JVI Journals.ASM.org

High Frequency of Herpesvirus-Specific Clonotypes in the Human T Cell Repertoire Can Remain Stable over Decades with Minimal Turnover

M. A. Neller,[a] J. M. Burrows,[a] M. J. Rist,[a,b] J. J. Miles,[a,b,c] S. R. Burrows[a,b]

Australian Centre for Vaccine Development, Queensland Institute of Medical Research, Brisbane, Australia[a]; School of Medicine, The University of Queensland, Brisbane, Australia[b]; and Institute of Infection and Immunity, Cardiff University School of Medicine, Heath Park, Cardiff, United Kingdom[c]

High-throughput T cell receptor sequencing on sequentially banked blood samples from healthy individuals has shown that high-frequency clonotypes can remain relatively stable for up to 18 years, with minimal inflation, deflation, or turnover. These populations included T cell expansions specific for Epstein-Barr virus. Thus, in spite of exposure to a barrage of microorganisms over the course of life, the dominant clonotypes in the mature peripheral T cell repertoire can alter surprisingly little.

The human immune system is thought to include approximately 10^{12} T cells expressing 1 million to 25 million unique T cell receptors (TCRs) (1–3). From the vast naive receptor pool, select clonotypes emerge to engage the series of infectious agents, aberrant cells, and occasional innocuous agents encountered over the course of life (4). Often these T cell expansions are further amplified by persistent pathogens that provide a constant stream of antigen (Ag) to the cellular compartment. In mice, chronic infection can lead to rapid memory inflation, T cell dysfunction, and "exhaustion" (5). To compensate, new naive T cells can be recruited into the Ag-specific repertoire (6). In humans, the effect of long-term persistent infection on the T cell repertoire is less well defined; however, the tracking of human memory clonotypes specific for viruses has shown long-term persistence and dominance in the peripheral circulation for many years (7–9). These studies, which employed conventional sequencing techniques and were limited to a small number of clonotypes specific for several viral epitopes, provide limited insight into the stability of the T cell repertoire *in totum*. To address this issue, we have utilized a sequential series of cryobanked peripheral blood mononuclear cell (PBMC) samples from healthy individuals with new high-throughput sequencing (HTS) technology to monitor global fluctuations in the T cell repertoire over many years.

***Ex vivo* cell sorting.** PBMCs were stained with antibodies to CD8, CD4, CD14, CD16, and CD19 (BioLegend) and Live/Dead fixable aqua dead cell stain (Molecular Probes). Approximately 10^6 $CD8^+$ T cells were sorted using a FACSAria II cell sorter (BD Biosciences). For HTS, 3 μg of DNA was isolated using the QIAamp DNA blood minikit (Qiagen). Donor parameters are detailed in Table 1. Written consent was obtained from the blood donors, and the study was approved by an institutional review committee.

High-throughput sequencing. TCR sequencing was performed using the ImmunoSEQ platform, which combines template-switch anchored reverse transcription-PCR (RT-PCR) with the Illumina HiSeq system (3). Data filtering and T cell receptor beta (TRB) gene annotation were performed using a microassembler and standard algorithms as described previously (3). TRBV, TRBD, TRBJ, and CDR3 parameters were delineated according to definitions from International ImMunoGeneTics collaboration (10).

TABLE 1 Donor parameters[a]

Donor	Gender	HLA-A	HLA-B	EBV serostatus	CMV serostatus
H01	Male	23, 25	18, 49	+	−
H02	Male	24, 29	4403, 4405	+	−

[a] HLA, human leukocyte antigen.

Epstein-Barr virus-specific T cell identification. Short-term T cell cultures specific for Epstein-Barr virus (EBV) were raised as previously described (11). EBV-specific T cells were isolated using a FACSAria II cell sorter (BD Biosciences) through autologous lymphoblastoid cell line (LCL) stimulation followed by a surface tumor necrosis factor (TNF) capture assay (12). Rearranged, functional αβ TCRs were identified using TRBV gene-specific RT-PCR, bacterial subcloning, and Sanger sequencing as previously described (11).

To examine the stability of the human $CD8^+$ T cell repertoire, the $CD8^+$ T lymphocyte subsets were sorted to >99% purity from PBMC samples from a healthy Caucasian male (H01), collected at four time points between the ages of 26 and 44. The donor remained EBV seropositive and cytomegalovirus (CMV) seronegative during the sampling schedule and could recall no significant health problems during this period. High-throughput TCR sequencing of the four populations yielded an average of 4.7 million reads and 41,000 unique TCR sequences per sample. Full sequence statistics are shown in Table S1 in the supplemental material.

The 40 most frequent TCR β-chain sequences from the 1993 time point accounted for 16.7% of the total $CD8^+$ repertoire and

Received 15 August 2012 Accepted 13 October 2012
Published ahead of print 17 October 2012
Address correspondence to S. R. Burrows, scott.burrows@qimr.edu.au, or J. J. Miles, john.miles@qimr.edu.au.
S.R.B. and J.J.M. contributed equally to this work.
Supplemental material for this article may be found at http://dx.doi.org/10.1128/JVI.02180-12.

doi:10.1128/JVI.02180-12

TABLE 2 Forty TCR β-chain sequences occurring at highest frequency in $CD8^+$ T cells from donor H01 in 1993 and their frequencies in $CD8^+$ T cells in 1997, 2008, and 2011

Rank	Amino acid sequence[b]	TRBV gene(s)	TRBJ gene	Frequency in $CD8^+$ T cells by yr (%)[a]			
				1993	1997	2008	2011
1	CASSFPGNEQFF	7-9	2-1	**5.312**	*0.579*	*0.344*	*0.162*
2	CASSEPGTSQETQYF	7-9	2-5	1.285	*0.160*	*0.292*	*0.239*
3	CASSSTGSGETQYF	7-9	2-5	*0.923*	*0.214*	*0.655*	*0.278*
4	CASSFGTSSYNEQFF	7-9	2-1	*0.651*	2.044	*0.349*	*0.156*
5	CASSLGHAEAFF	7-9	1-1	*0.645*	1.277	*0.896*	*0.446*
6	***CASSEQDGFNYGYTF***	7-9	1-2	*0.598*	*0.576*	*0.772*	*0.602*
7	CASSLGDYRGYTF	7-3	1-2	*0.458*	*0.379*	0.099	0.043
8	CASSSLEGVYDEQFF	7-9	2-1	*0.423*	*0.935*	*0.224*	*0.167*
9	CASSQGVRGQHSYNEQFF	4-2	2-1	*0.416*	*0.844*	*0.389*	*0.315*
10	CASSLASNGYTF	7-9	1-2	*0.390*	0.060	0.062	0.014
11	CASSPRQGTHNEQFF	19	2-1	*0.333*	0.040	0.040	0.009
12	CASSPKLGGEQYF	7-3	2-7	*0.284*	0.074	0.000	0.009
13	CASSDQGHRDEKLFF	6-1	1-4	*0.275*	*0.184*	*0.136*	*0.126*
14	CASSLLPRHTDTQYF	18	2-3	*0.274*	*0.161*	*0.129*	*0.132*
15	CASAPPPGEGARELFF	7-9/11-1	2-2	*0.257*	0.004	0.042	0.000
16	CASFPDRGYTGELFF	7-9	2-2	*0.238*	*0.153*	0.053	0.084
17	CASRRVMSGTDTQYF	7-8	2-3	*0.226*	0.017	0.002	0.006
18	CASATWAGATDTQYF	19	2-3	*0.218*	*0.472*	0.026	0.029
19	CASSPQSLGGYTF	18	1-2	*0.211*	*0.501*	*0.438*	*0.381*
20	CASSFVPGQPQHF	7-9	1-5	*0.203*	*0.436*	*0.322*	*0.208*
21	CASIAGSFDEQFF	7-9	2-1	*0.196*	0.056	0.028	0.000
22	CASSPLPRRDSHSPLHF	18	1-6	*0.189*	*0.113*	0.037	0.029
23	CASSPTGGSYNSPLHF	7-2/11-2/11-3	1-6	*0.188*	*0.277*	*0.323*	*0.331*
24	CASSLAGGYSYEQYF	7-6	2-7	*0.174*	*0.154*	0.019	0.073
25	CASSHSRDLDYEQYF	6-5/6-6	2-7	*0.171*	0.068	0.010	0.018
26	CASSLVPWSETTGDTDTQYF	7-6/7-7	2-3	*0.169*	0.095	0.082	0.098
27	CASSRGGNNEQFF	19	2-1	*0.165*	*0.272*	0.067	0.017
28	CASSLRDASYEQYF	7-9	2-7	*0.160*	0.086	0.016	0.000
29	CASSLGAGGLEQFF	7-6	2-1	*0.157*	0.003	0.040	0.001
30	***CASSYLTADGNQPQHF***	6-2/6-3	1-5	*0.156*	0.088	*0.199*	0.086
31	CASSPIFRGLYTEAFF	7-9/11-1	1-1	*0.152*	*0.167*	0.039	0.025
32	CACNNSPLHF	30	1-6	*0.147*	0.088	0.033	0.028
33	CASVLEGFNQPQHF	6-1/6-5/6-6	1-5	*0.146*	*0.844*	*0.147*	0.061
34	CAGGTGSDTQYF	5-4	2-3	*0.146*	*0.268*	*0.163*	*0.253*
35	CASSLWGTTYEQYF	7-9	2-7	*0.143*	0.082	0.039	0.002
36	CASSPVPATYEQYF	5-6	2-7	*0.137*	*0.306*	*0.234*	*0.332*
37	CASSPSSGPYEQYF	18	2-7	*0.137*	*0.204*	*0.225*	*0.276*
38	CASSPETGILSGYTF	7-6	1-2	*0.128*	*0.121*	0.037	0.045
39	***CASTARGNTGELFF***	6-1/6-5/6-6	2-2	*0.119*	*0.129*	*0.330*	*0.370*
40	CASSLVGHYEQYF	7-9	2-7	*0.113*	0.047	0.000	0.007
Total (%)				16.712	12.576	7.333	5.458

[a] Frequency of sequence in each sample relative to the total no. of productive sequences per sample. Key: unshaded, <0.001%; shaded and regular font (not underlined), 0.001 to 0.1%; shaded and italic, 0.1 to 1%; shaded and underlined, 1 to 5%; shaded and bold, >5%.

[b] Bold and italics indicate sequences corresponding to EBV-specific clonotypes identified by polychromatic flow cytometric sorting and TCR Sanger sequencing.

were tracked over the 18-year period. Surprisingly, 35 of these clonotypes could be identified across all time points, and all 40 β-chains could be observed at a significant frequency across at least three time points (Table 2; see also Table S2 in the supplemental material). The 40 most frequent TCR β-chain sequences from 2011, which accounted for 16.3% of the total $CD8^+$ repertoire, were also tracked at the earlier three time points (see Table S3). All 40 of these 2011 β-chain sequences were observed in 2008, while 37 and 32 of the sequences were observed in 1997 and 1993, respectively. In total, 17 TCR β-chain sequences persisted in the top 40 most frequent in both 1993 and 2011. These data illustrate that the apex of the $CD8^+$ T cell repertoire can remain surprisingly stable over many decades, with minimal turnover of highly expanded clonotypes.

Interestingly, each of the prominent β-chain amino acid sequences observed at these time points was encoded by a single dominant nucleotide sequence, with alternative coding sequences observed at very low frequencies or not at all at any time point (see Tables S2 and S3 in the supplemental material). These data indicate that the selection pressures that drive these clonal expansions

TABLE 3 Forty TCR β-chain sequences occurring at highest frequency in PBMCs from donor H02 in 2002 and their frequencies in 2011

Rank	Amino acid sequence[b]	TRBV gene(s)	TRBJ gene	Frequency in PBMCs by yr (%)[a]	
				2002	2011
1	***CASRYRDDSYNEQFF***	7-9	2-1	**5.019**	**5.664**
2	CASTPGRQSTRGNQPQHF	2	1-5	*0.981*	*2.707*
3	CASSLIGSGQSYNEQFF	7-9/11-1	2-1	*0.571*	*0.279*
4	CASSLAWGWKIDTQYF	7-9/11-1	2-3	*0.415*	*0.420*
5	CASSGGSGDADTQYF	6-2/6-3	2-3	*0.258*	*0.189*
6	CASSLMGGSETQYF	7-2	2-5	*0.253*	*0.388*
7	CASSSTLPGTTPHEQYF	6-5/6-6	2-7	*0.245*	*0.131*
8	CASSYGETQYF	7-9	2-5	*0.166*	*0.151*
9	CASSRQGANEQYF	7-9	2-7	*0.158*	0.026
10	CAWRGRGAAYEQYF	30	2-7	*0.147*	*0.557*
11	CASSLRLGGAHEQYF	5-1	2-7	*0.126*	*0.151*
12	CASHTGPGNSYEQYF	6-1/6-5/6-6	2-7	*0.124*	*0.262*
13	CASSPWDQETQYF	7-2	2-5	*0.115*	0.063
14	CASSNGPGQGASETQYF	18	2-5	*0.106*	0.018
15	CASSDSLPSLPAGGGNEQFF	Undefined	2-1	*0.104*	0.007
16	CASSLYGGTSYEQYF	7-9	2-7	*0.101*	0.024
17	CASSLGFTGELFF	5-1	2-2	*0.101*	*0.120*
18	CSAPDGTSGYNEQFF	20-1	2-1	0.085	*0.126*
19	CASSLPNIRNEQFF	7-9	2-1	0.082	*0.323*
20	CSAVGGRGYTF	29-1	1-2	0.076	*0.289*
21	CASSYSSGRVGYEQFF	6-2/6-3	2-1	0.073	0.021
22	CASSYRENTEAFF	6-2/6-3	1-1	0.072	0.085
23	CASSEGELSGAETQYF	6-1	2-5	0.071	0.013
24	CASSLNTGAPGELFF	7-6	2-2	0.068	0.047
25	CASSVGTGEQYF	7-9	2-7	0.065	0.007
26	CASSTGRSPDTQYF	18	2-3	0.056	0.020
27	CASSTPTSGRQTQYF	19	2-5	0.053	0.007
28	CASRSGLFSTDTQYF	7-2	2-3	0.053	0.015
29	CASSLAYQSETSYEQYF	7-9/11-1	2-7	0.048	0.003
30	CASSPRGYPEAYEQYF	7-9/11-1	2-7	0.048	0.003
31	CASSPRGPDTQYF	7-9	2-3	0.046	0.019
32	CASSPNGGATNEKLFF	18	1-4	0.044	0.000
33	CASLGPGRGLRGYTF	7-9	1-2	0.044	0.002
34	CASSPVAGDNEQFF	18	2-1	0.044	0.008
35	CASSFITDTQYF	7-2	2-3	0.043	0.000
36	CASSYSFSSVGYEQFF	6-2/6-3	2-1	0.043	0.018
37	CASSLLGQGKDTGELFF	7-6	2-2	0.043	0.042
38	CAWRYRGGNTEAFF	30	1-1	0.042	0.000
39	CASTDNTYEQYF	2	2-7	0.041	0.028
40	CASSVPTESSIQYF	19	2-4	0.041	0.000
Total (%)				10.272	12.234

[a] Frequency of sequence relative to the total no. of productive sequences. Key: unshaded, <0.001; shaded, in regular font and not underlined, 0.001 to 0.1; shaded and italic, 0.1% to 1%; shaded and underlined, 1% to 5%; shaded and bold, >5%.

[b] Bold and italics indicate sequence corresponding to EBV-specific clonotype identified previously (11).

have acted upon individual clonotypes, which, once expanded to large numbers, remove any further pressure for the expansion of additional clonotypes with the same specificity from the naive T cell repertoire.

We also performed TCR HTS on the PBMCs of a second healthy Caucasian donor (H02) at the ages of 58 and 67. This donor also remained EBV seropositive and CMV seronegative during the sampling schedule and had no significant health problems. The 40 most frequent TCR β-chain sequences from this donor in 2002 occupied 10.3% of the total T cell repertoire, and 36 of these β-chains were also found in 2011, many at high frequencies, filling a total of 12.2% of the repertoire at this recent time point (Table 3; see also Table S4 in the supplemental material). These data support the conclusion that high-frequency T cell clonotypes can persist for many years in adults. As with donor H01, each of the prominent β-chain amino acid sequences was encoded by a single dominant nucleotide sequence, with alternative coding sequences observed at very low frequencies or not at all (see Table S4).

The most frequent and persistent TCR β-chain in donor H02,

TABLE 4 TCR β-chain sequences and longitudinal analysis of EBV-specific clonotypes within the $CD8^+$ compartment of donor H01

CDR3 sequence	TRBV gene(s)	TRBJ gene	Frequency rank by yr[a]			
			1993	1997	2008	2011
CASSEQDGFNYGYTF	7-9	1-2	6	7	6	6
CASSYLTADGNQPQHF	6-2/6-3	1-5	30	61	23	47
CASTARGNTGELFF	6-1/6-5/6-6	2-2	39	37	14	10
CAISDPPGGVDEQYF	10-3	2-7	47	102	68	67
CASTSSAGLDTQYF	5-4	2-3	55	125	28	19
CSAYNRGDEAYEQYF	20-1	2-7	70	151	145	113
CASNQGGADTQYF	19	2-3	75	109	551	175
CSGGVPNTGELFF	29-1	2-2	1,446	614	ND	29,800
CASSLWETQYF	28	2-5	44,058	33,700	ND	ND
CASSPEGSFEPQHF	12-3/12-4	1-5	15,106	17,427	12,318	10,528
CSARSGTDRIEQYF	20-1	2-7	ND	ND	11,049	1,383
CASSYGEGAWNEQFF	6-2/6-3	2-1	ND	ND	ND	4,011
CSARDAGQEYEQYF	20-1	2-7	ND	ND	ND	28,921

[a] Frequency rank of β-chain sequence among all sequences identified within the $CD8^+$ subset using HTS. ND, not determined.

which was encoded by the TRBV7-9 and TRBJ2-1 gene segments, comprised more than 5% of the total repertoire in both 2002 and 2011. Interestingly, this β-chain corresponds to a T cell clonotype that recognizes the EBV nuclear antigen 3C and was previously identified for this donor using conventional sequencing techniques (8) and which has a high affinity for its target EBV epitope (13). It is well known that human herpesviruses trigger large $CD8^+$ memory T cell expansions which inflate with age (14–18). To investigate which of the TCR β-chain sequences identified by HTS in donor H01 were herpesvirus specific, we used LCL stimulation and TNF capture to sort EBV-specific T cells from the 2011 sample, which were then utilized in cloning and TCR Sanger sequencing. We identified 13 EBV-specific β-chains that corresponded to sequences from the HTS analysis of the total $CD8^+$ subset (Table 4). These included three highly prevalent β-chains that were within the 40 most frequently observed sequences from the 1993 or 2011 HTS data (Table 2; see also Tables S2 and S3 in the supplemental material). It is notable that the TNF capture technique may have underestimated the proportion of TCR β-chains that were EBV specific because only functionally "sound" $CD8^+$ T cells are detected, with the exclusion of T cells that fail to secrete TNF.

In summary, this study has investigated the evolution of the human T cell repertoire in peripheral blood and revealed that the high-frequency clonotypes can remain surprisingly stable over long periods of time. Furthermore, these high-frequency, ever-present clones occupy a considerable amount of immunological space and include several Epstein-Barr virus-specific expansions. Although our archived PBMC numbers were too limited for further sorting into smaller subsets, future studies aimed at characterizing the phenotype of the long-lived T cell clonotypes will be worth pursuing. It will also be important to investigate the mechanisms that control T cell longevity due to its potential importance in influencing the efficacy of T cell-based vaccines and adoptive therapies.

ACKNOWLEDGMENTS

J.J.M. is an NHMRC Career Development Fellow (APP1031652), and S.R.B. is an NHMRC Principal Research Fellow (APP1021452). This work was supported by an NHMRC program grant (APP389830).

REFERENCES

1. **Arstila TP, Casrouge A, Baron V, Even J, Kanellopoulos J, Kourilsky P.** 1999. A direct estimate of the human alphabeta T cell receptor diversity. Science **286**:958–961.
2. **Arstila TP, Casrouge A, Baron V, Even J, Kanellopoulos J, Kourilsky P.** 2000. Diversity of human alpha beta T cell receptors. Science **288**:1135.
3. **Warren RL, Freeman JD, Zeng T, Choe G, Munro S, Moore R, Webb JR, Holt RA.** 2011. Exhaustive T-cell repertoire sequencing of human peripheral blood samples reveals signatures of antigen selection and a directly measured repertoire size of at least 1 million clonotypes. Genome Res. **21**:790–797.
4. **Miles JJ, Douek DC, Price DA.** 2011. Bias in the alphabeta T-cell repertoire: implications for disease pathogenesis and vaccination. Immunol. Cell Biol. **89**:375–387.
5. **Barber DL, Wherry EJ, Masopust D, Zhu B, Allison JP, Sharpe AH, Freeman GJ, Ahmed R.** 2006. Restoring function in exhausted CD8 T cells during chronic viral infection. Nature **439**:682–687.
6. **Vezys V, Masopust D, Kemball CC, Barber DL, O'Mara LA, Larsen CP, Pearson TC, Ahmed R, Lukacher AE.** 2006. Continuous recruitment of naive T cells contributes to heterogeneity of antiviral CD8 T cells during persistent infection. J. Exp. Med. **203**:2263–2269.
7. **Barcy S, Huang ML, Corey L, Koelle DM.** 2005. Longitudinal analysis of herpes simplex virus-specific CD4+ cell clonotypes in infected tissues and blood. J. Infect. Dis. **191**:2012–2021.
8. **Miles JJ, Silins SL, Brooks AG, Davis JE, Misko I, Burrows SR.** 2005. T-cell grit: large clonal expansions of virus-specific CD8+ T cells can dominate in the peripheral circulation for at least 18 years. Blood **106**:4412–4413.
9. **van Bockel DJ, Price DA, Munier ML, Venturi V, Asher TE, Ladell K, Greenaway HY, Zaunders J, Douek DC, Cooper DA, Davenport MP, Kelleher AD.** 2011. Persistent survival of prevalent clonotypes within an immunodominant HIV gag-specific CD8+ T cell response. J. Immunol. **186**:359–371.
10. **Yousfi Monod M, Giudicelli V, Chaume D, Lefranc MP.** 2004. IMGT/JunctionAnalysis: the first tool for the analysis of the immunoglobulin and T cell receptor complex V-J and V-D-J JUNCTIONs. Bioinformatics **20**(Suppl 1):i379–i385.
11. **Miles JJ, Elhassen D, Borg NA, Silins SL, Tynan FE, Burrows JM, Purcell AW, Kjer-Nielsen L, Rossjohn J, Burrows SR, McCluskey J.** 2005. CTL recognition of a bulged viral peptide involves biased TCR selection. J. Immunol. **175**:3826–3834.
12. **Haney D, Quigley MF, Asher TE, Ambrozak DR, Gostick E, Price DA, Douek DC, Betts MR.** 2011. Isolation of viable antigen-specific CD8+ T cells based on membrane-bound tumor necrosis factor (TNF)-alpha expression. J. Immunol. Methods **369**:33–41.
13. **Archbold JK, Macdonald WA, Gras S, Ely LK, Miles JJ, Bell MJ, Brennan RM, Beddoe T, Wilce MC, Clements CS, Purcell AW, McCluskey J, Burrows SR, Rossjohn J.** 2009. Natural micropolymorphism in human leukocyte antigens provides a basis for genetic control of antigen recognition. J. Exp. Med. **206**:209–219.
14. **Crough T, Khanna R.** 2009. Immunobiology of human cytomegalovirus: from bench to bedside. Clin. Microbiol. Rev. **22**:76–98.
15. **Hislop AD, Taylor GS, Sauce D, Rickinson AB.** 2007. Cellular responses to viral infection in humans: lessons from Epstein-Barr virus. Annu. Rev. Immunol. **25**:587–617.
16. **Khan N, Hislop A, Gudgeon N, Cobbold M, Khanna R, Nayak L, Rickinson AB, Moss PA.** 2004. Herpesvirus-specific CD8 T cell immunity in old age: cytomegalovirus impairs the response to a coresident EBV infection. J. Immunol. **173**:7481–7489.
17. **Khan N, Shariff N, Cobbold M, Bruton R, Ainsworth JA, Sinclair AJ, Nayak L, Moss PA.** 2002. Cytomegalovirus seropositivity drives the CD8 T cell repertoire toward greater clonality in healthy elderly individuals. J. Immunol. **169**:1984–1992.
18. **Pourgheysari B, Khan N, Best D, Bruton R, Nayak L, Moss PA.** 2007. The cytomegalovirus-specific CD4+ T-cell response expands with age and markedly alters the CD4+ T-cell repertoire. J. Virol. **81**:7759–7765.