Risks and potential benefits of adoptively transferred virus-specific T cells

Huisman, W.


BENEFITS

CHAPTER 4

Public T-cell receptors (TCRs) revisited by analysis of the magnitude of identical and highly-similar TCRs in virus-specific T-cell repertoires of healthy individuals

Wesley Huisman
Lois Hageman
Didier A.T. Leboux
Alexandra Khmelevskaya
Grigory A. Efimov
Marthe C.J. Roex
Derk Amsen
J.H.Frederik Falkenburg
Inge Jedema

Frontiers in Immunology 2022; Vol.13

ABSTRACT

Since multiple different T-cell receptor (TCR) sequences can bind to the same peptide-MHC combination and the number of TCR-sequences that can theoretically be generated even exceeds the number of T cells in a human body, the likelihood that many public identical (PUB-I) TCR-sequences frequently contribute to immune responses has been estimated to be low. Here, we quantitatively analyzed the TCR-repertoires of 190 purified virus-specific memory T-cell populations, directed against 21 epitopes of Cytomegalovirus, Epstein-Barr virus and Adenovirus isolated from 29 healthy individuals, and determined the magnitude, defined as prevalence within the population and frequencies within individuals, of PUB-I TCR and of TCR-sequences that are highly-similar (PUB-HS) to these PUB-I TCR-sequences. We found that almost one third of all TCR nucleotide-sequences represented PUB-I TCR amino-acid (AA) sequences and found an additional 12% of PUB-HS TCRs differing by maximally 3 AAs. We illustrate that these PUB-I and PUB-HS TCRs were structurally related and contained shared core-sequences in their TCR-sequences. We found a prevalence of PUB-I and PUB-HS TCRs of up to 50% among individuals and showed frequencies of virus-specific PUB-I and PUB-HS TCRs making up more than 10% of each virus-specific T-cell population. These findings were confirmed by using an independent TCR-database of virus-specific TCRs. We therefore conclude that the magnitude of the contribution of PUB-I and PUB-HS TCRs to these virus-specific T-cell responses is high. Because the T cells from these virus-specific memory TCR-repertoires were the result of successful control of the virus in these healthy individuals, these PUB-HS TCRs and PUB-I TCRs may be attractive candidates for immunotherapy in immunocompromised patients that lack virus-specific T cells to control viral reactivation.

INTRODUCTION

Human virus-specific $CD8^{pos}$ T cells express heterodimeric alpha(α)/beta(β) TCRs that can specifically recognize viral peptides presented by HLA-class-I molecules(1). The TCRα- and the TCRβ-chain repertoires are highly variable due to the genetic recombination process involved in their generation. For the TCRβ-chains, recombination of 1 of 48 functional T-cell Receptor Beta Variable (TRBV), 1 of 2 functional T-cell receptor Beta Diversity (TRBD) and 1 of 12 functional T-cell Receptor Beta Joining (TRBJ) gene segments leads to a V-D-J reading frame(2). The TCRα-chains are generated by a similar recombination process with the exception of a diversity gene, resulting in a V-J reading frame(3). Insertion of template-independent nucleotides between the recombined segments (junctional region) results in a significant further increase in variability(4). The sequence around these junctions encodes for the Complementary Determining Region 3 (CDR3), a loop that reaches out and interacts with a peptide embedded in an HLA molecule, together with the loops of the CDR1 and CDR2 regions, which are fixed within the germline variable gene sequence(5, 6). It has been calculated that these gene rearrangements could potentially generate a repertoire of 10^{15}-10^{20} unique TCRs that may interact with all possible peptide-HLA complexes(7).

Pathogenic viruses like Cytomegalovirus (CMV), Epstein-Barr virus (EBV) and Adenovirus (AdV) can infect humans for life by staying latently present in target cells after a primary infection. In healthy individuals these latent viruses are controlled by the virus-specific T cells. As a result, reactivations of these latent viruses are observed frequently, but do not result in severe virus-associated disorders like malignancies and/or organ failure. However, in the absence of a competent immune system, these latent viruses remain uncontrolled and are associated with high-morbidity and mortality in immune-compromised patients, including patients after stem cells or organ tranplantation(8, 9). To control these viruses, antigen-experienced (central-memory and effector-memory) virus-specific T cells have to develop from the naïve T-cell repertoire. Due to the high diversity of the naïve T-cell repertoire(10), T-cell responses against the many potential viral epitopes presented in multiple HLA alleles may be composed of a large variety of different TCRs. Indeed, when naïve umbilical cord blood-derived T cells were stimulated *in vitro* to generate *de novo* responses against proteins from CMV or Human Immunodeficiency Virus (HIV), this resulted in responding virus-specific T-cell populations with a highly diverse repertoire of TCRs, recognizing many different CMV(11) or HIV-derived peptides(12). However, from *ex vivo* analyses in adults it became clear that *in vivo* the virus-specific memory T-cell populations are shaped during control and clearance of the infection and target only a limited number of viral-peptides, as was shown for T-cell populations specific for viruses like CMV(13), EBV(14), AdV(15), Influenza A(16) and also more recently SARS-Cov-2(17, 18). Nevertheless, the multiple viral-peptides that are targeted in the

various HLA alleles make it theoretically unlikely that individuals would frequently share exactly the same virus-specific TCR, unless T cells expressing certain TCRs would favor control of infections and would therefore dominate the responses. Since latent viruses like CMV, EBV and AdV are not fully eradicated, reactivations are frequent and trigger the expansion of antigen-experienced virus-specific memory T cells. This unique biology might contribute to the favor and skewing of specific TCRs expressed by T cells that dominate the response and control the virus.

Evidence for selection of certain virus-specific TCR-expressing T cells in controlling viruses has come from several reports identifying identical TCR amino-acid (AA) sequences in dominant virus-specific memory T-cell populations in different individuals, designated as public TCR-sequences (from here on referred to as public-identical (PUB-I) TCR-sequences). Most studies investigated the presence of PUB-I TCR-sequences in TCRβ-chains, since this is the most diverse TCR chain and the CDR3β sequence of CD8 T cells is positioned to interact with the antigenic peptide presented by HLA-class-I molecules. However, dominant virus-specific memory T-cell populations with PUB-I TCRα chains have also been described previously(19). Such PUB-I TCR sequences are most overtly observed in antigen-experienced memory virus-specific T cells due to the *in vivo* antigen-driven proliferation, but are also present within the naïve T-cell compartment, although at low frequency(20). PUB-I TCRβ sequences have been found in T-cell populations specific for latent viruses like CMV-specific T-cell responses(21-23), EBV-specific T-cell responses(24, 25), but also for non-latent viruses like Influenza-specific T-cell responses(16), respiratory syncytial virus-specific T-cell responses(26) and SARS-Cov-2-specific T-cell responses(17, 27). In addition, some of these virus-specific T-cell populations also contained TCR AA-sequences that were highly-similar to the identical shared TCR AA-sequence (from here on referred to as highly-similar to PUB-I (PUB-HS) TCR-sequences). However, the magnitude, defined as prevalence within the population and frequencies within virus-specific T-cell repertoires, of PUB-I and PUB-HS TCR-sequences is not known. A high probability to be generated during V-D-J recombination may play a role(28), but since virus-specific memory T-cell repertoires in the circulation are shaped based on antigen encounter and subsequent proliferation, the PUB-I and PUB-HS TCR-sequences most likely reflect highly functional T cells capable of antigen-driven proliferation.

We hypothesize that frequent induction of antigen-driven proliferation of virus-specific T cells targeting frequently reactivating latent viruses will increase the prevalence and frequencies of PUB-I and PUB-HS TCR-sequences within the repertoire of antigen-experienced virus-specific T-cells. Molecular analysis of these TCRs will add in the analysis of the development, presence and quality of memory T-cell responses, and tracking of virus-specific T-cell responses. Furthermore, Identification of dominant TCRs

with shared core-sequences may be utilized for the design of future immunotherapy purposes including TCR-gene transfer. Therefore, the aim of our study was to quantitively analyze the magnitude of PUB-I and PUB-HS TCRβ-sequences within the antigen-experienced virus-specific TCR-repertoires of CMV, EBV and AdV-specific $CD8^{pos}$ memory T cells. We confirmed that healthy individuals generate many different virus-specific TCRs, illustrated by the >3000 TCR nucleotide-sequences that were found *ex vivo* in virus-specific memory T-cell populations. However, a significant part of the virus-specific TCR-repertoires contained PUB-I and PUB-HS TCR nucleotide-sequences. The AAs of these PUB-HS TCRs varied on specific positions in the CDR3β-region, while maintaining a conserved core-AA-sequence that was also present in the respective PUB-I TCR.. We identified conserved TCR core-AA-sequences for each specificity that could be used for diagnostic purposes looking at anti-viral immune responses. Additionally, PUB-I or PUB-HS TCRs with the highest frequencies in healthy individuals may be utilized to develop off-the-shelf immunotherapeutics (using TCR-gene transfer) to effectively control CMV, EBV or AdV-infections or reactivations in immunocompromised patients.

MATERIALS AND METHODS

Collection of donor material

After informed consent according to the Declaration of Helsinki, healthy individuals (homozygously) expressing HLA-A*01:01 and HLA-B*08:01 or HLA-A*02:01 and HLA-B*07:02 were selected from the Sanquin database and the biobank of the department of Hematology, Leiden University Medical Center (LUMC). Peripheral blood mononuclear cells (PBMCs) were isolated by standard Ficoll-Isopaque separation and used directly or thawed after cryopreservation in the vapor phase of liquid nitrogen. Donor characteristics (HLA typing, CMV and EBV serostatus) are provided in **Table 1**.

Generation of peptide-MHC complexes to isolate virus-specific T cells

All viral peptides were synthesized in-house using standard Fmoc chemistry. Recombinant HLA-A*01:01, HLA-A*02:01, HLA-B*07:02 and HLA-B*08:01 heavy chain and human β2m light chain were in-house produced in Escherichia coli. MHC-class-I refolding was performed as previously described with minor modifications(29). Major histocompatibility complex (MHC)-class-I molecules were purified by gel-filtration using HPLC. Peptide-MHC(pMHC) tetramers were generated by labeling biotinylated pMHC-monomers with streptavidin-coupled phycoerythrin (PE; Invitrogen, Carlsbad, USA), allophycocyanin (APC, Invitrogen), brilliant violet 421 (BV421, Becton Dickinson (BD), Franklin Lakes, USA), brilliant violet 510 (BV510, BD) or peridinin-chlorophyll-protein complex (PerCP, Invitrogen). Complexes were stored at 4 °C. Formation of stable pMHC-monomers was performed using UVexchange technology(30) and according to a

previously described protocol(31).

Isolation and expansion of virus-specific T cells

Phycoerythrin (PE), allophycocyanin (APC), BV421, BV510 and/or peridinin-chlorophyll-protein (PerCP)-labeled pMHC-tetramer complexes were used for fluorescence-activated cell sorting (FACSorting). The pMHC-tetramers used are shown in Table 2. Per specificity, $30*10^6$ PBMCs were first incubated with pMHC-tetramers at 4°C for 30 min, followed by labeling with APC-H7 CD8 (BD) and fluorescein isothiocyanate-labeled (FITC) CD4 and CD14 (BD) antibodies at 4°C for 30 min. PeptideMHC-tetramer positive, $CD8^{pos}/CD4^{neg}$ T cells were FACsorted and seeded at 10,000 cells per well in U-bottom microtiter plates for the generation of bulk T-cell populations (**Supplementary Figure 1**). Peptide-MHC-tetramer^{pos} virus-specific T cells, targeting a single antigen, were first specifically expanded in the presence of 10^{-7}M of the specific peptide in T-cell medium: Iscove's Modified Dulbecco's Medium (IMDM; Lonza, Verviers, Belgium) containing 5% heat-inactivated fetal bovine serum (FBS; Invitrogen), 5% heat-inactivated human serum (ABOS; Sanquin Reagents, Amsterdam, The Netherlands), 100U/mL penicillin (Lonza), 100µg/mL streptavidin (Lonza) , 2.7mM L-glutamine (Lonza), and 100IU IL-2/mL (Chiron, Emeryville, USA) and with 5-fold 35 Gy irradiated autologous PBMCs as feeder cells. Initial specific stimulation and expansion with 10^{-6}M peptide was performed to stimulate preferential outgrowth of pMHC-tetramer^{pos} T cells. After two weeks of culture, pMHC-tetramer^{pos} T-cell populations were qualified as pure populations when they contained ≥97% pMHC-tetramer^{pos} cells. Sorting was performed on a FACS ARIA (BD) and analyzed using Diva software (BD). All analyses were performed on a FACS Calibur (BD), and analyzed using Flowjo Software (TreeStar, Ashland, USA).

TCRβ-library preparation

TCRβ-sequences were identified using ARTISAN PCR adapted for TCR PCR(32, 33). Total mRNA was extracted from 190 pMHC-tetramer^{pos} purified (**Supplementary Figure 2A**) virus-specific T-cell populations(34) using magnetic beads (Dynabead mRNA DIRECT kit; Invitrogen, Thermo Fisher Scientific). Ten µl (~1µg) of mRNA per sample was mixed with TCRβ constant region-specific primers (1µM final concentration) and SmartSeq2modified template-switching oligonucleotide (SS2m_TSO; 1µM final concentration) and denatured for 3 minutes at 72°C. After cooling, cDNA was synthesized for 90 minutes at 42°C with 170 U SMARTscribe reverse transcriptase (Takara, Clontech) in a total volume of 20µl containing 1.7U/µl RNasin (Promega), 1.7mM DTT (Invitrogen, Thermo Fisher Scientific), 0.8mM each of high-purity RNAse-free dNTPs (Invitrogen, Thermo Fisher Scientific) and 4µl of 5x first-strand buffer. During cDNA synthesis, a non-templated 3'polycytosine terminus was added (**Supplementary Figure 2B**), which created a template for extension of the cDNA with the TSO(35). PCR (2min at 98°C followed by 40 cycles of [1s at 98°C, 15s at 67°C, 15s at 72°C], 2 min at 72°C) of 5µl of cDNA was then performed using

Phusion Flash (Thermo Fisher Scientific) with anchor-specific primer (SS2m_For; 1µM final concentration) and each (1µM final concentration) of the nested primers specific for the constant regions of TCRβ constant 1 and TCRβ constant 2. Both forward and reverse PCR primers contained overhanging sequences suitable for barcoding. Amplicons were purified and underwent a second PCR (2min at 98°C followed by 10 cycles of [1s at 98°C, 15s at 65°C, 30s at 72°C], 2 min at 72°C) using forward and reverse primers (1µM final concentration) with overhanging sequences containing identifiers (sequences of 6 base-pairs) and adapter sequences appropriate for Illumina HiSeq platforms (or PacBio; Pacific Biosciences). Unique identifiers were used for each T-cell population targeting one epitope. Forward or reverse identifiers were shared between T-cell populations targeting different epitopes. For all primer sequences see **Supplementary Table 1**. For identifier sequences see **Supplementary Table 2**. Amplicons with identifiers were purified, quantified and pooled into one library for paired-end sequencing of 150bp on an Illumina HiSeq4000. Deep sequencing was performed at GenomeScan (Leiden, The Netherlands). Raw data were de-multiplexed and aligned to the matching TRBV, TRBD, TRBJ and constant (TRBC) genes. CDR3β-sequences were built using MIXCR software using a bi-directional approach (5'-3' and 3'-5' read)(36). CDR3β-sequences with a stop-codon were removed from the library. Bi-directional readings using MIXCR could result in out-of-frame CDR3β AA-sequences due to the even number of nucleotides. These sequences (n=392) were manually aligned with the germline TRBV and TRBJ-sequence. CDR3β-sequences were further processed using custom scripts in R to compare specificities and sharing of CDR3β-sequences.

Computational unbiased repertoire analysis

The following R-packages were used in R-software to generate a nodal plot of CDR3β AA-sequences with the levenshtein distance as parameter for similarity: "igraph" to create network objects, obtain the degree of a node and its betweenness(37), "data.table" to organize CDR3β-sequences; "stringdist" to calculate Levenshtein distances(38), "Biostrings" for fast manipulation of large biological sequences or sets of sequences(39), "dplyr" to arrange and filter data(40), "tibble" for providing opinionated data frames, "ggplot2" for generating figures(41) and "RColorBrewer" to create graphics(42). A levenshtein distance of 0.25 was added to visualize multiple identical sequences. Nodes with identical sequences (levenshtein distance of 0.25) were manually replaced by pie-charts using Adobe Illustrator CC 2018.

Table 1. HLA typing and CMV/EBV-serostatus of healthy donors.

#	Age	CMV	EBV	HLA-A		HLA-B		HLA-C		HLA-DR		HLA-DQ		HLA-DP	
1	62	Pos	Pos	01:01	30:01	08:01	13:02	07:01	06:02	03:01	07:01	02:01	02:02	04:01	09:01
2	48	Pos	Pos	01:01		08:01		07:01		03:01	01:02	02:01	05:01	01:01	04:01
3	53	Pos	Pos	01:01		08:01		07:01		03:01	15:01	02:01	06:02	N.D	
4	44	Pos	Pos	01:01		08:01		07:01		03:01		02:01		04:01	
5	65	Pos	pos	01:01	30:01	08:01	13:02	07:01	06:02	04:07	15:01	03:01	06:02	04:01	
6	26	Pos	Pos	01:01		08:01		07:01		03:XX		02:XX		N.D	
7	42	Pos	pos	01:01	30:01	08:01	13:02	07:01	06:02	03:01	04:01	02:01	03:01	04:01	
8	41	Pos	pos	02:01	24:02	08:01	35:01	07:01	11:01	02:02	03:01	02:02	03:01	02:01	13:01
9	27	Neg	Pos	01:01		08:01		07:01		03:01		02:01		04:01	
10	48	Neg	Pos	01:01		08:01		07:01		03:01		02:01		04:01	05:01
11	35	Neg	Pos	01:01		08:01		07:01		03:01		02:01		01:01	09:01
12	49	Neg	Pos	01:01		08:01		07:01		03:01		02:01		04:01	05:01
13	50	Neg	Pos	01:01		08:01		07:01		03:01		02:01		04:01	04:02/01
14	57	Pos	Pos	01:01	68:01	08:01	35:01	04:01	07:01	01:01	03:01	02	05:01	04:01	04:02
15	32	Neg	Pos	01:01		08:01		07:0		03:XX		02:XX		N.D	
16	25	Neg	pos	01:01	24:02	08:01	35:01	04:01	07:01	03:01	08:01	02:01	04:02	04:01	
17	38	Neg	Pos	01:01		08:01		07:01		03:XX		02:XX		N.D	
18	N.K.	Pos	Pos	02:01		07:02	44:02	07:02	05:01	15:01	04:01	06:02	03:01	04:XX	02:01
19	64	Pos	Pos	02:01		07:02		07:02		15:01		06:02		04:01	
20	33	Pos	Pos	02:01	03:01	07:02		07:02		15:01		06:02		04:01	03:01
21	53	Pos	Pos	02:01		07:02		07:02		15:01		06:02		04:01	05:01
22	62	Pos	Pos	02:01		07:02		07:02		15:01		06:02		02:01	04:01
23	43	Pos	Pos	02:01		07:02		07:02		15:01		06:02		04:01	
24	50	Pos	Pos	02:01	03:01	07:02	44:02	07:02	05:01	15:01	01:01	06:02	05:01	04:01	14:01
25	50	Pos	Pos	02:01		07:02		07:02		07:01	15:01	03:03	06:02	04:01	13:01
26	63	Pos	Pos	02:01		07:02		07:02		15:XX		06:XX		N.D	
27	29	Neg	Pos	02:01		07:02		07:02		15:01		06:02		04:01	

Table 1. Continued.

#	Age	CMV	EBV	HLA-A		HLA-B		HLA-C		HLA-DR		HLA-DQ		HLA-DP	
28	26	Neg	Pos	02:01		07:02		07:02		15:01		06:02		04:01	13:01
29	52	Neg	Pos	02:01		07:02		07:02		15:XX		06:XX		N.D	

Virus-specific T cells restricted to HLA-A*01:01/HLA-B*08:01 or HLA-A*02:01/HLA-B*07:02 were isolated from donors #1-17 and donors #18-29, respectively. CMV and EBV serostatus and age are indicated for each donor. HLA typing was determined either by serology, where the second digits could not be determined (XX), or with high resolution HLA typing unless indicated by N.D.. Blanks indicate homozygosity for the given allele. For one of the unrelated stem cell donors the age was not known (N.K.).

Sequence logo plots

To identify which positions of PUB-I and PUB-HS CDR3β AA-sequences were conserved and which were variable, all CDR3β AA-sequences with the most frequent CDR3 length were included and the AAs were stacked for each position in the sequence. The overall height of the stacks indicates the sequence conservation at that position, while the height of symbols within the stacks indicates the relative frequency of each AA at that position. AAs have colors according to their chemical properties; polar AAs (G, S, T, Y, C, Q, N) show as green, basic (K, R, H) blue, acidic (D, E) red, and hydrophobic (A, V, L, I, P, W, F, M) AAs as black(43).

Generation of independent TCR-database from virus-specific T-cell products generated for a clinical trial

As an independent TCR-database containing TCR-sequences from virus-specific T cells, we used the information obtained from the virus-specific T-cell products generated in the context of the phase I/II safety and feasibility study T Control (EudraCT-number 2014-003171-39) using the MHC-I-Streptamer isolation technology (Juno Therapeutics, Munich, Germany)(20, 44). Sequencing was performed as described above for all virus-specific T-cell populations per donor, resulting in unique identifiers for all virus-specific T-cell populations in the TCRβ-library.

Data deposition

TCR-Sequencing data is deposited to the Sequence Read Archive (SRA); submission: SUB10993301 (access to bioproject: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA803981)

RESULTS

Generation and validation of a library of TCR-sequences derived from FACsorted virus-specific T-cell populations

To examine the composition of the virus-specific TCR-repertoires in different individuals, the CDR3β-regions of purified expanded pMHC-tetramer-binding virus-specific T-cell populations were sequenced (**Supplementary Figure 2**). We analyzed the TCR-repertoires of CMV, EBV and AdV-specific T cells, restricted to four prevalent HLA alleles (HLA-A*01:01, HLA-A*02:01, HLA-B*07:02 and HLA-B*08:01) and specific for 21 different peptides (**Table 2**). Purified CMV, EBV and AdV-specific T-cell populations targeting CMV (n=8), EBV (n=10) or AdV (n=3)-derived peptides were isolated from 17 HLA-A*01:01/B*08:01^{pos} individuals and 12 HLA-A*02:01/B*07:02^{pos} individuals (**Table 1 and Table 2**). In total, 190 virus-specific T-cell populations, each targeting a single viral epitope, were successfully isolated and showed high purity (>97% pMHC-tetramer positive).

The mean precursor frequencies of the different T-cell specificities in the starting PBMC materials are shown in **Supplementary Figure 3**. Sequencing of the CDR3β-regions of these virus-specific T-cell populations resulted in 3346 CDR3β nucleotide-sequences that occurred at frequencies of more than 0.1%. 135 of these nucleotide-sequences were present at high frequencies (>5%) within one specificity, but were also found at low frequencies (around 0.1%) in another specificity, indicating contamination due to FACSorting impurities. These low frequency nucleotide-sequences were discarded from further analysis. 41 nucleotide-sequences were present at low frequencies in two unrelated specificities and could not be correctly annotated and these 41 duplicates (n=82) were also discarded from the analysis. Therefore, a total of 3129 nucleotide-sequences could be annotated. In total, 2224 (71%) of these nucleotide-sequences represented unique CDR3β AA-sequences that were found in only one individual and 905 nucleotide-sequences (29%) resulted in 131 different PUB-I CDR3β AA-sequences that were found in two or more unrelated individuals (**Flowchart; Figure 1**).

Table 2. Number of isolated virus-specific T-cell populations

				# of isolated T-cell populations / # of attempted isolations (%)		
Virus	**Antigen**	**HLA**	**Peptide**	**Per specificity**	**Per virus**	**Per HLA**
CMV	pp50	HLA-A*01:01	VTEHDTLLY	7/8 (88%)	**CMV:** 53/70 *(76%)*	**HLA-A*01:01:** 29/39 *(74%)*
	pp65	HLA-A*01:01	YSEHPTFTSQY	6/8 (75%)		
	pp65	HLA-A*02:01	NLVPMVATV	8/9 (89%)		
	IE-1	HLA-A*02:01	VLEETSVML	5/9 (56%)		
	pp65	HLA-B*07:02	TPRVTGGGAM	8/9 (89%)		
	pp65	HLA-B*07:02	RPHERNGFTVL	8/9 (89%)		**HLA-A*02:01:** 70/90 *(78%)*
	IE-1	HLA-B*08:01	ELRRKMMYM	5/9 (56%)		
	IE-1	HLA-B*08:01	QIKVRVDMV	6/9 (67%)		
EBV	LMP2	HLA-A*01:01	ESEERPPTPY	4/6 (67%)	**EBV:** 111/129 (86%)	
	LMP2	HLA-A*02:01	FLYALALLL	11/12 (92%)		
	LMP2	HLA-A*02:01	CLGGLLTMV	10/12 (83%)		
	EBNA3C	HLA-A*02:01	LLDFVRFMGV	7/12 (58%)		**HLA-B*07:02**: 33/42 *(79%)*
	BMLF1	HLA-A*02:01	GLCTLVAML	10/12 (83%)		
	BRLF1	HLA-A*02:01	YVLDHLIVV	11/12 (92%)		
	EBNA3A	HLA-B*07:02	RPPIFIRRL	11/12 (92%)		
	BZLF1	HLA-B*08:01	RAKFKQLL	17/17 (100%)		
	EBNA3A	HLA-B*08:01	FLRGRAYGL	13/17 (76%)		**HLA-B*08:01:** 58/69 *(84%)*
	EBNA3A	HLA-B*08:01	QAKWRLQTL	17/17 (100%)		
AdV	HEXON	HLA-A*01:01	TDLGQNLLY	12/17 (71%)	**AdV:** 26/41 (63%)	
	E1A	HLA-A*02:01	LLDQLIEEV	8/12 (67%)		
	HEXON	HLA-B*07:02	KPYSGTAYNAL	6/12 (50%)		

Seventeen donors were used to isolate HLA-A*01:01/B*08:01-restricted virus-specific T cells and 12 donors were used to isolate HLA-A*02:01/B*07:02-restricted virus-specific T cells.
CMV, Cytomegalovirus; EBV, Epstein-Barr virus; AdV, Adenovirus

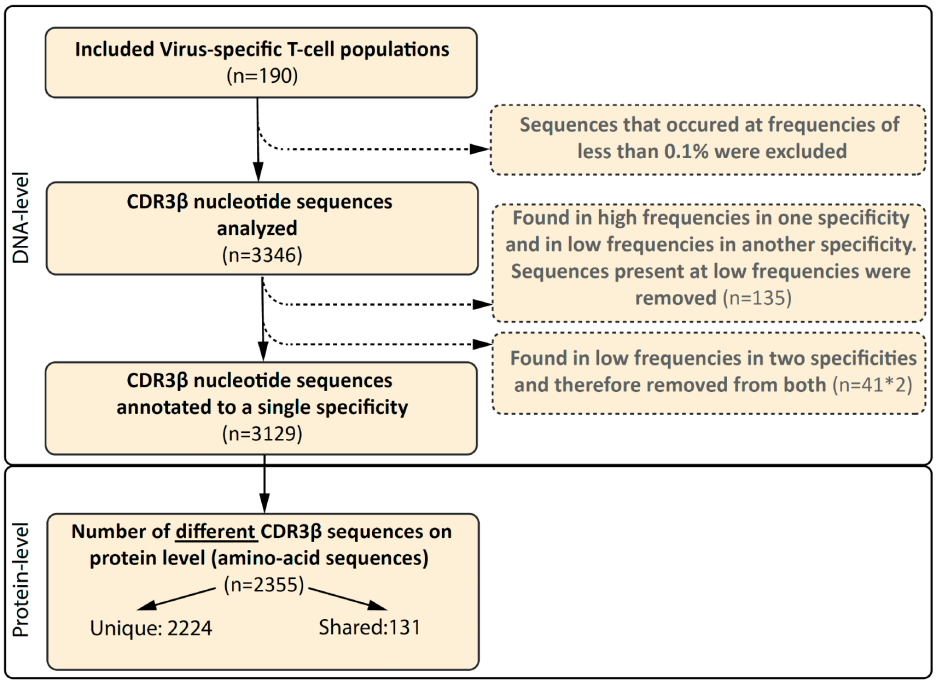


Figure 1. Flowchart of included and excluded CDR3β nucleotide and AA-sequences. In total, 190 different virus-specific T-cell populations were FACsorted using pMHC-tetramers, followed by a short-term *in vitro* stimulation. The CDR3β nucleotide-sequences were determined using next-gen Illumina sequencing. CDR3β nucleotide-sequences that occurred at a frequency of less than 0.1% in each sample were excluded. CDR3β nucleotide-sequences that were identical and present in two different specificities, but present at high frequencies in one specificity, were only removed from the specificity that contained the sequences at very low frequencies (0.1-0.5%; n=135). CDR3β nucleotide-sequences that were identical and present in two different specificities at low frequency were considered contamination and removed from the library (82 sequences, 41 different-sequences). The numbers of different CDR3β AA-sequences that were encoded by the CDR3β nucleotide-sequence are shown at protein level. We then assessed how many CDR3β-AA-sequences were found in multiple individuals (shared) and how many were only found in a single individual (unique).

To investigate the relationship between the numbers of CDR3β nucleotide-sequences and the translated number of CDR3β AA-sequences, we compared the nucleotide-sequences of the 131 PUB-I CDR3β AA-sequences found in different individuals. Different nucleotide-sequences can result in the same CDR3β AA-sequence, a phenomenon known as convergent recombination. We found that PUB-I CDR3β AA-sequences present at high (representative example**; Figure 2A**) or low frequencies (range 0.1-1%) (representative example; **Figure 2B**) could be encoded by different CDR3β nucleotide-sequences in the junctional regions of the CDR3β-regions in TCRs of T cells isolated from different individuals (**Figure 2C and 2D; Supplementary Figure 4**). Because the majority of nucleotide-sequences encoding the same CDR3β AA-sequences were different

between individuals, these data exclude contamination as an explanation for the finding of PUB-I CDR3β AA-sequences.

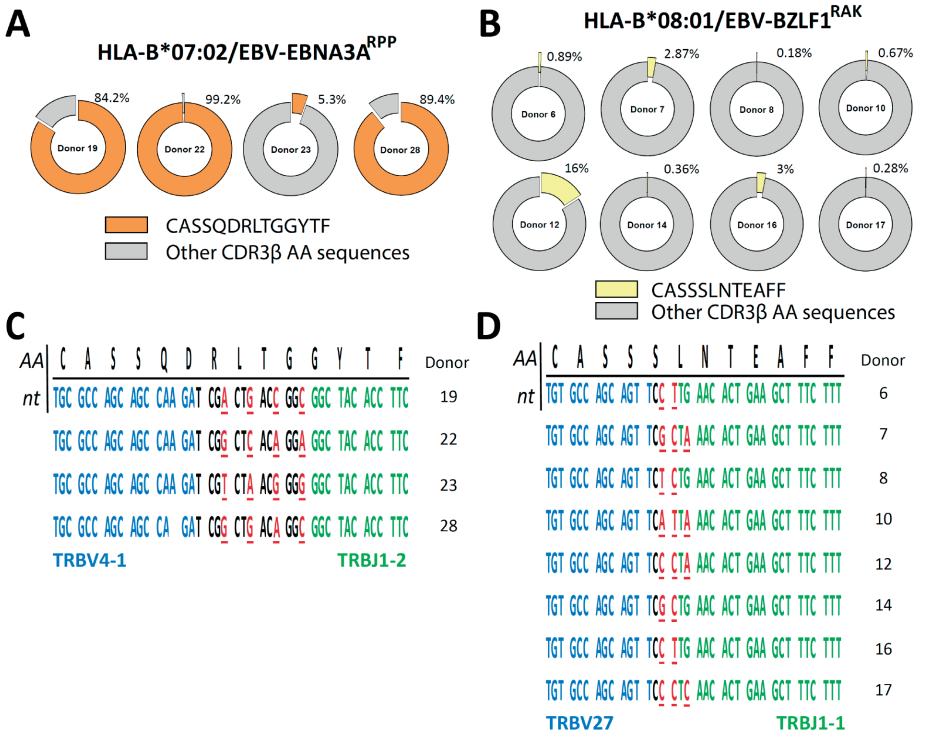


Figure 2. PUB-I CDR3β-sequences can be found in different individuals with small nucleotide-differences as a result of convergent recombination. The library of virus-specific CDR3β AA-sequences contained 131 sequences that were found in multiple different individuals (≥2). The nucleotide-sequences of these CDR3β AA-sequences were analyzed to investigate differences and similarities. **A and B**) Shown are two representative examples of the frequencies of the PUB-I CDR3β AA-sequences CASQDRLTGGYTF and CASSSLNTEAFF, that were specific for HLA-B*07:02-restricted EBV-EBNA3A^{RPP} and HLA-B*08:01-restricted EBV-BZLF1^{RAK}, respectively. **C and D**) Shown are the nucleotide-sequences of the CDR3β AA-sequences CASSQDRLTGGYTF and CASSSLNTEAFF that were shared by different individuals. Underlined nucleotides represent differences between the different individuals. Nucleotides in red represent differences to the consensus sequence. Nucleotide-sequences in blue and green represent perfect alignment with the germline sequences of the TRBV-gene and TRBJ-gene, respectively.
AA; amino-acids, nt; nucleotides

PUB-I and PUB-HS CMV-, EBV- and AdV-specific CDR3β AA-sequences are abundant in virus-specific T-cell populations

We then investigated the distribution of the 131 PUB-I CDR3β AA-sequences within the 21 different specificities and the prevalence among individuals for each of the PUB-I CDR3β AA-sequences per viral-epitope. T cells with PUB-I CDR3β AA-sequences were found

for 19 out of the 21 specificities (**Supplementary Table 3**). PUB-I CDR3β AA-sequences were not observed in AdV-IE1^{LLD} and EBV-LMP2^{ESE}-specific T-cell populations. Some T-cell populations (e.g. EBV-LMP2^{FLY}) contained many different PUB-I CDR3β AA-sequences (n=24) that were all highly-similar. For this reason, we investigated the distribution of PUB-I CDR3β AA-sequences with unique TRBV and TRBJ-gene usage. This resulted in 29 different PUB-I CDR3β AA-sequences, distributed over 19 specificities (**Figure 3A; grey bars**). Six specificities contained two or three different (expressing different TRBV and/or TRBJ-genes) PUB-I CDR3β AA-sequences that were highly prevalent among individuals. To investigate how often these PUB-I CDR3β AA-sequences could be found in our cohort of healthy donors, we quantified the prevalence of each of these 29 PUB-I CDR3β AA-sequence (**Figure 3A; grey bars**). Because we classified a PUB-I CDR3β AA-sequence as being present in at least 2 individuals, the prevalence among donors could not be less than 2 out of 17 (12%; 17 is maximum number of T-cell populations for 1 specificity). Only 4 out of 29 PUB-I CDR3β AA-sequences were found in only 2 individuals (**Figure 3A; grey bars**). Overall, these 29 PUB-I CDR3β AA-sequences had a prevalence of 33% among healthy individuals (median; range 12%-82%). Importantly, most PUB-I CDR3β AA-sequences were found in at least 25% of individuals and 5 were even present in more than half the donors.

We and others hypothesized that the binding/docking of TCRs to HLA-peptide complexes might allow for small changes/flexibility in the CDR3 AA-sequences without significantly changing the conformation or interaction(45). Therefore, we investigated if there were CDR3β AA-sequences present in our data set that were highly-similar (PUB-HS) to PUB-I CDR3β AA-sequences and differed by 1, 2 or 3 AAs. The 2224 unique TCR nucleotide-sequences identified in our previous analysis may contain PUB-HS CDR3β AA-sequences that are in fact part of the same public response as the respective PUB-I TCRs. In total, 379 PUB-HS CDR3β nucleotide-sequences were present that also resulted in 379 PUB-HS CDR3β AA-sequences that differed by 1, 2 or 3 AAs from one of the 131 PUB-I CDR3β AA-sequences. This shows that 41% of the total virus-specific TCR-repertoire contained PUB-I and PUB-HS CDR3β nucleotide-sequences. We investigated if these PUB-HS CDR3β AA-sequences were also present in individuals that did not contain the respective PUB-I CDR3β AA-sequences. PUB-HS CDR3β AA-sequences were present for 21 out of 29 PUB-I CDR3β AA-sequences (**Figure 3A; shaded orange bars**). When we include the PUB-HS CDR3β AA-sequences and quantified the 29 PUB-I and PUB-HS CDR3β AA-sequences, these had a median prevalence of 50% among healthy individuals (range 23%-100%). The AdV-IE1^{LLD} and EBV-LMP2^{ESE}-specific T-cell populations, where PUB-I CDR3β AA-sequences were not found, did contain PUB-HS CDR3β AA-sequences in multiple individuals at high frequencies(46) (**Supplementary Figure 5**). The frequencies of PUB-I combined with PUB-HS CDR3β AA-sequences were relatively high within each virus-specific T-cell population of each individual (**Figure 3B**). The frequencies of all PUB-I and PUB-HS CDR3β AA-sequences ranged from 0.1%-99.4% within the 19 different virus-

specific T-cell populations with a median of 13.1%. When combined, all but one PUB-I plus PUB-HS CDR3β AA-sequences were found in at least 25% of individuals and 3 were even found in over 75% of individuals. These data show that for many PUB-I CDR3β AA-sequences we found sequences that were similar (1, 2 or 3 AA-differences), making up more than 40% of the total virus-specific TCR-repertoire and together these sequences were found in the majority of individuals at high frequencies.

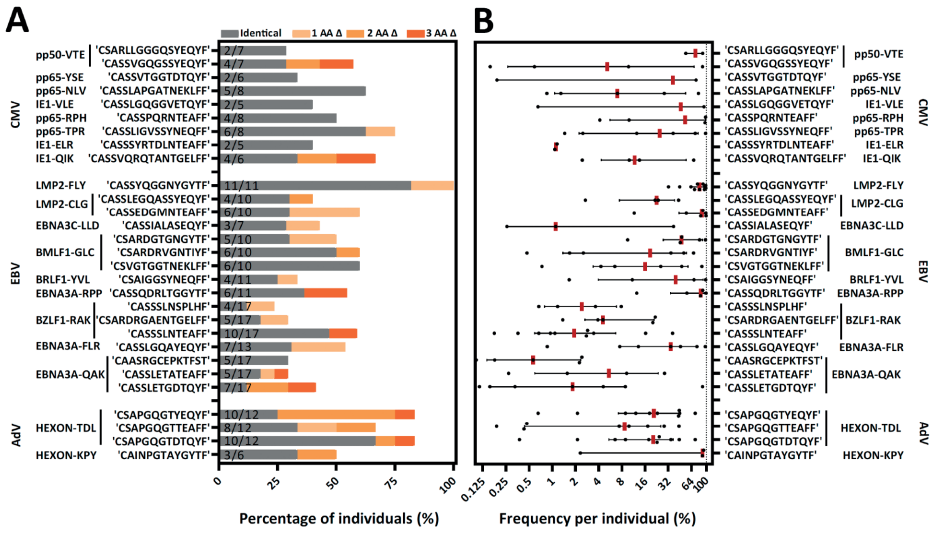


Figure 3. PUB-I and PUB-HS CMV, EBV and AdV-specific CDR3β AA-sequences are common in different individuals and of high frequencies within the T-cell specificities. **A**) Nine-teen different specificities contained PUB-I CDR3β-sequences, shared by at least 2 individuals. Six specificities contained 2 or 3 different (with different TRBV and/or TRBJ-genes) CDR3β AA-sequences that were found frequently in different individuals. The total numbers of different T-cell populations (different donors) that contained the PUB-I or PUB-HS CDR3β AA-sequences are indicated at the inner-side of the y-axis. The occurrence, shown as percentage among healthy individuals, is shown per CDR3β AA-sequence. PUB-I sequences are shown in grey. Individuals containing PUB-HS sequences with 1, 2 or 3 AA-differences (shown in light-dark orange) compared to the PUB-I sequence were stacked on top of the individuals where the PUB-I sequence was already identified. **B**) Shown is the sum of frequencies of the PUB-I and PUB-HS (1, 2 and 3 AA-differences) CDR3β AA-sequences per individual. Each dot is one individual, and the red-lines represent the medians with interquartile ranges.
AA; amino-acids, nt; nucleotides, Δ; difference(s)

Identical and highly-similar CDR3β AA-sequences contain conserved regions in the junctional region

To investigate how the PUB-HS CDR3β AA-sequences related to the PUB-I CDR3β AA-sequences, we analyzed if the PUB-HS (1, 2 or 3 AA-differences) CDR3β AA-sequences showed variations at random positions or at specific positions compared to the PUB-I

CDR3β AA-sequences. We hypothesized that if the binding/docking of PUB-HS TCRs was not significantly different, conserved regions and regions that allow for some variation could be identified in the CDR3β AA-sequences. As Illustrated in **Figure 4A** for the EBV-LMP2^{FLY}-specific PUB-I CDR3β AA-sequence CASSYQGGNYGYTF, two motifs were identified with AA-differences predominantly located at positions 5 and 9/10 of the CDR3β-region. The AAs [QGG] at positions 6-8 were conserved for both motifs. In total, the two motifs consisted of 86 PUB-HS CDR3β AA-sequences with 1 or 2 AA-differences. The majority (n=71) had the same CDR3 length of 14 AAs as the PUB-I CDR3β AA-sequence, implying that variations were caused by AA-substitutions. The remaining 15 PUB-HS CDR3β AA-sequences had a CDR3 length of 15 AAs, due to AA-inserts, compared to the PUB-I CDR3β AA-sequence (**Figure 4B**). Similar rules were found for the other 20 PUB-I CDR3β AA-sequences. Also here, some AA-positions were highly conserved, whereas others were variable. However, the precise locations of the variable AAs differed between specificities (Representative examples; **Figure 4C**). Interestingly, the corresponding CDR3-alpha sequences of a few highly-frequent PUB-I and PUB-HS CDR3β-sequences were also identical or highly-similar between different individuals (**Supplementary table 4**).

As a control, we assessed if these conserved motifs were predictive for the specificity when searching in our database of 2355 unique CDR3β AA-sequences. The requirement was that each motif should not be present in another specificity. We observed that some specificities contained motifs of only 3 or 4 AAs that were exclusive for that specificity and were not observed in any other specificity (**Table 3**). Altogether, these data show that the variations in the CDR3β AA-sequences were not random, but occurred at specific positions that resulted in conserved regions that were predictive for the specificities.

Computational analysis reveals conserved regions in CDR3β AA-sequences despite using different TRBJ-genes

We hypothesized that if the conserved junctional region is a crucial part of the peptide-HLA binding, virus-specific TCR-repertoires could also contain CDR3β AA-sequences with the same conserved region, while allowing different TRBJ-gene usage, as long as the 3-dimensional conformation would allow this. Since the TRBJ-regions often differ by more than 3 AAs, we were not able to include these as PUB-HS CDR3β AA-sequences. Such PUB-HS CDR3β AA-sequences that use different TRBJ-genes might even further increase the prevalence of PUB-I and PUB-HS CDR3β AA-sequences in the virus-specific TCR-repertoire. To investigate this, we performed a computational analysis using the levenshtein-distances (AA-differences) between all different CDR3β AA-sequences. For four different specificities (EBV-LMP2^{FLY}, EBV-EBNA3A^{RPP}, AdV-E1A^{LLD}, and AdV-HEXON^{TDL}) we observed clustering of CDR3β AA-sequences that expressed the same TRBV-genes while using different TRBJ-genes. For example (**Figure 5A**), the HLA-A*02:01-restricted

EBV-LMP2^{FLY}-specific CD8^{pos} T-cell repertoire contained 2 clusters within the cluster of TRBV6-5-expressing T cells (TRBV6-5/TRBJ1-2 and TRBV6-5/TRBJ2-1). The majority of CDR3β AA-sequences within the TRBJ1-2 cluster had a length of 14 AAs, while CDR3β AA-sequences from the TRBJ2-1 cluster had a length of 13 AAs (**Figure 5B**). Analysis of the junctional regions of the TRBV6-5/TRBJ1-2 and TRBV6-5/TRBJ2-1-encoded CDR3β AA-sequences revealed strong conservation of AAs [QGG] on positions 6-8, despite different TRBJ-usage and CDR3 lengths (**Figure 5C**). Similarly, the HLA-A*01:01-restricted AdV-HEXON^{TDL}-specific CD8^{pos} T-cell repertoire contained two large clusters of CDR3β AA-sequences, using TRBV20-1 or TRBV5-1 (**Figure 5D**), all with a CDR3 length of 13 AAs.

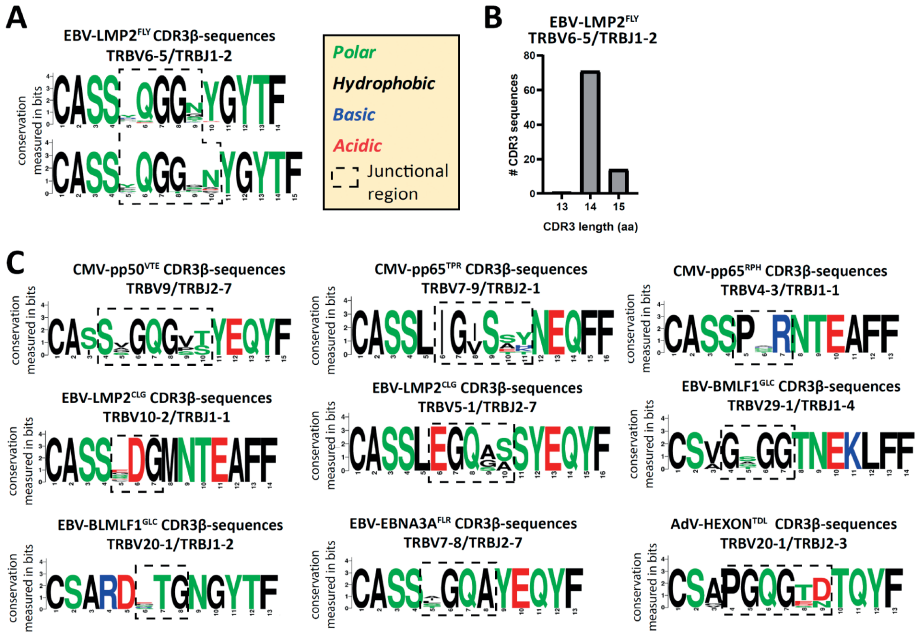


Figure 4. PUB-HS CDR3β AA-sequences show conserved regions and regions with high variability. The levenshtein distance was calculated (i.e. substitution, deletions or insertions of AAs) between CDR3β AA-sequences that express the same TRBV/TRBJ-genes as the PUB-I CDR3β AA-sequence. PUB-HS CDR3β AA-sequences were included with 1, 2 and 3 AA-differences compared to the PUB-I CDR3β AA-sequence. Sequence logos generated using WebLogo (http://weblogo.berkeley.edu/logo.cgi) show the relative frequency of each AA at each given position. The junctional regions (AAs that do not align with the germline TRBV or TRBJ-gene) are shown within the dotted-line box. **A**) The CDR3β AA-sequences specific for HLA-A*02:01-restricted EBV-LMP2^{FLY} with a CDR3 length of 14 and 15 AAs were stacked and show conserved and variable regions in the CDR3β-region. **B**) Shown are the CDR3 length distributions of CDR3β AA-sequences specific for EBV-LMP2^{FLY}-expressing TRBV6-5/TRBJ1-2. **C**) Shown are representative examples of PUB-HS (1, 2 or 3 AA differences) and PUB-I CDR3β AA-sequences that were stacked per specificity and TRBV/TRBJ-usage.

Table 3. Conserved motifs that predict the specificity

Specificity	TRBV	Shared TCR	TRBJ	Motif
CMV-pp50^{VTE}	9	CAS**S**V**GQG**SSYEQYF	2-7	**S**x**GQG**
CMV-pp65^{YSE}	9	CASS**VTGGT**DTQYF	2-3	**VTGGT**
CMV-pp65^{NLV}	7-6	CASS**LAPG**ATNEKLFF	1-4	**LAPG**
CMV-IE1^{VLE}	9	CASSL**GQGGV**ETQYF	2-5	**GQGGV**
CMV-pp65^{TPR}	7-9	CAS**SLIG**V**S**SYNEQFF	2-1	**SLIG**x**S**
CMV-pp65^{RPH}	4-3	CASS**P**Q**RNT**EAFF	1-1	**P**x**RNT**
CMV-IE1^{QIK}	9	CASS**V**Q**R**QT**ANT**GELFF	2-2	**V**x**R**xx**ANT**
CMV-IE1^{ELR}	27	CASSSYR**TDLN**TEAFF	1-1	**TDLN**
EBV-LMP2^{FLY}	6-5	CAS**S**Y**QGG**NYGYTF	1-2	**S**x**QGG**
EBV-LMP2^{CLG}	10-2	CASSE**DGMN**TEAFF	1-1	**DGMN**
	5-1	CASSL**EGQ**AS**S**YEQYF	2-7	**EGQ**xx**S**
EBV-EBNA3C^{LLD}	19	CASS**IAL**ASEQYF	2-7	**IAL**
EBV-BMLF1^{GLC}	20-1	CSA**RDR**V**G**NTIYF	1-3	**RDR**(V/T)**G**
	29-1	CSV**G**T**GGTN**EKLFF	1-4	**G**x**GGTN**
EBV-BRLF1^{YVL}	20-1	CS**AIGGS**YNEQFF	2-1	**AIGGS**
EBV-EBNA3A^{RPP}	4-1	CASSQD**RLTG**GYTF	1-2	**RLTG**
EBV-BZLF1^{RAK}	27	CA**SSSLN**TEAFF	2-1	**SSSLN**
	20-1	CSA**RDR**GA**EN**TGELFF	2-2	**RDR**xx**EN**
EBV-EBNA3A^{FLR}	7-8	CAS**S**L**GQA**YEQYF	2-7	**S**x**GQA**
EBV-EBNA3A^{QAK}	5-1	CASS**LET**GGYGYTF	1-2	**LET**(A/G)
AdV-HEXON^{TDL}	20-1	CSA**PGQ**GTDTQYF	2-3	(A/V)**PGQ**
AdV-E1A^{LLD}	20-1	CS**AR**P**GLA**DEQFF	2-1	**AR**x**GLA**
AdV-HEXON^{KPY}	10-3	CA**INP**GTAYGYTF	1-2	**INP**

We investigated which regions of PUB-I or PUB-HS CDR3β AA-sequences were predictive for the specificity. We searched for each motif in our library of 2355 CDR3β AA-sequences to determine what part of the junctional regions were unique for each specificity, without being present in any other specificity. Underscores with an x represent any of the 20 AAs. Some motifs contain two possible AAs that can be part of the motif which are shown between brackets. The minimum motifs are also shown in bold font in the original CDR3β AA-sequences.

The first cluster (TRBV20-1) contained sub-clusters of CDR3β AA-sequences using TRBJ1-1, TRBJ2-3 or TRBJ2-7 and the second cluster (TRBV5-1) contained CDR3β AA-sequences using TRBJ2-1 or TRBJ2-7. AdV-HEXON^{TDL}-specific CDR3β AA-sequences expressing TRBV20-1 revealed strong conservation of AAs [PGQG] on positions 4-7, which fell outside the region encoded by TRBJ (**Figure 5E**). Additionally, AdV-HEXON^{TDL}-specific CDR3β AA-sequences expressing TRBV5-1 revealed strong conservation of AAs [N__D] on positions 4 and 7, despite different TRBJ-usage. These examples illustrate that virus-specific TCR-repertoires can have conserved CDR3β-regions, while using different TRBJ-genes, allowing substantial variability at specific positions encoded by the TRBJ-region. This will further increase the prevalence of PUB-I and PUB-HS CDR3β AA-sequences in the total virus-specific TCR-repertoire.

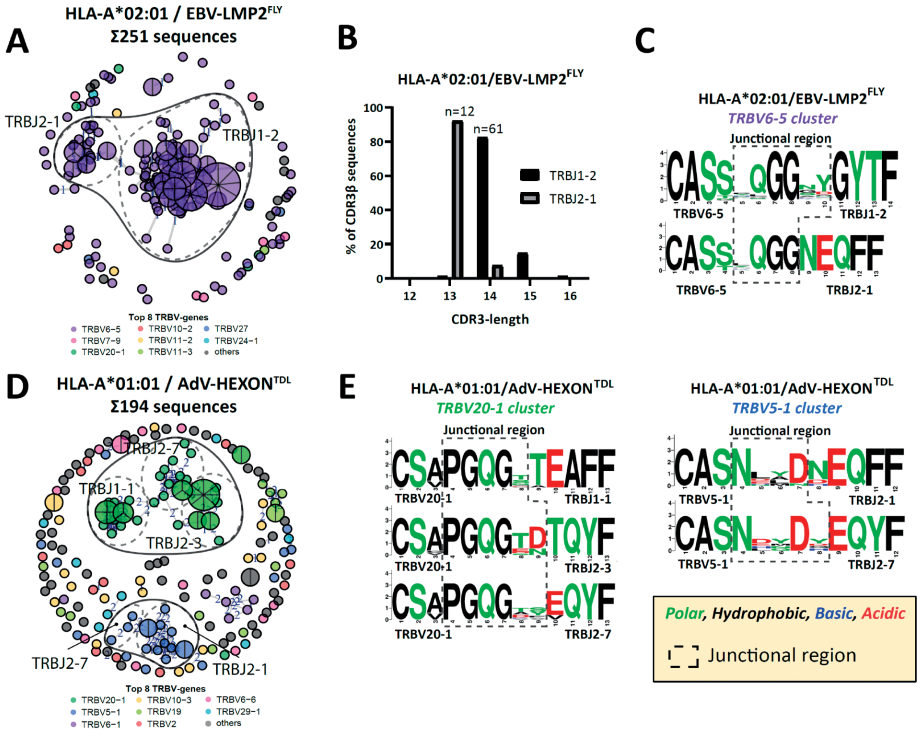


Figure 5. Computation analysis reveals clustering of PUB-HS CDR3β AA-sequences that contained conserved regions in the CDR3β-region. Computational analysis was performed using the levenshtein distance (differences in AAs) between CDR3β AA-sequences of one specificity. CDR3β AA-sequences were plotted and colored according to the top 8 most frequent TRBV-genes and were clustered and linked by a line if they were similar, with a number (levenshtein distance of 1, 2 or 3) representing the differences in AAs. PUB-I CDR3β AA-sequences were plotted as a pie-chart, whereby the size and number of slices indicate in how many individuals this CDR3β AA-sequence was present. Sequence logos generated using WebLogo (http://weblogo.berkeley.edu/logo.cgi) show the relative frequency of each AA at each given position. The junctional region (AAs that do not align with the germline TRBV or TRBJ-gene) are shown within the box with the dotted-line **A**) Shown is a representative example of a virus-specific $CD8^{pos}$ T-cell population, specific for EBV-$LMP2^{FLY}$ with overlapping clusters of sequences that express different TRBJ-genes, while expressing the same TRBV-gene. **B**) The lengths of the CDR3β-regions of the two clusters of EBV-$LMP2^{FLY}$-specific CDR3β-sequences are shown. Varying lengths of the CDR3β-region within a cluster would suggest deletions or insertions, whereby the same length would indicate AA substitutions. **C**) Shown are the sequence motifs of the two EBV-$LMP2^{FLY}$-specific clusters. **D**) Shown is a second representative example of a virus-specific $CD8^{pos}$ T-cell population, specific for AdV-$HEXON^{TDL}$, with overlapping clusters of sequences that express different TRBJ-genes, while expressing the same TRBV-gene. **E**) Shown are the sequence motifs of the TRBV20-1 and TRBV5-1-expressing EBV-$HEXON^{TDL}$-specific clusters.

Individuals with heterozygous HLA backgrounds contain the same shared identical and highly-similar CDR3β AA-sequences

To determine whether the magnitude of PUB-I and PUB-HS CDR3β AA-sequences was particular for our cohort of individuals with a homozygous HLA background, we investigated if the same phenomenon was also present in individuals with a heterogeneous HLA background. We performed the same analyses on virus-specific CD8^{pos} T-cell populations targeting 11 different viral epitopes that were generated and used in the context of a clinical study(20). A total of 1157 CDR3β nucleotide-sequences could be correctly annotated. In total, 695 (61%) nucleotide-sequences resulted in unique CDR3β AA-sequences, that were only found in one individual, and 462 nucleotide-sequences (39%) resulted in 89 different PUB-I CDR3β AA-sequences. From the 695 unique CDR3β AA-sequences, 134 PUB-HS CDR3β nucleotide-sequences were present that differed by 1, 2 or 3 AAs from one of the 89 PUB-I CDR3β AA-sequences.

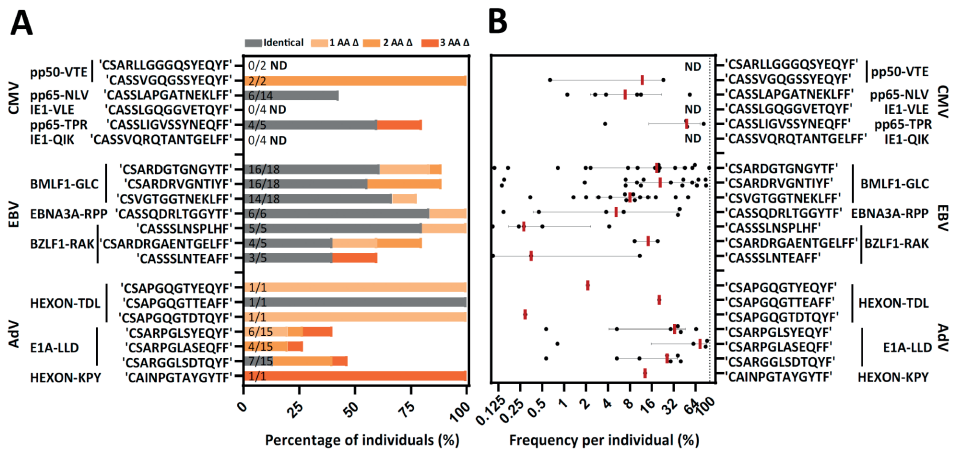


Figure 6. Individuals with a different HLA background from an independent database contain the same PUB-I and PUB-HS CDR3β AA-sequences. Virus-specific T-cell populations targeting 11 different viral-antigens, derived from an independent database, could be evaluated for the occurrence of PUB-I pr PUB-HS CDR3β AA-sequences. **A**) Nine out of the 11 specificities contained the same PUB-I or PUB-HS CDR3β AA-sequences as in our database. The occurrence, shown as percentages among healthy donors, is shown per CDR3β AA-sequence. PUB-I sequences are shown in grey. PUB-HS CDR3β AA-sequences from other individuals with 1, 2 or 3 AA differences, were stacked on top of the already identified public-identical sequences. The total numbers of different T-cell populations (different donors) that contained the PUB-I or PUB-HS CDR3β-sequences are indicated at the inner-side of the y-axis. **B**) Shown are the sum of frequencies of the PUB-I or PUB-HS CDR3β AA-sequences per donor. Each dot is one donor shown and the red-lines represent the medians with interquartile ranges.
AA; amino-acid, Δ; difference(s), ND; Not detected

This shows again that also in this cohort a large part (51%) of the total virus-specific TCR-repertoire contained PUB-I and PUB-HS CDR3β nucleotide-sequences. Because the targeted viral epitopes were not fully identical in both cohorts, we could investigate the prevalence of 20 out of 29 PUB-I CDR3β AA-sequences in this cohort. In total, 17 out of 20 CDR3β AA-sequences that were previously identified, could also be identified in this independent cohort. When we included the PUB-HS CDR3β AA-sequences and quantified the 17 PUB-I and PUB-HS CDR3β AA-sequences, these sequences had a similar high prevalence of a median of 89% among healthy individuals (range 26-100%). (**Figure 6A**). These CDR3β AA-sequences were also present at high frequencies within each virus-specific T-cell population (**Figure 6B**). These data show that the same PUB-I or PUB-HS CDR3β AA-sequences are also present in virus-specific T cells isolated from an independent cohort of individuals with a heterogeneous HLA background with a similar prevalence among donors and frequency within donors.

DISCUSSION

In this study, we quantitatively analyzed the magnitude, defined as prevalence within the population and frequencies within individuals, of public-identical (PUB-I) together with public-highly-similar (PUB-HS) TCRs in TCR-repertoires of CMV, EBV and AdV-specific $CD8^{pos}$ T-cell populations. In total, 2224 (71%) TCR-CDR3β nucleotide-sequences resulted in unique CDR3β AA-sequences, and 905 nucleotide-sequences (29%) resulted in 131 different PUB-I CDR3β AA-sequences that were found in two or more unrelated individuals. These PUB-I CDR3β AA-sequences were distributed over 19 out of 21 virus-specificities and contained 29 different PUB-I CDR3β AA-sequences that were often found in multiple individuals at high frequencies. The virus-specific T-cell populations additionally contained 12% PUB-HS CDR3β AA-sequences, which differed by 1, 2 or 3 AAs compared to the respective PUB-I CDR3β AA-sequences. PUB-HS CDR3β AA-sequences could be found in virus-specific T-cell populations of individuals who did not contain the PUB-I CDR3β AA-sequence as well as of individuals who already contained the PUB-I CDR3β AA-sequence. Analysis of the PUB-I and PUB-HS CDR3β AA-sequences revealed strong conservation of specific AA motifs in the junctional region together with variability of AAs at specific positions at the TRBV/TRBD- and/or TRBD/TRBJ-border regions. Positions with high variability were often adjacent to or even interspersed with the conserved motif. The conserved motifs that we identified were unique for each specificity, and could not be identified in any other specificity in our database. This makes it very likely that these motifs are important for binding of the TCRs to the peptide-HLA complexes. Combined, 41% of the total virus-specific TCR-repertoire consisted of PUB-I and PUB-HS CDR3β nucleotide-sequences. These findings were based on virus-specific T-cell populations derived from two homogeneous donor cohorts that homozygously

expressed HLA-A*01:01/HLA-B*08:01 or HLA-A*02:01/HLA-B*07:02. However, we found similar high percentages (51%) of PUB-I and PUB-HS CDR3β nucleotide-sequences within virus-specific T-cell populations from healthy donors with heterogeneous HLA-backgrounds that were generated for a recent clinical study(20). These dominant PUB-I and PUB-HS TCRs probably are a reflection of the viral-antigen-specific T-cell responses that most optimally encountered the peptide-HLA complexes on the infected target cells and could be utilized for the design of future immunotherapy purposes including TCR-gene transfer strategies.

Various explanations have been suggested to underlie the development of public TCRs in T-cell responses targeting the same antigenic epitope(47). One was a high probability that these PUB-I sequences can be generated during V-D-J recombination(28, 48). Furthermore, various nucleotide-sequences can result in the same TCR AA-sequences that further increase the probability(49). Selection *in vivo* by optimal antigen-specific proliferation may result in a dominant antigen-specific memory T-cell population(50). These determinants may also lead to TCRs that are highly-similar to the PUB-I sequence, although they were often not included in the analyses of such public T-cell responses. It has been shown that conserved AAs in the CDR3 loop provide a structural framework that is required for the maintenance of the three dimensional TCR-structure(51). A similar structural framework between the PUB-I and PUB-HS sequences can thus lead to a conserved engagement with the peptide/HLA complex(52). Our rationale is that the PUB-I and PUB-HS sequences are part of the same public T-cell response when the same peptide-HLA complex is targeted, the same variable gene is expressed to have identical CDR1 and CDR2 regions and contains the same conserved AAs in the CDR3 loop. With this set of rules, we were able to quantitatively analyze the public T-cell responses and showed that T cells expressing PUB-I TCRs together with T cells that express PUB-HS TCRs made up at least 41% of the total TCR-repertoire. To assess the role of the alpha chains in PUB-I and PUB-HS TCRs, we identified the CDR3α sequence usage of a selection of virus-specific T-cell populations that contained shared TCRβ sequences and the corresponding CDR3α sequences all showed to be identical or highly similar between individuals. However, the high percentages of shared TCR-sequences contradict the findings observed by Madi et al., 2017, where they performed immunization in mice with foreign ovalbumin (OVA)-derived peptide that resulted in dominant private TCR-repertoires and less public TCRs (53). Since virus-specific memory T-cell repertoires in the circulation are shaped based on antigen encounter and subsequent proliferation, the PUB-I and PUB-HS TCR-sequences most likely reflect highly functional T cells capable antigen-driven proliferation. For latent viruses such as CMV, EBV and AdV, virus-specific T cells frequently encounter antigen during episodes of viral reactivation. The presence of PUB-I and PUB-HS TCR-sequences for these virus-specific T cells could be rather high due to this frequent antigen encounter. However, multiple reports also observed shared

CDR3β sequences in T-cell populations specific for non-latent viruses such as Influenza, RSV, and SARS-CoV-2(16, 26, 54), suggesting that this phenomenon is not unique for latent viruses, although the unexpected high magnitude of PUB-I and PUB-HS TCR-sequences that we observed can be unique for latent viruses.

These percentages of PUB-I and PUB-HS TCRs within these virus specific T-cell responses may still be an underestimation since the prerequisite of the identification of a PUB-HS TCR was similarity to a PUB-I TCR that was present in at least 2 individuals. Highly similar TCRs with only mutual similarities without identity in at least 2 individuals were not included as PUB-HS TCRs. Therefore, some of the unique TCRs within the virus-specific T-cell repertoire may also be part of a public T-cell response. This was indeed illustrated by the growing percentages of PUB-I and PUB-HS sequences when including more sampled sequences(55). Although it was suggested that HLA polymorphisms might be a confounding factor that affect the sharing of TCRs(55, 56), we showed that our validation cohort with different HLA-backgrounds revealed frequencies of the PUB-I and PUB-HS TCRs with at least a similar magnitude. Our approach involved a short *ex vivo* expansion of the isolated virus-specific T cells that might have created a bias towards the expansion of the presently identified PUB-I and PUB-HS TCRs, indicating that the actual numbers of PUB-I and PUB-HS in unmanipulated peripheral blood may have even been higher.

In conclusion, our findings demonstrate that a large part of the virus-specific TCR-repertoire contains PUB-I and PUB-HS TCRs at high frequencies in multiple different individuals.. Because virus-specific memory T-cell repertoires in the circulation are shaped based on antigen encounter and subsequent proliferation, the PUB-I and PUB-HS TCR-sequences most likely reflect highly functional T cells capable of antigen-driven proliferation. Since it is plausible that the highly-similar TCRs with conserved motifs similarly dock to the peptide-HLA complex as the identical shared TCR-sequences, these PUB-I and PUB-HS sequences can be considered part of the same public T-cell response. Such public TCRs may then be utilized for diagnostic purposes or therapeutic benefit in TCR-gene transfer-based immunotherapy strategies to effectively control viral-reactivation in immunocompromised patients.

Acknowledgement
This work was supported by Sanquin Research and Landsteiner Laboratory for Blood Cell research [PPO 15-37/Lnumber 2101]. This study was in part also supported by research funding from Stichting den Brinker (The Netherlands, Zeist) that made a donation to the Leukemia fund from the Bontius Foundation (Leiden University Medical Center).

REFERENCES

1. Chien YH, Gascoigne NR, Kavaler J, Lee NE, Davis MM. Somatic recombination in a murine T-cell receptor gene. Nature. 1984;309(5966):322-6.
2. Lefranc MP. Nomenclature of the human T cell receptor genes. Curr Protoc Immunol. 2001;Appendix 1:Appendix 1O.
3. Padovan E, Casorati G, Dellabona P, Meyer S, Brockhaus M, Lanzavecchia A. Expression of two T cell receptor alpha chains: dual receptor T cells. Science. 1993;262(5132):422-4.
4. Davis MM, Bjorkman PJ. T-cell antigen receptor genes and T-cell recognition. Nature. 1988;334(6181):395-402.
5. Borg NA, Ely LK, Beddoe T, Macdonald WA, Reid HH, Clements CS, et al. The CDR3 regions of an immunodominant T cell receptor dictate the 'energetic landscape' of peptide-MHC recognition. Nat Immunol. 2005;6(2):171-80.
6. Garboczi DN, Ghosh P, Utz U, Fan QR, Biddison WE, Wiley DC. Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. Nature. 1996;384(6605):134-41.
7. Robins HS, Campregher PV, Srivastava SK, Wacher A, Turtle CJ, Kahsai O, et al. Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T cells. Blood. 2009;114(19):4099-107.
8. Hebart H, Einsele H. Clinical aspects of CMV infection after stem cell transplantation. Hum Immunol. 2004;65(5):432-6.
9. Englund J, Feuchtinger T, Ljungman P. Viral infections in immunocompromised patients. Biol Blood Marrow Transplant. 2011;17(1 Suppl):S2-5.
10. Hanley PJ, Cruz CR, Shpall EJ, Bollard CM. Improving clinical outcomes using adoptively transferred immune cells from umbilical cord blood. Cytotherapy. 2010;12(6):713-20.
11. Hanley PJ, Cruz CR, Savoldo B, Leen AM, Stanojevic M, Khalil M, et al. Functionally active virus-specific T cells that target CMV, adenovirus, and EBV can be expanded from naive T-cell populations in cord blood and will target a range of viral epitopes. Blood. 2009;114(9):1958-67.
12. Patel S, Chorvinsky E, Albihani S, Cruz CR, Jones RB, Shpall EJ, et al. HIV-Specific T Cells Generated from Naive T Cells Suppress HIV In Vitro and Recognize Wide Epitope Breadths. Mol Ther. 2018;26(6):1435-46.
13. Provenzano M, Mocellin S, Bettinotti M, Preuss J, Monsurro V, Marincola FM, et al. Identification of immune dominant cytomegalovirus epitopes using quantitative real-time polymerase chain reactions to measure interferon-gamma production by peptide-stimulated peripheral blood mononuclear cells. J Immunother. 2002;25(4):342-51.
14. Long HM, Meckiff BJ, Taylor GS. The T-cell Response to Epstein-Barr Virus-New Tricks From an Old Dog. Front Immunol. 2019;10:2193.
15. Schone D, Hrycak CP, Windmann S, Lapuente D, Dittmer U, Tenbusch M, et al. Immunodominance of Adenovirus-Derived CD8(+) T Cell Epitopes Interferes with the Induction of Transgene-Specific Immunity in Adenovirus-Based Immunization. J Virol. 2017;91(20).
16. Sant S, Grzelak L, Wang Z, Pizzolla A, Koutsakos M, Crowe J, et al. Single-Cell Approach to Influenza-Specific CD8(+) T Cell Receptor Repertoires Across Different Age Groups, Tissues, and Following Influenza Virus Infection. Front Immunol. 2018;9:1453.
17. Schultheiss C, Paschold L, Simnica D, Mohme M, Willscher E, von Wenserski L, et al. Next-Generation Sequencing of T and B Cell Receptor Repertoires from COVID-19 Patients Showed Signatures Associated with Severity of Disease. Immunity. 2020;53(2):442-55 e4.
18. Shomuradova AS, Vagida MS, Sheetikov SA, Zornikova KV, Kiryukhin D, Titov A, et al. SARS-CoV-2 Epitopes Are Recognized by a Public and Diverse Repertoire of Human T Cell Receptors. Immunity. 2020;53(6):1245-57.e5.
19. Kamga L, Gil A, Song I, Brody R, Ghersi D, Aslan N, et al. CDR3alpha drives selection of the immunodominant Epstein Barr virus (EBV) BRLF1-specific CD8 T cell receptor repertoire in primary infection. PLoS Pathog. 2019;15(11):e1008122.
20. Roex MCJ, van Balen P, Germeroth L, Hageman L, van Egmond E, Veld SAJ, et al. Generation and infusion of multi-antigen-specific T cells to prevent complications early after T-cell depleted allogeneic stem cell transplantation-a phase I/II study. Leukemia. 2020;34(3):831-44.
21. Moss PA, Moots RJ, Rosenberg WM, Rowland-Jones SJ, Bodmer HC, McMichael AJ, et al. Extensive conservation of alpha and beta chains of the human T-cell antigen receptor recognizing HLA-A2 and influenza A matrix peptide. Proc Natl Acad Sci U S A. 1991;88(20):8987-90.
22. Venturi V, Chin HY, Price DA, Douek DC, Davenport MP. The role of production frequency in the sharing of simian immunodeficiency virus-specific CD8+ TCRs between macaques. J Immunol. 2008;181(4):2597-609.
23. Huth A, Liang X, Krebs S, Blum H, Moosmann A. Antigen-Specific TCR Signatures of Cytomegalovirus Infection. J Immunol. 2019;202(3):979-90.
24. Burrows SR, Khanna R, Burrows JM, Moss DJ. 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. J Exp Med. 1994;179(4):1155-61.
25. Nguyen TH, Bird NL, Grant EJ, Miles JJ, Thomas PG, Kotsimbos TC, et al. Maintenance of the EBV-specific CD8(+) TCRalphabeta repertoire in immunosuppressed lung transplant recipients.

Immunol Cell Biol. 2017;95(1):77-86.
26. Billam P, Bonaparte KL, Liu J, Ruckwardt TJ, Chen M, Ryder AB, et al. T Cell receptor clonotype influences epitope hierarchy in the CD8+ T cell response to respiratory syncytial virus infection. J Biol Chem. 2011;286(6):4829-41.
27. Shomuradova AS, Vagida MS, Sheetikov SA, Zornikova KV, Kiryukhin D, Titov A, et al. SARS-CoV-2 epitopes are recognized by a public and diverse repertoire of human T-cell receptors. medRxiv. 2020:2020.05.20.20107813.
28. Venturi V, Price DA, Douek DC, Davenport MP. The molecular basis for public T-cell responses? Nat Rev Immunol. 2008;8(3):231-8.
29. Garboczi DN, Hung DT, Wiley DC. HLA-A2-peptide complexes: refolding and crystallization of molecules expressed in Escherichia coli and complexed with single antigenic peptides. Proc Natl Acad Sci U S A. 1992;89(8):3429-33.
30. Rodenko B, Toebes M, Hadrup SR, van Esch WJ, Molenaar AM, Schumacher TN, et al. Generation of peptide-MHC class I complexes through UV-mediated ligand exchange. Nat Protoc. 2006;1(3):1120-32.
31. Eijsink C, Kester MG, Franke ME, Franken KL, Heemskerk MH, Claas FH, et al. Rapid assessment of the antigenic integrity of tetrameric HLA complexes by human monoclonal HLA antibodies. J Immunol Methods. 2006;315(1-2):153-61.
32. Koning MT, Nteleah V, Veelken H, Navarrete MA. Template-switching anchored polymerase chain reaction reliably amplifies functional lambda light chain transcripts of malignant lymphoma. Leuk Lymphoma. 2014;55(5):1212-4.
33. van Bergen CA, van Luxemburg-Heijs SA, de Wreede LC, Eefting M, von dem Borne PA, van Balen P, et al. Selective graft-versus-leukemia depends on magnitude and diversity of the alloreactive T cell response. J Clin Invest. 2017;127(2):517-29.
34. Huisman W, Leboux DAT, van der Maarel LE, Hageman L, Amsen D, Falkenburg JHF, et al. Magnitude of Off-Target Allo-HLA Reactivity by Third-Party Donor-Derived Virus-Specific T Cells Is Dictated by HLA-Restriction. Front Immunol. 2021;12:630440.
35. Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD. Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. Biotechniques. 2001;30(4):892-7.
36. Bolotin DA, Poslavsky S, Mitrophanov I, Shugay M, Mamedov IZ, Putintseva EV, et al. MiXCR: software for comprehensive adaptive immunity profiling. Nat Methods. 2015;12(5):380-1.
37. Csardi G, Nepusz T. The Igraph Software Package for Complex Network Research. InterJournal. 2005;Complex Systems:1695.
38. Loo MPJ. The stringdist Package for Approximate String Matching. R Journal. 2014;6:111-22.
39. Aboyoun P, Gentleman R, Raoy S. Biostrings: String Objects Representing Biological Sequences and Matching Algorithms. 2009.
40. Wickham H, François R. dplyr: A Grammar of Data Manipulation2014.
41. Villanueva RAM, Chen ZJ. ggplot2: Elegant Graphics for Data Analysis (2nd ed.). Measurement: Interdisciplinary Research and Perspectives. 2019;17(3):160-7.
42. Neuwirth E, Neuwirth ME. Package 'RColorBrewer'. CRAN 2011-06-17 08: 34: 00. Apache License 2.0; 2011.
43. Lewin B. Genes V 5th ed. ed. Oxford, England: Oxford University Press; 1994.
44. Roex MCJ, Hageman L, Heemskerk MT, Veld SAJ, van Liempt E, Kester MGD, et al. The simultaneous isolation of multiple high and low frequent T-cell populations from donor peripheral blood mononuclear cells using the major histocompatibility complex I-Streptamer isolation technology. Cytotherapy. 2018;20(4):543-55.
45. Marrack P, Scott-Browne JP, Dai S, Gapin L, Kappler JW. Evolutionarily conserved amino acids that control TCR-MHC interaction. Annu Rev Immunol. 2008;26:171-203.
46. Huisman W, Gille I, van der Maarel LE, Hageman L, Morton LT, de Jong RCM, et al. Identification of Functional HLA-A*01 :01-Restricted EBV-LMP2-Specific T-cell Receptors. J Infect Dis. 2020.
47. Li H, Ye C, Ji G, Han J. Determinants of public T cell responses. Cell Res. 2012;22(1):33-42.
48. Robins HS, Srivastava SK, Campregher PV, Turtle CJ, Andriesen J, Riddell SR, et al. Overlap and effective size of the human CD8+ T cell receptor repertoire. Sci Transl Med. 2010;2(47):47ra64.
49. Quigley MF, Greenaway HY, Venturi V, Lindsay R, Quinn KM, Seder RA, et al. Convergent recombination shapes the clonotypic landscape of the naive T-cell repertoire. Proc Natl Acad Sci U S A. 2010;107(45):19414-9.
50. Price DA, Brenchley JM, Ruff LE, Betts MR, Hill BJ, Roederer M, et al. Avidity for antigen shapes clonal dominance in CD8+ T cell populations specific for persistent DNA viruses. J Exp Med. 2005;202(10):1349-61.
51. Bourcier KD, Lim D-G, Ding Y-H, Smith KJ, Wucherpfennig K, Hafler DA. Conserved CDR3 Regions in T-Cell Receptor (TCR) CD8⁺T Cells That Recognize the Tax11-19/HLA-A*0201 Complex in a Subject Infected with Human T-Cell Leukemia Virus Type 1: Relationship of T-Cell Fine Specificity and Major Histocompatibility Complex/Peptide/TCR Crystal Structure. J Virol. 2001;75(20):9836-43.
52. Szeto C, Lobos CA, Nguyen AT, Gras S. TCR Recognition of Peptide-MHC-I: Rule Makers and Breakers. Int J Mol Sci. 2020;22(1).
53. Madi A, Shifrut E, Reich-Zeliger S, Gal H, Best K, Ndifon W, et al. T-cell receptor repertoires share a restricted set of public and abundant CDR3 sequences that are associated with self-related immunity. Genome Res. 2014;24(10):1603-12.
54. Schultheiß C, Paschold L, Simnica D, Mohme M, Willscher E, von Wenserski L, et al. Next-Generation Sequencing of T and B Cell Receptor

Repertoires from COVID-19 Patients Showed Signatures Associated with Severity of Disease. Immunity. 2020;53(2):442-55.e4.
55. Madi A, Shifrut E, Reich-Zeliger S, Gal H, Best K, Ndifon W, et al. T-cell receptor repertoires share a restricted set of public and abundant CDR3 sequences that are associated with self-related immunity. Genome Res. 2014;24(10):1603-12.
56. Murugan A, Mora T, Walczak AM, Callan CG, Jr. Statistical inference of the generation probability of T-cell receptors from sequence repertoires. Proc Natl Acad Sci U S A. 2012;109(40):16161-6.

Supplementary Table 1. Primer sequences.

Description	Name	Nucleotide sequence 5'►3'
cDNA primer TRB constant region reverse transcription	TRB_RT	CACGTGGTCGGGGWAGAAGC
cDNA primer SmartSeq2modified template switching oligo	SS2m_TSO	AAGCAGTGGTATCAACGCAGAGTACAT(G)(G){G}
PCR primer SmartSeq2modified forward	SS2m_For	GAGTTCAGACGTGTGCTCTTCCGATCTAAGCAGTGGTATCAACGCAGAGTACAT*G
PCR primer TRBC1 reverse	TRBC1_rev	CCTACACGACGCTCTTCCGATCTGTGGGAACACCTTGTTCAGGTCCT*C
PCR primer TRBC1 reverse	TRBC2_rev	CCTACACGACGCTCTTCCGATCTGTGGGAACACGTTTTTCAGGTCCT*C
Barcode primer SS2m region, forward, backbone	BC_R7xx_For	CAAGCAGAAGACGGCATACGAGAT nnnnnn GTGACTGGAGTTCAGACGTGTGCTCTTCCGAT*C
Barcode primer TRBC region reverse, backbone	BC_R7xx_Rev	AATGATACGGCGACCACCGAGATCTACAC nnnnnn ACACTCTTTCCCTACACGACGCTCTTCCGATC*T

Abbreviations: TRB: T-cell Receptor Beta, SS2m: SmartSeq2Modified, TSO: Template Switching Oligo, TRBC: T-cell Receptor Beta Constant, For: Forward, Rev: Reverse, BC: Beta chain, nnnnnn: Identifier sequence
()=RNA, {}=LNA: Locked Nucleic Acid, *:phosphonothioate-binding

Supplementary Table 2. Identifier sequences.

Identifiers (For) Name	Identifiers (For) Seq	Identifiers (Rev) Name	Identifiers (Rev) Seq
BC_R701	ATCACG	BC_R725	ACTGAT
BC_R702	CGATGT	BC_R726	ATGAGC
BC_R703	TTAGGC	BC_R727	ATTCCT
BC_R704	TGACCA	BC_R728	CAAAAG
BC_R705	ACAGTG	BC_R729	CAACTA
BC_R706	GCCAAT	BC_R730	CACCGG
BC_R707	CAGATC	BC_R731	CACGAT
BC_R708	ACTTGA	BC_R732	CACTCA
BC_R709	GATCAG	BC_R733	CAGGCG
BC_R710	TAGCTT	BC_R734	CATGGC
BC_R711	GGCTAC	BC_R735	CATTTT
BC_R712	CTTGTA	BC_R736	CCAACA
BC_R713	AGTCAA	BC_R737	CGGAAT
BC_R714	AGTTCC	BC_R738	CTAGCT
BC_R715	ATGTCA	BC_R739	CTATAC
BC_R716	CCGTCC	BC_R740	CTCAGA
BC_R717	GTAGAG	BC_R741	GACGAC
BC_R718	GTCCGC	BC_R742	TAATCG
BC_R719	GTGAAA	BC_R743	TACAGC
BC_R720	GTGGCC	BC_R744	TATAAT
BC_R721	GTTTCG	BC_R745	TCATTC
BC_R722	CGTACG	BC_R746	TCCCGA
BC_R723	GAGTGG	BC_R747	TCGAAG
BC_R724	GGTAGC	BC_R748	TCGGCA

Supplementary Table 3. Occurrence and number of CDR3β amino-acid sequences that are shared between individuals.

Virus	Antigen	HLA	TRBV	CDR3	TRBJ	Occurrence (#)
CMV	pp50-VTE	A*01	TRBV20-1	CSARLLGGGQSYEQYF	TRBJ2-7	2/7
CMV	pp50-VTE	A*01	TRBV9	CASSVGQGSSYEQYF	TRBJ2-7	2/7
CMV	pp65-YSE	A*01	TRBV9	CASSVTGGTDTQYF	TRBJ2-3	2/6
CMV	pp65-NLV	A*02	TRBV7-6	CASSLAPGATNEKLFF	TRBJ1-4	5/8
CMV	pp65-NLV	A*02	TRBV7-6	CASSLAPGTTNEKLFF	TRBJ1-4	3/8
CMV	pp65-NLV	A*02	TRBV12-4	CASSSAYYGYTF	TRBJ1-2	2/8
CMV	IE1-VLE	A*02	TRBV7-3	CASSLGQGGVETQYF	TRBJ2-5	2/6
CMV	IE1-VLE	A*02	TRBV7-3	CASSPGQGGVETQYF	TRBJ2-5	2/6
CMV	pp65-TPR	B*07	TRBV7-9	CASSLIGVSSYNEQFF	TRBJ2-1	5/8
CMV	pp65-RPH	B*07	TRBV4-3	CASSPQRNTEAFF	TRBJ1-1	4/6
CMV	pp65-RPH	B*07	TRBV4-3	CASSPSRNTEAFF	TRBJ1-1	2/6
CMV	IE1-ELR	B*08	TRBV27	CASSSYRTDLNTEAFF	TRBJ1-1	2/5
CMV	IE1-QIK	B*08	TRBV9	CASSTQVSEPNTGELFF	TRBJ2-2	2/6
CMV	IE1-QIK	B*08	TRBV9	CASSVQRQTANTGELFF	TRBJ2-2	2/6
CMV	IE1-QIK	B*08	TRBV12-5	CASGPRAGAYNEQFF	TRBJ2-1	2/6
CMV	IE1-QIK	B*08	TRBV2	CASSGTGRLTMNTEAFF	TRBJ1-1	2/6
CMV	IE1-QIK	B*08	TRBV21-1	CASSKVAARVP-TLKLS	TRBJ1-1	2/6
CMV	IE1-QIK	B*08	TRBV7-9	CASSLTLAGNQPQHF	TRBJ1-5	2/6
EBV	BRLF1-YVL	A*02	TRBV10-1	CASSAGPDTQYF	TRBJ2-3	2/12
EBV	BRLF1-YVL	A*02	TRBV24-1	CATSDYGEDTQYF	TRBJ2-3	2/12
EBV	BRLF1-YVL	A*02	TRBV25-1	CASSEWTTDTQYF	TRBJ2-3	2/12
EBV	BRLF1-YVL	A*02	TRBV28	CASSKIMNTEAFF	TRBJ1-1	2/12
EBV	BRLF1-YVL	A*02	TRBV6-5	CASSQLLGSNQPQHF	TRBJ1-5	2/12
EBV	EBNA3A-RPP	B*07	TRBV4-1	CASSQDRLTGGYTF	TRBJ1-2	4/11
EBV	EBNA3A-RPP	B*07	TRBV4-1	CASSQDRLTGTQYF	TRBJ2-5	2/11
EBV	EBNA3A-RPP	B*07	TRBV4-1	CASSQEAFNYEQYF	TRBJ2-7	2/11
EBV	BZLF1-RAK	B*08	TRBV27	CASSSLNTEAFF	TRBJ1-1	8/17
EBV	BZLF1-RAK	B*08	TRBV27	CASSPLTDTQYF	TRBJ2-3	3/17
EBV	BZLF1-RAK	B*08	TRBV20-1	CSARDRGAENTGELFF	TRBJ2-2	3/17
EBV	BZLF1-RAK	B*08	TRBV20-1	CSARDRGGENTGELFF	TRBJ2-2	3/17
EBV	BZLF1-RAK	B*08	TRBV29-1	CSVGSGEGYEQYF	TRBJ2-7	3/17
EBV	BZLF1-RAK	B*08	TRBV4-1	CASSPGQGEGYEQYF	TRBJ2-7	3/17
EBV	BZLF1-RAK	B*08	TRBV7-9	CASSPTGAGNQPQHF	TRBJ1-5	3/17
EBV	BZLF1-RAK	B*08	TRBV27	CASSNLNTEAFF	TRBJ1-1	2/17
EBV	BZLF1-RAK	B*08	TRBV27	CASSDLNSPLHF	TRBJ1-6	2/17
EBV	BZLF1-RAK	B*08	TRBV27	CASSSLNSPLHF	TRBJ1-6	2/17
EBV	BZLF1-RAK	B*08	TRBV7-2	CASSLVLLGNSPLHF	TRBJ1-6	2/17
EBV	BZLF1-RAK	B*08	TRBV4-1	CASSRLAGDTDTQYF	TRBJ2-3	2/17
EBV	BZLF1-RAK	B*08	TRBV6-1	CASTGTASTDTQYF	TRBJ2-3	2/17
EBV	BZLF1-RAK	B*08	TRBV20-1	CSARDRGSENTGELFF	TRBJ2-2	2/17
EBV	BZLF1-RAK	B*08	TRBV20-1	CSARDRGTENTGELFF	TRBJ2-2	2/17
EBV	BZLF1-RAK	B*08	TRBV7-3	CASSSHSGINTGELFF	TRBJ2-2	2/17

Supplementary Table 3. Continued.

Virus	Antigen	HLA	TRBV	CDR3	TRBJ	Occurrence (#)	Virus	Antigen	HLA	TRBV	CDR3	TRBJ	Occurrence (#)
EBV	LMP2-FLY	A*02	TRBV6-5	CASSYQGGNYGYTF	TRBJ1-2	9/11	EBV	BZLF1-RAK	B*08	TRBV10-3	CATGLAGSTDTQYF	TRBJ2-3	2/17
EBV	LMP2-FLY	A*02	TRBV6-5	CASSRQGGNYGYTF	TRBJ1-2	7/11	EBV	BZLF1-RAK	B*08	TRBV4-1	CASSPGTGEGYEQYF	TRBJ2-7	2/17
EBV	LMP2-FLY	A*02	TRBV6-5	CASSLQGGNYGYTF	TRBJ1-2	5/11	EBV	BZLF1-RAK	B*08	TRBV7-2	CASSPGTGEGYEQYF	TRBJ2-7	2/17
EBV	LMP2-FLY	A*02	TRBV6-5	CASSGQGGNYGYTF	TRBJ1-2	4/11	EBV	BZLF1-RAK	B*08	TRBV7-2	CASSYHGSYEQYF	TRBJ2-7	2/17
EBV	LMP2-FLY	A*02	TRBV6-5	CASSKQGGGYGYTF	TRBJ1-2	3/11	EBV	BZLF1-RAK	B*08	TRBV7-6	CASSLAGEGYEQYF	TRBJ2-7	2/17
EBV	LMP2-FLY	A*02	TRBV6-5	CASSPQGGGYGYTF	TRBJ1-2	3/11	EBV	BZLF1-RAK	B*08	TRBV7-9	CASSSTGAGNQPQHF	TRBJ1-5	2/17
EBV	LMP2-FLY	A*02	TRBV6-5	CASSRQGGTYGYTF	TRBJ1-2	3/11	EBV	BZLF1-RAK	B*08	TRBV7-9	CASSSTGSGDQPQHF	TRBJ1-5	2/17
EBV	LMP2-FLY	A*02	TRBV6-5	CASSSQGGNYGYTF	TRBJ1-2	3/11	EBV	BZLF1-RAK	B*08	TRBV7-3	CASSLIASGGYNEQFF	TRBJ2-1	2/17
EBV	LMP2-FLY	A*02	TRBV6-5	CASSYSGGYYGYTF	TRBJ1-2	2/11							
EBV	LMP2-FLY	A*02	TRBV6-5	CASSDQGGGYGYTF	TRBJ1-2	2/11	EBV	EBNA3A-FLR	B*08	TRBV7-8	CASSLGQAYEQYF	TRBJ2-7	4/13
EBV	LMP2-FLY	A*02	TRBV6-5	CASSFQGGNYGYTF	TRBJ1-2	2/11	EBV	EBNA3A-FLR	B*08	TRBV7-8	CASSSGQAYEQYF	TRBJ2-7	4/13
EBV	LMP2-FLY	A*02	TRBV6-5	CASSPLGGAEGYTF	TRBJ1-2	2/11	EBV	EBNA3A-FLR	B*08	TRBV7-8	CASSTGQAYEQYF	TRBJ2-7	3/13
EBV	LMP2-FLY	A*02	TRBV6-5	CASSPQGGNYGYTF	TRBJ1-2	2/11	EBV	EBNA3A-FLR	B*08	TRBV4-3	CASSHGLAGILETQYF	TRBJ2-5	2/13
EBV	LMP2-FLY	A*02	TRBV6-5	CASSPQGGRDGYTF	TRBJ1-2	2/11	EBV	EBNA3A-FLR	B*08	TRBV4-3	CASSPTSGVAGELFF	TRBJ2-2	2/13
EBV	LMP2-FLY	A*02	TRBV6-5	CASSRQGGSYGYTF	TRBJ1-2	2/11	EBV	EBNA3A-FLR	B*08	TRBV4-1	CASSQGLAVSSYEQYF	TRBJ2-7	2/13
EBV	LMP2-FLY	A*02	TRBV6-5	CASSSQGGSNYGYTF	TRBJ1-2	2/11	EBV	EBNA3A-FLR	B*08	TRBV7-9	CASSWGPEQFF	TRBJ2-1	2/13
EBV	LMP2-FLY	A*02	TRBV6-5	CASSSQGGSYGYTF	TRBJ1-2	2/11							
EBV	LMP2-FLY	A*02	TRBV6-5	CASSYEGGYYGYTF	TRBJ1-2	2/11	EBV	EBNA3A-QAK	B*08	TRBV18	CAASRGCEPKTFST	TRBJ2-4	5/18
EBV	LMP2-FLY	A*02	TRBV6-5	CASSYQGGSYGYTF	TRBJ1-2	2/11	EBV	EBNA3A-QAK	B*08	TRBV28	CASSNLGVTELNTGELFF	TRBJ2-2	3/18
EBV	LMP2-FLY	A*02	TRBV6-5	CASNPQGGGGGYTF	TRBJ1-2	2/11	EBV	EBNA3A-QAK	B*08	TRBV5-1	CASSLELAVYNEQFF	TRBJ2-1	3/18
EBV	LMP2-FLY	A*02	TRBV6-5	CASNPQGGGNGYTF	TRBJ1-2	2/11	EBV	EBNA3A-QAK	B*08	TRBV5-1	CASSLETATEAFF	TRBJ1-1	3/18
EBV	LMP2-FLY	A*02	TRBV6-5	CASSYQGGNEQFF	TRBJ2-1	3/11	EBV	EBNA3A-QAK	B*08	TRBV5-1	CASSLETGGYGYTF	TRBJ1-2	3/18
EBV	LMP2-FLY	A*02	TRBV6-5	CASSLQGGNEQFF	TRBJ2-1	2/11	EBV	EBNA3A-QAK	B*08	TRBV27	CASSLYRDNQPQHF	TRBJ1-5	2/18
EBV	LMP2-FLY	A*02	TRBV6-5	CASTLQGGNEQFF	TRBJ2-1	2/11	EBV	EBNA3A-QAK	B*08	TRBV27	CASSPDRWETQYF	TRBJ2-5	2/18
							EBV	EBNA3A-QAK	B*08	TRBV28	CASSALSGLAGPGELFF	TRBJ2-2	2/18
EBV	LMP2-CLG	A*02	TRBV10-2	CASSEDGMNTEAFF	TRBJ1-1	3/10	EBV	EBNA3A-QAK	B*08	TRBV28	CASSKQGAPGHTGELFF	TRBJ2-2	2/18

Supplementary Table 3. Continued.

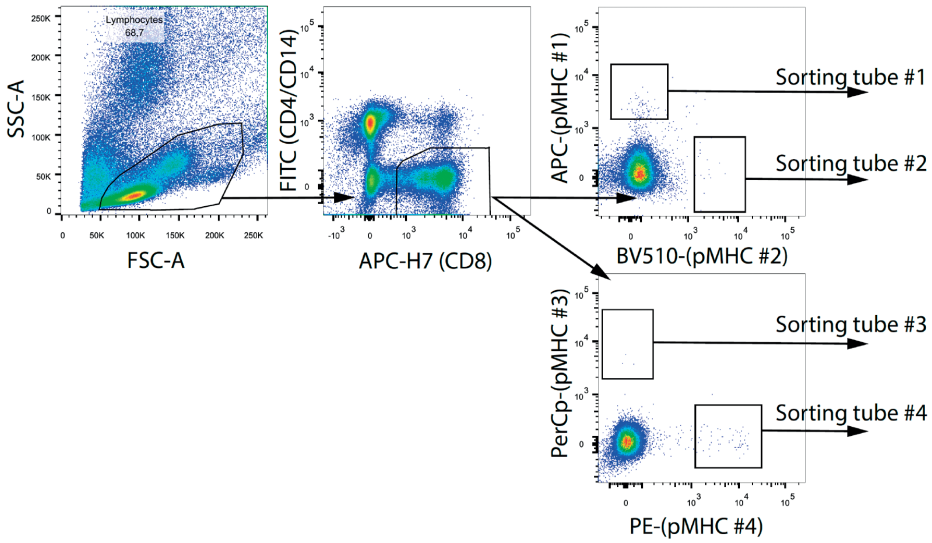
Virus	Antigen	HLA	TRBV	CDR3	TRBJ	Occurrence (#)	Virus	Antigen	HLA	TRBV	CDR3	TRBJ	Occurrence (#)
EBV	LMP2-CLG	A*02	TRBV10-2	CASSSDGMNTEAFF	TRBJ1-1	2/10	EBV	EBNA3A-QAK	B*08	TRBV28	CASSLLGARGLNEKLFF	TRBJ1-4	2/18
EBV	LMP2-CLG	A*02	TRBV10-2	CASSGDGMNTEAFF	TRBJ1-1	2/10	EBV	EBNA3A-QAK	B*08	TRBV28	CASSLLGTGGLSEKLFF	TRBJ1-4	2/18
EBV	LMP2-CLG	A*02	TRBV10-2	CASSQDGMNTEAFF	TRBJ1-1	2/10	EBV	EBNA3A-QAK	B*08	TRBV28	CASSQQGARSLSEKLFF	TRBJ1-4	2/18
EBV	LMP2-CLG	A*02	TRBV5-1	CASSLEGQASSYEQYF	TRBJ2-7	3/10	EBV	EBNA3A-QAK	B*08	TRBV4-2	CASSQDAGDRLAGVTGELFF	TRBJ2-2	2/18
							EBV	EBNA3A-QAK	B*08	TRBV5-1	CASSLETGDTQYF	TRBJ2-3	2/18
EBV	EBNA3C-LLD	A*02	TRBV19	CASSIALASEQYF	TRBJ2-7	2/7	EBV	EBNA3A-QAK	B*08	TRBV6-3	CASSLDPPGQSIRVNTGELFF	TRBJ2-2	2/18
EBV	BMLF1-GLC	A*02	TRBV29-1	CSVGTGGTNEKLFF	TRBJ1-4	6/10	AdV	HEXON-TDL	A*01	TRBV20-1	CSAPGQGTDTQYF	TRBJ2-3	8/12
EBV	BMLF1-GLC	A*02	TRBV20-1	CSARDRVGNTIYF	TRBJ1-3	5/10	AdV	HEXON-TDL	A*01	TRBV20-1	CSAPGQGTTEAFF	TRBJ1-1	4/12
EBV	BMLF1-GLC	A*02	TRBV20-1	CSARDGTGNGYTF	TRBJ1-2	3/10	AdV	HEXON-TDL	A*01	TRBV20-1	CSAPGQGTYEQYF	TRBJ2-7	3/12
EBV	BMLF1-GLC	A*02	TRBV20-1	CSARDRTGNGYTF	TRBJ1-2	3/10	AdV	HEXON-TDL	A*01	TRBV20-1	CSAPGQGSTEAFF	TRBJ1-1	3/12
EBV	BMLF1-GLC	A*02	TRBV29-1	CSVGAGGTNEKLFF	TRBJ1-4	3/10	AdV	HEXON-TDL	A*01	TRBV20-1	CSAPGQGEETQYF	TRBJ2-5	2/12
EBV	BMLF1-GLC	A*02	TRBV14	CASSQSPGGTQYF	TRBJ2-3	2/10	AdV	HEXON-TDL	A*01	TRBV20-1	CSAPGQGENTQYF	TRBJ2-3	2/12
EBV	BMLF1-GLC	A*02	TRBV20-1	CSARVGVGNTIYF	TRBJ1-3	2/10	AdV	HEXON-TDL	A*01	TRBV5-1	CASNDYDNEQFF	TRBJ2-1	2/12
EBV	BMLF1-GLC	A*02	TRBV29-1	CSAGSGGTNEKLFF	TRBJ1-4	2/10	AdV	HEXON-TDL	A*01	TRBV5-1	CASNLADDEQFF	TRBJ2-1	2/12
EBV	BMLF1-GLC	A*02	TRBV29-1	CSVGSGGTNEKLFF	TRBJ1-4	2/10	AdV	HEXON-TDL	A*01	TRBV10-3	CATQTGGSNQPQHF	TRBJ1-5	2/12
							AdV	HEXON-TDL	A*01	TRBV4-1	CASSQVVGQAHSPLHF	TRBJ1-6	2/12
EBV	BRLF1-YVL	A*02	TRBV20-1	CSAIGGSYNEQFF	TRBJ2-1	3/12	AdV	HEXON-TDL	A*01	TRBV6-6	CASSYPGNNSPLHF	TRBJ1-6	2/12
EBV	BRLF1-YVL	A*02	TRBV20-1	CSAPVPPYNEQFF	TRBJ2-1	2/12	AdV	HEXON-TDL	A*01	TRBV20-1	CSAR-ASVATSST	TRBJ2-7	2/12
EBV	BRLF1-YVL	A*02	TRBV20-1	CSARGTEFYEQYF	TRBJ2-7	2/12	AdV	HEXON-TDL	A*01	TRBV19	CATSSAAQETQYF	TRBJ2-5	2/12
EBV	BRLF1-YVL	A*02	TRBV28	CASSLFSNEQFF	TRBJ2-1	2/12							
							AdV	HEXON-KPY	B*07	TRBV10-3	CAINPGTAYGYTF	TRBJ1-2	2/8
							AdV	HEXON-KPY	B*07	TRBV18	CASSPGTPEQFF	TRBJ2-1	2/8

A total of 131 different shared identical CDR3β amino-acid sequences are shown. The number of T-cell populations that contain the shared identical CDR3β amino-acid sequences are shown per total number of T-cell populations of that respective specificity (#), reflecting the occurrence among donors.

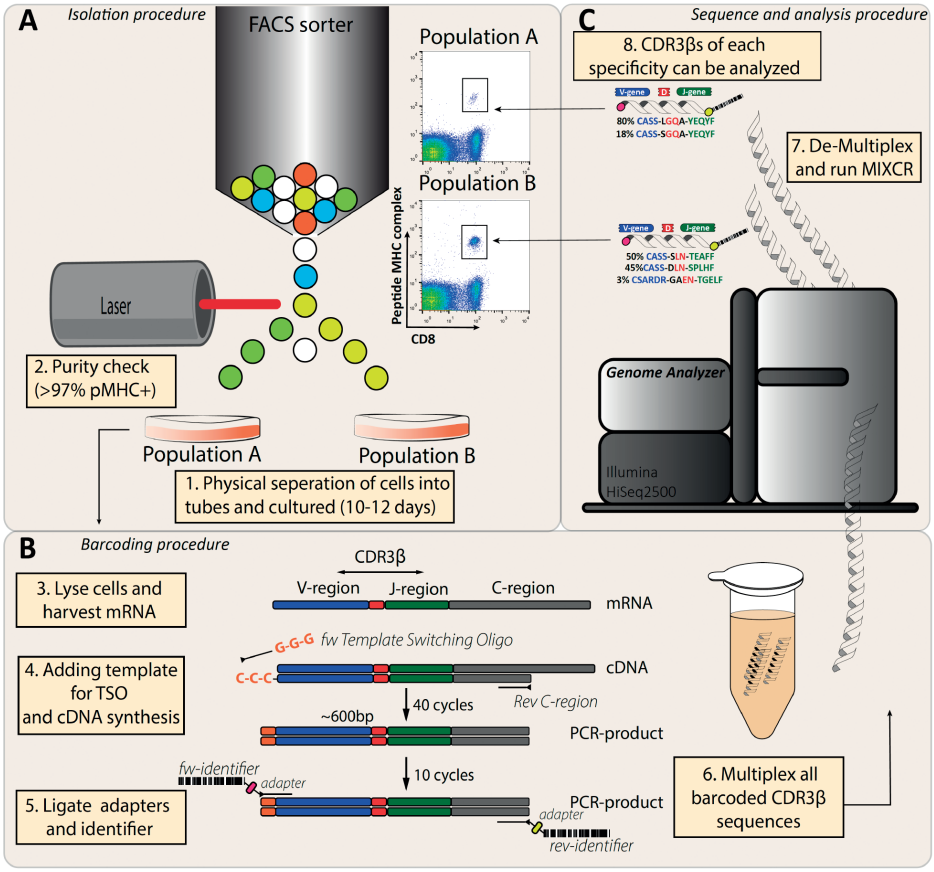
Supplementary Table 4. CDR3-alpha sequences of virus-specific T-cell populations with PUB-I and PUB-HS CDR3-beta sequences

Donor ID	Specificity	CDR3-beta	TRAV	CDR3-alpha	TRAJ
22	EBV-EBNA3A-RPP	CASSQDRLTGGYTF	TRAV24	CA**FS**SSNTGKLIF	TRAJ37
28	EBV-EBNA3A-RPP	CASSQDRLTGGYTF	TRAV24	CA**HG**SSNTGKLIF	TRAJ37
19	EBV-EBNA3A-RPP	CASSQDRLTGGYTF	TRAV24	CA**SS**SSNTGKLIF	TRAJ37
19	AdV-E1A-LLD	CSAR**A**GL**AE**TQYF	TRAV19	CALSDYGGYNKLIF	TRAJ4
28	AdV-E1A-LLD	CSAR**S**GL**SD**TQYF	TRAV19	CALSDYGGYNKLIF	TRAJ4
19	CMV-pp65-RPH	CASSPQRNTEAFF	TRAV23DV6	CAASIGNFGNEKLTF	TRAJ48
20	CMV-pp65-RPH	CASSPQRNTEAFF	TRAV23DV6	CAASIGNFGNEKLTF	TRAJ48
5	CMV-pp50-VTE	CSARLLGGGQSYEQYF	TRAV1-1	CAAPNNQGGKLIF	TRAJ23
7	CMV-pp50-VTE	CSARLLGGGQSYEQYF	TRAV1-1	CAAPNNQGGKLIF	TRAJ23
23	EBV-BRLF1-YVL	CSAIGGSYNEQFF	TRAV14DV4	CAMR**A**GGNFNKFYF	TRAJ21
18	EBV-BRLF1-YVL	CSAIGGSYNEQFF	TRAV14DV4	CAMR**S**GGNFNKFYF	TRAJ21
25	EBV-EBNA3A-CLG	CASSLEGQ**AS**SYEQYF	TRAV25	CAG**S**GAGSYQLTF	TRAJ28
29	EBV-EBNA3A-CLG	CASSLEGQ**GA**SYEQYF	TRAV25	CAG**L**GAGSYQLTF	TRAJ28
27	EBV-LMP2-FLY	CASS**S**QGGNYGYTF	TRAV17*01	CATEG**N**SGYSTLTF	TRAJ11*01
23	EBV-LMP2-FLY	CASS**Y**QGGNYGYTF	TRAV17*01	CATEG**D**SGYSTLTF	TRAJ11*01
29	EBV-LMP2-FLY	CASS**F**QGGNYGYTF	TRAV17*01	CA**S**EG**N**SGYSTLTF	TRAJ11*01

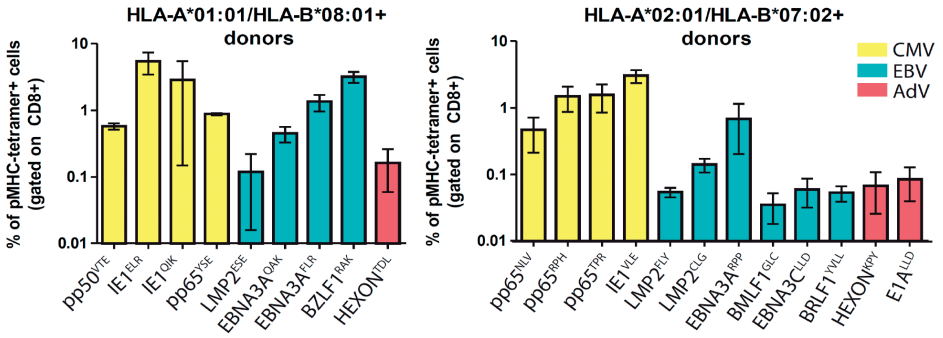
The CDR3-alpha usage of a selected number of virus-specific T-cell populations that contained a PUB-I or PUB-HS CDR3-beta sequence was analyzed. Amino-acids in bold represent differences between donors.



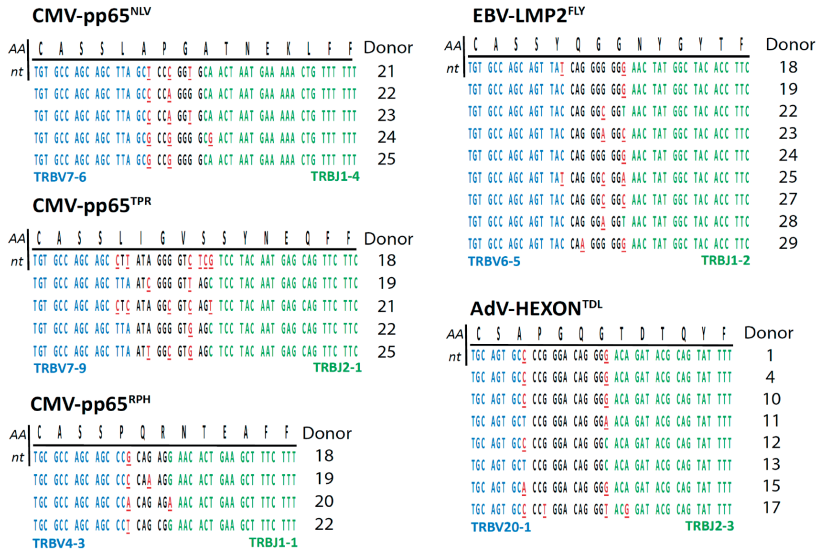
Supplementary Figure 1. Gating strategy for a 4-way single-pMHC-tetramer sort from PBMCs. In total, $30*10^6$ PBMCs were incubated with 4 different pMHC-tetramer complexes, followed by labeling with CD8, CD4 and CD14 monoclonal antibodies. Viable cells were gated based on FSC/SSC followed by gating of $CD8^{pos}$ and $CD4/CD14^{neg}$ T cells. Peptide-MHC-tetramer positive T cells were sorted simultaneously for 4 different specificities in bulk.



Supplementary Figure 2. Experimental setup to generate a library of CDR3β-sequences from virus-specific T-cell populations. **A**) A total of 190 virus-specific T-cell populations, restricted to HLA-A*01:01, HLA-A*02:01, HLA-B*07:02 or HLA-B*08:01 were isolated using 21 different peptideMHC-tetramers. **B**) Virus-specific T-cell populations were lysed and mRNA was harvested. In the first PCR step, primers specific for the C-region and template switching oligos were added to allow for cDNA synthesis and amplification. A second PCR step was performed with a single primer on each site, which adds unique forward and reverse identifiers (6 basepairs) to each PCR-product for each T-cell population. All 190 PCR-products were multiplexed and high-throughput sequenced. **C**) The library was de-multiplexed based on the unique identifiers. The CDR3β-region was determined using bi-directional readings with MIXCR.



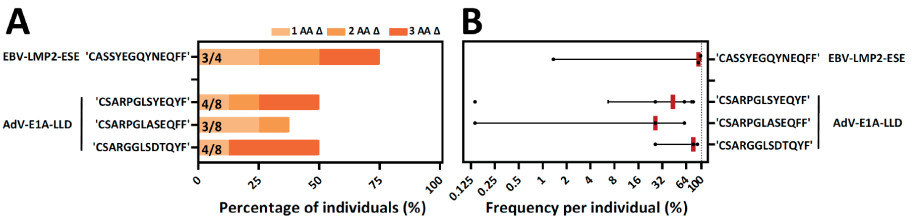
Supplementary Figure 3. Precursor frequencies of virus-specific T cells in healthy individuals. A total of 190 virus-specific T-cell populations, restricted to HLA-A*01:01, HLA-A*02:01, HLA-B*07:02 or HLA-B*08:01 were isolated using 21 different peptideMHC-tetramers. Percentages of CD8 positive pMHC-tetramer positive cells in the starting material are shown for each specificity. CMV and EBV-specific T-cell populations were only sorted from CMV and EBV-seropositive individuals, respectively.



Germline sequences

TRBV4-3	TGCGCCAGCAGCCAAGA	**TRBV7-9**	TGTGCCAGCAGCTTAGC	**TRBJ1-1**	TGAACACTGAAGCTTTCTTT	**TRBJ2-1**	CTCCTACAATGAGCAGTTCTTC
TRBV6-5	TGTGCCAGCAGTTACTC	**TRBV20-1**	TGCAGTGCTAGAGA	**TRBJ1-2**	CTAACTATGGCTACACCTTC	**TRBJ2-3**	AGCACAGATACGCAGTATTTT
TRBV7-6	TGTGCCAGCAGCTTAGC			**TRBJ1-4**	CAACTAATGAAAAACTGTTTTTT		

Supplementary Figure 4. **Identical shared CDR3β amino-acid sequences are found in different individuals with small nucleotide differences as a result of convergent recombination**. The CDR3β nucleotide sequences are shown per donor for 6 identical shared CDR3β amino-acid sequences. Underlined nucleotides in red resemble differences between the different individuals. Nucleotide sequences in blue and green represent perfect alignment with the germline sequences of the TRBV-gene and TRBJ-gene, respectively. The legend represents the germline sequences of (part of) the TRBV and TRBJ genes



Supplementary Figure 5. **Highly similar CDR3β amino-acid sequences in specific T-cell populations that did not contain an identical shared CDR3β amino-acid sequence**. **A**) For two specificities no identical shared CDR3β amino-acid sequences were found.. Individuals did contain highly similar sequences, and these were stacked with 1, 2 or 3 amino-acid differences. The occurrence, shown as percentages among healthy donors, is shown per CDR3β amino-acid sequence. The total number of different T-cell populations (different donors) for each specificity/CDR3β amino-acid sequence is shown at the inner-side of the y-axis. **B**) Shown is the sum of frequencies of the identical and highly similar (1,2 and 3 amino-acid differences) CDR3β amino-acid sequences per individual. Each dot is one individual, and the red-lines represents the medians with interquartile ranges
AA: amino-acids, nt: nucleotides, Δ: difference(s)