichment of specific TCRβ sequences against epitopes that

were not expected to be presented by certain donors, because these donors lacked the known matching HLA. An example are the 45 EFF-specific TCRβ sequences found in donor P19 who did not express the known restricting HLA-B*44:02. The online tool NetMHC 4.0 Server (http://www.cbs.dtu.dk/services/NetMHC/ (Nielsen et al. 2003; Andreatta and Nielsen 2015)) predicts strong binding affinity of the 9-mer "FFWDANDIY", which is fully contained within EFF, to two HLA class I alleles expressed by donor P19, namely HLA-B*35:01 (affinity = 20.97 nM) and HLA-C*04:02 (affinity = 1404.15 nM). Hence, it is possible that T cells enriched with EFF in donor P19 actually recognise the slightly different CMV peptide "FFWDANDIY" presented on HLA-B*35:01 or HLA-C*04:01. Having said that, it was beyond the scope of this study to analyse whether some of the unexpected combinations of HLA type and peptide were due to recognition of bona fide epitopes that are presently unknown. Consequently, specific TCRβ clonotypes identified from HLA-mismatched donors were excluded from further analyses in this thesis. Likewise, TCRβ sequences specific for the dual CMV peptide combinations DEE+EEA, KQI+QIK, and RPH+HER were excluded from further analyses, since such dual recognition of overlapping peptides by TCRβ clonotypes was rarely observed, and such TCRβ sequences were low in frequency.

4.8.3 Overlap of CMV-specific TCRβ repertoires identified by single peptide or peptide pool stimulation

PBMCs from donors P01 and P03 were stimulated both with single CMV peptides and with CMV peptide pools 1–8. Overall, peptide pool stimulation led to the identification of fewer epitope-specific TCRβ sequences than single peptide stimulation (see Table 4.2). In donor P01, sequencing of 1 ex vivo sample and 4 single peptide-stimulated samples resulted in the identification of 209 specific TCRβ sequences. In donor P03, sequencing of 1 ex vivo sample and 6 single peptide-stimulated samples resulted in the identification of 115 specific TCRβ sequences. With the pool stimulation assay, 1 ex vivo sample and 8 peptide pool-stimulated samples were sequenced per donor. Although the peptide pools contained additional peptides matching the HLA type of each donor, namely KLG and RIK (P01) and HER (P03), the number of specific TCRβ sequences found by pool stimulation was lower than by single peptide stimulation: 38 epitope-specific TCRβ sequences were found using pool stimulation in donor P01 and 98 specific TCRβ sequences were found in donor P03 (Table 4.5).

When comparing the specific TCRβ sequences identified by single peptide stimulation to those found by peptide pool stimulation, there was a considerable overlap between specific TCRβ sequences (Table 4.6). In all 4 epitopes studied in donor P01, and in three of the epitopes (CRV, FRC, NLV) studied in donor P03, most or all of the TCRβ sequences that were identified as specific in the sample of lower resolution (i.e. with the smaller number of specific TCRβ, either single peptide or peptide pool) were also identified as specific in the other sample. TCRβ sequences that were found in both

Table 4.6: Number of specific TCRβ sequences found by single peptide stimulation, pool stimulation, or both assays in donor P01 and donor P03. Numbers in parentheses indicate how many of the identified TCRβ sequences were also specifically enriched by stimulation with the matching mini-LCL. Since no UL29/28-expressing mini-LCL was available, FRC-specific TCRβ sequences could not be tested for recognition of endogenously processed peptide.

Peptide	Donor P01			Donor P03		
	Single peptide	Peptide pool	Both	Single peptide	Peptide pool	Both
CRV	119 (94)	9 (8)	9 (8)	39 (22)	8 (8)	8 (6)*
FRC	24	24	15	16	61	16
NLV				8 (4)	10 (8)	4 (4)
RPH	27 (20)	1 (1)	1 (1)	15 (6)	4 (0)	0
TPR	39 (36)	1 (1)	1 (1)	22 (2)	3 (1)	0
VLE				15 (4)	1 (0)	0

* In the pool stimulation assay, 2 TCRβ sequences were assigned different specificities (1× KQI+QIK, 1× VLE). They were not enriched by stimulation with IE1-expressing mini-LCL.

assays mostly also responded to endogenously processed antigen in the mini-LCL stimulations. For example, 9 CRV-specific TCRβ sequences were identified both by single peptide and by peptide pool stimulation in donor P01, and 8 of these 9 TCRβ sequences were specificially enriched by stimulation with the IE1-expressing mini-LCL. Conversely, equal or even higher numbers of TCRβ sequences specific for FRC or NLV were found by peptide pool stimulation than by single peptide stimulation. While 15 of 24 TCRβ sequences found by single peptide or peptide pool stimulation overlapped in donor P01, all 16 FRC-specific TCRβ sequences identified by single peptide stimulation in donor P03 were also found by peptide pool stimulation of the same donor. 10 NLV-specific TCRβ sequences were found by peptide pool stimulation in donor P03, including 4 TCRβ sequences that were specifically enriched by pp65 mini-LCL–stimulation but not by single peptide stimulation. All in all, these results show that CMV-specific TCRβ sequences can be identified by both single peptide stimulation and

peptide pool stimulation, and that a substantial number of these TCRβ sequences also respond to endogenously processed and presented antigen. The number of specific TCRβ sequences identified by peptide pool stimulation was, in most cases, lower than with single peptide stimulation, which may be a result of different cut-off criteria for the identification of CMV-specific TCRβ sequences in both assays. The cut-off criteria may have been more stringent for the peptide pool stimulation assay, because more control samples were included, and consequently fewer specific TCRβ sequences were identified.

In order to understand why the yield of specific TCRβ sequences was comparatively low in the peptide pool stimulation assay, the frequencies of epitope-specific TCRβ sequences in the single peptide-stimulated samples of donors P01 and P03 were compared to their frequencies in CMV peptide pools 1–8 (Figure 4.24). Generally, there was a positive correlation between clonotype frequencies in the single peptide-stimulated and peptide pool-stimulated samples: TCRβ sequences frequencies increased in all pools with increasing frequencies in the single peptide-stimulated samples. TCRβ sequences specific for FRC and NLV in donor P03 were well-separated and 2–3 orders of magnitude more frequent in the stimulations with the 3 pools containing the peptide than in the remaining 5 stimulations. The 4 lower-frequency NLV-specific TCRβ clonotypes from single peptide stimulation were not enriched with any of the 3 pools containing NLV. These four TCRβ sequences may not be specific for NLV; they were not enriched by stimulation with pp65 mini-LCLs either. The top FRC-specific TCRβ sequences derived from single peptide stimulation of donor P01 were also more frequent in the 3 samples stimulated with pools containing FRC than in the 5 samples stimulated with pools lacking FRC, but for the lower-frequency FRC-specific TCRβ sequences this was not seen as clearly. Most CRV-specific TCRβ sequences were often enriched with 6 of 8 pools (pools 1,2,3,5,7, and 8) and depleted in pools 4 and 6. This unexpected observation was made in both donors and raised the question of whether (a) these TCRs may be specific for more than one peptide, or (b) the three other pools may have been contaminated with the CRV peptide. Pools 1, 5, and 8 were supposed to contain the CRV peptide. Pools 2, 3 and 7 were the pools that contained the other HLA-C–restricted CMV-peptide FRC, and this peptide was handled immediately after CRV when lyophilised peptides were dissolved and distributed to pools. RPH-specific

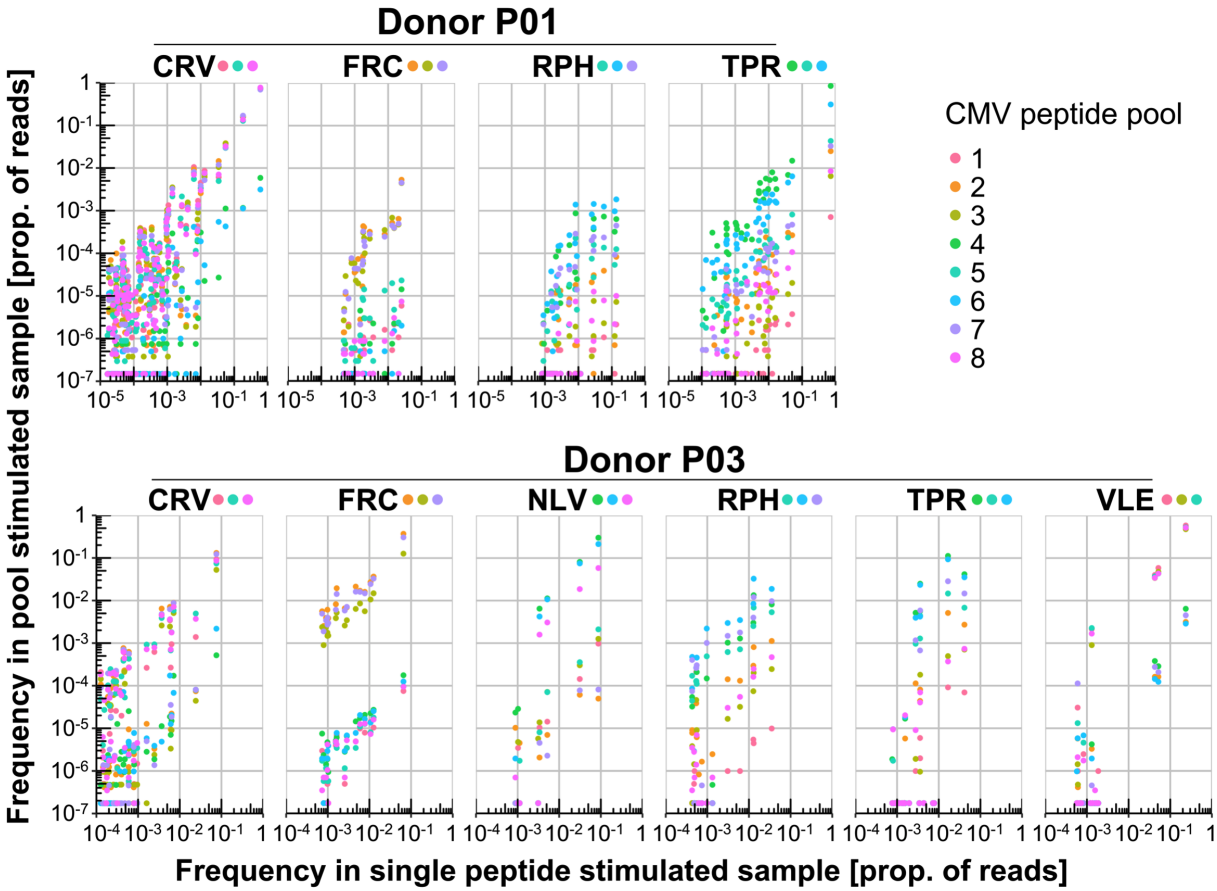


Figure 4.24: Frequencies of CMV-specific TCRβ clonotypes in the CMV peptide pool-stimulated samples of donor P01 and P03. All specific TCRβ sequences that were identified by single peptide stimulations in the respective donor are displayed in the plots. The proportional read frequency of each clonotype in the single peptide-stimulated sample are shown on the x-axis and the corresponding frequencies in each of the 8 pools are plotted on the y-axis. Consequently, each specific TCRβ clonotype identified by single peptide stimulation is depicted as 8 dots with identical x-axis values, but different y-axis values. TCRβ clonotypes that were not present in a pool were assigned a pseudofrequency of 0.5× the minimum proportion of reads a clonotype had in this analysis. The circles next to the peptide name in the plot headings correspond to the pools in which the peptide was contained.

and TPR-specific TCRβ sequences were enriched in pools 4,5,6, and 7, hinting at a potential cross-contamination of these 2 peptides in the CMV peptide pools; RPH was handled immediately after TPR at setup. VLE-specific TCRβ sequences identified by single peptide stimulation of donor P03 were enriched with pool 8 in addition to pools 1,3, and 5 in which VLE peptide was contained. If the reason for this was a contamination of another peptide with VLE, peptides CRV or KQI were likely candidates, since they appear only in pool 8 apart from two of the three pools 1, 3, and 5. Because TCRβ clonotypes against CRV, RPH, TPR, and VLE were enriched by stimulation with more

than the 3 pools containing the epitope, many TCRβ sequences specific for these epitopes could not be found using the bioinformatics data analysis pipeline established for peptide pool stimulations. Peptides FRC and NLV, however, appeared not affected by this hypothetical cross-contamination, which explains the high numbers of specific TCRβ sequences found for these epitopes using the peptide pool stimulation assay.

To test the hypothesis that cross-contamination of the peptides had occurred due to handling errors when solutions were mixed to prepare CMV peptide pools, T-cell clones specific for CRV or FRC were co-incubated with peptide stock solutions or peptide pools that contained either CRV, FRC, or neither of the two peptides (Figure 4.25). Older peptide stocks that were independently synthesised and handled were available for this experiment; these stocks had been used for all single peptide stimulation experiments, but not for the peptide pool stimulation experiments. The CRV-specific

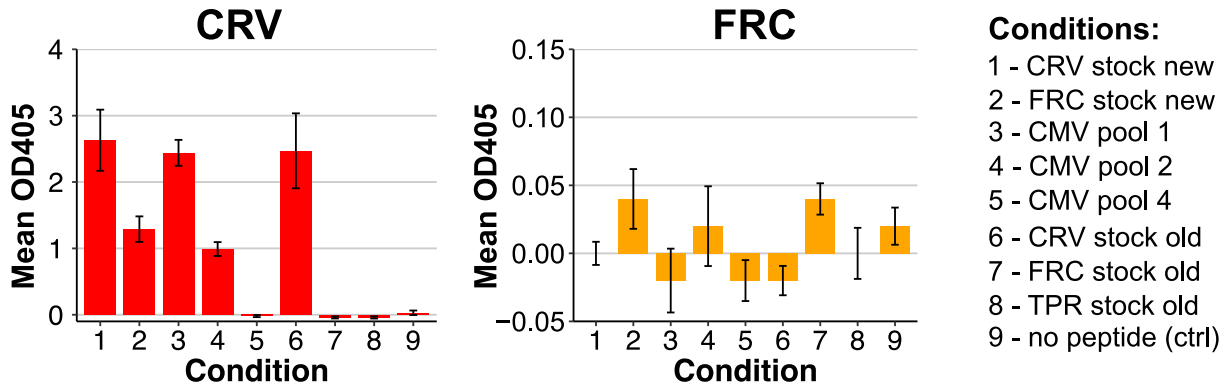


Figure 4.25: IFNγ secretion of a CRV-specific T-cell clone (made by Steffi Ameres) and an FRC-specific T-cell clone (made by Xiaoling Liang) in response to different CMV peptide stocks and pools. 9 different conditions were tested with both T-cell clones: the new CRV and FRC stock solutions used to make the peptide pools (1–2), three different CMV peptide pools, of which one only contained CRV (3), one only contained FRC (4), and one contained neither peptide (5), the old peptide stocks used for all single peptide stimulation assays (6–8), and a negative control containing no peptide. 10,000 T cells were co-incubated with 20,000 peptide-loaded CD40-stimulated B cells overnight to induce IFNγ secretion. Per well, 1 μg peptide was loaded onto the B cells and excess peptide was washed off. Each condition was tested in triplicates (n=3) for each T-cell clone, error bars indicate the standard deviation of measured OD405 values.

and the FRC-specific T-cell clone were produced and kindly provided by Steffi Ameres and Xiaoling Liang, respectively. IFNγ secretion was measured in an IFNγ-ELISA. As expected, the CRV-specific T-cell clone strongly reacted to both the old CRV stock solution used in the single peptide stimulations and the new stock solution used in the peptide pool stimulations. In addition, it secreted IFNγ in response to peptide pool 1,

which contained CRV peptide. IFNγ secretion was similar in all 3 samples containing CRV, as shown by the similar OD405 values (mean=2.4–2.6). However, the CRV-specific T-cell clone also secreted IFNγ in response to stimulation with the new FRC stock solution and peptide pool 2 containing FRC, even though the OD405 values were approximately halved compared to the intentionally CRV-containing samples. These results confirmed that the new FRC stock solution and the pools produced from it were contaminated with CRV peptide due to handling errors. This explains the enrichment of CRV-specific TCRβ clonotypes with pools 2, 3, and 7, all of which should not have contained CRV. The FRC-specific T-cell clone was barely functional, as demonstrated by the very low OD405 values measured in the IFNγ-ELISA. Nevertheless, there was a slight tendency of higher IFNγ-secretion in the samples containing FRC than in those were FRC was supposed to be absent. Together with the observation that FRC-specific TCRβ clonotypes were significantly more frequent in the 3 pools containing FRC than in the remaining 5 pools (Figure 4.24), it can be tentatively concluded that FRC peptide was exclusively contained in the 3 dedicated pools. The comparison between TCRβ clonotypes identified after single peptide and peptide pool stimulations (Table 4.6, Figure 4.24) together with the control experiment shown in Figure 4.25 suggested that, while identification of some specific TCRβ clonotypes was not possible due to peptide contamination, those TCRβ clonotypes that were identified as specific were most likely correctly assigned to their CMV epitope specificity. Therefore, the peptide pool-derived TCRβ data were further analysed, albeit with special caution.

There was a strong positive correlation of relative frequencies of TCRβ clonotypes that were deemed specific for CRV or FRC by both single peptide stimulation and peptide pool stimulation of donors P01 and P03 (Figure 4.26). Pearson correlation coefficients of log10-transformed frequency data ranged from r=0.66 for CRV-specific TCRβ sequences of donor P01 to r=0.95 for FRC-specific TCRβ sequences of donor P03. Average frequencies were at least 2 orders of magnitude higher after stimulation with the 3 pools containing CRV or FRC than with the remaining 5 pools. Notably, the frequency distribution patterns of specific TCRβ clonotypes in the 3 matching pool-stimulated samples were very similar. Because the CMV-specific TCRβ clonotypes that were found by both single peptide stimulation and peptide pool stimulation were at least 100-fold more frequent in the 3 pools containing the tested peptides than in the 5

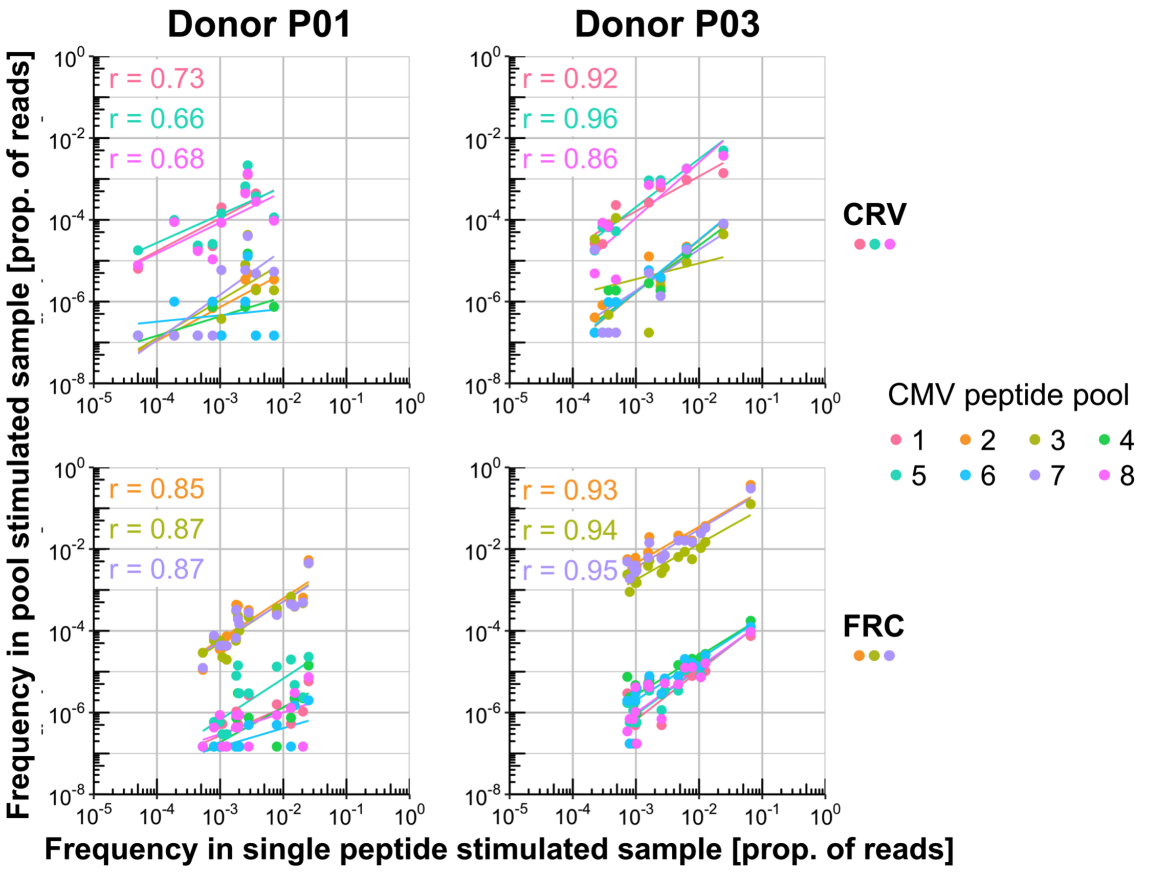


Figure 4.26: Frequency of TCRβ sequences that were identified as specific for CRV or FRC by single peptide stimulation and peptide pool stimulation of donor P01 and P03. The proportional read frequency of each clonotype in the single peptide-stimulated sample are shown on the x-axis, and the corresponding frequencies in each peptide pool-stimulated sample are plotted on the y-axis. The circles next to the peptide name in the plot headings indicate in which pools the peptide was contained. TCRβ clonotypes that were not present in a pool were assigned a pseudofrequency of 0.5× the minimum proportion of reads a clonotype had in this analysis. Pearson *r* values indicate the correlation of the log10-transformed frequencies in the single peptide stimulated samples and in the samples stimulated with one of the matching pools.

other pools, it is implied that stimulation with multiplexed peptides is a valid approach to identify epitope-specific TCRβ sequences.

In summary, there was cross-contamination of certain peptides, such as CRV and possibly other peptides, due to handling errors. Consequently, many TCRβ clonotypes with these specificities could not be identified bioinformatically. Other peptides, such as FRC, were not carried to other pools, which explains why identification of FRC-specific TCRβ clonotypes by peptide pool stimulation was better than for CRV-specific TCRβ clonotypes. There was a substantial overlap between CMV-specific TCRβ clonotypes

identified by peptide pool stimulation of donors P01 and P03 and those identified by single peptide stimulation and mini-LCL–stimulation; this observation provides preliminary validation of peptide pool stimulation as a suitable assay to characterise specific TCRβ clonotypes with multiple CMV peptide specificities simultaneously.

4.8.4 Cumulative frequency of CMV-specific TCRβ clonotypes obtained from peptide pool stimulation

Epitope-specific TCRβ clonotypes identified in 8 CMV-positive donors were traced back in the ex vivo sample and peptide pool-stimulated samples of each donor to assess the frequency of CMV-specific TCRβ clonotypes in each sample (Section 4.8.3, Figure 4.27). TCRβ clonotypes against epitopes ELK, FRC, NLV, and VTE were abun-

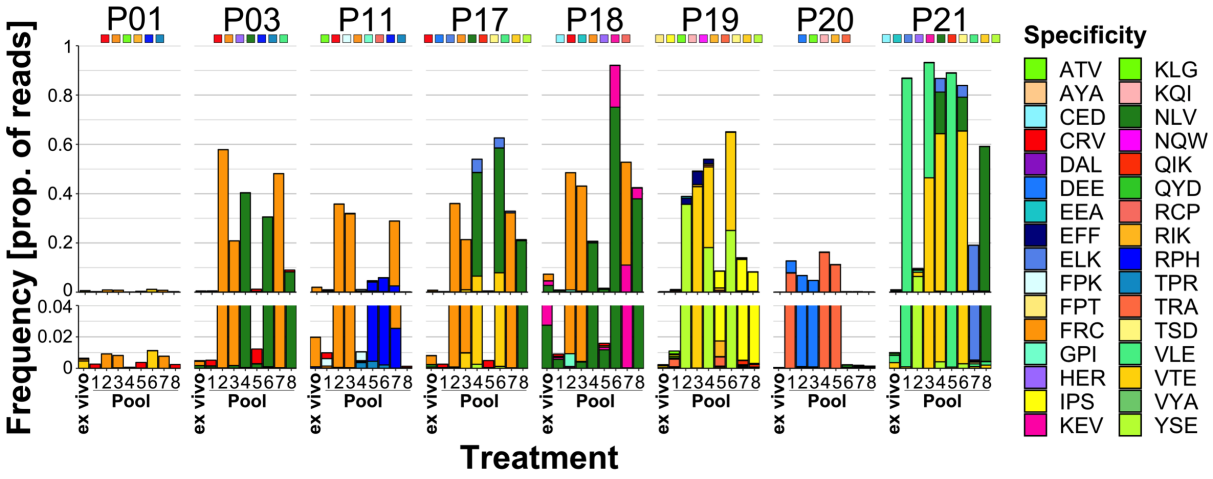


Figure 4.27: Frequencies of specific TCRβ sequences identified with the peptide pool stimulation assay in ex vivo samples and after pool stimulation. The upper panel shows the cumulative frequencies of TCRβ reads for each epitope specificity. The lower panel shows a close-up of the same bar chart. The numbers and colours of squares below each donor name indicate the number and kind of peptides that matched the donor's HLA type.

dant in specific peptide-containing pools and exclusively present in donors expressing the expected HLA molecule for epitope presentation. Cumulative read frequencies of TCRβ sequences specific for one of these four epitopes after peptide pool stimulation varied between 4% (P17, pool 6, ELK) and 65.1% (P21, pool 6, VTE). An exception to this were cumulative frequencies of ELK-specific TCRβ sequences in pool 7 and VTE-specific TCRβ sequences in pool 3 of donor P17, and FRC-specific TCRβ sequences in pools 2, 3, and 7 of donor P01: these specificities contributed less than 1% to the total TCRβ reads of the pool-stimulated samples. The reason for lower cumulative

frequencies of specific TCRβ sequences in these samples may be the presence of a strong competitor peptide which expands specific clonotypes that consume a majority of TCRβ reads in these samples. For example, we know from the single peptide stimulation assay that donor P01 mounts a very strong T-cell response against CRV and 97% of reads belong to CRV-specific TCRβ clonotypes in the stimulated sample (Table 4.2). Most of the 119 CRV-specific TCRβ clonotypes were also specifically enriched with endogenously processed antigen (Figure 4.20). Unfortunately, the majority of CRV-specific TCRβ clonotyes of donor P01 could not be identified by peptide pool stimulation. Despite the inability to identify CRV-specific TCRβ clonotypes in the peptide pool stimulation assay, they were still expanded in all 6 pools containing CRV peptide (pools 1, 2, 3, 5, 7, and 8) and likely accounted for most of the TCRβ reads in these pools. This was implied by the data presented in Figure 4.24, which shows the frequency of the 119 CRV-specific TCRβ clonotypes identified by single peptide stimulation in all 8 peptide-pool–stimulated samples. Because of the great expansion of CRV-specific clonotypes in donor P01, the relative frequency of FRC-specific TCRβ clonotypes in the samples was likely reduced; for in the single peptide stimulation assay, 11% of TCRβ sequences were FRC-specific after stimulation of PBMCs from donor P01 with FRC peptide (Table 4.3). To give more examples for competition between peptides, VTE- and YSE-specific TCRβ sequences were less frequent in samples derived from donor P19 stimulated with pool 4 and pool 6, which contained both peptides, than they were in pool 2 and pool 3, which contained only one of the peptides each. Likewise, the cumulative proportions of TCRβ sequences specific for ELK, NLV, VLE, or YSE were higher in samples stimulated with a pool containing only one peptide that donor P21 reacted to (pools 1, 2, 5, 7, 8) than in the remaining 3 pools, which additionally contained VTE. In conclusion, competition for reads between TCRβ clonotypes with different specificities likely modified cumulative frequencies of epitope-specific TCRβ clonotypes. Particularly the presence of strongly immunogenic peptides may have reduced the magnitude of the T-cell response observed for other CMV peptides in some of the tested donors, even when they could not be identified as specific by computational data analysis. TCRβ sequences specific for 6 CMV epitopes were frequent in pool-stimulated samples of one HLA-matched donor each: 1/1 donor mounted a strong response specific for IPS, 1/2 donors for DEE, KEV, or TRA, and

1/3 donors for VLE or YSE. Cumulative read frequencies of specific TCRβ sequences ranged between 0.04% (P18, pool 8, KEV) and 88.8% (P21, pool 5, VLE). Lack of specific TCRβ sequences against these epitopes in additional HLA-matched donors may be explained by general lack of epitope-specific T cells in these donors, since not all donors have T cells against all HLA-matched epitopes, or by enrichment of specific T cells in more than 3 pools so that they could not be identified as peptide-specific (see Figure 4.24, Figure 4.25).

Donor P18 was negative for HLA-A*02:01, but still mounted a strong T cell response specific for the HLA-A*02:01–restricted CMV epitope NLV: 22 NLV-specific TCRβ sequences accounted for 20–75% of reads in the samples stimulated with NLV-containing pools. Although the donor did not express HLA-A*02:01, they expressed the related alleles HLA-A*02:03 and HLA-A*02:06. This indicates that at least one of the two related HLA molecules is also able to present NLV and to activate NLV-specific T cells. An earlier study by Trivedi et al. (Trivedi et al. 2005) included a donor who expressed HLA-A*02:06, but not HLA-A*02:01; this donor had 0.2% of NLV-specific T cells in their PBMCs and 5% after stimulation with a pp65 peptide library consisting of overlapping 15-mers. Hence, HLA-A*02:06 is possibly the presenting HLA of NLV in donor P18, but a contribution of HLA-A*02:03 cannot be excluded.

In conclusion, the results from peptide pool stimulation show that selective enrichment and identification of epitope-specific TCRβ sequences worked well for some CMV epitopes, while specific TCRβ sequences against other epitopes could not be determined. Possible explanations for the lack of specific TCRβ sequences in donors expressing the expected HLA are (i) lack of a T cell response specific for an epitope in a particular donor, and (ii) cross-recognition of different peptides by the same TCRβ clonotype, which leads to expansion of this clonotype in more than 3 pools, leading to its elimination during quantitative analysis.

4.9 Common features of CMV peptide-specific TCRβ repertoires

In the previous sections, the CMV epitope-specific TCRβ sequences were examined separately for each donor. In the following sections, the epitope-specific TCRβ sequences identified by single peptide stimulation and peptide pools stimulation of PBMCs

derived from donors P01–P08, P11, P14, and P17–P21 will be compared between donors. TCRβ sequences with ambigous specificity derived from the single peptide stimulation assay were excluded from the analyses. Likewise, TCRβ sequences identified as specific for HLA-mismatched epitopes or as specific for overlapping epitopes DEE+EEA, KQI+QIK, and RPH+HER by peptide pool stimulation were excluded. The remaining 1350 monospecific TCRβ sequences from single peptide stimulation and 625 HLA-matched TCRβ sequences from pool stimulation were combined. This resulted in a total of 1921 monospecific TCRβ sequences, which were nucleotide-unique in the donor in which they were identified, but not necessarily nucleotide-unique between different donors (donor-nucleotide-unique). 52 of these 1921 TCRβ sequences were found both by single peptide stimulation and peptide pool stimulation of PBMCs derived from donors P01 and P03.

4.9.1 TCRβ sequences with the same epitope specificity use similar V genes

At first, the sharing of V and J genes used in TCRβ sequences with the same epitope specificity was investigated. For this purpose, VJ gene usage in TCRβ sequences targeting the 8 most intensively tested epitopes were analysed. This group consisted of four pairs of epitopes that were restricted by the same HLA class I molecule: The epitopes were VTE and YSE (HLA-A*01:01), NLV and VLE (HLA-A*02:01), RPH and TPR (HLA-B*07:02), and CRV and FRC (HLA-C*07:02). Figure 4.28 shows the overall VJ gene usages of donor-nucleotide-unique TCRβ sequences with the same epitope specificity and the use of particular VJ combinations. Certain V genes were preferably used by TCRβ sequences with the same specificity, but often in combination with a different J gene. In TCRβ clonotypes of some peptide specificities, a certain Vβ gene segment was nearly exclusively linked to a particular Jβ gene segment. For example, VTE-specific TCRβ sequences frequently used TRBV9 in combination with TRBJ2-7 (18%) and the most frequent VJ combination among RPH-specific TCRβ sequences was TRBV4-3 with TRBJ1-1 (11%). By contrast, the most frequent V genes used in TCRβ sequences with other specificities were more promiscuous in their J gene usage. For example, TRBV28 was with 39% the top V gene used in FRC-specific TCRβ sequences, but the 208 FRC-specific TCRβ clonotypes with TRBV28 used a variety of different J genes, such as TRBJ2-3, TRBJ2-1, TRBJ2-7, TRBJ2-5, and TRBJ1-

1, to comparable extents. Similarly, TRBV7-9 was the most frequently used V gene in TPR-specific TCRβ sequences. These 79 TPR-specific TCRβ clonotypes (33%) used TRBV7-9 in combination with TRBJ1-6, TRBJ1-1, TRBJ2-1, TRBJ2-2, and others. These observations imply that V gene usage is a stronger determinant for TCRβ clonotype specificity than J gene usage. TCRβ sequences with different epitope speci-

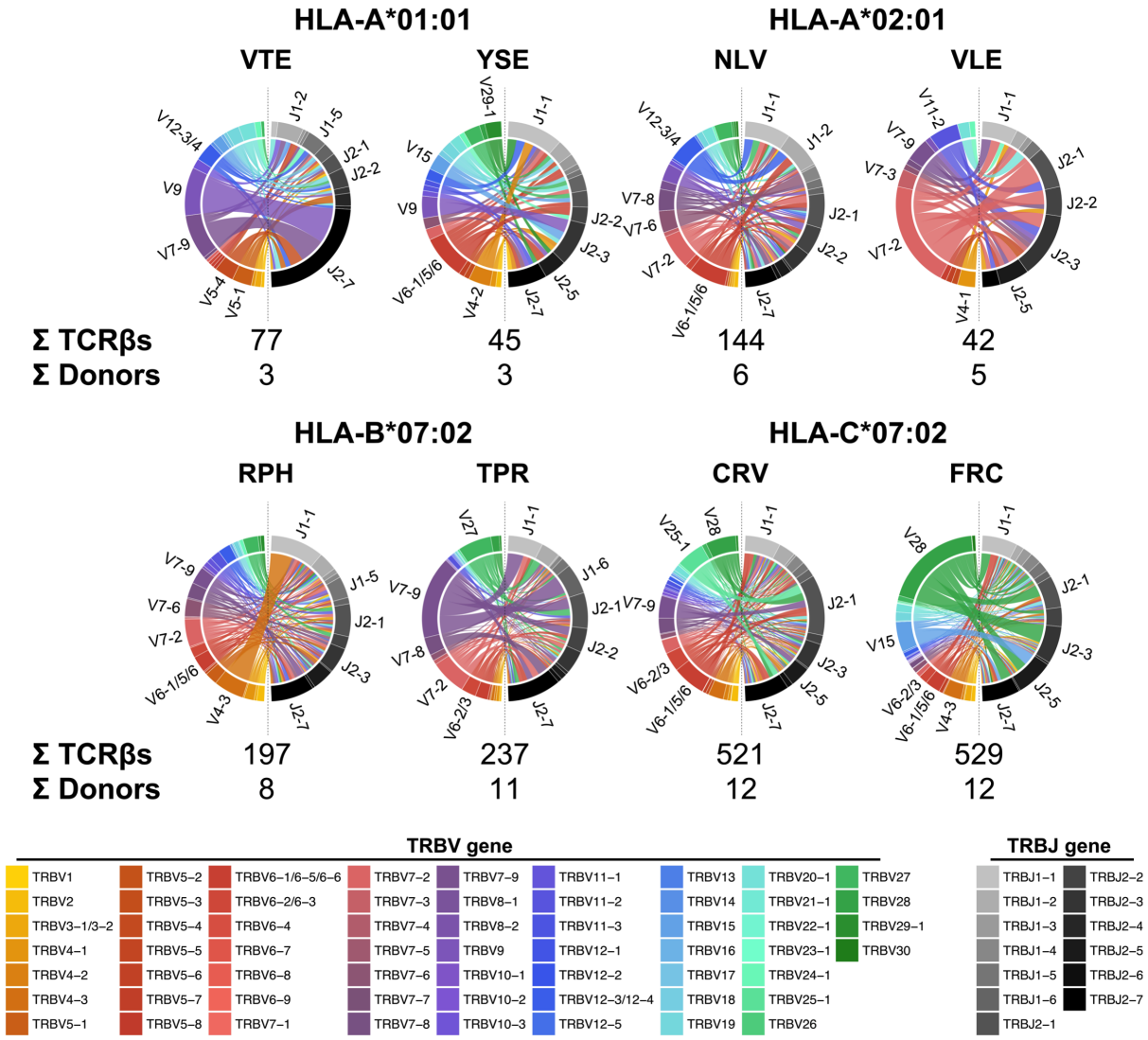


Figure 4.28: Vβ (TRBV) genes and Jβ (TRBJ) genes used in epitope-specific TCRβ sequences with identical CMV epitope specificity. Chord diagrams show the V gene usage (left semi-circle) and J gene usage (right semi-circle) of donor-nucleotide-unique TCRβ sequences specific for epitope CRV, FRC, NLV, RPH, TPR, VLE, VTE, or YSE. The sector widths correspond to the relative proportion of reads using a particular V or J gene. The chord thickness is proportional to the read frequencies of particular combinations of V and J genes in the sample. The top 5 most frequently used gene segments are labelled in each chord diagram. Number below the chord diagrams indicate how many donor-nucleotide-unique TCRβ sequences were found per specificity and from how many donors they were derived.

ficities preferably used different V genes. For instance, VLE-specific TCRβ sequences tended to prefer TRBV11-2 and TRBV7-3, while CRV-specific TCRβ sequences preferably used TRBV25-1 and TRBV6-1/6-5/6-6. FRC-specific TCRβ sequences frequently used TRBV28 and TRBV15. An exception to this was TRBV7-2, which was often used in TCRβ clonotypes with various epitope specificities. This V gene, however, was also the most frequent V gene in ex vivo repertoires of all tested donors and therefore may have been used in TCRβ sequences that were not truly epitope specific, but were rather background noise (Section 4.4.7, Section 4.4.8). Remarkably, one V gene was frequently used in TCRβ sequences specific for each pair of epitopes restricted by the same HLA allele. Many TCRβ sequences specific for HLA-A*01:01–restricted epitopes used TRBV9, while HLA-A*02:01-restricted TCRβ sequences often used TRBV7-2. HLA-B*07:02-restricted TCRβ sequences preferably used TRBV27 and HLA-C*07:02–restricted TCRβ sequences frequently used TRBV28. Since CDR1 and CDR2 of the TCRβ chain V region are mainly in contact with the MHC molecule and seldom interact with the presented peptide (Rudolph et al. 2006; Ishizuka et al. 2008; Merkle et al. 2017; Cole et al. 2017), conserved V gene usage of TCRβ sequences with the same epitope specificity indicates that TCRβ sequences specific for different CMV epitopes with identical HLA restriction may bind in a similar orientation to their target peptide-MHC complex.

4.9.2 Several TCRβ sequences were specifically enriched with the same CMV peptide in at least 2 donors

The analysis of specific TCRβ sequences was next taken to the amino acid level, because although the underlying nucleotide sequence may vary, the resulting protein sequence defines the specificity of a TCRβ chain. All in all, 1809 amino-acid-unique epitope-specific TCRβ sequences were found by single peptide and peptide pool stimulation. 45 TCRβ sequences were found to be specifically enriched with the same CMV peptide in 2–6 donors (Table 4.7). One additional TCRβ sequence was specifically enriched with two different peptides in two donors, namely with CRV in P07 and VLE in P08. This sequence was excluded from the following analyses. 12 specific TCRβ chains with identical CDR3β amino acid sequence were found more than once in the same donor, but with different underlying nucleotide sequences, suggesting that these TCRs originate from independently formed T-cell clones. The remaining 1751

Table 4.7: List of 46 amino acid-unique TCRβ clonotypes shared between multiple donors. The rightmost 3 columns show the sum of donors in which this TCRβ sequences was specifically enriched (Σ**Donors**), the total number of occurences among the 1921 monospecific TCRβ sequences (**Occ.**), and the number of nucleotide variants encoding the TCRβ amino acid sequence (**nt var.**).

V gene	CDR3 sequence	J gene	Specificity	ΣDonors	Occ.	nt var.
TRBV4-3	CASSPQRNTEAFF	TRBJ1-1	RPH	6	9	6
TRBV15	CATSREGGETQYF	TRBJ2-5	FRC	5	5	4
TRBV25-1	CASSPGDEQFF	TRBJ2-1	CRV	4	8	8
TRBV25-1	CASTPGDEQFF	TRBJ2-1	CRV	4	5	5
TRBV4-3	CASSPARNTEAFF	TRBJ1-1	RPH	4	4	4
TRBV7-8	CASSFRTVSSYEQYF	TRBJ2-7	TPR	4	4	4
TRBV7-9	CASSLIGVSSYNEQFF	TRBJ2-1	TPR	3	5	5
TRBV6-2/6-3	CASSYGLEAFF	TRBJ1-1	FRC	3	5	4
TRBV15	CATSRTGGETQYF	TRBJ2-5	FRC	3	4	4
TRBV7-6	CASSLAPGATNEKLFF	TRBJ1-4	NLV	3	4	3
TRBV12-3/12-4	CASSSVNEAFF	TRBJ1-1	NLV	3	3	3
TRBV6-2/6-3	CASSGGLEAFF	TRBJ1-1	FRC	3	3	3
TRBV6-1/6-5/6-6	CASSSGQKNTEAFF	TRBJ1-1	CRV	3	3	2
TRBV6-2/6-3	CASSLGLEAFF	TRBJ1-1	FRC	2	5	5
TRBV9	CASSAGQGVTYEQYF	TRBJ2-7	VTE	2	5	5
TRBV28	CASSFPDTQYF	TRBJ2-3	CRV	2	5	4
TRBV6-2/6-3	CASSYSSGELFF	TRBJ2-2	TPR	2	4	3
TRBV12-3/12-4	CASSSANYGYTF	TRBJ1-2	NLV	2	3	3
TRBV28	CASSLEQGAGETQYF	TRBJ2-5	FRC	2	3	3
TRBV28	CASTPWGAEAFF	TRBJ1-1	CRV	2	3	3
TRBV4-3	CASSPSRNTEAFF	TRBJ1-1	RPH	2	3	3
TRBV9	CASSVGTGSNTEAFF	TRBJ1-1	RIK	2	3	3
TRBV27	CASSLSPSTGNYGYTF	TRBJ1-2	TPR	2	3	2
TRBV10-2	CASSESTGFDGYTF	TRBJ1-2	CRV	2	2	2
TRBV15	CATSRVAGETQYF	TRBJ2-5	FRC	2	2	2
TRBV20-1	CSAPDWNNEQFF	TRBJ2-1	CRV	2	2	2
TRBV25-1	CASSPGDEQYF	TRBJ2-7	CRV	2	2	2
TRBV27	CASSLEGYTEAFF	TRBJ1-1	NLV	2	2	2
TRBV28	CASEDTGELFF	TRBJ2-2	FRC	2	2	2
TRBV28	CASSLADLNQPQHF	TRBJ1-5	FRC	2	2	2
TRBV28	CASSLGDTNQPQHF	TRBJ1-5	FRC	2	2	2
TRBV28	CASSPGFSGNTIYF	TRBJ1-3	FRC	2	2	2
TRBV28	CASSPISNEQFF	TRBJ2-1	CRV	2	2	2
TRBV28	CASSPVSNEQFF	TRBJ2-1	CRV	2	2	2
TRBV28	CASSSLNSNQPQHF	TRBJ1-5	FRC	2	2	2
TRBV4-3	CASSPHRNTEAFF	TRBJ1-1	RPH	2	2	2
TRBV4-3	CASSPNRNTEAFF	TRBJ1-1	RPH	2	2	2
TRBV6-1/6-5/6-6	CASSPGTPRDEQFF	TRBJ2-1	CRV	2	2	2
TRBV6-2/6-3	CASSYSGNTEAFF	TRBJ1-1	TPR	2	2	2
TRBV7-2	CASSSRGTVNTEAFF	TRBJ1-1	TPR	2	2	2
TRBV7-8	CASSFRTVNSYEQYF	TRBJ2-7	TPR	2	2	2
TRBV7-8	CASSLRTVSSYEQYF	TRBJ2-7	TPR	2	2	2
TRBV7-9	CASSFRQGVNTGELFF	TRBJ2-2	TPR	2	2	2
TRBV7-9	CASSLHTQGARTEAFF	TRBJ1-1	TPR	2	2	2
TRBV28	CASSLEQLSGNTIYF	TRBJ1-3	FRC	2	2	1
TRBV7-9	CASSLTQDQETQYF	TRBJ2-5	CRV;VLE*	2	2	1

* This TCRβ sequence was found in 2 different donors with diverging specificities. It was therefore excluded from further analyses.

TCRβ sequences were only found in one donor and with one nucleotide variant.
Up to 3 different nucleotide sequences translating to the same TCRβ amino acid sequence were found in a single donor, and up to 8 different nucleotide variants were found per amino acid-unique TCRβ sequence in all tested donors taken together. Of the 45 specific TCRβ sequences that were shared between multiple donors, 33 were found to be specific in 2 donors, 7 were found in 3 donors, 4 in 4 donors, 1 in 5 donors, and 1 in 6 donors. Several shared TCRβ sequences of the same specificity used the same V gene and a non-identical but similar CDR3 amino acid sequence, with just one or 2 amino acid exchanges. Such exchanges often appeared in the same position of the CDR3. For example, all RPH-specific shared TCRβ sequences used V gene TRBV4-3, had the same CDR3β length, and their CDR3 amino acid sequence followed the pattern ”CASSPXRNTEAFF”. In this project, a small set of up to 12 HLA-matched donors was tested per epitope, and only few donors were tested for epitopes that were not restricted by HLA-A*02:01, HLA-B*07:02, and HLA-C*07:02. Hence, TCRβ sequences with a lower extent of sharing may have been present in only one donor of this cohort and were therefore missed when searching for completely conserved sequences between donors.

4.9.3 Identification and clustering of shared and similar CMV peptide-specific TCRβ sequences

Based on the similarities found between 45 shared specific TCRβ sequences, the analysis of TCRβ sharing was extended to include non-identical but similar TCRβ sequences. Here, similar TCRβ sequences were defined as TCRβ sequences with (i) the same CMV peptide specificity, (ii) identical TRBV gene usage, (iii) identical CDR3 lengths, and (iv) a maximum of one divergent amino acid in the CDR3 amino acid sequence. Consequently, the 1809 amino-acid-unique TCRβ sequences were subjected to a second round of analysis, in which not only completely identical TCRβ sequences, but also similar TCRβ sequences were included. In total, 162 TCRβ sequences were identified that were similar to or identical with at least one other TCRβ sequence with the same epitope specificity in at least one other donor of the set (Table 4.8). Of the 162 shared TCRβ sequences, 20 were published with concordant epitope specificity and HLA restriction before this project started, 106 were published by us as epitope-specific in the frame of this project (Huth et al. 2019), and 36 have not been published yet. The

162 shared TCRβ sequences fell into 60 specificity clusters of similar TCRβ chains. The largest number of specific TCRβ clusters was found for HLA-C*07:02–restricted CMV epitopes CRV (21) and FRC (18), for which the largest number of HLA-matched donors were tested. The TCRβ clusters were of different size and comprised between 1 and 14 different CDR3β amino acid sequences.

Table 4.8: List of 162 shared and similar TCRβ sequences grouped into 60 specificity clusters. The table only lists donors in whom the TCRs were functionally identified as epitope-specific. The 'Reference' column shows the publication(s) in which a TCRβ sequence was first published with its precise epitope specificity and HLA restriction, blank spaces indicate unpublished TCRβ sequences. Data of this project were partially published in Huth et al. 2019.

V gene	CDR3 sequence	J gene	Cluster	Donors	Reference
TRBV25-1	CASSPGDEQFF	TRBJ2-1	CRV01	P01,P05,P06,P07	Huth et al. 2019
TRBV25-1	CASTPGDEQFF	TRBJ2-1	CRV01	P03,P06,P07,P14	Huth et al. 2019
TRBV25-1	CASSPGDEQYF	TRBJ2-7	CRV01	P01,P11	Huth et al. 2019
TRBV25-1	CASTLGDEQYF	TRBJ2-7	CRV01	P05	Huth et al. 2019
TRBV25-1	CASSAGDEQYF	TRBJ2-7	CRV01	P05	Huth et al. 2019
TRBV25-1	CASSFGDTQYF	TRBJ2-3	CRV01	P05	Huth et al. 2019
TRBV25-1	CASSLGDEQYF	TRBJ2-7	CRV01	P14	
TRBV25-1	CASSPGDTQYF	TRBJ2-3	CRV01	P06	Huth et al. 2019
TRBV25-1	CASTHGDEQFF	TRBJ2-1	CRV01	P05	Huth et al. 2019
TRBV25-1	CASTPGDEQYF	TRBJ2-7	CRV01	P06	Huth et al. 2019
TRBV25-1	CASTQGDEQFF	TRBJ2-1	CRV01	P06	Huth et al. 2019
TRBV25-1	CASTSGDEQFF	TRBJ2-1	CRV01	P07	Huth et al. 2019
TRBV25-1	CASTTGDEQFF	TRBJ2-1	CRV01	P03	Huth et al. 2019
TRBV25-1	CATSPGDEQYF	TRBJ2-7	CRV01	P01	Huth et al. 2019
TRBV6-1/6-5/6-6	CASSSGQKNTEAFF	TRBJ1-1	CRV02	P01,P07,P11	Huth et al. 2019
TRBV6-1/6-5/6-6	CASQPGQKNTEAFF	TRBJ1-1	CRV02	P08	Huth et al. 2019
TRBV6-1/6-5/6-6	CASSTGQKNTEAFF	TRBJ1-1	CRV02	P01	Huth et al. 2019
TRBV6-1/6-5/6-6	CASTPGQKNTEAFF	TRBJ1-1	CRV02	P04	Huth et al. 2019
TRBV6-1/6-5/6-6	CASTTGQKNTEAFF	TRBJ1-1	CRV02	P05	Huth et al. 2019
TRBV6-1/6-5/6-6	CASSPVGGGYTF	TRBJ1-2	CRV03	P02	Huth et al. 2019
TRBV6-1/6-5/6-6	CASSYVGGDYTF	TRBJ1-2	CRV03	P04	Huth et al. 2019
TRBV6-1/6-5/6-6	CASSYVGGGYTF	TRBJ1-2	CRV03	P08	Huth et al. 2019
TRBV6-1/6-5/6-6	CASSYVGSGYTF	TRBJ1-2	CRV03	P17	
TRBV7-9	CASSLMGLAGEETQYF	TRBJ2-5	CRV04	P06	Huth et al. 2019
TRBV7-9	CASSLSGLAGEETQYF	TRBJ2-5	CRV04	P05	Huth et al. 2019
TRBV7-9	CASSLSGLAGQETQYF	TRBJ2-5	CRV04	P07	Huth et al. 2019
TRBV7-9	CASSLVGLAGQETQYF	TRBJ2-5	CRV04	P05	Huth et al. 2019
TRBV28	CASSPISNEQFF	TRBJ2-1	CRV05	P01,P07	Huth et al. 2019
TRBV28	CASSPVSNEQFF	TRBJ2-1	CRV05	P01,P02	Huth et al. 2019
TRBV28	CASSPISNEQYF	TRBJ2-7	CRV05	P01	Huth et al. 2019
TRBV6-2/6-3	CASSAATPNTEAFF	TRBJ1-1	CRV06	P07	Huth et al. 2019
TRBV6-2/6-3	CASSQATPNTEAFF	TRBJ1-1	CRV06	P05	Huth et al. 2019
TRBV6-2/6-3	CASSTATPNTEAFF	TRBJ1-1	CRV06	P03	Huth et al. 2019
TRBV6-1/6-5/6-6	CASSGGQKNEKLFF	TRBJ1-4	CRV07	P05	Huth et al. 2019
TRBV6-1/6-5/6-6	CASSSGQKNEKLFF	TRBJ1-4	CRV07	P06	Huth et al. 2019
TRBV6-1/6-5/6-6	CASSTGQKNEKLFF	TRBJ1-4	CRV07	P02	Huth et al. 2019
TRBV7-9	CASSLRTDVYNEQFF	TRBJ2-1	CRV08	P05	Huth et al. 2019
TRBV7-9	CASSLRTETYNEQFF	TRBJ2-1	CRV08	P01	Huth et al. 2019
TRBV7-9	CASSLRTEVYNEQFF	TRBJ2-1	CRV08	P01	Huth et al. 2019
TRBV7-9	CASSLSGLDNEQFF	TRBJ2-1	CRV09	P03	Huth et al. 2019
TRBV7-9	CASSLSGLENEQFF	TRBJ2-1	CRV09	P05	Huth et al. 2019
TRBV7-9	CASSLSGLYNEQFF	TRBJ2-1	CRV09	P05	Huth et al. 2019
TRBV20-1	CSAPDWNNEQFF	TRBJ2-1	CRV10	P01,P02	Huth et al. 2019
TRBV20-1	CSAPDWGNEQFF	TRBJ2-1	CRV10	P08	Huth et al. 2019

Table 4.8: (continued)

V gene	CDR3 sequence	J gene	Cluster	Donors	Reference
TRBV11-3	CASSSWADTQYF	TRBJ2-3	CRV11	P04	Huth et al. 2019
TRBV11-3	CASSSWEDTQYF	TRBJ2-3	CRV11	P01	Huth et al. 2019
TRBV25-1	CASSAPRETQYF	TRBJ2-5	CRV12	P01	Huth et al. 2019
TRBV25-1	CASSEPRETQYF	TRBJ2-5	CRV12	P04	Huth et al. 2019
TRBV25-1	CASSEVLFEQFF	TRBJ2-1	CRV13	P06	Huth et al. 2019
TRBV25-1	CASSEVVFEQFF	TRBJ2-1	CRV13	P05	Huth et al. 2019
TRBV4-3	CASSQELAETYEQYF	TRBJ2-7	CRV14	P02	Huth et al. 2019
TRBV4-3	CASSQEWAETYEQYF	TRBJ2-7	CRV14	P08	Huth et al. 2019
TRBV6-1/6-5/6-6	CASSMGQKNQPQHF	TRBJ1-5	CRV15	P01	Huth et al. 2019
TRBV6-1/6-5/6-6	CASSPGQKNQPQHF	TRBJ1-5	CRV15	P11	
TRBV6-2/6-3	CASSYVAGGYEQYF	TRBJ2-7	CRV16	P06	Huth et al. 2019
TRBV6-2/6-3	CASSYVAGSYEQYF	TRBJ2-7	CRV16	P14	
TRBV7-9	CASSLLGTYSYNEQFF	TRBJ2-1	CRV17	P07	Huth et al. 2019
TRBV7-9	CASSLTGTYSYNEQFF	TRBJ2-1	CRV17	P05	Huth et al. 2019
TRBV28	CASSFPDTQYF	TRBJ2-3	CRV18	P01,P02	Huth et al. 2019
TRBV28	CASTPWGAEAFF	TRBJ1-1	CRV19	P04,P08	Huth et al. 2019
TRBV10-2	CASSESTGFDGYTF	TRBJ1-2	CRV20	P03,P18	Huth et al. 2019
TRBV6-1/6-5/6-6	CASSPGTPRDEQFF	TRBJ2-1	CRV21	P03,P05	Huth et al. 2019
TRBV15	CATSREGGETQYF	TRBJ2-5	FRC01	P05,P06,P08,P11,P14	Huth et al. 2019
TRBV15	CATSRTGGETQYF	TRBJ2-5	FRC01	P01,P03,P05	Huth et al. 2019
TRBV15	CATSRVAGETQYF	TRBJ2-5	FRC01	P06,P11	Huth et al. 2019
TRBV15	CATSRVGGETQYF	TRBJ2-5	FRC01	P11	
TRBV15	CATSAEGGETQYF	TRBJ2-5	FRC01	P08	Huth et al. 2019
TRBV15	CATSAFGGETQYF	TRBJ2-5	FRC01	P18	
TRBV15	CATSNIAGETQYF	TRBJ2-5	FRC01	P04	Huth et al. 2019
TRBV15	CATSNTGGETQYF	TRBJ2-5	FRC01	P14	
TRBV15	CATSQFGGETQYF	TRBJ2-5	FRC01	P11	
TRBV15	CATSQTGGETQYF	TRBJ2-5	FRC01	P14	
TRBV15	CATSRDAGETQYF	TRBJ2-5	FRC01	P06	Huth et al. 2019
TRBV15	CATSRDGGETQYF	TRBJ2-5	FRC01	P02	Huth et al. 2019
TRBV15	CATSRIAGETQYF	TRBJ2-5	FRC01	P17	
TRBV15	CATSRVAGEVQYF	TRBJ2-7	FRC01	P06	Huth et al. 2019
TRBV15	CATSSHAGETQYF	TRBJ2-5	FRC01	P07	Huth et al. 2019
TRBV15	CATSSHVGETQYF	TRBJ2-5	FRC01	P07	Huth et al. 2019
TRBV15	CATSSLAGETQYF	TRBJ2-5	FRC01	P18	
TRBV15	CATSSVAGETQYF	TRBJ2-5	FRC01	P11	
TRBV15	CATSVTGGETQYF	TRBJ2-5	FRC01	P02	Huth et al. 2019
TRBV6-2/6-3	CASSYGLEAFF	TRBJ1-1	FRC02	P03,P14,P17	
TRBV6-2/6-3	CASSGGLEAFF	TRBJ1-1	FRC02	P03,P07,P11	Huth et al. 2019
TRBV6-2/6-3	CASSLGLEAFF	TRBJ1-1	FRC02	P03,P17	Huth et al. 2019
TRBV6-2/6-3	CASSPGLEAFF	TRBJ1-1	FRC02	P14	
TRBV6-2/6-3	CASSYGLEAFF	TRBJ2-1	FRC02	P03	
TRBV15	CATSRVAGEELFF	TRBJ2-2	FRC03	P02	Huth et al. 2019
TRBV15	CATSRVAGEEQFF	TRBJ2-1	FRC03	P02	Huth et al. 2019
TRBV15	CATSRVAGEKLFF	TRBJ1-4	FRC03	P18	
TRBV15	CATSSSAGEEQFF	TRBJ2-1	FRC03	P08	Huth et al. 2019
TRBV15	CATSSVAGEEQFF	TRBJ2-1	FRC03	P02	Huth et al. 2019
TRBV19	CASSIGLEQFF	TRBJ2-1	FRC04	P11	
TRBV19	CASSLGLEQFF	TRBJ2-1	FRC04	P17	
TRBV19	CASSLGLEQYF	TRBJ2-7	FRC04	P17	
TRBV19	CASSYGLEQFF	TRBJ2-1	FRC04	P03	Huth et al. 2019
TRBV28	CASSLEQGAGETQYF	TRBJ2-5	FRC05	P08,P18	Huth et al. 2019
TRBV28	CASSFEQGAGETQYF	TRBJ2-5	FRC05	P11	
TRBV15	CATGPGQGPYEQYF	TRBJ2-7	FRC06	P17	
TRBV15	CATSPGQGPYEQYF	TRBJ2-7	FRC06	P07	Huth et al. 2019
TRBV15	CATSGTAGETQYF	TRBJ2-5	FRC07	P08	Huth et al. 2019
TRBV15	CATSHTAGETQYF	TRBJ2-5	FRC07	P11	
TRBV15	CATSRFGGEKLFF	TRBJ1-4	FRC08	P06	Huth et al. 2019
TRBV15	CATSRTGGEKLFF	TRBJ1-4	FRC08	P01	

Table 4.8: (continued)

V gene	CDR3 sequence	J gene	Cluster	Donors	Reference
TRBV28	CAIDTGDSPLHF	TRBJ1-6	FRC09	P18	
TRBV28	CAIDTGNSPLHF	TRBJ1-6	FRC09	P03	Huth et al. 2019
TRBV28	CASSLEALSSYNEQFF	TRBJ2-1	FRC10	P05	Huth et al. 2019
TRBV28	CASSLEFLSSYNEQFF	TRBJ2-1	FRC10	P18	
TRBV28	CASSLEVGGGLETQYF	TRBJ2-5	FRC11	P04	Huth et al. 2019
TRBV28	CASSLEVGGGQETQYF	TRBJ2-5	FRC11	P18	
TRBV4-3	CASSQDLGDTGELFF	TRBJ2-2	FRC12	P05	Huth et al. 2019
TRBV4-3	CASSQDLPDTGELFF	TRBJ2-2	FRC12	P14	
TRBV28	CASEDTGELFF	TRBJ2-2	FRC13	P05,P17	Huth et al. 2019
TRBV28	CASSLADLNQPQHF	TRBJ1-5	FRC14	P03,P17	
TRBV28	CASSLGDTNQPQHF	TRBJ1-5	FRC15	P01,P18	Huth et al. 2019
TRBV28	CASSPGFSGNTIYF	TRBJ1-3	FRC16	P11,P18	
TRBV28	CASSSLNSNQPQHF	TRBJ1-5	FRC17	P05,P17	Huth et al. 2019
TRBV28	CASSLEQLSGNTIYF	TRBJ1-3	FRC18	P04,P17	Huth et al. 2019
TRBV11-2	CASSLDPAGQGYEQYF	TRBJ2-7	IPS01	P19	Koning et al. 2013
TRBV11-2	CASSLDPTGQGYEQYF	TRBJ2-7	IPS01	P05	
TRBV12-3/12-4	CASSSVNEAFF	TRBJ1-1	NLV01	P03,P08,P21	Schwanninger et al. 2008
TRBV12-3/12-4	CASSIVNEAFF	TRBJ1-1	NLV01	P21	Dash et al. 2017
TRBV12-3/12-4	CASSSVTEAFF	TRBJ1-1	NLV01	P17	Venturi et al. 2008
TRBV7-6	CASSLAPGATNEKLFF	TRBJ1-4	NLV02	P03,P07,P21	Venturi et al. 2008
TRBV7-6	CASSLAPGSTNEKLFF	TRBJ1-4	NLV02	P21	Venturi et al. 2008
TRBV6-1/6-5/6-6	CASSPTTGTGAYGYTF	TRBJ1-2	NLV03	P03	
TRBV6-1/6-5/6-6	CASSPTTGTGNYGYTF	TRBJ1-2	NLV03	P21	
TRBV12-3/12-4	CASSSANYGYTF	TRBJ1-2	NLV04	P08,P17	Venturi et al. 2008
TRBV27	CASSLEGYTEAFF	TRBJ1-1	NLV05	P03,P21	Venturi et al. 2008
TRBV9	CASSVGTGSNTEAFF	TRBJ1-1	RIK01	P19,P20	
TRBV4-3	CASSPQRNTEAFF	TRBJ1-1	RPH01	P03,P04,P05,P06,P08,P11	Brennan et al. 2012
TRBV4-3	CASSPARNTEAFF	TRBJ1-1	RPH01	P03,P05,P08,P11	Weekes et al. 1999
TRBV4-3	CASSPSRNTEAFF	TRBJ1-1	RPH01	P03,P08	Weekes et al. 1999
TRBV4-3	CASSPHRNTEAFF	TRBJ1-1	RPH01	P03,P05	Weekes et al. 1999
TRBV4-3	CASSPNRNTEAFF	TRBJ1-1	RPH01	P03,P08	Brennan et al. 2012
TRBV4-3	CASSPGRNTEAFF	TRBJ1-1	RPH01	P03	Huth et al. 2019
TRBV4-3	CASSPTRNTEAFF	TRBJ1-1	RPH01	P08	Brennan et al. 2012
TRBV7-8	CASSFRTVSSYEQYF	TRBJ2-7	TPR01	P01,P03,P04,P11	Huth et al. 2019
TRBV7-8	CASSFRTVNSYEQYF	TRBJ2-7	TPR01	P02,P03	Huth et al. 2019
TRBV7-8	CASSLRTVSSYEQYF	TRBJ2-7	TPR01	P02,P04	Huth et al. 2019
TRBV27	CASSLGASPGELFF	TRBJ2-2	TPR02	P02	Huth et al. 2019
TRBV27	CASSLGGSPGELFF	TRBJ2-2	TPR02	P06	Huth et al. 2019
TRBV27	CASSLGSAPGELFF	TRBJ2-2	TPR02	P06	Huth et al. 2019
TRBV7-9	CASSLHDRGFRTEAFF	TRBJ1-1	TPR03	P02	Huth et al. 2019
TRBV7-9	CASSLHDRGSRTEAFF	TRBJ1-1	TPR03	P01	Huth et al. 2019
TRBV7-9	CASSLHDRGVRTEAFF	TRBJ1-1	TPR03	P02	Huth et al. 2019
TRBV7-9	CASSLIGVSSYNEQFF	TRBJ2-1	TPR04	P01,P02,P06	Miconnet et al. 2011
TRBV7-9	CASSLKGVSSYNEQFF	TRBJ2-1	TPR04	P06	Huth et al. 2019
TRBV7-9	CASSFRQGVNTGELFF	TRBJ2-2	TPR05	P01,P02	Huth et al. 2019
TRBV7-9	CASSFRQGSNTGELFF	TRBJ2-2	TPR05	P01	Huth et al. 2019
TRBV7-9	CASSLALGVAKNIQYF	TRBJ2-4	TPR06	P01	Koning et al. 2013
TRBV7-9	CASSLATGVAKNIQYF	TRBJ2-4	TPR06	P06	Huth et al. 2019
TRBV7-9	CASSSHDLTGMNTEAFF	TRBJ1-1	TPR07	P02	Huth et al. 2019
TRBV7-9	CASSSHDWTGMNTEAFF	TRBJ1-1	TPR07	P07	Huth et al. 2019
TRBV6-2/6-3	CASSYSSGELFF	TRBJ2-2	TPR08	P01,P08	Janbazian et al. 2012
TRBV27	CASSLSPSTGNYGYTF	TRBJ1-2	TPR09	P03,P06	Miconnet et al. 2011
TRBV7-2	CASSSRGTVNTEAFF	TRBJ1-1	TPR10	P03,P05	Huth et al. 2019
TRBV7-9	CASSLHTQGARTEAFF	TRBJ1-1	TPR11	P02,P07	Huth et al. 2019
TRBV11-2	CASSLNRGRTDTQYF	TRBJ2-3	VLE01	P03	
TRBV11-2	CASSLQRGRTDTQYF	TRBJ2-3	VLE01	P02	Koning et al. 2013

Table 4.8: (continued)

V gene	CDR3 sequence	J gene	Cluster	Donors	Reference
TRBV9	CASSAGQGVTYEQYF	TRBJ2-7	VTE01	P19,P21	
TRBV9	CASSSGQGVTYEQYF	TRBJ2-7	VTE01	P19	
TRBV9	CASSVGQGVTYEQYF	TRBJ2-7	VTE01	P21	Glanville et al. 2017

In most clusters, all TCRβ sequences used the same TRBJ gene. However, in clusters CRV01, CRV05, FRC02, FRC03, and FRC04, 2–3 different TRBJ genes were used. The most frequently interchanged TRBJ genes were TRBJ2-1 and TRBJ2-7, which led to a conservative amino acid exchange (F↔Y) in the penultimate CDR3β position. There were also TRBJ gene usage variations within specificity clusters that led to non-conservative amino acid exchanges, such as TRBJ1-4 and TRBJ2-1 (K↔E and L↔Q) in cluster FRC03. Specific TCRβ sequences in cluster FRC02 used either TRBJ1-1 or TRBJ2-1, but this differential Jβ gene usage did not alter the CDR3β sequence. These findings indicate that while the J gene segment is important for encoding parts of the CDR3β and for TCRα chain pairing, variations in J gene usage are tolerated as long as these variations do not lead to severe changes in the CDR3β amino acid sequence. TRBV gene usage of specificity clusters was similar to the overall TRBV gene usage of TCRβ sequences with the corresponding specificity (Section 4.9.1, Figure 4.28). For instance, the largest TCRβ cluster used TRBV25-1, which was also among the top 3 TRBV genes used in all 521 donor-nucleotide–unique CRV-specific TCRβ sequences. Likewise, many CRV clusters use TRBV6-1/6-5/6-6, TRBV7-9, or TRBV28, which are overrepresented V genes in CRV-specific TCRβ sequences.

A large extent of sharing between donors was observed for some TCRβ clusters. For example, in all 12 tested donors expressing HLA-C*07:02, at least one TCRβ sequence of cluster FRC01 was specifically enriched by single peptide or peptide pool stimulation. Likewise, one or more TCRβ sequences from cluster RPH01 were specifically enriched in 5 of 7 donors by single peptide stimulation. Donors P01 and P07 did not express any TCRβ sequence from cluster RPH01, but these donors have a TRBV4-3 deletion, as suggested by a lack of TRBV4-3 in their TCRβ repertoires. They can therefore not recombine TCRβ chains from cluster RPH01, which uses this particular Vβ gene. Other frequently shared TCRβ clusters were IPS01 (2/2 donors), NLV01 (4/6 donors), and VTE01 (2/3 donors). TCRβ sequences from the most common cluster specific for TPR, cluster TPR01, were found specifically enriched in 5 of 9 donors, and

the remaining clusters were shared to a lesser extent. Many donors expressed not only one, but multiple different epitope-specific TCRβ sequences of a given cluster. Donor P11, for instance, expressed 5 distinct TCRβ sequences of cluster FRC01 that were specifically enriched upon stimulation with FRC-containing peptide pools. Donors P05 and P06 each expressed 5 specific TCRβ sequences from cluster CRV01, and donor P03 even expressed 6 of 7 specific TCRβ sequences from cluster RPH01. All in all, the results show that there is extensive sharing of identical or highly similar specific TCRβ sequences.

The relationship architecture between TCRβ sequences in the 60 specificity clusters was vastly different (Figure 4.29). Some TCRβ clusters, like RPH01 or FRC02, were strongly interconnected and each TCRβ sequences in these clusters differed in maximum one amino acid from any other TCRβ sequence of this cluster. Other clusters, like CRV02 or FRC03, were rather linear with single connections between each neighboring TCRβ sequence and the TCRβ sequences at both ends differed in 2–4 amino acids from each other. Some clusters, i.e. NLV01 or TPR01, contained one frequently shared TCRβ sequence and two lesser shared variants each. Clusters CRV01 and FRC01 were most complex in their relationship architecture. In both clusters, there were 3 identical TCRβ sequences that were specifically enriched in multiple donors, and both clusters can be divided into two major subclusters. CDR3β sequences on the left hand side of cluster FRC01 had amino acids ”AG” on positions 7–8 and often an aliphatic amino acid (V, I, or L) on position 6. They clustered around the identically shared CDR3β sequence ”CATSRVAGETQYF”. CDR3β sequences on the right hand side of cluster FRC01 had the central motif ”GG” instead of ”AG”, which was generally accompanied by a polar or acidic amino acid (T or E) in position 6 and clustered around the to major CDR3β sequences of the cluster, which were ”CATSREGGETQYF” and ”CATSRTGGETQYF”. Similarly, CDR3β sequences in cluster CRV01 could be subdivided in TCRβ chains using TRBJ2-7 (right side) and TRBJ2-1 or TRBJ2-3 (left side). Some clusters with the same specificity contained TCRβ sequences that were quite similar across clusters, suggesting that there may be a missing link connecting these clusters that may be identified when testing more donors with our assays. An example for such sequence-related specificity clusters are CRV02 and CRV07, which used the same Vβ gene, but a different Jβ gene that substantially changed the last part of the

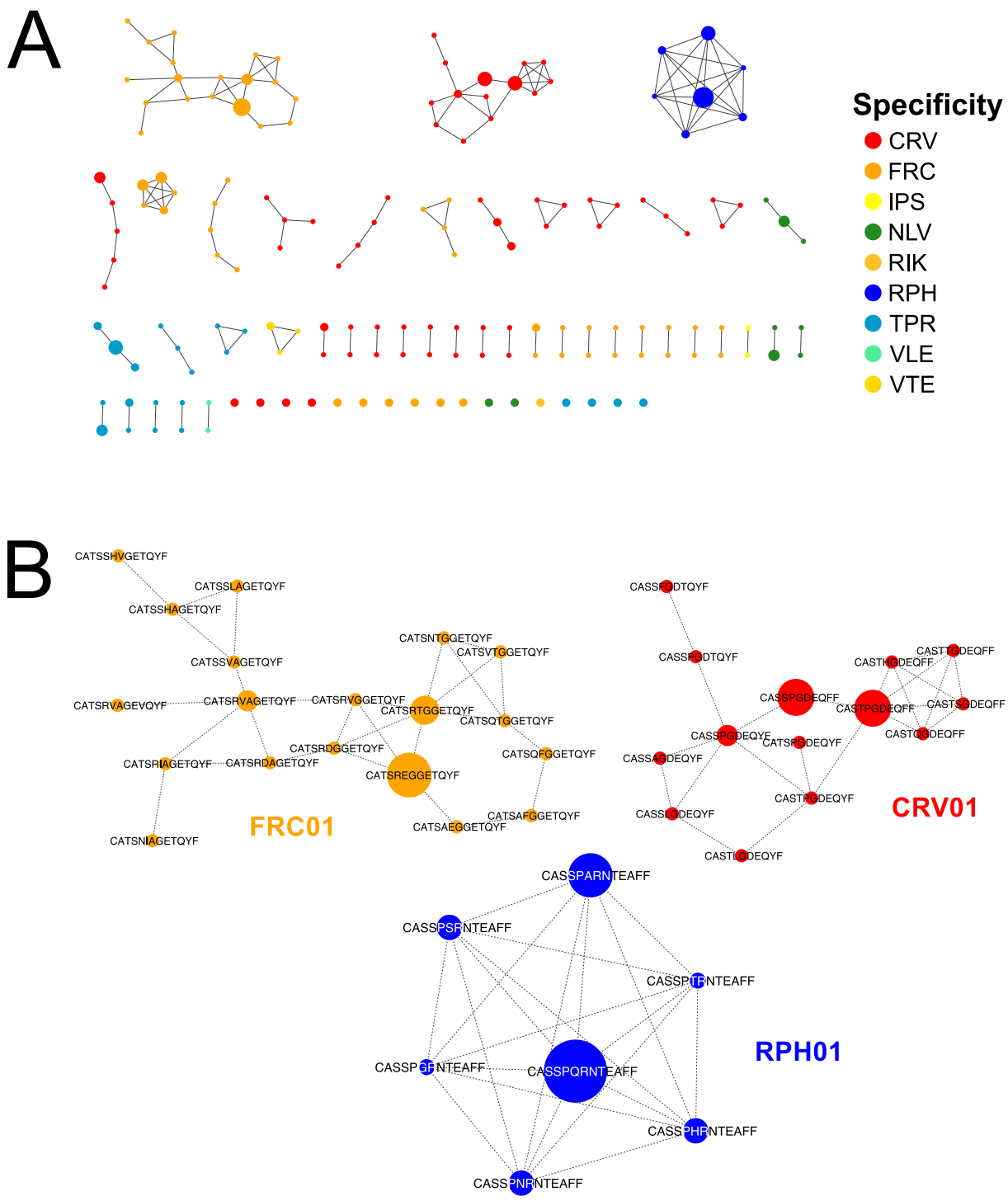


Figure 4.29: Sequence relationship architecture within 60 TCRβ specificity clusters. (**A**) Overview of all 60 specificity clusters identified in this thesis. The clusters are sorted by cluster size and then specificity. Each node represents one amino-acid-unique TCRβ sequence and the node sizes correspond to the number of donors in which a TCRβ sequence was specifically enriched (smallest = 1 donor, largest = 6 donors). The edges (black lines) connect CDR3β sequences that differ by one amino acid from each other. (**B**) Close-up of the 3 largest specificity clusters. Cluster nodes are labelled with the CDR3β amino acid sequences.

CDR3β. The cluster analysis of the similar or identical shared TCRβ sequences shows that while there were some CDR3βs that seem to be preferentially selected, be it because of superior function or increased recombination probability, other less shared sequence-related CDR3βs can also recognise the presented epitope and induce T-cell proliferation. Depending on the binding mode of a TCR to the peptide-MHC complex, the CDR3β sequences may be more restricted or more variable. For instance, amino acid exchanges in multiple CDR3β positions were tolerated in TCRβ sequences of cluster FRC01, but only CDR3β sequences with one amino acid exchange in position 5 were FRC-specific in cluster FRC02 (”CASSXGLEAFF”). Looking at positions and properties of tolerated amino acid exchanges in the CDR3β in more detail will shed more light on factors determining the specificity of TCRβ chains. At present, it is not possible to speculate in greater detail on the role of individual amino acids in HLA-C*07:02-restricted TCRs, since no X-ray structures of such TCRs in complex with their cognate epitope have been published yet.

The results presented in the preceding paragraphs demonstrate that TCRβ sequences of the same specificity are often distinct, but some have shared elements, such as same V gene usage, CDR3 length and similar CDR3 amino acid sequence. Recently, two research groups made an effort to try and identify common CDR3β sequence motifs rather than searching for conserved full-length CDR3β sequences, with the aim to predict TCRβ chain specificity (Dash et al. 2017; Glanville et al. 2017). In this project, only 162 of 1809 amino-acid-unique specific TCRβ sequences (9%) were similar or identical to one or more TCRβ sequences with the same epitope specificity (shared), which leaves 1647 specific TCRβ sequences to be donor-exclusive (private). Finding common motifs among TCRβ sequences of the same specificity could help to categorise not only shared, but also private specific TCRβ sequences according to their specificity. However, when comparing the CDR3β amino acid sequences of the largest 1–2 specificity clusters found here for epitopes CRV, FRC, NLV, RPH, and TPR, conserved sequence stretches that were not TRBV- or TRBJ-encoded were rare and often very short (Figure 4.30) comprising only 2–3 amino acids. Many specificity clusters, for example CRV01, FRC01, or RPH01, were highly variable in the central amino acid positions 5 and/or 6 of their CDR3β sequence and had conservative amino acid exchanges in more peripheral positions. Frequently observed amino acid

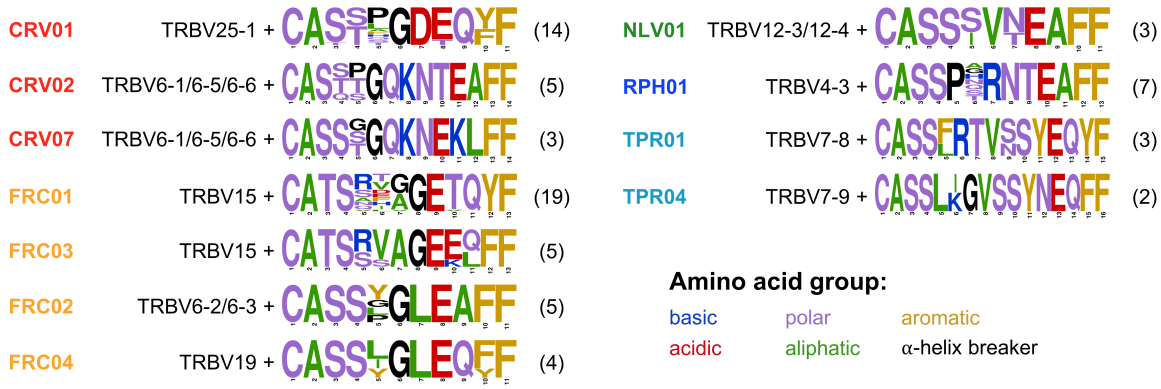


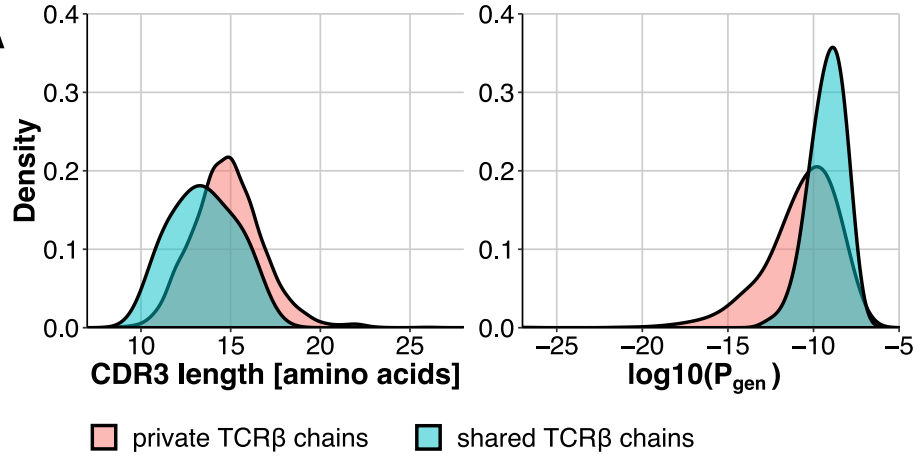
Figure 4.30: Amino acid logos of CDR3βs of the largest specificity clusters found in this project and similar clusters (CRV07, FRC03, FRC04, and TPR04). Amino acids are coloured based on their chemical properties. Names of the specificity clusters and the used Vβ genes are noted to the left of each logo. The number of TCRβ sequences in each cluster is shown in parentheses to the right. Letter sizes represent the proportion of TCRβ sequences in a cluster with the corresponding amino acid in a particular position.

exchanges were T↔S↔N and I↔V↔L. Despite of the overall diversity found in central CDR3β amino acid positions, some TCRβ specificity clusters were very similar to each other suggesting the presence of common motifs among TCRβ sequences with a particular specificity. For example, clusters CRV02 and CRV07 mainly differed in their CDR3β amino acid sequence due to the diverging TRBJ gene usage, but both had the central amino acid motif ”GQKN”. Likewise, clusters FRC01 and FRC03 often used motif ”VAGE”, and clusters TPR01 and TPR04 used motif ”VSSY”. When screening the 1809 amino acid-unique CMV-specific TCRβ sequences for CRV-motif ”GQKN”, 4 additional TCRβ sequences were using this motif were found, and all 4 were CRV-specific. However, the sole additional hit when searching for FRC-motif ”VAGE” was a CRV-specific TCR. Moreover, both hits when searching for TPR-motif ”VSSY” were not TPR-specific, but responded to NLV or VTE. This brief and random motif search implies that short CDR3 motifs of 3–4 amino acid length are not sufficient for determination of the specificity of a TCRβ sequence, since TCRβ sequences with different CMV epitope specificities made use of the same amino acid sequence motif stretches. The specificity-defining features of TCRβ sequences are likely more complex than short CDR3β amino acid motifs. Rather, the data presented here suggest that Vβ gene usage, CDR3β length, and a conserved CDR3β sequence, in which only few amino acid exchanges in defined positions are tolerated, determine the epitope specificity of a TCRβ chain.

4.9.4 Shared TCRβ sequences are shorter and more likely to be recombined

An interesting question raised by the epitope-specific TCRβ repertoires identified in this project is why some TCRβ sequences were specifically enriched in multiple donors (shared TCRβ sequences), while others were donor-exclusive (private). Is it because shared TCRβ sequences bind their cognate peptide-HLA complex with higher binding affinity than private TCRβ sequences? Are shared TCRβ sequences functionally superior to private TCRβ sequences? Or is it because the likelihood of somatic recombination of shared TCRβ sequences is higher? To address this question, the CDR3β amino acid sequence lengths and computed generation probabilities P_{gen} of the 162 shared and 1647 private amino acid-unique specific TCRβ sequences were compared (Figure 4.31). The CDR3βs of the shared TCRβs, starting with the canonical C and

A



B

	CDR3 length		P_{gen}	
	shared	private	shared	private*
Minimum	11	9	3.36×10^{-13}	1.15×10^{-27}
Maximum	17	26	2.73×10^{-8}	1.90×10^{-7}
Median	13	15	7.75×10^{-10}	4.51×10^{-11}
Mean	13.51	14.84	$6.19\times10^{-10\dagger}$	$1.98\times10^{-11\dagger}$

* 7 private TCRβ sequences were omitted here, because their P_{gen} was 0.
† Geometric mean

Figure 4.31: Comparison of CDR3β amino acid sequence lengths and computed generation probabilities P_{gen} of 162 shared and 1647 private TCRβ sequences. P_{gen} values were computed for each Vβ-CDR3β-Jβ combination on the amino acid level using the python-based algorithm OLGA (Sethna et al. 2019). (**A**) Density plots of the CDR3 amino acid sequence length distributions (left) and computed P_{gen} values (right) of shared and private TCRβ chains. (**B**) Hallmark values of CDR3 length and P_{gen} value distributions.

ending on the canonical F or, in case of TRBJ2-7*02, on V, were between 11–17 amino acids long, while the CDR3βs of the private TCRβs were 9–26 amino acids long. The median CDR3β lengths of shared and private TCRβs were 13 and 15 amino acids, respectively. On average, CDR3β lengths of shared TCRβs were shorter than those of private TCRβs, with mean lengths of 13.51 (shared TCRβs) and 14.84 (private TCRβs). The density plot curve of the CDR3β length distribution of shared TCRβs was shifted to shorter lengths compared to the curve of private TCRβs.

Generation probabilities of shared and private TCRβ sequences were calculated on the amino acid level with the python-based algorithm OLGA (Sethna et al. 2019). The P_{gen} values returned by OLGA reflect the likelihood of a CDR3β amino acid sequences to be recombined using the exact Vβ and Jβ genes with which it was identified here. P_{gen} values of shared TCRβ sequences ranged from 3.36×10^{-13} to 2.73×10^{-8}, with a geometric mean of 6.19×10^{-10} and a median P_{gen} of $\times10^{-10}$. Generation probabilities of private TCRβ sequences were more variable (1.15×10^{-27} to 1.90×10^{-7}), and the median and mean P_{gen} values were lower than those of shared TCRβ sequences (median(P_{gen})=4.51×10^{-11}; geometric mean(P_{gen})=1.98×10^{-11}). This also shows in the P_{gen} density plots, in which the curve of shared TCRβ sequences is shifted to higher P_{gen} values compared to the curve of private TCRβ sequences. In conclusion, shared TCRβ sequences tend to be shorter than private TCRβ sequences and have a higher likelihood to be generated by somatic recombination.

4.10 A CMV-specific TCRβ clonotype signature

The sequencing data produced and analysed in this project revealed that CMV epitope-specific TCRβ sequences are highly frequent in the ex vivo repertoire of the CMV-positive donors from which they were identified by single peptide stimulation (Section 4.4.9, Figure 4.13). However, most of the specific TCRβ sequences identified here were private and shared TCRβ clonotypes were often not the most frequent clonotype of a given specificity. Nontheless, a considerable number of shared TCRβ sequences were identified in this project. It was therefore tested whether these shared TCRβ sequences were sufficient to identify and track CMV epitope-specific T-cell responses in additional donors without testing each of them individually by CMV peptide stimulation. To do so, the cumulative read frequencies of the 162 shared and similar signature

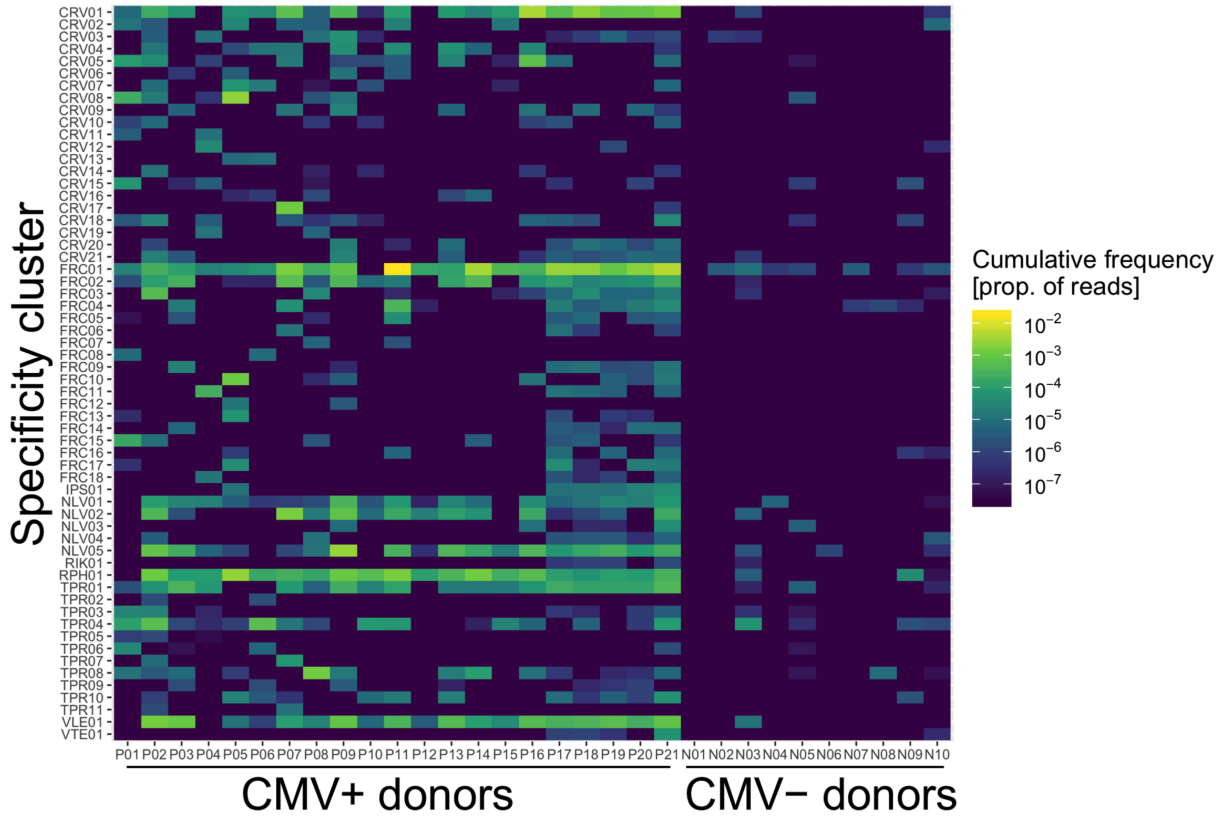


Figure 4.32: Frequency of the 162 shared signature TCRβ sequences in peripheral blood T-cell repertoires of 21 CMV-positive donors (P01–P21) and 10 CMV-negative donors (N01–N10). Cumulative proportions of signature TCRβ reads for each of the 60 individual specificity clusters are visualised as a heatmap. An overview of TCRβ sequences in each specificity cluster can be found in Table 4.8.

TCRβ sequences were calculated in the ex vivo TCRβ repertoires of donors P01–P21 and N01–N10 (this cohort), as well as in an additional cohort of 575 high-resolution HLA-typed donors with known CMV status published by Emerson et al. (Emerson et al. 2017) (designated ”cohort 1” in their study).

First, the cumulative TCRβ read frequencies of each specificity cluster (see Table 4.8) in the peripheral blood T-cell repertoires of donors P01–P21 and N01–N10 of this cohort were computed (Figure 4.32). Signature TCRβ sequences were highly frequent in the ex vivo repertoires of CMV-positive donors, but rare in CMV-negative donors. Cumulative TCRβ read frequencies of single specificity clusters in CMV-positive donors often exceeded 0.001, and the maximum cumulative frequency of any cluster in any donor was 0.0178 (FRC01 in P11). By contrast, sequences of the signature TCRβ set were seldom found in CMV-negative donors, and when they were found, the cumulative read frequencies did not exceed 6.3×10^{-5} (TPR04 in N03). Most CMV-positive donors had signature TCRβ sequences from multiple specificity clusters of the same epitope

specificity in their peripheral blood T cell repertoires, which highlights the relevance of signature TCRβ sequences with a lower extent of sharing. Donors P17–P21 did not express HLA-B*07:02 and donors P19–P21 did neither express HLA-C*07:02, which is why these donors were not expected to have TCRβ sequences against RPH and TPR, and CRV and FRC, respectively, in their ex vivo repertoires. Importantly, these donors exhibited a similar pattern of cumulative frequencies of signature TCRβ sequences on the heatmap and the same underlying signature TCRβ sequences were found at comparable frequencies. This was in contrast to all other donors who showed distinct peripheral blood TCRβ frequency patterns. Samples from donors P17–P21 were prepared and sequenced together, and it was sometimes observed that there was a spillover of dominant, highly frequent TCRβ sequences from peptide-stimulated samples to other samples sequenced on the same lane. For example, samples from donor N03 were sequenced together with samples from donor P06, whose most frequent TCRβ sequence specific for TPR was ”TRBV7-9–CASSLIGVSSYNEQFF–TRBJ2-1”. This TCRβ sequence held 37.2% of reads of all TCRβ reads in the TPR-stimulated sample of donor P06 and was found at a low read frequency (6.3×10^{-5}) with exactly the same nucleotide sequence in donor N03. It is likely that the presence of this TCRβ sequence in the ex vivo sample of donor N03 is due to a contamination of the sample with material from donor P06 and is not actually found in the peripheral blood of donor N03. Similarly, it is possible that HLA-mismatched signature TCRβ sequences were found at such a high frequency in donors P17–P21 because of sample contamination or barcode spillover rather than because they were actually present in the ex vivo repertoires of these donors. In addition to the signature TCRβ sequences found in mismatched donors P17–P21, several NLV- and VLE-specific signature TCRβ sequences, albeit more frequent in the HLA-A*02:01–expressing donors, were also found in HLA-mismatched CMV-positive donors P11 and P14. While TCRβ sequences of some specificity clusters were abundant in CMV-positive donors even in the absence of the required HLA allele, the TCRβ sequence from cluster RIK01, shared between donors P19 and P20, was not found in peripheral blood of 7 of 8 HLA-A*03:01–positive CMV carriers (absent in P01, P06, P08, P13, P14, P20). It was expected that not all donors had T cells against RIK in their CMV–specific repertoire, since another group reported that 2/4 HLA-matched donors responded to the peptide (Braendstrup et al.

2014). However, the proportion of donors with RIK-specific signature TCRβ sequences in their circulation was 4-fold lower in this project. This suggests that relevant RIK-specific signature TCRβ sequences, either belonging to cluster RIK01 or a different cluster, have not been identified yet. The shared RIK-specific TCRβ sequence was found to be specific in donor P20, but was below detection limit in the PBMC sample of that donor. Taken together, the data obtained in this project show that shared CMV epitope-specific TCRβ sequences are more frequent in the peripheral blood T-cell repertoires of CMV-positive donors than CMV-negative donors, although background signal in CMV-negative and HLA-mismatched donors was sometimes observed.

Next, the cumulative read frequencies of signature TCRβ sequences against the 6 best studied epitopes of this project, that is epitopes NLV and VLE (HLA-A*02:01–restricted), RPH and TPR (HLA-B*07:02–restricted), and CRV and FRC (HLA-C*07:02–restricted), were calculated in the donor cohort sequenced in this project (P01–P21, N01–N10) and 575 high-resolution HLA-typed donors with known CMV serostatus from cohort 1 (Emerson et al. 2017). High-resolution HLA-typing data for cohort 1 was only recently made available to the public (DeWitt et al. 2018).

Firstly, the impact of HLA matches on cumulative signature TCRβ read frequencies across donors was investigated. For this purpose, donors were grouped by the number of matching HLA loci (HLA-A*02:01, HLA-B*07:02, HLA-C*07:02). Donors expressing neither of these HLA alleles formed the group ”0 matched HLA loci”, while donors expressing all 3 HLA alleles were grouped to ”3 matched HLA loci”. Homozygosity in any of the 3 HLA loci was treated in the same way as heterozygosity. Figure 4.33A shows the cumulative read frequencies of 156 signature TCRβ sequences against the 6 epitopes restricted by the relevant HLA-alleles in peripheral blood of 31 donors of this cohort and 575 donors of cohort 1. The cumulative frequency of the 156 signature TCRβ sequences increased with increasing number of matched HLA loci in CMV-positive donors, while it remained constantly low in CMV-negative donors (median cumulative frequency approximately 10^{-5}). This shows that the ex vivo frequency of CMV epitope-specific signature TCRβ sequences is strongly dependent on the donor's HLA type and CMV serostatus. An exception were the high ex vivo signature TCRβ frequencies (ca. 3×10^{-3}) in donors P19 and P20 of this cohort, who had 0 matching HLA loci. These high cumulative frequencies may be a result of TCRβ library cross-contamination.

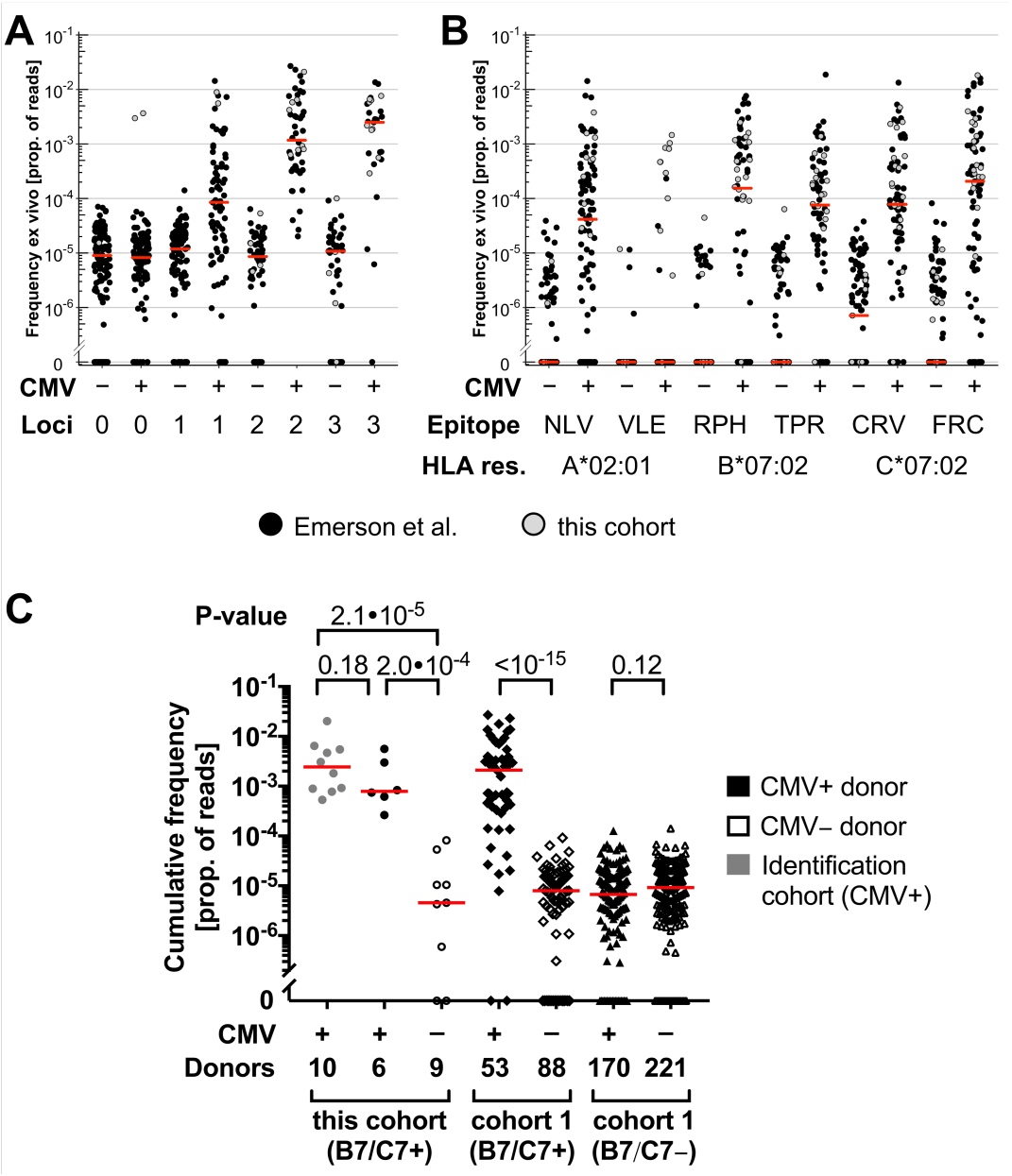


Figure 4.33: Frequency of CMV epitope-specific signature TCRβ sequences in peripheral T-cell repertoires of CMV-positive and CMV-negative donors from this cohort and 575 high-resolution HLA-typed donors with known CMV serostatus from cohort 1 published by Emerson et al. (Emerson et al. 2017) and DeWitt et al. (DeWitt et al. 2018). (**A**) Cumulative frequency of 156 TCRβ sequences restricted by HLA-A*02:01, HLA-B*07:02, or HLA-C*07:02 in ex vivo TCRβ repertoires of donors with 0, 1, 2, or 3 matching HLA loci, whereby homozygosity in an HLA allele is considered as 1 match. (**B**) Cumulative frequency of 156 TCRβ sequences in HLA-matched donors grouped by epitope specificity. HLA res.=HLA restriction (**A+B**) Grey dots belong to donors from this cohort and black dots to donors from cohort 1. Red lines show the median cumulative read frequencies. (**C**) Cumulative frequency of 145 TCRβ sequences restricted by HLA-B*07:02 or HLA-C*07:02 in peripheral blood of fully HLA-matched (B7/C7+) or double-negative donors (B7/C7–) from this cohort or cohort 1. Gray circles identify donors P01–P08, P11, and P14 from whose repertoires the TCRβ sequences were originally derived. Red lines show the median cumulative read frequencies. *P* values were calculated with a two-tailed Mann–Whitney U test.

Secondly, the cumulative read frequency for each of the 6 epitopes individually was computed exclusively in HLA-matched donors to see whether there was a difference in signature TCRβ read frequencies between epitopes (Figure 4.33B). The median cumulative read frequencies for all epitopes but CRV were 0 in CMV-negative donors, and for CRV it was well below 10^{-6}. Furthermore, median cumulative read frequencies of signature TCRβ sequences for each epitope, except for VLE, were at least 100-fold higher in CMV-positive than CMV-negative donors. Median cumulative read frequencies of the 4 best-studied epitopes (CRV, FRC, RPH, and TPR) with the highest number of signature TCRβ sequences in CMV-positive donors were in a similar range and varied between 7.6×10^{-5} and 2.1×10^{-4}. NLV-specific signature TCRβ sequences were slightly less frequent (median=4.2×10^{-5}) and the median cumulative frequency of VLE-specific signature TCRβ sequences was 0. However, fewer samples were analysed for these two specificities and, consequently, only 9 and 2 signature TCRβ sequences were found for these epitopes. Analysing more samples stimulated with epitopes NLV and VLE will likely increase the number of shared or similar specific TCRβ sequences and thereby the cumulative read frequencies in HLA-A*02:01–expressing CMV carriers. Importantly, it was expected that VLE-specific T cells were not found in all donors, since two studies by Khan et al. (Khan et al. 2002a; Khan et al. 2004) detected VLE-specific T-cell responses in approximately half of their tested CMV-positive donors. In conclusion, cumulative signature TCRβ read frequencies showed great potential to discriminate between CMV-positive and CMV-negative donors for all epitopes, although more signature TCRβ sequences need to be identified for HLA-A*02:01–restricted epitopes VLE and NLV.

Lastly, the four most intensively studied epitopes CRV, FRC, RPH, and TPR, all restricted by the HLA-B*07:02/HLA-C*07:02 haplotype, were put into focus of the analysis. Overall, 145 signature TCRβ sequences grouped to 53 specificity clusters were found for these epitopes. Cumulative frequencies of the 145 HLA-B*07:02/HLA-C *07:02–restricted signature TCRβ sequences were computed in peripheral T cell repertoires of 25 donors sequences in this project, all expressing the relevant HLA haplotype, and all donors from cohort 1 either expressing both (B7/C7+) or neither (B7/C7–) of the two HLA alleles of interest (Figure 4.33C). There was no significant difference in cumulative read frequencies in donors P01–P08, P11, and P14, from whom

the epitope-specific TCRβ sequences were identified (identification subcohort), and an additional cohort of 6 CMV-positive, HLA-matched donors (validation subcohort), confirming that there was no obvious bias favouring donors of the identification cohort. Median cumulative frequencies of the 145 signature TCRβ reads were 532-fold and 172-fold higher in the identification subcohort (median=2.24×10^{-3}) and the validation subcohort (median=7.9×10^{-4}) than in the CMV-negative donors of this cohort (median=4.6×10^{-6}). Accordingly, median cumulative frequencies of signature TCRβ sequences in B7/C7+ subcohort of cohort 1 were 263-fold higher in CMV-positive donors (median=2.1×10^{-3}) than in CMV-negative donors (median=8.0×10^{-6}). In contrast, the median cumulative frequency of signature TCRβ sequences in CMV-positive donors of the B7/C7– subcohort of cohort 1 was not different from that of HLA-mismatched CMV-negative donors, or from CMV-negative donors co-expressing HLA-B*07:02/HLA-C*07:02. A deliberately chosen cumulative frequency cutoff of 10^{-4} separates CMV-positive from CMV-negative donors with high precision in both HLA-matched cohorts. All 25 donors of this cohort were assigned their correct CMV serostatus with a cumulative read frequency cutoff of 10^{-4} (F1 score = 1). Similarly, 45 of 53 CMV-positive were correctly discriminated from 88 of 88 CMV-negative donors of the B7/C7+ subcohort of cohort 1 with regard to their CMV seropositivity (F1 score = 0.91). Taken together, the analysis of cumulative read frequencies demonstrates that the shared CMV epitope-specific signature TCRβ sequences identified in this project are strongly indicative of the CMV serostatus and HLA type of healthy donors.

4.11 CRV-specific T cells also expand with the strain variant peptide CRV-I

Protein sequences between different CMV strains are not perfectly conserved, and some of the peptides tested in this thesis differ in their sequence by 1 amino acid between different strains. One example for such an epitope polymorphism is found in the HLA-C*07:02–restricted epitope CRV derived from antigen IE1. The variant used in the single peptide stimulation assays is here abbreviated ”CRV” and the full length sequence is ”CRVLCCY**V**L”. This variant is present in many well-studied CMV strains, such as AD169 (Chee et al. 1990; Bradley et al. 2009), Towne (Bradley et al. 2009; Cui et al. 2012), and Merlin (Dolan et al. 2004). The second epitope variant, which is here

Table 4.9: Number of TCRβ clonotypes identified as specific for epitope variant CRV only, variant CRV-I only, or both variants in donors P01 and P04–P08.

Donor	CRV	CRV-I	both
P01	52	10	67
P04	8	10	37
P05	55	8	63
P06	16	2	33
P07	9	21	24
P08	10	2	15

abbreviated "CRV-I", has a conservative V→I amino acid exchange in position 8 and the full-length sequence is "CRVLCCY**I**L". The CRV-I variant is, for example, found in CMV strains TB40/E (Sinzger et al. 2008) and Toledo (Murphy et al. 2003). To test whether T cells with TCRβ clonotypes specific for variant CRV also respond to variant CRV-I, 6 CMV-positive donors P01 and P04–P08 were single-peptide–stimulated with CRV-I. CRV-I–specific TCRβ sequences were identified as described in Section 3.5.3.4 and Figure 3.5 using autologous TPR-stimulated samples and ex vivo samples as controls. Subsequently, CRV-specific TCRβ sequences were compared to CRV-I–specific TCRβ sequences in each donor (Table 4.9). In all donors, the majority of CRV-specific TCRβ sequences were also specifically enriched with peptide variant CRV-I. Between 15 and 67 TCRβ sequences responding to both epitope variants were found per donor. In addition, between 2 and 55 TCRβ sequences per donor were only specific for one of the epitope variants and not identified as specific with the other. The vast majority of the most frequent specific TCRβ sequences was specifically enriched with both epitope variants (Figure 4.34). Another group of exclusively CRV-specific or CRV-I–specific TCRβ clonotypes was similarly frequent in both stimulations and should therefore be specific for both epitope variants, but these clonotypes were likely narrowly below the specificity cutoffs for one of the variants. For example, one frequent TCRβ sequence of donor P07 had a proportional read frequency of 0.039 in the CRV-stimulated sample and 0.029 in the CRV-I–stimulated sample (Figure 4.34, blue "x") , yet it was only classified as specific for variant CRV. Some TCRβ sequences that were only specific for one epitope variant, however, lie outside the normal frequency distribution range of all TCRβ clonotypes found in the tested donors (marked by grey dots) and may indeed be specific for only one of the epitope variants. In conclusion, the data show that the most frequent CRV-specific TCRβ clonotypes generally recognised epitope variant

CRV-I as well. Therefore, a majority of CRV-specific T cells have antiviral potential against various different CMV strains.

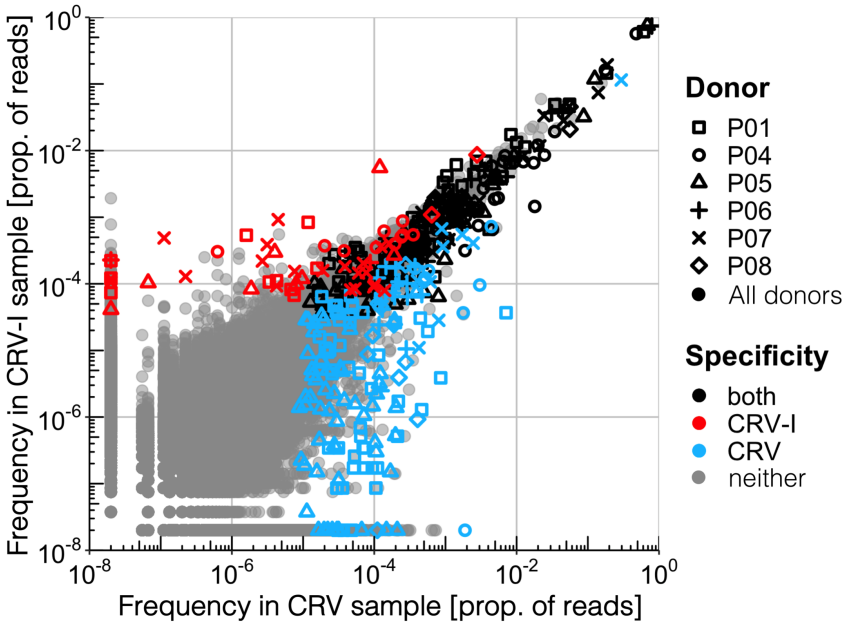


Figure 4.34: Frequencies of CRV- and/or CRV-I–specific TCRβ clonotypes in samples from donors P01 and P04–P08 stimulated with peptide CRV or peptide CRV-I. Clonotypes that were undetectable in one sample were assigned a pseudo-frequency of 2×10^{-8} to enable their display on a logarithmic axis. TCRβ clonotypes that were not specific for CRV or CRV-I, derived from any donor, are marked by grey dots. TCRβ clonotypes that were specific for either epitope variant or both are shown using hollow symbols and coloured in red (only CRV-I–specific), blue (only CRV-specific), or black (specific with both variants).

4.12 TCRα chain sequencing

Although TCRβ chain sequencing was highly efficient to identify epitope-specific TCRs and to trace virus-specific T cell responses in HLA-matched seropositive donors, I also sought to characterise the corresponding TCRα repertoire to see if some specific or shared TCRα sequences and, potentially, TCRαβ pairs can be identified.

4.12.1 Setup of the TCRα sequencing protocol

TCRα sequencing primers with the required overhangs (priming and barcoding sequences) were designed to resemble the binding properties of the TCRβ sequencing primers (see Section 3.5.1). Two different gene-specific reverse transcription primers (RT-TRAC A and RT-TRAC B), complementary to sequences in the TCRα C region, were designed and compared with the oligo-dT mix provided with the ”QuantiTect Re-

verse Transcription Kit". Two different reverse primers (CαA and CαB) that contained the required overhangs (priming sites for secondary PCR and Illumina sequencing, and barcode) were designed for and tested in the multiplex PCR step of library preparation. The first PCR was conducted with all 51 TCRα forward primers at equimolar concentrations, or with only one randomly chosen TRAV forward primer alone (TRAV12-1 from the set of newly designed primers), or with a previously published and established Vα primer TRAV19-1 (published as 12S1 in accordance with an older TCR nomenclature (Steinle et al. 1995)). Comparison of the different conditions by gel electrophoresis (Figure 4.35A) shows that TCRα products were obtained after reverse transcription with any of the tested primers, RT-TRAC A, RT-TRAC-B, or oligo-dT. Products of appropriate size were obtained with any combination of reverse primers CαA or CαB and forward primers TRAV12-1 or TRAV19-1. However, no product was obtained when the complete TRAV/TRDV mix was used with either of the reverse primers. Using the TRAV/TRDV primer mix together with reverse primer CαA, but not CαB, there was a bright band around 100 bp, where long primer by-products are expected to accumulate. This indicated that primer CαB works better that primer CαA. A possible explanation for this observation is that primer CαA may have a higher tendency to form hairpin structures, self dimers, or cross dimers with the forward primers than primer CαB. The high GC content of the binding portion of primer CαA (70%) as opposed to primer CαB (50%) may favour primer dimerisation by increasing the stability of primer dimers. The lack of visible bands for PCR product after the multiplex PCR with the TRAV/TRDV primer mix can potentially be explained by the differing lengths of the PCR products. When more TCRα chains than those expressing TRAV19-1 are amplified, the lengths of the PCR products will be more diverse. This leads to a more diffuse running behaviour of the PCR products on the gel and product bands may be too weak to be visualised by ethidium bromide staining. Next, the complete TCRα library preparation protocol was run with the same settings as the TCRβ protocol using primer RT-TRAC B as reverse transcription primer, TRAV/TRDV mix as forward primers in the multiplex PCR and CαB as reverse primer in the multiplex PCR. The purified product after the second PCR step (barcode PCR) was present at sufficient concentrations ($\geq$5 nM), as confirmed by Bioanalyzer measurement (Figure 4.35B), showing that the TCRα library preparation protocol with the primers designed in this project was effective.

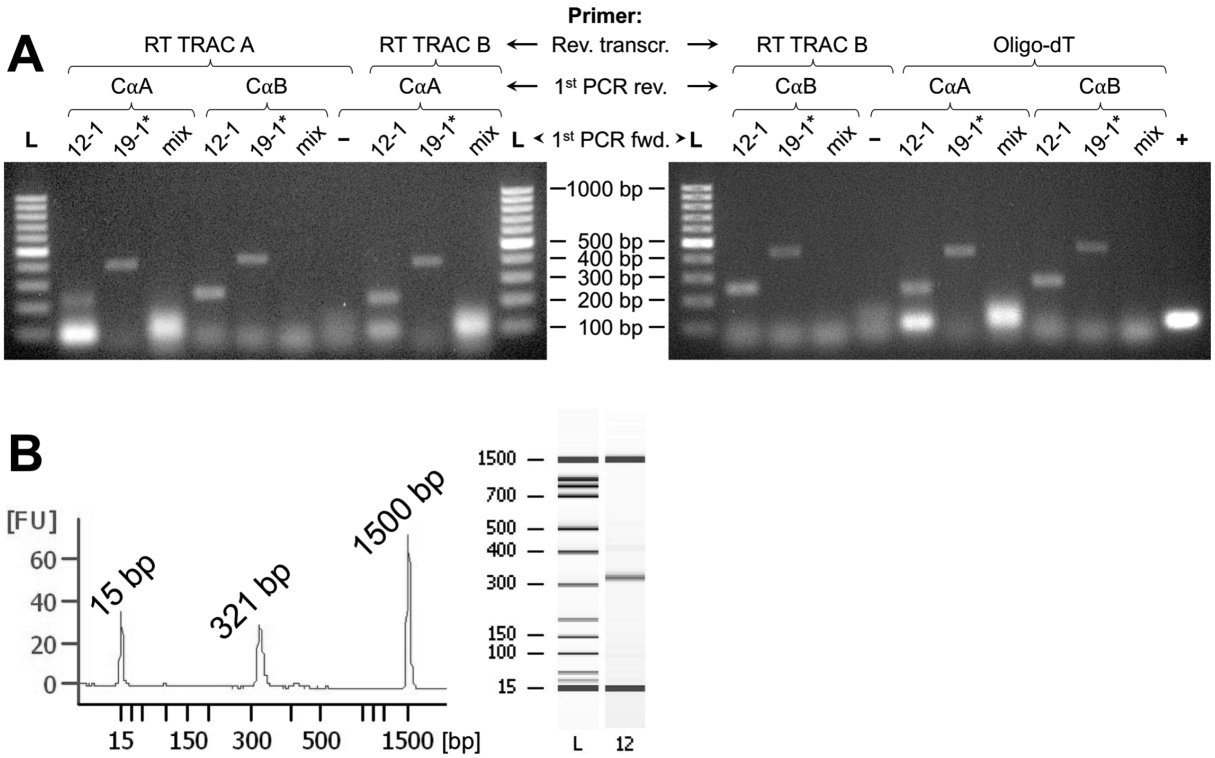


Figure 4.35: Validation of the TCRα sequencing library preparation protocol. (**A**) Gel electrophoresis after the first PCR step (multiplex PCR) with cDNA from 125 ng RNA to test different primers. Primers RT-TRAC A, RT-TRAC B, or oligo-dT were used for reverse transcription. The first PCR step was tested using only one of the TCRα primers designed in this project (TRAV12-1; here "12-1"), or all primers (TRAV/TRDV mix, here "mix"), or the established primer 12S1 binding to TRAV19-1 as positive control. The established primer TRAV19-1 was published by Steinle et al. (Steinle et al. 1995) using the V gene name 12S1 according to the nomenclature by Arden et al. (Arden et al. 1995). As reverse primer, one of the 2 primer candidates CαA or CαB were used. 26 cycles were run per reaction. Estimated product sizes with the new forward primers designed here were 225 bp with CαA and 250 bp with CαB as reverse primer. Estimated product sizes with control primer TRAV19-1 were 371 bp with CαA and 396 bp with CαB as reverse primer. L ladder, – negative control (no polymerase), + positive control (housekeeping gene GADPH) * control forward primer (**B**) Bioanalyzer electropherogram and virtual gel representation of the first TCRα library prepared from sample P07 stimulated with peptide CRV. The average PCR product size was 321 bp and the concentration of TCRα product after the second PCR step was sufficiently high for sequencing (9.7 nM).

4.12.2 VJ gene segment usage in TCRα repertoires

TCRα libraries were prepared from RNA of 4 samples from which TCRβ repertoires were previously sequenced. The 4 samples were ex vivo PBMCs and CRV-stimulated cells derived from donor P03, and CD8-enriched PBMCs and CRV-stimulated cells derived from donor P07. First, VJ gene usage in TCRα chains of the peripheral blood samples derived from donor P03 and P07 was assessed to see whether the TCRα

library preparation protocol introduced any amplification bias. As an independent control cohort, peripheral blood TCRα repertoires from 6 healthy donors published by Ruggiero et al. (Ruggiero et al. 2015) were analysed with regard to their TRAV and TRAJ gene usage. Figure 4.36 shows the proportions of TCRα clonotypes using a particular TRAV and TRAJ gene. The α-chain VJ usage of the two donors sequenced in this

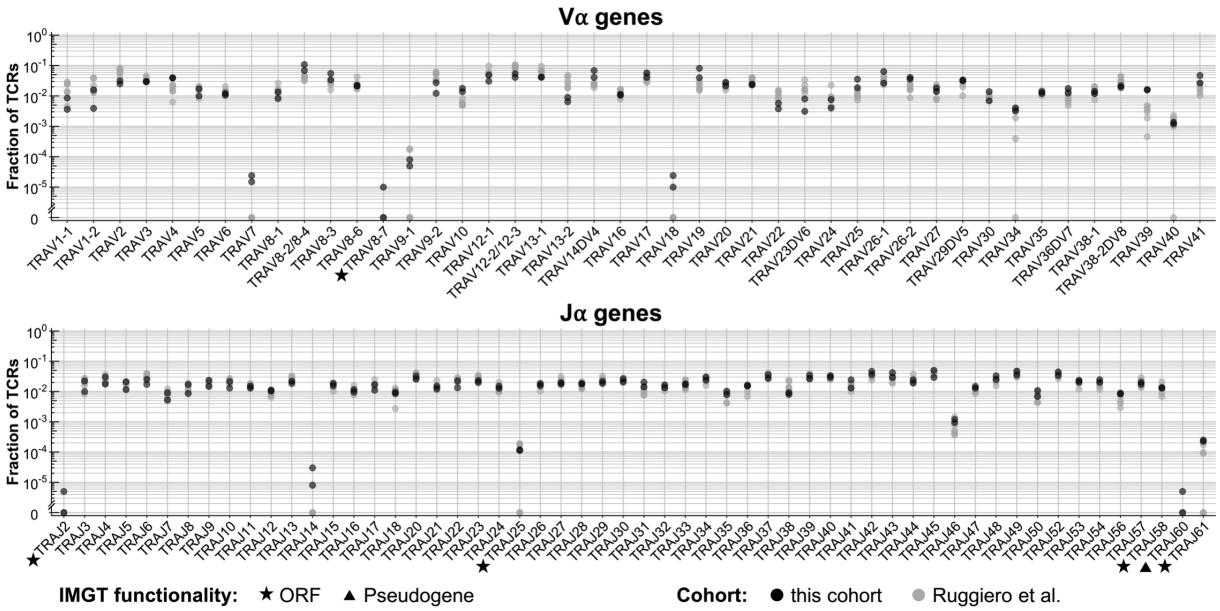


Figure 4.36: Proportion of TCRα clonotypes using a particular TRAV gene (upper panel) or TRAJ gene (lower panel) in ex vivo samples of donor P03 and donor P07 and in an independent cohort of 6 healthy donors (Ruggiero et al. 2015). Cells from donor P07 were CD8-enriched prior to RNA isolation. Grey dots correspond to donors from the cohort published by Ruggiero et al. and black dots correspond to donors P03 and P07 of this cohort. TRAV and TRAJ genes that are categorised as open reading frames or pseudogenes according to IMGT (Lefranc 2001) are marked with a star and a triangle, respectively.

project was highly similar to the α-chain VJ usage of the 6 healthy donors sequenced by Ruggiero et al. (Ruggiero et al. 2015), who used a non gene-specific cDNA amplification system with a biotinylated primer targeting the C region and a single-stranded linker cassette. The biotinylated primers were required for purification of TCR cDNA from the sample. The single stranded linker cassette was appended to the end of the cDNA and served as primer binding site for the subsequent PCR step. Functionality of each TRAV and TRAJ gene for this analysis was assigned in accordance with the IMGT TCR repertoire database (Lefranc 2001). Proportions of TCRα chains with each TRAV and TRAJ gene were between 10^{-3} and 10^{-1} for most functional Vα and Jα genes in both cohorts, and below 5×10^{-3} for ORFs and pseudogenes. Functional Vα genes

TRAV7, TRAV9-1, and TRAV18 were of low relative frequency in both tested donors of this cohort, suggesting that the Vα primers targeting these alleles may not have worked and need to be improved. However, TCRα sequences with TRAV7, TRAV9-1, and TRAV18 were also absent or of very low frequency in the 6 donors of the control cohort, indicating that these Vα genes may never or rarely be used in functional T cells. The relative frequency of functional gene TRAJ14 was remarkably low, while TRAJ58, which was annotated as an ORF (i.e. a gene segment of uncertain functionality) in the IMGT system, was used at a frequency typical for functional Jα segments in donors of both cohorts. This implies that the roles of these TRAJ genes may not have been correctly assigned in the IMGT database, and TRAJ14 may in fact be a non-functional Jα gene, whereas TRAJ58 may be functional. Because of the similar VJα gene usage patterns in the two cohorts, it can be concluded that the primers designed for TCRα library preparation in this project are functional and no discernible PCR bias was introduced in the multiplex PCR step.

4.12.3 Pairing of TCRα and TCRβ chains

TCRα and TCRβ chains are encoded by different chromosomes; these are chromosome 14 (TCRα) and chromosome 7 (TCRβ). There is currently no high-throughput method to sequence and reliably match TCR$\alpha\beta$ pairs at the high resolution of individual chain sequencing. Nevertheless, TCR$\alpha\beta$ pairs can be identified by single cell sequencing or sequencing of functionally selected clones obtained by limiting dilution. Alternatively, TCR$\alpha\beta$ pairs may be found in some cases by identifying TCRα and TCRβ chains with sufficiently matched frequencies in complex samples. Here, TCRα and TCRβ repertoires were sequenced separately from the same RNA sample. When plotting the clonotype frequencies in the unstimulated sample against their frequencies in the CRV-stimulated samples, the resulting TCRα and TCRβ clonotype distribution patterns were similar in each of the 2 donors (Figure 4.37A). For both chains, a population of distinctly enriched clonotypes was visible on the plots. Since a TCRα-sequenced control sample stimulated with a different peptide was missing, no specific TCRα sequences could be determined with the filtering strategy used above for TCRβ. However, because of the similar clonotype frequency distributions in TCRα and TCRβ samples, enriched populations of clonotypes for both chains were identified

using the enrichment cutoff values determined for TCRβ clonotypes in the identification of specific TCRβ sequences (Section 3.5.3.4). Accordingly, Figure 4.37B shows only TCRα and TCRβ clonotypes that were ≥67.12-fold (P03) and ≥4.24-fold (P07) enriched in the CRV-stimulated sample compared to the corresponding d0 sample and had a proportional read frequency $\geq 10^{-4}$. This resulted in 47 CRV-enriched TCRα and

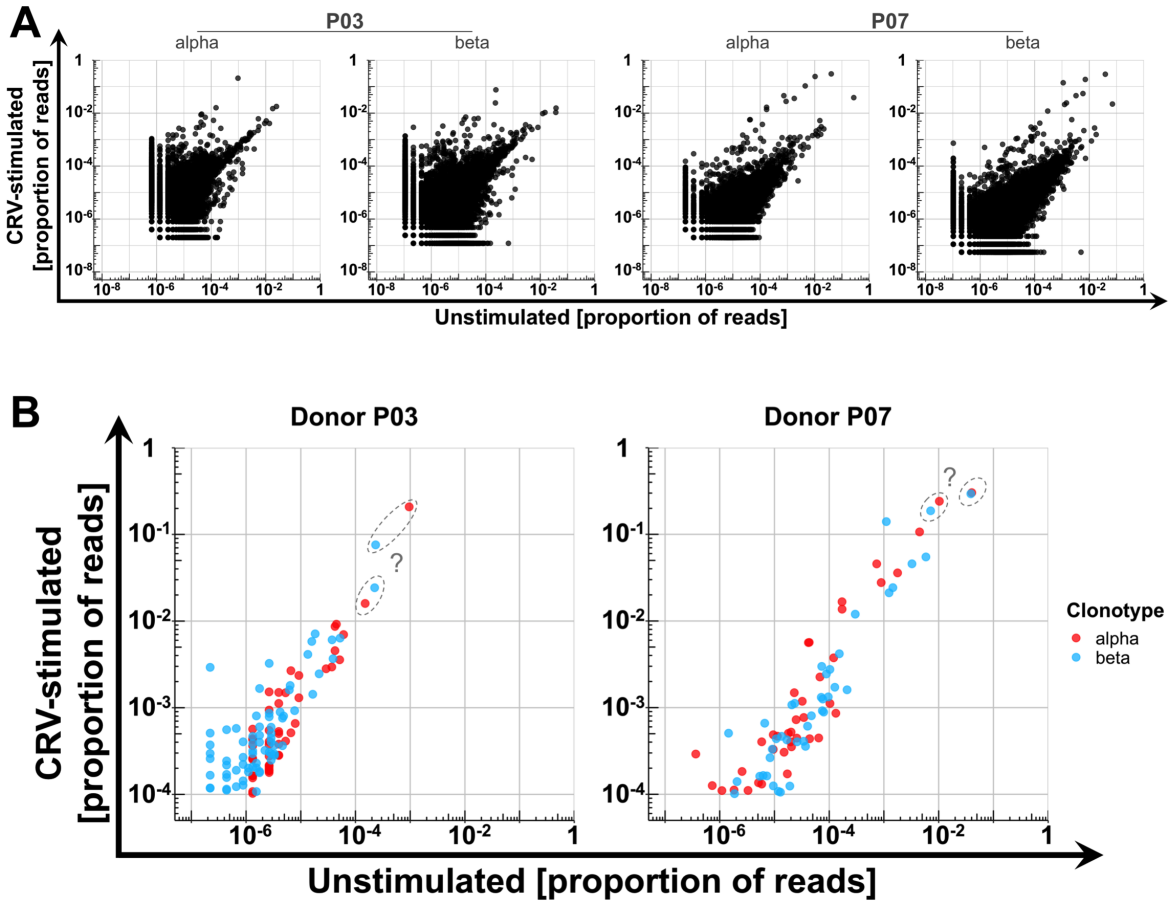


Figure 4.37: Comparison of TCRα and TCRβ repertoires derived from the same RNA samples of donors P03 and P07. (**A**) Juxtaposition of TCRα and TCRβ clonotype frequencies in ex vivo PBMCs and CRV-stimulated samples. Clonotypes that were undetectable in one sample were assigned a pseudo-frequency corresponding to 0.5 reads to enable their display on a logarithmic axis. Samples derived from donor P07 were CD8-enriched prior to RNA isolation. (**B**) Comparison of CRV-enriched TCRα and TCRβ clonotype frequencies. Displayed are all clonotypes that exceeded the enrichment cutoffs determined for TCRβ clonotypes in the identification of specific TCRβ sequences (Section 3.5.3.4) and had read frequencies $\geq 10^{-4}$. Concretely, the required enrichment factors for CRV-enriched TCRα and TCRβ clonotypes were ≥67.12 (P03) and ≥4.24 (P07). This way, 47 CRV-enriched TCRα and 71 TCRβ clonotypes were found for donor P03 and 39 CRV-enriched TCRα and 42 TCRβ clonotypes were found for donor P07. Circled clonotypes are putative TCRαβ-pairs based on similar frequency values.

71 TCRβ clonotypes in donor P03, and 39 CRV-enriched TCRα and 42 TCRβ clonotypes in donor P07. Although the clonotypes distribution patterns for both chains were similar, only the top 2 most frequent CRV-enriched TCRβ chains found in each donor could be assigned to their matching TCRα chain with high confidence. One of the 4 putative TCR$\alpha\beta$ pairs, namely the top CRV-enriched TCRα chain and TCRβ chain of donor P03, was previously found to form a functional CRV-specific TCR in our research group in T-cell clones generated by limiting dilution of cells derived from P03 (Ameres et al., unpublished). Although this confirms that matching of TCRα chains and TCRβ chains based on their frequencies in 2 samples was possible for the most frequent TCRs, TCR$\alpha\beta$ chain pairing for lower-frequency clonotypes was not feasible, since the accuracy of clonotype frequency values gets lower with decreasing relative frequency of a clonotype, and since the number of TCR chains whose frequencies are too close to each other increases. In addition, approximately 25 % of T cells express two distinct functional TCRα sequences together with one TCRβ chain (Malissen et al. 1992), which may further complicate assignment of TCR$\alpha\beta$ pairs from the two independent libraries. In order to find genuine TCR$\alpha\beta$ pairs, the samples could be analysed by single cell sequencing or by combinatory expression of TCRs, but both approaches strongly reduce the high resolution of the bulk sequencing approach presented here. Between the 2 donors, there were no shared CRV-enriched TCRα sequences, although the signature TCRβ chain TRBV25–CASTPGDEQFF–TRBJ2-1 was specifically enriched by CRV-stimulation in both donors. This indicates that there is TCRα chain variability in epitope-specific TCRs with the same TCRβ chain. Nonetheless, these observations are very limited in scope at this point. To gain insight into TCRα chain sharing between T cells with the same specificity, more samples from more donors will need to be TCRα-sequenced at high resolution.

5 DISCUSSION

In this PhD project, the CMV-specific human T-cell response was investigated in different assays combining short-term in vitro stimulation with immunogenic peptides or mini-LCLs expressing viral antigens with high-throughput TCRβ sequencing. Firstly, T cells from healthy virus carriers expressing the HLA-B*07:02/HLA-C*07:02 haplotype were stimulated with single antigenic CMV peptides to specifically enrich T cells. Enrichment of peptide-specific T cells was confirmed by flow cytometry. Bulk TCRβ sequencing and data analysis (three-sample comparison) identified 1350 donor-nucleotide-unique TCRβ clonotypes specific for 1 of 7 tested CMV epitopes from 9 CMV-positive donors P01–P08 and P14. Secondly, T cells of 3 of these CMV-positive donors (P01–P03) were co-cultured with autologous mini-LCLs that expressed a full length CMV antigen, pp65 or IE1, and presented derived peptides at physiological levels. The majority of TCRβ clonotypes specific for pp65-derived and IE1-derived peptides as identified by peptide stimulation also responded to endogenously processed and presented antigen on mini-LCLs. This confirmed that genuinely virus antigen-specific T cells can be enriched by short-term stimulation with defined CMV peptides, and that most of these T cells also respond to intracellularly processed levels of antigen. Thirdly, the assay was scaled up: 32 CMV peptides were distributed to 8 peptide pools, whereby each peptide was present in a unique combination of 3 pools and absent in the remaining 5 pools. Stimulation of 8 CMV-positive donors with various HLA types (P01, P03, P11, and P17–P21) with peptide pools identified an additional 625 TCRβ clonotypes specific for 1 of 26 tested CMV peptides. Despite the limited informative value of the data obtained by peptide pool stimulation due to peptide cross-contamination introduced by handling errors, the results imply that multiplexed stimulation with defined CMV peptides is suitable to identify TCRβ clonotypes with various specificities and restricted by many different HLAs at the same time. Lastly, 162 CMV-specific signature TCRβ sequences were identified that were shared between multiple donors. The frequency of these signature TCRβ sequences in peripheral blood was highly discriminative of a person's CMV status and HLA type in this cohort and a large independent validation cohort. Such signature TCRs will open up new possibilities for diagnosis, monitoring, and therapy of viral diseases.

5.1 Sample choice and experimental setup

Currently, the prevailing method to identify virus-specific TCRs is to label specific T cells with peptide-MHC multimers, enrich them by FACS, and perform high-throughput TCRα and/or TCRβ sequencing (Klarenbeek et al. 2012; Dash et al. 2017; Glanville et al. 2017; Attaf et al. 2018), sometimes after in vitro peptide stimulation (Chen et al. 2017; Miyama et al. 2017). This is problematic when researchers assume that all T-cell clonotypes identified from the multimer-sorted samples are CMV-specific. High-throughput sequencing is an extremely sensitive method and has the potential to identify TCR sequences even from very rare T cells in the sample and marker-positive T cell populations are not 100% pure. Some unspecific background signal, potentially caused by dominant T-cell clones of irrelevant specificity from the parent population, may be present in the marker-positive fractions. For no easily discernible reason, the purity of the marker-positive fractions that were analysed by high-throughput TCR sequencing was not stated in most of the aforementioned studies. In the two studies where the purity was stated, it was at least 95% (Klarenbeek et al. 2012; Chen et al. 2017). However, this still leaves up to 5% of T cells that are not specific for the tested epitopes in the marker-positive fraction; such impurities most probably lead to the identification of false-positive TCRs. Apart from sorting impurities, high-throughput TCR repertoires are heavily influenced by factors such as stochastical effects, polymerase errors, and sampling effects, all of which can be introduced during TCR mRNA or DNA amplification and during sequencing (Kebschull and Zador 2015). Consequently, a significant amount of TCR clonotypes obtained by high-throughput sequencing of the marker-positive fraction are likely not specfic for the tested peptides in these studies. This particularly affects TCR clonotypes with low read frequencies (see Section 4.4.1).

To reduce the background signal, a read cutoff can be introduced that eliminates low-frequency clonotyopes. As an example, Dziubianau et al. (Dziubianau et al. 2013) sequenced TCRβ libraries from T cells that were specifically enriched by stimulation with an overlapping pp65 peptide pool followed by FACS-based enrichment of activated T cells that were positive for CD40L. They introduced a frequency cutoff of 1% to exclude TCRβ sequences that were not antigen-specific; consequently only TCRβ clonotypes with a read frequency $\geq$1% were considered antigen-sepcific. While such a stringent frequency cutoff strongly reduces the background signal, it necessarily dis-

regards rarer TCRβ clonotypes that are antigen-specific. The data obtained in this project showed that the CMV-specific T-cell response was highly clonal and usually dominated by one or few abundant TCRβ clonotypes (Figure 4.10). Many additional TCRβ clonotypes with the same specificity were 100-fold–10000-fold less frequent in the peptide–stimulated samples of the respective donors; such TCRβ clonotypes would have been eliminated with a stringent $\geq$1% read frequency cutoff. It follows that a read frequency cutoff alone is not sufficient to identify specific TCRβ clonotypes from peptide-stimulated or multimer-sorted T-cell samples.

In a study published by Klinger et al. (Klinger et al. 2013), T cells from a CMV-positive donor expressing HLA-A*02:01 were labelled with pp65 NLV/HLA-A*02:01 multimers or stimulated with NLV peptide for 18 h in vitro and subsequently labelled with antibodies against the T-cell activation marker CD137. The labelled T cells were then enriched by FACS and TCRβ sequencing was performed on the multimer-positive or CD137-positive fraction, and, as control samples, on the multimer-negative or CD137-negative fraction, and on unsorted PBMCs. NLV-specific TCRβ clonotypes were identified by comparing clonotype frequencies in the positive fractions to their frequencies in both control samples. This way, 8 NLV-specific TCRβ clonotypes were identified by multimer-labelled FACS enrichment. All 8 NLV-specific TCRβ clonotypes were also identified by FACS enrichment of CD137-positive T cells, which identified 1 additional TCRβ clonotype as NLV-specific. The NLV-specific TCRβ clonotypes were approximately 100-fold more frequent in the marker-positive fractions than in PBMCs, and approximately 1000-fold more frequent compared to the marker-negative fraction. This is in line with the relative enrichment values observed in this project after single peptide stimulation (see Figure 3.5, Table 4.3, Figure 4.7), even though the three-sample-comparison performed here used a sample stimulated with a different peptide instead of a marker-negative fraction as control sample.

Sorting of multimer-positive T cells to identify CMV epitope-specific TCRβ clonotypes was also attempted in this PhD project, but enrichment of CMV multimer-positive T cells from PBMCs was done by MACS. This led to a less effective separation of peptide-specific TCRβ clonotypes and non-responder TCRβ clonotypes than enrichment by peptide stimulation done here (Figure 4.14) and FACS-based enrichment done by Klinger et al. (Klinger et al. 2013). The decreased discriminatory power of the three-

sample-comparison after MACS enrichment in this project compared to the FACS enrichment by Klinger et al. can be explained by the differences in purity of the sequenced samples. The purity of the sequenced multimer-positive fractions after MACS enrichment, as determined by FACS, was between 15.1% and 67.4% multimer-positive cells, whereas the purity of the positive fractions in the paper published by Klinger and colleagues, although not explicitly stated, was likely much higher. Consequently, the differences in frequency of specific TCRβ clonotypes in data from Klinger et al. were much larger in the marker-positive FACS fractions compared to the negative control samples than the differences in frequency found here by MACS enrichment. The proportion of peptide-specific T cells after single peptide stimulation in this project, which was measured by flow cytometry, was, on average, between 1.2% and 36% depending on the peptide. Despite this seemingly low purity, the peptide-specific TCRβ clonotypes were distinctively enriched in the peptide-stimulated samples compared to the respective controls. This suprisingly distinctive power of the peptide stimulation assay can, in part, be attributed to increased TCRβ mRNA levels several days after T-cell activation (Paillard et al. 1990). Concordantly, increased total mRNA levels after CMV peptide stimulation were also measured here (Figure 4.3). In other words, not only the count of specific T cells increases in peptide-stimulated samples (proliferation), but also the proportion of TCRβ mRNA in the activated, specific T cells increases (activation-induced transcription). Both aspects together lead to a more pronounced enrichment of peptide-specific TCRβ clonotypes compared to non-responder TCRβ clonotypes and thereby facilitate the identification of peptide-specific TCRβ clonotypes.

Klinger et al. (Klinger et al. 2013) identified NLV-specific TCRβ clonotypes in a three-sample-comparison; they demanded a $\geq$10-fold enrichment compared to both control samples and introduced a frequency cutoff equivalent to the frequency of $\geq$20 T cells in the sequenced sample. Very similar to that approach, CMV peptide-specific TCRβ clonotypes were identified in a three-sample-comparison in this project, but dynamic enrichment and frequency cutoff criteria were used to match the different TCRβ clonotype distributions obtained from different donors with different peptides (see Section 4.4.2, Table A2). In the discussed paper (Klinger et al. 2013), and in a follow-up paper of the same group (Klinger et al. 2015), in which they multiplexed peptide/HLA-A*02:01 multimers or antigenic HLA-A*02:01–restricted peptides derived from mutliple

pathogens, Klinger et al. identified 8–37 NLV-specific TCRβ clonotypes per CMV-positive, HLA-matched donor. Likewise, 8–26 NLV-specific TCRβ clonotypes were found in donors of this project by single peptide stimulation, and 10–31 NLV-specific TCRβ clonotypes were found in this project by CMV peptide pool stimulation, an assay that was developed based on the ”MIRA”-assay developed by Klinger et al (Klinger et al. 2015). Researchers who did not include control samples identified much higher numbers of peptide-specific TCRβ clonotypes, most of which are likely derived from sample impurities. For example, Chen et al. (Chen et al. 2017) identified almost 4000 NLV-specific TCRβ sequences from only 8 donors, a number that seems vastly exaggerated.

In essence, the major advantage of the assays presented in this project is that no physical isolation of peptide-specific T cells by FACS or MACS is required before TCRβ sequencing. Therefore, no expensive antibodies or multimers are needed for the identification of CMV peptide-specific TCRs and the assay can be performed with smaller cell numbers. Furthermore, short-term peptide stimulation increased the count of peptide-specific T cells in the sample, as well as the amount of total mRNA per cell (Figure 4.2B), which, in turn, increases the resolution of the assay and enables identification of rare peptide-specific TCRβ clonotypes in samples of limited size. A limitation of the peptide stimulation approach is that it identifies only such specific T cells that are capable of proliferating in vitro. However, it can be argued that these are likely the relevant T cells that play a role in pathogen control and protection from disease. Finally, peptide-MHC multimers for some peptides and HLAs are difficult to produce, despite recent advances in multimer technology (Braendstrup et al. 2013; Schlott et al. 2018), and many multimers are not readily available on the market. Besides, it was reported that multimers may not always stain the entire population of antigen-specific T cells (Rius et al. 2018), thus some antigen-specific T-cell clones may be missed using FACS-based enrichment. Taken together, these observations imply that peptide stimulation has clear advantages over enrichment of multimer-labelled cells by flow cytometry.

5.2 Sequence analysis and assignment of TCR specificity

In the experiments presented here, a potential PCR bias, which may have been introduced by the multiplex primer-based amplification system, was reduced by running only 10 cycles of the gene-specific multiplexed PCR. The multiplexed primers targetting the V region of the TCRα or TCRβ mRNA were used in excess, and because of the low number of cycles, all primers were present in sufficient amounts to ensure similar amplification efficiencies for all TCR transcripts, even in unbalanced libraries containing dominant TCR clonotypes, like those prepared from CMV peptide-stimulated samples. This does, of course, not exclude the possibility that some multiplexed V gene-specific primers bind their template better than others. However, the overall V gene and J gene usage in TCRα and TCRβ libraries prepared from PBMCs of donors in this cohort was highly similar to that of the cohort published by Ruggiero et al. (Ruggiero et al. 2015), who used an unbiased system without multiplexed V gene-specific primers (see Section 4.4.7 and Section 4.12.2). Furthermore, it was shown that omnipresent TCR repertoire skewing, which was also observed here after sequencing of a biological replicate (Section 4.4.1), was primarily caused by stochastical effects (Kebschull and Zador 2015). Primer bias, template switching, and polymerase errors only played a secondary role in TCR repertoire skewing. From these two observations, it can be deduced that the use of multiplexed primers has only minor effects, if any, on the obtained TCRβ repertoires in this project.

High-throughput TCR sequencing using the Illumina HiSeq platform is relatively error-prone, with an error rate of $\geq$0.1% (Glenn 2011). The CDR3 is a hypervariable gene region and single nucleotide polymorphisms between different T cells in a donor's T-cell repertoire will occur frequently, particularly for TCR chains that have a high formation probability due to few nucleotide insertions or deletions. Hence, it is important to reduce sequencing errors in the CDR3 region to obtain TCR repertoires that faithfully represent natural nucleotide variability. To reduce sequencing errors in the CDR3s, paired-end (bidirectional) TCR sequencing was performed in this project. TCR sequences with nucleotide mismatches in the CDR3 were either clustered to a highly similar and much more frequent parent clonotype, if such a parent population existed, or they were removed during clonotype assembly using the software MIXCR (Bolotin

et al. 2015). Barcode spillover - the assignment of reads to wrong samples due to amplification and sequencing errors in the indices - was observed, but was reduced over the course of the project by switching from a single 6-nucleotide index to a dual index system with two 8-nucleotide indices.

Enrichment of T cells by in vitro stimulation with externally loaded synthetic peptide is a somewhat artificial system and it can happen that T cells are unspecifically activated because of unnaturally high levels of peptide on the cells. Moreover, it was only partially addressed in this project if the identified peptide-specific TCRβ clonotypes functionally responded to CMV epitopes, mainly because the corresponding TCRα sequences remained unknown. Nevertheless, the data presented here strongly suggest that the CMV epitope-specific TCRβ clonotypes identified here are functionally active. Firstly, the majority of CMV peptide-specific TCRβ clonotypes was enriched both by stimulation with synthetic peptide and by stimulation with mini-LCL expressing the parent antigen (Figure 4.20). Consequently, these specific TCRβ clonotypes respond to physiological levels of endogenously processed antigen. But why is it that some CMV peptide-specific TCRβ clonotypes were not identified as antigen-specific in the mini-LCL stimulation assay? For the immunodominant CMV epitopes tested here, no specific T cells have been described in the literature that respond exclusively to the synthetic peptide, but not to the corresponding antigen. It is likely that some lower-frequency TCRβ clonotypes that were categorised as specific for CMV peptides may have been erroneously categorised as specific. Also, some CMV peptide-specific TCRβ clonotypes may have been lost due to the significantly longer stimulation period of the mini-LCL assay (30 days), with the possibility that some clonotypes were lost due to competition in cell culture. A second argument in favour of functionality of the identified CMV-specific TCRβ clonotypes is that many peptide-specific TCRβ sequences from single peptide stimulation or mini-LCL–stimulation were also identified as specific by peptide pool stimulation in donors P01 and P03. This dramatically decreases the likelihood of these TCRβ clonotypes being cell culture or stochastical artifacts. Thirdly, cumulative read frequencies of peptide-specific TCRβ sequences positively correlated with the frequency of multimer-labelled T-cells measured by flow cytometry, which showed that there is a concordance between phenotype and TCRβ mRNA expression level of peptide-specific T cells. Fourthly, many of the extensively

shared TCRβ clonotypes specific for well-studied CMV epitopes, such as NLV, RPH, and TPR, were previously found in independent studies of many different groups, in which T cell clones were functionally isolated and sequenced by classical Sanger sequencing (Weekes et al. 1999; Venturi et al. 2008; Schwanninger et al. 2008; Miconnet et al. 2011; Brennan et al. 2012; Koning et al. 2013). This validates the congruent TCRβ clonotypes and implies validity for additional private TCRβ clonotypes identified here. Moreover, it suggests functional activity, such as IFNγ secretion, for these TCRβ clonotypes. Finally, the cumulative read frequency of CMV peptide-specific signature peptide-specific TCRβ sequences was significantly higher in CMV-positive donors than CMV-negative donors in our cohort and a large, independent control cohort (Figure 4.33), implying that these TCRβ clonotypes are indeed virus-specific. Taken together, these results obtained throughout this PhD project confirmed the validity of the peptide-specific TCRβ clonotypes identified here and suggest antiviral function of such clonotypes.

Sequencing only the TCRβ chains of αβ T cells underestimates the diversity of epitope-specific T-cell responses, since a TCRβ chain can pair with different TCRα chains and still retain specificity for the same or an overlapping epitope (Miles et al. 2006; Dong et al. 2010; Misko et al. 1999). The results obtained in this PhD project, however, show that TCRβ sequencing alone is sufficient to inform about the CMV-specific T-cell status of donors. The recent large-scale study published by Emerson et al. (Emerson et al. 2017) identified some signature TCRβ sequences that were strongly associated with CMV seropositivity, but the CMV epitope or antigen specificity of these clonotypes was not determined in that study. In their study, ex vivo TCRβ sequencing and data analysis led to the identification of 164 CMV-associated signature TCRβ sequences from 641 healthy donors with known CMV status. Here, 162 antigen-specific signature TCRβ sequences were identified from only 15 CMV-positive donors by sequencing only 125 samples. Because the epitopes of the TCR sequences identified in this project were known, it was possible to include highly similar sequences with identical peptide specificity that were shared to a lesser extent. Such sequences were missed in the association-based approach by Emerson et al., which can only identify the most extensively shared TCRβ sequences. Rarer CMV-associated sequences with few incidences will be missed by their approach; they cannot be included based

on sequence similarity when the specificity is unknown, because there is a risk that similar TCRβ sequences have a completely different specificity. Another disadvantage of virus-associated TCRβ sequences as opposed to peptide-specific TCRβ sequences is that pathogens with overlapping epidemiology may negatively influence predictivity of the signature TCRβ sequences. For example, CMV-positive donors may be more likely to carry certain other viruses that CMV-negative donors, and in this case the association of CMV infection with certain TCR sequences may not indicate CMV specificity of these TCRs. In conclusion, knowledge of the antigen specificity enabled the identification of similar numbers of signature TCRβ sequences here, while using fewer resources and analysing fewer samples than Emerson et al. The 162 peptide-specific signature TCRβ sequences identified in this project distinguished CMV-positive donors from CMV-negative donors with high precision, confirming the predictive power of signature TCRβ sequences to detect virus-directed T-cell responses.

5.3 CMV-specific T cells are extremely frequent and target various viral epitopes

CMV-specific T cells against various epitopes tested in this study were highly frequent in PBMCs of healthy virus carriers (Figure 4.13), which shows how strongly CMV influences the T-cell repertoires of its hosts. The findings presented here add to previous publications reporting a high frequency of CMV-specific T cells in the circulation of virus-infected persons. For example, an early study found that 0.5%–4.5% of $CD8^+$ T cells secreted IFNγ in response to NLV or TPR peptide in CMV-positive donors expressing the appropriate HLA (Kern et al. 1999). Another group stained NLV-specific and TPR-specific circulating T cells with peptide/HLA tetramers and found a similar range of frequencies of specific $CD8^+$ T cells in their donors (0.13%–5%) (Gillespie et al. 2000). A subsequent large-scale study by Sylwester et al. extended these findings. They used overlapping peptide pools of 213 CMV ORFs to stimulate T cells derived from 33 virus carriers to assess the magnitude of the total CMV-specific T-cell response in humans by IFNγ secretion (Sylwester et al. 2005). CMV-specific T cells accounted for a median of 4.0% of $CD4^+$ T cells and 4.6% of $CD8^+$ T cells. Remarkably, in one subject, $>$30% of total CD8+ T cells were specific for the virus. Perhaps the frequency of CMV-specific T cells in the circulation of virus carriers becomes even

larger when all antigen-specific T cells are measured and not only those with active effector function at the time of analysis (Tan et al. 1999). In this project, comparable ex vivo frequencies were found in CMV-positive healthy donors: between 0.25% (donor P06, 3 peptides tested) and 6.66% (donor P02, 6 peptides tested) of total T cells were specific for the tested peptides in each donor as determined by flow cytometry. The variability of cumulative frequencies of CMV-specific T cells in the circulation of a person is influenced by several factors, including their HLA type (Höllsberg 2002), the epitopes that are recognised by their T-cell repertoire (Yewdell and Bennink 1999), the activation state of CMV (Khan et al. 2007), and the person's age (Hosie et al. 2017). Sylwester et al. showed that the CMV-specific T-cell response was not only strong, but also very complex. T cells were found to target between 5 and 55 different CMV antigens per person (Sylwester et al. 2005). Correspondingly, specific T cells and TCRβ clonotypes against a variety of different CMV epitopes derived from various antigens were found in ex vivo repertoires of donors in this cohort. Most of the epitopes used in the single peptide stimulation assay (CRV, FRC, IPS, NLV, RPH, TPR, and VLE) were previously reported to be immunodominant, i.e. to elicit immune responses in a majority of CMV-positive persons with the appropriate HLA haplotype (Khan et al. 2004). Notably, T-cell responses to epitopes FRC and VLE were shown to be a little less frequent than to the remaining epitopes: FRC elicited a tumour necrosis factor (TNF)α T-cell response in 74% (Hosie et al. 2017) and VLE elicited an IFNγ T-cell response in between 33%–54% of donors with the appropriate HLA haplotype (Khan et al. 2002a; Khan et al. 2004). The immunodominance of all 7 tested CMV epitopes was confirmed here. Remarkably, CMV peptide-specific TCRβ clonotypes were found in all HLA-matched donors for the tested peptides, even for peptides FRC and VLE. The sole exception was donor P05, who likely did not mount a TPR-specific T-cell response. This was implied by a decrease in TPR/HLA-B*07:02 multimer-positive T cells after stimulation of PBMCs from donor P05 with TPR, whereby inital levels of multimer-positive cells were already very low and likely resulted from background signal, by the lack of an enriched population of TCRβ clonotypes after peptide stimulation (plot not shown), and by the low cumulative read frequency of TCRβ sequences assigned to TPR after peptide stimulation (0.63%). Another oddity that deserves attention is the observed discrepancy between enrichment of CRV-specific T cells and TCRβ clonotypes

by single peptide stimulation of PBMCs from donors P02 and P14. In these donors, the frequency of CRV-specific T cells as measured by flow cytometry remained similar before and after stimulation with CRV peptide, suggesting that these donors did not mount a T-cell response against CRV. Nonetheless, a distinct population of enriched TCRβ clonotypes emerged after stimulation in scatter plots depicting the TCRβ sequencing data. The identified CRV-specific TCRβ clonotypes were frequent ex vivo and in the stimulated samples (see Table 4.3), constituting a paradox to what was observed by flow cytometry. A possible explanation for this paradox is that T-cell fratricide occured during peptide stimulation. T-cell fratricide has been reported in the context of adoptive T-cell therapy of tumours, where it was observed that T cells expressing a survivin-specific transgenic TCR restricted by HLA-A*02:01 killed survivin-expressing lymphocytes expressing this particular HLA (Leisegang et al. 2010). One key finding of this project is that CRV-specific TCRβ clonotypes are highly frequent in ex vivo repertoires of CMV-positive donors. Consequently, loading CRV peptide onto PBMCs will elicit a massive CRV-specific immune response. Since CRV-specific CD8$^+$ T cells express HLA class I molecules, they will also present CRV peptide and can be killed by other CRV-specific CD8$^+$ T cells. To circumvent this problem, only half of the donor PBMCs were loaded and incubated with CRV peptide and the remaining PBMCs were added after the washing steps right before starting the 10-day co-culture. This was done for all donors except donor P02 and P03, where this step was inadvertently left out. As a consequence, fratricide of CRV-specific T cells in culture is expected to be greatly reduced in donors with massive CRV-specific immune responses. Immune responses against CMV epitopes other than the 7 most intensively studied epitopes of this project were less immunodominant and elicited T-cell responses in fewer donors, which was in accordance with previously published data on these epitopes (see Section 4.8.2).

Apart from the extraordinary magnitude and diversity of the CMV-specific T-cell repertoire in seropositive donors, it has also been described that the frequency of CMV-specific memory CD8$^+$ T cells in the blood was higher in older virus carriers than in younger virus carriers, which is putatively due to an expansion of T cell memory over time following repeated episodes of CMV reactivation (Klenerman and Oxenius 2016). This CMV-dependent T-cell memory inflation was associated with greater clonality of

the CMV-specific CD8$^+$ T-cell repertoire (Khan et al. 2002b) and, potentially as a consequence of induced high clonality, may impair immunity to other pathogens in the elderly (Khan et al. 2004). Contrarily, a recent age-grouped analysis of the large data set from Emerson et al. (Emerson et al. 2017; Lindau et al. 2019) suggested that, while CMV-positive donors tended to have slightly more clonal overall TCRβ repertoires than CMV-negative donors, the clonality of the TCRβ repertoire was not significantly increased in elderly persons with the same virus status. In this PhD project, the overall frequency distribution of the top 100 most frequent TCRβ clonotypes was similar for CMV-positive and CMV-negative donors (Figure 4.13), but the abundance of CMV-specific clonotypes among the most frequent circulating clonotypes of virus carriers was strinking. No age-related analysis could be performed on this cohort, since the donors were anonymous and no age information was available. Recently, it was shown that particularly CD8$^+$ T-cells specific for HLA-C*07:02–restricted CMV epitopes are abundant in virus carriers and that such T cells dominate the CD8$^+$ T-cell compartment of the elderly (Hosie et al. 2017). Accordingly, the majority of expanded CMV peptide-specific TCRβ clonotypes, including the top most frequent one, were restricted by HLA-C*07:02 in 8 of 9 donors of this cohort that were subjected to the single peptide stimulation assay. While there is evidence that CMV infection may drive T-cell exhaustion and immunosenescene (Tu and Rao 2016), it was shown that CMV-specific HLA-C*07:02–restricted T cells retain their function in elderly individuals despite the considerable memory inflation (Hosie et al. 2017). The extraordinary high magnitude of the T-cell response specific for HLA-C*07:02–restricted CMV epitopes, together with the higher TCRβ clonotype diversity compared to other HLA restrictions (Table 4.2, Figure 4.10), suggests a special role for HLA-C*07:02–restricted T cells in virus control.

5.4 Unique features of the HLA-C*07:02–restricted T-cell response

Although it has long been established that CMV-specific CD8$^+$ T cells are associated with virus control (Quinnan et al. 1982; Cwynarski et al. 2001; Bunde et al. 2005; Sacre et al. 2008), it is still unknown which T cells are actually protective. This enigma is difficult to solve, because responses to different antigens and epitopes are associated with each other and it is difficult to examine them separately. Most studies investigated

the immune response directed towards pp65 epitopes, in particular the HLA-A*02:01–restricted epitope NLV. Less frequently, the immune response against epitopes derived from pp50 or IE1 was explored; however, mostly HLA-A and HLA-B–restricted T cells were studied. Recent data from this group suggests that not the nature of the presented peptide, but the restricting HLA determines protectivity of the T-cell response. Stefanie Ameres et al. showed that recognition of IE1-derived peptides by $CD8^+$ T cells in the presence of CMV immunoevasins was strongly HLA allotype-specific (Ameres et al. 2013). Cells infected with CMV were recognised much more efficiently by T cells restricted to HLA-C*07:02 than by those restricted by HLA-A or HLA-B alleles. Since the presence of HLA-C molecules is indispensable for prevention of NK-cell–mediated killing (Colonna et al. 1993), it can be speculated that during co-evolution of CMV and its human host, the virus has maintained HLA-C expression at the cost of being visible to specific $CD8^+$ T cells. All in all, the stable presentation of HLA-C*07:02, together with the observed higher magnitude and diversity of the HLA-C–restricted T-cell response in this project, imply that HLA-C*07:02–restricted CMV peptides may be a driving force in mobilising a protective T-cell response.

Apart from presentation of antigenic peptides to $CD8^+$ T cells, HLA-C also guides the cellular innate immune response mediated by NK cells. In doing so, HLA-C molecules interact with activating and inhibitory KIRs expressed on NK cells. HLA-C has evolved to be the dominant ligand of KIRs, and all HLA-C molecules expressed by humans contain either the C1 or C2 epitope required for inhibitory KIR recognition (Hilton and Parham 2017). Therefore, HLA-C plays a more broadly important role in protecting cells from NK-cell–mediated killing than the particular HLA-A or HLA-B alleles that can also interact with inhibitory KIRs. With their KIRs, NK cells can identify virus-infected cells by sensing reduced levels of surface expression of HLA-C, which may be caused by viral immunoevasion, or by sensing altered composition of peptides presented by HLA-C, which occurs when significant amounts of viral peptide are presented. Certain HLA-C and KIR combinations are associated with increased susceptibility to infectious diseases. Conversely, some HLA-C alleles, often in combination with particular KIRs, are associated with protection from infectious diseases. For instance, the interaction of KIR2DL2/3 with its cognate HLA-C epitope C1 was strongly associated with the development of cerebral malaria in persons infected with *Plasmodium falciparum* (Hirayasu

et al. 2012). The same combination of KIR2DL2/3 and HLA-C1 was also associated with improved resolution of hepatitis C infection (Khakoo et al. 2004).

Of the classical HLA class I loci, HLA-C was, and still is, the least studied locus (Rasmussen et al. 2014). However, given that HLA-C is relevant to both T-cell and NK-cell recognition and is therefore an important link between adaptive and innate immunity, it raises particularly interesting research questions. HLA-C is expressed at substantially lower levels compared to HLA-A and HLA-B, and the various HLA-C alleles are more closely related to each other than the HLA-A and HLA-B alleles (Zemmour and Parham 1992; Apps et al. 2015). Expression levels of HLA-C mRNA vary for different HLA-C alleles and influence NK-cell–mediated virus control, as shown for HIV (Thomas et al. 2009). Interestingly, of all HLA-C alleles tested in that study, HLA-C*07:02 was the least protective against acquired immune deficiency syndrome. In this project, it was found that the HLA-C*07:02–restricted T-cell response against CMV was extraordinarily strong and diverse. A comparably strong HLA-C–restricted immune response has not been described for other viruses yet, and only few HLA-C*07:02–restricted epitopes derived from viral pathogens other than CMV have been identified. Viral pathogens with known HLA-C*07:02–restricted epitopes include hepatitis B virus (Rivino et al. 2013), influenza A virus (Rasmussen et al. 2014), measles virus (Schellens et al. 2015), and yellow fewer virus (Rasmussen et al. 2014). Notably, one study found that, despite the reported lower expression of HLA-C alleles (Apps et al. 2015), one HLA-C*07:02–restricted measles epitope was expressed at extremely high copy numbers (Schellens et al. 2015); such ”supra-abundance” of epitopes was previously only attributed to HLA-A alleles. Despite recent advances by our group and others, research on HLA-C–restricted T-cell responses is still in the fledgling stage, and many surprising discoveries can be expected in this area in the future.

5.5 Patterns of TCR specificity

162 CMV peptide-specific TCRβ sequences were found to be shared between donor (”public”) in this project. These shared TCRβ sequences were not typically among the most frequent specific clonotypes in the repertoires obtained from PBMCs. Accordingly, the top most frequent peptide-specific TCRβ sequences were usually different for different donors and mostly exclusive to a certain donor (”private”). This is in con-

trast to what was observed for other viral infections, such as influenza (Lehner et al. 1995; Chen et al. 2017) or EBV (Miles et al. 2006), where some heavily conserved public TCRβ sequences dominated the virus-specific T-cell repertoires of its carriers. The CMV peptide-specific TCRβ repertoires identified here were mostly composed of clonotypes that were only moderately shared between donors or entirely private. Similar observations have already been made for CMV-specific HLA-A–restricted and HLA-B–restricted T cells (Price et al. 2005; Venturi et al. 2008; Miconnet et al. 2011; Dash et al. 2017; Glanville et al. 2017; Klinger et al. 2015; Trautmann et al. 2005). The results of this PhD thesis confirm these observations and extend them to HLA-C*07:02–restricted CMV-specific T cells. Thus, there is no CMV epitope known to date that activates universally dominant TCRβ clonotypes as it was observed for other pathogens.

In this project, 45 TCRβ sequences were found to be specificially enriched with the same peptide in multiple donors. Some of these 45 shared TCRβ sequences with the same peptide specificity were very similar to each other in that they had the same Vβ gene usage, CDR3β length, and 1 or 2 amino acid differences in the CDR3β. These observations expand on data from previous studies showing that virus-specific TCRβ chains restricted by HLA-A or HLA-B can be very similar in their amino acid sequence. Similarities among specific TCRβ sequences have been described for many epitopes derived from herpesviruses CMV (Weekes et al. 1999; Price et al. 2005; Trautmann et al. 2005; Venturi et al. 2008; Miconnet et al. 2011; Koning et al. 2013; Klinger et al. 2015; Dash et al. 2017; Glanville et al. 2017) and EBV (Lim et al. 2000; Fazou et al. 2001; Miles et al. 2005; Miles et al. 2006; Venturi et al. 2008; Miconnet et al. 2011; Koning et al. 2013), but also for epitopes derived from HIV (Miconnet et al. 2011) and influenza virus (Lehner et al. 1995; Chen et al. 2017). These TCRβ sequences used identical or closely related TRBV and TRBJ genes and showed exchanges in few amino acid positions. In an attempt to categorise specific T-cell responses, several groups searched not only for identically shared TCR sequences, but for short motifs within the CDR3 that may be indicative for a certain epitope specificity. Short motifs were found in the central portion of the CDR3 (Miconnet et al. 2011; Lehner et al. 1995), somewhere near residues 6 and 7. These residues were found to form strong bonds to the peptide-MHC complex and are therefore thought to significantly influence

the specificity of a TCR (Stadinski et al. 2016). The 162 signature TCRβ sequences identified here were most variable in positions 5–6 compared to other members of the same specificity cluster, but variations were also observed at more C-terminal positions, particularly in TCRβ sequences with long CDR3s (Figure 4.30, Table 4.8). Some TCRβ clusters only tolerated conservative amino acid exchanges, indicating that the chemical properties of the amino acids in these positions are indispensable for epitope recognition. For example, only a conservative S↔T exchange appeared to be tolerated on position 4 in TCRβ sequences cluster CRV01 without loss of specificity for CRV/HLA-C*07:02. Other TCRβ clusters tolerated more variation, like cluster RPH01, in which amino acids with diverse chemical properties were found, which implies that these residues may be less involved in peptide-MHC recognition. The search for common features of TCRβ sequences with the same epitope specificity was recently taken to the next level, when two research groups independently designed algorithms to cluster sequences from T cells of the same specificity by using local similarity (CDR3β motifs), or global similarity as discriminator (Dash et al. 2017; Glanville et al. 2017). Global similarity was a better predictor of TCRβ specificity than short sequence motifs, which is in line with results from this project indicating that the whole TCRβ sequence needs to be considered when investigating patterns of specificity. This makes sense, since while the CDR3 sequence is arguably a strong determinant of TCR specificity, CDR1 and CDR2, which are fully encoded by the V gene segment, also play a major role in epitope recognition by establishing bonds mainly to the peptide-presenting HLA. A brief CDR3β motif search in the data presented here, without taking into account the used TRBV gene, showed that most short motifs were not exclusive to the specificity from which they were identified, but also extended to other TCRβs with different CMV peptide specificity. Moreover, the short sequence motifs were generally located in positions of the CDR3β that were encoded, in parts or completely, by the Vβ, Dβ, and/or Jβ gene segment, which increases the general likelihood of such motifs to be present in a person's TCRβ repertoire. It seems questionable that TCRβ clonotype specificity can be pinpointed to a short 3–4 amino acid motif, as suggested by Glanville et al. and Dash et al. (Dash et al. 2017; Glanville et al. 2017), when there are an estimated 3×10^{11} naive T cells (Jenkins et al. 2009) and 10^{10} distinct TCRβ clonotypes in the human body. Chances are that such short motifs are frequently produced by indepen-

dent TCR recombination events. Thus, short motifs fail to reflect the true diversity of a specific T-cell response. Considering the usage of a certain TRBV gene in addition to a motif will likely improve the results, but to be able to predict TCRβ specificity, algorithms should rather focus on the identity of full-length CDR3βs with amino acid exchanges tolerated in particular positions. Applying the concept of "sequence specificity clusters", which was carved out in this thesis and is based on same Vβ gene usage, same CDR3 length, and a highly similar CDR3β amino acid sequence, will help to improve the predictive power of TCRβ specificity prediction algorithms compared to short CDR3 motifs. Because of the high prevalence of CMV and the exceptional magnitude of the CMV-specific T-cell response, the virus is suitable as a model to study specific TCRβ repertoires with regard to sequence similarities and differences, which will also expand our limited knowledge of the human T-cell response in general.

5.6 Application prospects of virus-specific TCRβ signatures

Even though public TCRβ clonotypes did not dominate the peptide-specific T-cell repertoire of the donors in which they were identified, the set of 162 signature TCRβ sequences identified here was highly discriminative of donor CMV status and HLA type. The signature TCRβ sequences were found with considerable cumulative read frequencies in CMV-positive donors expressing the relevant HLA haplotype, but rarely in donors that were either CMV-negative or lacked the appropriate HLA. Consequently, signature TCRβ sequences are unlikely to be found in T cells of irrelevant specificity. The data presented here also showed that including signature TCRβ sequences restricted to additional HLAs to cover more HLA loci increased the cumulative frequency of signature TCRβ sequences (Figure 4.33A). It would therefore be worthwile to extend our approach to more CMV epitopes restricted by many different HLAs in order to distinguish not only donors co-expressing HLA-B*07:02/HLA-C*07:02, but also other donors with different HLA types. The predictive power of the signature TCRβ sequences was confirmed in this study using this cohort and a large independent control cohort (Emerson et al. 2017). Therefore, our approach to identify signature TCRβ sequences can be extended to more epitopes, viruses, and other pathogens in the future.

Identification of a large set of specific signature TCRβ sequences covering multiple

pathogens and HLA types can find application in various areas, for example in clinical research, diagnostics, and therapy. Short-term in vitro stimulation combined with high-throughput sequencing is a convenient approach to analyse TCR repertoires and pave the way for novel disease management strategies and immunotherapies. Clinical immunologists recently tend to prefer in vitro stimulation over purification of specific T cells by multimer staining and flow cytometry to produce T-cell products for clinical application (Feuchtinger et al. 2004; Feuchtinger et al. 2006; Icheva et al. 2013; Papadopoulou et al. 2014; Tzannou et al. 2017; Gary et al. 2018), whereas researchers who perform TCR repertoire analyses still greatly rely on the more expensive, more intricate, and more limited flow cytometry approach. In clinical research, eptiope-specific signature TCRβ sequences can be used as a marker to track virus specific T-cell responses, for instance after hematopoietic stem cell transplantation. For example, if signature TCRβ sequences specific for CMV epitope CRV are found in a donor, they likely also have several T-cell clonotypes expressing private TCRβ sequences specific for this epitope. Because the presence of signature TCRβ sequences is indicative of the presence of an epitope-specific T-cell response, it is possible to correlate the presence of T cells of a certain specificity to viral load and cases of CMV disease to identify those CMV epitopes that elicit a protective immune response. Compared to multimer staining to identify specific T-cell responses in patients, TCRβ sequencing is more sensitive, fewer T cells are needed in a sample, and TCRs with the same epitope specificity, but potentially different function, can be distinguished. This is particularly important when investigating T cell responses in HSCT recipients, whose immune system takes a long time to recover, during which the amount of T cells in their circulation can be much lower than in healthy donors.

After the identification of protective T-cell responses, protection-associated signature TCRβ sequences will be suitable to guide treatment of patients at risk of CMV disease. If a patient already has protective CMV-specific $CD8^+$ T cells in their repertoire, it may not be necessary to treat them with antivirals, which can have severe side effects or lead to selection of resistant virus strains. Conversely, if a patient lacks such protective T cells, they may be in elevated need of an antiviral treatment regimen. Furthermore, signature TCRβ sequences of protective T-cell responses can be used to confirm the success of experimental CMV vaccines. If a novel CMV vaccine can induce virus-

specific T cells that were associated with protection from CMV disease, it is a highly promising candidate for clinical trials.

TCRβ sequencing could replace current CMV monitoring strategies, such as peptide-MHC multimer staining (Cwynarski et al. 2001; Gratama et al. 2001) or other immunoassays, as a highly parallelisable and economic standalone method to identify virus-specific T cells in healthy persons and patients in need of antiviral regimens. Furthermore, this TCRβ sequencing-based monitoring approach could be extended to explore T-cell responses to various different epitopes, HLAs, and pathogens simultaneously. Much like a detectable antibody response can be detected by immunoassays like ELISA, pathogen-specific T-cell repertoire signatures can be read by high-troughput TCRβ sequencing, and TCRβ sequencing is more precise and sensitive than other immunoassays. Because TCRβ sequencing is a highly sensitive method, it only requires a small amount of blood sample. T-cell responses of HSCT recipients may be more easily identified using genetical methods, such as TCRβ sequencing, than with flow cytometry, since the T-cell activation status will be more heterogenous than in healthy patients. This can, in turn, lead to irregular multimer staining, putatively because of TCR internalisation, downregulation of other T-cell markers, or increased background signal. Like many other newly developed technologies, it can be expected that the costs of high-throughput sequencing will decrease over the following years, making TCRβ sequencing economically more attractive as a diagnostic tool. In addition, TCRβ sequencing for diagnostic purposes is highly parallelisable, as signature TCRβ sequences against many pathogens can be identified from only one sequencing sample. This way, the protection status of a person against selected pathogens can be determined to identify patients in need of therapy, such as adoptive T-cell therapy or antiviral regimen, which would, in turn, reduce costs and adverse effects of unnecessary treatments. Moreover, the readout of signature TCRβ against multiple pathogens from a single blood sample would minimise the amount of required patient material for diagnostic purposes. To expedite the development of TCRβ sequencing-based diagnosis and patient monitoring, specific signature TCRβ to a multitude of pathogens and HLA types need to be identified. Many groups have already studied pathogen-specific TCRβ repertoires by traditional or high-thoughput sequencing. Their results are collected in the data base vdjdb (Shugay

et al. 2017), a curated repository for antigen-specific TCR sequences. Since TCR sequencing has become a popular technology to assess T-cell responses, several groups have developed tools to simplify and standardise the complex task that is processing of TCR sequencing data (Bolotin et al. 2015; Nazarov et al. 2015). Simultaneously, data bases for T-cell epitopes (www.iedb.org (Vita et al. 2018)) and crystal structures of TCRs in complex with their cognate peptide-MHC (www.rcsb.org (Berman et al. 2000)), and epitope prediction algorithms (www.syfpeithi.de (Rammensee et al. 1999), http://www.cbs.dtu.dk/services/NetMHC-4.0/ (Nielsen et al. 2003; Andreatta and Nielsen 2015)) will guide the identification of novel pathogen-derived epitopes and specific signature TCRs. Published high-throughput TCR sequence datasets (Ruggiero et al. 2015; Emerson et al. 2017) will be extremely valuable in order to identify and validate novel signature TCRs specific for many different epitopes and pathogens.

Extensively shared CMV-specific TCRs may be of special interest for adoptive T-cell therapy. In this PhD project, some CMV peptide-specific TCRβ sequences were enriched in multiple donors with various HLA backgrounds, which shows that such TCRs are frequently present in the self-tolerant TCR repertoires of healthy donors. Consequently, such TCRs are likely nonresponsive to human self-antigens in diverse genetic backgrounds. The use of extensively shared specific TCRs in adoptive T-cell therapy will greatly reduce the risk of allo-HLA cross-reactivity, which is frequently observed with virus-specific $CD8^+$ T cells (Amir et al. 2010), and thus reduce the risk of graft-versus-host disease. Therefore, such TCRs are suited best for use in immunotherapy of viral reactivation and disease with TCR-transgenic T cells (Schub et al. 2009). Antiviral activity of adoptively transferred T cells produced from donor material hast been observed in several clinical trials (Walter et al. 1995; Einsele et al. 2002; Cobbold et al. 2005; Leen et al. 2006; Micklethwaite et al. 2007; Papadopoulou et al. 2014; Gary et al. 2018). Using TCR-transgenic CMV-specific T cells (Schub et al. 2009) equipped with a TCR that tolerates multiple HLA backgrounds will enable the production of CMV-specific T-cell grafts even when the donor is seronegative for the virus and with the best chance of avoiding allo-reactivity.

5.7 Future objectives of TCR sequencing

Since the initial description of high-throughput TCR sequencing in 2009 (Robins et al. 2009), the method already transformed and significantly accelerated many areas of immunological research. TCR sequencing is a powerful, sensitive, and gradually more economical method to characterise complex T-cell responses against a multitude of pathogens and autoantigens in various vertebrates. With growing data sets of specific TCRs, we will eventually be able to answer some long-standing questions about the human pathogen-specific T-cell response. For example, it would be interesting to identify common TCR sequence features of the T-cell response against certain epitopes, as undertaken by Glanville at al. and Dash et al. (Glanville et al. 2017; Dash et al. 2017). Based on these results, computational algorithms could be developed to predict the specificity of a T-cell based on its TCR sequence. However, the human T-cell response is highly diverse due to the variability of the TCR itself, the extreme polymorphism of the MHC loci, and the multitude of antigens and derived epitopes. The TCR sequencing data obtained in this PhD project imply that there is little conservation of TCRβ sequences with the same specificity. Conserved amino acids of TCRβ sequences with the same specificity were mostly germline encoded. Short CDR3β sequence motifs were not exclusive to one CMV specificity, and Vβ gene usage was diverse among TCRβ sequences with the same specificity, even though some Vβ genes were preferentially used. Only 162 of 1809 specific TCRβ sequences were found to be shared or similar in multiple donors. Consequently, it is questionable whether computational algorithms will ever be able to fully encompass and categorise the immense diversity of the human T-cell response.

High-throughput TCR sequencing can also be employed to identify T cells and antigens that cause autoimmune diseases. The MHC locus is the most polymorphic gene region in the human genome, with $>$7200 HLA class I alleles and $>$2000 HLA class II alleles presently known (Robinson et al. 2014). A popular hypothesis on what caused the immense diversity of the MHC locus is that it was evolutionarily necessary for slowly evolving vertebrates to diversify its MHC locus in order to remain protected against more rapidly evolving pathogenic microorganisms and viruses (Sommer 2005). A person's individual T-cell repertoire is shaped by the HLA alleles that they express. Certain HLA alleles are strongly linked to autoimmune diseases (Shiina et al. 2009), such

as celiac disease (HLA-DQ2 (Sollid et al. 1989)), multiple sclerosis (HLA-DR15 and HLA-DQ6 (Terasaki et al. 1976; Compston et al. 1976; Compston and Coles 2008)), narcolepsy (HLA-DQ1 (Mignot et al. 1994)), or psoriarsis (HLA-Cw6 (Tiilikainen et al. 1980)), or to susceptibility to viral diseases (Smeraldi et al. 1986; Lenzi et al. 1998; Bohl et al. 2005). Although associations of HLA alleles to some diseases were identified decades ago, the identities of autoreactive T cells interacting with these HLAs is often not known yet. Certain autoimmune diseases are linked to viral infections or vaccinations against viruses. For instance, infection with EBV or HHV-6 was found to be linked to multiple sclerosis (Challoner et al. 1995; Lünemann et al. 2006; Ascherio and Munger 2007; Jilek et al. 2008; Farrell et al. 2009), and vaccination with the influenza vaccine Pandemrix increased the incidence of narcolepsy approximately 9-fold (Nohynek et al. 2012). One of the proposed mechanisms for this correlation of viral infections and autoimmune diseases is a process termed ”molecular mimicry”. This hypothesis is based on the assumption that some viral peptides are highly similar to self peptides and that presentation of such viral peptides can activate autoreactive T cells (Oldstone 1998). High-throughput TCR sequencing can help to identify such cross-reactive T cells, as well as other T cells causing autoimmune diseases.

Another interesting aspect to investigate using TCR sequencing is why TCR repertoires against viruses and other pathogens are so variable between individuals. In this thesis, it was implied that shared TCRβ sequences are more likely to be recombined than private TCRβ sequences (Figure 4.31). Still, it is unlikely that TCRβ sequences, even those with high recombination probabilities, are produced in all humans. Some shared TCRβ sequences cannot be made by donors with a common deletion in the TRBV gene locus (Brennan et al. 2012). For example, TCRβ sequences of specificity cluster RPH01 cannot be made by such donors, because the Vβ gene used in these TCRs, TRBV4-3, was deleted in these donors. Therefore, some TCRs cannot be produced in certain human beings due to genetic predisposition.

The different HLA combinations of different persons also enhance the overall diversity of specific TCR repertoires. Some pathogen-specific TCRs may cross-react with self-peptides in the context of a different HLA and are therefore removed by negative selection in the thymus. This hypothesis predicts that TCR repertoires against a viral peptide presented by a certain HLA will be more similar in persons whose other HLAs

are more similar than in persons whose other HLAs are mostly different. Negative selection was implied as one driving force of T-cell diversity in a set of publications studying the human TCR$\alpha\beta$ repertoire against an immunodominant HLA-B*08:01–restricted epitope (FLRGRAYGL) derived from EBV antigen EBNA3A (Burrows et al. 1994; Argaet et al. 1994; Burrows et al. 1995). The authors initially showed that T cells specific for the FLRGRAYGL epitope cross-reacted with HLA-B*44:02. Conversely, the majority of alloreactive T cells against HLA-B*44:02 produced in vitro also recognised the HLA-B*08:01–restricted EBV epitope. In a follow-up study, they found striking TCR publicness between EBV-positive donors: 4 of 5 tested donors had T cells with the exact same TCR against epitope FLRGRAYGL in their repertoire; the fitfth donor had a highly similar TCRβ chain. A year later, the authors demonstrated that donors expressing both HLA-B*08:01 and HLA-B*44:02 lacked the public FLRGRAYGL-specific TCR, implying that it was deleted by negative selection in the thymus. Recently, it was shown that the alloreactive FLRGRAYGL-specific T cells bind to a self peptide (EEYLQAFTY) that only shares one conserved and two similar amino acids with the EBV epitope (Macdonald et al. 2009), suggesting that conformational mimicry of epitopes is more important than sequence identity for cross-recognition of alloreactive T cells. The role that the HLA type plays in the deletion and diversification of TCRs of the CMV epitope-specific TCRβ repertoires was beyond the scope of this project, but should be addressed in future TCR repertoire studies.

Another aspect that shapes the human TCR repertoire is the infection history of a person. Infection with unrelated pathogens can potentially activate T cells that cross-react with other epitopes derived from pathogens of subsequent infections, a process termed ”heterologous immunity”. As a consequence of different infection histories between individuals, different cross-reactive T cells may be activated that differentially influence the pathogen-specific TCR repertoires of these individuals. Few cases of cross-reactivity of cytotoxic $CD8^+$ T cells with different epitopes from different viral pathogens were initially described in mice (Yang et al. 1989; Kulkarni et al. 1993; Selin et al. 1994), and later also in humans (Misko et al. 1999; Nilges et al. 2003; Wedemeyer et al. 2001; Clute et al. 2005). Cross-reactivity in humans was observed for T cells specific for HLA-A*02:01–restricted epitopes; examples are epitopes derived from the virus pairs influenza A virus and hepatitis C virus (Wedemeyer et al. 2001), influenza A

virus and EBV (Clute et al. 2005), or human papilloma virus and coronavirus (Nilges et al. 2003). Moreover, one study reported cross-reactivity of HLA-B*08:01–restricted EBV epitope-specific $CD8^+$ T cells with a self peptide and a bacterial peptide from *Staphylococcus aureus* (Misko et al. 1999). The pairs of cross-recognised epitopes were moderately sequence-related: between 33% and 66% of sequence identity was found. Despite these reports, it is unclear how common cross-reactivity of $CD8^+$ T cells really is across the repertoire, and whether structural identity or sequence identity of epitopes is the major driver of cross-reactivity. In the data presented in this thesis, 90 TCRβ clonotypes were found to respond to mutliple tested CMV peptides in the same donor. These clonotypes, however, often responded to epitopes restricted by different HLA alleles, and were of comparatively low frequency in all or all but one samples. Thus, they were likely enriched in the other samples because of stochastical effects rather than because of true multi-specificity. Although only epitopes derived from CMV, and not across different pathogens, were studied here, the data implies that cross-reactive $CD8^+$ T cells are not a common phenomenon in humans. In line with this, a recent study failed to find cross-reactive $CD8^+$ T cells specific for epitopes derived from CMV, EBV, and influenza A virus (Rowntree et al. 2018). Importantly, the group found no cross-reactive $CD8^+$ T cells against an HLA-A*02:01–restricted epitope pair derived from EBV and influenza A virus, for which cross-reactivity was previously described (Clute et al. 2005). Additionally, the group reported that they found no structural similarities between the tested epitopes of different viruses that would support cross-reactive $CD8^+$ T-cell responses. They concluded that T cell cross-reactivity between different pathogens may not be a phenomenon as common as previously assumed.
Cross-reactivity of $CD8^+$ T cells with homologous epitope variants derived from different strains of the same virus, however, seems to be a widespread phenomenon. For example, it was shown that conserved sequence-related, non-identical epitopes from different influenza A strains and even different influenza viruses were cross-recognised by the same T cells (Gras et al. 2010; Koutsakos et al. 2019). T cells cross-reacting to multiple strain variants of an epitope often expressed the same TCRs within donors (Grant et al. 2018). In both studies on influenza A virus (Gras et al. 2010; Grant et al. 2018), it was shown that the cross-reactive influenza A epitopes have a highly similar conformation, which further corroborates the hypothesis of cross-strain recognition of

epitopes by $CD8^+$ T cells. The data from this PhD project confirmed these findings: The majority of TCRβ clonotypes specific for CMV epitope CRV also specifically expanded with strain variant epitope CRV-I (Table 4.9, Figure 4.34). These two epitope variants cover all CMV strains known to date. T cells expressing TCRs that are able to recognise different strain variant epitopes may be particularly important in protection from the virus. Furthermore, conserved epitopes that are recognized in donors with different, yet related HLA alleles deserve special attention. In this cohort, NLV-specific TCRβ sequences were found in all donors expressing HLA-A*02:01, the most frequent HLA-A allele in persons of European descent, but also in an additional donor of Asian descent, who expressed the closely related HLA alleles HLA-A*02:03 and HLA-A*02:06. It was previously described that donors expressing HLA-02:06 in the absence of HLA-A*02:01 mobilised an NLV-specific T-cell response (Trivedi et al. 2005; Miyama et al. 2017). No similar NLV-specific TCRβ sequences were found to be shared between the donors expressing HLA-A*02:01 and the donor expressing HLA-A*02:03 and HLA-A*02:06. However, if more NLV-specific TCRβ clonotypes from donors expressing HLA-A2 alleles other than HLA-A*02:01 were to be found in the future, it would be interesting to search for cross-reactive TCRs that recognise NLV presented on different HLA alleles. Such cross-reactive TCRs would be highly interesting for basic research, but also for the development of TCR-engineered CMV-specific T-cell grafts for adoptive transfer, since they can find application in many patients of different ethnicities. Co-infection with different CMV strains is common, but it remains controversial if the diversity of different CMV genomes a person harbours contributes to higher viral loads and disease progression. For EBV, it was shown that different strains with different biological properties circulate in a person, and that some epitopes are highly variable between strains, while others are conserved (Palser et al. 2015; Cirac et al. 2018). Renzette et al. demonstrated that CMV genomes in congenitally infected neonates were highly variable, with only 25% of the genome being fully conserved, and that CMV strain occurrence between different sample types, peripheral blood and urine, was highly variable and resembled the variance of the same sample type between different donors (Renzette et al. 2013; Renzette et al. 2015). Another group showed that diversity hotspots between different CMV strains were mostly found in genes that other members of the herpesvirus family do not possess, whereas conserved regions of the

CMV genome could be mapped to genes that are found in many other herpesviruses (Sijmons et al. 2015). Smith et al. recently investigated the specific $CD8^+$ T-cell response for pairs of epitope variants derived from IE1 of different CMV strains in patients who underwent HSCT (Smith et al. 2016). In their cohort, 52% of patients were infected with multiple CMV strain variants as distinguished by polymorphisms in the IE1-encoding gene UL123. They found that some homologous epitope variant pairs induced comparably strong T-cell responses. Contraily, only one of two tested variants of other homologous epitope pairs was capable to induce a variant-specific or cross-reactive T-cell response in patients, although both variants were present. In summary, it appears that the CMV-specific T-cell repertoire of a carrier is shaped by the virus strains they are infected with and the epitope variants that are processed and presented. This has important implications on adoptive T-cell therapy, since limited cross-reactivity of a T-cell graft will impair its antiviral potential. T cells expressing cross-reactive TCRs that can respond to multiple strain variants of viral peptides, such as those recognising both CRV and strain variant CRV-I, are of special interest for the development of vaccines and adoptive T-cell therapies, because they can confer protection against many virus strains at once.

REFERENCES

Acquaye-Seedah, E., Frye, Z. P., and Maynard, J. A. (2015). Immunotherapeutic approaches to prevent cytomegalovirus-mediated disease. *Antibodies for Infectious Diseases*. American Society of Microbiology, 273–288.

AdisInsight database (2018). Drug profile VCL 6365. Springer. URL: https://adisinsight.springer.com/drugs/800032591 (visited on 04/05/2019).

Adler, S. P., Starr, S. E., Plotkin, S. A., Hempfling, S. H., Buis, J., Manning, M. L., and Best, A. M. (1995). Immunity induced by primary human cytomegalovirus infection protects against secondary infection among women of childbearing age. *Journal of Infectious Diseases* 171(1), 26–32.

Adler, S. P., Manganello, A.-M., Lee, R., McVoy, M. A., Nixon, D. E., Plotkin, S., Mocarski, E., Cox, J. H., Fast, P. E., Nesterenko, P. A., Murray, S. E., Hill, A. B., and Kemble, G. (2016). A Phase 1 Study of 4 Live, Recombinant Human Cytomegalovirus Towne/Toledo Chimera Vaccines in Cytomegalovirus–Seronegative Men. *The Journal of Infectious Diseases* 214(9), 1341–1348.

Alcami, A. and Koszinowski, U. H. (2000). Viral mechanisms of immune evasion. *Trends in Microbiology* 8(9), 410–418.

AlphaVax (a). Clinical experience. URL: https://www.alphavax.com/clinical-experience.html (visited on 04/05/2019).

AlphaVax (b). Pipeline. URL: https://www.alphavax.com/clinical-experience.html (visited on 04/05/2019).

Ameres, S. (2013). The T cell repertoire specific for the IE-1 protein of human cytomegalovirus: diversity, function and evasion. PhD thesis. Universitätsbibliothek der Ludwig-Maximilians-Universität.

Ameres, S., Mautner, J., Schlott, F., Neuenhahn, M., Busch, D. H., Plachter, B., and Moosmann, A. (2013). Presentation of an immunodominant immediate-early CD8+ T cell epitope resists human cytomegalovirus immunoevasion. *PLOS Pathogens* 9(5), e1003383.

Ameres, S., Besold, K., Plachter, B., and Moosmann, A. (2014). CD8 T Cell–Evasive Functions of Human Cytomegalovirus Display Pervasive MHC Allele Specificity, Complementarity, and Cooperativity. *The Journal of Immunology* 192(12), 5894–5905.

Ameres, S., Liang, X., Wiesner, M., Mautner, J., and Moosmann, A. (2015). A diverse repertoire of CD4 T cells targets the immediate-early 1 protein of human cytomegalovirus. *Frontiers in Immunology* 6, 598.

Amir, A. L., D'Orsogna, L. J. A., Roelen, D. L., Loenen, M. M. van, Hagedoorn, R. S., Boer, R. de, Hoorn, M. A.W. G. van der, Kester, M. G. D., Doxiadis, I. I. N., Falkenburg, J. H. F., Claas, F. H. J., and Heemskerk, M. H. M. (2010). Allo-HLA reactivity of virus-specific memory T cells is common. *Blood* 115(15), 3146–3157.

Andreatta, M. and Nielsen, M. (2015). Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics* 32(4), 511–517.

Apps, R., Meng, Z., Del Prete, G. Q., Lifson, J. D., Zhou, M., and Carrington, M. (2015). Relative expression levels of the HLA class-I proteins in normal and HIV-infected cells. *The Journal of Immunology* 194(8), 3594–3600.

Arden, B., Clark, S. P., Kabelitz, D., and Mak, T. W. (1995). Human T-cell receptor variable gene segment families. *Immunogenetics* 42(6), 455–500.

Argaet, V. P., Schmidt, C. W., Burrows, S. R., Silins, S. L., Kurilla, M. G., Doolan, D. L., Suhrbier, A., Moss, D. J., Kieff, E., Sculley, T. B., and Misko, I. S. (1994). Dominant selection of an invariant T cell antigen receptor in response to persistent infection by Epstein-Barr virus. *Journal of Experimental Medicine* 180(6), 2335–2340.

Ascherio, A. and Munger, K. L. (2007). Environmental risk factors for multiple sclerosis. Part I: the role of infection. *Annals of Neurology* 61(4), 288–299.

Astellas Pharma (2018). Astellas and Vical Announce Top-Line Results for Phase 3 Trial of Cytomegalovirus Vaccine ASP0113 in Hematopoietic Stem Cell Transplant Recipients. URL: https://newsroom.astellas.us/2018-01-22-Astellas-and-Vical-Announce-Top-Line-Results-for-Phase-3-Trial-of-Cytomegalovirus-Vaccine-ASP0113-in-Hematopoietic-Stem-Cell-Transplant-Recipients.

Attaf, M., Malik, A., Severinsen, M. C., Roider, J., Ogongo, P., Buus, S., Ndung'u, T., Leslie, A., Kløverpris, H. N., Matthews, P. C., Sewell, A. K., and Goulder, P. (2018). Major TCR Repertoire Perturbation by Immunodominant HLA-B*44:03-Restricted CMV-Specific T Cells. *Frontiers in Immunology* 9, 2539.

Baroco, A. L. and Oldfield, E. C. (2008). Gastrointestinal cytomegalovirus disease in the immunocompromised patient. *Current Gastroenterology Reports* 10(4), 409–416.

Beck, S. and Barrell, B. G. (1988). Human cytomegalovirus encodes a glycoprotein homologous to MHC class-I antigens. *Nature* 331(6153), 269.

Beltran, P. M. J. and Cristea, I. M. (2014). The life cycle and pathogenesis of human cytomegalovirus infection: lessons from proteomics. *Expert Review of Proteomics* 11(6), 697–711.

Belzile, J.-P., Stark, T. J., Yeo, G. W., and Spector, D. H. (2014). Human cytomegalovirus infection of human embryonic stem cell-derived primitive neural stem cells is restricted at several steps but leads to the persistence of viral DNA. *Journal of Virology* 88(8), 4021–4039.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000). The protein data bank. *Nucleic Acids Research* 28(1), 235–242.

Bernstein, D. I., Reap, E. A., Katen, K., Watson, A., Smith, K., Norberg, P., Olmsted, R. A., Hoeper, A., Morris, J., Negri, S., Maughan, M. F., and Chulay, J. D. (2009). Randomized, double-blind, Phase 1 trial of an alphavirus replicon vaccine for cytomegalovirus in CMV seronegative adult volunteers. *Vaccine* 28(2), 484–493.

Bernstein, D. I., Munoz, F. M., Callahan, S. T., Rupp, R., Wootton, S. H., Edwards, K. M., Turley, C. B., Stanberry, L. R., Patel, S. M., Mcneal, M. M., Pichon, S., Amegashie, C., and Bellamy, A. R. (2016). Safety and efficacy of a cytomegalovirus glycoprotein B (gB) vaccine in adolescent girls: a randomized clinical trial. *Vaccine* 34(3), 313–319.

Bianconi, E., Piovesan, A., Facchin, F., Beraudi, A., Casadei, R., Frabetti, F., Vitale, L., Pelleri, M. C., Tassani, S., Piva, F., Perez-Amodio, S., Strippoli, P., and Canaider, S. (2013). An estimation of the number of cells in the human body. *Annals of Human Biology* 40(6), 463–471.

Biron, C. A., Byron, K. S., and Sullivan, J. L. (1989). Severe herpesvirus infections in an adolescent without natural killer cells. *New England Journal of Medicine* 320(26), 1731–1735.

Blum, J. S., Wearsch, P. A., and Cresswell, P. (2013). Pathways of antigen processing. *Annual Review of Immunology* 31, 443–473.

Boeckh, M. and Geballe, A. P. (2011). Cytomegalovirus: pathogen, paradigm, and puzzle. *The Journal of Clinical Investigation* 121(5), 1673–1680.

Boeckh, M., Gooley, T. A., Myerson, D., Cunningham, T., Schoch, G., and Bowden, R. A. (1996). Cytomegalovirus pp65 antigenemia-guided early treatment with ganciclovir versus ganciclovir at engraftment after allogeneic marrow transplantation: a randomized double-blind study. *Blood* 88(10), 4063–4071.

Bohl, D. L., Storch, G. A., Ryschkewitsch, C., Gaudreault-Keener, M., Schnitzler, M. A., Major, E. O., and Brennan, D. C. (2005). Donor origin of BK virus in renal transplantation and role of HLA C7 in susceptibility to sustained BK viremia. *American Journal of Transplantation* 5(9), 2213–2221.

Bolotin, D. A., Poslavsky, S., Mitrophanov, I., Shugay, M., Mamedov, I. Z., Putintseva, E. V., and Chudakov, D. M. (2015). MiXCR: software for comprehensive adaptive immunity profiling. *Nature Methods* 12(5), 380.

Boppana, S. B. and Britt, W. J. (1995). Antiviral antibody responses and intrauterine transmission after primary maternal cytomegalovirus infection. *The Journal of Infectious Diseases* 171(5), 1115–1121.

Bradley, A. J., Lurain, N. S., Ghazal, P., Trivedi, U., Cunningham, C., Baluchova, K., Gatherer, D., Wilkinson, G. W. G., Dargan, D. J., and Davison, A. J. (2009). High-throughput sequence analysis of variants of human cytomegalovirus strains Towne and AD169. *Journal of General Virology* 90(Pt 10), 2375.

Braendstrup, P., Justesen, S., Østerbye, T., Nielsen, L. L. B., Mallone, R., Vindeløv, L., Stryhn, A., and Buus, S. (2013). MHC class II tetramers made from isolated recombinant α and β chains refolded with affinity-tagged peptides. *PLOS ONE* 8(9), e73648.

Braendstrup, P., Mortensen, B. K., Justesen, S., Østerby, T., Rasmussen, M., Hansen, A. M., Christiansen, C. B., Hansen, M. B., Nielsen, M., Vindeløv, L., Buus, S., and Stryhn, A. (2014). Identification and HLA-tetramer-validation of human CD4+ and CD8+ T cell responses against HCMV proteins IE1 and IE2. *PLOS ONE* 9(4), e94892.

Brennan, R. M., Petersen, J., Neller, M. A., Miles, J. J., Burrows, J. M., Smith, C., McCluskey, J., Khanna, R., Rossjohn, J., and Burrows, S. R. (2012). The impact of a large and frequent deletion in the human TCRβ locus on antiviral immunity. *The Journal of Immunology* 188(6), 2742–2748.

Britt, W. J., Vugler, L., Butfiloski, E. J., and Stephens, E. B. (1990). Cell surface expression of human cytomegalovirus (HCMV) gp55-116 (gB): use of HCMV-recombinant vaccinia virus-infected cells in analysis of the human neutralizing antibody response. *Journal of Virology* 64(3), 1079–1085.

Bunde, T., Kirchner, A., Hoffmeister, B., Habedank, D., Hetzer, R., Cherepnev, G., Proesch, S., Reinke, P., Volk, H.-D., Lehmkuhl, H., and Kern, F. (2005). Protection from cytomegalovirus after transplantation is correlated with immediate early 1–specific CD8 T cells. *Journal of Experimental Medicine* 201(7), 1031–1036.

Burrows, S. R., Khanna, R., Burrows, J. M., and Moss, D. J. (1994). An alloresponse in humans is dominated by cytotoxic T lymphocytes (CTL) cross-reactive with a single Epstein-Barr virus CTL epitope: implications for graft-versus-host disease. *Journal of Experimental Medicine* 179(4), 1155–1161.

Burrows, S. R., Silins, S. L., Moss, D. J., Khanna, R., Misko, I. S., and Argaet, V. P. (1995). T cell receptor repertoire for a viral epitope in humans is diversified by tolerance to a background major histocompatibility complex antigen. *Journal of Experimental Medicine* 182(6), 1703–1715.

Burrows, S. R., Elkington, R. A., Miles, J. J., Green, K. J., Walker, S., Haryana, S. M., Moss, D. J., Dunckley, H., Burrows, J. M., and Khanna, R. (2003). Promiscuous CTL recognition of viral epitopes on multiple human leukocyte antigens: biological validation of the proposed HLA A24 supertype. *The Journal of Immunology* 171(3), 1407–1412.

Cai, X., Lu, S., Zhang, Z., Gonzalez, C. M., Damania, B., and Cullen, B. R. (2005). Kaposi's sarcoma-associated herpesvirus expresses an array of viral microRNAs in latently infected cells. *Proceedings of the National Academy of Sciences* 102(15), 5570–5575.

Camara, R. de la (2016). CMV in hematopoietic stem cell transplantation. *Mediterranean Journal of Hematology and Infectious Diseases* 8(1), e2016031.

Cao, K., Hollenbach, J., Shi, X., Shi, W., Chopek, M., and Fernández-Viña, M. A. (2001). Analysis of the frequencies of HLA-A, B, and C alleles and haplotypes in the five major ethnic groups of the United States reveals high levels of diversity in these loci and contrasting distribution patterns in these populations. *Human Immunology* 62(9), 1009–1030.

Challoner, P. B., Smith, K. T., Parker, J. D., MacLeod, D. L., Coulter, S. N., Rose, T. M., Schultz, E. R., Bennett, J. L., Garber, R. L., and Chang, M. (1995). Plaque-associated expression of human herpesvirus 6 in multiple sclerosis. *Proceedings of the National Academy of Sciences* 92(16), 7440–7444.

Charoudeh, H. N., Schmied, L., Gonzalez, A., Terszowski, G., Czaja, K., Schmitter, K., Infanti, L., Buser, A., and Stern, M. (2012). Quantity of HLA-C surface expression and licensing of KIR2DL+ natural killer cells. *Immunogenetics* 64(10), 739–745.

Chee, M. S., Bankier, A. T., Beck, S., Bohni, R., Brown, C. M., Cerny, R., Horsnell, T., Hutchison, C. A., Kouzarides, T., Martignetti, J. A., Preddie, E., Satchwell, S. C., Tomlinson, P., Weston, K. M., and Barrell, B. G. (1990). Analysis of the protein-coding content of the sequence of human cytomegalovirus strain AD169. *Cytomegaloviruses*. Ed. by J. K. McDougall. Springer, 125–169.

Chemaly, R. F., Ullmann, A. J., Stoelben, S., Richard, M. P., Bornhäuser, M., Groth, C., Einsele, H., Silverman, M., Mullane, K. M., Brown, J., Nowak, H., Kölling, K., Stobernack, H. P., Lischka, P., Zimmermann, H., Rübsamen-Schaeff, H., Champlin, R. E., and Ehninger, G. (2014). Letermovir for cytomegalovirus prophylaxis in hematopoietic-cell transplantation. *New England Journal of Medicine* 370(19), 1781–1789.

Chen, G., Yang, X., Ko, A., Sun, X., Gao, M., Zhang, Y., Shi, A., Mariuzza, R. A., and Weng, N.-p. (2017). Sequence and structural analyses reveal distinct and highly diverse human CD8+ TCR repertoires to immunodominant viral antigens. *Cell Reports* 19(3), 569–583.

Chen, K., Cheng, M. P., Hammond, S. P., Einsele, H., and Marty, F. M. (2018). Antiviral prophylaxis for cytomegalovirus infection in allogeneic hematopoietic cell transplantation. *Blood Advances* 2(16), 2159–2175.

Chou, S. (2015). Rapid in vitro evolution of human cytomegalovirus UL56 mutations that confer letermovir resistance. *Antimicrobial Agents and Chemotherapy* 59(10), 6588–6593.

Cirac, A., Stützle, S., Dieckmeyer, M., Adhikary, D., Moosmann, A., Körber, N., Bauer, T., Witter, K., Delecluse, H.-J., Behrends, U., and Mautner, J. (2018). Epstein–Barr virus strain heterogeneity impairs human T-cell immunity. *Cancer Immunology, Immunotherapy* 67(4), 663–674.

Clute, S. C., Watkin, L. B., Cornberg, M., Naumov, Y. N., Sullivan, J. L., Luzuriaga, K., Welsh, R. M., and Selin, L. K. (2005). Cross-reactive influenza virus–specific CD8+ T cells contribute to lymphoproliferation in Epstein-Barr virus–associated infectious mononucleosis. *The Journal of Clinical Investigation* 115(12), 3602–3612.

Cobbold, M., Khan, N., Pourgheysari, B., Tauro, S., McDonald, D., Osman, H., Assenmacher, M., Billingham, L., Steward, C., Crawley, C., Olavarria, E., Goldman, J., Chakraverty, R., Mahendra, P., Craddock, C., and Moss, P. A. H. (2005). Adoptive transfer of cytomegalovirus-specific CTL to stem cell transplant patients after selection by HLA–peptide tetramers. *Journal of Experimental Medicine* 202(3), 379–386.

Cole, D. K., Fuller, A., Dolton, G., Zervoudi, E., Legut, M., Miles, K., Blanchfield, L., Madura, F., Holland, C. J., Bulek, A. M., Bridgeman, J. S., Miles, J. J., Schauenburg, A. J. A., Beck, K., Evavold, B. D., Rizkallah, P. J., and Sewell, A. K. (2017). Dual molecular mechanisms govern escape at immunodominant HLA A2-restricted HIV epitope. *Frontiers in Immunology* 8, 1503.

Colonna, M., Borsellino, G., Falco, M., Ferrara, G. B., and Strominger, J. L. (1993). HLA-C is the inhibitory ligand that determines dominant resistance to lysis by NK1-and NK2-specific natural killer cells. *Proceedings of the National Academy of Sciences* 90(24), 12000–12004.

Compston, A. and Coles, A. (2008). Multiple sclerosis. *The Lancet* 372(9648), 1502–1517.

Compston, D. A. S., Batchelor, J. R., and McDonald, W. I. (1976). B-lymphocyte alloantigens associated with multiple sclerosis. *The Lancet* 308(7998), 1261–1265.

Compton, T., Nowlin, D. M., and Cooper, N. R. (1993). Initiation of human cytomegalovirus infection requires initial interaction with cell surface heparan sulfate. *Virology* 193(2), 834–841.

Crough, T. and Khanna, R. (2009). Immunobiology of human cytomegalovirus: from bench to bedside. *Clinical Microbiology Reviews* 22(1), 76–98.

Crough, T., Burrows, J. M., Fazou, C., Walker, S., Davenport, M. P., and Khanna, R. (2005). Contemporaneous fluctuations in T cell responses to persistent herpes virus infections. *European Journal of Immunology* 35(1), 139–149.

Cui, X., Adler, S. P., Davison, A. J., Smith, L., Habib, E.-S. E., and McVoy, M. A. (2012). Bacterial artificial chromosome clones of viruses comprising the towne cytomegalovirus vaccine. *Journal of Biomedicine and Biotechnology* 2012, 428498.

Cwynarski, K., Ainsworth, J., Cobbold, M., Wagner, S., Mahendra, P., Apperley, J., Goldman, J., Craddock, C., and Moss, P. A. H. (2001). Direct visualization of cytomegalovirus-specific T-cell reconstitution after allogeneic stem cell transplantation. *Blood* 97(5), 1232–1240.

Dash, P., Fiore-Gartland, A. J., Hertz, T., Wang, G. C., Sharma, S., Souquette, A., Crawford, J. C., Clemens, E. B., Nguyen, T. H. O., Kedzierska, K., La Gruta, N. L., Bradley, P., and Thomas, P. G. (2017). Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature* 547(7661), 89.

DeAngelis, M. M., Wang, D. G., and Hawkins, T. L. (1995). Solid-phase reversible immobilization for the isolation of PCR products. *Nucleic Acids Research* 23(22), 4742.

DeWitt, W. S., Smith, A., Schoch, G., Hansen, J. A., Matsen Frederick A, I. V., and Bradley, P. (2018). Human T cell receptor occurrence patterns encode immune history, genetic background, and receptor specificity. *eLife* 7. Ed. by A. M. Walczak, A. K. Chakraborty, Y. Elhanati, and B. Gerritsen, e38358.

Diamond, D. J., York, J., Sun, J.-Y., Wright, C. L., and Forman, S. J. (1997). Development of a candidate HLA A*0201 restricted peptide-based vaccine against human cytomegalovirus infection. *Blood* 90(5), 1751–1767.

Diamond, D. J., La Rosa, C., Chiuppesi, F., Contreras, H., Dadwal, S., Wussow, F., Bautista, S., Nakamura, R., and Zaia, J. A. (2018). A fifty-year odyssey: prospects for a cytomegalovirus vaccine in transplant and congenital infection. *Expert Review of Vaccines* 17(10), 889–911.

Dolan, A., Cunningham, C., Hector, R. D., Hassan-Walker, A. F., Lee, L., Addison, C., Dargan, D. J., McGeoch, D. J., Gatherer, D., Emery, V. C., Griffiths, P. D., Sinzger, C., McSharry, B. P., Wilkinson, G. W. G., and Davison, A. J. (2004). Genetic content of wild-type human cytomegalovirus. *Journal of General Virology* 85(5), 1301–1312.

Dong, L., Li, P., Oenema, T., McClurkan, C. L., and Koelle, D. M. (2010). Public TCR Use by Herpes Simplex Virus-2–Specific Human CD8 CTLs. *The Journal of Immunology* 184(6), 3063–3071.

Dupont, L. and Reeves, M. B. (2016). Cytomegalovirus latency and reactivation: recent insights into an age old problem. *Reviews in Medical Virology* 26(2), 75–89.

Dziubianau, M., Hecht, J., Kuchenbecker, L., Sattler, A., Stervbo, U., Rödelsperger, C., Nickel, P., Neumann, A. U., Robinson, P. N., Mundlos, S., Volk, H. D., Thiel, A., Reinke, P., and Babel, N. (2013). TCR repertoire analysis by next generation sequencing allows complex differential diagnosis of T cell–related pathology. *American Journal of Transplantation* 13(11), 2842–2854.

Einsele, H., Roosnek, E., Rufer, N., Sinzger, C., Riegler, S., Löffler, J., Grigoleit, U., Moris, A., Rammensee, H.-G., Kanz, L., Kleihauer, A., Frank, F., Jahn, G., and Hebart, H. (2002). Infusion of cytomegalovirus (CMV)–specific T cells for the treatment of CMV infection not responding to antiviral chemotherapy. *Blood* 99(11), 3916–3922.

Elkington, R., Walker, S., Crough, T., Menzies, M., Tellam, J., Bharadwaj, M., and Khanna, R. (2003). Ex vivo profiling of CD8+-T-cell responses to human cytomegalovirus reveals broad and multispecific reactivities in healthy virus carriers. *Journal of Virology* 77(9), 5226–5240.

Emerson, R. O., DeWitt, W. S., Vignali, M., Gravley, J., Hu, J. K., Osborne, E. J., Desmarais, C., Klinger, M., Carlson, C., Hansen, J. A., Rieder, M., and Robins, H. S. (2017). Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. *Nature Genetics* 49(5), 659.

Farrell, R. A., Antony, D., Wall, G. R., Clark, D. A., Fisniku, L., Swanton, J., Khaleeli, Z., Schmierer, K., Miller, D. H., and Giovannoni, G. (2009). Humoral immune response to EBV in multiple sclerosis is associated with disease activity on MRI. *Neurology* 73(1), 32–38.

Fazou, C., Yang, H., McMichael, A. J., and Callan, M. F. C. (2001). Epitope specificity of clonally expanded populations of CD8+ T cells found within the joints of patients with inflammatory arthritis. *Arthritis & Rheumatism* 44(9), 2038–2045.

Feuchtinger, T., Lang, P., Hamprecht, K., Schumm, M., Greil, J., Jahn, G., Niethammer, D., and Einsele, H. (2004). Isolation and expansion of human adenovirus–specific CD4+ and CD8+ T cells according to IFN-γ secretion for adjuvant immunotherapy. *Experimental Hematology* 32(3), 282–289.

Feuchtinger, T., Matthes-Martin, S., Richard, C., Lion, T., Fuhrer, M., Hamprecht, K., Handgretinger, R., Peters, C., Schuster, F. R., Beck, R., Schumm, M., Lotfi, R., Jahn, G., and Lang, P. (2006). Safe adoptive transfer of virus-specific T-cell immunity for the treatment of systemic adenovirus infection after allogeneic stem cell transplantation. *British Journal of Haematology* 134(1), 64–76.

Fish, K. N., Soderberg-Naucler, C., Mills, L. K., Stenglein, S., and Nelson, J. A. (1998). Human cytomegalovirus persistently infects aortic endothelial cells. *Journal of Virology* 72(7), 5661–5668.

Full, F., Lehner, M., Thonn, V., Goetz, G., Scholz, B., Kaufmann, K. B., Mach, M., Abken, H., Holter, W., and Ensser, A. (2010). T cells engineered with a cytomegalovirus-specific chimeric immunoreceptor. *Journal of Virology* 84(8), 4083–4088.

Garcia, K. C., Degano, M., Stanfield, R. L., Brunmark, A., Jackson, M. R., Peterson, P. A., Teyton, L., and Wilson, I. A. (1996). An $\alpha\beta$ T cell receptor structure at 2.5 Å and its orientation in the TCR-MHC complex. *Science* 274(5285), 209–219.

García-Sastre, A. and Biron, C. A. (2006). Type 1 Interferons and the Virus-Host Relationship: A Lesson in Détente. *Science* 312(5775), 879–882.

Gardiner, C. M. (2008). Killer cell immunoglobulin-like receptors on NK cells: the how, where and why. *International Journal of Immunogenetics* 35(1), 1–8.

Gardner, T. J. and Tortorella, D. (2016). Virion glycoprotein-mediated immune evasion by human cytomegalovirus: a sticky virus makes a slick getaway. *Microbiology and Molecular Biology Reviews* 80(3), 663–677.

Garrone, P., Neidhardt, E.-M., Garcia, E., Galibert, L., Van Kooten, C., and Banchereau, J. (1995). Fas ligation induces apoptosis of CD40-activated human B lymphocytes. *Journal of Experimental Medicine* 182(5), 1265–1273.

Gary, R., Aigner, M., Moi, S., Schaffer, S., Gottmann, A., Maas, S., Zimmermann, R., Zingsem, J., Strobel, J., Mackensen, A., Mautner, J., Moosmann, A., and Gerbitz, A. (2018). Clinical-grade generation of peptide-stimulated CMV/EBV-specific T cells from G-CSF mobilized stem cell grafts. *Journal of Translational Medicine* 16(1), 124.

Gavin, M. A., Gilbert, M. J., Riddell, S. R., Greenberg, P. D., and Bevan, M. J. (1993). Alkali hydrolysis of recombinant proteins allows for the rapid identification of class I MHC-restricted CTL epitopes. *The Journal of Immunology* 151(8), 3971–3980.

Gerna, G., Lilleri, D., Fornara, C., Comolli, G., Lozza, L., Campana, C., Pellegrini, C., Meloni, F., and Rampino, T. (2006). Monitoring of human cytomegalovirus-specific CD4+ and CD8+ T-cell immunity in patients receiving solid organ transplantation. *American Journal of Transplantation* 6(10), 2356–2364.

Gibson, L., Dooley, S., Trzmielina, S., Somasundaran, M., Fisher, D., Revello, M. G., and Luzuriaga, K. (2007). Cytomegalovirus (CMV) IE1-and pp65-specific CD8+ T cell responses broaden over time after primary CMV infection in infants. *The Journal of Infectious Diseases* 195(12), 1789–1798.

Gillespie, G. M. A., Wills, M. R., Appay, V., O'Callaghan, C., Murphy, M., Smith, N., Sissons, P., Rowland-Jones, S., Bell, J. I., and Moss, P. A. H. (2000). Functional heterogeneity and high frequencies of cytomegalovirus-specific CD8+ T lymphocytes in healthy seropositive donors. *Journal of Virology* 74(17), 8140–8150.

Glanville, J., Huang, H., Nau, A., Hatton, O., Wagar, L. E., Rubelt, F., Ji, X., Han, A., Krams, S. M., Pettus, C., Haas, N., Lindestam Arlehamn, C. S., Sette, A., Boyd, S. D., Scriba, T. J., Martinez, O. M., and Davis, M. M. (2017). Identifying specificity groups in the T cell receptor repertoire. *Nature* 547(7661), 94.

Glenn, T. C. (2011). Field guide to next-generation DNA sequencers. *Molecular Ecology Resources* 11(5), 759–769.

Goodrum, F. (2016). Human cytomegalovirus latency: approaching the Gordian knot. *Annual Review of Virology* 3, 333–357.

Gragert, L., Madbouly, A., Freeman, J., and Maiers, M. (2013). Six-locus high resolution HLA haplotype frequencies derived from mixed-resolution DNA typing for the entire US donor registry. *Human Immunology* 74(10), 1313–1320.

Grant, E. J., Josephs, T. M., Loh, L., Clemens, E. B., Sant, S., Bharadwaj, M., Chen, W., Rossjohn, J., Gras, S., and Kedzierska, K. (2018). Broad CD8+ T cell cross-recognition of distinct influenza A strains in humans. *Nature Communications* 9(1), 5427.

Gras, S., Kedzierski, L., Valkenburg, S. A., Laurie, K., Liu, Y. C., Denholm, J. T., Richards, M. J., Rimmelzwaan, G. F., Kelso, A., Doherty, P. C., Turner, S. J., Rossjohn, J., and Kedzierska, K. (2010). Cross-reactive CD8+ T-cell immunity between the pandemic H1N1-2009 and H1N1-1918 influenza A viruses. *Proceedings of the National Academy of Sciences* 107(28), 12599–12604.

Gratama, J. W., Esser, J. W. J. van, Lamers, C. H. J., Tournay, C., Löwenberg, B., Bolhuis, R. L. H., and Cornelissen, J. J. (2001). Tetramer-based quantification of cytomegalovirus (CMV)–specific CD8+ T lymphocytes in T-cell–depleted stem cell grafts and after transplantation may identify patients at risk for progressive CMV infection. *Blood* 98(5), 1358–1364.

Gratama, J. W., Brooimans, R. A., Holt, B. van der, Sintnicolaas, K., Doornum, G. van, Niesters, H. G., Löwenberg, B., and Cornelissen, J. J. (2008). Monitoring cytomegalovirus IE-1 and pp65-specific CD4+ and CD8+ T-cell responses after allogeneic stem cell transplantation may identify patients at risk for recurrent CMV reactivations. *Cytometry Part B: Clinical Cytometry* 74(4), 211–220.

Grey, F., Meyers, H., White, E. A., Spector, D. H., and Nelson, J. (2007). A human cytomegalovirus-encoded microRNA regulates expression of multiple viral genes involved in replication. *PLOS Pathogens* 3(11), e163.

Hahn, G., Revello, M. G., Patrone, M., Percivalle, E., Campanini, G., Sarasini, A., Wagner, M., Gallina, A., Milanesi, G., Koszinowski, U., Baldanti, F., and Gerna, G. (2004). Human cytomegalovirus UL131-128 genes are indispensable for virus growth in endothelial cells and virus transfer to leukocytes. *Journal of Virology* 78(18), 10023–10033.

Hebart, H., Daginik, S., Stevanovic, S., Grigoleit, U., Dobler, A., Baur, M., Rauser, G., Sinzger, C., Jahn, G., Loeffler, J., Kanz, L., Rammensee, H.-G., and Einsele, H. (2002). Sensitive detection of human cytomegalovirus peptide–specific cytotoxic T-lymphocyte responses by interferon-γ–enzyme-linked immunospot assay and flow cytometry in healthy individuals and in patients after allogeneic stem cell transplantation. *Blood* 99(10), 3830–3837.

Hegde, N. R., Dunn, C., Lewinsohn, D. M., Jarvis, M. A., Nelson, J. A., and Johnson, D. C. (2005). Endogenous human cytomegalovirus gB is presented efficiently by MHC class II molecules to CD4+ CTL. *Journal of Experimental Medicine* 202(8), 1109–1119.

Heineman, T. C., Schleiss, M., Bernstein, D. I., Spaete, R. R., Yan, L., Duke, G., Prichard, M., Wang, Z., Yan, Q., Sharp, M. A., Klein, N., Arvin, A. M., and Kemble, G. (2006). A phase 1 study of 4 live, recombinant human cytomegalovirus Towne/Toledo chimeric vaccines. *The Journal of Infectious Diseases* 193(10), 1350–1360.

Hengel, H., Brune, W., and Koszinowski, U. H. (1998). Immune evasion by cytomegalovirus–survival strategies of a highly adapted opportunist. *Trends in Microbiology* 6(5), 190–197.

Hilton, H. G. and Parham, P. (2017). Missing or altered self: human NK cell receptors that recognize HLA-C. *Immunogenetics* 69(8-9), 567–579.

Hirayasu, K., Ohashi, J., Kashiwase, K., Hananantachai, H., Naka, I., Ogawa, A., Takanashi, M., Satake, M., Nakajima, K., Parham, P., Arase, H., Tokunaga, K., Patarapotikul, J., and Yabe, T. (2012). Significant association of KIR2DL3-HLA-C1 combination with cerebral malaria and implications for co-evolution of KIR and HLA. *PLOS Pathog* 8(3), e1002565.

Höllsberg, P. (2002). Contribution of HLA Class I Allele Expression to CD8+ T-Cell Responses against Epstein–Barr Virus. *Scandinavian Journal of Immunology* 55(2), 189–195.

Holtappels, R., Pahl-Seibert, M.-F., Thomas, D., and Reddehase, M. J. (2000). Enrichment of immediate-early 1 (m123/pp89) peptide-specific CD8 T cells in a pulmonary CD62Llo memory-effector cell pool during latent murine cytomegalovirus infection of the lungs. *Journal of Virology* 74(24), 11495–11503.

Hosie, L., Pachnio, A., Zuo, J., Pearce, H., Riddell, S., and Moss, P. (2017). Cytomegalovirus-Specific T Cells Restricted by HLA-Cw* 0702 Increase Markedly with Age and Dominate the CD8+ T-Cell Repertoire in Older People. *Frontiers in Immunology* 8, 1776.

Hoz, R. E. De la, Stephens, G., and Sherlock, C. (2002). Diagnosis and treatment approaches of CMV infections in adult patients. *Journal of Clinical Virology* 25, 1–12.

Huber, M. T. and Compton, T. (1998). The human cytomegalovirus UL74 gene encodes the third component of the glycoprotein H-glycoprotein L-containing envelope complex. *Journal of Virology* 72(10), 8191–8197.

Huppa, J. B. and Davis, M. M. (2003). T-cell-antigen recognition and the immunological synapse. *Nature Reviews Immunology* 3(12), 973.

Huth, A., Liang, X., Krebs, S., Blum, H., and Moosmann, A. (2019). Antigen-Specific TCR Signatures of Cytomegalovirus Infection. *The Journal of Immunology* 202(3), 979–990.

Icheva, V., Kayser, S., Wolff, D., Tuve, S., Kyzirakos, C., Bethge, W., Greil, J., Albert, M. H., Schwinger, W., Nathrath, M., Schumm, M., Stevanovic, S., Handgretinger, R., Lang, P., and Feuchtinger, T. (2013). Adoptive transfer of epstein-barr virus (EBV) nuclear antigen 1-specific t cells as treatment for EBV reactivation and lymphoproliferative disorders after allogeneic stem-cell transplantation. *Journal of Clinical Oncology* 31(1), 39–48.

Isaacson, M. K. and Compton, T. (2009). Human cytomegalovirus glycoprotein B is required for virus entry and cell-to-cell spread but not for virion attachment, assembly, or egress. *Journal of Virology* 83(8), 3891–3903.

Ishizuka, J., Stewart-Jones, G. B., Merwe, A. van der, Bell, J. I., McMichael, A. J., and Jones, E. Y. (2008). The structural dynamics and energetics of an immunodominant T cell receptor are programmed by its Vβ domain. *Immunity* 28(2), 171–182.

Jackson, S. E., Mason, G. M., Okecha, G., Sissons, J. G. P., and Wills, M. R. (2014). Diverse specificities, phenotypes, and antiviral activities of cytomegalovirus-specific CD8+ T cells. *Journal of Virology* 88(18), 10894–10908.

Janbazian, L., Price, D. A., Canderan, G., Filali-Mouhim, A., Asher, T. E., Ambrozak, D. R., Scheinberg, P., Boulassel, M. R., Routy, J.-P., Koup, R. A., Douek, D. C., Sekaly, R.-P., and Trautmann, L. (2012). Clonotype and repertoire changes drive the functional improvement of HIV-specific CD8 T cell populations under conditions of limited antigenic stimulation. *The Journal of Immunology* 188(3), 1156–1167.

Jenkins, M. K., Chu, H. H., McLachlan, J. B., and Moon, J. J. (2009). On the composition of the preimmune repertoire of T cells specific for peptide–major histocompatibility complex ligands. *Annual Review of Immunology* 28, 275–294.

Jilek, S., Schluep, M., Meylan, P., Vingerhoets, F., Guignard, L., Monney, A., Kleeberg, J., Le Goff, G., Pantaleo, G., and Du Pasquier, R. A. (2008). Strong EBV-specific CD8+ T-cell response in patients with early multiple sclerosis. *Brain* 131(7), 1712–1721.

Jonjić, S., Pavić, I., Polić, B., Crnković, I., Lucin, P., and Koszinowski, U. H. (1994). Antibodies are not essential for the resolution of primary cytomegalovirus infection but limit dissemination of recurrent virus. *Journal of Experimental Medicine* 179(5), 1713–1717.

Kaeuferle, T., Krauss, R., Blaeschke, F., Willier, S., and Feuchtinger, T. (2019). Strategies of adoptive T-cell transfer to treat refractory viral infections post allogeneic stem cell transplantation. *Journal of Hematology & Oncology* 12(1), 13.

Karrer, U., Sierro, S., Wagner, M., Oxenius, A., Hengel, H., Koszinowski, U. H., Phillips, R. E., and Klenerman, P. (2003). Memory inflation: continuous accumulation of antiviral CD8+ T cells over time. *The Journal of Immunology* 170(4), 2022–2029.

Kebschull, J. M. and Zador, A. M. (2015). Sources of PCR-induced distortions in high-throughput sequencing data sets. *Nucleic Acids Research* 43(21), e143.

Kern, F., Khatamzas, E., Surel, I., Frömmel, C., Reinke, P., Waldrop, S. L., Picker, L. J., and Volk, H.-D. (1999). Distribution of human CMV-specific memory T cells among the CD8pos. subsets defined by CD57, CD27, and CD45 isoforms. *European Journal of Immunology* 29(9), 2908–2915.

Khakoo, S. I., Thio, C. L., Martin, M. P., Brooks, C. R., Gao, X., Astemborski, J., Cheng, J., Goedert, J. J., Vlahov, D., Hilgartner, M., Cox, S., Little, A.-M., Alexander, G. J., Cramp, M. E., O'Brien, S. J., Rosenberg, W. M. C., Thomas, D. L., and Carrington, M. (2004). HLA and NK cell inhibitory receptor genes in resolving hepatitis C virus infection. *Science* 305(5685), 872–874.

Khan, N. (2007). The immunological burden of human cytomegalovirus infection. *Archivum Immunologiae et Therapiae Experimentalis* 55(5), 299.

Khan, N., Cobbold, M., Keenan, R., and Moss, P. A. H. (2002a). Comparative analysis of CD8+ T cell responses against human cytomegalovirus proteins pp65 and immediate early 1 shows similarities in precursor frequency, oligoclonality, and phenotype. *The Journal of Infectious Diseases* 185(8), 1025–1034.

Khan, N., Shariff, N., Cobbold, M., Bruton, R., Ainsworth, J. A., Sinclair, A. J., Nayak, L., and Moss, P. A. H. (2002b). Cytomegalovirus seropositivity drives the CD8 T cell repertoire toward greater clonality in healthy elderly individuals. *The Journal of Immunology* 169(4), 1984–1992.

Khan, N., Hislop, A., Gudgeon, N., Cobbold, M., Khanna, R., Nayak, L., Rickinson, A. B., and Moss, P. A. H. (2004). Herpesvirus-specific CD8 T cell immunity in old age: cytomegalovirus impairs the response to a coresident EBV infection. *The Journal of Immunology* 173(12), 7481–7489.

Khan, N., Best, D., Bruton, R., Nayak, L., Rickinson, A. B., and Moss, P. A. H. (2007). T cell recognition patterns of immunodominant cytomegalovirus antigens in primary and persistent infection. *The Journal of Immunology* 178(7), 4455–4465.

Kharfan-Dabaja, M. A., Boeckh, M., Wilck, M. B., Langston, A. A., Chu, A. H., Wloch, M. K., Guterwill, D. F., Smith, L. R., Rolland, A. P., and Kenney, R. T. (2012). A novel therapeutic cytomegalovirus DNA vaccine in allogeneic haemopoietic stem-cell transplantation: a randomised, double-blind, placebo-controlled, phase 2 trial. *The Lancet Infectious Diseases* 12(4), 290–299.

Kim, S., Lee, S., Shin, J., Kim, Y., Evnouchidou, I., Kim, D., Kim, Y.-K., Kim, Y.-E., Ahn, J.-H., Riddell, S. R., Stratikos, E., Kim, V. N., and Ahn, K. (2011). Human cytomegalovirus microRNA miR-US4-1 inhibits CD8+ T cell responses by targeting the aminopeptidase ERAP1. *Nature Immunology* 12(10), 984.

Kirchner, A., Hoffmeister, B., Cherepnev-G, G., Fuhrmann, S., Streitz, M., Lachmann, R., Bunde, T., Meij, P., Schönemann, C., Hetzer, R., Lehmkuhl, H. B., Volkmer-Engert, R., Volk, H.-D., Gratama, J. W., and Kern, F. (2008). Dissection of the CMV specific T-cell response is required for optimized cardiac transplant monitoring. *Journal of Medical Virology* 80(9), 1604–1614.

Klarenbeek, P. L., Remmerswaal, E. B. M., Ten Berge, I. J. M., Doorenspleet, M. E., Van Schaik, B. D. C., Esveldt, R. E. E., Koch, S. D., Ten Brinke, A., Van Kampen, A. H. C., Bemelman, F. J., Tak, P. P., Baas, F., Vries, N. de, and Lier, R. A. W. van (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), e1002889.

Klenerman, P. and Oxenius, A. (2016). T cell responses to cytomegalovirus. *Nature Reviews Immunology* 16(6), 367.

Klinger, M., Kong, K., Moorhead, M., Weng, L., Zheng, J., and Faham, M. (2013). Combining next-generation sequencing and immune assays: A novel method for identification of antigen-specific T cells. *PLOS ONE* 8(9), e74231.

Klinger, M., Pepin, F., Wilkins, J., Asbury, T., Wittkop, T., Zheng, J., Moorhead, M., and Faham, M. (2015). Multiplex identification of antigen-specific T cell receptors using a combination of immune assays and immune receptor sequencing. *PLOS ONE* 10(10), e0141561.

Knoll, B. M., Seiter, K., Phillips, A., and Soave, R. (2019). Breakthrough cytomegalovirus pneumonia in hematopoietic stem cell transplant recipient on letermovir prophylaxis. *Bone Marrow Transplantation* 54(6), 911–912.

Kondo, E., Akatsuka, Y., Kuzushima, K., Tsujimura, K., Asakura, S., Tajima, K., Kagami, Y., Kodera, Y., Tanimoto, M., Morishima, Y., and Takahashi, T. (2004). Identification of novel CTL epitopes of CMV-pp65 presented by a variety of HLA alleles. *Blood* 103(2), 630–638.

Koning, D., Costa, A. I., Hoof, I., Miles, J. J., Nanlohy, N. M., Ladell, K., Matthews, K. K., Venturi, V., Schellens, I. M. M., Borghans, J. A. M., Keşmir, C., Price, D. A., and Baarle, D. van (2013). CD8+ TCR repertoire formation is guided primarily by the peptide component of the antigenic complex. *The Journal of Immunology* 190(3), 931–939.

Kotton, C. N. (2013). CMV: prevention, diagnosis and therapy. *American Journal of Transplantation* 13(s3), 24–40.

Koutsakos, M., Illing, P. T., Nguyen, T. H. O., Mifsud, N. A., Crawford, J. C., Rizzetto, S., Eltahla, A. A., Clemens, E. B., Sant, S., Chua, B. Y., Wong, C. Y., Allen, E. K., Teng, D., Dash, P., Boyd, D. F., Grzelak, L., Zeng, W., Hurt, A. C., Barr, I., Rockman, S., Jackson, D. C., Kotsimbos, T. C., Cheng, A. C., Richards, M., Westall, G. P., Loudovaris, T., Mannering, S. I., Elliott, M., Tangye, S. G., Wakim, L. M., Rossjohn, J., Vijaykrishna, D., Luciani, F., Thomas, P. G., Gras, S., Purcell, A. W., and Kedzierska, K. (2019). Human CD8+ T cell cross-reactivity across influenza A, B and C viruses. *Nature Immunology* 20(5), 613–625.

Kulkarni, A. B., Morse, H. C., Bennink, J. R., Yewdell, J. W., and Murphy, B. R. (1993). Immunization of mice with vaccinia virus-M2 recombinant induces epitope-specific and cross-reactive Kd-restricted CD8+ cytotoxic T cells. *Journal of Virology* 67(7), 4086–4092.

Kuzushima, K., Hayashi, N., Kimura, H., and Tsurumi, T. (2001). Efficient identification of HLA-A* 2402–restricted cytomegalovirus-specific CD8+ T-cell epitopes by a computer algorithm and an enzyme-linked immunospot assay. *Blood* 98(6), 1872–1881.

Lacey, S. F., Martinez, J., Gallez-Hawkins, G., Thao, L., Longmate, J., Haq, W., Spielberger, R., Forman, S. J., Zaia, J. A., and Diamond, D. J. (2005). Simultaneous reconstitution of multiple cytomegalovirus-specific CD8+ cell populations with divergent functionality in hematopoietic stem-cell transplant recipients. *The Journal of Infectious Diseases* 191(6), 977–984.

Laydon, D. J., Bangham, C. R. M., and Asquith, B. (2015). Estimating T-cell repertoire diversity: limitations of classical estimators and a new approach. *Philosophical Transactions of the Royal Society B: Biological Sciences* 370(1675), 20140291.

Leen, A. M., Myers, G. D., Sili, U., Huls, M. H., Weiss, H., Leung, K. S., Carrum, G., Krance, R. A., Chang, C.-C., Molldrem, J. J., Gee, A. P., Brenner, M. K., Heslop, H. E., Rooney, C. M., and Bollard, C. M. (2006). Monoculture-derived T lymphocytes specific for multiple viruses expand and produce clinically relevant effects in immunocompromised individuals. *Nature Medicine* 12(10), 1160—1166.

Lefranc, M.-P. (2001). IMGT, the international ImMunoGeneTics database. *Nucleic Acids Research* 29(1), 207–209.

Lehner, P. J., Wang, E. C., Moss, P. A., Williams, S., Platt, K., Friedman, S. M., Bell, J. I., and Borysiewicz, L. K. (1995). Human HLA-A0201-restricted cytotoxic T lymphocyte recognition of influenza A is dominated by T cells bearing the V beta 17 gene segment. *Journal of Experimental Medicine* 181(1), 79–91.

Leisegang, M., Wilde, S., Spranger, S., Milosevic, S., Frankenberger, B., Uckert, W., and Schendel, D. J. (2010). MHC-restricted fratricide of human lymphocytes expressing survivin-specific transgenic T cell receptors. *The Journal of Clinical Investigation* 120(11), 3869–3877.

Lenzi, M., Frisoni, M., Mantovani, V., Ricci, P., Muratori, L., Francesconi, R., Cuccia, M., Ferri, S., and Bianchi, F. B. (1998). Haplotype HLA-B8-DR3 confers susceptibility to hepatitis C virus-related mixed cryoglobulinemia. *Blood* 91(6), 2062–2066.

Lim, A., Trautmann, L., Peyrat, M.-A., Couedel, C., Davodeau, F., Romagné, F., Kourilsky, P., and Bonneville, M. (2000). Frequent contribution of T cell clonotypes with public TCR features to the chronic response against a dominant EBV-derived epitope: application to direct detection of their molecular imprint on the human peripheral T cell repertoire. *The Journal of Immunology* 165(4), 2001–2011.

Lindau, P., Mukherjee, R., Gutschow, M. V., Vignali, M., Warren, E. H., Riddell, S. R., Makar, K. W., Turtle, C. J., and Robins, H. S. (2019). Cytomegalovirus exposure in the elderly does not reduce CD8 T cell repertoire diversity. *The Journal of Immunology* 202(2), 476–483.

Link, E. K., Brandmüller, C., Suezer, Y., Ameres, S., Volz, A., Moosmann, A., Sutter, G., and Lehmann, M. H. (2017). A synthetic human cytomegalovirus pp65-IE1 fusion antigen efficiently induces and expands virus specific T cells. *Vaccine* 35(38), 5131–5139.

Lischka, P., Michel, D., and Zimmermann, H. (2015). Characterization of cytomegalovirus breakthrough events in a phase 2 prophylaxis trial of letermovir (AIC246, MK 8228). *The Journal of Infectious Diseases* 213(1), 23–30.

Ljunggren, H.-G. and Kärre, K. (1990). In search of the 'missing self': MHC molecules and NK cell recognition. *Immunology Today* 11, 237–244.

Ljungman, P, De La Camara, R, Cordonnier, C, Einsele, H, Engelhard, D, Reusser, P, Styczynski, J, and Ward, K (2008). Management of CMV, HHV-6, HHV-7 and Kaposi-sarcoma herpesvirus (HHV-8) infections in patients with hematological malignancies and after SCT. *Bone Marrow Transplantation* 42(4), 227.

Loewendorf, A. and Benedict, C. A. (2010). Modulation of host innate and adaptive immune defenses by cytomegalovirus: timing is everything. *Journal of Internal Medicine* 267(5), 483–501.

Longmate, J., York, J., La Rosa, C., Krishnan, R., Zhang, M., Senitzer, D., and Diamond, D. J. (2001). Population coverage by HLA class-I restricted cytotoxic T-lymphocyte epitopes. *Immunogenetics* 52(3-4), 165–173.

Lünemann, J. D., Edwards, N., Muraro, P. A., Hayashi, S., Cohen, J. I., Münz, C., and Martin, R. (2006). Increased frequency and broadened specificity of latent EBV nuclear antigen-1-specific T cells in multiple sclerosis. *Brain* 129(6), 1493–1506.

Lythe, G., Callard, R. E., Hoare, R. L., and Molina-París, C. (2016). How many TCR clonotypes does a body maintain? *Journal of Theoretical Biology* 389, 214–224.

Macdonald, W. A., Chen, Z., Gras, S., Archbold, J. K., Tynan, F. E., Clements, C. S., Bharadwaj, M., Kjer-Nielsen, L., Saunders, P. M., Wilce, M. C. J., Crawford, F., Stadinsky, B., Jackson, D., Brooks, A. G., Purcell, A. W., Kappler, J. W., Burrows, S. R., Rossjohn, J., and McCluskey, J. (2009). T cell allorecognition via molecular mimicry. *Immunity* 31(6), 897–908.

Malissen, M., Trucy, J., Jouvin-Marche, E., Cazenave, P.-A., Scollay, R., and Malissen, B. (1992). Regulation of TCR α and β gene allelic exclusion during T-cell development. *Immunology Today* 13(8), 315–322.

Manley, T. J., Luy, L., Jones, T., Boeckh, M., Mutimer, H., and Riddell, S. R. (2004). Immune evasion proteins of human cytomegalovirus do not prevent a diverse CD8+ cytotoxic T-cell response in natural infection. *Blood* 104(4), 1075–1082.

Marshall, N. B. and Swain, S. L. (2011). Cytotoxic CD4 T cells in antiviral immunity. *BioMed Research International* 2011.

Marty, F. M., Ljungman, P., Chemaly, R. F., Maertens, J., Dadwal, S. S., Duarte, R. F., Haider, S., Ullmann, A. J., Katayama, Y., Brown, J., Mullane, K. M., Boeckh, M., Blumberg, E. A., Einsele, H., Snydman, D. R., Kanda, Y., DiNubile, M. J., Teal, V. L., Wan, H., Murata, Y., Kartsonis, N. A., Leavitt, R. Y., and Badshah, C. (2017). Letermovir prophylaxis for cytomegalovirus in hematopoietic-cell transplantation. *New England Journal of Medicine* 377(25), 2433–2444.

Masuoka, M., Yoshimuta, T., Hamada, M., Okamoto, M., Fumimori, T., Honda, J., Oizumi, K., and Itoh, K. (2001). Identification of the HLA-A24 peptide epitope within cytomegalovirus protein pp65 recognized by CMV-specific cytotoxic T lymphocytes. *Viral Immunology* 14(4), 369–377.

McVoy, M. A. (2013). Cytomegalovirus vaccines. *Clinical Infectious Diseases* 57(suppl_4), S196–S199.

Merkle, P. S., Irving, M., Hongjian, S., Ferber, M., Jørgensen, T. J. D., Scholten, K., Luescher, I., Coukos, G., Zoete, V., Cuendet, M. A., Michielin, O., and Rand, K. D. (2017). The T-Cell Receptor Can Bind to the Peptide-Bound Major Histocompatibility Complex and Uncomplexed β2-Microglobulin through Distinct Binding Sites. *Biochemistry* 56(30), 3945–3961.

Micklethwaite, K., Hansen, A., Foster, A., Snape, E., Antonenas, V., Sartor, M., Shaw, P., Bradstock, K., and Gottlieb, D. (2007). Ex vivo expansion and prophylactic infusion of CMV-pp65 peptide-specific cytotoxic T-lymphocytes following allogeneic hematopoietic stem cell transplantation (HSCT). *Biology of Blood and Marrow Transplantation* 13(6), 707–714.

Miconnet, I., Marrau, A., Farina, A., Taffé, P., Vigano, S., Harari, A., and Pantaleo, G. (2011). Large TCR diversity of virus-specific CD8 T cells provides the mechanistic basis for massive TCR renewal after antigen exposure. *The Journal of Immunology* 186(12), 7039–7049.

Middleton, D., Meenagh, A., Moscoso, J., and Arnaiz-Villena, A. (2008). Killer immunoglobulin receptor gene and allele frequencies in Caucasoid, Oriental and Black populations from different continents. *Tissue Antigens* 71(2), 105–113.

Mignot, E., Lin, X., Arrigoni, J., Macaubas, C., Olive, F., Hallmayer, J., Underhill, P., Guilleminault, C., Dement, W. C., and Grumet, F. C. (1994). DQB1*0602 and DQA1*0102 (DQ1) are better markers than DR2 for narcolepsy in Caucasian and black Americans. *Sleep* 17(suppl_8), S60–S67.

Miles, J. J., Elhassen, D., Borg, N. A., Silins, S. L., Tynan, F. E., Burrows, J. M., Purcell, A. W., Kjer-Nielsen, L., Rossjohn, J., Burrows, S. R., and McCluskey, J. (2005). CTL recognition of a bulged viral peptide involves biased TCR selection. *The Journal of Immunology* 175(6), 3826–3834.

Miles, J. J., Borg, N. A., Brennan, R. M., Tynan, F. E., Kjer-Nielsen, L., Silins, S. L., Bell, M. J., Burrows, J. M., McCluskey, J., Rossjohn, J., and Burrows, S. R. (2006). TCRα genes direct MHC restriction in the potent human T cell response to a class I-bound viral epitope. *The Journal of Immunology* 177(10), 6804–6814.

Miller, G. and Lipman, M. (1973). Release of infectious Epstein-Barr virus by transformed marmoset leukocytes. *Proceedings of the National Academy of Sciences* 70(1), 190–194.

Miller, G., Shope, T., Lisco, H., Stitt, D., and Lipman, M. (1972). Epstein-Barr virus: Transformation, cytopathic changes, and viral antigens in squirrel monkey and marmoset leukocytes. *Proceedings of the National Academy of Sciences* 69(2), 383–387.

Misko, I. S., Cross, S. M., Khanna, R., Elliott, S. L., Schmidt, C., Pye, S. J., and Silins, S. L. (1999). Cross-reactive recognition of viral, self, and bacterial peptide ligands by human class I-restricted cytotoxic T lymphocyte clonotypes: implications for molecular mimicry in autoimmune disease. *Proceedings of the National Academy of Sciences* 96(5), 2279–2284.

Miyama, T., Kawase, T., Kitaura, K., Chishaki, R., Shibata, M., Oshima, K., Hamana, H., Kishi, H., Muraguchi, A., Kuzushima, K., Saji, H., Shin-I, T., and Suzuki Ryuji & Ichinohe, T. (2017). Highly functional T-cell receptor repertoires are abundant in stem memory T cells and highly shared among individuals. *Scientific Reports* 7(1), 3663.

Moosmann, A., Khan, N., Cobbold, M., Zentz, C., Delecluse, H.-J., Hollweck, G., Hislop, A. D., Blake, N. W., Croom-Carter, D., Wollenberg, B., Moss, P. A. H., Zeidler, R., Rickinson, A. B., and Hammerschmidt, W. (2002). B cells immortalized by a mini–Epstein-Barr virus encoding a foreign antigen efficiently reactivate specific cytotoxic T cells. *Blood* 100(5), 1755–1764.

Murphy, E., Yu, D., Grimwood, J., Schmutz, J., Dickson, M., Jarvis, M. A., Hahn, G., Nelson, J. A., Myers, R. M., and Shenk, T. E. (2003). Coding potential of laboratory and clinical strains of human cytomegalovirus. *Proceedings of the National Academy of Sciences* 100(25), 14976–14981.

Murphy, E., Vaníček, J., Robins, H., Shenk, T., and Levine, A. J. (2008). Suppression of immediate-early viral gene expression by herpesvirus-coded microRNAs: implications for latency. *Proceedings of the National Academy of Sciences* 105(14), 5453–5458.

Murphy, K. and Weaver, C. (2016). Janeway's Immunobiology. Garland Science.

Murrell, I., Bedford, C., Ladell, K., Miners, K. L., Price, D. A., Tomasec, P., Wilkinson, G. W. G., and Stanton, R. J. (2017). The pentameric complex drives immunologically covert cell–cell transmission of wild-type human cytomegalovirus. *Proceedings of the National Academy of Sciences* 114(23), 6104–6109.

Nastke, M.-D., Herrgen, L., Walter, S., Wernet, D., Rammensee, H.-G., and Stevanović, S. (2005). Major contribution of codominant CD8 and CD4 T cell epitopes to the human cytomegalovirus-specific T cell repertoire. *Cellular and Molecular Life Sciences CMLS* 62(1), 77–86.

Navarro, D. (2016). Expanding role of cytomegalovirus as a human pathogen. *Journal of Medical Virology* 88(7), 1103–1112.

Nazarov, V. I., Pogorelyy, M. V., Komech, E. A., Zvyagin, I. V., Bolotin, D. A., Shugay, M., Chudakov, D. M., Lebedev, Y. B., and Mamedov, I. Z. (2015). tcR: an R package for T cell receptor repertoire advanced data analysis. *BMC Bioinformatics* 16(1), 175.

Neuenhahn, M., Albrecht, J., Odendahl, M., Schlott, F., Dössinger, G., Schiemann, M., Lakshmipathi, S., Martin, K., Bunjes, D., Harsdorf, S., Weissinger, E. M., Menzel, H., Verbeek, M., Uharek, L., Kröger, N., Wagner, E., Kobbe, G., Schroeder, T., Schmitt, M., Held, G., Herr, W., Germeroth, L., Bonig, H., Tonn,

T., Einsele, H., Busch, D. H., and Grigoleit, G. U. (2017). Transfer of minimally manipulated CMV-specific T cells from stem cell or third-party donors to treat CMV infection after allo-HSCT. *Leukemia* 31(10), 2161.

Nielsen, M., Lundegaard, C., Worning, P., Lauemøller, S. L., Lamberth, K., Buus, S., Brunak, S., and Lund, O. (2003). Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Science* 12(5), 1007–1017.

Nilges, K., Höhn, H., Pilch, H., Neukirch, C., Freitag, K., Talbot, P. J., and Maeurer, M. J. (2003). Human papillomavirus type 16 E7 peptide-directed CD8+ T cells from patients with cervical cancer are cross-reactive with the coronavirus NS2 protein. *Journal of Virology* 77(9), 5464–5474.

Nohynek, H., Jokinen, J., Partinen, M., Vaarala, O., Kirjavainen, T., Sundman, J., Himanen, S.-L., Hublin, C., Julkunen, I., Olsén, P., Saarenpää-Heikkilä, O., and Kilpi, T. (2012). AS03 adjuvanted AH1N1 vaccine associated with an abrupt increase in the incidence of childhood narcolepsy in Finland. *PLOS ONE* 7(3), e33536.

Noriega, V., Redmann, V., Gardner, T., and Tortorella, D. (2012). Diverse immune evasion strategies by human cytomegalovirus. *Immunologic Research* 54(1-3), 140–151.

Oldstone, M. B. A. (1998). Molecular mimicry and immune-mediated diseases. *The FASEB Journal* 12(13), 1255–1265.

Pachnio, A., Ciaurriz, M., Begum, J., Lal, N., Zuo, J., Beggs, A., and Moss, P. (2016). Cytomegalovirus infection leads to development of high frequencies of cytotoxic virus-specific CD4+ T cells targeted to vascular endothelium. *PLOS Pathogens* 12(9), e1005832.

Paillard, F., Sterkers, G., and Vaquero, C. (1990). Transcriptional and post-transcriptional regulation of TcR, CD4 and CD8 gene expression during activation of normal human T lymphocytes. *The EMBO journal* 9(6), 1867–1872.

Palser, A. L., Grayson, N. E., White, R. E., Corton, C., Correia, S., Ba abdullah, M. M., Watson, S. J., Cotten, M., Arrand, J. R., Murray, P. G., Allday, M. J., Rickinson, A. B., Young, L. S., Farrell, P. J., and Kellam, P. (2015). Genome diversity of Epstein-Barr virus from multiple tumor types and normal infection. *Journal of Virology* 89(10), 5222–5237.

Papadopoulou, A., Gerdemann, U., Katari, U. L., Tzannou, I., Liu, H., Martinez, C., Leung, K., Carrum, G., Gee, A. P., Vera, J. F., Krance, R. A., Brenner, M. K., Rooney, C. M., Heslop, H. E., and Leen, A. M. (2014). Activity of broad-spectrum T cells as treatment for AdV, EBV, CMV, BKV, and HHV6 infections after HSCT. *Science Translational Medicine* 6(242), 242ra83.

Pass, R. F., Zhang, C., Evans, A., Simpson, T., Andrews, W., Huang, M.-L., Corey, L., Hill, J., Davis, E., Flanigan, C., and Cloud, G. (2009). Vaccine prevention of maternal cytomegalovirus infection. *New England Journal of Medicine* 360(12), 1191–1199.

Peggs, K. S. (2009). Adoptive T cell immunotherapy for cytomegalovirus. *Expert Opinion on Biological Therapy* 9(6), 725–736.

Peggs, K. S., Verfuerth, S., Pizzey, A., Khan, N., Guiver, M., Moss, P. A., and Mackinnon, S. (2003). Adoptive cellular therapy for early cytomegalovirus infection after allogeneic stem-cell transplantation with virus-specific T-cell lines. *The Lancet* 362(9393), 1375–1377.

Perng, G.-C., Jones, C., Ciacci-Zanella, J., Stone, M., Henderson, G., Yukht, A., Slanina, S. M., Hofman, F. M., Ghiasi, H., Nesburn, A. B., and Wechsler, S. L. (2000). Virus-induced neuronal apoptosis blocked by the herpes simplex virus latency-associated transcript. *Science* 287(5457), 1500–1503.

Pfeffer, S., Zavolan, M., Grässer, F. A., Chien, M., Russo, J. J., Ju, J., John, B., Enright, A. J., Marks, D., Sander, C., and Tuschl, T. (2004). Identification of virus-encoded microRNAs. *Science* 304(5671), 734–736.

Price, D. A., Brenchley, J. M., Ruff, L. E., Betts, M. R., Hill, B. J., Roederer, M., Koup, R. A., Migueles, S. A., Gostick, E., Wooldridge, L., Sewell, A. K., Connors, M., and Douek, D. C. (2005). Avidity for antigen shapes clonal dominance in CD8+ T cell populations specific for persistent DNA viruses. *Journal of Experimental Medicine* 202(10), 1349–1361.

Prod'homme, V., Griffin, C., Aicheler, R. J., Wang, E. C. Y., McSharry, B. P., Rickards, C. R., Stanton, R. J., Borysiewicz, L. K., López-Botet, M., Wilkinson, G. W. G., and Tomasec, P. (2007). The human cytomegalovirus MHC class I homolog UL18 inhibits LIR-1+ but activates LIR-1- NK cells. *The Journal of Immunology* 178(7), 4473–4481.

Quinnan, G. V., Kirmani, N., Rook, A. H., Manischewitz, J. F., Jackson, L., Moreschi, G., Santos, G. W., Saral, R., and Burns, W. H. (1982). Cytotoxic t cells in cytomegalovirus infection: HLA-restricted T-

lymphocyte and non-T-lymphocyte cytotoxic responses correlate with recovery from cytomegalovirus infection in bone-marrow-transplant recipients. *New England Journal of Medicine* 307(1), 7–13.

Rammensee, H.-G., Bachmann, J., Emmerich, N. P. N., Bachor, O. A., and Stevanović, S. (1999). SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* 50(3-4), 213–219.

Rancan, C., Schirrmann, L., Hüls, C., Zeidler, R., and Moosmann, A. (2015). Latent membrane protein LMP2A impairs recognition of EBV-infected cells by CD8+ T cells. *PLOS Pathogens* 11(6), e1004906.

Rasmussen, M., Harndahl, M., Stryhn, A., Boucherma, R., Nielsen, L. L., Lemonnier, F. A., Nielsen, M., and Buus, S. (2014). Uncovering the peptide-binding specificities of HLA-C: a general strategy to determine the specificity of any MHC class I molecule. *The Journal of Immunology* 193(10), 4790–4802.

Reeves, M. B., MacAry, P. A., Lehner, P. J., Sissons, J. G. P., and Sinclair, J. H. (2005). Latency, chromatin remodeling, and reactivation of human cytomegalovirus in the dendritic cells of healthy carriers. *Proceedings of the National Academy of Sciences* 102(11), 4140–4145.

Rehm, A., Engelsberg, A., Tortorella, D., Körner, I. J., Lehmann, I., Ploegh, H. L., and Höpken, U. E. (2002). Human cytomegalovirus gene products US2 and US11 differ in their ability to attack major histocompatibility class I heavy chains in dendritic cells. *Journal of Virology* 76(10), 5043–5050.

Renzette, N., Gibson, L., Bhattacharjee, B., Fisher, D., Schleiss, M. R., Jensen, J. D., and Kowalik, T. F. (2013). Rapid intrahost evolution of human cytomegalovirus is shaped by demography and positive selection. *PLOS Genetics* 9(9), e1003735.

Renzette, N., Pokalyuk, C., Gibson, L., Bhattacharjee, B., Schleiss, M. R., Hamprecht, K., Yamamoto, A. Y., Mussi-Pinhata, M. M., Britt, W. J., Jensen, J. D., and Kowalik, T. F. (2015). Limits and patterns of cytomegalovirus genomic diversity in humans. *Proceedings of the National Academy of Sciences* 112(30), E4120–E4128.

Retière, C., Prod'homme, V., Imbert-Marcille, B.-M., Bonneville, M., Vié, H., and Hallet, M.-M. (2000). Generation of cytomegalovirus-specific human T-lymphocyte clones by using autologous B-lymphoblastoid cells with stable expression of pp65 or IE1 proteins: a tool to study the fine specificity of the antiviral response. *Journal of Virology* 74(9), 3948–3952.

Revello, M. G. and Gerna, G. (2010). Human cytomegalovirus tropism for endothelial/epithelial cells: scientific background and clinical implications. *Reviews in Medical Virology* 20(3), 136–155.

Rist, M., Smith, C., Bell, M. J., Burrows, S. R., and Khanna, R. (2009). Cross-recognition of HLA DR4 alloantigen by virus-specific CD8+ T cells: a new paradigm for self-/nonself-recognition. *Blood* 114(11), 2244–2253.

Rius, C., Attaf, M., Tungatt, K., Bianchi, V., Legut, M., Bovay, A., Donia, M., Straten, P. T., Peakman, M., Svane, I. M., Ott, S., Connor, T., Szomolay, B., Dolton, G., and Sewell, A. K. (2018). Peptide–MHC class I tetramers can fail to detect televant functional T cell clonotypes and underestimate antigen-reactive T cell populations. *The Journal of Immunology* 200(7), 2263–2279.

Rivino, L., Tan, A. T., Chia, A., Kumaran, E. A. P., Grotenbreg, G. M., MacAry, P. A., and Bertoletti, A. (2013). Defining CD8+ T cell determinants during human viral infection in populations of Asian ethnicity. *The Journal of Immunology* 191(8), 4010–4019.

Robins, H. S., Campregher, P. V., Srivastava, S. K., Wacher, A., Turtle, C. J., Kahsai, O., Riddell, S. R., Warren, E. H., and Carlson, C. S. (2009). Comprehensive assessment of T-cell receptor β-chain diversity in $\alpha\beta$ T cells. *Blood* 114(19), 4099–4107.

Robinson, J., Halliwell, J. A., and Marsh, S. G. E. (2014). IMGT/HLA and the immuno polymorphism database. *Immunoinformatics*. Springer, 109–121.

Roddie, C. and Peggs, K. S. (2017). Immunotherapy for transplantation-associated viral infections. *The Journal of Clinical Investigation* 127(7), 2513–2522.

Rossetto, C. C., Tarrant-Elorza, M., and Pari, G. S. (2013). Cis and trans acting factors involved in human cytomegalovirus experimental and natural latent infection of CD14 (+) monocytes and CD34 (+) cells. *PLOS Pathogens* 9(5), e1003366.

Rowntree, L. C., Nguyen, T. H. O., Halim, H., Purcell, A. W., Rossjohn, J., Gras, S., Kotsimbos, T. C., and Mifsud, N. A. (2018). Inability To Detect Cross-Reactive Memory T Cells Challenges the Frequency of Heterologous Immunity among Common Viruses. *The Journal of Immunology* 200(12), 3993–4003.

Rudolph, M. G., Stanfield, R. L., and Wilson, I. A. (2006). How TCRs bind MHCs, peptides, and coreceptors. *Annual Review of Immunology* 24, 419–466.

Ruggiero, E., Nicolay, J. P., Fronza, R., Arens, A., Paruzynski, A., Nowrouzi, A., Ürenden, G., Lulay, C., Schneider, S., Goerdt, S., Glimm, H., Krammer, P. H., and Schmidt Manfred & von Kalle, C. (2015). High-resolution analysis of the human T-cell receptor repertoire. *Nature Communications* 6, 8081.

Ryckman, B. J., Jarvis, M. A., Drummond, D. D., Nelson, J. A., and Johnson, D. C. (2006). Human cytomegalovirus entry into epithelial and endothelial cells depends on genes UL128 to UL150 and occurs by endocytosis and low-pH fusion. *Journal of Virology* 80(2), 710–722.

Ryckman, B. J., Rainish, B. L., Chase, M. C., Borton, J. A., Nelson, J. A., Jarvis, M. A., and Johnson, D. C. (2008). Characterization of the human cytomegalovirus gH/gL/UL128-131 complex that mediates entry into epithelial and endothelial cells. *Journal of Virology* 82(1), 60–70.

Sacre, K., Nguyen, S., Deback, C., Carcelain, G., Vernant, J.-P., Leblond, V., Autran, B., and Dhedin, N. (2008). Expansion of human cytomegalovirus (HCMV) immediate-early 1-specific CD8+ T cells and control of HCMV replication after allogeneic stem cell transplantation. *Journal of Virology* 82(20), 10143–10152.

Saulquin, X., Ibisch, C., Peyrat, M.-A., Scotet, E., Hourmant, M., Vie, H., Bonneville, M., and Houssaint, E. (2000). A global appraisal of immunodominant CD8 T cell responses to Epstein-Barr virus and cytomegalovirus by bulk screening. *European Journal of Immunology* 30(9), 2531–2539.

Savaryn, J. P., Reitsma, J. M., Bigley, T. M., Halligan, B. D., Qian, Z., Yu, D., and Terhune, S. S. (2013). Human cytomegalovirus pUL29/28 and pUL38 repression of p53-regulated p21CIP1 and caspase 1 promoters during infection. *Journal of Virology* 87(5), 2463–2474.

Saverino, D., Ghiotto, F., Merlo, A., Bruno, S., Battini, L., Occhino, M., Maffei, M., Tenca, C., Pileri, S., Baldi, L., Fabbi, M., Bachi, A., De Santanna, A., Grossi, C. E., and Ciccone, E. (2004). Specific recognition of the viral protein UL18 by CD85j/LIR-1/ILT2 on CD8+ T cells mediates the non-MHC-restricted lysis of human cytomegalovirus-infected cells. *The Journal of Immunology* 172(9), 5629–5637.

Schellens, I. M., Meiring, H. D., Hoof, I., Spijkers, S. N., Poelen, M., Brink, J. van Gaans-van den, Costa, A. I., Vennema, H., Keşmir, C., Baarle, D. van, and Els, C. A.C. M. van (2015). Measles virus epitope presentation by HLA: novel insights into epitope selection, dominance, and microvariation. *Frontiers in Immunology* 6, 546.

Schlott, F., Steubl, D., Ameres, S., Moosmann, A., Dreher, S., Heemann, U., Hösel, V., Busch, D. H., and Neuenhahn, M. (2018). Characterization and clinical enrichment of HLA-C*07:02-restricted cytomegalovirus-specific CD8+ T cells. *PLOS ONE* 13(2), e0193554.

Schmidt, A. H., Baier, D., Solloch, U. V., Stahr, A., Cereb, N., Wassmuth, R., Ehninger, G., and Rutt, C. (2009). Estimation of high-resolution HLA-A,-B,-C,-DRB1 allele and haplotype frequencies based on 8862 German stem cell donors and implications for strategic donor registry planning. *Human Immunology* 70(11), 895–902.

Schub, A., Schuster, I. G., Hammerschmidt, W., and Moosmann, A. (2009). CMV-specific TCR-transgenic T cells for immunotherapy. *The Journal of Immunology* 183(10), 6819–6830.

Schust, D. J., Tortorella, D., Seebach, J., Phan, C., and Ploegh, H. L. (1998). Trophoblast class I major histocompatibility complex (MHC) products are resistant to rapid degradation imposed by the human cytomegalovirus (HCMV) gene products US2 and US11. *Journal of Experimental Medicine* 188(3), 497–503.

Schwanninger, A., Weinberger, B., Weiskopf, D., Herndler-Brandstetter, D., Reitinger, S., Gassner, C., Schennach, H., Parson, W., Würzner, R., and Grubeck-Loebenstein, B. (2008). Age-related appearance of a CMV-specific high-avidity CD8+ T cell clonotype which does not occur in young adults. *Immunity & Ageing* 5(1), 14.

Selin, L. K., Nahill, S. R., and Welsh, R. M. (1994). Cross-reactivities in memory cytotoxic T lymphocyte recognition of heterologous viruses. *Journal of Experimental Medicine* 179(6), 1933–1943.

Sethna, Z., Elhanati, Y., Callan Curtis G, J., Walczak, A. M., and Mora, T. (2019). OLGA: fast computation of generation probabilities of B- and T-cell receptor amino acid sequences and motifs. *Bioinformatics*, btz035.

Shiina, T., Hosomichi, K., Inoko, H., and Kulski, J. K. (2009). The HLA genomic loci map: expression, interaction, diversity and disease. *Journal of Human Genetics* 54(1), 15.

Shugay, M., Bagaev, D. V., Zvyagin, I. V., Vroomans, R. M., Crawford, J. C., Dolton, G., Komech, E. A., Sycheva, A. L., Koneva, A. E., Egorov, E. S., Eliseev, A. V., Van Dyk, E., Dash, P., Attaf, M., Rius, C., Ladell, K., McLaren, J. E., Matthews, K. K., Clemens, E., Douek, D. C., Luciani, F., van Baarle, D.,

Kedzierska, K., Kesmir, C., Thomas, P. G., Price, D. A., Sewell, A. K., and Chudakov, D. M. (2017). VDJdb: a curated database of T-cell receptor sequences with known antigen specificity. *Nucleic Acids Research* 46(D1), D419–D427.

Sijmons, S., Thys, K., Mbong Ngwese, M., Van Damme, E., Dvorak, J., Van Loock, M., Li, G., Tachezy, R., Busson, L., Aerssens, J., Van Ranst, M., and Maes, P. (2015). High-throughput analysis of human cytomegalovirus genome diversity highlights the widespread occurrence of gene-disrupting mutations and pervasive recombination. *Journal of Virology* 89(15), 7673–7695.

Sinzger, C., Hahn, G., Digel, M., Katona, R., Sampaio, K. L., Messerle, M., Hengel, H., Koszinowski, U., Brune, W., and Adler, B. (2008). Cloning and sequencing of a highly productive, endotheliotropic virus strain derived from human cytomegalovirus TB40/E. *Journal of General Virology* 89(2), 359–368.

Sissons, J. G. P. and Carmichael, A. J. (2002). Clinical aspects and management of cytomegalovirus infection. *Journal of Infection* 44(2), 78–83.

Slezak, S. L., Bettinotti, M., Selleri, S., Adams, S., Marincola, F. M., and Stroncek, D. F. (2007). CMV pp65 and IE-1 T cell epitopes recognized by healthy subjects. *Journal of Translational Medicine* 5(1), 17.

Smeraldi, R. S., Lazzarin, A., Moroni, M., Fabio, G., Eisera, N. B., and Zanussi, C. (1986). HLA-associated susceptibility to acquired immunodeficiency syndrome in Italian patients with human-immunodeficiency-virus infection. *The Lancet* 328(8517), 1187–1189.

Smith, C and Khanna, R (2013). Immune regulation of human herpesviruses and its implications for human transplantation. *American Journal of Transplantation* 13(s3), 9–23.

Smith, C., Brennan, R. M., Tey, S.-K., Smyth, M. J., Burrows, S. R., Miles, J. J., Hill, G. R., and Khanna, R. (2016). Coinfection with Human Cytomegalovirus Genetic Variants in Transplant Recipients and Its Impact on Antiviral T Cell Immune Reconstitution. *Journal of Virology* 90(16). Ed. by K. Frueh, 7497–7507.

Sollid, L. M., Markussen, G., Ek, J., Gjerde, H., Vartdal, F., and Thorsby, E. (1989). Evidence for a primary association of celiac disease to a particular HLA-DQ alpha/beta heterodimer. *Journal of Experimental Medicine* 169(1), 345–350.

Sommer, S. (2005). The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Frontiers in Zoology* 2(1), 16.

Stadinski, B. D., Shekhar, K., Gómez-Touriño, I., Jung, J., Sasaki, K., Sewell, A. K., Peakman, M., Chakraborty, A. K., and Huseby, E. S. (2016). Hydrophobic CDR3 residues promote the development of self-reactive T cells. *Nature Immunology* 17(8), 946.

Steinle, A., Reinhardt, C., Jantzer, P., and Schendel, D. J. (1995). In vivo expansion of HLA-B35 alloreactive T cells sharing homologous T cell receptors: evidence for maintenance of an oligoclonally dominated allospecificity by persistent stimulation with an autologous MHC/peptide complex. *Journal of Experimental Medicine* 181(2), 503–513.

Stenberg, R. M., Fortney, J., Barlow, S. W., Magrane, B. P., Nelson, J. A., and Ghazal, P. (1990). Promoter-specific trans activation and repression by human cytomegalovirus immediate-early proteins involves common and unique protein domains. *Journal of Virology* 64(4), 1556–1565.

Stinski, M. F. (1978). Sequence of protein synthesis in cells infected by human cytomegalovirus: early and late virus-induced polypeptides. *Journal of Virology* 26(3), 686–701.

Storch, G. A., Buller, R. S., Bailey, T. C., Ettinger, N. A., Langlois, T., Gaudreault-Keener, M., and Welby, P. L. (1994). Comparison of PCR and pp65 antigenemia assay with quantitative shell vial culture for detection of cytomegalovirus in blood leukocytes from solid-organ transplant recipients. *Journal of Clinical Microbiology* 32(4), 997–1003.

Suni, M. A., Ghanekar, S. A., Houck, D. W., Maecker, H. T., Wormsley, S. B., Picker, L. J., Moss, R. B., and Maino, V. C. (2001). CD4+ CD8dim T lymphocytes exhibit enhanced cytokine expression, proliferation and cytotoxic activity in response to HCMV and HIV-1 antigens. *European Journal of Immunology* 31(8), 2512–2520.

Sylwester, A. W., Mitchell, B. L., Edgar, J. B., Taormina, C., Pelte, C., Ruchti, F., Sleath, P. R., Grabstein, K. H., Hosken, N. A., Kern, F., Nelson, J. A., and Picker, L. J. (2005). Broadly targeted human cytomegalovirus-specific CD4+ and CD8+ T cells dominate the memory compartments of exposed subjects. *Journal of Experimental Medicine* 202(5), 673–685.

Tan, L. C., Gudgeon, N., Annels, N. E., Hansasuta, P., O'Callaghan, C. A., Rowland-Jones, S., McMichael, A. J., Rickinson, A. B., and Callan, M. F. C. (1999). A re-evaluation of the frequency of CD8+ T cells specific for EBV in healthy virus carriers. *The Journal of Immunology* 162(3), 1827–1835.

Taylor-Wiedeman, J., Sissons, P., and Sinclair, J. (1994). Induction of endogenous human cytomegalovirus gene expression after differentiation of monocytes from healthy carriers. *Journal of Virology* 68(3), 1597–1604.

Terasaki, P. I., Park, M. S., Opelz, G., and Ting, A. (1976). Multiple sclerosis and high incidence of a B lymphocyte antigen. *Science* 193(4259), 1245–1247.

Terhune, S. S., Moorman, N. J., Cristea, I. M., Savaryn, J. P., Cuevas-Bennett, C., Rout, M. P., Chait, B. T., and Shenk, T. (2010). Human cytomegalovirus UL29/28 protein interacts with components of the NuRD complex which promote accumulation of immediate-early RNA. *PLOS Pathogens* 6(6), e1000965.

Thomas, R., Apps, R., Qi, Y., Gao, X., Male, V., O'hUigin, C., O'Connor, G., Ge, D., Fellay, J., Martin, J. N., Margolick, J., Goedert, J. J., Buchbinder, S., Kirk, G. D., Martin, M. P., Telenti, A., Deeks, S. G., Walker, B. D., Goldstein, D., McVicar, D. W., Moffett, A., and Carrington, M. (2009). HLA-C cell surface expression and control of HIV/AIDS correlate with a variant upstream of HLA-C. *Nature Genetics* 41(12), 1290.

Tiilikainen, A., Lassus, A., Karvonen, J., Vartiainen, P., and Julin, M. (1980). Psoriasis and HLA-Cw6. *British Journal of Dermatology* 102(2), 179–184.

Tomazin, R., Boname, J., Hegde, N. R., Lewinsohn, D. M., Altschuler, Y., Jones, T. R., Cresswell, P., Nelson, J. A., Riddell, S. R., and Johnson, D. C. (1999). Cytomegalovirus US2 destroys two components of the MHC class II pathway, preventing recognition by CD4+ T cells. *Nature Medicine* 5(9), 1039.

Tortorella, D., Gewurz, B. E., Furman, M. H., Schust, D. J., and Ploegh, H. L. (2000). Viral subversion of the immune system. *Annual Review of Immunology* 18(1), 861–926.

Trautmann, L., Rimbert, M., Echasserieau, K., Saulquin, X., Neveu, B., Dechanet, J., Cerundolo, V., and Bonneville, M. (2005). Selection of T cell clones expressing high-affinity public TCRs within human cytomegalovirus-specific CD8 T cell responses. *The Journal of Immunology* 175(9), 6123–6132.

Trivedi, D., Williams, R. Y., O'Reilly, R. J., and Koehne, G. (2005). Generation of CMV-specific T lymphocytes using protein-spanning pools of pp65-derived overlapping pentadecapeptides for adoptive immunotherapy. *Blood* 105(7), 2793–2801.

Tu, W. and Rao, S. (2016). Mechanisms underlying T cell immunosenescence: aging and cytomegalovirus infection. *Frontiers in Microbiology* 7, 2111.

Tzannou, I., Papadopoulou, A., Naik, S., Leung, K., Martinez, C. A., Ramos, C. A., Carrum, G., Sasa, G., Lulla, P., Watanabe, A., Kuvalekar, M., Gee, A. P., Wu, M.-F., Liu, H., Grilley, B. J., Krance, R. A., Gottschalk, S., Brenner, M. K., Rooney, C. M., Heslop, H. E., Leen, A. M., and Omer, B. (2017). Off-the-shelf virus-specific T cells to treat BK virus, human herpesvirus 6, cytomegalovirus, Epstein-Barr virus, and adenovirus infections after allogeneic hematopoietic stem-cell transplantation. *Journal of Clinical Oncology* 35(31), 3547.

Umbach, J. L., Kramer, M. F., Jurak, I., Karnowski, H. W., Coen, D. M., and Cullen, B. R. (2008). MicroRNAs expressed by herpes simplex virus 1 during latent infection regulate viral mRNAs. *Nature* 454(7205), 780.

Venturi, V., Chin, H. Y., Asher, T. E., Ladell, K., Scheinberg, P., Bornstein, E., Bockel, D. van, Kelleher, A. D., Douek, D. C., Price, D. A., and Davenport, M. P. (2008). TCR β-chain sharing in human CD8+ T cell responses to cytomegalovirus and EBV. *The Journal of Immunology* 181(11), 7853–7862.

Vincenti, F., Budde, K., Merville, P., Shihab, F., Ram Peddi, V., Shah, M., Wyburn, K., Cassuto-Viguier, E., Weidemann, A., Lee, M., Flegel, T., Erdman, J., Wang, X., and Lademacher, C. (2018). A randomized, phase 2 study of ASP 0113, a DNA-based vaccine, for the prevention of CMV in CMV-seronegative kidney transplant recipients receiving a kidney from a CMV-seropositive donor. *American Journal of Transplantation* 18(12), 2945–2954.

Vita, R., Mahajan, S., Overton, J. A., Dhanda, S. K., Martini, S., Cantrell, J. R., Wheeler, D. K., Sette, A., and Peters, B. (2018). The immune epitope database (IEDB): 2018 update. *Nucleic Acids Research* 47(D1), D339–D343.

Walker, S., Fazou, C., Crough, T., Holdsworth, R., Kiely, P., Veale, M., Bell, S., Gailbraith, A., McNeil, K., Jones, S., and Khanna, R. (2007). Ex vivo monitoring of human cytomegalovirus-specific CD8+ T-cell responses using QuantiFERON®-CMV. *Transplant Infectious Disease* 9(2), 165–170.

Walter, E. A., Greenberg, P. D., Gilbert, M. J., Finch, R. J., Watanabe, K. S., Thomas, E. D., and Riddell, S. R. (1995). Reconstitution of cellular immunity against cytomegalovirus in recipients of allogeneic bone marrow by transfer of T-cell clones from the donor. *New England Journal of Medicine* 333(16), 1038–1044.

Wang, D. and Shenk, T. (2005). Human cytomegalovirus virion protein complex required for epithelial and endothelial cell tropism. *Proceedings of the National Academy of Sciences* 102(50), 18153–18158.

Wedemeyer, H., Mizukoshi, E., Davis, A. R., Bennink, J. R., and Rehermann, B. (2001). Cross-reactivity between hepatitis C virus and Influenza A virus determinant-specific cytotoxic T cells. *Journal of Virology* 75(23), 11392–11400.

Weekes, M. P., Wills, M. R., Mynard, K., Carmichael, A. J., and Sissons, J. G. P. (1999). The memory cytotoxic T-lymphocyte (CTL) response to human cytomegalovirus infection contains individual peptide-specific CTL clones that have undergone extensive expansion in vivo. *Journal of Virology* 73(3), 2099–2108.

Weißbrich, B. (2016). T cell receptor binding avidity of antigen-specific CD8+ cytotoxic T cells in chronic infection. PhD thesis. Technical University of Munich.

Wieczorek, M., Abualrous, E. T., Sticht, J., Álvaro-Benito, M., Stolzenberg, S., Noé, F., and Freund, C. (2017). Major histocompatibility complex (MHC) class I and MHC class II proteins: conformational plasticity in antigen presentation. *Frontiers in Immunology* 8, 292.

Wiesner, M., Zentz, C., Hammer, M. H., Cobbold, M., Kern, F., Kolb, H.-J., Hammerschmidt, W., Zeidler, R., and Moosmann, A. (2005). Selection of CMV-specific CD8+ and CD4+ T cells by mini-EBV-transformed B cell lines. *European Journal of Immunology* 35(7), 2110–2121.

Wiesner, M., Zentz, C., Mayr, C., Wimmer, R., Hammerschmidt, W., Zeidler, R., and Moosmann, A. (2008). Conditional immortalization of human B cells by CD40 ligation. *PLOS ONE* 3(1), e1464.

Wills, M. R., Carmichael, A. J., Mynard, K., Jin, X., Weekes, M. P., Plachter, B., and Sissons, J. G. (1996). The human cytotoxic T-lymphocyte (CTL) response to cytomegalovirus is dominated by structural protein pp65: frequency, specificity, and T-cell receptor usage of pp65-specific CTL. *Journal of Virology* 70(11), 7569–7579.

Yang, H. Y., Dundon, P. L., Nahill, S. R., and Welsh, R. M. (1989). Virus-induced polyclonal cytotoxic T lymphocyte stimulation. *The Journal of Immunology* 142(5), 1710–1718.

Yewdell, J. W. (2005). The seven dirty little secrets of major histocompatibility complex class I antigen processing. *Immunological Reviews* 207(1), 8–18.

Yewdell, J. W. and Bennink, J. R. (1999). Immunodominance in major histocompatibility complex class I–restricted T lymphocyte responses. *Annual Review of Immunology* 17(1), 51–88.

Yu, X., Jih, J., Jiang, J., and Zhou, Z. H. (2017). Atomic structure of the human cytomegalovirus capsid with its securing tegument layer of pp150. *Science* 356(6345), eaam6892.

Zemmour, J. and Parham, P. (1992). Distinctive polymorphism at the HLA-C locus: implications for the expression of HLA-C. *Journal of Experimental Medicine* 176(4), 937–950.

Zhong, J., Rist, M., Cooper, L., Smith, C., and Khanna, R. (2008). Induction of pluripotent protective immunity following immunisation with a chimeric vaccine against human cytomegalovirus. *PLOS ONE* 3(9), e3256.

Zhou, D., Srivastava, R., Grummel, V., Cepok, S., Hartung, H.-P., and Hemmer, B. (2006). High throughput analysis of TCR-β rearrangement and gene expression in single T cells. *Laboratory Investigation* 86(3), 314.

ACKNOWLEDGEMENTS

I would like to express my gratitude to all the amazing people who supported and reinforced me over the course of this PhD project.

First and foremost, I thank PD Dr. Andreas Moosmann for his excellent supervision. You devised this exciting project, guided me through my time as a PhD student, challenged me to think outside the box, and always trusted in my scientific abilities. I am deeply grateful for your valuable input, the countless discussions, and, above all, your kindness and ingenuity.

I sincerely thank Prof. Dr. Horst Domdey for his support as the faculty supervisor of my PhD thesis.

My special thanks goes to the members of my Thesis Committee, Prof. Dr. Wolfgang Hammerschmidt, PD Dr. Reinhard Obst, and PD Dr. Josef Mautner. Your innovative ideas and constructive criticism helped me a great deal throughout this project.

I also thank the members of the SFB-TR36 ”Principles and Applications of Adoptive T Cell Therapy”, in particular the Graduate College coordinator Prof. Dr. Elfriede Nößner. You gave me the opportunity to take part in many conferences, retreats, and seminars, where I learned much about immunology and met outstanding researchers.

Moreover, I would like to thank the blood donors, who heroically endured physical pain for the progress of my research project.

To my coworkers at the Helmholtz Center Munich (AGV), especially Wiebe, Markus, and Alex: Thank you for cheering me on and up, helping me with words and deeds, and discussing scientific and social matters with me over coffee. You made my PhD experience truly enjoyable and memorable.

My dear family and friends, thank you for your endless emotional support and for always believing in me. I could not have gone through this challenging, yet inspiring time without you. I dedicate this thesis to my parents and my sister, for it is because of you that I can follow my dreams and pursue a career I am passionate about.

APPENDIX

Table A1: List of high-throughput sequencing samples used to identify CMV-specific TCRβ sequences from single peptide stimulations of CMV-positive donors.

Donor	Sample ID	Treatment	Cell type	Total productive reads	Number of clonotypes
P01	0073	unstimulated	CD8	12043137	584630
P01	0086	unstimulated	all	14515877	583111
P01	0080	CRV	CD8	18182271	84775
P01	0182	FRC	all	4333494	363333
P01	0079	RPH	CD8	7438191	222019
P01	0078	TPR	CD8	8134892	37234
P02	0188	unstimulated	all	4461887	403105
P02	0190	CRV	all	4334321	779479
P02	0191	FRC	all	2893655	507295
P02	-	NLV	not done	not done	not done
P02	-	RPH	not done	not done	not done
P02	0192	TPR	all	4708088	623779
P02	0193	VLE	all	5043299	710836
P03	0162	unstimulated	all	4536124	584506
P03	0184	CRV	all	4138428	538597
P03	0164	FRC	all	5240440	305592
P03	0165	NLV	all	4486573	274060
P03	0185	RPH	all	3547381	386160
P03	0175	TPR	all	4710517	337660
P03	0167	VLE	all	4412951	318528
P04	0015	unstimulated	CD8	963210	212125
P04	0017	CRV	CD8	1592670	59739
P04	0021	FRC	CD8	1440428	75470
P04	0018	NLV	CD8	2153113	234318
P04	0020	RPH	CD8	1375678	200167
P04	0016	TPR	CD8	1425865	84142
P04	-	VLE	not done	not done	not done
P05	0035	unstimulated	CD8	13028045	283478
P05	0036	CRV	CD8	15000704	81340
P05	0040	FRC	CD8	25588668	25723
P05	0041	IPS	CD8	11332377	60300
P05	0039	RPH	CD8	16294617	132349
P05	0038	TPR	CD8	13010720	122888
P06	0025	unstimulated	CD8	1042812	183707
P06	0026	CRV	CD8	2425994	55471
P06	0030	FRC	CD8	1899999	31431
P06	0029	RPH	CD8	938285	60619
P06	0028	TPR	CD8	1802570	64580
P07	0049	unstimulated	CD8	4821591	256394
P07	0054	CRV	CD8	8951471	121384
P07	0056	FRC	CD8	7263340	97686
P07	0050	NLV	CD8	5694601	109483
P07	0053	RPH	CD8	6471832	154897
P07	0052	TPR	CD8	5359364	119764
P07	0051	VLE	CD8	7565271	128115
P08	0057	unstimulated	CD8	7334612	111989
P08	0062	CRV	CD8	6748656	92462
P08	0064	FRC	CD8	6042090	21546

Table A1: (continued)

Donor	Sample ID	Treatment	Cell type	Total productive reads	Number of clonotypes
P08	0058	NLV	CD8	6948206	72414
P08	0061	RPH	CD8	6210355	53741
P08	0060	TPR	CD8	7362160	77073
P08	0059	VLE	CD8	5124377	67568
P14	0158	unstimulated	CD8	4744864	29679
P14	0159	CRV	CD8	4066750	30068
P14	0161	FRC	CD8	5403673	14181
P14	0160	TPR	CD8	5148835	40137
Statistics					
Median reads:		512437		Median clonotypes:	128115
Min. reads:		938285		Min. clonotypes:	21546
Max. reads:		25588668		Max. clonotypes:	779479

Table A2: Computed cutoff values for the identification of specific TCRβ clonotypes by single peptide stimulation of PBMCs derived from CMV-positive donors. Abbreviations: ctrl = control peptide, spec = specific, enr = enrichment, d0 = unstimulated sample.

Donor	Peptide	Ctrl	Spec. TCRβs	Read cutoff	Enr. cutoff (ctrl)	Enr. cutoff (d0)
P01	CRV	TPR	119	297	5.12	0.57
P01	FRC	TPR	24	1949	1116.99	23
P01	RPH	CRV	27	4565	3323.06	210.77
P01	TPR	CRV	39	824	199.11	11.08
P02	CRV	TPR	31	396	20.14	5.91
P02	FRC	TPR	25	908	12.95	11.56
P02	NLV	not done	not done	not done	not done	not done
P02	RPH	not done	not done	not done	not done	not done
P02	TPR	CRV	38	618	9.04	15.35
P02	VLE	TPR	5	2280	47.95	24.2
P03	CRV	TPR	40	513	17.5	67.12
P03	FRC	TPR	17	3017	37.88	14.68
P03	NLV	CRV	9	3575	110.13	55.74
P03	RPH	CRV	17	1473	101.02	52.55
P03	TPR	CRV	23	3597	31.08	44.63
P03	VLE	TPR	16	2288	7.39	32.78
P04	CRV	TPR	45	582	15.34	3.95
P04	FRC	TPR	13	2194	249.29	34.87
P04	NLV	CRV	14	2083	2046.05	158.55
P04	RPH	CRV	19	986	70.62	100.35
P04	TPR	CRV	21	572	303.54	41.24
P04	VLE	not done	not done	not done	not done	not done
P05	CRV	TPR	118	132	2.05	1.16
P05	FRC	TPR	99	521	2.53	1.6
P05	IPS	CRV	30	2371	738.66	48.12
P05	RPH	CRV	75	3974	312.29	4.64
P05	TPR	CRV	22	2385	13173.6	16.35
P06	CRV	TPR	49	233	15.25	7.1
P06	FRC	TPR	36	914	25.65	6.15
P06	RPH	CRV	21	1468	137.7	11.63
P06	TPR	CRV	21	2067	170.05	59.32

Table A2: (continued)

Donor	Peptide	Ctrl	Spec. TCRβs	Read cutoff	Enr. cutoff (ctrl)	Enr. cutoff (d0)
P07	CRV	TPR	33	1970	19.21	4.24
P07	FRC	TPR	45	1107	14.08	25.68
P07	NLV	CRV	31	822	114.53	21.59
P07	RPH	CRV	40	1793	32.88	26.78
P07	TPR	CRV	11	1983	96.77	9.17
P07	VLE	TPR	14	2054	141.45	65.79
P08	CRV	TPR	25	438	15.84	17.52
P08	FRC	TPR	18	161	10.7	1.82
P08	NLV	CRV	33	2100	62.68	5.59
P08	RPH	CRV	23	479	78.07	22.1
P08	TPR	CRV	18	1388	63.12	9.6
P08	VLE	TPR	3	2052	78.98	5.55
P14	CRV	TPR	26	2265	18	11.72
P14	FRC	TPR	52	680	7.03	2.94
P14	TPR	CRV	55	693	28.28	74.8

Table A3: List of high-throughput sequencing samples used to identify epitope-reactive TCRβ sequences from single peptide stimulations of CMV-negative donors.

Donor	Sample ID	Treatment	Cell type	Total productive reads	Number of clonotypes
N03	0031	unstimulated	CD8	869916	214838
N03	0032	NLV	CD8	1427925	111428
N03	0033	CRV	CD8	1109306	113611
N03	0034	TPR	CD8	1474310	120585
N05	0042	unstimulated	CD8	5755743	738851
N05	0043	TPR	CD8	10157300	1083931
N05	0044	RPH	CD8	8815592	976338
N05	0045	IPS	CD8	9468082	954464
N05	0046	CRV	CD8	10898408	1086147
N05	0048	FRC	CD8	9574205	877361
N04	0096	unstimulated	all	1676344	192539
N06	0097	unstimulated	all	1669857	319087
N07	0098	unstimulated	all	1530288	180502
N08	0099	unstimulated	all	1533634	104796
N01	0100	unstimulated	all	1748374	157616
N02	0101	unstimulated	all	1403545	269192
N04	0102	CRV	all	2115559	90908
N04	0103	FRC	all	1829487	87970
N04	0104	TPR	all	2060695	113698
N06	0105	CRV	all	2107213	156250
N06	0106	FRC	all	2152171	187758
N06	0107	TPR	all	1758352	115160
N07	0109	CRV	all	4853025	174541
N07	0110	FRC	all	4203024	226460
N07	0111	TPR	all	4729117	160043
N08	0112	CRV	all	3577046	98734
N08	0113	FRC	all	4677728	129521
N08	0114	TPR	all	4366374	157027
N01	0115	CRV	all	3854037	194343

Table A3: (continued)

Donor	Sample ID	Treatment	Cell type	Total productive reads	Number of clonotypes
N01	0117	TPR	all	3568847	160237
N02	0118	CRV	all	3682471	329210
N02	0119	FRC	all	3304195	329450
N02	0120	TPR	all	3247774	286225
N01	0154	FRC	all	2907567	161402
N03	0201	FRC	CD8	5268659	457776
Statistics					
Median reads:		3247774		Median clonotypes:	180502
Min. reads:		869916		Min. clonotypes:	87970
Max. reads:		10898408		Max. clonotypes:	1086147

Table A4: Computed cutoff values for the identification of specific TCRβ clonotypes by single peptide stimulation of PBMCs derived from CMV-negative donors. Abbreviations: ctrl = control peptide, spec = specific, enr = enrichment, d0 = unstimulated sample.

Donor	Peptide	Ctrl	Spec. TCRβs	Read cutoff	Enr. cutoff (ctrl)	Enr. cutoff (d0)
N01	CRV	TPR	5	1019	101.11	69.9
N01	TPR	CRV	0	4426	14.47	104.76
N01	FRC	TPR	10	2247	74.46	7.12
N02	CRV	TPR	25	1955	34.47	5.31
N02	FRC	TPR	10	3079	39.3	5.97
N02	TPR	CRV	2	6054	49.02	7.2
N03	NLV	CRV	6	746	14.32	22.94
N03	CRV	TPR	6	628	96.11	35.18
N03	TPR	CRV	1	1175	104.04	35.38
N03	FRC	TPR	4	5557	214.62	27.35
N04	CRV	TPR	23	939	66.93	2.96
N04	FRC	TPR	22	842	24.9	7.81
N04	TPR	CRV	4	2829	32.91	3.99
N05	TPR	CRV	30	1900	16.43	101.41
N05	RPH	CRV	49	1175	105.38	47.28
N05	IPS	CRV	12	2681	109.37	12.16
N05	CRV	TPR	26	1551	10.55	85.86
N05	FRC	TPR	60	1073	11.56	11.46
N06	CRV	TPR	24	606	90.05	34.16
N06	FRC	TPR	22	2303	72.22	12.33
N06	TPR	CRV	11	1851	17.28	22.95
N07	CRV	TPR	7	3976	26.47	37.98
N07	FRC	TPR	5	3739	89	63.14
N07	TPR	CRV	24	3086	18.17	4.05
N08	CRV	TPR	4	1395	77.29	38.34
N08	FRC	TPR	6	1499	103.55	35.56
N08	TPR	CRV	1	3650	129.39	24.23

Table A5: Proportions of reads that CMV-reactive TCRβ sequences hold in unstimulated (d0) and peptide-stimulated (d10) samples of donors N01–N08 in percent [%] of total reads and overall enrichment of these TCRβ sequences. Stimulations of donors N03 and N05 were performed by Xiaoling Liang and reanalysed in this thesis. Geometric means of the read proportions and the enrichment (d10÷d0) are noted in the bottom row. All values below 100 were rounded to two significant digits. Enr. = enrichment.

Donor	CRV			FRC			TPR		
	d0 [%]	d10 [%]	Enr.	d0 [%]	d10 [%]	Enr.	d0 [%]	d10 [%]	Enr.
N01	0.001	0.38	382	0.017	1.1	63	0*	0*	-
N02	0.048	3.3	70	0.015	1.6	103	0.0001	0.59	5939
N03†,‡	0.0029	0.51	176	0.0031	0.57	184	0.0001	0.094	944
N04†,§,‖	0.022	2.1	99	0.018	1.5	83	0	1.1	-
N05	0.0017	1	607	0.011	1.7	147	0.0021	2.2	1043
N06	0.016	2.8	180	0.025	7.2	293	0.013	3.2	247
N07	0.0056	0.94	169	0.002	0.81	405	0.059	3.5	594
N08	0.0003	0.21	712	0.0058	0.49	85	0	0.18	-
Geometric mean	**0.0044**	**0.98**	**223**	**0.0091**	**1.3**	**140**	**0.0017**	**0.83**	**611**

* No epitope-reactive TCRβ sequences found in this donor for this peptide.
† All samples of this donor were CD8-enriched before RNA isolation and TCRβ sequencing library preparation.
‡ NLV stimulation: d0 = 0.0066%; d10 = 0.72%; Enrichment = 110.
§ IPS stimulation: d0 = 0.0022%; d10 = 3.2%; Enrichment = 1463.
‖ RPH stimulation: d0 = 0.0018%; d10 = 2.734%; Enrichment = 1519.

Table A6: List of high-throughput sequencing samples that were CMV multimer-sorted before TCRβ sequencing and the corresponding unsorted control samples.

Donor	Sample ID	Treatment	Cell type	Total productive reads	Number of clonotypes
P01	0086	unstimulated	all	14515877	583111
P01	0076	CRV-sorted	all	7979341	7947
P01	0074	TPR-sorted	all	9024836	9392
P04	0089	unstimulated	all	18297841	298620
P04	0067	CRV-sorted	all	2960351	27141
P04	0065	TPR-sorted	all	10572282	33353
Statistics					
Median reads:		9798559		Median clonotypes:	30247
Min. reads:		2960351		Min. clonotypes:	7947
Max. reads:		18297841		Max. clonotypes:	583111

Table A7: List of high-throughput sequencing samples used to identify CMV antigen-specific TCRβ sequences from mini-LCL stimulations of CMV-positive donors P01–P03.

Donor	Sample ID	Treatment	Cell type	Total productive reads	Number of clonotypes
P01	0148	unstimulated	all	3327630	627081
P01	0140	pp65 mini-LCL	all	4159034	5018
P01	0141	IE1 mini-LCL	all	4876559	3956
P01	0142	empty mini-LCL	all	3251600	5669
P02	0188	unstimulated	all	4461887	403105
P02	-	pp65 mini-LCL	not done	not done	not done
P02	0194	IE1 mini-LCL	all	5168770	61537
P02	0195	empty mini-LCL	all	3761372	49042
P03	0162	unstimulated	all	4536124	584506
P03	0168	pp65 mini-LCL	all	4767526	137946
P03	0169	IE1 mini-LCL	all	4869584	71045
P03	0170	empty mini-LCL	all	5252146	124923
Statistics					
Median reads:		4536124		Median clonotypes:	71045
Min. reads:		3251600		Min. clonotypes:	3956
Max. reads:		5252146		Max. clonotypes:	627081

Table A8: Computed cutoff values for the identification of specific TCRβ clonotypes by mini-LCL stimulation of PBMCs derived from CMV-positive donors P01–P03. Abbreviations: ctrl = control peptide, spec = specific, enr = enrichment, d0 = unstimulated sample.

Donor	Antigen	Ctrl	Spec. TCRβs	Read cutoff	Enr. cutoff (ctrl)	Enr. cutoff (d0)
P01	pp65	empty mini-LCL	139	88	256.82	5
P01	IE1	empty mini-LCL	193	29	19.67	5
P02	pp65	not done	not done	not done	not done	not done
P02	IE1	empty mini-LCL	590	106	46	5
P03	pp65	empty mini-LCL	507	364	49.98	5
P03	IE1	empty mini-LCL	143	264	12.03	5

Table A9: List of high-throughput sequencing samples used to identify CMV epitope-specific TCRβ sequences from peptide pool stimulations of CMV-positive donors P01, P03, P11, and P17–P21.

Donor	Sample ID	Treatment	Cell type	Total productive reads	Number of clonotypes
P01	0247	unstimulated	all	3457397	497105
P01	0248	CMV pool 1	all	1885494	24022
P01	0249	CMV pool 2	all	1441626	15767
P01	0250	CMV pool 3	all	2644115	28002
P01	0251	CMV pool 4	all	1337003	36999
P01	0252	CMV pool 5	all	3385459	33525
P01	0253	CMV pool 6	all	2004841	222909
P01	0254	CMV pool 7	all	1861126	28779

Table A9: (continued)

Donor	Sample ID	Treatment	Cell type	Total productive reads	Number of clonotypes
P01	0255	CMV pool 8	all	2308324	22154
P03	0220	unstimulated	all	4675940	709835
P03	0221	CMV pool 1	all	2027312	115666
P03	0222	CMV pool 2	all	2435917	192033
P03	0223	CMV pool 3	all	2100551	110560
P03	0224	CMV pool 4	all	2135187	214906
P03	0225	CMV pool 5	all	1739644	109697
P03	0226	CMV pool 6	all	1028891	103088
P03	0227	CMV pool 7	all	2202455	154399
P03	0228	CMV pool 8	all	2870789	147115
P11	0238	unstimulated	all	4016000	234555
P11	0239	CMV pool 1	all	1882969	223593
P11	0240	CMV pool 2	all	2402198	160455
P11	0241	CMV pool 3	all	2103641	126658
P11	0242	CMV pool 4	all	2160096	58530
P11	0243	CMV pool 5	all	1744034	70787
P11	0244	CMV pool 6	all	1744828	76739
P11	0245	CMV pool 7	all	1714674	130340
P11	0246	CMV pool 8	all	5604062	456527
P17	0229	unstimulated	all	5309262	427449
P17	0230	CMV pool 1	all	1924212	60416
P17	0231	CMV pool 2	all	1595723	63257
P17	0232	CMV pool 3	all	1901659	50053
P17	0233	CMV pool 4	all	2726130	118721
P17	0234	CMV pool 5	all	2485994	81067
P17	0235	CMV pool 6	all	1602103	79106
P17	0236	CMV pool 7	all	2163102	75709
P17	0237	CMV pool 8	all	2628999	84041
P18	0202	unstimulated	all	4370127	566357
P18	0203	CMV pool 1	all	2712898	123294
P18	0204	CMV pool 2	all	2851549	84229
P18	0205	CMV pool 3	all	2247059	47768
P18	0206	CMV pool 4	all	1623487	33931
P18	0207	CMV pool 5	all	1806345	84803
P18	0208	CMV pool 6	all	1848378	43407
P18	0209	CMV pool 7	all	2299591	46224
P18	0210	CMV pool 8	all	3083352	106600
P19	0211	unstimulated	all	4311069	358585
P19	0212	CMV pool 1	all	2195646	206573
P19	0213	CMV pool 2	all	3493755	233768
P19	0214	CMV pool 3	all	4461729	228553
P19	0215	CMV pool 4	all	1871649	79072
P19	0216	CMV pool 5	all	2612966	284753
P19	0217	CMV pool 6	all	2080054	106067
P19	0218	CMV pool 7	all	2040106	188041
P19	0219	CMV pool 8	all	2124318	154135
P20	0256	unstimulated	all	3047784	197536
P20	0257	CMV pool 1	all	1747643	231810
P20	0258	CMV pool 2	all	1979802	253671
P20	0259	CMV pool 3	all	1675717	210592
P20	0260	CMV pool 4	all	2086259	247566
P20	0261	CMV pool 5	all	1360568	209766

Table A9: (continued)

Donor	Sample ID	Treatment	Cell type	Total productive reads	Number of clonotypes
P20	0262	CMV pool 6	all	1842610	257657
P20	0263	CMV pool 7	all	1907590	270000
P20	0264	CMV pool 8	all	2145825	241465
P21	0265	unstimulated	all	3979008	172574
P21	0266	CMV pool 1	all	1856237	47622
P21	0267	CMV pool 2	all	1515965	94851
P21	0268	CMV pool 3	all	2134357	38911
P21	0269	CMV pool 4	all	1724689	32672
P21	0270	CMV pool 5	all	1984094	51092
P21	0271	CMV pool 6	all	1695427	26804
P21	0272	CMV pool 7	all	1452737	40025
P21	0273	CMV pool 8	all	2222118	54682
Statistics					
Median reads:		2102096		Median clonotypes:	110128.5
Min. reads:		1028891		Min. clonotypes:	15767
Max. reads:		5604062		Max. clonotypes:	709835