Whitepaper

Cerba Research

Decoding a T-Cell Fingerprint

All you need to know about TCR sequencing

Revised April 2023



The authors

Goran Marinković, Ilse Wolters, Irene Chan, Dorota Talady, Nestor Vazquez, Henk-Jan van den Ham, Wilco Knetsch, Dirk van Alewijk

The adaptive component of the immune system, also known as the acquired immunity, in humans is composed of a highly specialized and highly diverse populations of B and T lymphocytes poised to recognize and eliminate invading pathogens or other sources of unknown antigens.

To prime and activate the T-cell mediated adaptive component of the immune response, antigens are to be displayed on the surface of infected, mutated, or antigen presenting cells. To ensure the efficient recognition, antigen derived peptides are presented in a form bound to human leukocyte antigen (HLA) complex. Once exposed in this way, T-cells can engage antigens via its own membrane bound T-cell receptors or TCRs. To ensure detection of a large variety of antigens, originating from invading pathogens as well as the abnormal ones originating from the host, there is an estimated 10^{15} to 10^{19} unique TCRs rearrangements in humans[1,2]. This broad TCR diversity is known as T-cell repertoire.

Next-generation sequencing (NGS) technology enables us to detect these T-cell clones, measure and utilize parameters such as the diversity and richness and provide much needed answers in the field of oncology, immuno-oncology, cell and gene therapy, and vaccine development. The power of NGS to decode the immune repertoire is used for preclinical, clinical and diagnostic purposes, development of new therapeutics, evaluation of current immune status and monitoring therapy effect.

Dose finding

More direct effect monitoring on immune system

Antibody selection

Find more candidates with less developmental liabilities

Target discovery

Enable high–throughput discovery capabilities e.g., for T-cell based therapies

Patient stratification

Predict responders and identify non–responders to immunotherapy to accelerate clinical development

Treatment monitoring

Longitudinal patient monitoring in clinical trials

Companion diagnostics

Obtain market authorization for clinical decision support

How is a TCR Diversity Created?

A T-cell receptor (TCR) is a cell surface protein, anchored in the membrane of the T lymphocyte and responsible for the recogition of the specific antigens presented by the HLA complex.

Every T-cell receptor clone has a unique combination of alpha/beta heterodimer polypeptide chains that contribute to the overall diveristy and enable high selectivity in binding to a specific antigen.



Figure 1: Gene rearrangement in T-cell receptor, type of reads, and downstream applications

The Diversity of TCRs arise from the process called somatic or V(D)J recombination.

Somatic recombination is a type of genetic rearrangement that alters original DNA information. In case of the T-cell receptor, such rearrangement takes place between variable (V), diversity (D) and joining (J) gene segments prior to transcription (Figure 1). The insertion or deletion of nucleotides at the junction points of these rearranged V(D) J gene segments generates even greater diversity of the TCR repertoire.

Each polypeptide chain of a TCR alpha/beta heterodimer is composed of two extracellular domains: a variable (V) and a constant (C) domain, as well as a transmembrane domain and a short cytoplasmic tail. About 95% of T-cells have TCRs with an alpha and beta chain. The remaining 5% have TCRs composed of delta and gamma chains. Human TCRs undergo somatic recombination within the variable domain of three gene segments of the beta chain (V, D, J) and two gene segments of alpha chain (V, J). Each variable domain of both the TCR-alpha and TCR-beta consist of three complementary-determining regions (CDRs) that recognize and bind to the antigen/HLA complex.



Depending on the scientific question and the quality of the examined sample it is possible to obtain two types of NGS reads, short or long reads.

Short reads of a sequence relate to sequences within the CDR3 domain, and can be used when lacking high quality material, for example, to inspect tumor heterogeneity in formalin fixed paraffin embedded (FFPE) material. Long reads enable the decoding of an entire T-cell receptor sequence but can be used only when high quality material is available and for example, to examine effectiveness of an immunotherapy **(Figure 1).**

Cerba Research: one-stop shop

Cerba Research powers up with its partner ENPICOM to provide tailored, End to End solutions to its customers in decoding immune repertoires (Figure 2).

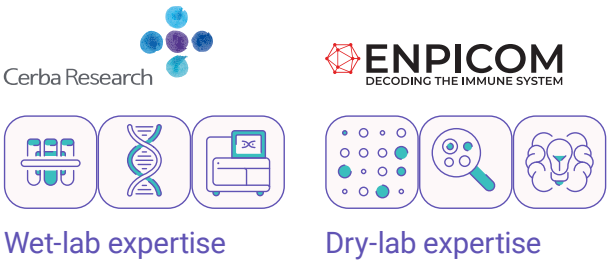
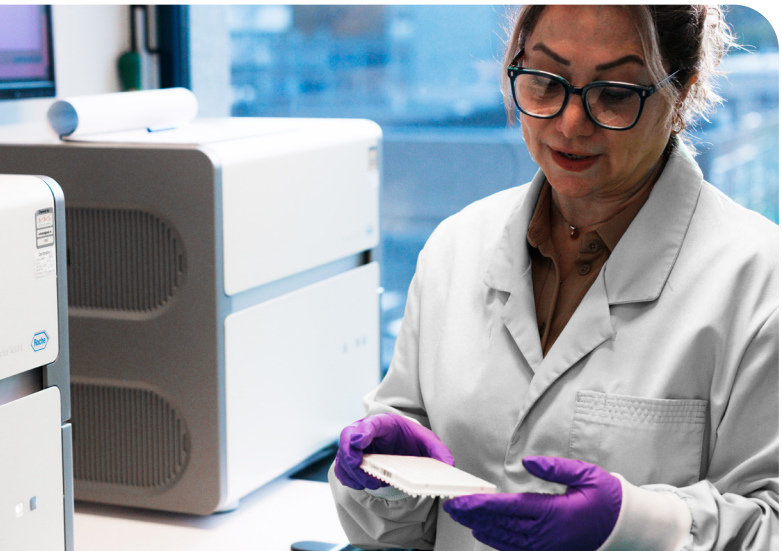


Figure 2: Schematic overview of the End-to-End workflow at Cerba Research

Cerba Research utilizes the Takara SMARTer TCR Profiling Kit that couples SMART (Switching Mechanism at 5' End of RNA Template) technology (Figure 10) with 5' RACE approach, for unbiased gene and clonal amplification of TCR repertoire sequences. Repertoires are sequenced on Illumina MiSeq sequencers and analyzed using the IGX platform developed by our partner ENPICOM for sensitive, accurate and comprehensive TCR profiling. Cerba Research will also support the development of a tailor-made novel research, therapeutic and diagnostic experimental designs, accompanied by advanced bioinformatic analysis.



TCR alpha/beta chain analysis: An example

Several factors can affect quality of the TCR repertoire profiling. Due to the immense diversity of T-cell clones in the peripheral blood, the recovered repertoire can vary to a significant degree even when taken from the same sample. The purpose of the following TCR profiling experiment example is to assess the effect of different sampling points (e.g. biological and technical replicates) on the similarity of the obtained repertoire profiles. For this demonstration, RNA material was isolated from human peripheral blood mononuclear cells (PBMCs), and aliquots were made to mimic biological and technical replicates.

Following this, RNA quality assessment was performed (Figure 3). To address the question, a standard input amount of 140 ng of RNA was used. To assess quantity and ensure quality of the isolated material, each sample was examined using the RiboGreen assay and Bioanalyzer. RNA samples with a RNA integrity number (RIN) value of 7 or higher are considered of a sufficient quality to proceed with unbiased amplification of the TCR repertoire. As shown in figure 3 all PBMC samples are of good RNA quality and have sufficient RNA yield for downstream processing.

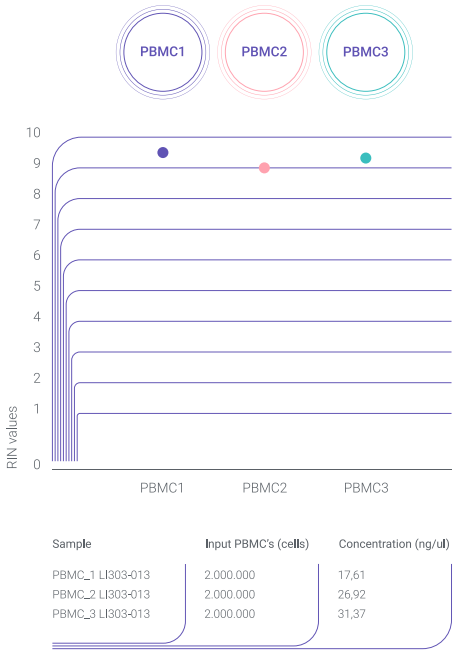


Figure 3: All samples in the experiment had higher than required quality of RNA material reflected in the RIN value higher than 7.

TCR profiling: the analysis

Once sequencing of your samples has been performed, decoding the TCR repertoire profile data is not a straightforward process and requires knowledge and practical expertise in several different disciplines including immunology, molecular biology, and bioinformatics.

The Cerba Research/ENPICOM team harbors the excellence in all relevant areas to provide our customers with accurate and understandable immune repertoire sequencing data. This can be either in a standard analysis format or tailored to address specific customer questions.

At Cerba Research we report the TCR repertoire in a standard data analysis format. Our standard data analysis is comprised of 10 parameters and includes:

- Raw/trimmed sequencing data quality
- Number of reads per UMI
- Total/unique clones
- Correlation analysis
- CDR3 length
- Richness
- Diversity
- V(D)J usage (top clones)
- AA sequence (top clones)
- Clonal distribution

Some of these standard reporting elements that are illustrating the immense diversity of TCRs are described more in detail below using the human PBMC sample as an example.

Number of reads per UMI & CDR3 length as a quality check

To increase the amount of available cDNA molecules needed for a successful repertoire profile sequencing run, the library preparation process includes a PCR amplification step. This amplification can preferentially multiply some cDNA molecules over others and introduce amplification bias toward a certain clone otherwise not overwhelmingly present in the sample. Additionally, PCR can stochastically introduce errors and propagate them further throughout the library preparation process and sequencing run. To avoid this, our workflow introduces unique molecular identifiers (UMIs) at an early stage of the process.

Unique molecular identifiers (UMIs) are random oligonucleotide barcodes added in early during the reverse transcription step. In this way, every unique, newly synthesized cDNA molecule is UMI-labeled and every cDNA-UMI represents a unique transcript.



UMI labeling strategy in combination with next generation sequencing (NGS) yields multiple reads of transcript per UMI. This ensures that every transcript can be corrected for amplification and sequencing errors.

Conclusion: In our experimental example we observed that the number of reads per UMI in all samples (with the input of 9 samples) averages around 10 **(Figure 4).** This is an optimal number of reads. As expected, all technical replicates have similar number of reads per UMI with a slight deviation of one technical replicate (PBMC1-1).

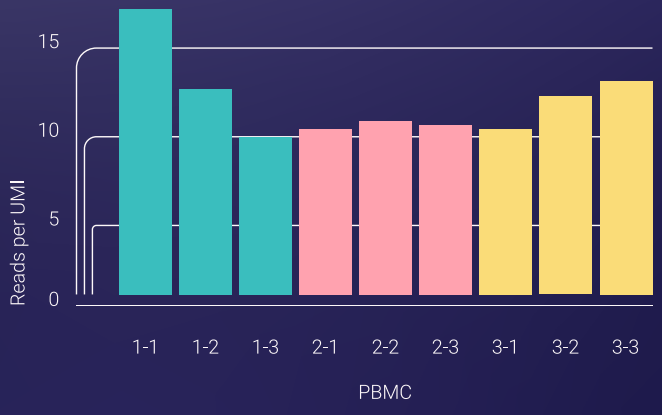


Figure 4: TCR; average number of reads per UMI.

As mentioned above, alpha-beta chains composed T-cell receptors can engage only antigens presented by the HLA complex. Affinity of TCR for the HLA-antigen complex depends greatly on the structure and sequence of the complementary determining regions (CDRs). The immune receptor complementarity determining region-3 (CDR3) is a domain that determines specificity of T-cell receptors for the respective antigen. The length of an average CDR3 domain varies between 4 and 22 amino acids[3]. Therefore, a lot of information about T-cell clones can be obtained by decoding the CDR3 sequence.

In the field of immuno-oncology, the CDR3 sequence can be used to assess tumor mutational burden (TMB).

Therefore, it can have a prognostic value. By providing insights into sequence of this region, we can better understand the nature of its interaction with its antigen and by that potentiate the prospects for successful development of tumor-infiltrating lymphocyte (TIL)-based therapies and other immunotherapies. Due to the varying length of the region, within the determined range, it can be used as a quality check parameter too. To assess the length of the recovered CDR3 domain, a spectratype plot is made (Figure 5).

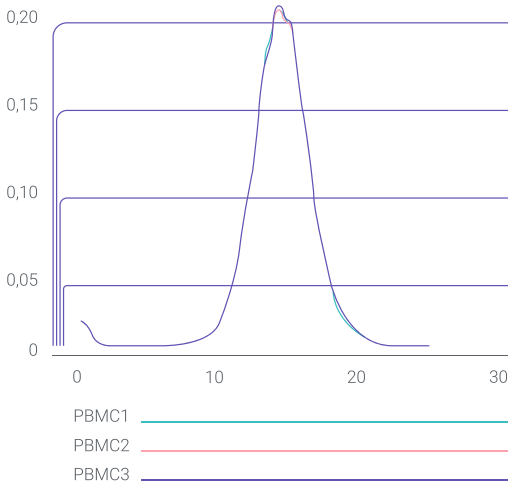


Figure 5: T-cell receptor beta spectratype, CDR3 length distribution is equal in all three biological samples.

Conclusion: Average length of the CDR3 domain varies between 8 and 21 amino acids and in all three biological samples analyzed in this example it is within the expected range.



Number of total and **unique clones** detected

The estimated number of different T-cell receptors that can be obtained through VDJ recombination ranges between 10^{15} and 10^{19} possible combinations[1,2]. Taking into consideration this enormous potential for biological diversity in the TCR repertoire, it is important to know whether different spatio-temporal sampling points can have the effect on size and the diversity of the recovered repertoire. For this experiment, an equal amount of RNA material (140 ng) was used to evaluate sampling effect on the obtained TCR profile. Using three technical replicates per sample, it was expected that similar number of both, total and unique clones, would be recovered with a similar ratio between two.

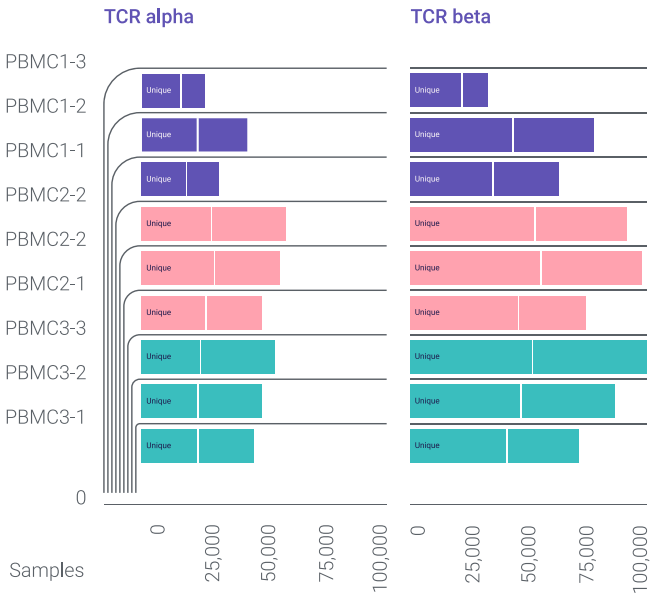


Figure 6: T-cell receptor beta spectratype, CDR3 length distribution is equal in all three biological samples.



Conclusion: The total number of recovered receptors, although similar, shows a degree of variability, particularly regarding the first technical replicate of the first sample **(Figure 6).** This is the same sample that had noticeably a higher average number of reads per UMI. Consequently, this might result in the observed decrease in the number of total and unique clones. However, the number of unique recovered receptors is highly similar between replicates in all other samples, for both the alpha and beta chains.

A significantly larger number of recovered beta chains, both total and unique, reproduces an expected higher biological diversity of this chain. In conclusion, we observe a highly similar number of recovered clones with slight variations which can be ascribed to the vast diversity poll of the TCR receptors in a PBMC sample.

Similarity between technical replicates

Evaluation of similarity between technical replicates is performed by calculating the Jaccard Similarity Coefficient.

The similarity of both alpha and beta chains was compared between samples (Figure 7).



Figure 7: T-cell receptor alpha and beta chain similarity plots: Samples are plotted against each other, and the Jaccard Similarity Coefficient was calculated.

Conclusion: A high degree of similarity is noticeable between all samples, although it is slightly higher in the sequenced TCR beta chain pool compared to the TCR alpha chain pool. This can be explained by the higher biological diversity leading subsequently to the larger number of reads of this chain over the alpha.

When comparing individual samples, a slightly lowered similarity coefficient is noticeable when plotting the first technical replicate of the first sample against any other. This can be attributed to the lower number of recovered receptors in this sample (Figure 6).



Diversity & richness

In theory, samples with one or more dominant clones will have lower diversity and higher clonality than a sample where no dominant clones are detected.

Parameters such as diversity and richness are of pivotal importance in the field of immuno-oncology as it is noted that certain treatments can affect them significantly.

For example, they can be effectively utilized as a metric of the patient response to immune check point therapy (ICT). Here, we aim to demonstrate the effect of sampling on the similarity of the obtained richness and diversity profiles. Using three technical replicates per sample, similar profile of these parameters is expected between technical replicates. To assess diversity and richness between biological and technical replicates, we plotted the number of unique clones recovered **(Figure 8)**.



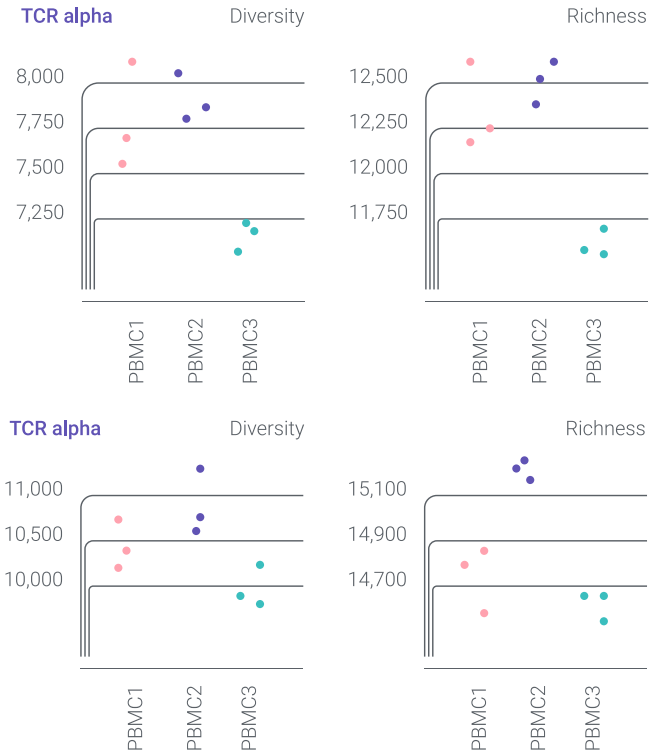


Figure 8: T-cell receptor alpha and beta chain diversity and richness plots. Reproducible recovery is observed in all samples

Conclusion: Recovered diversity and richness reflected in the number of observed clones is comparable between biological samples and highly similar between technical replicates.

Top 20 T-Cell clones, CDR3 amino acid sequence

Analysis of the top 20 clones is performed to examine the clonal distribution within the samples, depth of the reads per clone, and the similarity in the number of reads between technical replicates. The top twenty most abundant clones, according to their CDR3 domain (amino acid sequence), were chosen for the graphical representation of the results.

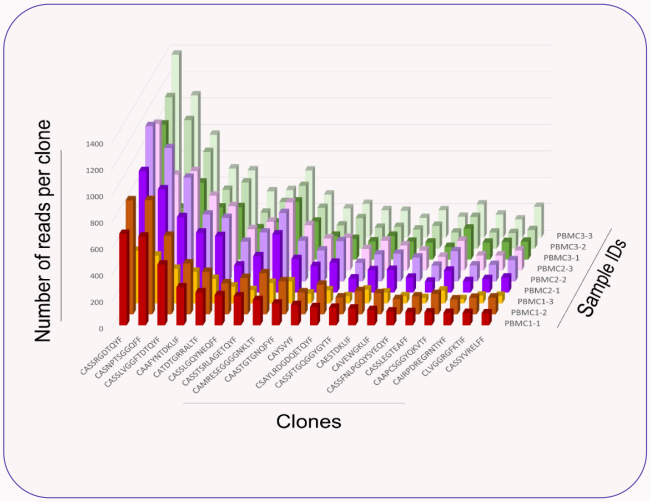


Figure 9: Top 20 clones (CDR3 region). Same top clones are recovered from all nine replicates.

Conclusion: A high overlap among most abundant clones is observed between technical replicates. The number of reads for each of the top twenty recovered clones is highly similar between technical replicates (Appendix - Table 1). The same dominant clones are recovered from the first technical replicate of the first sample, regardless of the lower number of reads previously observed.

Conclusion

In this whitepaper we present the TCR sequencing workflow utilized at Cerba Research and discuss its features and benefits.

Experimental setup of the presented case demonstrates our standard TCR repertoire sequencing workflow, expected quality of the generated data, and accuracy of our data analysis and interpretation.

Take home messages:

- Parameters to be considered as an internal quality check such as reads per UMI or CDR3 length are reproducible and highly similar between both biological and technical replicates.
- The number of both total and unique reads is comparable between replicates.
- Using the Jaccard Similarity Coefficient, a high similarity is confirmed among all technical replicates.
- Obtained diversity and richness is highly similar within replicates of interest.
- Same top clones are detected in all the samples and the number of reads per every clone is comparable between technical replicates.

Cerba Research's end-to-end repertoire sequencing service provides reproducible high quality TCR profiling data. The RNA input range and spike-in controls for an optimal recovery of the unique clones, diversity and richness is under investigation and will be finalized soon.

If you want to know more about our BCR repertoire profiling services, check out the whitepaper: **Decoding a B cell fingerprint**

Appendix

The Takara Human TCR Profiling kit utilizes SMART technology (Switching Mechanism at 5’ End of RNA Template) by combining NGS with a 5’ RACE system for rapid amplification (Figure 10).

In this approach, first strand cDNA synthesis starts from the 3’ RNA poly A tail which is dT-primed. During this step template switching reverse transcription reaction generates cDNAs with a sequence introduced by a template switching oligo attached to the 3' end of the cDNA (5' end of the transcript). Following this step, the first strand cDNA undergoes an enrichment process in which two rounds of gene-specific PCR amplifications are performed by using semi-nested PCR. The nested PCR reduces variability and allows for priming from the constant region of alpha and beta chains. This method generates highly sensitive and reproducible T-cell repertoire profiling and allow to capture complete V(D)J variable regions of TCR transcripts.

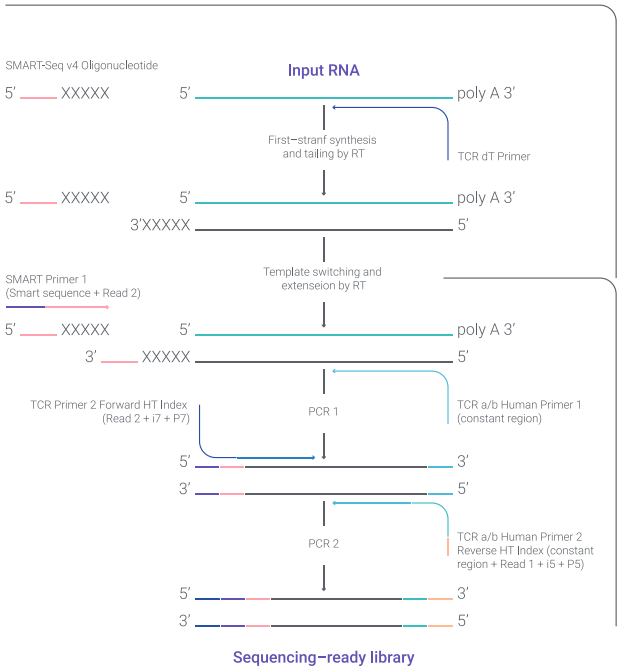


Figure 10: Schematic overview of the Takara SMARTer TCR 5’ RACE approach

Sample ID	CDR3 AA Seq	PBMC1-1	PBMC1-2	PBMC1-3	PBMC2-1	PBMC2-2	PBMC2-3	PBMC3-1	PBMC3-2	PBMC3-3
1	CASSRGDTQYF	693	864	399	919	1117	1112	1023	1148	1387
2	CASNPTSGGQFF	672	866	162	781	1011	727	394	975	1078
3	CASSLVGGFTDTQYF	460	596	262	570	787	753	584	733	780
4	CAAFYNTDKLIF	288	385	239	453	508	567	395	446	525
5	CATDTGRRALTF	250	322	186	427	481	488	398	502	511
6	CASSLGQYNEQFF	228	236	128	204	301	313	237	272	352
7	CASSTSRLAGETQYP	219	276	100	274	371	370	281	353	361
8	CAMRESEGGGGNKLTF	191	311	155	439	519	517	440	475	509
9	CAASTGTGNQFYF	165	251	167	252	308	344	288	309	327
10	CAYSVVF	154	167	74	200	235	243	181	176	222
11	CSAYLRDGDQETQYF	138	225	99	225	306	248	157	231	257
12	CASSFTGQGGYGYTF	133	128	55	106	141	164	138	158	209
13	CAESTDKLIF	127	176	105	169	205	226	181	165	204
14	CAVEWGKLIF	116	160	83	171	211	191	136	145	150
15	CASSFNLPGQYSYEQYF	105	116	55	112	182	151	130	173	208
16	CASSLEGTEAFF	100	136	41	82	120	104	97	122	159
17	CAAPCSGGYQKVTF	97	154	97	165	229	225	235	241	253
18	CAIRPDREGRNTIYF	94	113	33	86	122	114	129	122	176
19	CLVGGRGFKTIF	93	126	59	101	126	114	134	118	139
20	CASSYVRELFF	91	127	58	113	161	152	136	141	234

Table 1: Clonal distribution & the number of reads per clone.

Top 10 clones re stratified based on sample PBMC-5--1*
NA – Particular clone in respective dilution was not in its own top ten*e.

Cerba Research

About Cerba Research

Cerba Research is a leading specialty laboratory services provider with the capacity and breadth of a global central laboratory network. Our highly qualified scientists provide insight on the latest biomarkers, assays, and testing approaches and develop innovative solutions for unique challenges across all research phases for pharmaceutical, biotechnology, medical device, government, public health, and CRO organizations.

Cerba Research's extensive capability in laboratory testing and global logistics, including Bioanalysis, Flow Cytometry, HistoCytopathology, and Next-Generation Sequencing, enables us to drive operational agility at scale in a wide range of therapeutic areas, with recognized expertise in Virology, Immunology, Oncology, and Cell & Gene Therapy.

Cerba Research is part of the Cerba HealthCare Group with 15,000 employees on five continents, driven to advance diagnosis and health.

Interested in how we can shape your laboratory strategy for better outcomes?

Get in touch today and see how Cerba Research will support your next program.

cerbaresearch.com



References

1. Nikolich-Žugich, J., Slifka, M. & Messaoudi, I. The many important facets of T-cell repertoire diversity. Nat Rev Immunol 4, 123–132 (2004). doi.org/10.1038/nri1292-8
2. Pai, J. A., & Satpathy, A. T. (2021). High-throughput and single-cell T-cell receptor sequencing technologies. Nature Methods. doi:10.1038/s41592-021-01201-8
3. E. P. Rock , P. R. Sibbald, M. M. Davis, Y. H. Chien. CDR3 length in antigen-specific immune receptors. J. Exp. Med. 1994 Jan 1;179(1):323-8. doi: 10.1084/jem.179.1.323.

© Cerba Research 2026

2026-05-29_WP_DTCFC_V1