Supplementary materials

1. **Supplementary Table 1.** Optimized primer set for the human TCR alpha and TCR beta libraries preparation with UMI-barcoded RACE.

2. **Supplementary Table 2.** TCR profiling of small PBMC and CD8 memory T cells samples.

3. **Supplementary Table 3**. UMI-based analysis eliminates residual PCR and sequencing errors.

4. **Supplementary Table 4**. Top-10 clonotypes ranks and concentrations in technical replicas of 4000 PBMC (2,000 T cells).

Supplementary Table 1. Optimized primer set for the human TCR alpha and TCR beta libraries preparation with UMI-barcoded RACE.

Primer name	Application	Sequence
1st strand cDNA synthesis performed jointly for the TCR alpha and beta libraries		
ACR_st4	Primers for TCR alpha and TCR beta cDNA synthesis	GTCTAGCACAGTTTTGTC
BC_R4short		GTATCTGGAGTCATTGA
SmartNNNa	Template switch adapter with unique molecular identifier	AAGCAGUGGTAUCAACGCAGAGUNNNNUNNNNUNNNNUCTT(rG)$_4$
1st PCR amplification performed jointly for the TCR alpha and beta libraries		
M1ss	Primer anneals on the template switch adapter	AAGCAGTGGTATCAACGCA
ACR_st1	Reverse primers for TCR alpha and beta chains	GTCACTGGATTTAGAGTC
BC2R		TGCTTCTGATGGCTCAAACAC
2nd PCR amplification performed separately for the TCR alpha and beta libraries		
M1s_iNNN	Primer with sample barcode, anneals on the switch adapter, nested compared to M1ss	(N)$_{2-4}$(XXXXX)CAGTGGTATCAACGCAGAG
acj_H_i*	Reverse primer for TCR alpha chain with sample barcode	(N)$_{2-4}$(XXXXX)GGGTCAGGGTTCTGGATAT
bcj_H*	Reverse primer for TCR beta chain with sample barcode	(N)$_{2-4}$(XXXXX)ACACSTTKTTCAGGTCCTC

* In the first experiment these primers were used without sample barcodes.

U = dU

XXXXX = sample barcode

$(N)_{2-4}$ = from 2 to 4 random nucleotides introduced in order to generate diversity for better cluster differentiation on Illumina. These nucleotides also protect sample barcode from damage by polymerase proof-reading activity or during adapter ligation.

Supplementary Table 2. TCR profiling of small PBMC and CD8 memory T cells samples.

a. TCR profiling of 8 replicas of ~4,000 PBMC (~2,000 T cells).

Replicate library	sequencing reads with CDR3, MiTCR analysis	sequencing reads used in MIGEC analysis (%)*	in-frame cDNA, MIGEC analysis*	functional clones diversity, MIGEC analysis*
TCR beta 1	344,740	303,121 (87,93%)	1,944	882
TCR beta 2	351,908	294,307 (83,63%)	1,520	690
TCR beta 3	575,639	475,617 (82,62%)	3,473	1,414
TCR beta 4	617,786	481,992 (78,02%)	6,395	2,558
TCR beta 5	614,788	521,669 (84,85%)	3,077	1,285
TCR beta 6	574,116	493,165 (85,90%)	2,564	1,098
TCR beta 7	702,195	576,752 (82,14%)	4,296	1,860
TCR beta 8	652,733	523,044 (80,13%)	3,154	1,410
TCR alpha 1	61,860	56,625 (91,54%)	658	349
TCR alpha 2	81,391	75,540 (92,81%)	489	276
TCR alpha 3	118,577	111,420 (93,96%)	1,060	542
TCR alpha 4	218,138	182,907 (83,85%)	1,830	812
TCR alpha 5	109,766	93,269 (84,97%)	1,147	505
TCR alpha 6	85,013	71,730 (84,38%)	888	404
TCR alpha 7	112,672	93,301 (82,81%)	1,246	583
TCR alpha 8	66,042	52,012 (78,76%)	965	463

* only in-frame TCR variants, each cDNA sequenced at least 4 times.

b. TCR profiling of sorted replicas of 500 and 1,000 CD8 memory T cells.

Replicate library	T cells sorted	sequencing reads with CDR3, MiTCR analysis	sequencing reads used in MIGEC analysis (%)*	in-frame cDNA, MIGEC analysis*	functional clones diversity, MIGEC analysis*
TCR beta 1	1000	180,630	166,889 (92,4)	440	70
TCR beta 2	1000	1,087,649	1,049,782 (96,5)	583	81
TCR beta 3	1000	1,911,724	1,846,015 (96,6)	550	78
TCR beta 4	1000	1,292,448	1,246,884 (96,5)	513	86
TCR beta 5	500	417,645	395759 (94,8)	262	46
TCR beta 6	500	669,493	643,004 (96,0)	279	50
TCR beta 7	500	194,914	183,315 (94,0)	260	50
TCR beta 8	500	1,414,272	1,327,419 (93,9)	300	49
TCR alpha 1	1000	259,257	225,466 (87,0)	122	39
TCR alpha 2	1000	212,934	202,727 (95,2)	193	51
TCR alpha 3	1000	644,902	615,095 (95,4)	196	52
TCR alpha 4	1000	379,840	360,776 (95,0)	172	47
TCR alpha 5	500	71,162	67,176 (94,4)	87	29
TCR alpha 6	500	469,045	427,413 (91,1)	108	27
TCR alpha 7	500	87,607	82,075 (93,7)	82	26
TCR alpha 8	500	193,983	157,914 (81,4)	63	17

* only in-frame TCR variants, each cDNA sequenced at least 5 times.

Supplementary Table 3. UMI-based analysis eliminates residual PCR and sequencing errors.

CDR3 nucleotide sequence	CDR3 amino acid sequence	Standard analysis, reads	UMI-based analysis, cDNA events (reads used)
Major CMV-HLA*A02-NLV specific TCR beta clonotype			
TGTGCCAGCAGCTTAGCGCCGGGAGCAACTAATGAAAAACTGTTTTTT	**CASSLAPGATNEKLFF**	**7,917**	**55 (7,235)**
TGTGCCAGCGGCTTAGCACCGGGAGCAACTAATGAAAAACTGTTTTTT	CASGLAPGATNEKLFF	4	-
TGTACCAGCAGCTTAGCGCCGGGAGCAACTAATGGAAAACTGTTTTTT	CTSSLAPGATNGKLFF	2	-
TGTGCCAGCAGCTTAGCGCCGGGAGCAGCTAATGAAAAACCGTTTTTT	CASSLAPGAANEKPFF	2	-
TGTACCAGCAGCTTAGCGCCGGGAGCAACTAATGAAAAACTGCTTTTT	CTSSLAPGATNEKLLF	1	-
TGTACCAGCAGCTTAGCGCCGGGGGCAACTAGTGAAAAACTGTTTTTT	CTSSLAPGATSEKLFF	1	-
TGTGCCAGCAGCTTAGCGCCGAGAGCAACTAATGAAGAACTGTTTTTT	CASSLAPRATNEELFF	1	-
TGTGCCAGCAGCTTAGCGCCGGGAGCAACAAATGAAAAACTGGTTTTT	CASSLAPGATNEKLVF	1	-
TGTGCCAGCAGCTTAGCGCCGGGAGCAACAAATGAAGAACTGTTTTTT	CASSLAPGATNEELFF	1	-
TGTGCCAGCAGCTTAGCGCCGGGATCAACTAATGAAGAACTGTTTTTT	CASSLAPGSTNEELFF	1	-
TGTGCCAGCGGCTTAGCGCCGGGAGCAACTGATGAAAAACTGTTTTTT	CASGLAPGATDEKLFF	1	-
Minor CMV-HLA*A02-NLV specific TCR beta clonotype			
TGTGCCAGCAGCTTAGCCCCCGGGGCAACTAATGAAAAACTGTTTTTT	**CASSLAPGATNEKLFF**	**201**	**1 (184)**
TGTGCCAGCAGCTTAGCCCCCGGGGCAACTAACAAAAAACTGTTTTTT	CASSLAPGATNKKLFF	1	-

Accumulated errors in two CMV-specific TCR beta CDR3 clonotypes are shown for the single replica of 4,000 PBMC. True TCR beta sequences studied in previous works[2, 3] are shown bold. Undercorrected errors are shaded gray. TRBV, TRBD, and TRBJ gene segments are shown red, blue, and green, correspondingly. Although MiTCR efficiently removed erroneous subvariants with single mismatches using frequency-based clustering, those with double mismatches remained unfiltered according to the MiTCR algorithm that aims to preserve natural TCR diversity. In contrast, UMI-based MIGEC analysis does not employ any frequency-based clusterisation of homologous variants (i.e. fully preserves natural diversity of the sample) but eliminates all artificial subvariants.

Supplementary Table 4. Top-10 clonotypes ranks and concentrations in technical replicas of 4000 PBMC (2,000 T cells).

TCR beta CDR3 amino acid sequence	Average clonotype concentration, %	Geometric deviation	Clonotype rank (color) and concentration (number) in technical replicas							
			1	2	3	4	5	6	7	8
Standard analysis										
CSVEIWDSSYNEQFF	11.49	1.2	13.6	11.4	9.7	10.6	11.5	15.3	10.6	9.5
CASSLGENIQYF	8.73	1.1	10.3	9.1	9.4	9.0	9.6	7.5	7.1	7.6
CSARTTYGTDIISQHF	2.69	1.2	2.6	2.5	2.9	2.7	2.9	2.1	3.5	3.2
CASSLAPGATNEKLFF	2.39	1.2	2.4	2.5	2.6	2.5	2.4	2.1	2.4	2.1
CSVEEWASRYNEQFF	1.61	1.4	1.7	2.5	1.6	1.5	2.0	2.0	1.6	1.8
CSVGTSEAYEQYF	1.39	1.4	1.4	2.2	1.4	1.1	1.8	1.6	1.5	1.2
CASTVDSLDTEAFF	1.18	1.6	1.2	2.2	1.0	0.9	1.5	1.4	1.4	1.1
CASSVALGLNYEQYF	1.00	1.5	1.2	2.0	1.0	0.8	1.2	1.4	1.4	1.1
CSVADSTYEQYF	0.78	2.0	1.1	1.2	0.9	0.8	1.1	1.2	0.9	1.1
CASSFGTFGAYGYTF	0.73	1.6	1.1	1.1	0.9	0.6	1.0	1.0	0.8	0.9
UMI-based analysis										
CASSLGENIQYF	12.21	1.1	11.4	13.5	13.4	10.6	12.0	12.5	12.0	12.2
CSVEIWDSSYNEQFF	7.29	1.2	7.0	8.3	4.8	7.6	7.3	9.1	7.4	6.6
CASSLAPGATNEKLFF	3.65	1.1	4.2	3.6	3.7	2.7	4.2	3.6	3.6	3.6
CASSVALGLNYEQYF	1.51	1.1	1.6	1.5	1.6	1.7	1.9	1.6	1.9	1.5
CASTVDSLDTEAFF	1.48	1.3	1.5	1.3	1.6	1.6	1.6	1.1	1.3	1.3
CSARTTYGTDIISQHF	1.30	1.2	1.5	1.1	1.5	1.5	1.3	1.1	1.2	1.0
CASSKARWDFTANVLTF	0.85	1.2	1.2	1.1	1.1	0.8	0.9	1.1	0.9	1.0
CSVGTSEAYEQYF	0.80	1.2	1.1	1.0	0.9	0.8	0.9	1.1	0.8	1.0
CSVEEWASRYNEQFF	0.77	1.3	1.0	0.9	0.8	0.8	0.9	0.9	0.7	0.9
CASSFGTFGAYGYTF	0.75	1.3	0.8	0.9	0.7	0.8	0.8	0.9	0.6	0.9

References

1. Li, Y. et al. An improved one-tube RT-PCR protocol for analyzing single-cell gene expression in individual mammalian cells. *Analytical and bioanalytical chemistry* **397**, 1853-1859 (2010).
2. Britanova, O.V. et al. First autologous hematopoietic SCT for ankylosing spondylitis: a case report and clues to understanding the therapy. *Bone Marrow Transplant* **47**, 1479-1481 (2012).
3. Turchaninova, M.A. et al. Pairing of T-cell receptor chains via emulsion PCR. *Eur J Immunol* **43**, 2507-2515 (2013).