Liquid biopsy-based T-cell receptor profiling in lung cancer patients undergoing immunotherapy

Master's thesis in Biotechnology

ALMA GLAVENIUS

DEPARTMENT OF MATHEMATICAL SCIENCES

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026
www.chalmers.se

Master's thesis 2026

Liquid biopsy-based T-cell receptor profiling in lung cancer patients undergoing immunotherapy

ALMA GLAVENIUS

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Mathematical Sciences
Division of Applied Mathematics and Statistics
Chalmers University of Technology
Gothenburg, Sweden 2026

Liquid biopsy-based T-cell receptor profiling in lung cancer patients undergoing immunotherapy
ALMA GLAVENIUS

Supervisor: Anna Rohlin, Department of Clinical Genetics and Genomics, Sahlgrenska University Hospital and Department of Laboratory Medicine, Institute for Biomedicine, Sahlgrenska Academy, University of Gothenburg

Supervisor: Johanna Svensson, Department of Clinical Genetics and Genomics, Sahlgrenska University Hospital and Department of Laboratory Medicine, Institute for Biomedicine, Sahlgrenska Academy, University of Gothenburg

Examiner: Erik Kristiansson, Division of Applied Mathematics and Statistics, Department of Mathematical Sciences, Chalmers University of Technology

Master's Thesis 2026
Department of Mathematical Sciences
Division of Applied Mathematics and Statistics
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2026

Liquid biopsy-based T-cell receptor profiling in lung cancer patients undergoing immunotherapy

ALMA GLAVENIUS
Department of Mathematical Sciences
Chalmers University of Technology

Abstract

Non-small cell lung cancer (NSCLC) is the most common type of lung cancer, and immune checkpoint inhibitors targeting the PD-1/PD-L1 pathway have significantly improved patient outcomes. However, only a subset of patients respond to treatment, highlighting the need for predictive biomarkers. The T-cell receptor (TCR) repertoire reflects the adaptive immune response and has emerged as a promising candidate biomarker for immunotherapy response.

The aim of this study was to investigate whether the characteristics of the peripheral blood TCR repertoire are associated with response to anti-PD-1/PD-L1 therapy in NSCLC patients. A total of 96 longitudinal blood samples from n=18 NSCLC patients undergoing anti-PD-1/PD-L1 treatment were analysed using TCR-beta sequencing with the SiMSen-seq method. Following quality control and bioinformatic processing with a molecular identifier groups-based error correction (MIGEC) pipeline, repertoire diversity was evaluated using the number of unique clonotypes, the D50 index, and the Shannon diversity index. In addition, clonal expansion patterns were assessed by tracking highly variable clonotypes over time.

Responders consistently exhibited higher TCR repertoire diversity than non-responders. At the pre-treatment baseline, responders showed a 55% higher median number of unique clonotypes, as well as higher median D50 and Shannon index values. Across all time points, responders maintained a more diverse repertoire, with a 33% higher median number of clonotypes, an 8 percentage points higher D50 index, and a 9% higher Shannon index compared with non-responders. Analysis of individual clonotype expansion revealed highly heterogeneous patterns and no clear distinction between responders and non-responders.

These findings suggest that overall TCR repertoire diversity is associated with improved response to anti-PD-1/PD-L1 therapy in NSCLC, whereas monitoring individual clonotypes alone appears to be insufficient as a predictive biomarker. Larger studies are required to validate these observations and assess the clinical utility of TCR repertoire profiling in immunotherapy response prediction.

Keywords: Non-small cell lung cancer, T-cell receptor repertoire, TCR-beta sequencing, Immune checkpoint inhibitors, Biomarkers, T-cell diversity.

Acknowledgements

I would like to express my sincere gratitude to everyone who has contributed to this thesis and supported me throughout this project.

First of all, I would like to thank my supervisor Anna Rohlin for giving me the opportunity to work on this project. Thank you for your valuable guidance, feedback and encouragement throughout my work. Your support has helped me grow in confidence, develop new skills, and gain a deeper understanding of research.

I would also like to thank my co-supervisor Johanna Svensson, for your continuous support throughout the project. Thank you for always taking the time to answer my questions, discuss results, provide feedback, and guide me through the experimental work. Your patience, encouragement, and day-to-day mentorship have been invaluable, and I am truly grateful for everything I have learned from working with you.

I would like to thank everyone at the Department of Clinical Genetics and Genomics at the University of Gothenburg and Sahlgrenska University Hospital for creating a welcoming and inspiring environment. A special thank you to Firaol Tamiru Kabede for valuable discussions about PCR and primer design, and for always taking the time to answer my questions during the laboratory work. Thank you to Tobias Österlund for your assistance with MIGEC and for teaching me everything I needed to know about the data analysis. I am also grateful to Sukanya Raghavan for many valuable and engaging discussions about immunology and the TCR repertoire.

To the rest of the BioLung group, thank you for welcoming me into your work and for all your support throughout the project. Finally, I would like to thank the patients who contributed samples to this study. Without their participation, this work would not have been possible.

This project has been challenging, but also incredibly informative, rewarding, and inspiring. I have learned a great deal and genuinely enjoyed the experience.

Alma Glavenius, Gothenburg, June 2026

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis listed in alphabetical order:

AC	Adenocarcinoma
APC	Antigen-Presenting Cells
CDR	Complementarity-Determining Region
D-segment	Diversity segment
gDNA	Genomic DNA
J-segment	Joining segment
MHC	Major Histocompatibility Complex
MIGEC	Molecular Identifier Groups-based Error Correction
NGS	Next-Generation Sequencing
NSCLC	Non-Small Cell Lung Cancer
OS	Overall Survival
PCR	Polymerase Chain Reaction
PD-1	Programmed Death-1
PD-L1	Programmed Death Ligand-1
pMHC	peptide-MHC
PFS	Progression-Free Survival
SCC	Squamous Cell Carcinoma
SiMSen-seq	Simple, multiplexed, PCR-based barcoding of DNA for sensitive mutation detection using sequencing
TCR	T-Cell Receptor
TMB	Tumour Mutational Burden
TRBC	T-cell Receptor Beta-chain C-segment
TRBD	T-cell Receptor Beta-chain D-segment
TRBJ	T-cell Receptor Beta-chain J-segment
TRBV	T-cell Receptor Beta-chain V-segment
UMI	Unique Molecular Identifier
V-segment	Variable segment

Contents

List of Figures

List of Tables

1
Introduction

Non-small cell lung cancer (NSCLC) is the most common form of lung cancer, which is the disease with the highest mortality rate and the second highest amount of reported cases of all cancers [4]. Immunotherapy has demonstrated long-term benefits in approximately 20–50% of patients with NSCLC [5]. However, a substantial proportion of patients fail to respond to treatment, highlighting the need for reliable biomarkers that can identify those most likely to benefit.

2
Theory

T-cells are central components of the adaptive immune system, and their activity is mediated by unique T-cell receptors (TCRs) that recognise specific antigens found on antigen presenting cells (APCs) [6]. The receptor has a heterodimeric structure made up of two peptide chains, each of which is divided into a constant and a variable domain. There are two subsets of T-cells depending on what peptide chains the TCR is composed of, the alpha-beta ($\alpha\beta$) and the gamma-delta ($\gamma\delta$), where alpha-beta T-cells make up the majority [7][8].

The variable region of the TCR is generated by somatic rearrangement of DNA segments called V (Variable), D (Diversity), and J (Joining), as well as by random deletion and addition of template and non-template nucleotides at the V(D)J junctions [6]. These regions are involved in antigen recognition and binding. The alpha chain is determined by V and J segments, while the beta chain is determined by all three segments. T-cells are divided into clonotypes based on the combination of V(D)J segments, where each clonotype shares the same combination.

The purpose of this project is to identify immune-signature response patterns in patients in the early treatment phase. This will be done by optimising the sequencing method of the TCR beta chain, and evaluate the diversity and frequency of T-cell clonotypes in each individual. The data will be compared with how patients respond to treatment, which can help better predict patient responses and select individuals who would benefit from immunotherapy.

2.1 Non-small cell lung cancer

The most common form of lung cancer is non-small cell lung cancer (NSCLC), comprising up to 85% of the cases. It has three main subtypes based on histology appearance, with adenocarcinoma (AC) and squamous cell carcinoma (SCC) being the most dominant (40% and 25-30% respectively) [4][9][10]. The third subtype is large cell carcinoma, making up 5-10%. A common risk factor for NSCLC across all subtypes is tobacco smoking. However, up to 25% of cases occur from non-smoking factors, such as exposure to carcinogenic chemicals and heavy metals, previous lung disease, and genetic ancestry, and make up approximately 10-15% of AC cases [11][4].

There are oncogenic alterations associated with lung cancer that drive tumour evolution and progression [4]. The total number of somatic alterations per megabase of tumour DNA is referred to as tumour mutational burden (TMB). Most lung

cancers have high genomic complexity with several alterations contributing to tumour growth, while some ACs have low genomic diversity and only one alteration that drives the disease progression. The most common oncogenic drivers associated with NSCLC are *EGFR*, *KRAS*, *ROS1*, *ALK*, *MET*, *BRAF*, *HER2*, *NTRK*, *RET*, and *NRG1* [12]. They are explored as therapeutic molecular targets for precision medicine and for predicting patient response to treatment.

2.1.1 Treatment for NSCLC

Over the past few decades, the treatment for NSCLC has been surgery, chemotherapy and/or radiotherapy, with platinum-based chemotherapy serving as the most common first-line therapy, with a median survival rate of 9-12 months [5]. For patients harbouring genomic alterations, targeted treatment approaches, including epidermal growth factor receptor (*EGFR*) tyrosine kinase inhibitors (TKIs) and anaplastic lymphoma kinase (*ALK*) inhibitors, have significantly improved treatment efficiency. However, since only a minority of patients have such alterations, treatments such as immune checkpoint inhibitor (ICI) therapy were developed to combat this issue.

Such treatment targets an inhibitory immune checkpoint molecule to reactivate the immune system to target tumour cells. ICIs used in NSCLC treatment are monoclonal antibodies which target the programmed death-ligand 1 (PD-L1) located on tumour cells, and its receptor programmed death-1 (PD-1) located on T-cells [5][13]. When tumour cells increase the expression of PD-L1, the ligands engage PD-1 receptors expressed on T-cells and deactivate the T-cells and suppress an immune response [14]. This enables further proliferation of tumour cells. ICIs work by blocking the PD-1 or PD-L1 immune function, reactivating the T-cells and restoring their cytotoxic functions. Anti-PD-1/PD-L1 therapy has demonstrated durable clinical responses and overall survival in NSCLC patients [15], and is used in combination with chemotherapy or as monotherapy [4]. Initially, anti-PD-1/PD-L1 therapy was used as monotherapy in the second-line setting, but today nearly all patients, who do not have targetable oncogenes, receive it in the first-line [16].

2.2 Biomarkers to predict treatment response

To date, there are only two recognised biomarkers used for clinical prediction of response to anti-PD-1/PD-L1 treatment, namely PD-L1 expression on tumour tissue and tumour mutational burden (TMB) [17][15]. PD-L1 is approved as a biomarker for clinical practice by both the European Medicines Agency (EMA) and the Food and Drug Administration (FDA) in the United States, while TMB is only approved for use by the FDA.

2.2.1 Established biomarkers

For monotherapy with anti-PD-1/PD-L1, a PD-L1 expression of $\geq$50% is the most reliable threshold for selection of patients likely to benefit [17]. The progression-

free survival (PFS) and overall survival (OS) have not been shown to improve for this patient group with the addition of chemotherapy. For patients with PD-L1 <50%, the standard approach is to combine ICIs and chemotherapy. However, the PD-L1 biomarker is limited due to several reasons. There are patients above or below the 50% threshold with different sensitivity to ICIs, meaning that patients that might benefit from treatment might be overlooked [15]. In the case of combined chemo- and immunotherapy, the PD-L1 biomarker is of little to no predictive value. Additionally, the biopsy site of the tumour can significantly influence the PD-L1 value due to its high variation in expression across the tumour [17]. Even though TMB shows promise as a predictive biomarker, a lack of universally standardized assays, thresholds, and analysis pipelines makes clinical implementation difficult. TMB results can also vary significantly in analyses of the same patient sample or between samples due to differences in quality and quantity.

2.2.2 Emerging biomarkers

There are new emerging biomarkers that are being investigated or employed to characterise NSCLC and predict response to ICI treatment. Genomic alterations in oncogenic driver genes such as *EGFR* and *ALK*, and co-variants in *STK11* and *KEAP1* have been associated with resistance or poor response to ICI treatment [18]. On the other hand, the oncogenic driver *KRAS* has been associated with better response, and *KRAS* in combination with other variants such as *LRP1B* and *TP53* has recently been liked to favourable outcomes.

Moreover, the T-cell repertoire has gained increasing interest for its potential as a biomarker for ICI therapy, since anti-PD-1/PD-L1 therapy directly targets T-cells [19]. Patients with high T-cell diversity before the start of immunotherapy treatment and an increased clonality after treatment, have been shown to have a better response and longer progression-free survival (PFS). Thus, T-cell diversity and clonality might serve as predictors of anti-PD-1/PD-L1 treatment, although more research is needed on the subject.

2.3 T-cell receptor

The T-cell receptor (TCR) is located on T-cells [20]. These immune cells can recognise and bind to foreign substrates, such as abnormal cells, tumour cells, cells from other organisms, and cells infected with a virus or other microorganism. The mechanism behind this is that TCRs can bind to antigen peptides that are presented on major histocompatibility complex (MHC) molecules by these antigen presenting cells (APCs). When MHC binds to an antigen peptide, it forms the peptide-MHC complex (pMHC), which can be recognised by the TCRs [21]. This facilitates a cytotoxic immune response by T-cells.

The TCR has a heterodimeric structure and is composed of two peptide chains. There are two groups of TCRs, the alpha-beta ($\alpha\beta$) and the gamma-delta ($\gamma\delta$), referring to the paired peptide chains [1]. Alpha-beta TCRs make up the majority

of peripheral T-cells, which are matured T-cells located outside of the bone marrow in the lymphatic system [22]. Alpha-beta TCRs recognise antigens presented on the MHC, while gamma-delta TCRs are not MHC-restricted.

2.3.1 V(D)J Recombination

The TCR peptide chains are divided into a genetically constant and a variable region as seen in Figure 2.1 A. The variable region of the TCR is generated by somatic rearrangement of V(D)J DNA segments, generating a theoretical number of 10^{18} unique receptors [6]. Therefore, there is a minimal chance of finding identical TCRs in two individuals. The alpha and gamma chains have V (Variable) and J (Joining) segments, while the beta and delta chains also have a D (Diversity) segment. The variability of the TCR is mostly due to three complementarity determining regions (CDRs) located at the ends of each of the chains in the shape of hairpin loops. The six loops sit at the membrane-distal end of the TCR extracellular domain, and together they form the antigen-binding site, Figure 2.1 C. The CDR1 and CDR2 regions are encoded by the V-segment. The CDR3 region is encoded by the V and J-segments on the alpha chain, and by the V, J, and D-segments on the beta chain. Due to the CDR3-beta region offering the most information on antigen specificity and being genomically hypervariable, it is a common target for TCR sequencing [1].

The TCR-beta locus *trb* consists of 48 TCR-beta variable (TRBV) segments, two TCR-beta diversity (TRBD) segments, 12 TCR-beta joining (TRBJ) segments, and two TCR-beta constant (TRBC) segments [6]. The recombination of the TCR-beta chain is a two-step process, starting with recombination of the D and J segments. This is done by combining either the TRBD1 with one of many TRBJ1 segments, or TRBD2 with one of many TRBJ2 segments. Then the V-segment is included in the second step with the already combined D-J segments, by allowing the D-J segment to recombine with one of the many TRBV segments. After splicing, the transcript of the TCR-beta chain is obtained with V, D, J, and C segments. The V(D)J recombination process is visualised in Figure 2.1 B.

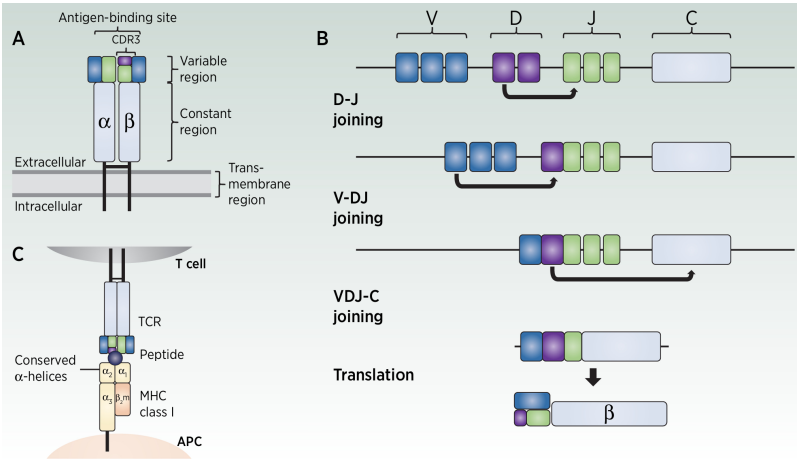


Figure 2.1: Overview of the TCR, V(D)J recombination of TCR-beta chain, and the TCR-MHC binding. *Reprinted from* [1].

2.4 Measures of TCR diversity

A way to describe the TCR repertoire is to consider its diversity and clonality. Diversity is used to determine immune fitness in response to treatment and is commonly used in immune profiling studies, including in inflammatory diseases, viral infections, and cancer [23]. In TCR analysis, the diversity takes into account the richness and evenness of the repertoire. Richness is described as the number of unique TCR sequences, corresponding to the number of unique clonotypes. Evenness is described as the distribution of the clonotypes, i.e. their relative abundance and how the number of clones are distributed between the clonotypes. Evenness is highest when all clonotypes have the same number of clones, and lowest when a small number of clonotypes contain the highest number of clones. In terms of diversity, high richness and evenness gives high diversity, and vice versa. TCR clonality is reversely related to diversity, and refers to the number and frequency of TCRs that are observed multiple times in a sample. Figure 2.2 illustrates the relationship between clonality and diversity. When diversity is high, clonality is low, and vice versa.

When describing the diversity of the TCR repertoire, there are several diversity indices to choose from. Some of them describe only richness, some only evenness, and some describe both. The D50 index is the number of unique CDR3 sequences that account for 50% of the total number of sequences observed in the repertoire, and is thereby related to evenness [24]. The Shannon index also considers the distribution of clones across the clonotypes, but it takes all clonotypes into account, including rare ones, which D50 does not [25]. It is calculated by:

$$H = -\sum_{i=1}^{S} p_i \ln(p_i) \tag{2.1}$$

where S is the number of unique clonotypes in the sample and (p_i) is the proportion of total reads belonging to the i-th clonotype. For repertoires with a few dominant clonotypes, the Shannon index is closer to 0. When the sequences are evenly distributed, the index reaches its maximum, which corresponds to the logarithm of the number of unique sequences.

Clonality is defined as the inverse of diversity and is calculated by:

$$Clonality = 1 - H/ln(S) \tag{2.2}$$

where H is the Shannon index and S is the number of unique clonotypes in the sample [19]. A clonality score equal to 0 indicates that the repertoire is perfectly diverse, and a score equal to 1 indicates a monoclonal repertoire.

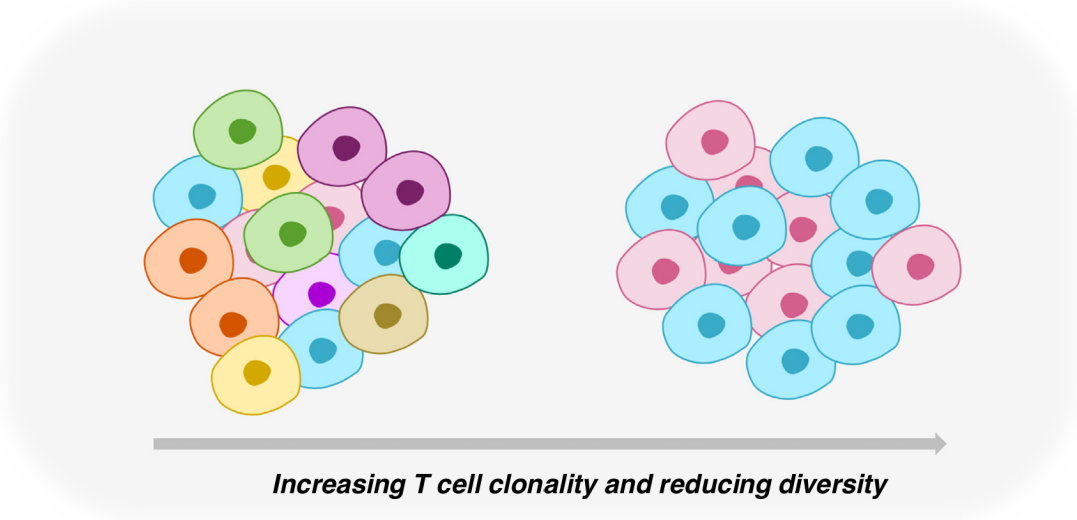


Figure 2.2: TCR clonality and diversity. *Reprinted from* [2].

2.5 SiMSen-seq

Simple, multiplexed, PCR-based barcoding of DNA for sensitive mutation detection using sequencing (SiMSen-seq) is a method developed for DNA analysis using next-generation sequencing (NGS) [26]. The method uses minimal DNA input, three cycles of barcoding PCR followed by adapter PCR, and purification using magnet beads to generate a targeted barcoded library ready for sequencing. The highly sensitive method can detect genetic variants at <1% frequency. Following the sequencing, a bioinformatic analysis pipeline is applied, which uses the barcodes, or unique molecular identifiers (UMIs), for near absolute error correction using consensus reads [27]. The method was developed to efficiently overcome errors introduced by DNA polymerases during library construction or sequencing, which are the major causes of background in NGS.

2.5.1 Library preparation

PCR (Polymerase Chain Reaction) is a laboratory method used for amplification of nucleic acids, like DNA and RNA [28]. In general, the process includes targeting specific fragments of DNA or RNA that have been extracted from a sample, and amplifying them through repeating cycles of denaturation, annealing, and extension. In the SiMSen-seq method, PCR is performed in two steps, namely barcoding PCR (PCR1) and adapter PCR (PCR2) as described in Figure 2.3 [26]. During barcoding PCR, the primers anneal to the target DNA, in this case the CDR3 region of the TCR-beta chain. The forward primers anneal to the V-segment and the reverse primers anneal to the J-segment, with the D-segment located between them. This step is performed in three cycles, ideally creating six copies of each original CDR3 sequence. The forward primers are designed to contain a hair-pin loop in which the barcode sequence, or unique molecular identifier (UMI) sequence, is located as

well as an adapter sequence. This structure protects the UMIs during the initial amplification and prevents them from annealing to other UMIs, which creates errors during the down-stream bioinformatic analysis. The hair-pin loop structure breaks apart after the cycles to prepare for the next PCR step.

Next, the products from the barcoding PCR go through adapter PCR, where the hair-pin structure on the forward primers is straightened out as a result of an increase in temperature. In this step, specific Illumina sequencing indexes anneal to the adapters located on the forward and reverse primers from the barcoding PCR, to facilitate amplification during sequencing. Every DNA sequence in the same sample receives the same Illumina index. This makes it possible to determine which sequence belongs to which sample when all samples are pooled and sequenced together.

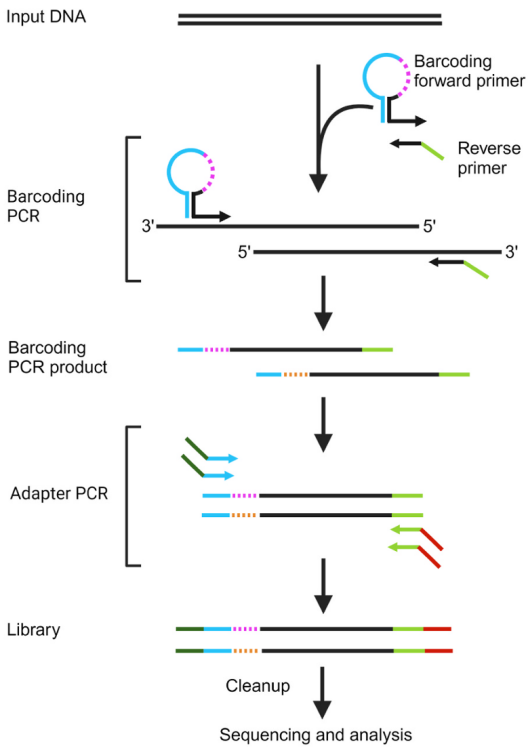


Figure 2.3: Overview of the SiMSen-seq library preparation using PCR in two steps. Adapter sequences are shown in blue and green, UMIs are shown in pink and orange, and sequencing indexes are shown in dark green and red. *Adapted from* [3].

2.5.2 Bioinformatic analysis

Sequencing of the TCR repertoire generates millions of reads per sample that contain information about its diversity and clonality . Since the sequences can be near-identical and exist in various concentrations, errors that occur during PCR and sequencing might make minor clonotypes indistinguishable, thus making the results less accurate [27]. MIGEC (molecular identifier groups-based error correction) is a bioinformatic pipeline that is nearly error free in correcting for such errors. During PCR1, each forward primer sequence contains a unique UMI sequence, usually

consisting of 8-12 randomised nucleotides [29]. After amplification, each original DNA molecule will have thousands of copies with the same UMI sequence. In the MIGEC method, the sequences with the same UMI are grouped together, and a consensus read is generated for each group to correct for polymerase-induced errors and uneven amplification, see Figure 2.4. Each consensus read corresponds to an original TCR-beta CDR3 sequence and is counted only once, allowing accurate quantification of each T-cell clonotype without errors from PCR duplicates.

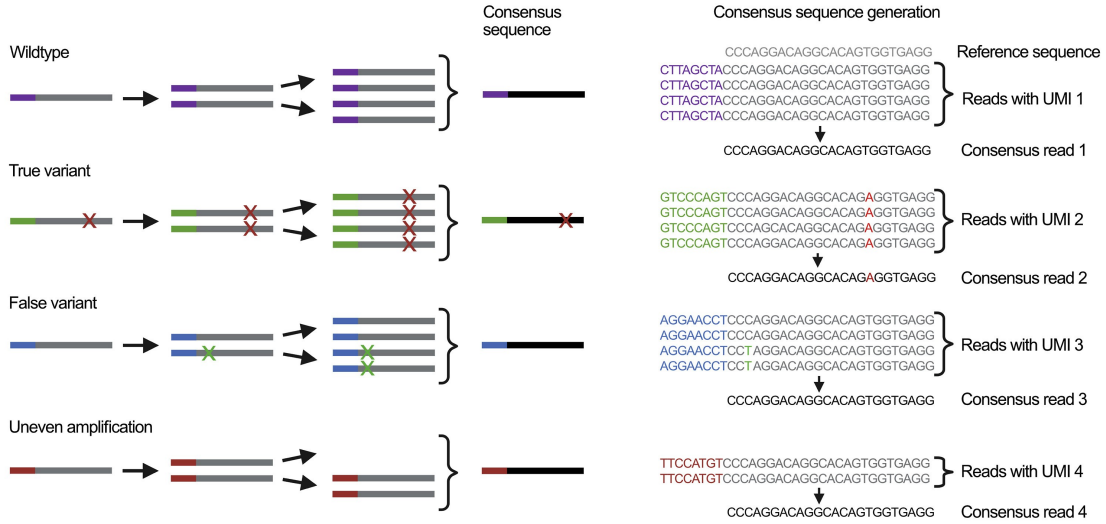


Figure 2.4: Overview of how DNA target sequences (grey) are analysed using UMIs and consensus reads (purple, green, blue, red) according to the MIGEC method. *Adapted from* [3].

2.6 Aims

The general aim of the thesis is to provide information on the ability of the TCR repertoire to be applied as a predictive biomarker for ICI treatment response in NSCLC patients. To achieve this overall aim, it has been divided into partial aims.

- The first partial aim is to apply and optimise the existing next-generation sequencing method SiMSen-seq for TCR-beta chain CDR3 sequencing.
- The second partial aim is to analyse the TCR repertoire of lung cancer patients. This will be done by considering the diversity and expansion of clonotypes, and the presence of TRBV segments among the patients. The results will be correlated to patient treatment response to ICI therapy.

3
Materials and methods

3.1 Patient cohort

The cohort in this study consisted of 18 patients from the BioLung study at Sahlgrenska University Hospital, Department of Clinical Genetics and Genomics. Blood samples from 18 patients were taken at up to 6 time points during treatment with ICI therapy. The sampling process followed the scheme illustrated in Figure 3.1, and began with a baseline sample before the start of treatment (time point A). During the treatment period, samples were collected with three-week intervals coinciding with treatment occasions (time points B through E). In the case of clinical progress of the cancer, a progress sample (time point F) was taken, which could be collected anytime during the treatment process. For patients were no progress was seen, a follow-up sample (time point I) was collected one year after the baseline sample was collected. For the patient cohort in this study, the results from this sampling process generated a total of 96 samples to be sequenced and further analysed.

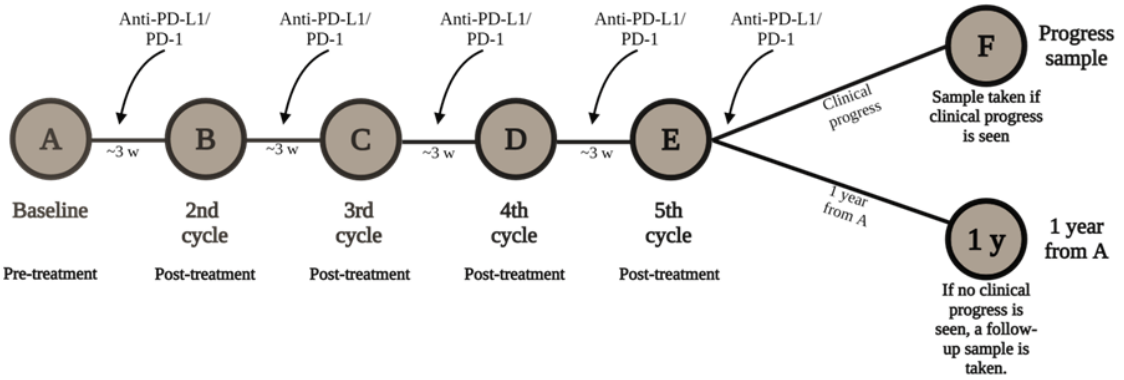


Figure 3.1: Sampling process, from pre-treatment to clinical progress or 1 year follow-up, with three week intervals during the first five treatment cycles. *Image courtesy of the BioLung Study Consortium.*

All patients in the cohort were treated with anti-PD-1/PD-L1 as first-line therapy, either as mono-therapy or in combination with chemotherapy. Table 3.1 contains patient information. The median age was 68 years old, 9 (50%) were female, and 14 (78%) had adenocarcinoma (AC) histology, and the rest had squamous cell carcinoma (SCC). Regarding smoking history, 7 (39%) were active smokers, 9 (50%) were former smokers, and 2 (11%) had never smoked before. Two of the patients (11%) had stage III NSCLC, and the rest had stage IV. There were 8 (45%) re-

sponders to treatment and 10 (55%) non-responders, based on clinical response seen from radiological imaging. For this study, the cut-off for response was set to nine months. As the only clinically applied biomarker, PD-L1% was also reported for all patients. The value ranged between 0-95% for the cohort, with 6 (33%) patients having a value above the threshold of $\geq$50% and 12 (67%) patients having a value below the threshold of $<$50%. The patient metadata are summarised in Table 3.2.

Patient-ID	Sex	Response	Age	Smoking	Histology	Stage	PD-L1 (%)
BL-005	Male	Responder	54	Smoker	SCC	IV	1
BL-009	Female	Responder	68	F. Smoker	AC	IV	70
BL-014	Female	Non-responder	65	Smoker	AC	III	10
BL-015	Male	Non-responder	69	Smoker	AC	IV	70
BL-021	Male	Non-responder	74	Smoker	SCC	IV	40
BL-025	Male	Responder	79	F. Smoker	AC	IV	95
BL-026	Female	Non-responder	61	F. Smoker	AC	IV	20
BL-028	Female	Responder	75	F. Smoker	SCC	III	5
BL-029	Male	Non-responder	76	F. Smoker	AC	IV	5
BL-032	Male	Non-responder	61	N. Smoker	AC	IV	1
BL-038	Female	Non-responder	63	F. Smoker	AC	IV	90
BL-045	Male	Responder	65	Smoker	AC	IV	95
BL-046	Male	Responder	76	Smoker	AC	IV	1
BL-047	Female	Non-responder	80	F. Smoker	AC	IV	60
BL-063	Female	Non-responder	68	F. Smoker	AC	IV	0
BL-071	Female	Responder	38	F. Smoker	AC	IV	10
BL-101	Male	Non-responder	52	N. Smoker	AC	IV	5
BL-122	Female	Responder	78	Smoker	SCC	IV	3

Table 3.1: Patient information. Abbreviations: F. Smoker, Former Smoker; N. Smoker, Never Smoker; SCC, squamous cell carcinoma; AC, adenocarcinoma.

Sex	Male	9 (50%)
	Female	9 (50%)
Respons	Responder	8 (45%)
	Non-responder	10 (55%)
Age	Range	38-80
	Median	68 years old
Smoking	Smoker	7 (39%)
	F. Smoker	9 (50%)
	N. Smoker	2 (11%)
Histology	AC	14 (78%)
	SCC	4 (22%)
Stage	III	2 (11%)
	IV	16 (89%)
PD-L1	$\geq 50\%$	6 (33%)
	$<50\%$	12 (67%)

Table 3.2: Metadata for the patient cohort (n = 18).

3.2 SiMSen-seq

The project used the established SiMSen-seq method for library preparation, next-generation sequencing, and bioinformatic analysis of the T-cell receptor repertoire of the patients in the cohort [26]. Blood samples from the patients were prepared for sequencing using the workflow outlined in Figure 3.2. The sequencing data was error corrected using the MIGEC method, and analysed and correlated to patient information in R (version 4.3.3). The SiMSen-seq workflow was also evaluated and optimised for TCR-beta CDR3 sequencing.

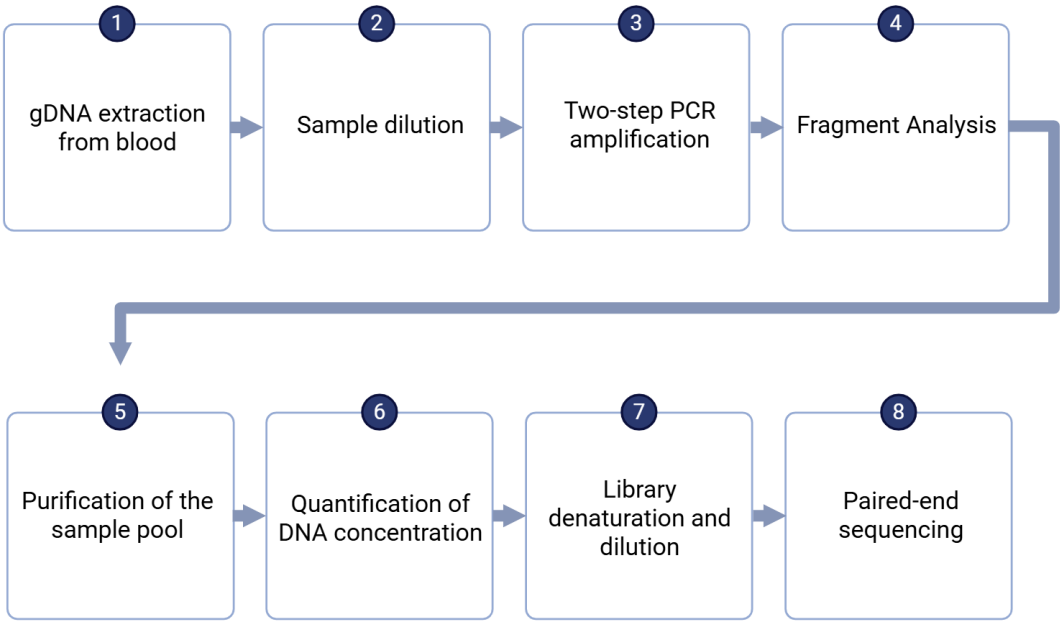


Figure 3.2: Overview of the workflow for the library preparation process for sequencing using the SiMSen-seq method. *Generated by the author.*

3.2.1 Library preparation

Genomic DNA (gDNA) from blood samples was extracted manually or automatically. Manual extraction was performed using a QIAamp DNA Blood Mini Kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions. Automatic extraction of DNA was performed using an EZ2 instrument with a EZ1&2 DNA Blood 200µl kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions.

The concentration of the extracted DNA samples was measured. For the PCR, a final amount of 200ng DNA in each well was desired. Since all samples have different starting DNA concentration, it is required to dilute them with nuclease-free or MQ water.

The PCR was performed in two steps, barcoding PCR (PCR1) and adapter PCR

(PCR2). In PCR1, primers containing a unique molecular identifier (UMI) and an adapter anneal to the DNA molecule, which were amplified in three thermal cycles. The UMI and adapter are protected in a hair-pin structure to prevent primer-dimers. Two thirds of the product from PCR1 was transferred to PCR2. In PCR2, the hair-pin loops were broken and straightened out as a result of an increase in temperature, allowing Illumina sequencing indexes to anneal to the adapters. The DNA fragments were amplified in 27 thermal cycles.

For this work, the primers used for PCR amplification were obtained from the EuroClonality Consortium [30]. They are designed as forward and reverse primers for paired-end sequencing. The forward primer anneals to the V-segment of the TCR region, and the reverse primer anneals to the TCR J-segment. The V-segment is divided into 30 subgroups, (named TRBV1 to TRBV 30), which are in turn divided into genes, (TRBV3-1, TRBV3-2 etc.), and alleles (TRBV3-1*01 etc.) according to the international ImMunoGeneTics information system [31]. The J-segment is divided into two subgroups (TRBJ1 and TRBJ2), which are also divided into genes (TRBJ1-1, TRBJ1-2 etcetera) and alleles (TRBJ1-1*01) [32].

After amplification, the length of the DNA fragments were analysed to ensure correct amplification. This was done using a Fragment Analyzer (Agilent Technologies). The samples were diluted 1:10 in TE buffer. The Fragment Analyzer plate was prepared using NGS Fragment kit (1-6000bp), according to the manufacturer's protocol. The resulting data was analysed in Agilent ProSize data analysis software. The software provides information about fragment length and concentration. The fragments of interest are approximately 350bp long. From this data, the samples with fragments of correct length were selected for the next steps in the analysis.

Before pooling, the samples were normalised to ensure an equal representation in the pool, achieving uniform quantity for all components and an equal number of sequencing reads per UMI and sample. The samples from the PCR2 plate were pooled in an Eppendorf tube.

Purification was performed in two steps using AmPureXP magnet beads (Beckman Coulter, Brea, CA, USA) and the Pippin Prep system (Sage Science, Beverly, MA, USA). For bead purification, double the volume of beads was added to the pool and 70% EtOH was used for clean up. The bead pellet was then resuspended in 60µl of nuclease-free water. For Pippin Prep, loading solution was added to the tube, vortexed, and spun down. The pool was separated on a gel plate to isolate fragments between 250–450bp, resulting in 40µl of purified eluate.

The purified pool was then analysed on TapeStation (Agilent Technologies, Santa Clara, CA, USA) with a High sensitivity D1000 kit and Qubit (Thermo Fisher Scientific, Waltham, MA, USA) with a 1X dsDNA High Sensitivity (HS) kit, according to the manufacturer's instructions, to ensure the purification steps were successful. For TapeStation, 2µl of buffer solution and 2µl of pool or ladder were added to sample tubes, vortexed and spun down. After analysis, the average fragment length and concentration of the pool were obtained. The pool with the best result, i.e. the highest peak, was then measured in triplicate on Qubit, and the average concentration was calculated.

3.2.2 Sequencing using NextSeq (Illumina)

The pool was diluted to 1nM. The pool and PhiX were then denatured separately with NaOH, neutralised with Tris-HCl and prepared to a 20pM denatured pool or PhiX. The last step in the preparation was to mix the library pool solution with PhiX to reach a final concentration of 20% PhiX in the library pool solution. The pool was then ready to be loaded into the Illumina NextSeq instrument and start sequencing. This was done using a NextSeq 550 Dx (Illumina, San Diego, CA, USA) Kit: NextSeq 500/550 Mid Output Kit v2.5 (300 Cycles). A total of 96 samples were sequenced and divided between two runs with 48 samples on each run.

3.3 Error correction and analysis of data

Quality scores and parameters from the sequencing runs were obtained as MultiQC reports. The raw sequencing results were error corrected using a MIGEC pipeline [29]. The outputs from the pipeline were obtained as .txt files, which were compiled and analysed using R (version 4.3.3) with the *immunarch* package.

For each patient and time point, information about all discovered unique CDR3 sequences was obtained, including their nucleotide and amino acid sequence and V/D/J-segments used. Every unique CDR3 sequence corresponds to a unique T-cell clonotype. The sequencing information was compiled to show a summary of the total number of unique clonotypes, the total number of clones, as well as the amount of times every V-segment was found in each sample. Several diversity indices were also calculated for each sample, including Chao1, Shannon, Simpson, Gini-Simpson, and D50. Ultimately, the D50 and Shannon indices were selected for the report. For each patient, the top 10 clonotypes with the highest variation across the time points were analysed. This was done by counting the amount of times each clonotype was observed in each time point, and calculating the standard deviation for each clonotype between the time points. The 10 clonotypes with the highest standard deviation were selected and their frequencies at each time point were visualised. These results gave insight into clonotype diversity and clonal expansion, which were correlated with clinical patient information. To simultaneously evaluate relationships between TRBV segment usage and longitudinal sample profiles, unsupervised two-way hierarchical clustering was performed using the *pheatmap* package in R. Both rows (unique V-segments) and columns (sample time points) were clustered using Euclidean distance and complete linkage, allowing for the identification of coordinated V-segment expression modules.

4

Results and discussion

In total, 18 patients were included in the cohort, and resulted in a total of 96 different samples. For some patients, certain time points were not included due to missing samples. An overview of the data regarding total number of clones, total number of unique clonotypes, D50 and Shannon diversity indices for each patient and time point is compiled in Table A.1 in Appendix A.

4.1 Sequencing quality parameters

SiMSen-seq requires a high read depth to ensure that each unique UMI receives multiple reads, which enables error correction using consensus reads through the MIGEC method [26]. By using 200ng of input DNA to each sample, the number of required reads per sample was calculated to be approximately 2,3M. The sequencing kit used has a capacity for 130M single reads (260M paired-end reads) and with a Phix concentration of 20% in the library pool, the required amount of reads per sample was calculated to be approximately 2,2M [33]. Tables A.2 and A.3 in Appendix 1 include the general information provided from the MultiQC reports from the two sequencing runs of the patient samples. Table 4.1 shows information from the MultiQC reports. The average total number of reads per sample was 2,9M and the median was 2,7M, including the values from both runs in the metrics. The lowest and highest number of total reads were 1,5M and 7,8M, respectively. The sum of all total number of reads per sample was approximately 282M, which is higher than the 260M paired-end reads the sequencing kit provides. The cluster densities on the flow cells for the two sequencing runs were 255 $\mathrm{K/mm^2}$ and 233 $\mathrm{K/mm^2}$. These are higher than the optimal values between 170-220 $\mathrm{K/mm^2}$, which might have a negative effect on data quality [34]. However, the high reads depths indicate that the SiMSen-seq workflow followed in this study was successful for TCR-beta sequencing.

	Number of samples	**M reads (average)**	**M reads (range)**	**Cluster density** [$\mathrm{K/mm^2}$]
Run 1	48	3,00	1,5 - 5,9	255
Run 2	48	2,88	1,7 - 7,8	233

Table 4.1: Summary of the number of reads per sample for the two sequencing runs performed, and cluster densities on the flow cells.

4.1.1 Evaluation of primer sequences

Table 4.2 gives an overview of the V-segments covered by the EuroClonality primers. The second column provides information about the V-segments found in the sequencing data after error correction using the MIGEC pipeline and further analysis in R. To begin, not all of the known V-segments were covered by the primers, which means they were not amplified and sequenced. This would lead to uncompleted results when analysing the TCR repertoires for the patients, since not all possible CDR3 sequences could be amplified. Moreover, the V-segments marked in red in the output column are V-segments that were found in the data without having a primer containing the matching sequence. It should not be possible for these V-segments to show up in the data, since they shouldn't have been amplified in the PCR steps without a matching primer sequence. This indicates that the primers were able to amplify more V-segments than initially reported. There are also V-segments that were represented by the primers, but not found in the data. These segments might be rare, which could be an explanation as to why they are not found in the patient cohort.

Table 4.3 contains information about the J-segments covered by the EuroClonality primers, and the J-segments found in the sequencing output. As seen in the table, there are no mismatches between the primer sequences and the segments found in the data.

Primers	Sequencing output
TRBV2	TRBV2
TRBV3-1	TRBV3-1
TRBV3-2	
TRBV4-1	TRBV4-1
TRBV4-2	TRBV4-2
TRBV4-3	TRBV4-3
TRBV5-1	TRBV5-1
TRBV5-3	
TRBV5-4	TRBV5-4
TRBV5-5	TRBV5-5
TRBV5-6	TRBV5-6
TRBV5-7*01	
TRBV5-8	TRBV5-8
	TRBV6-1
TRBV6-2	TRBV6-2
	TRBV6-3
TRBV6-4	TRBV6-4
	TRBV6-5
TRBV6-6	TRBV6-6
TRBV6-7	
	TRBV6-9
	TRBV7-2
TRBV7-3	TRBV7-3
	TRBV7-4
TRBV7-5	
	TRBV7-6
TRBV7-7	TRBV7-7
TRBV7-8	TRBV7-8
TRBV7-9	TRBV7-9
TRBV9	TRBV9
	TRBV10-1
TRBV10-2	TRBV10-2
TRBV10-3	TRBV10-3
TRBV11-1	TRBV11-1
	TRBV11-2
	TRBV11-3
TRBV12-3	TRBV12-3
TRBV12-4	TRBV12-4
TRBV12-5	TRBV12-5
TRBV13	TRBV13
TRBV14	TRBV14
TRBV15	TRBV15
TRBV16	TRBV16
TRBV18	TRBV18

TRBV19	TRBV19
TRBV20-1	TRBV20-1
TRBV21-1	
TRBV23-1	
TRBV24-1	TRBV24-1
TRBV25-1	TRBV25-1
TRBV27	TRBV27
TRBV28	TRBV28
TRBV29-1	TRBV29-1
TRBV30	TRBV30

Table 4.2: Comparison of V-segments represented by primers, and V-segments found in sequencing output. Annotations TRBVX-Y*Z: TRBV, T-cell Receptor Beta V-segment; X, subgroup of V-segment; Y, subgroup gene; Z, allele. V-segments marked in red indicate a lack of matching primer sequence.

Primers	**Sequencing output**
TRBJ1-1	TRBJ1-1
TRBJ1-2	TRBJ1-2
TRBJ1-3	TRBJ1-3
TRBJ1-4	TRBJ1-4
TRBJ1-5	TRBJ1-5
TRBJ1-6	TRBJ1-6
TRBJ2-1	TRBJ2-1
TRBJ2-2	TRBJ2-2
TRBJ2-3	TRBJ2-3
TRBJ2-4	TRBJ2-4
TRBJ2-5	TRBJ2-5
TRBJ2-6*01,*02	TRBJ2-6
TRBJ2-7	TRBJ2-7

Table 4.3: Comparison of J-segments represented by primers, and J-segments found in sequencing output. Annotations TRBJX-Y*Z: TRBJ, T-cell Receptor Beta-chain J-segment; X, subgroup of J-segment; Y, subgroup gene; Z, allele.

4.1.2 V-segment presence

Figure 4.1 shows the hierarchical clustering of the V-segments for all patients at all time points, and the corresponding frequencies. The x-axis shows the samples for each patient and time point and the y-axis shows the V-segments found in the sequencing data. The heatmap shows the frequency of each V-segment in the samples, and the dendrogram groups both samples and V-segments based on similarity. As expected, the general trend is that the time points from the same patients cluster together, meaning that the variation in V-segment frequency varies more between patients than between time points in the same patient. This would be expected due to the highly personalised nature of the TCR-beta repertoire. The patient samples did not necessarily cluster in chronological time point order, reflecting the dynamic variations in V-segment frequency during immunotherapy treatment. However, a few patient time points broke away from their patient clusters, although they did not show any specific increased frequencies for any V-segments, indicating an increase in the diversity of the TCR-beta repertoire rather than increased clonality.

The large darker band corresponding to low frequency close to zero, shows the V-segments that do not seem to be used during T-cell response to treatment. The V-segments with the highest frequencies, marked in orange and red, seem to be patient-specific, only occurring in high frequency in one patient, while being low in the rest. Towards the bottom of the heatmap are the V-segments that are used in frequencies closer to 0.1, which can be interpreted as the ones most affected during immunotherapy treatment. There seems to be no apparent clustering between responders and non-responders, again indicating the patient-specific nature of the TCR-beta repertoire.

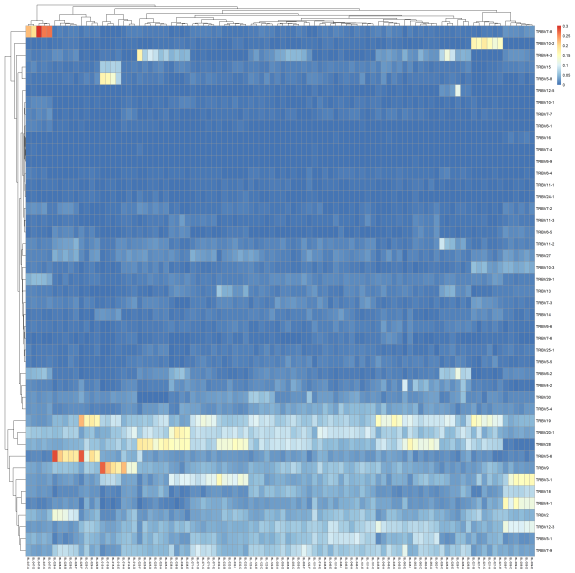


Figure 4.1: Hierarchical clustering heatmap over all TRBV-segments included in the sequencing data from the patient cohort.

4.2 T-cell diversity

As mentioned, there is a connection between T-cell diversity and response to treatment [35]. It has been found that patients that have a high number of unique T-cell clonotypes (richness) have responded well to treatment. It is also said that patients with a more even distribution of clones over the different clonotypes (evenness) respond better to treatment as well.

4.2.1 Number of unique clonotypes

For the patient cohort in this study, there is a clear connection between the number of clonotypes and response to treatment. This can be seen in Figure 4.2, where the non-responders are marked in red and the responders are marked in blue. The trend for responders is that they have a much higher amount of unique T-cell clonotypes already from the start of treatment. Table 4.4 contains the differences in the median values between responders and non-responders for each time point. The values were calculated as:

$$Difference^{i}(\%) = [median^{i}_{responder} - median^{i}_{non-responder}]/median^{i}_{responder} \quad (4.1)$$

where i corresponds to a time point A through F/1 year. The far right column shows the difference in median for all samples from all time points. The general trend for both non-responders and responders is that the median number of clonotypes increases from start point A (before treatment). However, it increases until approximately time point D for non-responders (9 weeks after start of treatment), before it decreases to point E (12 weeks after start of treatment). For the time point of progress (F), the median number of clonotypes has decreased to a point below the starting point.

For responders, the median number of clonotypes increases slightly from the starting point (A) until sample point E (12 weeks after treatment start). For the sample after 1 year, the trend is the same as for non-responders, namely that the number of clonotypes significantly drops to a level below the starting point. Compared to the non-responder group, the responders have a 33% higher median number of clonotypes overall, and a 55% higher median value at baseline. These results are in line with the literature [35][19], which reports responders having a higher number of unique clonotypes at baseline. The better response might be since a higher number of unique clonotypes at the start of treatment can be activated by ICI therapy and be active towards tumour cells.

When considering individual plots of the number of clonotypes for each patient, Figure 4.3 to 4.5 for responders and non-responders, there is no apparent correlation between number of unique clonotypes and response to treatment. This indicates that this measure is not sufficient on its own to predict response to treatment, and highlights the need to combine with other measures.

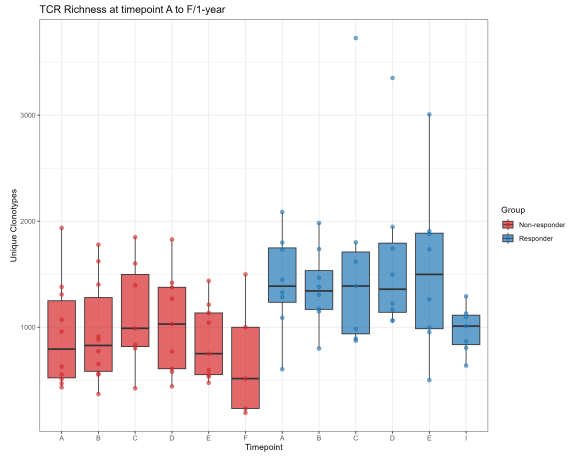
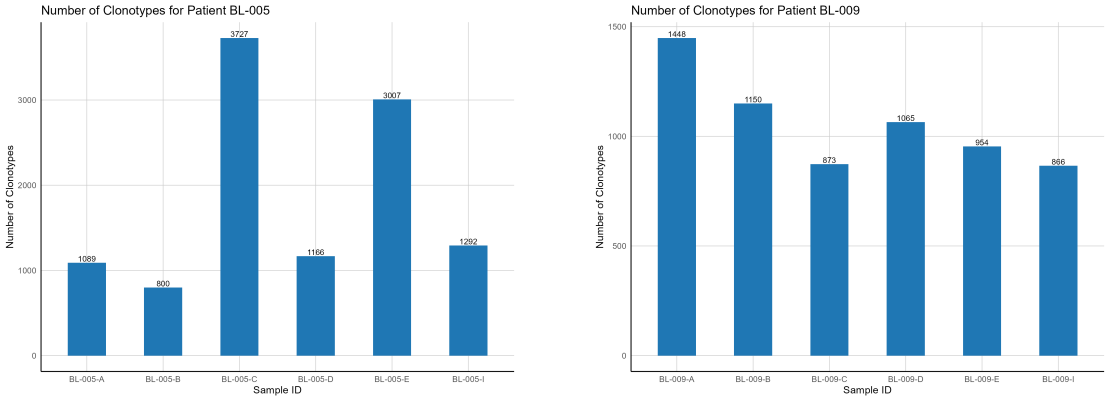


Figure 4.2: Number of unique clonotypes in non-responders vs responders.

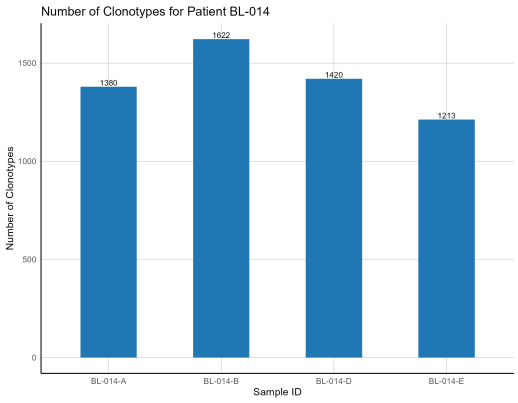
Unit %-points	**A**	**B**	**C**	**D**	**E**	**F/1 year**	**Total**
Non-responders	628	829	990	1030	751	516	859
Responders	1388	1343	1389	1359	1499	1011	1288
Difference	55%	38%	29%	24%	50%	49%	33%

Table 4.4: Difference in median values of the D50 index expressed in percentage. Compared for time points A through F/1 year for responders and non-responders, and the difference in median values across all time points.

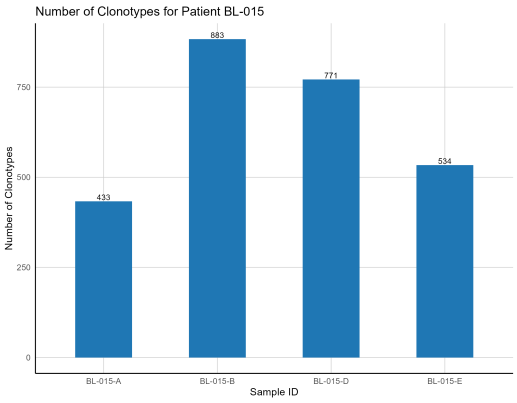


BL-005 (Responder)

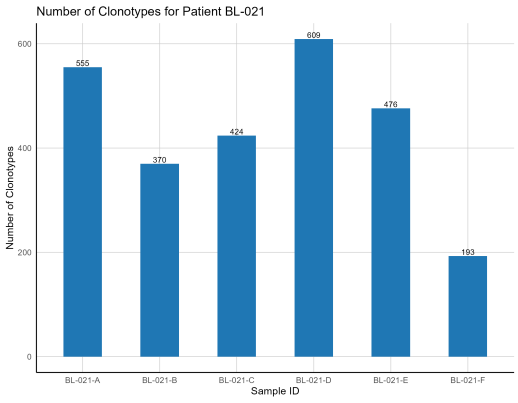
BL-009 (Responder)



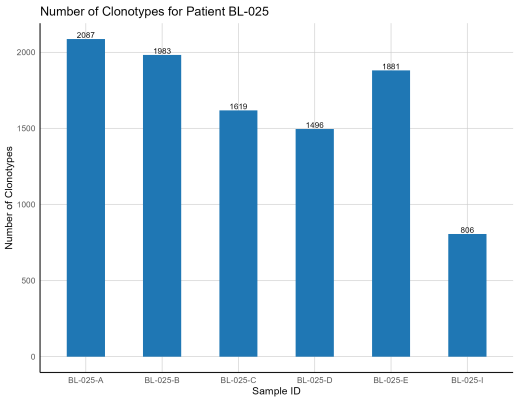
BL-014 (Non-responder)



BL-015 (Non-responder)

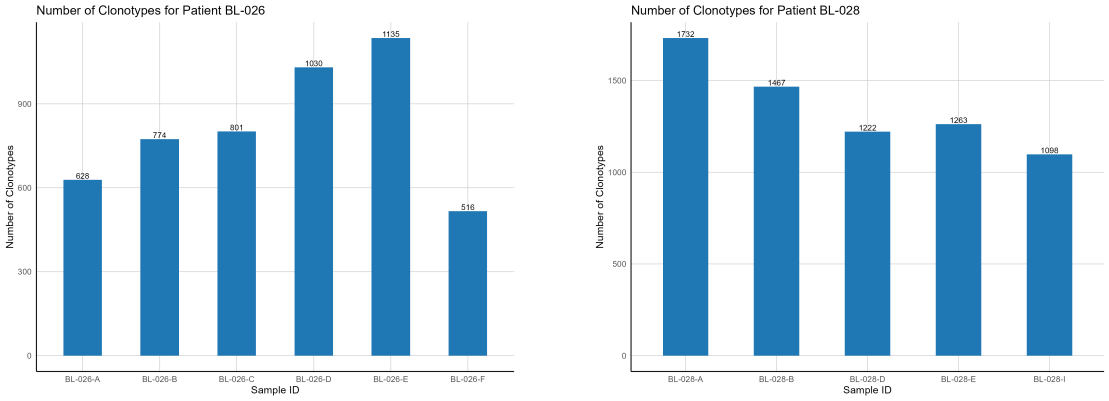


BL-021 (Non-responder)



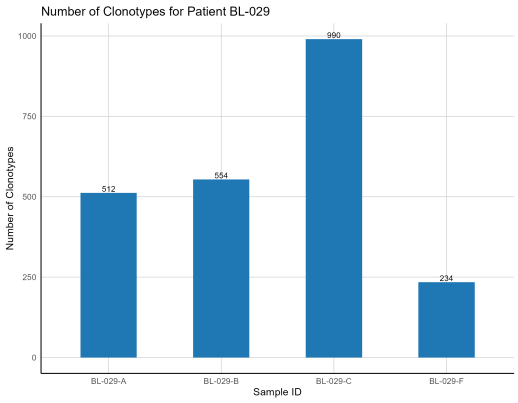
BL-025 (Responder)

Figure 4.3: Number of unique clonotypes in patients BL-005 to BL-025

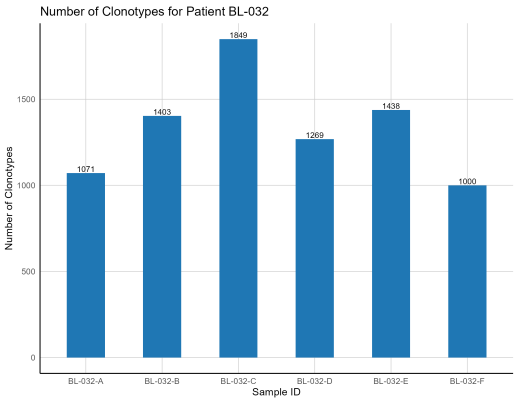


BL-026 (Non-responder)

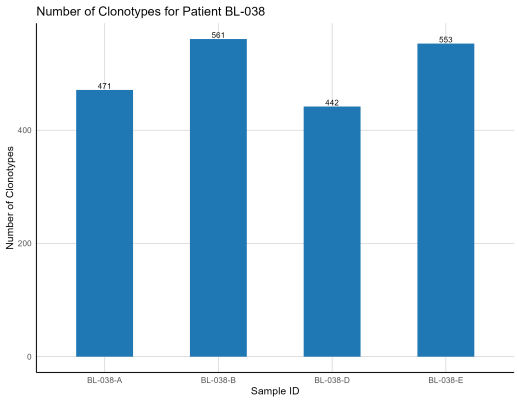
BL-028 (Responder)



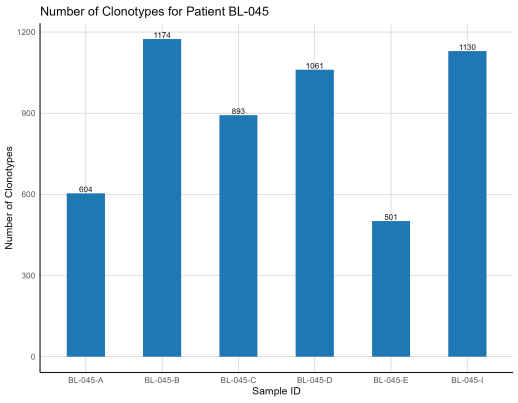
BL-029 (Non-responder)



BL-032 (Non-responder)

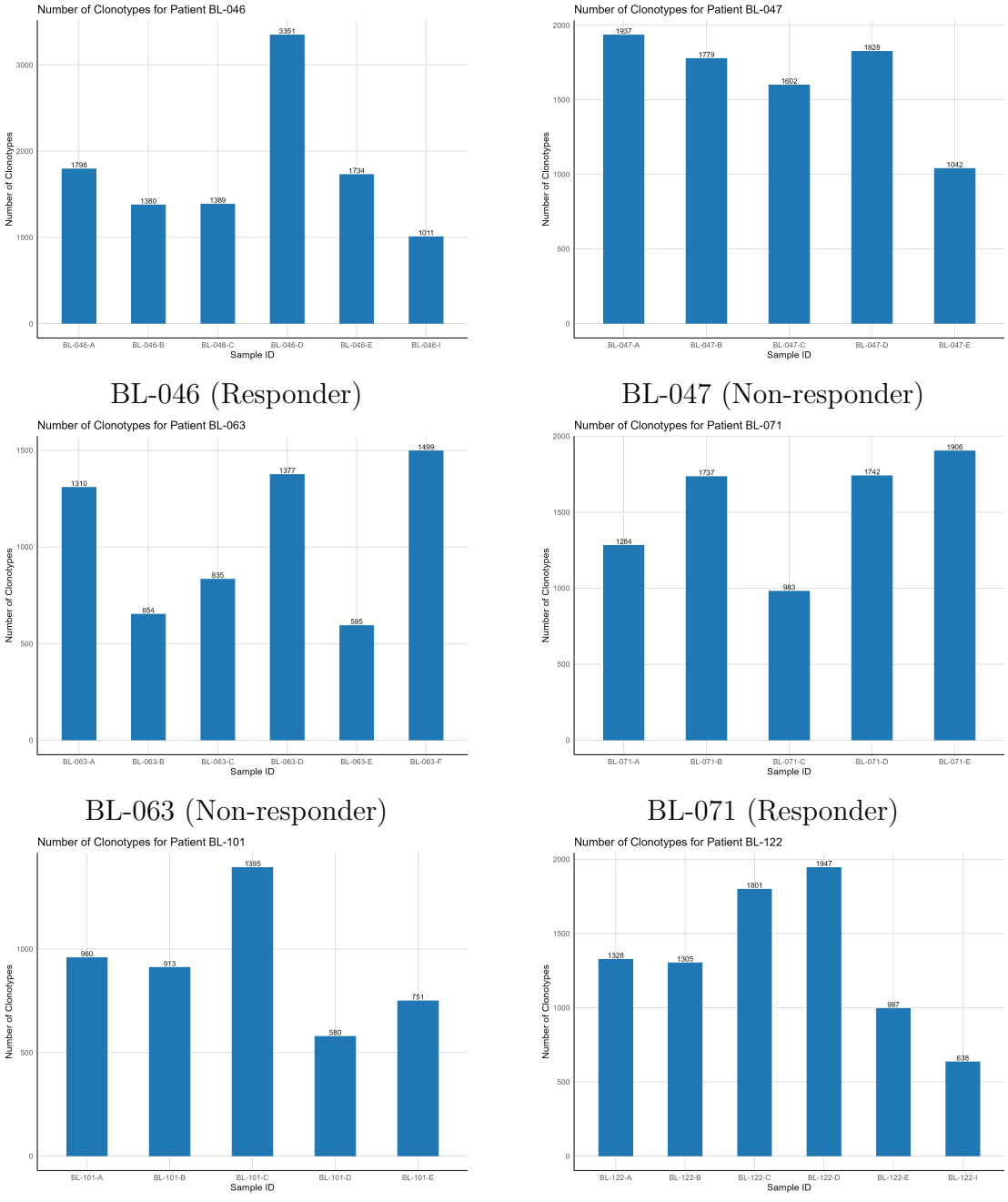


BL-038 (Non-responder)



BL-045 (Responder)

Figure 4.4: Number of unique clonotypes in patients BL-026 to BL-045



BL-046 (Responder)

BL-047 (Non-responder)

BL-063 (Non-responder)

BL-071 (Responder)

BL-101 (Non-responder)

BL-122 (Responder)

Figure 4.5: Number of unique clonotypes in patients BL-046 to BL-122.

4.2.2 D50 index

Another metric for diversity is the evenness of the repertoire, which was measured and analysed using the D50 index. Figure 4.6 shows the D50 index for non-responders and responders across the sample time points. A higher D50 index has been correlated with patient response to immunotherapy treatment [35], and this trend can also clearly be seen in the Figure. The median of the D50 index across all samples for responders is 8 percentage points higher than for non-responders, as seen in Table 4.5. Before start of treatment at time point A, the responders have a D50 index median that is 12 percentage points higher than non-responders. The D50 median increases from time point A to B for non-responders and remains at a higher level than baseline for all time points. This indicates a more diverse T-cell repertoire, but since these patients are non-responders, the increased clonotypes might be non-specific to tumour response. On the other hand, the responders have a median D50 index that decreases from time point A to the 1 year point. The decrease in diversity might indicate an increase in tumour specific T-cell clones, which would be an expected response to treatment.

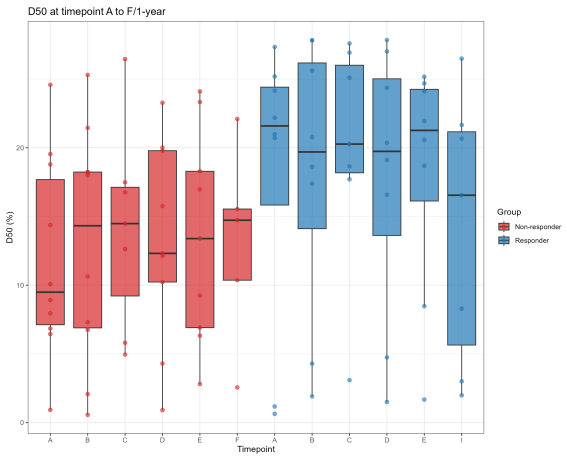


Figure 4.6: D50 index for non-responders vs responders.

	A	B	C	D	E	F/1 year	Total
Non-responders	9.49	14.33	14.48	12.32	13.39	14.73	13.01
Responders	21.59	19.70	19.46	19.74	21.27	16.55	20.62
Difference*	12	5	5	7	8	2	8

Table 4.5: Difference in median values of the D50 index expressed in percentage. Compared for time points A through F/1 year for responders and non-responders, and the difference in median values across all time points. * The units are expressed as percentage points.

4.2.3 Shannon index

A third diversity index is the Shannon index, which considers both richness and evenness at the same time. An increase of the Shannon index means that the

repertoire consists of more evenly distributed clones among all the clonotypes. As seen in Figure 4.7, it follows the same trend as for the richness and D50 measures, namely that the responders have higher median values than non-responders, also seen in Table 4.6. At time point A, responders have a 10% higher median value than non-responders, and a total median value that is 9% higher. The higher Shannon index corresponds with higher T-cell repertoire diversity, which has been correlated to better response to ICI treatment [19]. For responders, the index is steady at values close to the value at time point A, before it decreases for the 1 year point. This is an indication of a more stable, diverse repertoire for responders during the initial phase of treatment. The lower value after 1 year is a sign of decreased diversity, which might indicate an expansion of clonotypes engaged in the long-term anti-tumour response. For non-responders, the index fluctuates more around the value at time point A, which might indicate a less diverse repertoire.

Together, the number of unique clonotypes, and D50 and Shannon indices describe different parameters of diversity of the T-cell repertoire in the patients. The results indicate that a higher diversity of the repertoire is associated with a better response to anti-PD-1/PD-L1 treatment, which is also supported by the literature.

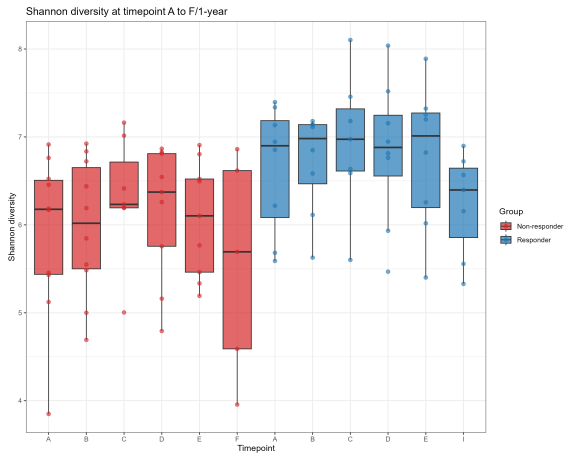


Figure 4.7: Shannon index for non-responders vs responders.

	A	B	C	D	E	F/1 year	Total
Non-responders	6.177	6.018	6.233	6.373	6.102	5.693	6.193
Responders	6.899	6.981	6.974	6.880	7.012	6.397	6.836
Difference	10 %	14%	11%	7%	13%	11%	9%

Table 4.6: Difference in median values of the Shannon index expressed in percentage. Compared for time points A through F/1 year for responders and non-responders, and the difference in median values across all time points.

4.3 TCR clonal expansion

Patients with increased clonal expansion of T-cells from pre-treatment to post-treatment have been reported to respond well to anti-PD-1/PD-L1 treatment [19].

This would be expected since a re-activation of T-cells by the therapy facilitates tumour cell recognition and an immune response, when the T-cells proliferate abundantly. The frequencies of the top 10 most variable clonotypes are tracked across all time points for each patient and visualised and included in Appendix 2. Figures B.1 to Figure B.8 are responders and Figure B.9 to Figure B.18 are non-responders. When considering the clonal expansions for both responders and non-responders, there seems to be no apparent pattern, pointing to a highly individualised immune response for the patients. Some responders such as Figures B.2, B.4, and B.7 seem to have little expansion, with the clonotypes having roughly the same frequency across all time points. For others such as patients in Figures B.1, B.3, B.5, B.6, and B.8, there seems to be a pattern of low frequency for the initial time points, and an increase in frequency for the later time points, indicating immune response. However this can be explained by variation in when the patients responds to treatment, since not all respond after the initial treatment cycle. Interestingly, patient BL-009, Figure B.2, was classified as a complete responder to treatment across all time points, and does not show any clonal expansion, but rather a uniform distribution already from before start of treatment. Since all 10 clonotypes are represented in all time points, it indicates a diverse repertoire, which would be in line with previous findings, namely that responders have a high diversity.

For non-responders, there are similar patterns of patients with low clonal expansion, as seen in Figures B.9, B.10, B.11, B.13, B.14, B.15, B.16, and B.17. The frequencies of the clonotypes can vary, either decreasing over time or staying roughly the same. This would be a more expected result for non-responders since it would indicate low immune response to treatment. Figures B.12, and B.18 have a greater expansion of certain clones over time. However, since not all clonotypes are represented for each time point, it indicates a lower diversity, which again would be consistent with previous results that connect non-responders with low repertoire diversity.

These ambiguous results for the individual patients in the cohort show that tracking specific clonotypes in an individual is not a sufficient predictor for response to treatment. Rather, a broader perspective over the entire T-cell repertoire would be better to consider.

4.4 Future outlooks

While this study provides valuable insights into how the T-cell repertoire changes during immunotherapy, more research is needed to use these metrics in clinical practice. First, future studies should include a larger number of NSCLC patients. Expanding the patient cohort is necessary to confirm the patterns observed in the Shannon and D50 indices, and the number of clonotypes, and to ensure the results apply to a broader population.

Furthermore, it would be highly valuable to compare blood samples with tissue biopsies from the tumour microenvironment (TME). By comparing tumour-infiltrating lymphocytes (TILs) directly with T-cells in the blood, the exact clonotypes that are

actively fighting the tumour could be identified. They can then be tracked in the blood over time to see for e.g. clonal expansion. If responding patients share the same active clonotypes, clinical tests could focus strictly on tracking these specific clones, which would remove a lot of background noise from the data.

Additionally, future research should correlate TCR data with other health conditions or diseases that patients have at the same time as their cancer. Since T-cells expand and change in response to many different illnesses, knowing the patients' full medical history is essential to separate tumour responses from unrelated immune events in the blood. Ultimately, more comprehensive research is still needed to develop reliable biomarkers that can accurately predict response to anti-PD-1/PD-L1 therapy.

5
Conclusion

This study investigated the relationship between T-cell receptor repertoire characteristics and response to anti-PD-1/PD-L1 immunotherapy in patients with non-small cell lung cancer (NSCLC).

The results demonstrate that patients who responded to treatment generally exhibited higher T-cell repertoire diversity than non-responders. Responders showed a greater number of unique clonotypes as well as higher D50 and Shannon diversity indices throughout the treatment period. These findings are consistent with previous studies suggesting that a diverse T-cell repertoire might enhance the ability of the immune system to recognize and respond to tumour-associated antigens.

In contrast, analysis of clonal expansion patterns revealed substantial inter-patient variability and no consistent differences between responders and non-responders. This suggests that monitoring individual clonotypes alone is unlikely to provide a robust predictor of treatment outcome. Instead, global measures of repertoire diversity appear to provide more informative indicators of therapeutic response.

Overall, the findings support the potential of TCR repertoire diversity as a biomarker for immunotherapy response in NSCLC. However, the limited cohort size and the heterogeneity of patient responses highlight the need for larger validation studies. Future work should integrate TCR sequencing with tumour tissue analysis and additional clinical parameters to further evaluate the utility of repertoire-based biomarkers in precision medicine.

Bibliography

[1] Meredith L. Frank, Kaylene Lu, Can Erdogan, Yi Han, Jian Hu, Tao Wang, John V. Heymach, Jianjun Zhang, and Alexandre Reuben. T-cell receptor repertoire sequencing in the era of cancer immunotherapy. Clinical Cancer Research, 27(12):3325–3339, 2021.

[2] Kroopa Joshi, Martina Milighetti, and Benjamin M. Chain. Application of t cell receptor (tcr) repertoire analysis for the advancement of cancer immunotherapy. Current Opinion in Immunology, 74:1–8, 2022.

[3] Daniel Andersson, Firaol T. Kabede, Maria Escobar, Tobias Österlund, and Anders Ståhlberg. Principles of digital sequencing using unique molecular identifiers. Molecular Aspects of Medicine, 96, 2024.

[4] Lizza E. L. Hendriks, Jordi Remon, Corinne Faivre-Finn, Marina C. Garassino, John V. Heymach, Keith M. Kerr, Daniel S. W. Tan, Giulia Veronesi, and Martin Reck. Non-small-cell lung cancer. Nature Reviews Disease Primers, 10(1):71, 2024.

[5] Hongshu Sui, Ningxia Ma, Ying Wang, Hui Li, Xiaoming Liu, Yanping Su, and Jiali Yang. Anti-pd-1/pd-l1 therapy for non-small-cell lung cancer: Toward personalized medicine and combination strategies. Journal of Immunology Research, (1), August 2018.

[6] M. Attaf, M. Legut, D. K. Cole, and A. K. Sewell. The t cell antigen receptor: the swiss army knife of the immune system. Clinical and Experimental Immunology, 181(1):1–18, 2015.

[7] Elsevier. Alpha beta t cell. `https://www.sciencedirect.com/topics/immunology-and-microbiology/alpha-beta-t-cell`, 2014. Topic overview on alpha/beta T cells from ScienceDirect Topics.

[8] British Society for Immunology. Gamma delta ($\gamma\delta$) t cells. `https://www.immunology.org/public-information/bitesized-immunology/cells/gamma-delta-gd-t-cells`, 2022. Definition and overview of T cells from BiteSized Immunology.

[9] Bing-Yen Wang, Jing-Yang Huang, Heng-Chung Chen, Ching-Hsiung Lin, Sheng-Hao Lin, Wei-Heng Hung, and Ya-Fu Cheng. The comparison between adenocarcinoma and squamous cell carcinoma in lung cancer patients. Journal of Cancer Research and Clinical Oncology, 146(1):43–52, 2020.

[10] Raj Shah, Senthil Sabanathan, Alan J. Mearns, and John Richardson. Non-small cell lung cancer (nsclc): a review of risk factors, diagnosis, and treatment. Medicine, 102(8):e32750, 2023.

[11] Venkatkiran Kanchustambham and Sanjeev Sharma. Non-small cell lung cancer (NSCLC): a review of risk factors, diagnosis, and treatment. StatPearls Publishing, Treasure Island, FL, 2026. Accessed: 2026-05-24.

[12] Alex Friedlaender, Maurice Perol, Giuseppe Luigi Banna, Kaushal Parikh, and Alfredo Addeo. Oncogenic alterations in advanced nsclc: a molecular superhighway. Biomarker Research, 12(1):24, 2024.

[13] Renwang Liu, Yiming Meng, and Mohamed Rahouma. Editorial: Advancing nsclc treatment: overcoming challenges in immune checkpoint inhibitor therapy. Frontiers in Pharmacology, 17:1744580, 2026.

[14] Md Mohiuddin. Anti-pd-1/pd-l1 immunotherapy as a potential treatment option for lung cancer: A perspective analysis of opportunities and challenges. Health Science Reports, 9(1):e71749, 2026.

[15] Andreas Hallqvist, Anna Rohlin, and Sukanya Raghavan. Immune checkpoint blockade and biomarkers of clinical response in non–small cell lung cancer. Scandinavian Journal of Immunology, 92(6):e12980, 2020.

[16] Martin Reck, Jordi Remon, and Matthew D. Hellmann. First-line immunotherapy for non-small-cell lung cancer. Journal of Clinical Oncology, 40(6):586–597, 2022.

[17] Eleonora Gariazzo, Francesca Colamartini, Martina Ubaldi, Valentina Santo, Leonardo Brunetti, Chiara Tomarelli, Daniele Ognissanti, Jouana Nassar, Silvia Costabile, Luca Romano, Emanuela De Vita, Ilenia Scorpiniti, Miriam Macrì, Roberta Porreca, Maria Francesca Currà, Alessio Cortellini, Biagio Ricciuti, and Giulio Metro. Emerging predictive biomarkers of immunotherapy sensitivity in patients with non-small cell lung cancer. ImmunoTargets and Therapy, 15:567238, 2026.

[18] Ella A. Eklund, Johanna Svensson, Louise S. Näslund, Maria Yhr, Sama I. Sayin, Clotilde Wiel, Levent M. Akjürek, Per Torstensson, Volkan I. Sayin, Andreas Hallqvist, Sukanya Raghavan, and Anna Rohlin. Comprehensive genetic variant analysis reveals combination of kras and lrp1b as a predictive biomarker of response to immunotherapy in patients with non-small cell lung cancer. Journal of Experimental Clinical Cancer, 44(75), 2025.

[19] Jiefei Han, Duan Jianchun, Hua Bai, Yuqi Wang, Rui Wan, Xin Wang, Si Chen, Yanhua Tian, Di Wang, Kailun Fei, Zhuoran Yao, Shuhang Wang, Zhimin Lu, Zhijie Wang, and Jie Wang. Tcr repertoire diversity of peripheral pd-1+cd8+ t cells predicts clinical outcomes after immunotherapy in patients with non–small cell lung cancer. Cancer Immunology Research, 8:146–154, 2020.

[20] National Cancer Institute. Tcr. `https://www.cancer.gov/publications/dictionaries/cancer-terms/def/tcr`, 2026. Definition of TCR (T-cell receptor) from the NCI Dictionary of Cancer Terms.

[21] Yanan Wu, Nianzhi Zhang, Keiichiro Hashimoto, Chun Xia, and Johannes M. Dijkstra. Structural comparison between mhc classes i and ii; in evolution, a class-ii-like molecule probably came first. Frontiers in Immunology, 12:621153, 2021.

[22] National Cancer Institute. Peripheral t-cell lymphoma, 2026. Definition of peripheral T-cell lymphoma from the NCI Dictionary of Cancer Terms.

[23] Justyna Mika, Alicja Polanska, Kim R. M. Blenman, Lajos Pusztai, Joanna Polanska, Serge Candéias, and Michal Marczyk. A comprehensive evaluation of diversity measures for tcr repertoire profiling. BMC Biology, 23(1):133, 2025.

[24] Ryan Bromley, Sandra Vranic, Debra Sinner, Sarah DiGiulio, Beth Hall, Laura Rassenti, Thomas Kipps, Victor Adleff, Dana Chudova, and Richard B. Lanman. Ep16.01-030 peripheral t-cell receptor repertoire profiling in non-small cell lung cancer using an amplicon-based sequencing assay. Journal of Thoracic Oncology, 17(9, Supplement):S925, 2022.

[25] Johanna Chiffelle, Raphael Genolet, Marta A. S. Perez, George Coukos, Vincent Zoete, and Alexandre Harari. T-cell repertoire analysis and metrics of diversity and clonality. Current Opinion in Biotechnology, 65:284–295, 2020.

[26] Anders Ståhlberg, Paul M. Krzyzanowski, Matthew Egyud, Stefan Filges, Lincoln Stein, and Tony E. Godfrey. Simple, multiplexed, pcr-based barcoding of dna for ultrasensitive mutation detection by next-generation sequencing. Nature Protocols, 12(4):664–682, 2017.

[27] Mikhail Shugay, Olga V. Britanova, Ekaterina M. Merzlyak, Maria A. Turchaninova, I. Z. Mamedov, T. R. Tuganbaev, Dmitry A. Bolotin, Dmitry B. Staroverov, Ekaterina V. Putintseva, Karla Plevova, et al. Towards error-free profiling of immune repertoires. Nature Methods, 11(6):653–655, 2014.

[28] Nimrat Khehra, Inderbir S. Padda, and Muhammad Zubair. Polymerase Chain Reaction (PCR). StatPearls Publishing, Treasure Island, FL, 2025. StatPearls [Internet]. Last update July 7, 2025.

[29] Tobias Österlund, Stefan Filges, Gustav Johansson, and Anders Ståhlberg. Umierrorcorrect and umianalyzer: Software for consensus read generation, error correction, and visualization using unique molecular identifiers. Clinical Chemistry, 68(11):1425–1435, 2022.

[30] EuroClonality Consortium. Euroclonality official website. `https://euroclonality.org/`, 2026. Accessed: 2026-06-25.

[31] IMGT, the international ImMunoGeneTics information system. Imgt repertoire: Gene table - trbv (homo sapiens). `https://www.imgt.org/IMGTrepertoire/index.php?section=LocusGenes&repertoire=genetable&species=human&group=TRBV`, 2026. Accessed: 2026-05-31.

[32] IMGT, the international ImMunoGeneTics information system. Imgt repertoire: Gene table - trbj (homo sapiens). `https://www.imgt.org/IMGTrepertoire/index.php?section=LocusGenes&repertoire=genetable&species=human&group=TRBJ`, 2026. Accessed: 2026-05-31.

[33] Illumina. Nextseq 550 system specification sheet. `https://emea.illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/nextseq-550-system-spec-sheet-m-gl-01298/nextseq-550-system-spec-sheet-m-gl-01298.pdf`, 2025. Accessed: 2026-05-26.

[34] Illumina. Optimal cluster density. `https://support-docs.illumina.com/SHARE/ClusterOptimize/Content/SHARE/ClusterOptimize/OptimalClusterDensity.htm`, 2021. Accessed: 2026-05-26.

[35] Mehmet Altan, Ruoxing Li, Ziyi Li, Runzhe Chen, Ajay Sheshadri, Hai T. Tran, Latasha Little, Joshua Baguley, Jefferson Sinson, Natalie Vokes, Saumil

Gandhi, Mara B. Antonoff, Stephen G. Swisher, Greg Lizee, Alexandre Reuben, John V. Heymach, and Jianjun Zhang. High peripheral t cell diversity is associated with lower risk of toxicity and superior response to dual immune checkpoint inhibitor therapy in patients with metastatic nsclc. Journal for Immunotherapy of Cancer, 12(12), 2024.

A

Appendix A

This Appendix contains a summary of the patient data from sequencing, including total number of clones, total number of unique clonotypes, as well as D50 and Shannon diversity indices. The quality results from the two sequencing runs are also displayed.

Patient	Timepoint	Clones	Clonotypes	D50 index	Shannon index
BL-005	A	1818	1089	263	6.853
	B	1124	800	205	6.585
	C	4708	3727	935	8.103
	D	1569	1166	284	6.946
	E	4152	3007	725	7.889
	I	1878	1292	267	6.898
BL-009	A	3919	1448	17	5.590
	B	2763	1150	22	5.628
	C	1964	873	27	5.601
	D	2654	1065	16	5.468
	E	2430	954	16	5.401
	I	1967	866	26	5.556
BL-014	A	3067	1380	139	6.184
	B	2537	1622	292	6.924
	D	2034	1420	281	6.866
	E	1627	1213	283	6.907
BL-015	A	2024	433	4	3.850
	B	2665	883	5	4.692
	D	2299	771	7	4.792
	E	1194	534	15	5.192
BL-021	A	1002	555	38	5.456
	B	802	370	25	5.000
	C	1011	424	21	5.004
	D	1046	609	75	5.756
	E	1134	476	44	5.334
	F	448	193	20	4.590
BL-025	A	2858	2087	438	7.337
	B	2937	1983	369	7.142
	C	2309	1619	302	6.974
	D	2135	1496	248	6.814
	E	2490	1881	387	7.201
	I	2023	806	16	5.329
BL-026	A	1052	628	118	6.170
	B	1205	774	166	6.438
	C	1621	801	116	6.192
	D	1824	1030	125	6.373
	E	2098	1135	152	6.495
	F	838	516	76	5.693
BL-028	A	3794	1732	11	5.682
	B	3278	1467	63	6.113
	D	2571	1222	58	5.933
	E	2203	1263	107	6.256
	I	1878	1098	91	6.155
BL-029	A	2021	512	33	5.123

	B	1705	554	59	5.484
	C	1697	990	125	6.233
	F	1350	234	6	3.956
BL-032	A	1597	1071	154	6.456
	B	2128	1403	256	6.835
	C	3023	1849	310	7.014
	D	1887	1269	254	6.810
	E	2315	1438	263	6.804
	F	1653	1000	221	6.617
BL-038	A	1039	471	42	5.431
	B	1139	561	41	5.549
	D	1118	442	19	5.161
	E	1390	553	35	5.462
BL-045	A	906	604	134	6.217
	B	1694	1174	244	6.849
	C	1273	893	181	6.588
	D	1510	1061	216	6.763
	E	768	501	110	6.018
	I	1640	1130	187	6.567
BL-046	A	2276	1798	453	7.395
	B	1587	1380	384	7.180
	C	1614	1389	374	7.181
	D	4102	3351	905	8.038
	E	2176	1734	428	7.322
	I	1486	1011	219	6.722
BL-047	A	3372	1937	154	6.524
	B	3910	1779	37	5.846
	C	2773	1602	93	6.195
	D	3433	1828	187	6.545
	E	2146	1042	72	5.767
BL-063	A	1779	1310	256	6.914
	B	908	654	119	6.191
	C	1164	835	146	6.414
	D	1942	1377	217	6.825
	E	813	595	101	6.102
	F	2222	1499	233	6.861
BL-071	A	1847	1284	266	6.945
	B	2501	1737	302	7.139
	C	1379	983	174	6.636
	D	2783	1742	333	7.156
	E	2753	1906	356	7.256
BL-101	A	1299	960	236	6.761
	B	1178	913	231	6.723
	C	1671	1395	369	7.163
	D	763	580	135	6.259
	E	982	751	181	6.522

BL-122	A	1563	1328	363	7.135
	B	1533	1305	363	7.114
	C	2013	1801	497	7.457
	D	2288	1947	542	7.519
	E	1276	997	251	6.822
	I	759	638	169	6.397

Table A.1: A summary of data for all patients in the cohort, divided by the time point for sampling.

Sample Name	% Duplicates (fw/rev)	% Failed (fw/rev)	M Sequences
BL-005-A	91.6 / 94.0	45 / 45	3.2
BL-005-B	93.4 / 94.2	45 / 36	4.6
BL-005-C	91.7 / 96.0	18 / 36	2.7
BL-005-D	91.6 / 93.9	36 / 36	2.6
BL-005-E	90.6 / 94.9	36 / 45	2.9
BL-005-F	92.1 / 95.7	18 / 36	2.0
BL-021-D	92.6 / 94.2	45 / 45	3.2
BL-021-E	92.5 / 93.9	45 / 45	4.6
BL-026-B	91.9 / 93.0	45 / 36	2.6
BL-026-C	91.6 / 93.9	45 / 45	3.6
BL-026-D	92.7 / 94.8	45 / 45	4.0
BL-026-E	92.3 / 94.7	45 / 45	4.2
BL-026-F	94.1 / 95.7	27 / 45	2.7
BL-029-A	91.9 / 93.2	45 / 45	4.7
BL-029-B	93.5 / 93.8	45 / 45	5.0
BL-029-C	92.4 / 94.7	45 / 45	3.6
BL-029-F	92.6 / 93.5	45 / 45	4.6
BL-038-A	93.1 / 94.2	36 / 45	5.9
BL-038-B	92.1 / 93.4	36 / 45	2.7
BL-038-D	93.1 / 93.8	36 / 45	5.1
BL-038-E	92.6 / 93.9	45 / 45	4.5
BL-045-A	93.2 / 95.1	27 / 45	1.8
BL-045-B	92.7 / 95.5	18 / 36	1.9
BL-045-C	93.7 / 95.5	18 / 36	2.2
BL-045-D	93.6 / 95.8	18 / 36	2.7
BL-045-E	95.0 / 95.8	27 / 45	2.3
BL-045-I	93.2 / 95.7	18 / 36	2.4
BL-046-A	92.9 / 95.5	18 / 36	2.7
BL-046-B	93.8 / 95.8	18 / 36	2.5
BL-046-C	93.1 / 95.1	18 / 36	2.2
BL-046-D	92.0 / 95.9	18 / 36	3.4
BL-046-E	93.0 / 95.5	18 / 36	2.2
BL-046-I	94.4 / 95.5	18 / 36	3.9
BL-063-A	91.7 / 94.9	18 / 36	1.5
BL-063-B	94.2 / 95.3	27 / 45	2.1
BL-063-C	93.9 / 94.9	18 / 36	1.9
BL-063-D	93.7 / 96.1	18 / 36	2.0
BL-063-E	94.6 / 95.5	18 / 36	1.5
BL-101-B	93.0 / 95.3	18 / 36	1.8
BL-101-C	93.5 / 95.7	18 / 36	1.9
BL-101-D	94.4 / 94.9	27 / 45	2.2
BL-101-E	94.3 / 94.9	27 / 45	2.4
BL-122-A	93.1 / 95.1	18 / 36	2.2
BL-122-B	93.6 / 95.4	18 / 36	2.4
BL-122-C	93.0 / 95.1	18 / 36	2.8
BL-122-D	93.7 / 96.0	18 / 36	3.3
BL-122-E	94.5 / 95.3	27 / 45	3.6
BL-122-F	94.3 / 94.0	36 / 45	3.4

Table A.2: QC report general info from sequencing run 1 of 48 samples.

Sample Name	% Duplicates (fw/rev)	% Failed (fw/rev)	M Sequences
BL-009-A	92.8 / 96.7	18 / 36	4.7
BL-009-B	92.8 / 96.5	18 / 36	3.0
BL-009-C	92.6 / 96.1	18 / 36	2.3
BL-009-D	92.8 / 96.6	27 / 36	2.5
BL-009-E	93.1 / 96.5	27 / 36	2.9
BL-009-I	93.5 / 96.8	18 / 36	3.1
BL-014-A	94.1 / 95.9	27 / 45	7.8
BL-014-B	92.6 / 96.5	18 / 36	2.7
BL-014-D	93.3 / 96.1	27 / 45	3.3
BL-014-E	94.1 / 96.4	27 / 45	2.9
BL-015-A	92.7 / 96.2	36 / 45	3.9
BL-015-B	92.1 / 96.5	27 / 36	2.6
BL-015-D	91.6 / 96.2	18 / 36	2.2
BL-015-E	93.4 / 96.5	27 / 45	2.3
BL-021-A	93.2 / 95.4	27 / 45	2.0
BL-021-B	93.9 / 95.7	27 / 36	2.3
BL-021-C	94.4 / 96.1	27 / 45	3.1
BL-021-F	94.4 / 95.4	36 / 45	2.6
BL-025-A	91.8 / 96.1	18 / 36	2.3
BL-025-B	92.4 / 96.3	18 / 36	2.6
BL-025-C	93.4 / 96.2	18 / 36	3.5
BL-025-D	93.1 / 96.5	18 / 36	2.2
BL-025-E	93.1 / 95.8	27 / 45	2.9
BL-025-F	92.8 / 96.5	18 / 36	2.2
BL-026-A	95.0 / 96.4	36 / 45	4.5
BL-028-A	91.3 / 96.1	18 / 36	1.9
BL-028-B	92.8 / 96.6	18 / 36	3.4
BL-028-D	93.9 / 94.8	18 / 36	3.3
BL-028-E	92.4 / 96.0	18 / 36	1.9
BL-028-F	92.5 / 96.0	18 / 36	1.7
BL-032-A	93.4 / 96.2	27 / 45	2.5
BL-032-B	94.0 / 96.4	18 / 36	3.3
BL-032-C	92.6 / 96.5	18 / 36	2.6
BL-032-D	93.2 / 96.0	27 / 45	2.2
BL-032-E	92.8 / 96.6	18 / 36	2.4
BL-032-I	93.2 / 96.1	27 / 45	3.5
BL-047-A	92.5 / 96.1	18 / 36	3.1
BL-047-B	92.2 / 97.0	18 / 36	2.7
BL-047-C	92.5 / 96.1	27 / 45	2.3
BL-047-D	93.1 / 96.9	18 / 36	3.8
BL-047-E	94.0 / 96.2	27 / 45	4.7
BL-063-F	93.6 / 96.3	18 / 36	2.3
BL-071-A	93.3 / 96.3	18 / 36	1.9
BL-071-B	92.5 / 96.5	18 / 36	2.0
BL-071-C	93.6 / 96.0	27 / 45	2.1
BL-071-D	92.7 / 96.6	18 / 36	2.8
BL-071-E	92.5 / 96.4	18 / 36	2.1
BL-101-A	94.0 / 95.4	27 / 45	3.1

Table A.3: QC report general info from run 2 of 48 samples.

B

Appendix B

This Appendix contains plots of the top 10 most variable clonotypes for each of the patients in the cohort, from the time points A to F (non-responders) and A to 1 year (responders). The numbers 1-10 in the legend refer to the most variable clonotypes across all sampling time points, with 1 being the most variable. Each number corresponds to a unique clonotype, i.e., a unique CDR3 sequence, and the same 10 clonotypes are tracked across all time points. When choosing the most variable clonotypes, the standard deviation of the abundance of each clonotype between the available time points was calculated. The 10 clonotypes with the highest standard deviation were chosen. The x-axis in the plots shows the sampling time points and the y-axis shows frequency of the clonotypes.

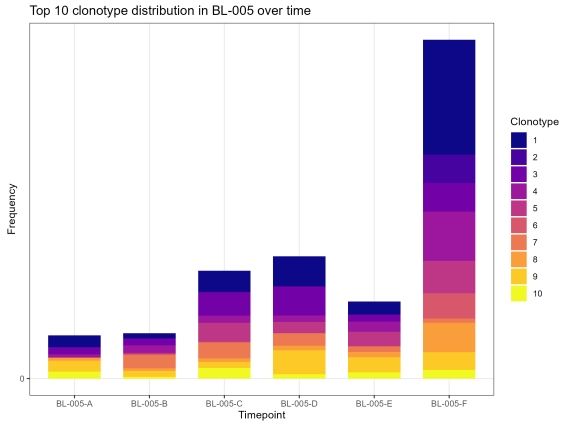


Figure B.1: Clonal expansion for patient BL-005 (responder)

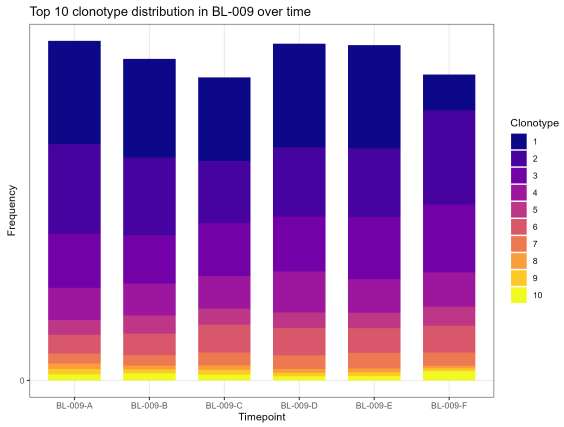


Figure B.2: Clonal expansion for patient BL-009 (responder)

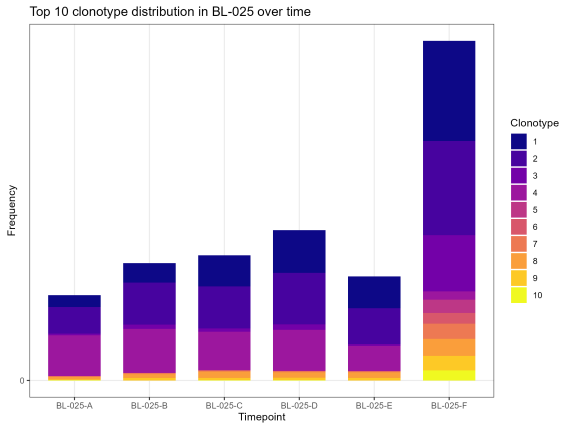


Figure B.3: Clonal expansion for patient BL-025 (responder)

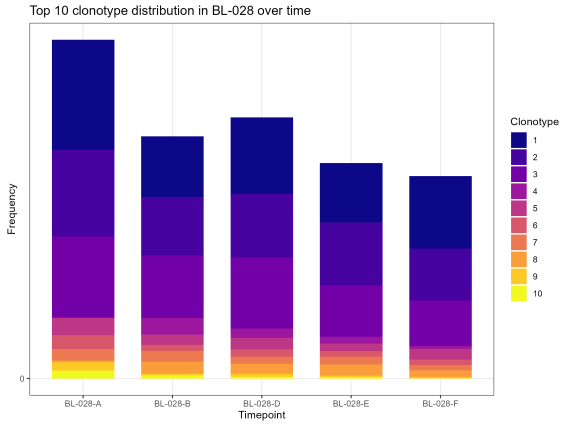


Figure B.4: Clonal expansion for patient BL-028 (responder)

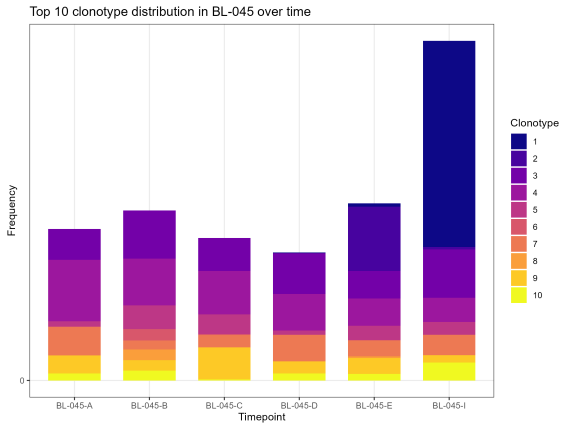


Figure B.5: Clonal expansion for patient BL-045 (responder)

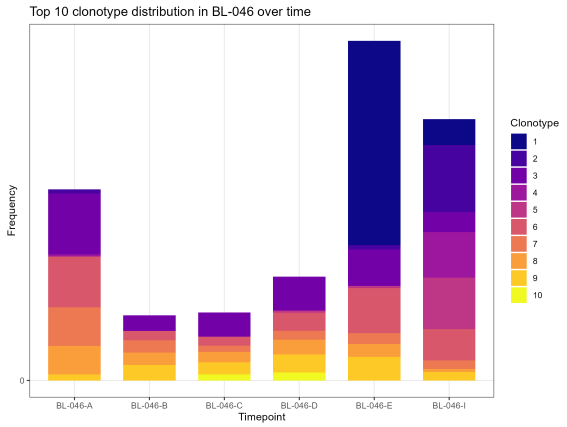


Figure B.6: Clonal expansion for patient BL-046 (responder)

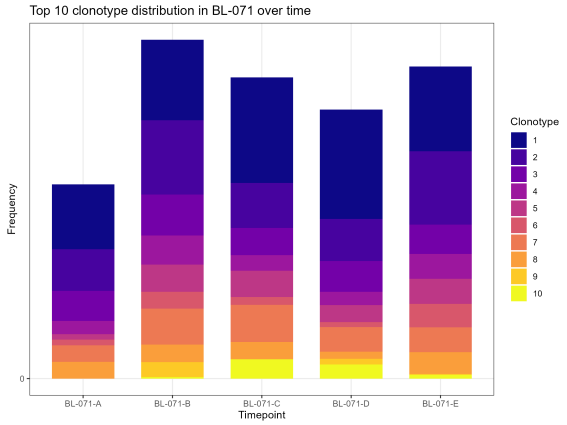


Figure B.7: Clonal expansion for patient BL-071 (responder)

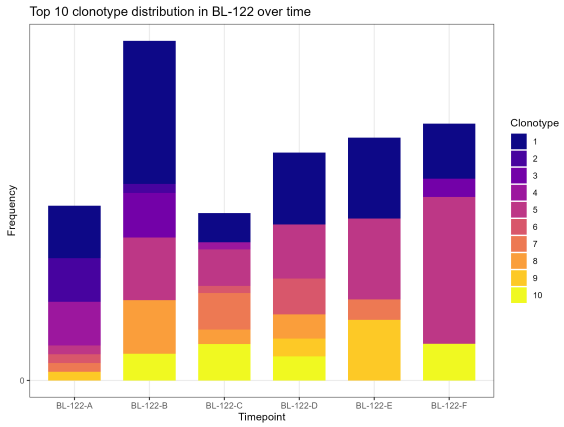


Figure B.8: Clonal expansion for patient BL-122 (responder)

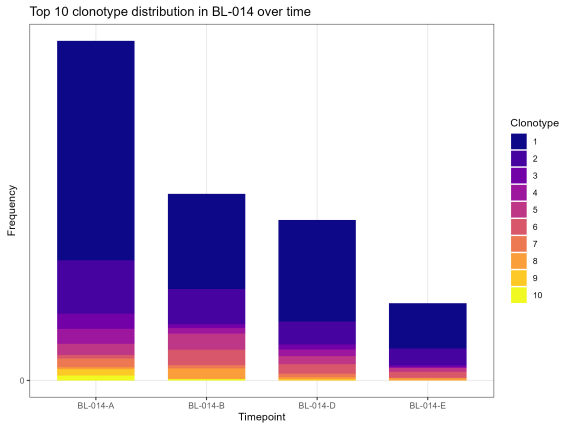


Figure B.9: Clonal expansion for patient BL-014 (non-responder)

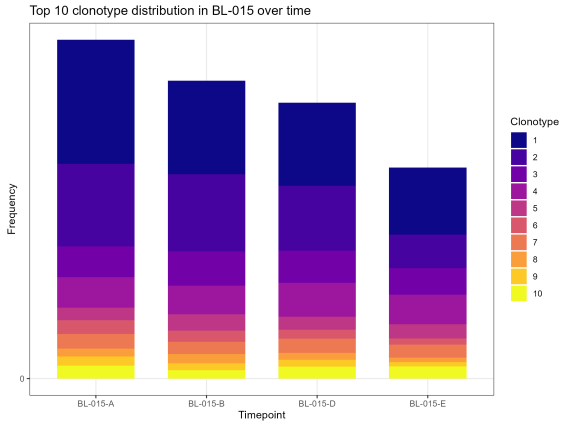


Figure B.10: Clonal expansion for patient BL-015 (non-responder)

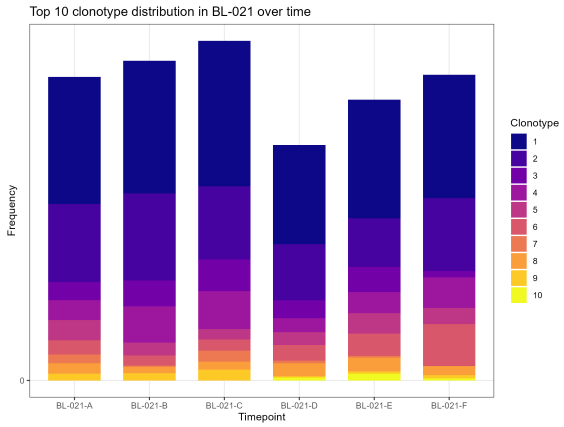


Figure B.11: Clonal expansion for patient BL-021 (non-responder)

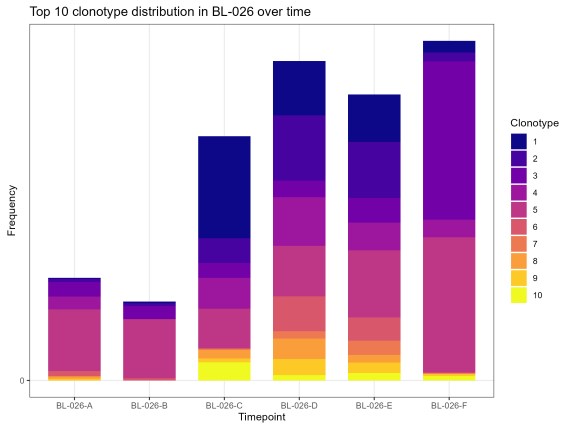


Figure B.12: Clonal expansion for patient BL-026 (non-responder)

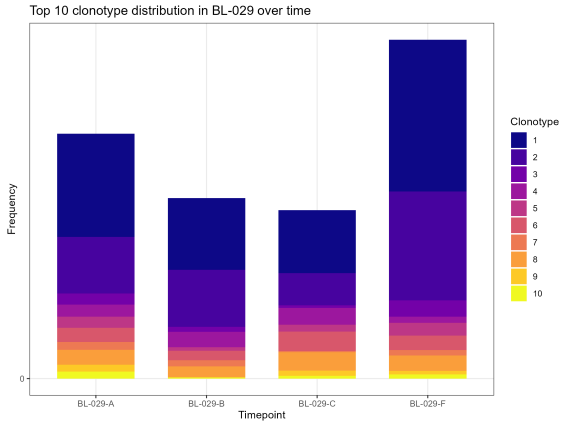


Figure B.13: Clonal expansion for patient BL-029 (non-responder)

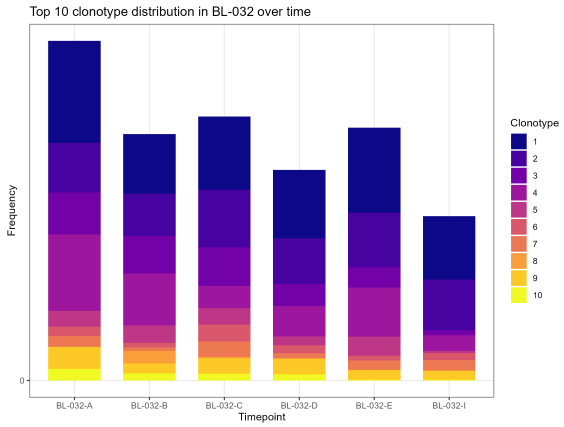


Figure B.14: Clonal expansion for patient BL-032 (non-responder)

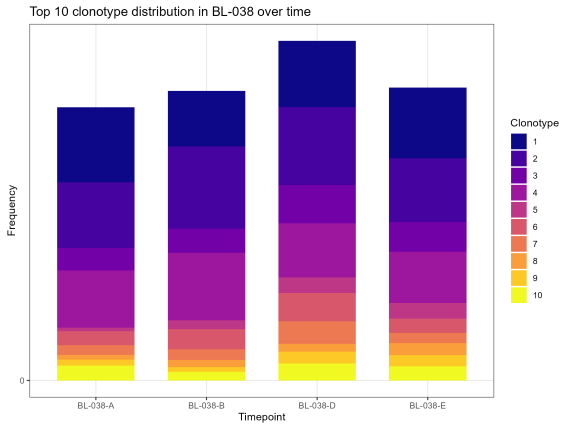


Figure B.15: Clonal expansion for patient BL-038 (non-responder)

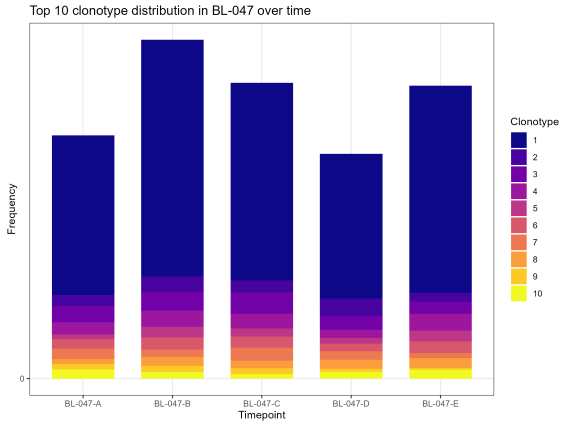


Figure B.16: Clonal expansion for patient BL-047 (non-responder)

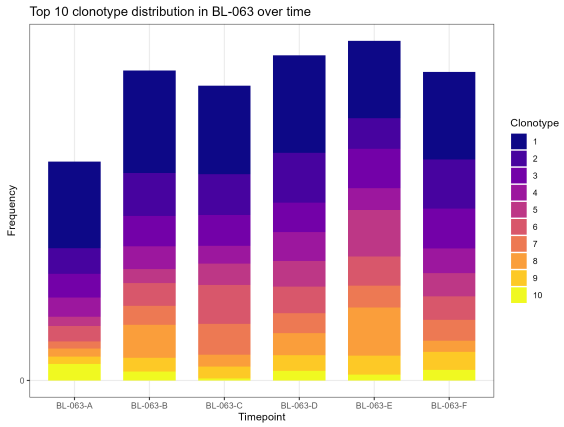


Figure B.17: Clonal expansion for patient BL-063 (non-responder)

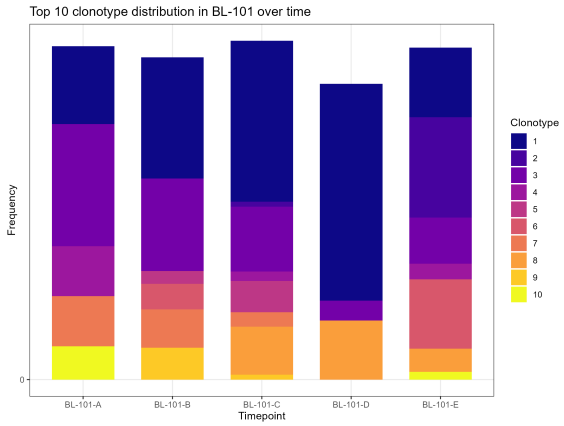


Figure B.18: Clonal expansion for patient BL-101 (non-responder)

DEPARTMENT OF MATHEMATICAL SCIENCES
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY