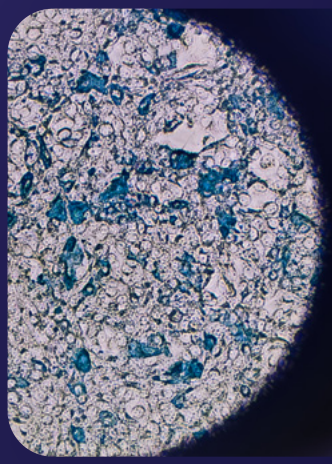
Whitepaper

Cerba Research

Decoding a B Cell Fingerprint

All you need to know about BCR sequencing

April 2023



The adaptive component of the immune system in humans is composed of a diverse population of B and T cells, which are a paramount for the effective immune response upon exposure to foreign antigens or mutated host ones (derived from pathogens, or mutated proteins expressed by cancer cells).

In the case of B cells, antigens are presented on the surface of an invading pathogen or mutated cell and are recognized by a B cell receptor (BCR). Like a fingerprint being unique for any given person in the world, every BCR is unique for every B cell clone. Collectively, a pool of these clones generates a great variety of BCRs and antibodies, called the B-cell repertoire.

The application of next generation sequencing (NGS) technologies enables us to profile the B-cell receptor repertoire with a high resolution. Profiling the BCR repertoire is typically done by assessing parameters such as clonality, diversity and richness of a large number of B cells by sequencing their genetic code. Obtained genetic information is used to predict biophysical and other properties of a BCR or an antibody. This greatly enhances our ability to harness the potential of the selected clones for therapeutic, diagnostic or research applications.

B-cell repertoire profiling holds great promise to boost our efforts in fields of antibody development, immunotherapeutics, vaccine development and disease monitoring. The ultra-deep high throughput sequencing we offer combined with state-of-the-art analysis developed by our partner is truly unparalleled.

The Authors

Goran Marinković
Strategic Research and Development Leader

Ilse Wolters
Research and Development Technician

Irene Chan
Research and Development Technician

Dorota Talady
Business Development Director

Nestor Vazquez
Scientific Solutions Consultant

Henk-Jan van den Ham
Team Lead, Genomics and Target Discovery

Wilco Knetsch
Director, Research and Development

Dirk van Alewijk
Business Development Director

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 B Cell Repertoire Generated?

B-cell receptors are essentially cell-surface-bound antibodies composed of two light and two heavy chains. Every B cell is expressing only one type of BCR and subsequently during the plasma cell stage, only one type of antibody. B-cell receptors are generated through the process of genetic rearrangement of the large immunoglobulin gene segment.

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

The gene loci of these receptors are organized as gene segment families comprising a variable (V) gene, a diversity (D) gene, a joining (J) gene and a constant (C) gene **(Figure 1).** V(D)J recombination presumes that rearrangement of these segments is followed by random nucleotide insertion and/or deletion at the junction points which creates even higher diversity of the B-cell repertoire.

Both the heavy and light chains are composed of two domains: a variable (V) domain and a constant (C) domain. Each variable domain consists of three complementarydetermining regions (CDRs). The CDR3 region is of specific interest as this region interacts with antigens and determines receptor specificity. The collection of all B-cells receptors genetically rearranged for different antigen specificity is known as the BCR repertoire.

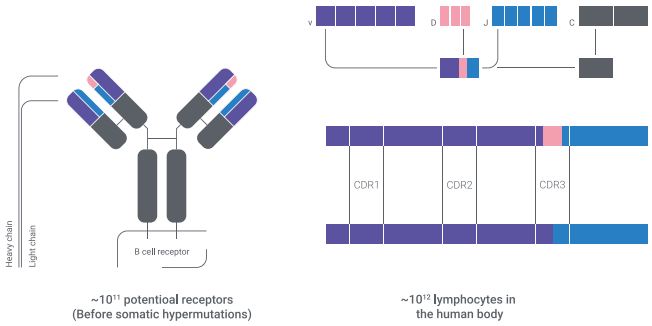


Figure 1: Gene rearrangement in B-cell receptor, heavy chain (light chain also depicted).

Cerba Research: One-Stop Shop

Cerba Research together with its partner Enpicom, provides tailored End to End solutions to its customers in decoding the immune repertoire (Figure 2).

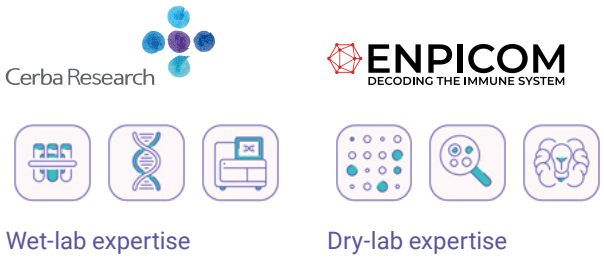


Figure 2: Schematic overview of the End-to-End workflow at Cerba Research

Cerba Research utilizes the Takara SMARTer TCR Profiling Kit that couples SMART (Switching Mechanism at 5' End of RNA Template) technology (Figure 9) 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.

BCR Light/Heavy Chain Analysis: An Example

Quality control is essential to obtain useful data – RNA quality.

Various factors may influence the quality of BCR profiling, for example the quality of samples being analyzed. The purpose of the example experiment outlined below is to determine what are the optimal sample amounts needed to obtain consistent, high-quality results.

Three separate isolations were performed on a human peripheral blood mononuclear cell (PBMC) sample after which aliquots were made for each isolation to mimic biological and technical replicates. To address the question of how much input material is needed for successful BCR profiling, the aliquots were made with different inputs of RNA ranging from 50 to 800 ng.

To assess quantity and ensure quality of the isolated material, each sample is examined using the Bioanalyzer. RNA samples with an RNA integrity number (RIN) of 7 or higher are considered of sufficient quality to proceed with unbiased amplification of the BCR whole repertoire.
All samples in the experiment met the quality criteria of RIN value of 7 or higher and were included in the BCR profiling assay **(Figure 3).**

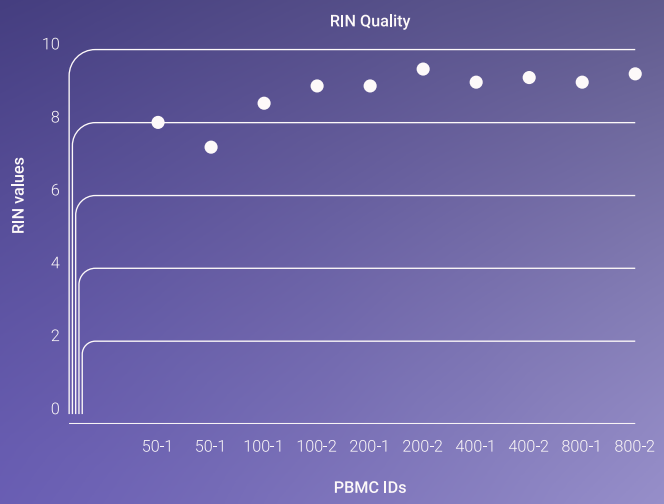


Figure 3: All samples in the experiment had higher than required quality of RNA, reflected in a RIN value higher than 7.

Quality control is essential to obtain useful data – Spike in controls to generate reproducible data

To ensure the quality of our workflow, control oligonucleotides (spike-in controls) are designed and spiked into the samples (Appendix, Figure 10). This provides a within-sample standard for between sample comparison and allows us to identify and correct for any amplification bias.

A spike-in control should never interfere with the sample itself because this will lead to a biased and unreliable results. Therefore, a minimal / maximal input concentration is determined. For each spike-in control concentration, the clone count fraction is determined (Figure 4). In the reactions spiked with 0.075 ng of spike-in control the fraction of reads used to sequence the spike-in control is 0.04, which is equal to 4% of the total reads. This amount is the upper limit in our runs.

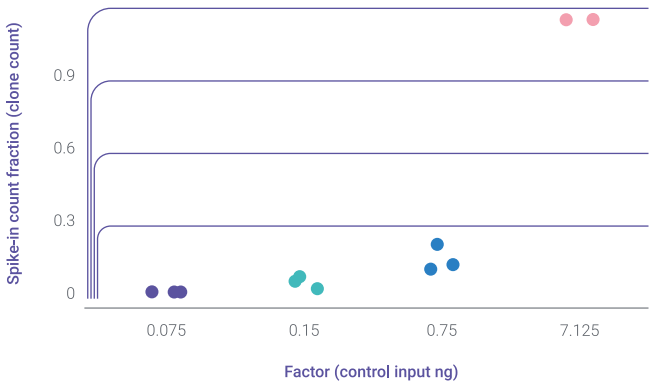
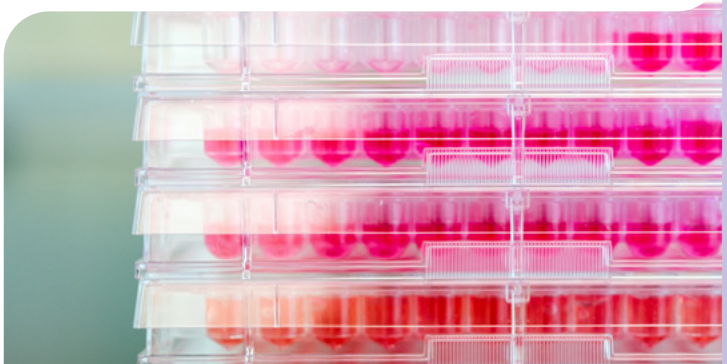


Figure 4: Recovery of spike-in reads. Three concentrations of spike-in control are spiked in 200 ng PBMC RNA. Three samples with 0.075 ng spike-in control (blue), three samples with 0.15 ng spike-in control (red), three samples with 0.75 ng spike-in control (green). Two reactions with maximum spike-in input, 7.125 ng, are the controls (purple). The three spike-in concentrations need respectively 4, 8, 15 and 100% of the total reads per sample.



BCR Profiling: The Analysis

The analysis of BCR repertoire sequencing data is not straightforward and requires different disciplines, including immunology and bioinformatics.

Cerba Research harbors the knowledge and expertise in all relevant areas to provide our customers with solid and understandable immune repertoire sequencing data. This can be either in a standard analysis format or tailored to address specific customer questions. Our standard data analysis pipeline consists of the following elements:

- 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 the standard reporting elements are described in more detail below using human PBMC samples as an example.

Number of Total and Unique Clones Recovered

For this experiment, an increasing amount of input RNA material was used to evaluate the range of nucleic acid input that provides accurate BCR profiling results. It is expected that with increasing nucleic acid input, an increasing number of BCR receptors (both total number of BCR receptors and unique BCR receptors) will be recovered. To evaluate reproducibility, technical replicates were included. With an increasing concentration of RNA input we observe a steady increase in the number of total clones detected.

An input of 400 ng resulted in the highest number of unique clones, although a significant part of unique receptors was already recovered using 50 ng of RNA input **(Figure 5).**

The number of recovered total and unique clones is found to be highly similar between technical replicates.

Conclusion:

As expected, the higher RNA input results in a higher number of receptors recovered. Using amounts between 200 ng and 400 ng of RNA input resulted in recovery of the highest number of unique receptors, therefore our recommendation is to use amounts in this range for successful BCR repertoire profiling. Finally, our results show that the workflow is highly reproducible in numbers of receptors recovered using technical replicates.

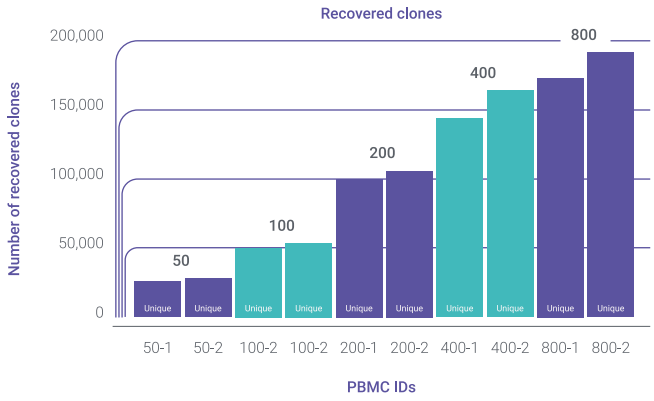


Figure 5: BCR receptors recovered using samples with increasing amounts of RNA (range between 50 ng/µL and 800 ng/µL). The full bar represents the absolute number of BCR receptors recovered, the part in bold represents the unique BCR receptors recovered from the data.

High Level of Similarity Between Technical Replicates

To evaluate the reproducibility of our BCR profiling workflow in more depth, we evaluated similarity between technical replicates. The similarity in BCR clones identified between technical replicates is addressed by calculating the Ruzicka Similarity Coefficient. This coefficient takes clone count into consideration when assessing similarity between samples. We observe high similarity between technical replicates, which is indicated in Figure 6. Similarity increases with an increasing amount of input material. As expected, 800 ng of input RNA represents the highest similarity (upper right corner), whereas > 200 ng of input RNA acceptable similarity is observed.

Conclusion:

Higher concentration of RNA input results in higher similarity between replicates. Based on similarity it is recommended to have at least 200 ng RNA as input.

Figure 6: Sample similarity plot: Technical replicates are plotted against each other and Ruzicka similarity coefficient was calculated. Similarity increases with an increased sample input amount (darker shades of purple).

Diversity & Richness

To assess diversity and richness as a function of RNA input amount, we plotted the number of unique clones recovered. We observe that recovered diversity and richness steadily rose with an increase of the input RNA up to 200 ng. This was accompanied by high reproducibility between technical replicates **(Figure 7).**

Conclusion:

The best recovery for both diversity and richness was obtained using an input amount of 200 ng of RNA. Considering other analyzed parameters, our recommendation is to use sample amount varying between 200 and 400 ng RNA. 800 ng of RNA input material results in a reduction of diversity and richness.

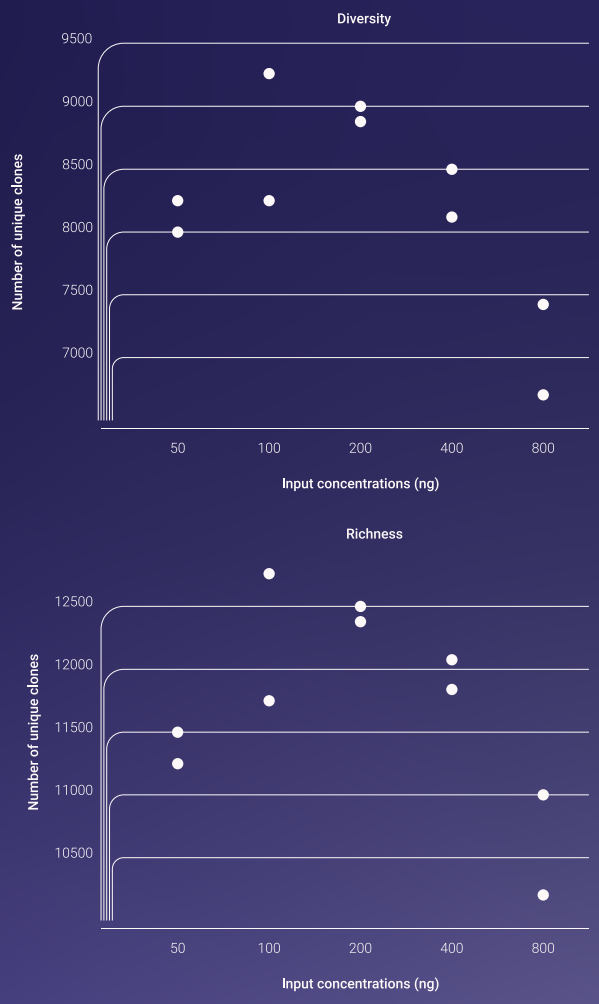


Figure 7: Diversity and richness as a function of the amount of the input RNA. Solid and reproducible recovery was observed using 200 and 400 ng of RNA.

Top-10 B Cell Clones by CDR3 Amino Acid Sequence

Analysis of the top clones was 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 in respect to the most abundant clones.

The top 10 clones according to their CDR3 domain (amino acid sequence) are represented in Figure 8.

Conclusion:

All five dilutions have a high overlap among the most abundant clones detected. For every top 10 clone, the number of reads recovered is highly similar between technical replicates (Appendix, Table 1). As expected, the number of reads per clone increases with the increasing amount of the input material.

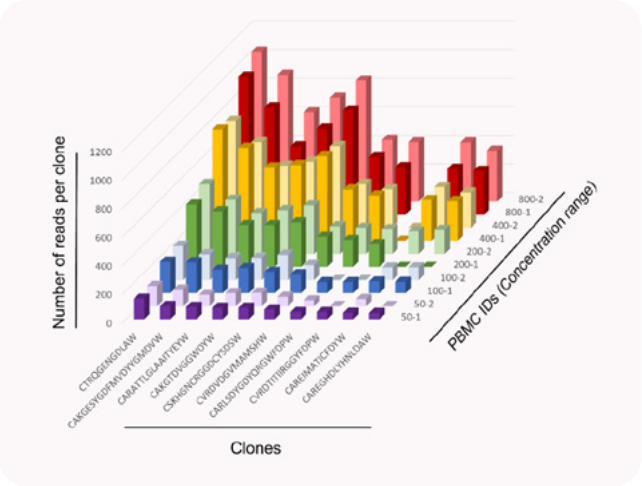


Figure 8: Top 10 clones. Top clones are recovered from all five different input RNA amounts. Number of reads per clone was closely correlated with the RNA input amount.

Conclusion

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

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

Take home messages:

- More RNA input results in a higher number of receptors recovered. 200 and 400 ng of RNA input yielded the highest recovery of unique receptors.
- Our results show that the BCR profiling workflow is highly reproducible in numbers of receptors recovered using technical replicates.
- Higher concentration of RNA input results in higher similarity between replicates. Based on similarity, it is recommended to have at least 200 ng of RNA as input. This also depends on the underlying scientific question and other parameters of interest.
- Diversity and richness are both highly similar between replicates. Best recovery is obtained with 200 ng RNA input. Our recommendation is to use between 200 to 400 ng of RNA for the optimal recovery of these parameters
- Finally, we examined clonal distribution for the most abundant clones according to amino acid sequence of the CDR3 domain. Same top clones are detected in all the samples, regardless of the amount of input material. Increasing the amount of the RNA material results in the higher number of reads per clone.
- In conclusion, our Cerba Research repertoire sequencing service provides reproducible high quality BCR profiling data. Optimal recovery of the unique clones, diversity and richness is obtained when using an RNA input ranging between 200 and 400 ng. Spike-in controls we use, complete our end-to-end solution. A one-stop shop for Concentrations (ng) of input material your immune profiling request.

Appendix

Technology in use

Takara Human BCR 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 9)**.

In this approach, first strand cDNA synthesis starts from the 3' RNA poly A tail and it 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, 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 heavy or light chains. This method generates highly sensitive and reproducible B cell repertoire profiling and allows to capture complete V(D)J variable regions of BCR transcripts.

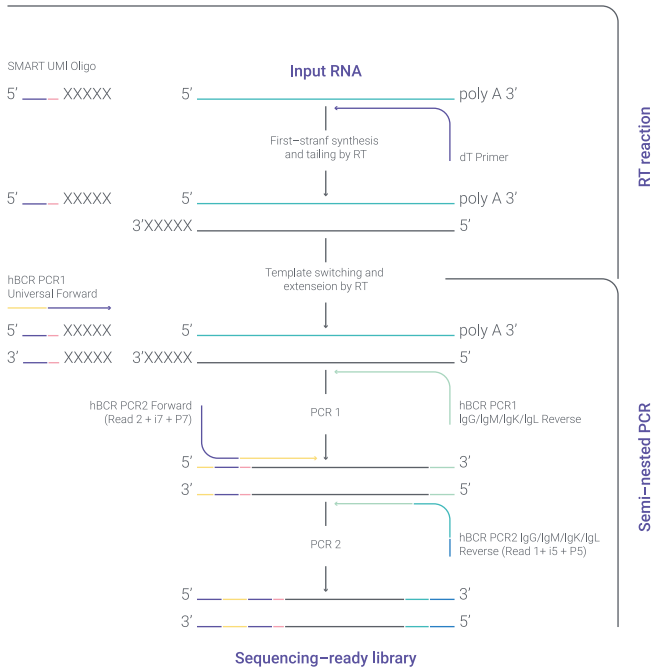
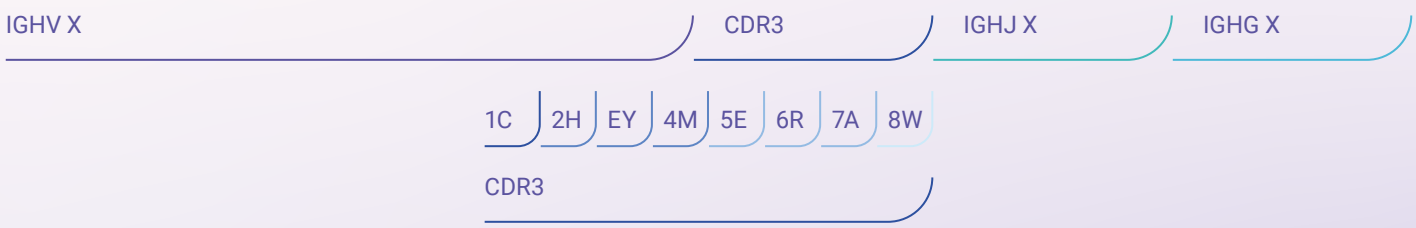
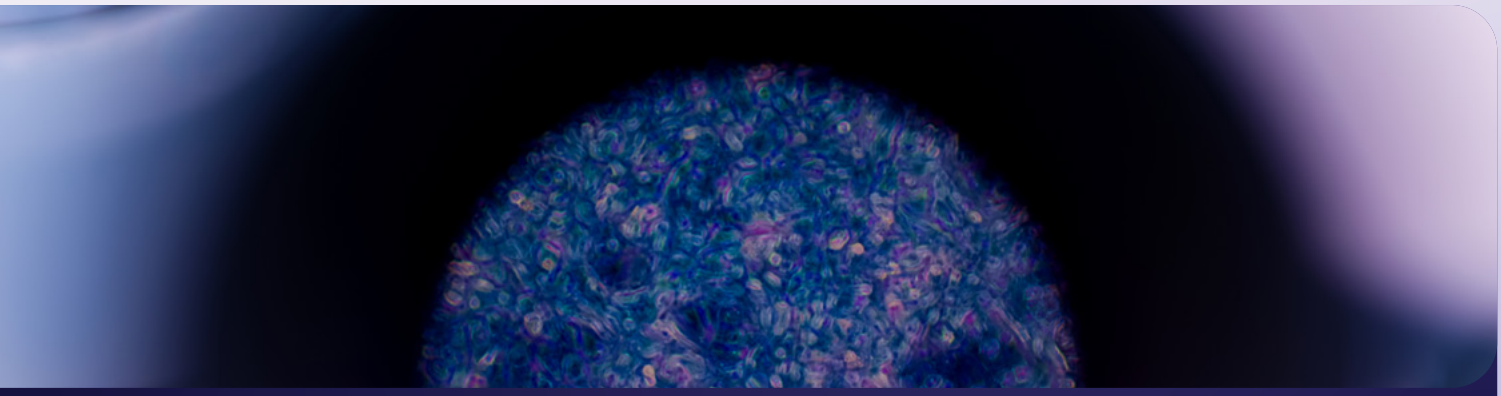


Figure 9: Schematic overview of the Takara SMARTer BCR 5' RACE approach.

Spike-In Control Design

Figure 10: Schematic overview of the IVTX1, V-gene/CDR3 (incl. AA sequence)/J-gene/C-gene





Clonal Distribution and The Number of Reads Per Clone

Table 1: Clonal distribution and 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*

Sample ID	CDR3 AA Seq	PBMC-50-1	PBMC-50-2	PBMC-100-1	PBMC-100-2	PBMC-200-1	PBMC-200-2	PBMC-400-1	PBMC-400-2	PBMC-800-1	PBMC-800-2
1	CTRQGENGDLAW	144	141	220	239	239	239	780	749	749	1048
2	CAKGESYGDFMVDYYGMDVW	92	116	216	190	190	190	651	599	599	889
3	CARATTLGLAAITYEYW	86	81	163	153	153	153	514	435	435	625
4	CAKGTDVGGWDYW	80	92	175	155	155	155	531	463	463	728
5	CSKHGNCRGGDCYSDSW	80	94	148	176	176	176	594	571	571	848
6	CVRDVDGVMAMSHW	69	66	132	105	105	105	359	306	306	433
7	CARLSDYGDYQRGWFDPW	51	41	77	NA	NA	NA	317	269	269	415
8	CVRDTITIIRGGYFDPW	51	NA	78	NA	NA	NA	NA	NA	NA	NA
9	CAREIMATICFDYW	45	50	82	88	88	88	286	286	286	414
10	CAREGHDLYHNLDAW	44	NA	79	89	89	89	277	247	247	357

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.[2]

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

© Cerba Research 2026

2026-06-02_WP_DABCF_V1