From sample to discovery: Applications, challenges, and the path ahead in full-service repertoire sequencing

3 October 2023

Cerba Research
Transforming research, advancing health together.

ENPICOM
DECODING THE IMMUNE SYSTEM

Interview with Goran Marinkovic, Dirk van Alewijk, Ilse Wolters, and Henk-Jan van den Ham

The immune system consists of a large variety of cells that fight off diseases and maintain body homeostasis. T and B lymphocytes are educated to recognize specific threats to the body through their B- and T-cell receptors (BCR and TCR). This education system generates a huge variety of T and B cell receptors that are collectively called the immune repertoire. Using a technique called 'immune repertoire sequencing' or Rep-Seq, the immune repertoire can be characterized using high-throughput sequencing. Since B and T lymphocytes also form memory cells in response to threats, the immune repertoire reflects both current and past immune responses; this is why Rep-Seq is also known as 'immune fingerprinting'.

Rep-Seq has proven to be a valuable tool in many research areas, including basic research, pre-clinical and clinical studies, diagnostics, and drug discovery. By integrating sequencing data with experimental data, researchers can gain new insights into fields such as oncology, infectious disease, vaccine development, and antibody discovery.

Today, together with the experts from Cerba Research and ENPICOM, we discuss the challenges of analyzing repertoire data, the importance of quality control, promising Rep-Seq applications, and future developments.

## Can you explain the concept of immune repertoire sequencing and its significance in immunology research?

Goran Marinkovic: Immune repertoire profiling is a powerful technique that relies heavily on a technology known as next-generation sequencing (NGS). Its primary focus is on the adaptive component of the immune system, specifically lymphocytes. These cells protect us from foreign pathogens and mutated cancer cells. To do so, B and T lymphocytes have to be able to recognize and interact with millions of foreign-origin molecules, also known as antigens. This recognition occurs through their T and B cell receptors, through which every lymphocyte and its progeny are specialized to recognize only one type of antigen. Collectively, these lymphocytes have a tremendously large repertoire of receptors that are created in a process called VDJ-recombination. Because of this, it is evident that insights into immune repertoire profiles hold immense value not simply in addressing basic immunological research questions but also when conducting pre-clinical and clinical immunomodulatory studies, and in diagnostic applications.
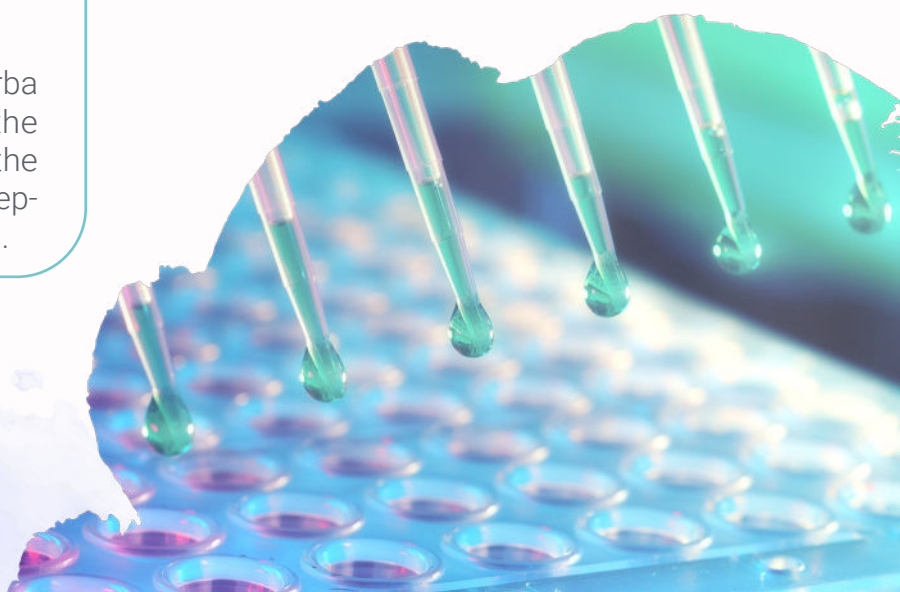
What techniques or methods do you utilize for immune repertoire sequencing? Can you describe the workflow involved in this process?

Ilse Wolters: The process of immune repertoire sequencing involves a series of individual techniques, ultimately resulting in successful NGS data analysis. The first step is RNA isolation, followed by a quality control check. Once high-quality RNA is confirmed, sample libraries are prepped which involves numerous smaller steps, such as processing RNA to cDNA and preparing DNA libraries for sequencing. These libraries are also subjected to a quality control check before being sequenced by a NGS system. The data files are also subjected to a quality control check to ensure high-quality sequencing data output. Finally, the last step is the analysis of the FastQ files.

Sample
Blood or tumor tissue

Isolate
mRNA

Amplify
Variable regions of receptors

Sequence
To determine receptor sequence

Process
Repertoire sequencing data

Analyze
Clonotypes

Discover
Insights to guide drug discovery and R&D

How do you ensure the accuracy and reliability of immune repertoire sequencing data? What quality control measures do you employ?

Ilse Wolters: Together with our partners at ENPICOM, we've developed spike-in controls for both BCR and TCR. These controls consist of synthetic RNA fragments with known sequences, allowing us to incorporate them into a sample, ensuring the high quality of a run. The controls help us assess not only the RNA quality but also detect PCR and NGS errors. We've meticulously determined the optimal concentration of these spike-ins for repertoire sequencing runs, ensuring the most efficient performance.

Additionally, we've implemented various quality control steps throughout our workflow. For RNA isolation, we use the Bioanalyzer and RiboGreen techniques to assess both the quality and quantity of the RNA. Once we've prepared the sample libraries through all the necessary steps, we utilize gel electrophoresis to gauge the sample's length and PicoGreen assays to determine its quantity.

For NGS on our part, we meticulously monitor quality by evaluating QC 30 scores, read lengths, and the total number of reads generated during the sequencing run.

Having multiple quality checkpoints between each step is essential to guarantee the highest quality data possible.

What are the biggest challenges researchers encounter while working with immune repertoire sequencing? How can they overcome them?

Dirk van Alewijk: This is a very intriguing question, as researchers are facing all kinds of new challenges every day. But I think sample quality is one of the most significant challenges and it's crucial to create awareness among our customers about the importance of using high-quality materials for accurate results. We have conducted tests to determine the minimum quality standards required for analysis. It is important to remember that "garbage in, garbage out" applies here, as poor-quality samples can lead to a huge amount of inaccurate data and a waste of resources. For instance, we perform an incoming quality check on a sample to ensure a successful repertoire profiling. This approach has helped our clients understand the quality of materials they can provide and has improved the overall workflow. RNA measurements are another way to ensure that you have good-quality samples. We can do that with a Bioanalyzer as Ilse mentioned and the spike-in for in-process controls. And finally, the diversity of our immune system is a challenge in terms of reproducibility.

I think nowadays we can generate reproducible data, but this depends on the experimental design of a study.

What bioinformatics tools or software do you use for analyzing immune repertoire sequencing data? Can you explain their functionalities and advantages?

Henk-Jan van den Ham: We have various software tools for different stages of data analysis. One of our first steps is to extract the specific bits of data we need, such as immune receptors. We use our software, called **IGX-Profile**, for this task. By developing our own tool, we have complete control over the process and can ensure high code quality and output consistency. After this initial step, there are several analysis workflows available, including our full-service workflow that we developed in our partnership with Cerba Research. This workflow is a combination of wet- and dry- lab techniques to obtain high-quality Rep-Seq data. The samples are sent to Cerba Research, who then perform amplification and sequencing and transfer the data to us, to ENPICOM. We then process the data and provide a report on the number of receptors, sequences, and other relevant data. This pipeline includes some of the well-known quality control parameters that are used for assessing high-throughput sequencing data quality. There are also a few parameters that are quite specific and that we developed by utilizing the combined strengths of our wet-lab and dry-lab expertise. One of them, for instance, is receptor recovery rate. This is something we've used quite frequently because it allows you to evaluate the efficiency of the whole pipeline and ensure that your sequencing setup is producing the desired immune receptors. So, if you sequence a lot but you get very few receptors, you're wasting your sequencing space or something's not going right in your protocol.

I think that's the biggest advantage of working together. You can integrate both wet-lab and dry-lab expertise into a streamlined workflow.

How do you handle the large volume of sequencing data generated from immune repertoire studies? What strategies do you employ for data management and storage?

Henk-Jan van den Ham: We currently work with MiSeq sequencing data, but NextSeq data generation is next on our list. This means that the number of sequences and samples will increase by 5 to 10 times, posing a challenge to how we currently process and handle the data. But, since we develop our own software and data storage solutions, we can accommodate the increased volume of data. Our software scales well and our analysis platform runs in the cloud, making us future proof. Data storage is not a problem per se, but searching the data can be tricky due to the large number of receptors. If in a more distant future, when large sequencing data sets become available for immune repertoire sequencing, we will meet the challenge head on.

Are there any specific applications or research areas where immune repertoire sequencing has provided valuable insights? Can you share any examples of your work in these areas?

Goran Marinkovic: There is a broad spectrum of applications ranging from basic research to preclinical clinical applications and diagnostics. However, if we focus on BCR profiling or TCR profiling in particular, these have different applications. BCR profiling is primarily used in the field of infectious diseases, whether for vaccine development or discovering antigens or antibodies. TCR profiling is mainly used in the field of oncology for assessing the effects of oncological therapy or to analyze the immunological profile of a tumor. It is also used in the development and monitoring of cell-based gene therapies such as CAR-T or TCR-T, and the detection of minimal residual disease (MRD). In terms of the practical examples of our work in the area, many sponsors have contacted us to help them with immune repertoire profiling, mainly for vaccine development, MRD detection, and tumor-infiltrating lymphocytes (TILs) profiling.

To illustrate some of the added benefits of repertoire profiling, let's talk about MRD detection. Flow cytometry is currently the gold standard for detecting minimal residual disease, but it has limitations in terms of the number of cells that can be analyzed. You can detect an event per 100,000 analyzed cells. However, next-generation sequencing and immune repertoire profiling have a 50 to 100 times higher resolution and can detect events in one in 5-10 million cells.

In the next 5 to 10 years, as the oncology field moves towards immuno-oncological solutions, profiling immune responses will become more important. Understanding how a patient's immune system responds to treatment is critical to understanding the possible outcomes of the treatment. Next-generation sequencing and immune repertoire profiling are significantly more powerful than other technologies and will become more accessible in terms of price as time goes on. Although there are technologies that can complement next-generation sequencing, there is currently no technology more powerful than it.

Are there any other medical fields of application?

Dirk van Alewijk: I think that the repertoire profiling technology might also be applicable in fields of aging, allergy, and autoimmune diseases like arthritis, lupus, and multiple sclerosis.

Henk-Jan van den Ham: There are many different applications related to lymphocytes. Minimal residual disease was already mentioned and, although not FDA approved yet, there is growing evidence that IRS can be used for MRD detection being demonstrably more sensitive than any other technique available to date. Other fields are still in the 'research use only' phase, but there are many developments happening currently. These applications are often related to patient monitoring for immunotherapy or vaccine research. Another application worth noting and where our analysis platform is frequently applied is the discovery of new antibody-based drugs. For instance, new therapeutics that bind certain tumor cells or targets that have immunomodulatory effects. These targets are often found through repertoire sequencing, which is a quick way to screen large pools of antibodies. This is one of the main applications of B cell receptor sequencing. It's interesting to note that this is usually a biotechnological application rather than a patient-based application at the moment, because you work with animal or synthetic libraries, often in combination with techniques like phage display, yeast, or hybrid screening.

What are some emerging trends or advancements in the field of immune repertoire sequencing that you find particularly exciting or promising?

Dirk van Alewijk: I find the topics of immunotherapy and detection of minimal residual disease fascinating. The potential of this technology in these research areas holds the biggest promise in the near future. Cancer remains a leading cause of death worldwide, but we must not forget that our body harbors unharnessed potential to fight tumors, through the immune system. Thus, the development of therapies that are potentiating our own immune response against cancer is gaining momentum. These immunomodulatory approaches can make the treatment of cancer highly precise, more effective, and less harmful to patients compared to other treatments such as chemotherapy, radiotherapy, and hormone therapies. I had the opportunity to attend the ASGCT conference in LA recently and saw the huge progress that's being made in cancer treatment over the last few years. I am hopeful that these advancements will bring us closer to finding a cure for cancer. In this cure for cancer, detection of MRD is crucial in preventing tumors from returning. I believe that immune repertoire profiling, even though it is an indirect method, can be a powerful technique for detecting any remaining tumor cells.

Collaboration is often crucial in scientific research. What are the most interesting collaborations regarding immune repertoire sequencing you've participated in?

Goran Marinkovic: It's a bit difficult to name a specific project. I'll share two most interesting examples that I have encountered since I've been working at Cerba Research. One project involved developing an assay to detect minimal residual disease, while another one focused on profiling tumor-infiltrating lymphocytes (TILs). MRD detection is used for diagnosis, while TILs profiling is used to monitor tumors and their response to therapy. Both applications are very challenging. In the case of MRD, its detection is often done with flow cytometry, which is a far less powerful technology compared to next-generation sequencing and repertoire sequencing. Using NGS to detect MRD is an exciting direction for further development. TILs profiling has a very promising role in diagnostics, but also it can also be important from the perspective of learning from the immune system.

Immune repertoire profiling data is used as the basis for designing efficient CAR-T or TCR-T therapies. These two applications are among my favorites.
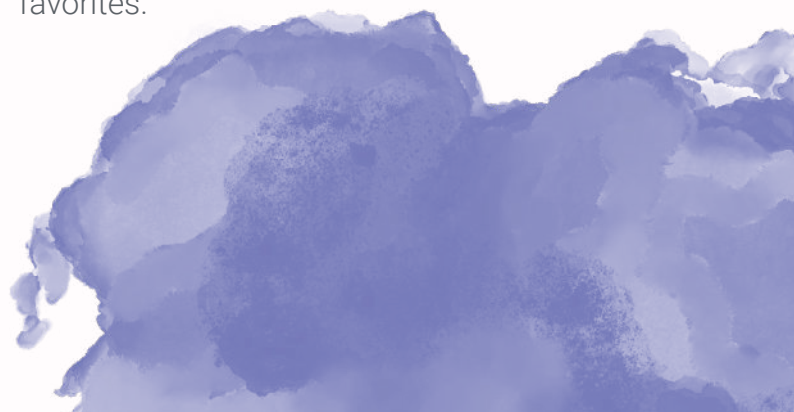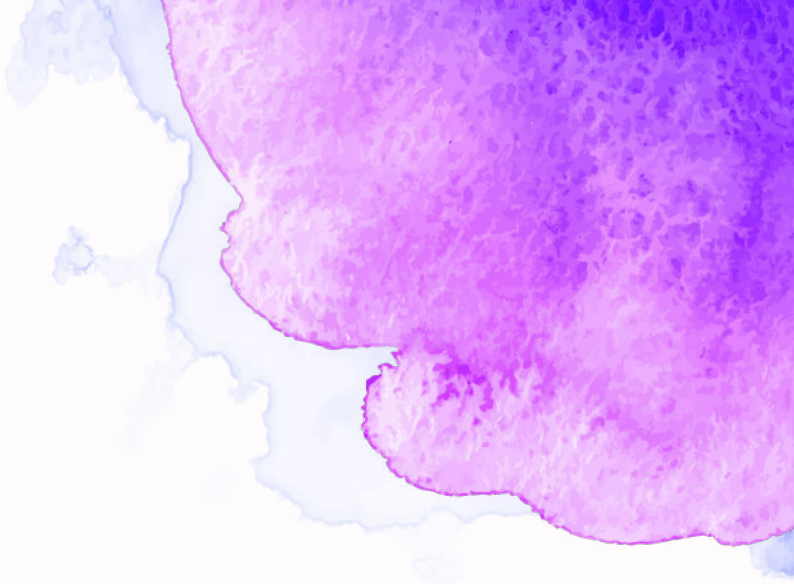
Henk-Jan van den Ham: Let me add another example that I find interesting. ENPICOM is collaborating with academic and corporate research groups in a large European consortium called **Inno4Vac,** and specifically the sub-group VaxPred. This consortium aims to rationalize vaccine design and production. And specifically for immune repertoire, the aim is to use repertoire data to better predict whether a candidate vaccine is good at inducing immune responses. This is crucial because if a vaccine fails to elicit an immune response, it simply doesn't work. To achieve this, our partners are developing a computational pipeline that will incorporate repertoire data. I think it is exciting to note that Cerba Research is also part of this project now, helping generate B cell receptor data. Overall, this collaboration highlights the potential of repertoire data in rational vaccine design.

In your opinion, what are the key future directions for immune repertoire sequencing? What are some of the unanswered questions or areas that require further investigation?

Ilse Wolters: There are several areas to explore in repertoire sequencing, such as tumor-infiltrating lymphocytes, as Goran mentioned earlier. We have isolated the RNA from the FFPE material, and we are just beginning to sequence them. There are many possibilities for analyzing the immune repertoire with this type of material and the RNA obtained from it. For example, we can look into repertoire sequencing for species other than human, as we have done with mouse samples in the past. Next, we currently perform TCR sequencing on the alpha and beta chains, but we can also explore gamma and delta chains in the future. Our immune repertoire sequencing is currently bulk sequencing, where all types of B or T lymphocytes are sequenced at once. And single-cell sequencing is an area that requires further investigation and is something we may explore in the future.

What is unique about your full-service proposition?

Henk-Jan van den Ham: I think the unique aspect of our collaboration is the integration of wet and dry labs. Our teams meet regularly, and we've learned to work with specifics of each other's fields of expertise. This collaboration allows us to identify and address issues more effectively, making for a far better service proposition. In contrast to traditional sequencing projects, where the wet and dry labs work independently, our approach ensures better data quality as both sides are aligned and communicate regularly. Communication is key, and our team talks a lot, especially when something doesn't seem right. We have not lost our curiosity, which has been instrumental in setting up a good service.

Dirk van Alewijk: With our scientific backgrounds, we look into the data through an analytical lens. Cerba Research offers an end-to-end concept for clinical trial testing: from sample collection on site where labs all over the world are partnering with us, and where extensive data analysis is performed on request. While we provide standard reports, we involve the ENPICOM team for more complex questions that require further data analysis. Also, it's worth noting that Cerba Research excels in three areas of expertise: infectious diseases, immune-oncology, and cell- and gene therapy. Within the organization, the Rijswijk site is considered the center of excellence in next-generation sequencing.

Goran Marinkovic: Our specialty is providing custom-made services. Basically, each request is unique to the sponsor's needs. There's always a slight variation in what they are interested in. We have a tailored approach to answering their requests. We have extensive experience in servicing clinical trials. Additionally, we have a global network of labs and experts, so regardless of your location in Europe, the US, or Asia, we can provide a similar level of service and deliver high-quality reported data.

The world of immune repertoire sequencing is fascinating and fast-growing. Its applications range from oncology to vaccine development, and its potential is far-reaching. As sequencing continues to become more cost-effective, more opportunities for NGS start to open up. We would like to thank Goran Marinkovic, Dirk van Alewijk, Ilse Wolters, and Henk-Jan van den Ham for sharing their insights with us. If you are interested in learning how to get the most out of your data and efficiently move from sample to discovery, **talk to our experts about your research project today.**

Dirk van Alewijk, PhD

Dirk completed his PhD at the Erasmus Medical Center in Rotterdam and then worked as a Postdoc at the Netherlands Cancer Institute in Amsterdam and at a biotech company called PamGene in Den Bosch. He has been a Senior Scientist at Cerba Research in Rijswijk (previously Viroclinics-DDL) for almost 20 years. In this role, he oversees various clinical trial related development projects for leading pharma customers. Dirk finds it thrilling to see his expertise and personal interests aligning in immune repertoire sequencing, a technology with great potential for immunology, oncology, virology, and other fields in the near future.

Goran Marinković, PhD

After completing his PhD at the University of Amsterdam in 2014, Goran moved to Israel where he worked as a Postdoc at the Weizmann Institute of Science. After that, Goran moved to Sweden to work on the development of immunomodulatory therapy for the treatment of myocardial infarction. For the past year and a half, Goran has been employed as an immunologist at Cerba Research in Rijswijk.

Ilse Wolters

Ilse Wolters is a technician at Cerba Research in Rijswijk. She has been employed there for three years and completed her Master's degree in Molecular Life Sciences in August 2022. Since September 2020, Ilse has been focused on immune repertoire sequencing and has gained extensive experience and knowledge in all technical aspects of the process.

Henk-Jan van den Ham, PhD

As a Research Team Lead at ENPICOM, Henk-Jan focuses on the latest developments in immunology and bioinformatics, particularly in the analysis of adaptive immune repertoire. He obtained his PhD in Theoretical Immunology & Bioinformatics at Utrecht University, and subsequently worked as postdoc and staff scientist at Erasmus MC, dept Viroscience in Rotterdam. In his current role, Henk-Jan is responsible for leading academic and industry projects, collaborations, and grant applications at the intersection of immunology, bioinformatics, and software engineering.

References

1. https://cerbaresearch.com/specialty-lab-biomarker-services/genomics/irs/
2. https://www.inno4vac.eu/partners
3. https://enpicom.com