Research &
Pharma Solutions

T-Cell Receptor Sequencing

Tech Note

T-Cell Receptor Sequencing

T-cells play a crucial role in the control of immune-related diseases as main players of the adaptive immune system. The adaptive immune system is equipped with a powerful tool to protect us against pathogenic organisms and cancer by generating a highly diverse T-cell repertoire. This diversity is achieved by a broad range of unique receptors presented on the surface of T-cells. These highly diverse heterodimers are mostly composed of two subunits, the alpha and beta chains, and to a low percentage of gamma and delta chains. The T-cell receptor (TCR) recognizes small peptides presented by the major histocompatibility complex (MHC) molecules on antigen-presenting cells.

The diversity of the TCR complex is produced by the recombination process of V and J gene segments in the alpha chain, and an additional D gene in the beta chain. The resulting complementary-determining region 3 (CDR3), which is extremely diverse, is an attractive target to assess the overall TCR repertoire diversity as it is unique to each TCR beta variant.

In this technote, we discuss:

- T-cell receptor sequencing and its potential use in a variety of applications.
- Challenges of T-cell receptor sequencing, such as different sample types or input material, and how to overcome them.
- A use case of CeGaT´s T-cell receptor sequencing, highlighting the analysis of changes in the TCR repertoire over time and under different conditions.

What is T-cell receptor sequencing?

T-cell receptor sequencing describes the TCR repertoire profiling via high-throughput sequencing. It has become an indispensable tool for understanding the complex T-cell receptor repertoire. The main challenge in studying the TCR repertoire is its diversity. Clonotypes are usually defined as unique combination of the V gene, CDR3 amino acid sequence, and J gene. The clonotypes build the unique immune repertoire of an individual. In humans, the diversity is estimated to comprise around 1013 different clonotypes. Studies showed that TCR development is not random, and that there are common TCRs, shared in the general population. However, most of the TCRs are rare. This is one of the reasons why accurate techniques are necessary to investigate individual TCR repertoires. Sequencing-based techniques, such as bulk and single cell TCR sequencing are particularly attractive, since, in contrast to other methods like mass spectrometry, they are not restricted by a limited number of selection markers.

Applications

The high resolution TCR sequencing is applicable for different questions involving the immune system. Monitoring a specific adaptive immune response to pathogens or vaccines to analyze the induction of a persistent T-cell memory is one important application of TCR sequencing.

Areas of application for TCR sequencing are (Calis and Rosenberg 2014, Gohil et al. 2021):

- Immune system research
- Cancer immunotherapy and vaccine development
- Cancer diagnostic, treatment, and minimal residual diseases monitoring
- Personalized medicine

This technote presents an NGS-based workflow for TCR repertoire analysis. Furthermore, we aim at providing an insight into different options for subsequent data analysis.

Input material and its challenges

A challenging parameter in TCR sequencing is constituted by the original sample material. Different aspects should be considered when planning a TCR sequencing experiment: Analyzing blood or peripheral blood nuclear cells (PBMCs) in contrast to tissue delivers data that is not fully comparable between the different sample types because of their physical differences. In addition, it is possible to use genomic DNA as well as RNA to determine clonotypes in these samples. Especially in RNA analysis, a good quality of the starting material is essential for receiving reliable results. Further, the input amount of DNA or RNA into the library preparation of TCR sequencing libraries is not neglectable. Using deviating amounts of input material for different samples of one project or study, could result in a biased TCR repertoire result. The choice of starting material for TCR sequencing can determine the options in T-cell receptor analysis and is linked to certain advantages, but also limitations. As shown in table 1 (next site), genomic DNA has a higher stability compared to RNA and is thus, often a preferred sample material in the clinical context. An important aspect of using DNA as TCR template is that it allows for a quantitative analysis of the TCR repertoire, since the presence of one single template can be assigned to one cell. In contrast, it can be assumed that one cell contains multiple TCR transcripts, which then results in a higher probability of identifying rare receptor sequences when analyzing RNA. The higher sensitivity is thus one of the major advantages of performing RNA-based TCR sequencing.

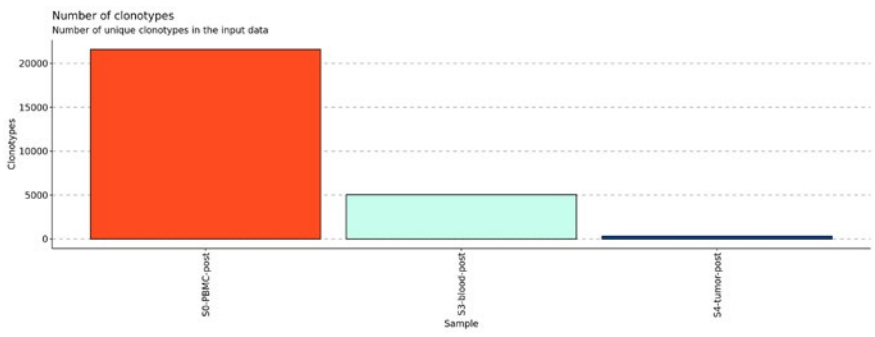

Figure 1: *Comparison of sample types. Depending on the sample type, the number of clonotypes differs. In this comparison, PBMC, whole blood, and FFPE tumor samples after treatment („post") are compared. The PBMC sample shows the most clonotypes followed by the blood sample. In the tumor sample the number of clonotypes is the lowest.*

	DNA	RNA
Stability	Relatively stable	Less stable, quick degradation
Quantification	Number of input fragments corresponds to number of T-cells in sample	Number of TCR templates does not correspond to number of T-cells → differences in expression levels → might confound quantification
Number of clonotypes	Less clones with same amount of input	More clones with same amount of input
Clonality	Higher	Lower
Diversity	Lower; Only small amount of input DNA relevant for PCR amplification → difficult to capture full diversity	Higher; Proportion of TCR template molecules higher → diversity is better captured (less „genomic background)
Sequence	Introns	No introns

Table 1: *DNA versus RNA as target for TCR sequencing*

Different tissue types or blood-based input materials determine the TCR analysis regarding the output of extracted DNA or RNA. There is a huge difference between the compositions of cells in the samples. This difference affects the diversity, and thus the number of clonotypes, of the repertoire. The difference in clonotypes is visualized in figure 1: Using PBMCs as sample type results in the highest number of clonotypes, followed by whole blood and formalin-fixed, paraffin-embedded (FFPE) tissue of the tumor. At CeGaT, we offer TCR analysis of different starting materials, including high-quality DNA and RNA, as well as degraded DNA or RNA from FFPE material.

Sample types include:

- Fresh frozen tissue
- FFPE tissue
- Whole blood
- PBMCs
- Sorted cells
- Isolated DNA or RNA

Case Study

In this Technote, we discuss a common application of TCR sequencing: Monitoring changes in the TCR repertoire during treatment of a tumor patient. The characterization of the T-cell repertoire in the tumor and the changes in tumor patients' blood can broaden the understanding of the immune system's response to tumorigenesis and therapy.

In this case study, we analyzed whole blood and FFPE tumor tissue, both before and after treatment with individualized peptide-based cancer vaccines. TCR sequencing was performed using DNA to monitor quantitative changes in the TCR repertoire.

This approach allows answering frequent questions, such as:

- How many different TCR clones are present in the tumor and the blood?
- Are there changes in the TCR repertoire before and after the treatment?
- How do selected TCRs develop over time and between samples?
- What are the relative abundances of clones in the different samples

Method

DNA was extracted from whole blood and FFPE sections of the tumor. After quality control of the isolated DNA, sequencing libraries were generated using our adjusted and validated protocol, compatible with Illumina. The samples were sequenced on one of our NovaSeq 6000 systems with a read-length of 2x100 bp and a sequencing depth of 2 M paired reads.

The steps of the computational analysis for the reconstruction of the TCR clone set are summarized in figure 2. The FASTQ files are downsampled to 2 million read pairs before further bioinformatic processing to ensure comparability between samples. Overlapping paired reads are merged into single reads before reconstruction of the CDR3β regions. Only read pairs that could be successfully merged into a single read are kept for the analysis. The reconstruction of T-cell receptor sequences returns a set of functional clones on one side, and a second set of clones that is discarded as the clones are considered non-functional due to alternative reading frames or premature stop codons. The non-functional clones are not used in further analysis of the reconstructed TCR repertoires. Unexpectedly long CDR3β sequences (>29 amino acids) are removed from the functional clone set, because they are considered assembly artefacts. Further analyses of the TCR repertoire are performed using the R package immunarch (ImmunoMind Team 2019).

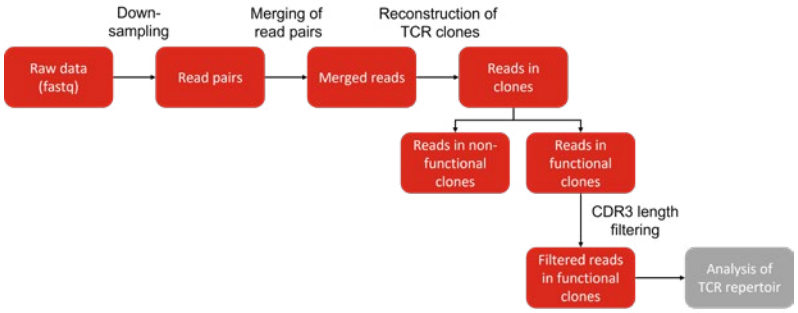

Figure 2: *Steps involved in the computational analysis of the raw sequencing reads to the final clone sets.*

Results

Number of clonotypes

The number of unique clonotypes provides information about the diversity of a TCR repertoire. The clonotype count could be a first parameter to compare changes in the TCR repertoire between different samples or time points. Within a study, we recommend bringing in the same amount of starting DNA or RNA into library preparation to allow a meaningful comparison. To demonstrate differences in sample types, we analyzed blood and tumor samples from a cancer patient before and after an individualized peptide-based tumor vaccine treatment. In figure 3, an obvious difference in clonotype counts between blood and tumor samples can be observed. This originates, among other things, from the higher proportion of T-cells in the blood compared to the tissue.

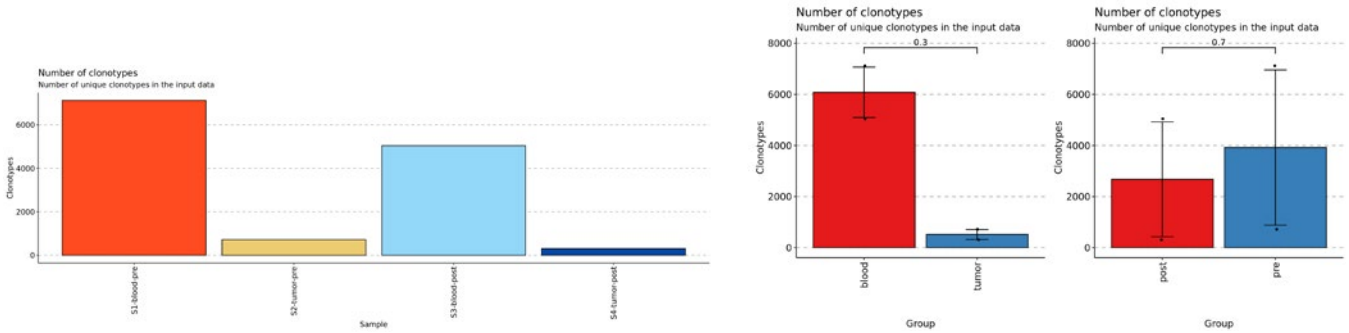

Figure 3: *Number of unique clonotypes over different sample types and time points: Blood samples show a higher number of clonotypes in contrast to tumor samples. We can additionally perform several group comparisons based on provided metadata. For every comparison, statistical tests can be calculated. The lower left graph shows the number of clonotypes grouped by sample type. The lower right graph shows blood samples, grouped before („pre") and after („post") treatment.*

Relative abundance

Seeing the relative abundance as a parameter for the measure of clonality within the clonotype repertoire, figure 4 shows the frequency distribution between the blood and tumor samples before and after treatment. The data indicates that the TCR repertoire in the blood samples contains more different clonotypes (small clonotype group). In contrast, the tumor samples have a larger proportion of the large clonotype group. After treatment, the tumor sample reveals an even larger proportion of hyperexpanded clonotypes.

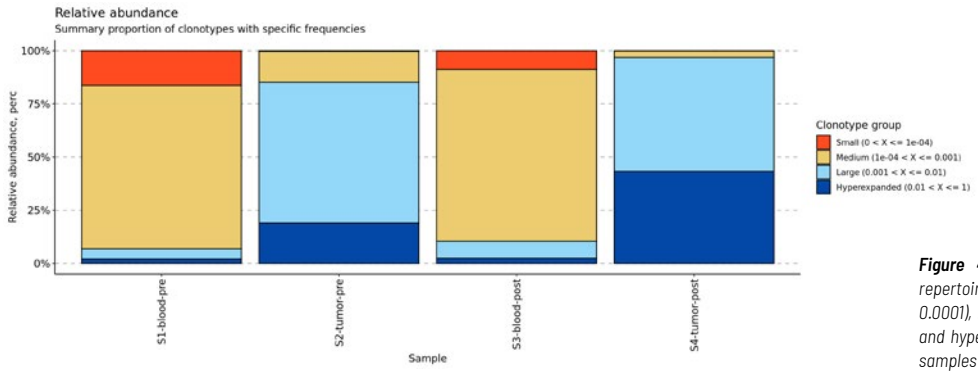

Figure 4: *Relative abundance of clonotype groups. TCR repertoire grouped by their proportion of small ($0 < X \leq 0.0001$), medium ($0.0001 < X \leq 0.001$), large ($0.00\ 1 < X \leq 0.01$), and hyperexpanded ($0.01 < X \leq 1$) clonotype groups. The blood samples show a larger proportion of the small clonotype group. The proportion of the hyperexpanded clonotype group increased after the treatment.*

Clonotype tracking

The shift in clonotypes is even more apparent for the clonotypes in the tumor: The 30 most frequent clonotypes in the tumor before the treatment are eradicated (figure 5A). However, new clonotypes are formed (figure 5B).

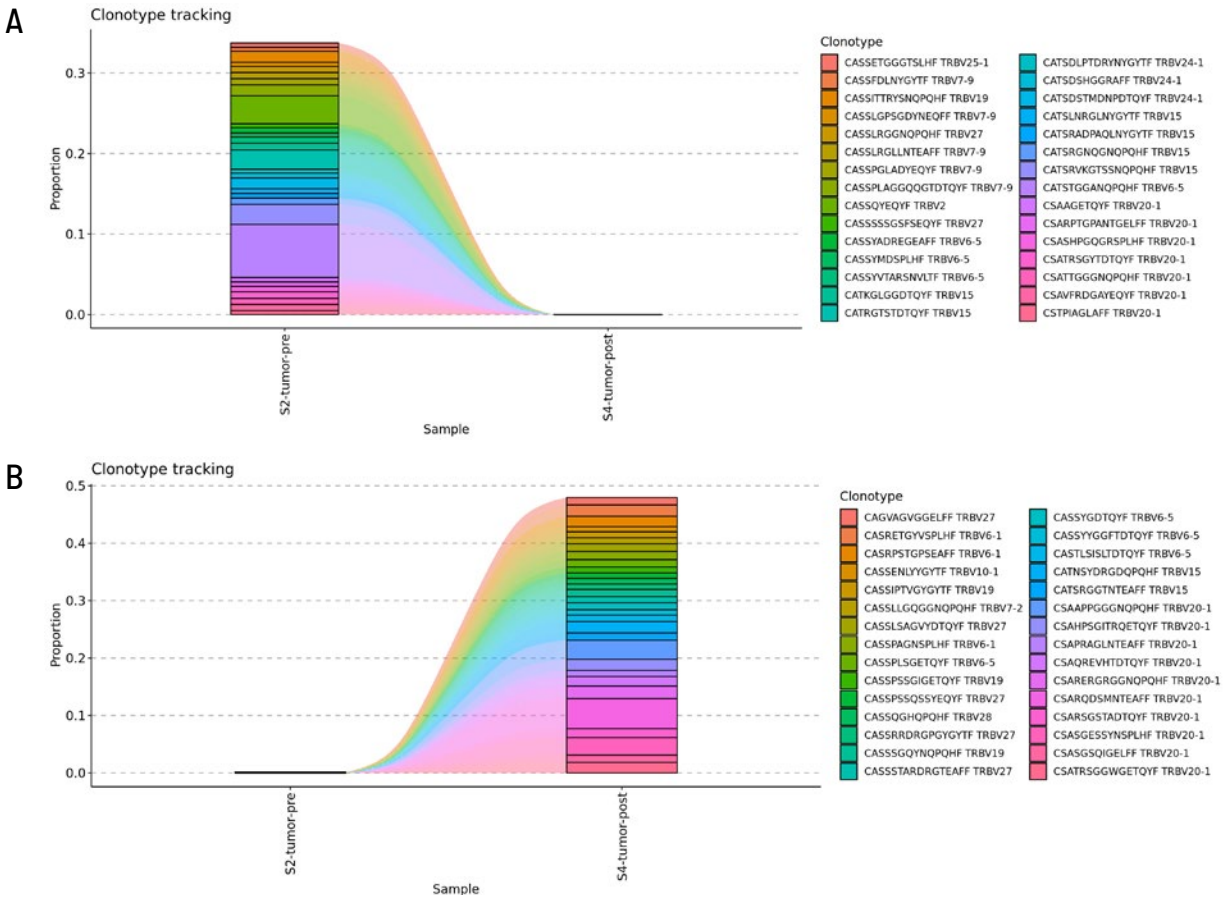

***Figure 5**: Changes in 30 clonotypes in patients tumor samples before and after treatment. (A) Top 30 clonotypes in tumor sample before treatment. (B) Top 30 clonotypes in tumor sample after treatment.*

To track the new clones before and after treatment in all samples, table 2 comprises the number of certain clones in all samples. Such a table can also be created for only a subset of samples, e.g., for all tumor samples.

CDR3.aa	V.name	S1-blood-pre	S2-tumor-pre	S3-blood-post	S4-tumor-post
CASSLEVLVLRQGASGNTIYF	TRBV11-3	189	218	56	720
CASSLSAGVYDTQYF	TRBV27	13480	2071	16273	17880
CASSVPAGGWQTSGTDTQYF	TRBV9	1639	392	1822	880
CAWSRTGDDFYNSPLHF	TRBV30	357	1685	322	1908
CASGLKDISYEQYF	TRBV12-5	2722	NA	1097	941
CASQLGRGGPDTQYF	TRBV4-3	639	NA	5648	1934
CASSAATSRRDTQYF	TRBV10-2	382	706	863	NA
CASSALWTQETQYF	TRBV6-1	16228	2194	21007	NA
CASSDSTSGYNEQFF	TRBV6-4	318	NA	933	4640
CASSLLLAYEQYF	TRBV28	1081	1222	1234	NA
CASSLSRNLNNEQFF	TRBV28	3409	2516	2604	NA
CASSLSTKGETQYF	TRBV7-9	763	7212	1673	NA
CASSPGLYNEQFF	TRBV14	61	340	56	NA
CASSPPLGTGYEQFF	TRBV7-2	NA	1229	176	542
CASSPPTSPFFEQYF	TRBV7-2	NA	2027	490	4830

***Table 2:** Clonotype occurrence in all samples. The CDR3 amino acid sequence and the name of the V gene is provided, followed by the absolute number of clonotypes in the respective sample.*

Interesting clonotypes can be tracked over all samples, as seen in figure 6. By this visualization, the shift in clonotype occurrence can be tracked and reconstructed. The tracking can be applied to any clonotype and sample of interest.

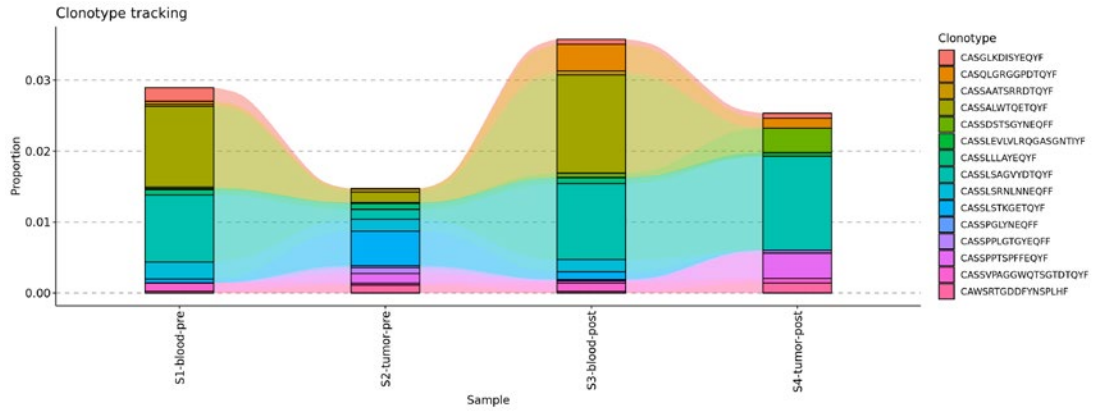

Figure 6: *Clonotypes tracking. Clonotypes of interest are tracked over all samples.*

Repertoire overlap

Besides tracking clonotypes over different samples, we can calculate the overlap of the clonotype repertoires. This information can give valuable insight into repertoire similarity for example between different tumor entities or between responders and non-responders. Figure 7 indicates a similarity between the blood samples, as expected.

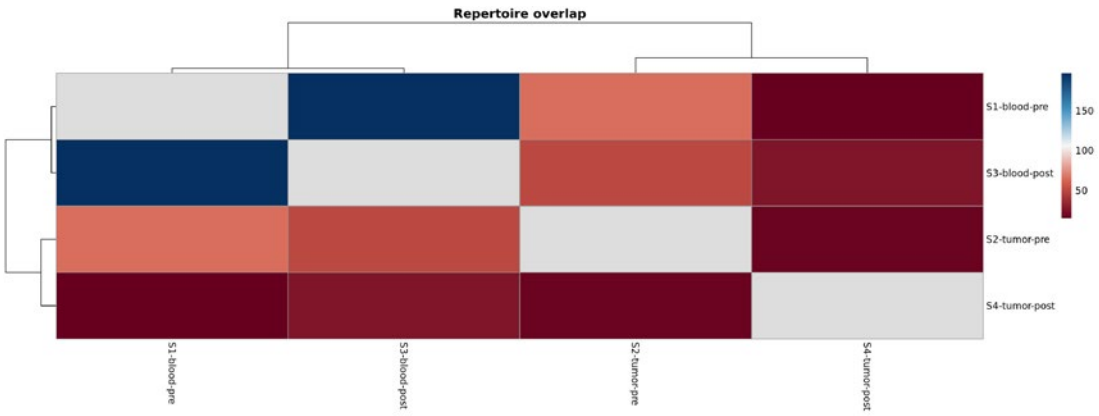

Figure 7: *Repertoire overlap between blood and tumor samples over different time points.*

Diversity and clonality

Diversity and clonality are measures to describe the composition and frequency distribution of a TCR repertoire. We use the Shannon Entropy as a measure of diversity of a TCR repertoire, i.e., how diverse are the sequences contained in the repertoire. The Shannon Entropy is defined as: $H = -\Sigma piNi = 1log2(pi)$, where pi is the proportional abundance of clone i. Simpson Clonality is a single measure for the frequency distribution of the clones within the TCR repertoire and is defined as: $Simpson\ Clonality = \sqrt{\Sigma pi2}$, where pi is the proportional abundance of clone i. The higher the value, the higher the clonal expansion of the TCR repertoire. This means that a few frequent clones are occupying a large fraction of the repertoire.

Conclusion

Our TCR sequencing protocol allows for the analysis of various starting materials including whole blood, but also more challenging material like FFPE tissue. In our use case study, we demonstrate examples of how TCR sequencing data can be processed and presented. The graphical representation can be adapted, depending on the underlying clinical or research question. In this technote, we showed the possible analyses to gain more insights into the TCR repertoire and its changes, without attempting clinical interpretation of the findings. Nonetheless, achieving meaningful clinical interpretation is particularly important in T-cell receptor sequencing. Our aim is to provide you with the best support and recommendations for your TCR sequencing study.

References

Calis, Jorg J.; & Rosenberg, Brad R. (2014). Characterizing immune repertoires by high throughput sequencing: strategies and applications. In Trends in immunology, 35(12), 581–590. https://doi.org/10.1016/j.it.2014.09.004.

Gohil, Satyen H.; Iorgulescu, J. Bryan; Braun, David A.; Keskin, Derin B.; & Livak, Kenneth J. (2021). Applying high-dimensional single-cell technologies to the analysis of cancer immunotherapy. In Nature reviews. Clinical oncology, 18(4), 244–256. https://doi.org/10.1038/s41571-020-00449-x.

ImmunoMind Team. (2019). immunarch: An R Package for Painless Bioinformatics Analysis of T-Cell and B-Cell Immune Repertoires. Zenodo. http://doi.org/10.5281/zenodo.3367200-

Rosati, Elisa; Dowds, C Marie; Liaskou, Evaggelia; Kemsdal Hendriksen, Eva Kristine; Karlsen, Tom H; Franke, Andre (2017): Overview of methodologies for T-cell receptor repertoire analysis. In BMC Biotechnology 17, p 61.

Wickham, Hadley (2009): ggplot2. Elegant graphics for data analysis. New York: Springer.

About us

CeGaT is a global provider of genetic analyses for a wide range of medical, research, and pharmaceutical applications. The company combines its next-generation sequencing (NGS) process and analysis pipelines with its medical expertise – dedicated to identifying the genetic cause of disease and supporting patient management.

Genetic mutations can trigger a wide range of diseases, from epilepsy to Parkinson's. Through the use of NGS, it is possible to analyze all genes associated with a disease phenotype simultaneously – both fast and effectively. An interdisciplinary team of scientists and physicians evaluates the data and summarizes the findings in a comprehensive medical report. All services are performed in-house.

CeGaT, founded in 2009 and based in Tübingen, Germany, is accredited according to CAP, CLIA and DIN EN ISO/IEC 17025:2018.

CeGaT GmbH
Research & Pharma Solutions
Paul-Ehrlich-Str. 23
72076 Tübingen
Germany

Phone: +49 7071 56544-333
Fax: +49 7071 56544-56
Email: rps@cegat.com
Web: www.cegat.com

DAkkS Deutsche Akkreditierungsstelle D-PL-13206-01-00

Accredited by DAkkS according to DIN EN ISO/IEC 17025:2018

CAP ACCREDITED COLLEGE of AMERICAN PATHOLOGISTS

CLIA CERTIFIED ID: 99D2130225

2022/09