High-throughput sequencing of human B cell receptor repertoires at single-cell level with preservation of the native antibody heavy and light chain pairs

A Dissertation

Submitted in Partial Fulfilment of the
Requirements for the Degree of
Doctor rerum naturalium (Dr. rer. nat.)

to the Department of Biology, Chemistry and Pharmacy
of the Freie Universität Berlin

by

Praneeth Reddy Devulapally
from Hyderabad, India

Berlin, July 2017

Supervisor:	Prof. Dr. rer. nat. Marie-Laure Yaspo Otto Warburg Laboratory Research Group Gene Regulation and Systems Biology of Cancer Max Planck Institute for Molecular Genetics Ihnestr 63-73 14195 Berlin
Second examiner:	Prof. Dr. Sutapa Chakrabarti Institute of Chemistry and Biochemistry Freie Universität Berlin Takustr 6 14195 Berlin
Date of defense:	12th October 2017

Declaration

I herewith declare that all the experiments reported in this thesis were performed by myself and are not published or submitted elsewhere. Any technical assistance for the experiments has been acknowledged by name and all references are cited accordingly.

All experiments stated in this thesis were performed between 1st August 2013 to July 2017 in the Gene Regulation and Systems Biology research group of Dr. Marie-Laure Yaspo, Otto Warburg Laboratory, under the direct supervision of Dr. Hans-Jörg Warnatz at the Max Planck Institute for Molecular Genetics, Berlin, Germany.

Acknowledgements

Firstly, I would like to thank my mentor Dr. Marie-Laure Yaspo for supporting me and giving me the freedom to accomplish this thesis under her guidance without any objections, while providing insightful and scientific inputs to the project. Furthermore, I would like to thank Prof. Dr. Hans Lehrach, former director of the Department for Vertebrate Genomics at the Max Planck Institute for Molecular Genetics for providing me an opportunity to start my research work in his department. I would also like to thank Dr. Zoltán Konthur for fruitful discussions on my project. Besides, I would like to thank Prof. Dr. Sutapa Chakrabarti for agreeing to be my second examiner as part of my inaugural dissertation thesis at the Department of Biochemistry of the Freie Universität Berlin.

Foremost, I would like to express my deepest and sincere gratitude to my direct supervisor Dr. Hans-Jörg Warnatz for providing me an excellent research platform, guiding, motivating and nurturing me into a better scientist and person. Hans-Jörg has always encouraged and challenged me to think innovatively and question the know-how of every experiment designed. I am very grateful to him for his patience, critical thinking, invaluable support and insightful discussions which led to successful accomplishment of this project. Finally, he is one of the friendliest and smartest supervisors that I have known, and I look forward to him as a scientist and as a person.

I would like to extend my thanks to the present and former members of our group for providing a wonderful atmosphere and support to conduct my research work. In particular, I would like to thank my colleagues (the two angels) Sabine and Carola for being there since the day I started my Ph.D. and supporting me during the up/down phase of my life and my career. I would also like to thank Alex, Matthias, Daniela, Simon, Nilofar, Thomas, Slava, Sören, Jenny, René and Ingrid for support and technical assistance whenever needed.

My heartfelt thanks to the EU Marie-Curie FP7-funded Initial Training Network "VacTrain" consortium, for offering me an opportunity to be part of their Ph.D. program and providing me with valuable resources and world-class training with activities ranging from e-learning modules to attending consortium meetings, courses and conferences, which eventually helped me to become a trained next-generation vaccinologist. I thoroughly enjoyed my time to be part of this international and diverse consortium and made very good friendships.

I wish to thank all my friends who kept me sane and were extremely supportive during the ups and downs of my life, encouraged and motivated me to pursue my goals.

It would have not been easy to reach this point without the support, sacrifices, unconditional love and affection of my parents, my brothers Deepak and Shiva, and my grandparents. Words fall short to express my love and gratitude to you. I am the person I am today because of you. Thank you for being there with me throughout my life, for guiding me in the right path, for your belief in me, and being very supportive with all my decisions. From the bottom of my heart, I love you all. Lastly, the last couple of years here in Berlin have been very joyful and exciting in my life, as I found the love of my life in this city. My dear Dr. Charanya— Thank you for being there during the most critical and important phases of my life and for all your support and motivation, to reach this point. Can't wait to start the new phase of our lives as a married couple.

Finally, I would like to thank everyone who is involved in this wonderful journey.

Abstract

The adaptive immune system in vertebrates provides protection against diverse antigens and facilitates long-lasting and highly specific immunological memory. One of the most important cell types of the adaptive immune system are B cells. These express B cell receptors (BCRs; also called antibodies or immunoglobulins) on their surface which are encoded by pairs of unique heavy chains (HC) and light chains (LC), that allow binding of BCRs to a diverse array of antigens. Each B cell undergoes somatic recombination that randomly combines germline-encoded variable (V), diversity (D) and joining (J) gene segments to produce a unique BCR expressed on its surface. Random V(D)J recombination, combined with additional junctional diversity and somatic hypermutations (SHM) in activated B cells, leads to an estimated BCR diversity of $>10^{14}$ in humans.

Early antibody repertoire studies, employing single-cell RT-PCR and Sanger sequencing, have provided a glimpse into antibody repertoire diversity and allowed screening B cells producing desired antigen-specific monoclonal antibodies (mAbs). However, these techniques were limited by low cell numbers ($>10^2$ B cells), were laborious and involved high costs. Currently, large-scale characterization of antibody repertoires is performed by next generation DNA sequencing (NGS) platforms. Initial antibody repertoire studies using NGS have provided valuable sequencing data on one of the two antibody chains (mostly HC), missing the endogenous paired LC sequence information due to bulk lysis of B cells. The preservation of native HC-LC pairs from single B cells is very crucial for cloning and expression of antibodies and to understand the antibody repertoire diversity. Thus, retaining native antibody HC-LC pairs from single B cells, processing millions of cells in parallel, remained a major obstacle for some time. To this end, very recently several single-cell paired antibody repertoire sequencing techniques were reported. Although substantial, these methods require physical separation of single cells, and are limited by low throughput, low cell numbers (400-10^5 B cells), and require laborious construction of complex flow focusing or

microfluidic instruments, which are expensive and require experienced personnel for operation.

In contrast to the existing paired immune repertoire sequencing methods, this thesis reports the development, application and validation of a simple, cost-effective and innovative technique, employing only common laboratory equipment. In our approach, we prepared rough endoplasmic reticulum (rER) microsomes from B cells (>1 million B cells per experiment), each comprising HC and LC mRNAs from a single B cell. As a positive control for our method, we prepared microsomes from the ARH-77 cell line that expresses known HC and LC sequences. We spiked 10% (v/v) of ARH-77 microsomes to B cell-derived microsomes and passed these microsomes suspension into water-in-oil emulsion droplets containing RT-PCR master mix and multiplex primers. Within each emulsion droplet, the HC and LC mRNAs were reverse transcribed, linked, and clonally amplified by overlap-extension RT-PCR. Subsequently, a nested PCR amplification was performed on the assembled HC-LC DNA, followed by Illumina MiSeq long read sequencing. As a proof-of-concept, our first results showed that paired Ig sequences from individual B cells can be obtained and ARH-77 cell microsomes demonstrated a high HC-LC pairing accuracy of >88%. Next, we applied and validated our method by sequencing $CD19^+$ B cells isolated from human peripheral whole blood samples, obtained before (day 0) and 7 days after Tetanus Toxoid/Diphtheria Toxoid booster vaccination (Td). Applying our method, we identified two previously and four novel TT antigen-specific mAbs. These results demonstrate the application of our technology for discovery of high-affinity mAbs.

Taken together, the method presented in this thesis does not require the physical separation of single B cells and construction of complex flow focusing or microfluidic instruments. The whole process – from B cell isolation to sequencing and analyzing paired antibody repertoires – spans only four days. Antibody validation can be performed within two weeks' time after sequencing data acquisition. In conclusion, our method provides a platform for rapid discovery of high-affinity antigen-specific mAbs with minimal costs and can applied for various aspects of immune biology.

Zusammenfassung

Das adaptive Immunsystem von Wirbeltieren ermöglicht den Schutz vor diversen Antigenen und beinhaltet ein langanhaltendes und hochspezifisches immunologisches Gedächtnis. Einer der wichtigsten Zelltypen des adaptiven Immunsystems sind B-Zellen. Diese exprimieren B-Zell-Rezeptoren (BCR, auch bekannt als Antikörper oder Immunoglobuline) auf ihrer Oberfläche, die durch Paare von einzigartigen schweren Ketten (HC) und leichten Ketten (LC) kodiert sind. Dies erlaubt die Bindung von BCRs zu einer Vielzahl von Antigenen. Jede B-Zelle durchläuft somatische Rekombinationen, die zufällig keimbahnkodierte variable (V), diverse (D) und verbindende (J) Gensegmente kombinieren, um einen einzigartigen BCR auf ihrer Oberfläche zu produzieren. Zufällige V(D)J-Rekombinationen in Verbindung mit resultierender CDR3-Diversität und somatischen Hypermutationen (SHM) in aktivierten B-Zellen führen im Menschen zu einer geschätzten BCR-Diversität von $>10^{14}$.

Erste Antiköper-Repertoire-Studien erlaubten unter Verwendung von Einzellzell-Reverser Transkriptase (RT)-PCR und Sequenzierung nach Sanger einen kleinen Einblick in Antikörper-Diversitäten und ermöglichten die Untersuchung von B-Zellen, die erwünschte antigenspezifische monoklonale Antikörper (mAbs) produzieren. Allerdings waren diese Techniken aufwendig und kostenintensiv und durch kleine Zellmengen von $>10^2$ B-Zellen limitiert. Heutzutage erfolgen umfangreiche Untersuchungen von Antikörper-Repertoires anhand von Hochdurchsatz-Sequenzierung (NGS). Frühe Studien zu Antikörper-Repertoires mittels NGS stellten wertvolle Sequenzdaten zu einer der beiden Antikörper-Ketten (meist HC) bereit. Aufgrund der unspezifischen Lyse von B-Zellen fehlte jedoch die Sequenzinformation über die endogen gepaarten LC. Der Erhalt der ursprünglichen HC-LC-Paare von einzelnen B-Zellen ist fundamental für die Klonierung und Expression von Antikörpern und zum Verständnis der Antikörper-Repertoire-Diversität. Dementsprechend war der Erhalt der nativen Antikörper-HC-LC Paare der einzelnen B-Zellen, bei paralleler Prozessierung von Millionen von Zellen, für einige Zeit ein großes Hindernis. Daher wurden verschiedene Repertoire-Sequenzierungstechniken zur Erhaltung der nativ gepaarten Antikörperketten

entwickelt. Obgleich diese Methoden sehr hilfreich sind, so weisen sie doch einige Nachteile auf. Neben der physischen Separierung einzelner Zellen sind diese Techniken durch einen geringen Durchsatz von 400-10^5 B Zellen limitiert und benötigen eine aufwendige Konstruktion von komplexen mikrofluidischen oder durchfluss-fokussierenden Geräten, die kostenintensiv sind und deren Anwendung erfahrenen Personals bedarf.

Diese Dissertation berichtet von der Entwicklung, Validierung und Anwendung einer innovativen, einfachen und kosteneffektiven Technik, die mit gewöhnlicher Labor-ausrüstung angewendet werden kann. Dabei präparierten wir Mikrosomen des rauen endoplasmatischen Retikulums (rER) von B-Zellen (>1 Million B-Zellen pro Experiment), wobei jedes Mikrosom HC- und LC-mRNAs einer einzigen B-Zelle enthält. Als Positivkontrolle verwendeten wir Mikrosomen der Zelllinie ARH-77, die bekannte HC und LC Sequenzen exprimiert. Wir fügten 10% (v/v) der ARH-77-Mikrosomen zu aus B-Zellen isolierten Mikrosomen hinzu und überführten diese Mikrosomen-Suspension in Wasser-in-Öl-Tröpfchen, die einen RT-PCR-Mastermix und einen multiplexen Primermix enthielten. Innerhalb jedes Emulsions-Tröpfchens wurden die HC- und LC-mRNAs durch Overlap-Extension-RT-PCR revers transkribiert, verbunden und klonal amplifiziert. Anschließend erfolgte mit der assemblierten HC-LC-DNA eine Nested-PCR-Amplifikation und eine darauffolgende Illumina MiSeq Long-Read-Sequenzierung. Als Nachweis der Machbarkeit dienten erste Resultate, die einerseits zeigten, dass gepaarte Ig-Sequenzen von individuellen B-Zellen ermittelt werden können und andererseits, dass bei ARH-77-Mikrosomen eine hohe HC-LC-Paarungsgenauigkeit von >88 % vorlag. Als nächstes erfolgte die Anwendung und Validierung der Methode anhand der Sequenzierung von $CD19^+$ B-Zellen, die aus menschlichen peripheren Vollblutproben isoliert wurden. Die Blutentnahme erfolgte vor (Tag 0) und sieben Tage nach einer Tetanus Toxoid/Diphteria Toxoid Boosterimpfung (Td). Unter Anwendung unserer Methode entdeckten wir zwei bereits bekannte und vier neue TT Antigen-spezifische mAbs. Diese Ergebnisse demonstrieren die Verwendbarkeit unserer neuen Technik zur Entdeckung hochaffiner mAbs.

Zusammenfassend lässt sich sagen, dass die in dieser Arbeit vorgestellte Methode keine physische Separierung einzelner B-Zellen fordert und keine Konstruktion elaborierter Geräte erfordert. Der gesamte Prozess – von der Isolierung von B-Zellen über die Sequenzierung zur Analyse gepaarter Antikörper-Repertoires – benötigt nur vier Tage. Im Anschluss an die Generierung der Sequenzdaten kann eine Validierung der Antikörper innerhalb von zwei Wochen erfolgen. Unsere Methode stellt eine kosteneffektive und in vielen Gebieten der Immunbiologie einsetzbare Plattform zur schnellen Entdeckung hochaffiner Antigen-spezifischer mAbs zur Verfügung.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Innate and adaptive immunity

During the course of our lifetime, we face an innumerable number of diseases, triggered by extrinsic sources (microorganisms, viruses, and toxic agents such as chemicals and radiation) and intrinsic sources (for example, autoimmune processes and tumor cells). The ability to survive in a potentially hostile environment depends on three lines of defense systems which collectively form our immunity. The epithelial barriers serve as the first line of defense and prevent disease-causing agents from entering the body[1]. The innate immune system acts as the second line of defense against pathogens that have bypassed the epithelial barriers[2], providing a rapid immune response by recognizing certain pathogen-associated molecular patterns (PAMPS) and secreting induced interferons and antimicrobial enzymes such as lysozyme for digestion of bacterial cell walls[3, 4]. The phagocytes (e.g. macrophages, dendritic cells and mast cells) of the innate immune system can distinguish between self-antigens and nonself-antigens by recognizing certain molecular patterns on microbial cells and provide protection by activating the so-called complement system for killing and internalizing foreign antigens[5, 6, 7, 8].

Although defense against microbial pathogens by the innate immune system is crucial, certain foreign antigens can evade this defense mechanism. Thus, the third line of defense, also called the adaptive/acquired immune system, has been developed by vertebrates to provide enhanced and strategic immune responses against a wide array of self- and non-self-antigens[9, 10]. In contrast to the innate immune system, an adaptive immune system is more specific, facilitates a long lasting memory and generates a rapid immune response upon re-exposure to the same antigen[11, 12]. The adaptive immune responses are mainly carried out by two different lymphocyte populations, namely B lymphocytes and T lymphocytes. B lymphocytes originate and mature in bone marrow and provide humoral/antibody-mediated immune responses by secreting large amounts of antibodies for specific binding and neutralization of invading bacteria or viruses. In

contrast, T lymphocytes also originate in bone marrow but travel via the blood stream to the thymus for maturation[13, 14]. T lymphocytes provide cell-mediated immunity by recognizing antigens that are displayed by antigen-presenting cells (APCs) such as macrophages and dendritic cells, activating cytotoxic T lymphocytes ($CD8^+$ T cells) and producing several cytokines in response to an antigen. Both B and T cells express cell surface receptors that specifically bind to an antigen and are encoded by different gene segments. These gene segments undergo somatic recombination during early phases of B and T cell maturation and generate diverse cell surface receptors to identify, process and provide protection against a wide range of foreign antigens[15].

Analysis of the B cell repertoire composition, particularly in response to vaccination, and the discovery of antigen-specific monoclonal antibodies (mAbs) is the major focus of this thesis and therefore the B cell repertoires are described in more detail in the next sections.

1.2 Antibody discovery, structure and composition

Antibodies were first discovered in 1890 by Emil Behring and Shibasabura Kitasato[16]. They showed that serum from an immunized animal against diphtheria and tetanus toxin can be used for neutralization of the toxins in another animal. However, it took until the late 1950s for the discovery that only B lymphocytes can produce antibodies[17]. Subsequent investigations on antibody structure and diversity resulted in major breakthroughs revealing the biological and therapeutic importance of antibodies in combating diseases[18, 19, 20, 21, 22]. Antibodies produced by B cells exist in two physical forms: a membrane-bound form expressed on the surface of the cells called the B cell receptor (BCR), and a soluble form that is secreted into the blood plasma. The BCRs are responsible for antigen binding and activation of B cells for subsequent proliferation and differentiation of antibody-secreting plasma cells and memory cells[23, 24, 25]. A typical human B cell possesses around 50,000 – 100,000 antibodies with the same specificity expressed on its surface.

Antibodies, also called immunoglobulins (Ig), are Y-shaped molecules belonging to the immunoglobulin superfamily[26]. Each antibody molecule consists of two identical heavy chain (HC) and light chain (LC) pairs joined together by a disulfide bridge. The antibody molecule can be divided into three equal sized portions (two F(ab) regions and one Fc region) linked together by a flexible hinge region [27, 28] (Figure 1). The tip of the antibody (F(ab) region) binds to the epitope of an antigen with varying affinity and is highly variable among different antibodies, thus providing efficient protection against a diverse range of antigens [29]. In contrast, the Fc region is identical among several antibodies and provides a binding site for Fc receptors expressed on effector cells (e.g. phagocytes) for efficient destruction of antibody-bound antigens[30]. The effector function and the class of immunoglobulin (IgM, IgD, IgG, IgA, and IgD) are defined based on the structure of the heavy chains they possess. For example, naïve B cells express IgM/IgD antibody isotypes on their surface, allowing the B cells to bind antigens followed by killing the pathogens by activating immune complement. On the other hand, IgG isotypes are the most abundant class of antibodies comprising ~70% of the total antibody pool and are mostly produced during secondary immune challenges. Several isotype-specific subclasses exist (e.g. IgG1-4, IgA1-2 etc.) which vary in their ability to induce different effector functions (Table 1).

HC isotype	IgM	IgD	IgG1	IgG2	IgG3	IgG4	IgA1	IgA2	IgE
Serum level mg/ml	1.5	0.03	9	3	1	0.5	3	0.5	0.00005
Complement activation	+++	-	+++	+	+++	-	-	-	-
Molecular weight (kDa)	970	184	146	146	165	146	160	160	188
Half-life in serum (days)	10	3	21	20	7	21	6	6	2

Table 1: Isotype classes and their differences in activating the complement system: This table is adapted from Ken Murphy, P.T., Mark Walport Janway´s Immunobiology seventh edition. 2008. "+" represents the strength of activating the complement system after binding to the antigens and "-" demonstrates no complement activation.

Each antibody producing B cell contains HC and LC encoding genes located on three different loci in human genome. The HC gene is located on chromosome 14q32.33. Two types of LC genes named kappa and lambda (denoted by the Greek letters κ and λ, named k and l here) are located on chromosome 2p11.2 (Igk)[31] and 22q11.2 (Igl)[32], respectively. A given immunoglobulin HC is always paired with either Igk or Igl light chains[33]. The discovery that each antibody has a unique amino acid (aa) sequence in the variable region suggested that antibodies are encoded by separate genes to generate a vast diversity of Ig molecules[34]. The variable regions in the antibody molecules are formed by random rearrangements of several gene segments in the variable domains of HC (V, D, and J) and LC (V and J) during B cell development[35], [36, 37]. The variable region is further divided into four regions of relatively conserved framework region (FR1-4) containing antiparallel beta sheets, and three regions of most diverse hypervariable/complementarity determining regions (CDR1-3). The framework regions are proposed to support the binding of the CDR to antigens[38, 39] and are involved in maintaining the overall structure of the Fv region [40, 41]. In contrast, the CDR3 regions are highly diverse and are responsible for binding numerous antigens with high

specificities[42, 43, 44]. Since antibodies contain a pair of heavy and light chain variable domains, there are twelve antigen receptors that can collectively come in contact with an antigen. The part of the variable region that recognizes and binds to an antigen is called the paratope. The V(D)J recombination process for generating antibody diversity is described in more detail in the chapter below.

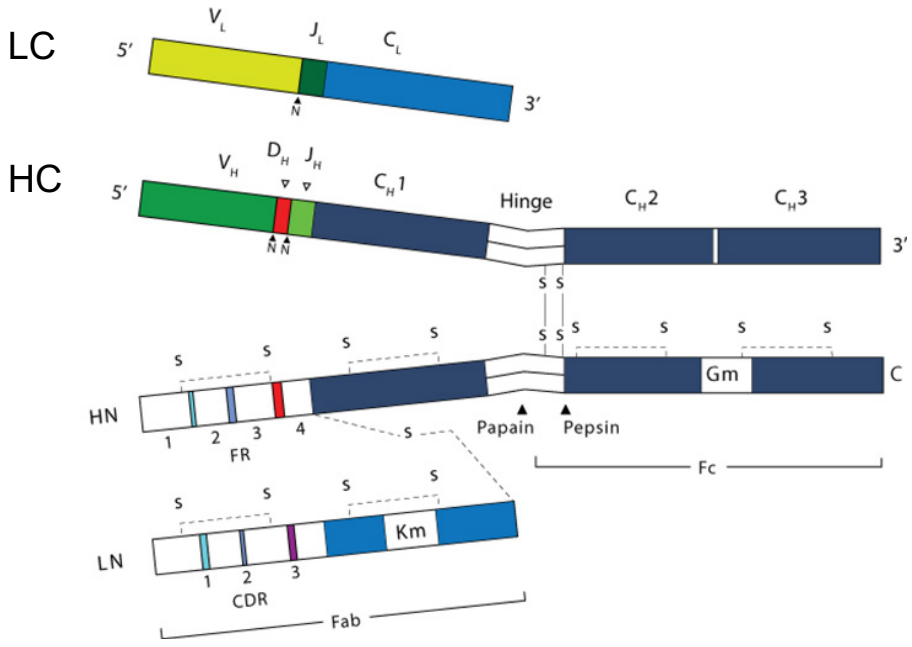

Figure 1: Schematic illustration of the two-dimensional model of an IgG antibody.
Each antibody (IgG) consists of a pair of HC and LC molecules expressed on the surface of a B cell. The upper panel represents the HC and LC at the gene level while the lower panel represents the protein level. The variable regions of the antibody molecule (V_H, V_L) bind to the antigen while the constant regions (Fc region) determine the function of an antibody. N indicates indels of nucleotides along V_H, D_H, J_H of the HC region and V_L and J_L of the LC region, demonstrating the enormous diversity in the V(D)J region of an antibody for efficient binding to diverse antigen molecules. The hinge region connects the C_{H2} in the Fc region with C_{H1} in the Fab arm. An antibody can be fragmented into antigen-binding fragments (Fab) and class-determining fragment (Fc) at the hinge region by proteolytic enzyme-mediated digestion (for e.g. using papain or pepsin). The protein schematic of the antibody from the amino-terminus (N) to the carboxy-terminus (C) specifies the linkage of HC and LC fragments by disulphide (S=S) bridges. The isotype-specific markers in the antibody are denoted by Gm for HC and Km for LC, respectively. The CDR3 regions in the Fab arm are colored while the framework regions are displayed in white color. This figure is modified from Schroeder *et al.* [45]

1.3 Immunoglobulin V(D)J recombination

The diversity in the paratope of antibody is a result of V(D)J somatic gene recombination (V=variable, D=diverse, and J=joining) at chromosomal level (Figure 1). The HC and LC V(D)J recombinations primarily take place in developing B cells within bone marrow. The human Ig HC chromosomal region contains 55 V, 27 D, and 6 J gene segments[46, 47, 48] while the LC chromosomal region together contains 64 V and 9 J gene segments, but lacks the D gene segments[49, 50]. Thus, based on these rearranged V(D)J gene segment counts, the antibody diversity in humans is estimated to range from 10^{12} to 10^{14} [51, 52, 53]

In developing B cells, first the heavy chain V(D)J rearrangement occurs, followed by VJ gene rearrangement of the light chain. The V(D)J recombination in the pro-B cell is activated by products of recombinase activated gene 1 and 2 (RAG 1 and RAG2). The RAG enzyme recognizes conserved noncoding DNA segments and introduces double-strand breaks between V, D and J gene segment located adjacent to recombination signal sequences (RSS). A RSS is composed of a conserved heptamer and a nanomer sequence separated by nonconserved 12/23 bp spacer sequences[54, 55]. After the double-strand break of V(D)J segments at RSS sites, one D and one J gene segments of the HC are joined together forming a DJ complex followed by joining of a V gene segment forming a recombined V(D)J gene segment (Figure 2). The joining of V(D)J gene segments is carried out by an enzyme called terminal deoxynucleotidyl transferase (TdT) which is a member of the nonhomologous end joining protein family (NHEJ)[56]. In addition, the TdT enzyme also adds non-templated nucleotides to the DJ or VJ junctions during the recombination process, increasing antibody diversity[51]. Deficiency in either RAG1 or RAG2 gene products results in inhibition of the V(D)J recombination process in early stages of pro- and pre-B cell development[57]. Likewise, deficiency in TdT enzyme results in reduced junctional diversity[57]. After V(D)J recombination, the pre-B cells express the recombined HC on their surface and undergo expansion (up to 8 cell divisions), generating multiple B cell clones with the same heavy chain[58]. Next, Ig LC recombination takes place in the similar manner as described

above. However, in LC recombination, a V gene segment joins with one J gene segment forming a rearranged VJ complex before the addition of LC constant regions (Figure 2). The recombination events are stopped when the newly formed Ig μ HC and surrogate LC are expressed on the surface of the immature B cell[59, 60]. The immature B cells expressing IgM antibody on their surface pass through an immune tolerance checkpoint in bone marrow, where the B cells that recognize self-antigens undergo clonal deletion[61]. However, the B cells that recognize self-antigens can be rescued from clonal deletion by undergoing receptor editing that may replace self-reactive receptors with nonself-reactive receptors. After passing this checkpoint, the immature B cells with surface expressed IgM and IgD antibodies travel to the spleen where they recognize and bind to diverse range of antigens [62, 63].

Upon binding to an antigen, the precursor B cells get activated and proliferate into plasma B cells or memory B cells (Figure 3). During the proliferation stage, activation-induced cytidine deaminase (AID) induces SHM in the CDR regions of B cells [64] enhancing the ability of B cells to recognize a wide array of foreign antigens[65]. If the activated B cells receive molecular signals via their CD40 and cytokine receptors, the antibodies produced by mature B cells (IgM and IgD) can class-switch to other isotype forms (e.g. IgG, IgA, and IgE). During class switch recombination (CSR), only the constant region of the antibody changes while the variable region remains the same, allowing the activated B cells to produce antibodies with different subtypes that can effectively neutralize the respective foreign antigens[66].

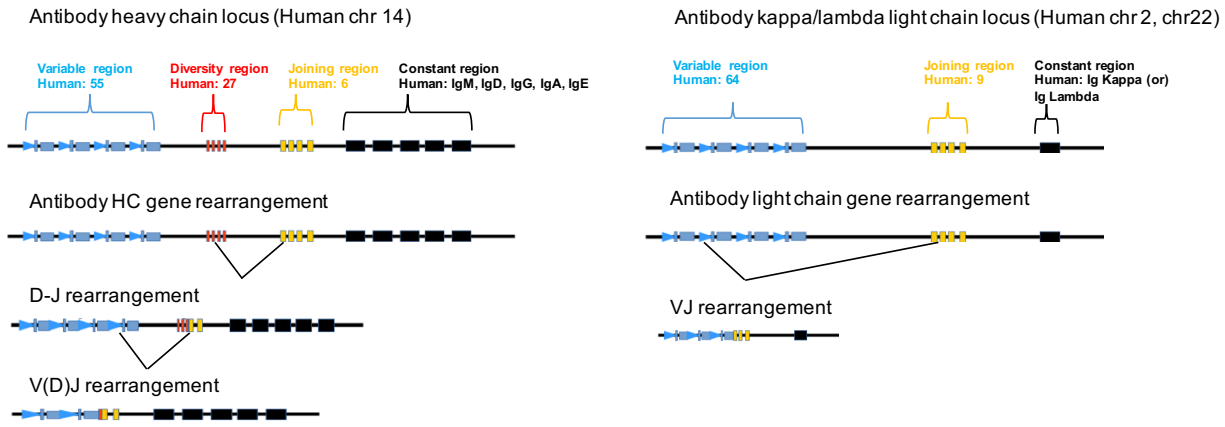

Figure 2: Schematic overview of immunoglobulin heavy chain and light chain gene rearrangement: The V(D)J recombination event in developing B cells plays an important role to enable an enormous diversity in the antibody repertoire to detect and bind to a wide array of pathogens. The immunoglobulin (Ig) HC and LC loci are located on separate chromosomes. The Ig HC locus is on chromosome 14, while the Ig Kappa LC locus is on chromosome 2 and the Ig Lambda locus is on chromosome 22. In a developing B cell, first the HC gene arrangement occurs, followed by the LC gene rearrangement. The first recombination event occurs in the HC locus, where one D and one J segment are joined together, followed by joining of one V segment to the DJ complex to form a fully rearranged V(D)J gene. The second recombination event occurs in the Ig Kappa or Lambda locus, where one V and one J segment are joined together to form a fully rearranged VJ gene. Finally, the assembly of a Igμ HC with one of the LC (Kappa/Lambda) results in a membrane-bound form of IgM antibody expressed on the surface of an immature B cell. This figure is modified from Dr. Stephen Peluse's overview on antibody gene rearrangement. http://mainebiotechnology.com/generation-of-antibody-diversity-through-gene-rearrangement-and-somatic-hypermutation/

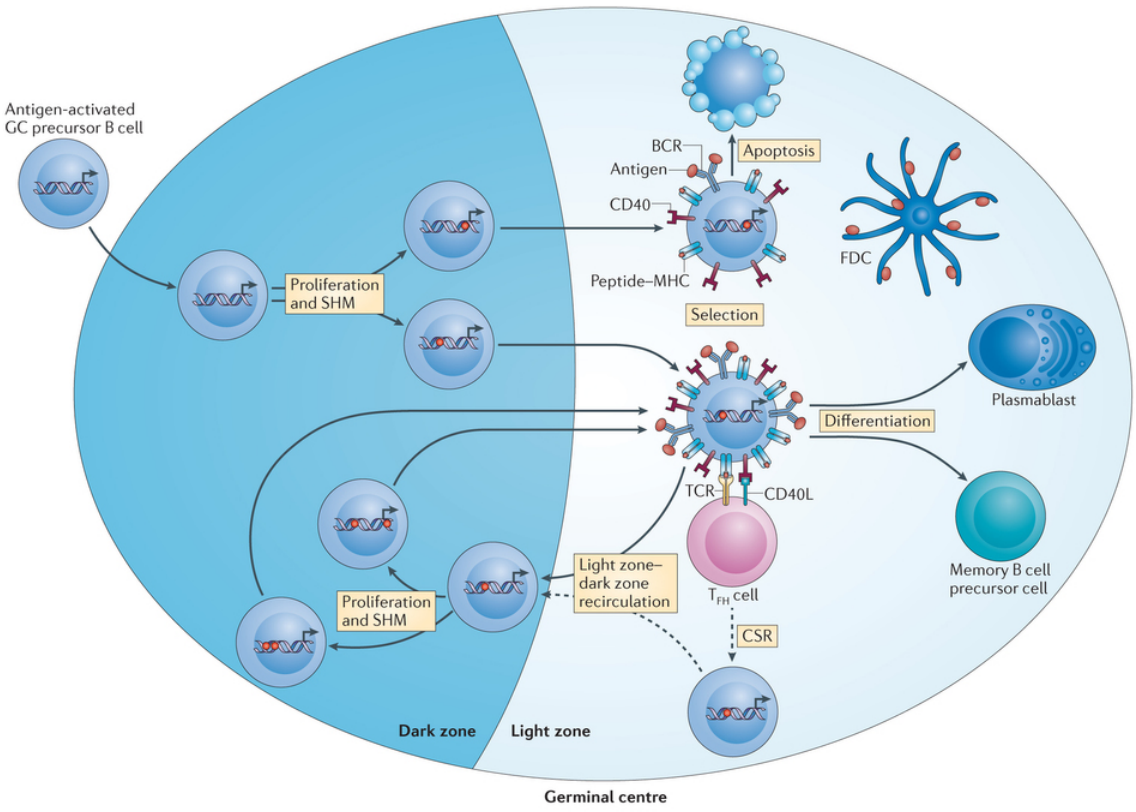

Figure 3: Dynamics of germinal center reactions, where selection, proliferation, isotype class switching and differentiation of B cells occur: An antigen-activated precursor B cell undergoes proliferation and somatic hypermutation (SHM) in the recombined antibody V(D)J gene in the dark zone of the germinal center. Then, the B cell proliferates to the light zone where the modified B cell surface receptor binds to the antigens presented by follicular dendritic cells (FDCs) or T follicular helper cells (TFH cells) with high affinities. In the light zone, the B cell receptors that bind to self-antigens undergo apoptosis. The B cells that receive molecular signals via CD40 ligands expressed on the surface of T cells undergoes class switch recombination (CSR) to produce antibodies with other isotype forms. After binding to an antigen with high specificities the CSR, B cells undergo differentiation into antibody-producing plasmablasts or memory B cells. This figure is modified from Nilushi S. De Silva *et al.*[67]

1.4 Traditional methods for investigating antibody repertoire diversity

Based on the techniques used, immune cell receptor repertoires can be studied at different biological levels for e.g. studying secreted or membrane-bound proteins and investigating the DNA/mRNA transcripts of the rearranged VDJ genes. At the protein

level, the first antibody repertoire studies used isoelectric focusing on polyacrylamide gels [68]. This allowed resolving the heterogeneity of serum IgG antibodies into discrete bands, demonstrating the application of this technique to look into the composition of heterogeneous and complex antibody mixtures. At the genomic DNA level, CDR3 spectratyping, also called immunoscope technology, was developed for analysis of CDR3 length and usage[69, 70] providing insights into antibody repertoires after vaccination [71].

Furthermore, the advent of Sanger sequencing method[72] has paved a way to study antibody repertoires at single-nucleotide level, permitting to determine exact nucleotide sequences of antibody variable and hypervariable regions. Repertoire studies using Sanger sequencing method has provided a great insight into understanding antibody structure, composition and diversity[73, 74]. Next, the advent of fluorescence activated single-cell sorting (FACS) has allowed studying B cell populations of interest (e.g. plasma B cells, memory B cells and naïve B cells) at the single-cell level. Early single-cell studies on B cell receptors were performed by FACS sorting and clonal cell expansion in culture[75] or by using single-cell reverse transcription PCR[76, 77] for generating adequate amounts of DNA for Sanger sequencing and/or for cloning experiments, for discovery and production of mAbs. Given the enormous diversity in antibody receptor repertoires as described above, single-cell RT-PCRs followed by Sanger sequencing of B cell repertoires were often performed using 10^2 cells within 96-well or 384-well plates, which was laborious and expensive. However, these technologies were fruitful in providing a first glimpse into the complete antibody repertoire[78].

The above-mentioned methods were mostly used to investigate mAbs produced against a wide variety of antigens. These methods also showed that activated B cell clones against a particular antigen differ between individuals [76, 79]. Although limited by the number of B cells that were investigated, these methods pioneered in providing valuable knowledge on some antigen-specific sequences in response to vaccinating antigens[79].

1.6 Immune repertoire studies using next generation sequencing technologies

The assessment of antibody diversity using Sanger sequencing method was highly robust, delivering long sequencing read lengths with low error rates. However, studying the Ig repertoires by processing millions of B cells using traditional methods was too laborious and very expensive[80, 81]. The advent of next generation sequencing (NGS) technologies in the last decade has advanced biological research tremendously, enabling researchers to sequence millions of DNA molecules simultaneously while exponentially reducing the sequencing costs (**Figure 4**). Although powerful, NGS technologies are limited by relatively high error rates occurring due to PCR or sequencing, and short read lengths of DNA sequences. Several NGS technology platforms are currently available in the field, possessing different strengths, characteristics, and bottlenecks. A comparison of current high-throughput sequencing technology platforms is listed in **Table 2**.

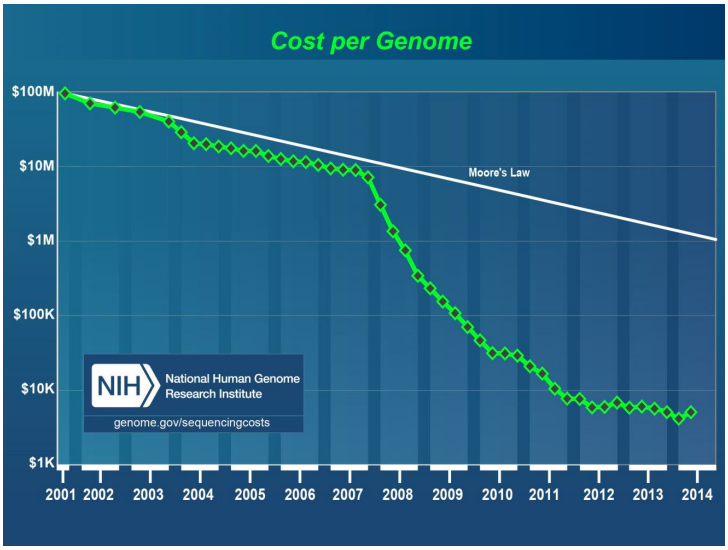

Figure 4: Overview of drop in DNA sequencing costs: DNA sequencing using next generation sequencing platforms developed in the recent years has witnessed a log-scale reduction of sequencing costs. This reduction in sequencing costs has directly enabled to sequence and analyze antibody repertoires at unprecedented depth, accelerating translational research ranging from antibody discovery to cancer immunotherapies. Graph obtained from *National Human Genome Research* Institute with permission granted for presentation and teaching purposes.
http://www.genome.gov/sequencingcosts

Sequencing platform	Instrument cost ($US)	Sequencing cost per Gb ($US, approx.)	Chemistry	Read length (bp)	Sequencing depth	Throughput	Error rate/type	Run time
454 GS FLX Titanium XL+ (Roche)	N/A§	$9.5 K‡	Pyrosequencing	Up to 1000; 700 average (SE, PE)	~1 M*	700 Mb*	1%, indel‡	23 h*
SOLiD 5500 xl	$251K‡	$70‡	Sequencing by ligation	50 -75 (SE)	~1.4 B*	240 Gb*	≤0.1%, AT bias‡	10 d*
Illumina MiSeq	$128K‡	$142‡	Sequencing by synthesis	250-300 (PE)*	24-50 M*	7.5-15 Gb*	0.1%, substitution‡	39-56 h*
Illumina HiSeq	$690K‡	$40‡	Sequencing by synthesis	250 (PE)*	600 M (PE)*	125-150 Gb*	0.1%, substitution‡	60 h*
Ion Torrent PGM	$80K‡	$25-3.5K‡	Ion semiconductor	400 (SE)*	3-5 M*	1-2 Gb*	1%, indel‡	7.3 h*
PacBio RS	$695K‡	$1K‡	Single molecule in realtime sequencing	~8-20 Kb	500 Mb-1 Gb*	500 Mb-7 Gb*	13% single pass, ≤1% circular consensus read, indel‡	4 h*

Table 2: Comparison of selected next generation sequencing platforms:

N/A, not available; Gb, gigabase pairs; approx., approximate; bp, base pair; SE, single-end sequencing; PE, paired-end sequencing; Kb, kilo bases; M, million; B, billion; Mb, mega base pairs; AT, adenine and thymine; h, hours; d, days.

§ Costs for this instrument are not available as this instrument will be discontinued.

‡ Rounded from field guide to next-generation DNA sequencing

* According to manufacturer's data.

This table is modified from Sara Goodwin *et al*. [82]

NGS technologies, also called high-throughput sequencing technologies, have transformed antibody repertoire sequencing studies by permitting sequencing and analysis of a large number of B cell and T cell receptor repertoires[83, 84, {Arnaout, 2011 #171}]. The first study to determine antibody repertoire diversity using high-throughput sequencing technologies was conducted in zebrafish[85]. Zebrafish contains very few antibody-producing B cells (~300,000), five orders of magnitude less than human antibody repertoires, thus making it an ideal model organism to study antibody HC and LC VDJ gene rearrangements and understand antibody repertoire diversity[85]. The measurement of complete human antibody repertoire diversity is still impossible due to limitations in tissue sampling (for e.g. peripheral whole blood, secondary lymphoid organs and bone marrow) and ethical concerns[86, 87]. First human antibody repertoire studies using NGS technologies were performed in 2009 using DNA isolated from tissue and blood samples[84]. These studies on human antibody diversity led to the discovery of frequencies and similarities in VDJ gene segment usage of human HC antibody repertoires[88,89]. In the last eight years, a huge rise in the application of high-throughput sequencing platforms to study human antibody repertoires has been witnessed and was applied to characterize human antibody repertoires of diseased and healthy states[90,91] and, very importantly, to rapidly identify antigen-specific high-affinity monoclonal antibodies[92,93]. Thus, human immune repertoire studies using NGS platforms have proven to be extremely useful and are providing a deeper insight into understanding adaptive immune repertoire diversity.

1.7 Limitations and recent developments in immune repertoire sequencing studies using NGS

While there are several remarkable developments in the NGS platforms and library preparation protocols for immune repertoire sequencing, this field still possesses several experimental and computational challenges to fully understand and make meaningful inferences of the antibody repertoire sequencing data. Here are some of the limitations and recent advancements in the immunoglobulin repertoire sequencing technologies.

1.7.1 Addressing the PCR biases using experimental approaches

The first limitation in analyzing immune repertoire sequencing data is to identify sequencing errors generated by the NGS platforms, PCR amplification (e.g. multiplex PCRs, RT-PCR) and/or during library preparation methods. The experimental errors could arise due to the inability of reverse transcriptase and/or DNA polymerase to flawlessly transcribe and amplify the genomic DNA/total RNA used in the PCR reactions. Furthermore, higher PCR amplification cycles result in nucleotide substitution errors, which are generally observed in PCR protocols. Besides, the use of degenerate multiplex PCR primers in PCR reactions, that targets the V-region of HC and LC genes results in a bias towards preferentially amplified V-segments. Thus estimating the original amounts of templates used in the PCR reaction based on the sequencing read counts[94,95] will become difficult. However, these problems can be mitigated by performing few number of amplification cycles, use of high-fidelity DNA polymerase/reverse transcriptase enzymes with proofreading activity[96] and use of template-switching reactions that allows a single forward primer to be used in PCR protocols for e.g. 5' RACE[97, 98]. Although employing the above-mentioned steps could considerably reduce the PCR amplification bias, it should, however, be noted that the certain degree of substitution errors and preferential amplification biases still occurs[99,100].

1.7.2 Addressing the sequencing errors using computational approaches

Given the enormous diversity in antibody repertoires, the misinterpretation of artefacts in the sequencing data will directly impact the understanding of immunological responses. Thus a sophisticated computational approach is required to correctly assess the sequencing data. Several research groups are working towards this goal, and complications underlying DNA sequencing errors using NGS were reviewed recently[101, 100, 102, 103].

Briefly, NGS platforms generate a huge amount of sequencing data with different Phred quality scores for each read[104]. Filtering the sequencing reads based on Phred quality scores is one way to remove the false-positive reads from the

sequencing data and to estimate the diversity in the analyzed Ig repertoire[105]. A Phred quality score of 20 means one error per 100 bp (p=10-Q/10), where p=probability of erroneous base call while, Q= Phred quality scores associated with that base[106]. Another approach for correcting the sequencing errors is by clustering the highly similar sequences together based on their CDR3 sequence or sometimes even complete V(D)J sequence[107] and building a consensus sequence based on all sequencing reads in a given cluster. Although, this method is proven to be feasible in application with TCR-sequencing, this is less effective in application with Ig repertoire sequencing. This is because, the antigen-experienced B cells which underwent SHM and clonal expansion shares similar CDR3 sequences but might undergo some mutations in V(D)J region. Clustering and error corrections based on CDR3 sequence alone will result in removal of unique somatic variants, reducing the repertoire diversity[100]. Several immune repertoire error correction software's have been developed very recently[108-110]. For e.g. very recently, MiXCR software was developed, which considers sequence quality, corrects PCR errors and specifically distinguishes SHM in the Ig repertoire sequencing data containing both single- and paired –end sequencing data[109]. The developments of these software's help in efficiently handling and interpreting the NGS data and make meaningful inferences of Ig repertoire sequences.

Another unique and powerful way to correct sequencing errors and quantify an absolute number of template molecules is based on using unique molecular identifiers (UMIs)[110, 111]. UMIs are random nucleotide sequences of 6-16 bases (e.g. NNNNNN) that rely on tagging each mRNA or DNA molecule prior to library preparation steps, thereby facilitating the quantification of template molecules which is unaffected by amplification biases[112]. Further, clustering the sequencing reads according to their cDNA facilitates to generate a consensus sequence, circumventing the PCR errors. However, currently available protocols require template switching during the RT step, which could lead to the loss of low-expression transcripts[113, 114].

1.7.3 Loss of endogenous heavy and light chain pairing information

Early antibody repertoire studies using NGS were limited by the loss of native chain pairings (HC and LC) from individual B cells due to bulk lysis of tissue or B cell populations of interest[84, 85, 115]. The native antibody HC-LC pairing information is important for understanding natural antibody VDJ gene rearrangements and usage, for generating high-affinity antigen-specific mAbs, for investigating immune responses to cancer, and for many other aspects of the biology of the immune system. Several low-throughput single-cell antibody receptor repertoire sequencing technologies were developed in the recent years to obtain native antibody HC-LC pairs for efficient cloning and discovery of monoclonal antibodies[116-118]. These single-cell technologies were often limited by cell number (<200-500 cells), physically sorted by FACS sorting into 96-well or 384-well plates and sometimes required fabrication of complex microfluidic devices which are expensive and commercially not available[119-122].

1.7.4 Emulsion-based paired antibody repertoire sequencing

Encapsulation of template fragments into minute, physically separate, water-in-oil emulsion droplets is a proven method to perform RT and PCR amplification of complex genomic libraries[123].Each template molecule (DNA/total RNA) encapsulated in a separate reaction compartment undergoes amplification until all the resources (e.g. RT and/or PCR master mix, degenerate multiplex primers) within each emulsion droplet are completely used. Therefore, a large-scale competition of resources for amplification of each template molecule, erroneous artefact amplification, and preferential amplification biases can be minimized using the emulsion PCR amplification method. Thus, employing emulsions for RT and PCR amplifications to generate immune repertoire libraries offers a strategy to retain paired immune receptor repertoire sequences. When I started this Ph.D. project DeKosky *et.al* have reported a way to retain native antibody HC-LC pairs from single B cells using around 500,000 cells per experiment[124]. In their approach, single-cells were captured inside PDMS slides containing beads for cell lysis and mRNA capture to the beads. These beads attached mRNA were emulsified followed by overlap-

extension RT-PCR amplification within each emulsion droplet. Later, the HC-LC assembled products were extracted from the emulsion droplets and the libraries were sequenced by MiSeq 2x250 bases paired end sequencing. Although this method proved to be feasible, it is limited by low throughput and requires PDMS microtiter plates which are laborious to construct and are commercially not available. In another approach, Turchaninova *et.al* encapsulated whole cells (T cells) into water-in-oil emulsion droplets which were followed by lysis of cells based on high temperatures, reverse transcription and amplification of mRNA templates within each emulsion droplet[122]. Although this method was applied to process >106 cells per experiment, this was limited by low-throughput, entrapment of cells between individual emulsion droplets and inadequate amplification of templates by reverse transcriptase and/or DNA polymerase enzymes. However, very recently, two other emulsion based high-throughput paired antibody repertoire sequencing technologies were reported which uses >2-3x10^6 B cells per experiment[125]. Although powerful, these methods are limited by the use of very expensive microfluidic devices or elaborate construction of a flow-focusing apparatus which requires expertise for operation.

Based on the above-mentioned applications and limitations of high-throughput sequencing technologies for immune repertoire analysis, a need for the development of a more high-throughput sequencing method to characterize and retain native antibody pairs at single-cell level using millions of B cells was evident. This thesis describes the development and validation of a novel simple and cost-effective paired antibody HC-LC sequencing method for antibody discovery and cloning. Furthermore, this thesis addresses the previously reported limitations such as B cell number, low throughput, high costs, and combines the above mentioned computational developments in analyzing Ig repertoire sequencing data. The knowledge gained from sequencing paired antibody HC-LC repertoires can be applied to accelerate translational research ranging from antibody discovery and autoimmune disease research to cancer immunotherapies.

Chapter 2: Aims

Aim 1:

Development of a simple and cost-effective technology to retain and sequence native antibody heavy- and light-chain pairs from single B cells.

The primary aim of this thesis is focused towards developing a simple, cost-effective and innovative technology that enables to retain high-throughput sequencing information of native antibody HC-LC pairs from single B cells, processing millions of B cells isolated from human peripheral whole blood samples. The amplicons generated using this approach should contain the native assembled HC-LC pairs from single B cells and should address the earlier reported limitations in antibody repertoire sequencing studies (as mentioned above, Chapter 1). Furthermore, the bioinformatic pipeline used for analysis of paired antibody repertoire sequences should carefully consider PCR and sequencing errors, group and extract identical sequences into a human-readable text file with final clonotype information.

Aim 2:

Application and validation of the developed method to study antibody repertoire changes in response to vaccination and for discovery of novel antigen-specific monoclonal antibodies.

Native antibody HC-LC pairs of $CD19^+$ B cell populations isolated from human peripheral whole blood samples of a healthy individual, obtained on day 0 before and 7 days after Tetanus Toxoid/Diphtheria Toxoid vaccination should to be analyzed. The analysis of native antibody HC-LC pairs should be used to identify and track clonally expanded B cell populations in response to vaccination, showing a potential application for discovery of high-affinity antigen-specific monoclonal antibodies in response to vaccination and diseases.

Chapter 3: Materials and Methods

Methods performed in this section are mostly based on manufacturer's instructions obtained for commercially available kits or based on protocols obtained from "Molecular Cloning: A Laboratory Manual" (Sambrook et al. 1989). However, some protocols were slightly modified according to the experimental needs, these protocols are described in detail in this chapter. Unless stated, the chemicals were purchased from either Sigma-Aldrich Chemie GmbH (Germany) or Thermo Fisher Scientific (Germany).

Materials

Chemicals/enzymes/kits/equipment	Company
1M Tris Buffer	Ambion
26 G needle	BD
5 ml Syringes	BD
5810R Centrifuge	Eppendorf
5815R Centrifuge	Eppendorf
5424 Centrifuge	Eppendorf
Acrylamide	BIO RAD
Agencourt AMPure XP beads	Beckman Coulter
APS	BIO RAD
ARH-77 Cell Line	ATCC®
Axiovert 40 CFL Microscope	Carl Zeiss
BssHII	New England Biolabs
Blotting-Grade Blocker	BIO RAD
C-Chip Disposable Hemacytometers	INCYTO
CD19 S-pluriBeads	pluriSelect Life Science
Clostridium tetani, tetanus toxin (tetanospasmin, TeNT) antibody [b416M]	GeneTex
Cryogenic Vial	VWR
dART 1-Step RT-PCR Kit	roboklon
Dimethylsulfoxide (DMSO)	CARL ROTH
DNA LoBind Tubes	Eppendorf

dNTP Set	Amersham
DynaMag-2 Magnet	Invitrogen
Gerhardt RMO 76 Hot Plate	LabX
GloMax Multi Detection System	Promega
Glycine Buffer pH 3.0	Jena Biosciences
Goat anti-human kappa LC-HRP secondary antibody	Invitrogen
HindIII	New England Biolabs
Illumina MiSeq paired end Kit	Illumina
Instant sticky end DNA ligase master mix	New England Biolabs
LB Broth	BD
Low-range Ultrapure Agarose Powder	Bio Rad
Micellula DNA Emulsion and Purification Kit	roboklon
Micro Bio-Spin® Chromatography Columns	BIO RAD
Mini ProteanIII electrophoresis unit	BIO RAD
Mouse IgG HC-HRP conjugated secondary antibody	Abcam
NEB 2.1 Buffer	New England Biolabs
NheI	New England Biolabs
Nuclease Free Water	Ambion
Competent *E.coli* Top10 cells	Iba-Solutions for Life Sciences
PCR primer cocktail	Illumina
PCR tubes	Eppendorf
Pellet Paint NF Co-precipitant	Merck
pluriPlix Two-Armed Universal Rotation Mixer	pluriSelect Life Sciences
Protein A Agarose Beads/Resin	Sino Biologicals
Protein LoBind Tubes	Eppendorf
PTC-200 Peltier Thermocycler	MJ Research
Q5® High-Fidelity PCR Kit	New England Biolabs
Qiagen MinElute DNA Kit	QIAGEN
QIAprep Spin Miniprep Kit	QIAGEN
Qubit 3.0 Flurometer	Life technologies
Real-Time PCR System	Applied Biosystems

RiboLock RNAse inhibitor	Thermo Fisher Scientific
S-pluriBead Mini Reagent Kit	pluriSelect Life Sciences
Sodium Dodecyle Sulfate (SDS)	CARL ROTH
SYBR® Select Master Mix	Applied Biosystems
Tetanus Toxoid in-vitro	Staten Serum Institute
TRIzol reagent	Invitrogen
TruSeq ChIP Sample Preparation Kit	Illumina
Turbo DNA-free Kit	Ambion|Life Technologies
Ultrapure agarose	Invitrogen
Vortex-Genie 2T	Scientific Industries
Waterbath WNB	Memmert GmbH
Zymo DNA Clean & Concentrator Kit	Zymo Research
Zymoclean Gel DNA recovery Kit	Zymo Research

Table 3: Materials used in this thesis

3.1 Cell culture, vaccination, B cell isolation and microsome preparation

3.1.1 ARH-77 cell culture

ARH-77 cells derived from a IgG plasma cell leukemia patient were obtained from the American Type Culture Collection (ATCC). Complete medium for cell growth was prepared by adding 10% FBS and 1% P/S to Hybridoma SFM medium. Thawing, growth, subculturing and cryopreservation of cells was performed according to the ARH-77 (ATCC CRL1621) product manual.

3.1.2 Vaccination and collection of human peripheral whole blood samples

Human peripheral whole blood samples used in this project were purchased from in.vent Diagnostica GmbH (Hennigsdorf, Germany) after obtaining informed consent from the healthy individual. Ethical approval was obtained according to Freiburg Ethics Commission International, Feki code 011/1763.

For this study, a healthy individual donated 50 ml of human peripheral whole blood samples from before (day 0) and 7 days after Td booster immunization (Tetanus Toxoid (TT)/Diphtheria Toxoid (DT)) [Td-pur®; 20 International units (IU) TT and 2 IU DT; Novartis, Basel, Switzerland]. Blood was drawn intravenously into Vacutainer K_2EDTA plasma tubes (SARSTEDT-Monovette) and was delivered on the same day.

3.1.3 $CD19^+$ B cell isolation from human peripheral whole blood samples

$CD19^+$ B cells used in this study were isolated from freshly derived human peripheral whole blood samples (day 0 and 7 days post Td booster immunization) using CD19 S- pluriBead anti-hu and S-pluriBead Mini Reagent Kit (pluriSelect GmbH, Leipzig, Germany) following the manufacturer's instructions detailed in the link below.
(http://www.tataa.com/wp-content/uploads/2012/10/Product-Catalog-2015-V1_1.pdf)

After B cell isolation, the cell pellet was resuspended in 1 ml of 1x wash buffer, and cell count was performed using C-Chip Disposable Hemacytometers cells under an

Axiovert 40 CFL Microscope. Finally, the desired number of cells (e.g. $1.5x10^6$ cells) were transferred into fresh 1.5 ml ProteinLoBind tubes and stored on ice for cycloheximide treatment.

3.1.4 Cycloheximide treatment of freshly isolated $CD19^+$ B cells and ARH-77 cells

Freshly isolated $CD19^+$ B cells from both before (day 0) and 7 days after Td booster immunization and cultured ARH-77 cells were treated with cycloheximide to preserve the mRNAs at the ribosomes translocon complexes and to further inhibit the translation of mRNAs into proteins[126,127].

In this protocol, the cell pellets ($CD19^+$ B cells and ARH-77 cells) were resuspended in 1 ml ice cold PBS (no calcium, no magnesium, pH 7) containing 50 µg/ml cycloheximide (final conc. 36 mM) and 5.5 µl of Ribolock RNase inhibitor (40U/L). The cell suspension was incubated on ice for 10 min to preserve stalled ribosomes with associated mRNAs on rough endoplasmic reticulum (rER). After incubation, cells were centrifuged at 300 g for 10 min at 4˚C, and the supernatants were discarded without disturbing the pellet. One cell pellet (~1.5 million cells) from each sample was stored on ice for preparation of microsomes, while the remaining cell pellets were quick-frozen in liquid nitrogen and stored at -80˚C.

3.1.5 Buffer compositions for preparation of rER microsomes

All buffer components used in the microsome preparation protocol were purchased from Sigma-Aldrich. Buffers were prepared to the final concentrations described below, filter-sterilized using 0.2 µm filters, aliquoted in 1.5 ml DNA LoBind tubes and stored at -80˚C for long term.

3.1.5.1 Preparation of High-Density Buffer (HD buffer)

Component	Final concentration
Nuclease-free water	N/A
HEPES-KOH pH 7.2	25 mM
KOAc	110 mM
$Mg(OAc)_2$ (see note 1)	5 mM
EGTA pH 8.0 (see note 2)	1 mM
Sucrose (w/w)	25 % = 0.81 M
Glycerol (v/v)	5 %
DTT	1 mM
cOmplete EDTA-free	1 x
Cycloheximide (see note 3)	0.1 mg/ml

The HD buffer was filter-sterilized using 0.22 µm filters, 400 µl buffer was aliquoted into fresh 1.5 ml DNA LoBind tubes, quick-frozen in liquid nitrogen and stored at -80˚C.

3.1.5.2 Preparation of Wash Buffer

Component	Final concentration
Nuclease-free water	N/A
HEPES-KOH pH 7.2	25 mM
KOAc	110 mM
$Mg(OAc)_2$ (see note 1)	2.5 mM
EGTA pH 8.0 (see note 2)	1 mM
DTT	1 mM
cOmplete EDTA-free	1 x
Cycloheximide (see note 3)	0.1 mg/ml

The wash buffer was filter-sterilized using 0.22 µm filters, 550 µl buffer was aliquoted into fresh 1.5 ml DNA LoBind tubes, quick-frozen in liquid nitrogen and stored at -80˚C.

Note 1: The nuclei of some cell lines tend to be more sensitive to low magnesium levels relative to other cell lines tested during hypotonic lysis. This can lead to rupture of the nuclear membranes and cause contamination of rough microsome fractions with DNA. This can be easily remedied by adjusting the Mg concentration of hypotonic lysis solution to 5 mM Mg(OAc).

Note 2: 10 mM EGTA was used in original protocol for plants[126,127], this was adjusted to 1 mM for mammalian cells. 10 mM EDTA was also part of the original protocol for plants, this was omitted for mammalian cells.

Note 3: To recover polyribosomes, add a final concentration of 0.1 mg/ml (0.36 mM) cycloheximide fresh to all buffers. 100x cycloheximide can be prepared as a filer-sterilized 10 mg/ml stock solution in water[128] and stored in aliquots at -80˚C.

3.1.6 Microsome preparation for enrichment of rough endoplasmic reticulum

This protocol was modified from a microsome preparation protocol used for plant material[127,129]. All pipetting steps were performed on ice and centrifugation steps were carried out at 4˚C using Eppendorf centrifuge 5810R with fixed angle rotar F-45-30-11. Protein LoBind tubes (1.5 ml) were used in all steps to reduce the cell/microsomes adhesion to the tube walls.

3.1.6.1 First protocol for microsome preparation

Note 1: Before proceeding to the microsome preparation protocol, unfreeze one 400 µl aliquot of HD buffer (Chapter 3.1.5.1), add 1.2 µl digitonin (5 % w/v; 1:333 dilution) and 0.8 µl RNase inhibitor (40 U/µl; 1:500 dilution)

Note 2: Before proceeding to washing step in microsome preparation protocol, unfreeze one 550 µl aliquot wash buffer (Chapter 3.1.5.2), add 0.44 µl of digitonin (5 % w/v; 1:1250 dilution) and 1.1 µl RNase inhibitor (40 U/µl; 1:500 dilution)

Following cycloheximide treatment of $CD19^+$ B cells and ARH-77 cells (chapter 3.1.4), rER microsomes from each cell pellet (corresponding to ~$1x10^6$ cells) were

prepared in three steps: (i) cell and organelle lysis, (ii) enrichment of microsomes and (iii) washing of microsomes.

1) Cell and organelle lysis

Cell pellets ($CD19^+$ B cells and ARH-77 cells) from each sample (B cells from before and after vaccination) were immediately suspended in 120 µl of high-density buffer containing digitonin (5 % w/v; 1:333) and RiboLock RNase inhibitor (40 U/µl; 1:400). The cell suspension was pipetted up/down 15 times and incubated on ice for 5 min to allow cell and organelle lysis. After incubation, 2x 55 µl of homogenates from each sample were transferred into fresh 1.5 ml Protein LoBind tubes and centrifuged at 600 g for 3 min at 4˚C to pellet nuclei and cell debris.

2) Enrichment of microsomes

After centrifugation of homogenates, supernatants (~40 µl) containing the membranes and cytosolic fractions were transferred into pre-chilled fresh 1.5 ml Protein LoBind tubes, and sucrose concentrations was diluted to 0.37-0.40 M (12-13% w/w) by addition of equal volumes of nuclease-free water. Subsequently, the microsomes suspension was mixed by pipetting up/down 5 times followed by centrifugation of samples at 20,800 g for 90 min at 4˚C.

3) Washing of microsomes

After centrifugation, the supernatants containing cytosol were discarded and the microsomal pellets (invisible to the naked eye) were resuspended in 85 µl of wash buffer containing digitonin (5% w/v; 1:1250) and RiboLock RNAse inhibitor (40 U/µl; 1:500). Then, the microsome suspension was gently mixed by pipetting up/down 10 times, centrifuged at 20,800 g for 45 min at 4˚C, supernatants were discarded and the resulting microsome pellet was stored on ice.

In total two microsome pellets were prepared per sample, where each pellet corresponded to ~500,000 B cells. As an internal control, ~10% ARH-77 cell microsomes (corresponding to ~50,000 cells) were spiked to one B cell microsome

pellet. For immediate use in emulsion RT-PCR assembly reactions (Chapter 3.2.2), freshly prepared B cell rER microsomes and B cells with 10% ARH-77 spike-in samples were resuspended in 24 µl of wash buffer containing digitonin and RNAse inhibitor and were stored on ice until further use. For long term storage (e.g. ARH-77 cells microsomes), the enriched microsomes were quick-frozen in liquid nitrogen and stored at -80˚C.

3.1.6.2 Optimized protocol for microsome preparation

Note 1: Before proceeding to the microsome preparation protocol, unfreeze one 400 µl aliquot of HD buffer (Chapter 3.1.5.1), add 1.2 µl digitonin (5 % w/v; 1:333 dilution) and 4 µl RNase inhibitor (40 U/µl; 1:100 dilution)

Note 2: Before proceeding to washing step in microsome preparation protocol, unfreeze one 550 µl aliquot wash buffer (Chapter 3.1.5.2), add 0.44 µl of digitonin (5 % w/v; 1:1250 dilution) and 5.5 µl RNase inhibitor (40 U/µl; 1:500 dilution)

Following cycloheximide treatment (Chapter 3.1.4), microsomes were freshly prepared from 1 million ARH-77 cells and 1.5 million $CD19^+$ B cells from both before (day 0) and 7 days after Td booster vaccination samples. This protocol follows the similar experimental conditions as mentioned in Chapter 3.1.6.1 with minor modifications in the following steps:

1) Cell and organelle lysis

In this step, the cell suspension was incubated for 10 min on ice to allow cell and organelle lysis.

2) Enrichment of microsomes

In this step, the centrifugation of samples was performed at 20,800 g for 2 hours.

3) Washing of microsomes

The centrifugation of samples was performed at 20,800 g for 1 hour.

In total, two microsome pellets were prepared per sample, where each pellet corresponded to ~750,000 B cells. As an internal control, ~0.5% ARH-77 cell microsomes (corresponding to ~3,800 cells) were spiked to one B cell microsome pellet. For immediate use in emulsion RT-PCR assembly reactions (Chapter 3.2.4), freshly prepared B cells rER microsomes and B cells with 0.5% ARH-77 spike in samples from both before and after Td booster vaccination were resuspended in 24 µl of wash buffer containing digitonin and RNAse inhibitor and were stored on ice until further use. For long term storage (e.g. ARH-77 cells microsomes), the enriched microsomes were quick-frozen in liquid nitrogen and stored at -80˚C.

3.2. Paired immunoglobulin heavy and light chain amplification and DNA purification methods

The following protocols were developed using total RNA obtained from human peripheral blood leukocytes to link and amplify human immunoglobulin (Ig) heavy chain (HC) and light chain (LC) mRNAs. These protocols were later adapted for application with freshly prepared B cell/ARH-77 cell rER microsome-associated mRNAs (Chapter 3.1.6) to retain endogenous HC-LC pairs from each individual B cell/ARH-77 cell. The RT-PCR assembly and nested primers used in the following protocols were designed and modified from prior studies[124,130] by Dr. Hans-Jörg Warnatz. The primer sequences are listed in Appendix table 1. Two different overlap extension emulsion RT-PCR assembly reactions using different oil components were tested (herein described as first protocol and optimized protocol).

3.2.1 First protocol; overlap-extension emulsion RT-PCR assembly using leukocyte total RNA

All steps were performed on ice and the emulsions were generated according to a previously described method[131]. Human peripheral blood leukocyte total RNA (100 ng/µl) was used as starting material and RT-PCR master mix (aqueous phase) was prepared using dART 1-step RT-PCR Kit following the manufacturer's instructions.

The assembly primer mixes (IgMG-VH1…7-VK1…6-IgKC) were prepared to a final concentration as described in **Appendix table 1**.

The oil phase was prepared to a final volume of 25 ml in 50 ml centrifuge tube composed of 95.05% mineral oil, 4.5% Span 80, 0.4% Tween 80 and 0.05% Triton X-100. The oil phase was mixed thoroughly by continuous stirring using a magnetic stirrer (4.5 x 12 mm) and was prepared freshly for every use. The aqueous phase prepared to 50 µl reaction volumes consisted of: 1x master buffer (Roboklon), 0.5 µg/ml BSA, 100 µg/ml cycloheximide, assembly primer mixes (IgMG-VH1…7-VK1…6-IGKC; final concentrations are described in Appendix table 1), 4 ng/µl leukocyte RNA and 1x master enzyme mix* (Roboklon; master enzyme mix includes dART Reverse Transcriptase, OptiTaq DNA Polymerase and Ribonuclease Inhibitor; see below). Next, water-in-oil emulsions were prepared by drop-wise addition of 50 µl of the aqueous phase to 800 µl of oil phase placed in 50 ml centrifuge tube. The drop-wise addition step was performed over a period of 7 mins, while the mixture was continuously stirred on Gerhardt stirrer with output 10. After water-in-oil emulsion formation, the mixture was continuously stirred for additional 3 min, and the emulsions were aliquoted into 6 PCR tubes, each containing 100 µl of emulsion mix. Then, the reaction tubes were placed in a thermocycler and following thermal cycling conditions were used:

Step	**Temp [˚C]**	**hh:mm:ss**	**Cycles**
{cDNA synthesis	55	00:30:00}	1x
{RT inactivation	95	00:03:00}	1x
{Denaturation	95	00:00:20	
Annealing	56	00:00:30	
Extension	72	00:02:00}	3x
{Denaturation	95	00:00:20	
Annealing	56	00:00:30	
Extension	72	00:04:00}	20x
{Final extension	72	00:05:00}	1x
{Storage	12	Forever}	1x

After thermal cycling, the emulsion reactions were visually inspected for flocculation and creaming to ensure emulsion stability. The emulsions were broken using isobutanol and DNA was recovered as mentioned in Chapter 3.2.7. Afterwards, the purified DNA was loaded onto 1% agarose gel and the gel run was performed at 100 V for 45 mins. HC-LC assembled products of ~740 bp were size-selected using X-tracta gel extraction tools and DNA from the size-selected agarose gel slices was purified using the Pellet paint NF Co-precipitant protocol as described in Chapter 3.2.8.

* In another experiment following similar experimental conditions, higher amounts of dART 1-Step RT-PCR master enzyme (final conc. 2x) was added to the aqueous phase and RT-PCR reaction and emulsion DNA recovery was performed as described above. Later, DNA from the size-selected agarose gel slices was purified using Zymo Gel DNA recovery Kit, following the manufacturer's instructions.

3.2.2 First protocol; Overlap-extension emulsion RT-PCR assembly using freshly prepared microsomes (B cells + 10% ARH-77 cell microsomes)

The above mentioned emulsion RT-PCR assembly (Chapter 3.2.1) protocol was adapted for application with freshly prepared microsomes with slight modifications as follows:

The aqueous phase for each emulsion RT-PCR reaction with microsomes prepared to a final volume of 200 µl contained: 1x master buffer (Roboklon), 0.5 µg/ml BSA, 100 µg/ml cycloheximide, assembly primer mixes (IgMG-VH1…7-VK1…6-IGKC; final concentrations are described in Table 3) and 1x master enzyme mix (Roboklon; master enzyme mix includes dART Reverse Transcriptase, Opti Taq DNA Polymerase and Ribonuclease Inhibitor). Then, to each independent RT-PCR reaction (aqueous phase), 20 µl of freshly prepared microsome pellets from each sample (Chapter 3.1.6.1) was added. Water-in-oil emulsions and RT-PCR reactions were performed as mentioned above (Chapter 3.2.1). The emulsions were broken using isobutanol and DNA was extracted as described in Chapter 3.2.7 and Chapter 3.2.8.

3.2.3 Optimized protocol; overlap-extension emulsion RT-PCR assembly using leukocyte total RNA

All steps were performed on ice and emulsions were generated using the Micellula DNA Emulsion & Purification Kit. Experimental conditions were optimized using human peripheral blood leukocytes total RNA (100 ng/µl), and RT-PCR master mix (aqueous phase) was prepared using the dART 1-step RT-PCR Kit following the manufacturer's instructions. The assembly primer mixes (IgMG-VH1…7-VK1…6-IgKC) were prepared to a final concentration as described in Table 3. This protocol was modified from the manufacturer's manual according to the experimental needs and was set up under the following conditions:

The oil phase was prepared to a final volume of 840 µl in a 50 ml centrifuge tube composed of ~73% Emulsion component T-1, ~7% Emulsion component T-2 and ~20% Emulsion component T-3. The oil phase was mixed thoroughly by continuous stirring using a magnetic stirrer (4.5 x 12 mm) and was prepared freshly for every use. The aqueous phase prepared to 50 µl reaction volumes consisted of: 1x master buffer (Roboklon), 0.5 µg/ml BSA, 100 µg/ml cycloheximide, assembly primer mixes (IgMG-VH1…7-VK1…6-IGKC), 4 ng/µl leukocyte RNA and 2x master enzyme mix (Roboklon; master enzyme mix includes dART Reverse Transcriptase, OptiTaq DNA Polymerase and Ribonuclease Inhibitor). Next, water-in-oil emulsions were prepared as described in Chapter 3.2.1 and following thermal cycling conditions were used:

Step	**Temp [˚C]**	**hh:mm:ss**	**Cycles**
{cDNA synthesis	55	00:30:00}	1x
{RT inactivation	95	00:03:00}	1x
{Denaturation	95	00:00:20	
Annealing	56	00:00:30	
Extension	72	00:02:00}	3x
{Denaturation	95	00:00:20	
Annealing	56	00:00:30	
Extension	72	00:04:00}	20x
{Final extension	72	00:05:00}	1x
{Storage	12	Forever}	1x

After thermal cycling, the emulsion RT-PCR reactions were visually checked to ensure emulsion stability. The emulsions were broken using isobutanol and DNA was extracted as described in 3.2.7. Afterwards, the purified DNA was loaded onto 1% agarose gel and a gel run was performed at 100 V for 45 mins. HC-LC assembled products of ~740 bp were size-selected using X-tracta gel extraction tool and DNA from the size-selected agarose gel slices was purified using a Zymo Gel DNA Extraction Kit following the manufacturer's instructions.

3.2.4 Optimized protocol; overlap-extension emulsion RT-PCR assembly using freshly prepared microsomes (B cells + 0.5% ARH-77 cell microsomes)

The above mentioned emulsion RT-PCR assembly (Chapter 3.2.3) protocol was adapted for application with freshly prepared microsomes with slight modifications as follows:

The RT-PCR master mix (aqueous phase) prepared to a final volume of 100 µl contained 1x master buffer mix, 0.5 µg/ml BSA, 100 µg/ml cycloheximide, 2x master enzyme mix (Roboklon; master enzyme mix includes dART Reverse Transcriptase, OptiTaq DNA Polymerase and Ribonuclease Inhibitor) and isotype-specific HC and LC assembly primers (IgM-VH1…7-VK1…6-IgK; IgG-VH1…7-VK1…6-IgK). The primer set for overlap extension RT-PCR assembly reaction is mentioned in Appendix table 1. Then, 20 µl of freshly prepared microsomes from each sample (Chapter 3.1.6.2) was passed into water-in-oil emulsion droplets for assembly and amplification in two separate reaction tubes with primers specific for IgM and IgG isotypes, respectively. The emulsion RT-PCR assembly was performed as mentioned in Chapter 3.2.4. The emulsion DNA recovery and subsequent DNA recovery from agarose gel slices were performed as mentioned above (Chapter 3.2.3).

Note: In this experiment, B cells microsomes suspensions spiked with 0.5% ARH-77 cell microsomes were added to an aqueous phase containing IGG-VH1…7-VK1…6-IgK assembly primer mix (as ARH-77 cells express IGG and IGκ sequences).

3.2.5 Nested PCR protocols

First protocol; Nested PCR amplification

All steps were performed on ice and PCR master mix was prepared using a Q5 High-Fidelity DNA Polymerase Kit following the manufacturer's instructions. The nested primer sequences (IgM_n, IgG_n, IgK_n) and final concentrations used in nested PCR reactions are listed in Appendix table 1 and PCR reactions were set up as follows:

The nested PCR master mix prepared to a final volume of 50 µl contained Q5 Reaction Buffer (1x), 200 µM dNTP mix, nested primer mixes (0.4 µM each; IgM_n+IgK_n and IgG_n+IgK_n), Q5 High-Fidelity DNA Polymerase (0.02 U/µl) and 3 µl of emulsion purified HC-LC assembled DNA (Chapter 3.2.1 and 3.2.2). PCR amplification was performed under following conditions:

Step	Temp [˚C]	hh:mm:ss	Cycles
{Initial Denaturation	98	00:01:00}	1x
{Denaturation	98	00:00:07	
Annealing/Extension	71	00:00:30}	35x
{Final Extension	72	00:05:00}	1x
{Hold	12	Forever}	1x

After PCR amplification, 12 µl of amplified products were analyzed on 1% agarose gel. Samples demonstrating bands were purified from the remaining PCR mixture using Agencourt AMPure XP beads (Chapter 3.2.9). Subsequently, Illumina ChIP TruSeq DNA libraries were prepared (Chapter 3.3.1) for 2x250 bases paired end Illumina MiSeq sequencing.

Optimized protocol; Nested PCR amplification

All steps were performed on ice and master mix was prepared as mentioned above (First protocol; Nested PCR amplification) using emulsion purified HC-LC assembled

DNA from Chapter 3.2.3 and 3.2.4 and the thermal cycling conditions were as follows:

Step	Temp [˚C]	hh:mm:ss	Cycles
{Initial Denaturation	98	00:03:00}	1x
{Denaturation	98	00:00:30	
Annealing/Extension	71	00:01:00}	34x*
{Final Extension	72	00:05:00}	1x
{Hold	12	Forever}	1x

* To determine the lowest number of PCR cycles and reduce subsequent mutations introduced during PCR amplification, samples of 12.5 µl were collected after the annealing/extension step in cycles 28, 31 and 34 and were visualized on 1% agarose gel.

Samples demonstrating bands with lowest cycle number were determined and nested PCR amplification was repeated once again with the lowest cycle number for large-scale amplification of the assembled DNA. Subsequently, the samples were purified from the reaction mixture using AMPure beads (Chapter 3.2.9) and Illumina ChIP TruSeq libraries were prepared (Chapter 3.3.1) for 2x250 bases paired end sequencing on the Illumina MiSeq sequencing platform.

3.2.6 Analytical PCR for verification of Ig HC and LC diversity

To verify the B cell repertoire diversity in the HC-LC assembled DNA, an analytical PCR amplification was performed using primers specific for variable heavy (VH) and variable kappa (VK) families (VH1…7 and VK1…6) and isotype-specific nested primers. The VH and VK primer sequences and isotype-specific nested PCR primers are listed in Appendix table 1.

Briefly, 0.2 ng/µl purified nested PCR products from Chapter 3.2.5 were added to each 25 µl analytical PCR reaction tube containing Q5 Reaction Buffer (1x); 200 µM dNTP mix; and 0.02 U/µl Q5 High-Fidelity DNA Polymerase. Primer pairs corresponding to 0.4 µM each of variable family and isotype specific nested primers

were added to each reaction tube separately (IgM_n+VH1/7, IgM_n+VH2, IgM_n+VH3, IgM_n+VH4, IgM_n+VH5, IgM_n+VH6; IgG_n+VH1/7, IgG_n+VH2, IgG_n+VH3, IgG_n+VH4, IgG_n+VH5, IgG_n+VH6 and IgK_n+VK1, IgK_n+VK246, IgK_n+VK5, IgK_n+VK6) and thermal cycling was performed under following conditions:

Step	**Temperature [˚C]**	**hh:mm:ss**	**Cycles**
{Initial Denaturation	98	00:01:00}	1x
{Denaturation	98	00:00:07	
Annealing/Extension	71	00:00:20}	15x
{Final Extension	72	00:05:00}	1x
{Hold	12	Forever}	1x

After PCR amplification, 15 µl of all samples were visualized on 1.5% agarose gel.

3.2.7 Emulsion breaking and DNA recovery

DNA purification from emulsion PCR samples (Chapter 3.2.1…4) was performed as previously described[132] using Zymo Clean & Concentrator Kit, with some modifications. Briefly, two emulsion PCR samples were pooled together into one 1.5 ml DNA LoBind tube and 1 ml of isobutanol was added to each tube followed by short vortexing (~15 seconds) until the solution was clear. Next, 200 µl of Zymo DNA binding buffer was added to the broken emulsions and were mixed with gentle agitation for 2 min on a shaker. Then, the oil and aqueous phase was separated by centrifugation at 16,000 g for 2 min. The upper oil phase (~1 ml) was discarded and the resulting aqueous phase was loaded onto a Zymo spin column for purification of amplified DNA following manufacturer's instructions. The remaining washing steps were performed according to the manufacturer's instructions and DNA was eluted in 6 mM Tris-Cl pH 8 in sterile 1.5 ml DNA LoBind tubes.

3.2.8 DNA recovery from agarose gel slices using Pellet Paint NF Co-Precipitant

DNA from size-selected agarose gel pieces (Chapter 3.2.1 and 3.2.2) were extracted following pellet paint NF Co-precipitant protocol according to the manufacturer's instructions with some modifications. Briefly, >3 volumes of agarose-dissolving buffer (ADB) from the Zymo Gel DNA extraction kit was added to one volume of agarose excised from the gel and was incubated at 25˚C for 15 minutes with shaking at 750 rpm until the gel slice was completely dissolved. To the dissolved agarose gel, 2 µl of Pellet Paint NF Co-Precipitant was added to each sample followed by addition of 0.1 volumes of 3M Sodium Acetate and short vortexing. Next, two volumes of 70% Ethanol were added to the samples, vortexed briefly, incubated at RT for 2 min followed by centrifugation of the samples at 15,800 g for 5 min. After centrifugation, the supernatant was discarded, and the pellet (Violet color) was rinsed with 70% ethanol followed by brief vortexing and centrifugation at 15,800 g for 5 min. The supernatant was discarded and the pellet was rinsed once again with 500 µl of 100% ethanol and was centrifuged at 15,800 g for 5 min. Next, the pellet was air-dried for 5 min, suspended in the desired volume of 6 mM Tris-Cl buffer at pH 8, frozen on dry ice and stored at -20˚C until further use.

3.2.9 DNA purification using Agencourt AMPure XP beads

For maximum recovery of PCR-amplified amplicons (Chapter 3.2.5) and to remove primer dimers (<100 bp), salts and enzymes, I employed a simple DNA purification method using Agencourt AMPure XP beads. Briefly, carrier DNA (~10 ng) was added to the sample volumes to enhance DNA recovery. The Agencourt AMPure XP bead suspension was vortexed until the beads were thoroughly dispersed. Beads were added at a ratio of 1:1.2 to the total sample volume and mixed by pipetting 10 times. Briefly, the samples were incubated for 15 min at RT followed by 5 min incubation on magnetic rack. After incubation, the supernatants were discarded and the beads were washed two times with 80% ethanol, followed by air-drying the samples for 7 minutes at RT. Next, samples were removed from magnetic rack, and DNA was eluted by addition of 20 µl of Qiagen elution buffer (EB) to each sample

and were mixed thoroughly by pippeting 10 times. Samples were then incubated for 2 minutes at RT followed by 5 minutes' incubation on magnetic rack for separation of beads from the solution. The supernatants containing the eluted DNA (~18 µl) was collected in fresh 1.5 ml DNA LoBind tubes and were stored at -20˚C.

3.2.10 Quantification of DNA

DNA and RNA concentrations and purity in a mixture were determined at absorbance 260 nm using a NanoDrop spectrophotometer and a Qubit 3.0 Fluorometer following the manufacturer's manual.

3.3 Illumina library preparation, MiSeq Sequencing and bioinformatic analysis

3.3.1 Library preparation for MiSeq sequencing

AMPure-purified nested PCR samples of ~200 ng were used as input and sequencing libraries were prepared following the Illumina TruSeq ChIP Sample Preparation guide according to manufacturer's instructions. Briefly, nested PCR samples of ~710 bp were first subjected to end repair using an End Repair Mix followed by adenylation of 3' ends. Next, to enable multiplexing of several samples, different Illumina barcode adapters were ligated to the ends of the DNA fragments for hybridization of multiplexed samples onto a flow cell. AMPure bead cleanup was performed (Chapter 3.2.9) after end repair and adapter ligation for removal of exonucleases and unligated adapters. Verification of adapter ligation to DNA fragments and determination of the number of PCR cycles to retain library representation was performed by qPCR. Each qPCR master mix prepared to a final volume of 20 µl contained: 2x SYBR Green Master Mix (10 µl), 10 µM Primer 1 (0.8 µl) and 10 µM Primer 2 (0.8 µl), 1 µl of adapter-ligated DNA sample. Thermal cycling conditions for qPCR were as follows: Initial denaturation at 95˚C for 10 mins followed by 40 cycles of denaturation at 95˚C for 15 seconds, annealing at 65˚C for 30 seconds and extension at 72˚C for 30 seconds followed by final extension at 72˚C for 5 minutes. Next, DNA fragments possessing adapter molecules on both sides

were selectively enriched by PCR amplification. The PCR master mix prepared to a final volume of 50 µl contained: 5 µl of PCR Primer Cocktail, 25 µl of PCR Master Mix (Supplied by Illumina) and 20 µl of AMPure-purified adapter-ligated sample. Thermal cycling was performed under following conditions: Initial denaturation at 98˚C for 30 seconds followed by 10-20 (depending on qPCR results) of denaturation at 98˚C for 10 seconds, annealing at 60˚C for 30 seconds, extension at 72˚C for 30 seconds. A final extension step was performed at 72˚C for 5 min. The PCR-amplified products were purified using AMPure beads and were size-separated on 1% Low-range Ultrapure agarose gel. The adapter-ligated samples of ~830 bp were size-selected and DNA was purified using the Qiagen MinElute Gel Extraction Kit following the manufacturer's instructions.

3.3.2 Illumina MiSeq 2x250 bases paired end sequencing

Illumina MiSeq 2x250 bases paired end sequencing was performed by a dedicated sequencing team in our laboratory (Alexander Kovacsovics and Matthias Linser). Briefly, the procedure involves dilution of all gel-purified adapter-ligated sequencing libraries (Chapter 3.3.1) to 10 µM concentrations, creating a sequencing library pool and determining the cluster density for loading onto the flow cell by performing qPCR, and a final denaturation step with 1 mM NaOH for library hybridization to the flow cell. The final products of this procedure were then sequenced on the Illumina MiSeq sequencing platform. The adapter-ligated libraries were demultiplexed using the Generate FASTQ workflow on the MiSeq sequencing machine, and the output was produced in FASTQ file format.

3.3.3 Bioinformatics analysis for analysis of paired Ig repertoire sequencing data

The bioinformatic pipeline used in this thesis to process Illumina MiSeq 2x250 bases paired end sequencing data was written and executed by Dr. Hans-Jörg Warnatz. A schematic flow chart summarizing the bioinformatic pipeline is shown in (Figure 16). Briefly, MiSeq sequencing data results in FASTQ format which is a text-based format used for storing nucleotide sequence and its Phred quality scores, in ASCII

characters. Conventionally, each FASTQ sequence entry is represented by four lines where the first line starts with the "@" character followed by a sequence identifier, the second line contains the nucleotide sequence, the third line starts with a "+" character and the fourth line contains the Phred quality scores of the nucleotide sequence in line 2. The bioinformatic pipeline first performs quality filtering on raw 2x250 MiSeq sequencing data for minimum Phred quality scores of 20 over 50% nucleotide identity using the FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) to ensure high quality reads in the CDR3-containing region of the paired HC and LC sequences. Then, the reverse complement sequences of of the filtered FASTQ entries were generated using the FASTX toolkit, followed by creating two tab-delimited text files of the individual sequence reads (READ_1 and READ_2) and their identifiers. After these steps, READ_1 And READ_2 sequences with the same identifier were combined into a single tab-delimited text file followed by annotating the combined sequences based on isotype-specific constant regions (e.g. IgM, IgG and Igκ) to determine the antibody isotypes of the heavy and light chains. Later, the read pairs annotated with HC-HC or LC-LC structures or without constant region match(es) were filtered out. After these filtering steps, the annotated HC-LC paired end reads were converted back into FASTQ format and used as input file to annotate V, D, J and C genes of B cell receptors using the MiXCR software[109]. The MiXCR software can efficiently handle millions of paired-end reads by accurately aligning them to reference human germline V, D, J and C gene sequences from IMGT. The MiXCR work flow contains three main processing steps, (i) alignment of reads to reference human V, (D), J and C germline genes by considering sequence quality and mismatches, (ii) Extraction of CDR3 regions and clustering of reads with identical V, (D), J and C genes as well as CDR3 sequences into clonotype clsuters including correction of PCR and sequencing errors, and (iii) exporting the final aligned HC and LC clonotypes into tab-delimited human-readable text files. After processing the data using the MiXCR software, the pipeline was further extended to create a pairing statistics file containing separate statistics for each heavy chain isotype to examine the frequency of VH-VK gene families and to compare B cell clonotype frequencies among different samples. To visualize the pairing statistics data, a heat map was generated using R and was graphically represented using

ggplot2. The R script used to generate heat maps was written and executed with assistance from Sören Matzk and is presented in the Appendix of this thesis.

3.3.4 Determination of most frequent LC-CDR3 sequence paired to a given HC-CDR3 sequence

Analysis of reads with MiXCR software results in output files containing all V, (D), J and C annotated heavy and light chain sequences in separate files. However, this analysis does not give the information on paired HC-LC sequences. In order to obtain the information on the most frequent LC-CDR3 sequences paired to a given HC-CDR3 sequence, and to calculate the pairing accuracies, a custom UNIX shell script was used (shown in the Appendix of this thesis) which results in an output in text file format. Pairing accuracies (%) were calculated by dividing total number of reads for a given HC-CDR3 sequence with total number of reads for most frequently paired LC-CDR3 sequence (for e.g. Top1).

3.4 Antibody selection, cloning, expression and binding studies for discovery of monoclonal antibodies

3.4.1 Selection, PCR strategy and primer design for generation of full-length paired HC-VDJ and LC-VJ sequences

Fourteen highly induced HC-LC pairs (Refered as Top1 to Top14 clonotypes) obtained from IgG sequencing post Td booster immunization were selected based on their frequency, pairing accuracy, and fold difference between Top1 LC-CDR3 and Top2 LC-CDR3 paired to a given HC-CDR3 sequence. Next, a two-step PCR based amplification method (Figure 24) was designed for selective amplification of highly induced clonotypes and to incorporate restriction digestion sites at 5’ and 3’ ends to generate HC and LC amplicons with complete V(D)J sequence for efficient cloning into respective expression vectors. Clonotype-specific HC and LC forward and reverse primers were designed according to the following steps:

1) First, 2x250 bases sequence information of fourteen selected clonal pairs obtained from MiSeq sequencing was provided as input sequence to the IMGT/V-QUEST program[133] for annotation of HC and LC V(D)J germline gene sequences.

2) First step RT-PCR primers for selected HC-LC pairs were designed such that the forward primer encodes 22 nucleotides of the respective V gene family-specific FR1 region sequence (sequence information obtained from IMGT/V-QUEST) along with 5 nucleotides spacer, BssHII restriction site, and 6 nucleotides spacer overhang at the 3' end. The reverse primers encoded a part of the respective CDR3 sequence and 18 nucleotides of the respective FR4 region (sequence information annotated by IMGT/V-QUEST). The forward and reverse primers used in the first-step RT-PCR are listed in Appendix table 2 and the primers melting temperatures range between 62-68˚C.

3) Second-step PCR primers for selected HC and LC pairs were designed such that the forward primers encoded the 6 nucleotide spacer, BssHII restriction site, 5 nucleotide spacer and 3 nucleotides of the respective V region. The reverse primers encoded the complete respective FR4 region (sequence information annotated by IMGT/V-QUEST) with a BssHII/NheI restriction site at the 3' end. The forward and reverse primers used in this step are listed in Appendix table 2 and the primer melting temperatures range between 65-67˚C.

4) All the primers were checked for self- or hetero-dimerization and for internal hairpin formation using the OligoAnalyzer 3.1 tool from IDT (Integrated DNA Technologies).

3.4.2 Total RNA isolation and DNase treatment

Frozen $CD19^+$ B cells from post-Td booster immunization were used for extraction of total RNA using TRIzol reagent from Invitrogen. The TRIzol RNA purification protocol mentioned here was slightly modified from the original version.

$CD19^+$ B cells corresponding to ~1.5 million cells were thawed on ice and 1 ml of TRIzol was added to the cells. Homogenization of cells was performed by passing the cells 2-3 times through a 1 ml syringe using a 26G needle followed by incubation of samples for 5 minutes at RT. Next, 200 µl of cold chloroform was added to the homogenized samples, strongly agitated for 15 secs, incubated at RT for 3 min and then centrifuged at 12,000 g for 15 min at 4˚C. The upper aqueous phase (~500 µl) was transferred into fresh 1.5 mL Protein LoBind tubes, 1 µl Glycogen and 500 µl of Isopropanol was added to the tubes, incubated for 10 min at RT, and then centrifuged at 12,000 g for 10 min at 4˚C. The supernatant was completely removed and 70% ethanol was added to the pellet and centrifuged once again at 12,000 g for 5 min at 4˚C. Then, the supernatant was completely removed, the pellet was air dried and RNA was dissolved in 90 µl of RNase free water. The isolated RNA was further subjected to DNase treatment with the Turbo DNA Free Kit for removal of contaminating DNA, DNase and divalent cations.

DNase treatment was performed by addition of 10 µl of 10x Turbo DNase Buffer and 1 µl of Turbo DNase (2 U/µl) to 90 µl of total RNA. The reaction was mixed by flicking the tube and was incubated in water bath at 37˚C for 20 min. For inactivation of DNase activity, 10 µl of DNase Inactivation Reagent was added to the reaction tube, incubated at 25˚C for 5 min at 1000 rpm and centrifuged for 1.5 minutes at 10,000 g. The supernatant was transferred into a fresh 1.5 ml DNALoBind tube and the reaction volume was adjusted to 150 µl with RNase free water. Next, 15 µl of 3 M NaOAc and 1µl Glycogen were added to the reaction tube and briefly vortexed. Ice-cold ethanol (100%; ~375 µl) was added to the reaction tubes and incubated overnight at -20˚C. On the following day, the tubes were centrifuged at 14,000 rpm for 15 min at 4˚C and the supernatant was discarded. The pellet was washed with 500 µl 70% ethanol and centrifuged once again at 14,000 g for 5 min, the supernatant was discarded and the pellet was air-dried for 10-15 min. The air-dried pellet was resuspended in 12 µl of RNase-free water, concentrations were measured using a Qubit RNA measurement kit, the RNA was aliquoted in desired volumes into fresh Protein LoBind tubes, quick-frozen on dry ice and stored at -80˚C until further use.

3.4.3 First-step: RT-PCR amplification of selected HC and LC clonotypes

Total RNA isolated from frozen $CD19^+$ B cells after Td booster immunization (Chapter 3.4.2) was used as template and RT-PCR reactions for each selected clonotype were carried out in separate reaction tubes using the dART 1-step RT-PCR Kit. The RT-PCR master mix prepared to a final volume of 25 µl contained: 1x master buffer mix, 1x master enzyme mix (0.2 U/µl) and 0.4 µM each of forward and reverse primers for each selected clonotype (Appendix table 2). A total of ~400 ng of total RNA was distributed to all reaction tubes containing clonotype-specific primers and RT-PCR master mix. The thermal cycling conditions were as follows:

Step	Temp [˚C]	hh:mm:ss	Cycles
{cDNA synthesis	55	00:30:00}	1x
{RT inactivation	95	00:03:00}	1x
{Denaturation	95	00:00:20	
Annealing	56	00:00:30	
Extension	72	00:01:30}	23x
{Final extension	72	00:05:00}	1
{Storage	12	Forever}	1

After thermal cycling, the RT-PCR amplified products were purified by AMPure beads as previously described (Chapter 3.2.9)

3.4.4 Second-step: PCR amplification for generation of full-length V(D)J regions for cloning

In the second step, PCR amplification for each clonotype was performed in a separate reaction tube using AMPure-purified RT-PCR products from the above step (Chapter 3.4.3). The PCR master mix was prepared using the Q5 HF DNA Polymerase Kit and the master mix prepared to a 50 µl reaction volume contained: 1x Q5 reaction buffer, 200 µM of dNTP mix, 6 µl of AMPure-purified RT-PCR products, 0.4 µM each of forward and reverse primers for selected clonotypes

(Appendix table 2) and 0.02 U/µl Q5 HF DNA polymerase. The thermal cycling conditions were as follows:

Step	Temp [˚C]	hh:mm:ss	Cycles
{Initial denaturation	98	00:03:00}	1x
{Denaturation	98	00:00:30}	1x
{Annealing	69	00:00:30	
Extension	72	00:01:00}	16x
{Final extension	72	00:05:00}	1x
{Hold	12	Forever}	1x

After thermal cycling, samples were analyzed on a 1.5 % agarose gel and the DNA was gel-extracted using a Zymo Gel DNA Extraction Kit following the manufacturer's instructions.

3.4.5 Bacterial cell culture (*E. coli)*

E. coli strains containing plasmids (pCMV-CD40-4IE3-HC and pCMV-CD40-4IE3-LC) were inoculated into 200 ml 2x YT-Broth (16 g/l Tryptone, 10 g/l Yeast Extract and 5 g/l NaCl at pH 7) containing 100 µg/ml ampicillin and 2% (v/v) glucose. The culture medium was incubated overnight at 37˚C with continuous shaking at 220 rpm. To store bacterial cultures for long time, glycerol stocks were prepared by addition of 50% (v/v) glycerol to 50% of overnight grown culture in a sterile 2 ml cryopreservation tube, frozen on dry ice and stored at -80˚C.

3.4.6 Plasmid DNA isolation from *E. coli* cells

HC and LC plasmid DNA from cultured *E. coli* cells was isolated using a QIAprep Spin Miniprep Kit following the manufacturer's instructions. DNA elution was performed by addition of 40 µl EB (10 mM Tris-Cl, pH 8.5) to the center of each column, incubation for 1 min and centrifugation for 1 min at 13,000 rpm. DNA concentrations were measured using a NanoDrop fluorimeter, and the eluted DNA was stored at -20˚C.

3.4.7 Restriction digestion, ligation and transformation

HC and LC plasmid DNAs (Chapter 3.4.6) and agarose gel-purified full length HC and corresponding LC DNAs (Chapter 3.4.4) were double digested following the NEB double digestion protocol (https://www.neb.com/protocols/2014/05/07/double-digest-protocol-with-standard-restriction-enzymes) using appropriate restriction enzymes as described below :

Insert	Restriction enzymes	Corresponding vector	Restriction enzymes
HC	BssHII and NheI	pCMV-CD40-4IE3-HC	BsshII and NheI
LC	BssHII and HindIII	pCMV-CD40-4IE3-LC	BssHII and HindIII

After restriction digestion, samples were loaded on a 2% agarose gel, and the HC vector (pCMV-CD40-4IE3-LC ~5.9 kb) and the LC vector (pCMV-CD40-4IE3-LC, ~5.7 kb) were size-selected and the DNA was isolated using a Zymo Gel DNA purification kit following the manufacturer's instructions. The restriction-digested HC and LC inserts corresponding to ~370 bp and ~330 bp were purified using AMPure XP beads (Chapter 3.2.9) and the concentrations were measured using a Qubit 3.0 flurometer. After purification, the amounts (concentration) of vectors and corresponding inserts used for ligation were calculated using the formula obtained from http://molbi.de/protocols/ligation_v1_0.htm and ligation of vectors and inserts was performed using the Instant Sticky-End DNA Ligase Master Mix following the manufacturer's instructions. The ligation mixes were transformed into *E. coli* Top10 chemically competent cells (iba solutions) following the manufacturer's protocol. After transformation, the bacteria were plated on 2YT agar plates with ampicillin (1:1000) and incubated overnight at 37˚C.

3.4.8 Colony PCR for verification of ligation transformed clones

For verification of inserts in transformed colonies, a colony PCR was performed using insert-specific primers Appendix table 2 with annealing temperatures of ~69˚C. The PCR master mix prepared using Q5 High-Fidelity DNA polymerase Kit to a final volume of 25 µl contained: 1x Q5 reaction buffer, 200 µM dNTP mix, 0.02 U/µl Q5

HF DNA polymerase and 0.4 µM each of insert-specific forward and reverse primers. After preparation of the master mix, a single bacterial colony was picked using a 1 ml pipette tip, dipped for about 2 seconds in 25 µl PCR mix with isotype-specific primers, and transferred into 5 ml 2YT medium with ampicillin (1:1000) for replication of the picked colony. The 2YT medium was cultured overnight at 37˚C with continuous shaking at 180 rpm.

Thermal cycling for colony PCR were performed under following conditions:

Step	Temp [˚C]	hh:mm:ss	Cycles
{Initial denaturation	98	00:03:00	1x
{Denaturation	98	00:00:30	
Annealing	69	00:00:30	
Extension	72	00:01:00}	25x
{Final extension	72	00:05:00}	1x
{Hold	12	Forever}	1x

For verification of positive clones, 15 µl of all samples were analyzed on 1.5 % agarose gels and overnight grown bacterial cultures from positive clones demonstrating the correct insert size were used to prepare glycerol stocks and minipreps of plasmid DNA.

3.4.9 HEK293T cell culture

Human embryonic kidney 293T cells (HEK 293T cells) were cultured according to the instructions given by the American Type Culture Collection (ATCC). Complete medium for cell growth was prepared by addition of 10% Ultra-low IgG FCS and 1% P/S to high-glucose D-MEM medium. Thawing, growth and cryopreservation of HEK 293T cells was performed in a similar way to that for ARH-77 cells (Chapter 3.1.1), except that HEK 293T cells were transferred into T75 adherent flasks for culturing.

3.4.10 Transient transfection with polyethylenimine (PEI)

For transfection of plasmid DNA into mammalian HEK 293T cells, polyethylenimine (PEI) was used because of its low costs and easy handling. The protocol described

here is applied for transient transfection of paired HC and LC plasmids into HEK 293T cells in 6-well plates. For transfections in 150 mm^2 petri dishes, the cells, PEI: DNA mixture and medium were scaled up accordingly. The number of cells, plasmid DNA and PEI mixture used for transfection in 6-well plates and 150 mm^2 plates are mentioned in the table below.

Plate	Number of cells	Total DNA (µg)	PEI (µg)
6-well plate	$1x10^6$ cells	2	6
15 cm dish	$12x10^6$ cells	18	54

PEI transfection of HEK 293T cells with paired HC and LC plasmid DNA

Briefly, a day before transfection, $1x10^6$ HEK 293T cells were transferred into each well in a 6-well plate and cultured for ~24 hours in 2 ml high-glucose DMEM medium with 10% ultra-low IgG FCS and 1% P/S antibiotics, so that the cells were about 80-90% confluent. Prior to transfection, 2 µg of total plasmid DNA (1 µg of each HC and corresponding paired LC plasmid DNA) were diluted in 200 µl serum-free DMEM medium in a Protein LoBind tube. 6 µg pf PEI (3:1 ratio of PEI (µg): total DNA (µg)) were added to the reaction mixture, vortexed briefly, and incubated at RT for 15 minutes. After incubation, PEI:DNA mixtures from each sample were added drop-wise into each well with HEK 293T cells, followed by gentle swirling and incubation for 7 days at 37˚C with 5% CO_2. The cell culture supernatants (2 ml) were harvested on day 2, day 4 and day 6 and were immediately replaced with 2 ml DMEM medium with 10% ultra-low IgG FCS and 1% P/S antibiotics after each harvesting. The supernatants were stored at 4˚C and were used to perform indirect ELISA to identify antigen-specific monoclonal antibodies.

3.4.11 Enzyme Linked Immunosorbant Assay (ELISA)

A standard indirect ELISA protocol was used to detect antigen-specific monoclonal antibodies using harvested cell culture supernatants from day 4.

Briefly, Nunc-Immuno MicroWell 96-well plates were coated with Tetanus Toxoid (TT) antigen at a concentration of 10 µg/ml in 50 mM carbonate-bicarbonate buffer, pH 9.6, covered with adhesive plastic and incubated overnight at 4˚C. The coating solution was removed by washing the plates three times with 1x PBST (PBS + 0.1% Tween) and the plates were blocked with 400 µl of PTM (PBS + 0.1% Tween 20 + 2% Milk) for 2 hours 30 minutes at RT. After incubation, the plates were washed two times with 1x PBST and 120 µl of 1:2 pre-diluted transfected cell culture supernatants in PTM was added to sample wells. While, for positive control with TT antigen, 350 ng of mouse TT mAb (GeneTex, CA, USA) was added to control well. The samples were incubated at RT for one hour followed by three times washing of the plate with 1xPBST. After washing, 50 µl of 1:2000-diluted goat anti-human kappa LC-HRP secondary antibody was added to each sample well and mouse IgG HC-HRP conjugated secondary antibody was added to the control well. The plate was incubated for 2 min at RT, washed three with 1xPBST, and 50 µl of 1-step Ultra TMB substrate (Thermo Fisher Scientific, MA, USA) was added to the wells and incubated at 25˚C for 5 min. Reactions were stopped by addition of 50 µl 2M H_2SO_4 to each well, and O.D. values were measured at 450 nm using the GloMax Multi Detection System (Promega, Wisconsin, USA).

Each ELISA assay was performed in triplicates and the values were normalized by removing the background signals from the negative controls. Errors were represented as standard deviations from the mean (S.D.).

3.4.12 Purification of IgG antibodies from HEK 293T cell supernatants

This protocol utilizes protein A agarose beads that selectively bind to IgG monoclonal antibodies and its subclasses enabling the purification of IgG antibodies from HEK cell culture supernatants. A gravity flow column procedure was established using Micro Bio-Spin Columns, dialysis bags with MW cut-off of 3 kDa, 1x PBS, 100 mM Glycine-HCl at pH 2.2 and 1 M Tris buffer at pH 8 for high-yield recovery of IgG monoclonal antibodies. IgG monoclonal antibody purification was performed in the following steps:

1) Equilibration of HEK cell culture supernatants using dialysis bags:
Transfected HEK 293T cell culture supernatants of ~25 ml harvested on day 4 were transferred into dialysis bags with MW cut-off of 3 kDa and equilibrated overnight in 1x PBS (>50x sample volume).

2) Packing of Micro-Bio-Spin columns with IgA agarose resin:
Following the dialysis step, Micro-Bio-Spin columns were packed by addition of ~200 µl IgA agarose slurry (capacity to bind ~6 mg of IgG antibody) to the spin column. The residual buffer in the settled slurry was allowed to flow out by gravity and the column was washed two times with 1 ml 1x PBS.

3) Binding of IgG antibodies to agarose resin:
For binding of IgG antibodies to agarose resin, 25x 1 ml of dialyzed sample was applied to the spin columns and the flow through was collected. The binding step was repeated two more times by addition of flow through from the previous step to the packed columns, allowing maximal binding of IgG antibodies to the agarose resin. For removal of unbound proteins, the columns were washed 3x with 1 ml 1x PBS and the flow through was discarded.

4) Elution of bound IgG antibodies:
For elution of bound IgG antibodies, a cap was placed tightly at the bottom of each column and 1 ml of 100 mM Glycine-HCl buffer at pH 2.5 was added.
The IgA agarose resin with Glycine-HCl mixture was mixed by inverting the tube up/down 10 times followed with incubation for 5 minutes at RT. After incubation, the bottom cap was removed and the eluted fractions were collected in 1.5 ml Protein LoBind tubes and were neutralized with 150 µl of 1 M Tris buffer, pH 8.0. The elution step was repeated three times for maximal recovery of bound IgG mAbs.

Finally, the eluted IgG proteins were analyzed on SDS-PAGE and protein concentrations were measured using a Qubit Protein Assay Kit on a Qubit 3.0 flurometer.

Chapter 4: Results

This chapter deals with the protocol development of a novel technique to identify native antibody heavy chain (HC) and light chain (LC) pairings from millions of B cells at single-cell level using rough endoplasmic reticulum (rER) microsome-associated mRNAs. Further, this chapter deals with the application of this technique for discovery of antigen-specific monoclonal antibodies, followed by cloning, expression and binding studies of natively paired antibodies.

4.1 First sequencing results of paired HC-LC antibody repertoires

During the course of my Ph.D., I have tested several PCR amplification and sequencing protocols, which culminated in the first protocol I used, to produce high-throughput sequencing information of native antibody HC-LC pairs from individual B cells. In this section, I present the development of our first protocol and results obtained from MiSeq 2x250 bases paired-end sequencing using the Illumina MiSeq sequencing platform.

4.1.1 Simple HC-LC amplification from leukocyte RNA using emulsion overlap-extension RT-PCR

The initial overlap-extension RT-PCR (OE RT-PCR) experiment was kept relatively simple by using leukocyte RNA as a template because of its high abundance of molecules encoding HC and LC with constant, variable heavy (VH) and variable light (VK) chain genes of B cells. Water-in-oil emulsion reactions were generated according to a previously described method[131] with some modifications (Chapter 3.2.1). Multiplex overlap extension primers were designed and modified from previous studies[124,130] to amplify all VH and VK gene families. The emulsion OE RT-PCR assembly procedure involves cDNA synthesis from HC and LC mRNAs using isotype-specific primers (for e.g. IgM/IgG), followed by linkage and amplification with overlap extension VH and VK primers, to generate paired HC-LC assembly products. A schematic representation of the OE RT-PCR amplification is presented in Figure 5. After OE RT-PCR amplification, the emulsions were broken, DNA was

extracted (Chapter 3.2.7) and products were loaded onto agarose gels for verification. Due to the complex mixture of PCR products amplified during OE RT-PCR assembly reactions, the HC-LC assembled products of interest were not detectable on an agarose gel. Thus, the HC-LC assembly products of ~740 bp were size-selected (Figure 6) and DNA recovery was evaluated by three different DNA extraction methods (i) Zymo Gel DNA Recovery following manufacturers' instructions, (ii) Pellet Paint Co-Precipitant purification detailed in Chapter 3.2.8, and (iii) AMPure Beads Purification detailed in Chapter 3.2.9 to determine the best purification method for maximal recovery of assembled products.

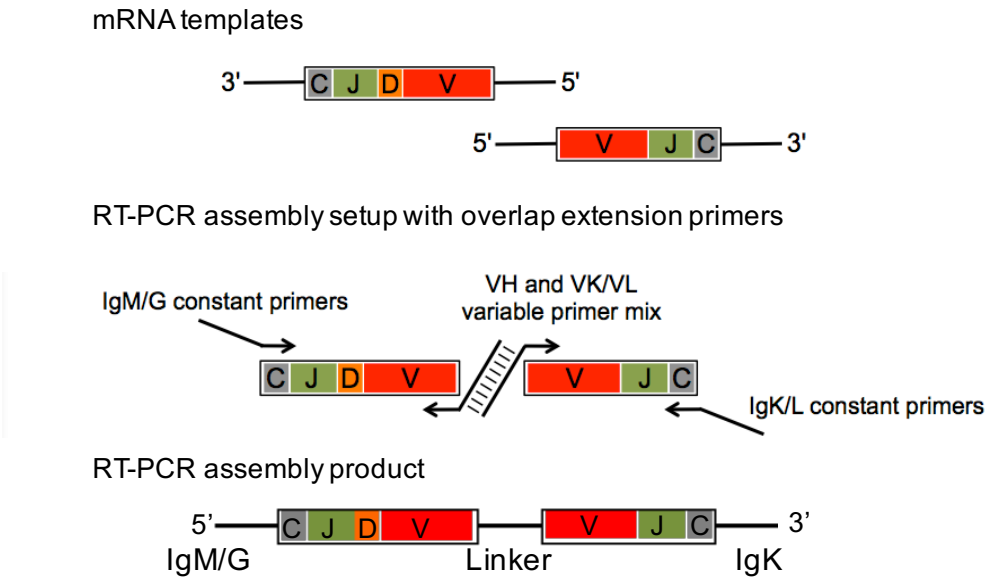

Figure 5: Overlap extension RT-PCR setup to link antibody heavy and light chains: To retain immunoglobulin HC and LC pairing information from individual B cells, we have setup RT-PCR assembly reactions with overlap extension primers specific for HC (IgM/IgG) and LC (IgK) constant region, and VH and VK gene families. In our approach, the HC and LC mRNAs from B cells are reverse transcribed with Ig HC and kappa LC constant region primers, linked and amplified with VH and VK variable primers. The OE RT-PCR results in assembly product of ~740 bp.

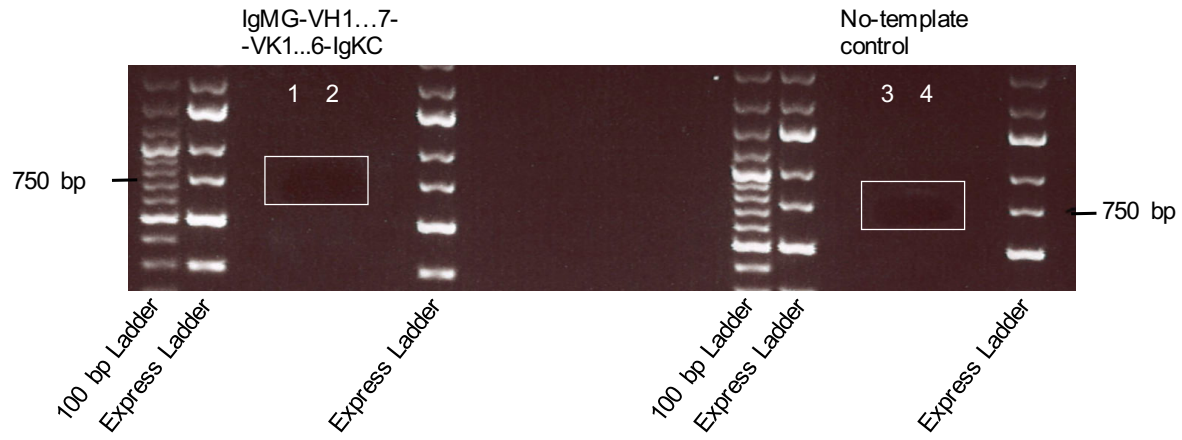

Figure 6: Agarose gel demonstrating size-selection process of OE RT-PCR assembly products: Total RNA from human peripheral blood leukocytes was used as templates for OE RT-PCR assembly reactions in emulsion using the assembly primer set composed of IgM and IgG heavy chain constant primers, VH and VK family primers, and kappa light chain constant primers. After OE RT-PCR assembly reaction, the emulsions were broken, DNA was extracted and loaded on 1% agarose gel. Lanes 1 and 2 demonstrate the size-selection process of OE RT-PCR assembly products (white box; IgMG-VH1…7-VK1…6-IgKC) of ~740 bp. Lanes 3 and 4 show the size-selection process of no-template control (white box). Similarly, two other emulsion OE RT-PCR assembly reactions were performed using leukocyte RNA under similar conditions, and DNA was size-selected as mentioned above. 100 bp DNA ladder and Express DNA ladder were used for size estimation of assembly products.

4.1.2 Nested PCR amplification using gel purified OE RT-PCR assembly products

To obtain enough assembled antibody HC-LC products for high-throughput sequencing, nested PCR amplification was performed using gel-extracted HC-LC assembled DNA recovered from three different DNA extraction methods with isotype-specific nested primers (Figure 7, Appendix table 1, Chapter 3.2.5). The nested PCR amplification resulted in HC-LC assembly products of ~710 bp, and no background signals were detected in the no-template control from the three gel extraction methods (Figure 8). However, we observed that using Pellet Paint Co-Precipitant for gel DNA purification results in maximum recovery of HC-LC assembly products (Figure 8, lanes 7, 8 and 9), followed by Zymo Gel DNA Recovery (Figure 8, lanes 1, 2 and 3) and AMPure Beads Purification (Figure 8, lanes 13, 14 and 15).

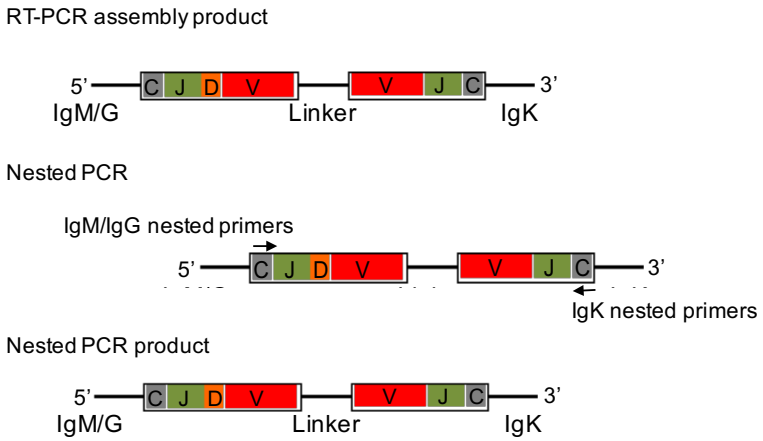

Figure 7: Nested PCR amplification: HC-LC assembled DNA from emulsion-purified OE RT-PCR assembly reaction were used as templates to perform nested PCR amplification using primers with HC (IgM and IgG) and LC (IgK) constant region sequences. The resulting amplified nested PCR product had a size of ~710 bp.

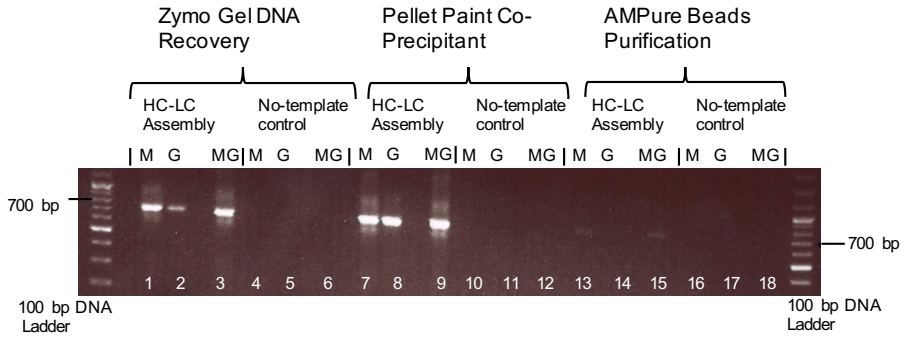

Figure 8: Agarose gel demonstrating nested PCR products of HC-LC assembly: Size-selected, gel-purified products from three different DNA extraction methods were used as templates to perform nested PCR amplification using IgM/IgG and IgK nested primers. Lanes 1-6 demonstrate nested PCR amplification of HC-LC assembled DNA extracted using the Zymo Gel DNA Recovery method. Lanes 1 and 2 show separate amplification of assembly products for IgM and IgG isotypes, respectively. Lane 3 show the amplification of two heavy chains and light chain isotype primers combined together in one reaction. Lanes 4,5 and 6 show no background signals with no-template control when amplified using isotype-specific primer mixes (IgM/IgG/IgK) and also when both the HC and LC isotype primer mixes are combined together in one reaction. Lanes 7-12 demonstrate nested PCR amplification of HC-LC assembled DNA extracted using the Pellet Paint Co-Precipitant method. Lanes 13-18 demonstrate nested PCR amplification of HC-LC assembled DNA extracted using the AMPure Beads Purification method.

4.1.3 Analytical PCR for verification of HC and LC diversity in the assembled DNA

To verify the presence of all VH and VK gene families in the HC-LC assembled DNA, an additional PCR amplification step was performed using IgM/IgG nested PCR products recovered by three different DNA purification methods. Isotype-specific nested primers encoding HC (IgM and IgG) and LC (IgK) constant regions were combined with different VH (VH1/7, VH2, VH3, VH4, VH5, VH6) and VK (VK1, VK246, VK3, VK5) family primers and were amplified using analytical PCR (Chapter 3.2.6). The results from the analytical PCR confirmed the presence of all VH and VK family genes of ~370 bp in the assembled DNA extracted with Pellet Paint Co-Purification and Zymo Gel DNA purification methods (Figure 9). AMPure purification of assembled DNA resulted in amplification of VH family genes only with IgM isotype (VH1, VH3, VH4, VH5, VH6), while VH families with IgG isotypes (VH3 and VH5) were poorly amplified and no bands were detected for remaining VH and VK family genes (Figure 10). The results from this experiment led us to the use of the Pellet Paint Co-Precipitant Gel DNA purification method for the following experiment using B cell microsomes.

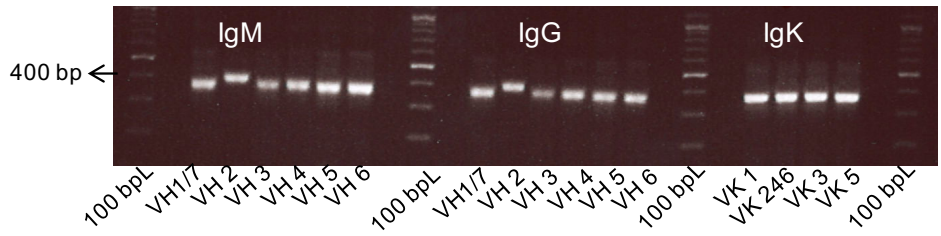

Figure 9: Analytical PCR demonstrating the presence of all VH and VK family genes in the emulsion assembled DNA: The nested PCR amplified products with IgMG heavy chain isotype (Figure 4, lanes 3) was purified using Pellet Paint Co-Precipitant purification method, which was used as template to perform analytical PCR with primers amplifying different VH and VK gene families. Fragments of ~370 bp results from IgM, IgG nested primers combined with primers for different VH families. Also products of ~370 bp resulted from kappa nested primers combined with different VK family primers. Similar results were obtained with IgMG heavy chain isotype (Figure 4, lane 9) purified using the Zymo Gel DNA purification method. These results demonstrate the presence of all VH and VK gene families in the assembled DNA.

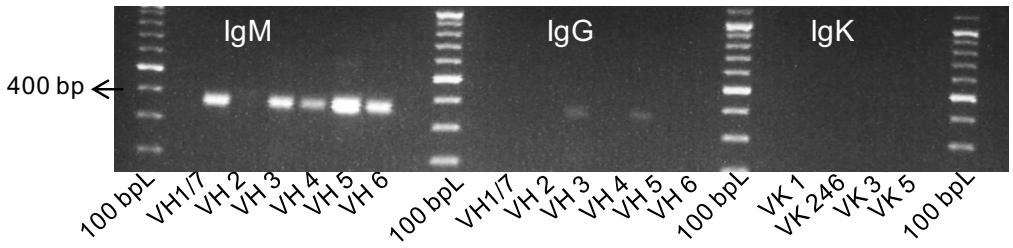

Figure 10: Analytical PCR showing VH and VK gene families in the emulsion assembled DNA purified using AMPure purification method: Nested PCR amplified products with IgMG heavy chain isotype (Figure 4, lanes 15) was purified using the AMPure Beads Purification method and was used as template to perform analytical PCR with primers amplifying different VH and VK gene families. Fragments of ~370 bp resulted from IgM nested primers combined with primers for the different VH gene families, while only poor amplification of two VH gene families (VH3 and VH5) were evident with IgG isotype and no bands were detected for the remaining VH and VK gene families.

4.1.4 Paired HC-LC amplification from CD19^{+} B cells of a healthy individual

The procedure described here to retain native antibody HC-LC pairs from single B cells is based on pioneering studies of Palade and co-workers illustrating the importance of endoplasmic reticulum (ER) in mRNA translation and protein synthesis [134,135]. The concept behind this approach is that each B cell contains rough endoplasmic reticulum (rER)-bound ribosomes with heavy and light chain mRNAs at translocon complexes (Figure 11a). These rER-bound mRNAs are further translated into membrane-bound or secretory antibodies that are expressed on the B cell surface. We hypothesized that native HC and LC mRNAs from single B cells should be preserved in the resulting enriched microsomes derived after cell lysis. Consequently, these enriched microsome-associated mRNAs could be used for clonal assembly and amplification of native antibody HC-LC pairs from single B cells, when separated into individual water-in-oil emulsions.

4.1.4.1 Microsome preparation method

The microsome preparation method described in this section was developed by Hans-Jörg Warnatz and Jörn Glökler using HEK 293T cells (Figure 11c) based on a previously published protocol for plant material[126]. Using this approach, rER microsomes were prepared using one million CD19+ B cells freshly isolated from a human peripheral whole blood sample of a healthy individual, following the pluriBead positive non-magnetic cell isolation protocol (Chapter 3.1.3). After isolation, cells were immediately treated with the protein synthesis inhibitor cycloheximide[136,137] to halt mRNA translation into proteins and to retain stalled ribosome-associated mRNAs at rER translocon complexes (Chapter 3.1.4). The cycloheximide-treated cells were incubated in high-density buffer (Chapter 3.1.5) containing sucrose and 5% digitonin for cell and organelle lysis. The sucrose provides higher density inside the lysed microsomes, allowing easy isolation of rER microsomes (Figure 11b) using a cooled table-top centrifuge. Two different centrifugation steps were performed to obtain enriched rER microsomes from the lysed B cells. First, cell debris, nuclei and mitochondria were removed by centrifugation at 600 g. Next, microsomes-containing supernatants (cytosol) was diluted with water enabling sedimentation of rER microsomes by centrifugation at 20,800 g. Finally, the supernatant (cytosol) was removed and the microsomes were resuspended in wash buffer (Chapter 3.1.5), centrifuged again at 20,800 g to obtain enriched microsomes with associated mRNAs for further analysis. As a positive control, microsomes from the ARH-77 cell line expressing a known IgG heavy and IgK (kappa) light chain were prepared and 10% (v/v) of ARH77 microsomes were spiked to B cell derived microsomes from one million cells.

4.1.4.2 Emulsion RT-PCR assembly and nested PCR amplification

An emulsion-based OE RT-PCR assembly reaction followed by nested PCR amplification was performed to retain the native HC-LC pairing information from single B cells. In this approach, the microsome suspensions from B cells with or without 10% ARH-77 cells were added to separate RT-PCR master mixes (aqueous phase). Water-in-oil emulsions were generated by dropwise addition of the aqueous

phase into an oil phase with continuous stirring using a magnetic stirrer. Based on Poisson statistics, the majority of individual microsomes are captured in separate emulsion droplet containing RT-PCR master mix with degenerate assembly primers (Appendix table 1, Figure 11d). Within each emulsion droplet, the microsome-associated mRNAs from single B cells (with/without ARH-77 cells) were reverse transcribed (RT) with isotype-specific primers specific for HC and LC genes, combined with VH/VK OE primers and were clonally amplified (Figure 11e) by PCR. To test the RNA integrity of the microsomes, an open RT-PCR reaction (without emulsion) with assembly primer mixes was performed as positive control. The PCR products were extracted from the emulsion droplets and the assembly products from both open and emulsion samples were loaded onto an agarose gels, and HC-LC assembled DNA of ~740 bp was size-selected as shown in Figure 6. DNA from size-selected products was extracted following the Pellet Paint Co-Purification method and was selectively amplified using nested PCR primers to increase the specificity of further amplification (Figure 11f). The results from nested PCR amplification demonstrates successful amplification of overlap extended HC-LC assembly products of ~710 bp from both open and emulsion assembled PCR products (Figure 12 and 13). However, nested PCR amplification using emulsion assembled products from B cell microsomes with ARH-77 cell microsomes resulted in stronger amplification (Figure 12) in comparison to that of B cells microsomes (Figure 13).

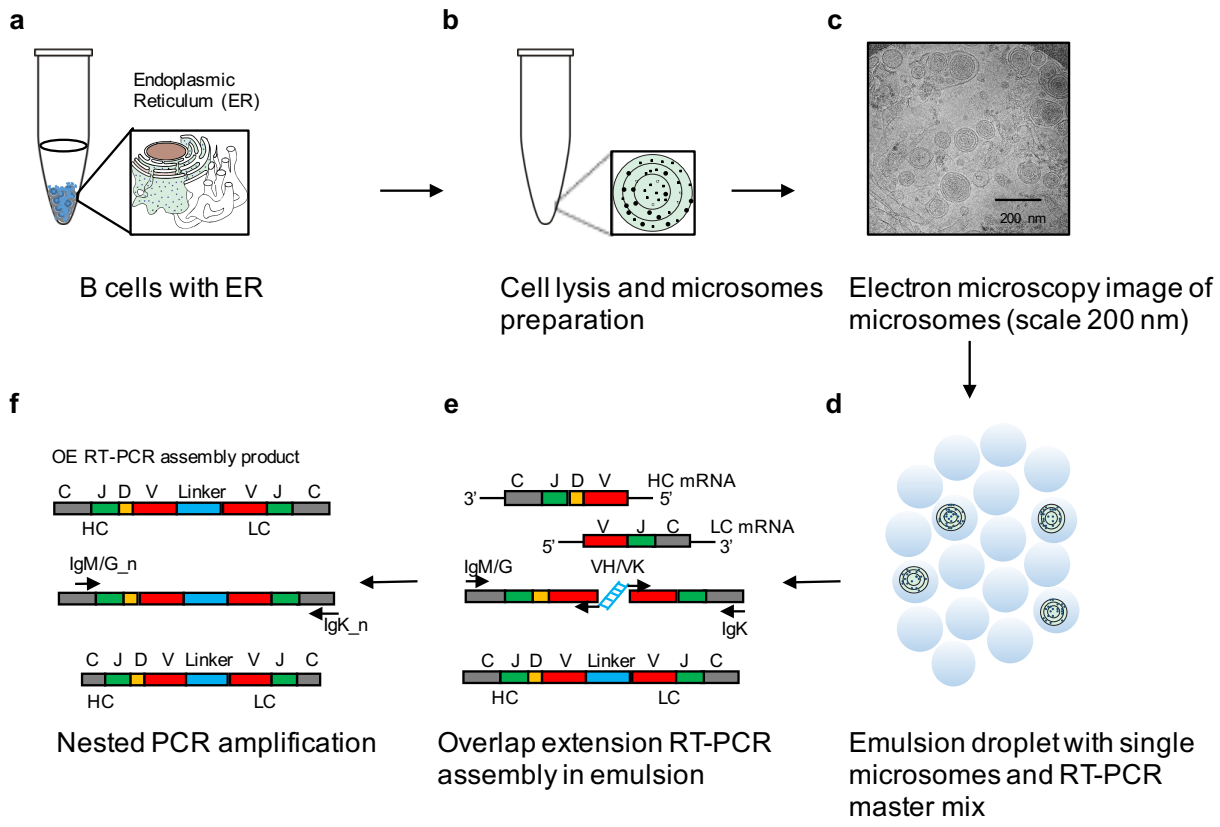

Figure 11: Overview of paired antibody HC-LC amplification using microsomes in water-in-oil emulsion droplets: (a) Each B cell contains endoplasmic reticulum (ER) studded with ribosomes encoding both HC and LC mRNAs at translocon complexes. (b) Freshly isolated B cells were lysed and microsomes were enriched using a differential centrifugation method. (c) The transmission electron microscopic image reveals enriched rER microsomes with multilamellar and unilamellar structures. The image was acquired from HEK 293T microsomes used for establishment of the microsome preparation method. Scale bar represents 200 nm. (d) Single B cell microsomes with degenerate primers and RT-PCR master mix present in individual emulsion droplets. (e) OE RT-PCR assembly of HC and LC mRNAs within individual emulsion droplets. (f) Nested PCR amplification using OE RT-PCR assembly products and isotype-specific primers (IgM_n, IgG_n, and IgK_n).

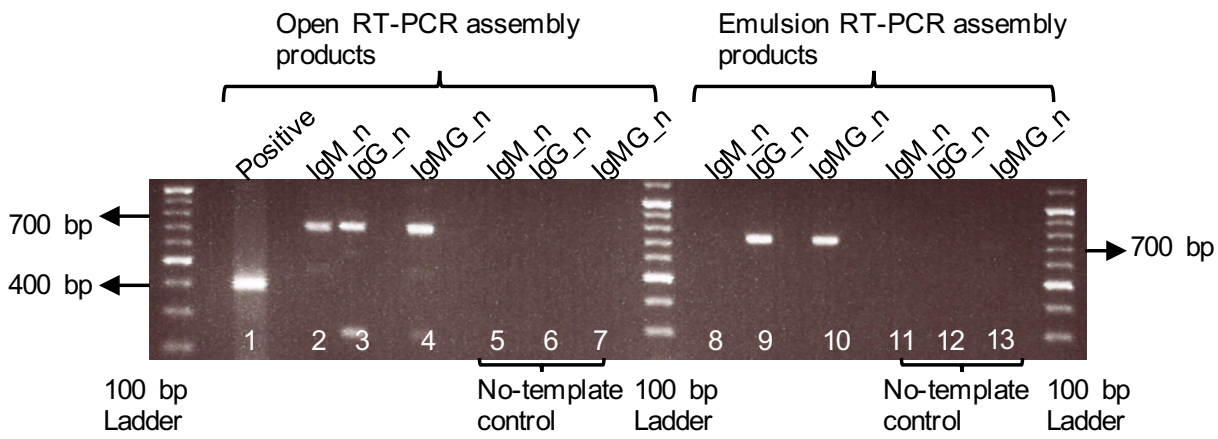

Figure 12: Agarose gel showing nested PCR products of HC and LC assemblies from B cell microsomes spiked with 10% ARH-77 microsomes: Gel-purified open and emulsion RT-PCR assembly products from B cells spiked with 10% ARH-77 microsomes were used as a template to perform nested PCR using IgM/IgG and IgK nested primers. Separate amplification of assembly products for IgG was observed both in open and emulsion PCR samples (lanes 3 and 9) while amplification of assembly product with IgM isotype was detected only in the open PCR sample (lane 2). Also, the amplification of assembly products with open and emulsion PCR samples worked when both the HC and LC isotype primers were combined in one reaction (lanes 4 and 10). No background signals were observed in the no-template control with primers specific for IgM and IgG heavy chains and also when both the HC and LC isotype primers were combined in one reaction for both open and emulsion assembly products (lanes 5, 6, 7, 11, 12 and 13). The open RT-PCR positive control for RNA integrity with only one HC and one VH family primer (IgG_n + VH4) resulted in an amplified product of ~400 bp (lane 1).

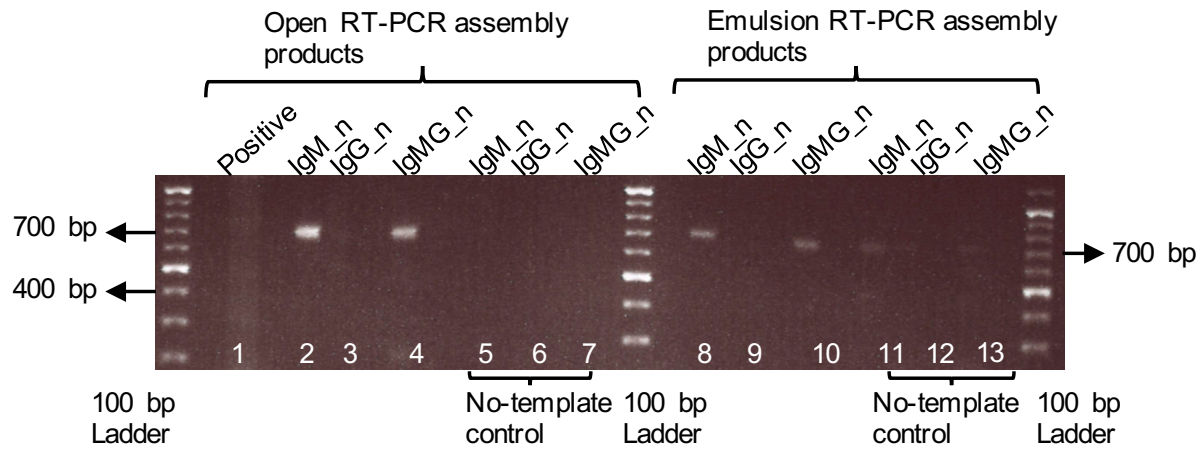

Figure 13: Agarose gel showing nested PCR products of HC and LC assemblies from B cells microsomes: Gel-purified open and emulsion RT-PCR assembly products from B cells microsomes were used as a template to perform nested PCR amplification using IgM/IgG and IgK nested primers. Separate amplification of assembly products for IgM isotype was observed in both open and emulsion PCR samples (lanes 2 and 8), while amplified assembly product with IgG isotype was not detected (lanes 3 and 9). A band of ~710 bp was detected when both HC and LC isotype primers were combined in one reaction in both open and emulsion PCR samples (lanes 4 and 10). No background signals were observed in open PCR sample (lanes 5,6 and 7) while background signals of ~710 were detected in no-template control (lanes 11,12 and 13). Also, the open RT-PCR positive control for RNA integrity with only one HC and one VH family primer (IgG_n + VH4) did not result in any amplified product (lane 1).

4.1.4.3 Analytical PCR for verification of the presence of the different VH and VK gene families

As a quality control and to verify the presence of amplified VH and VK gene families in the assembled DNA, an analytical PCR (Chapter 3.2.6) was performed using IgM/IgG-amplified nested PCR products from emulsion assembled samples (B cells spiked with 10% ARH-77 cells, and only B cell microsomes). The PCR result with 10% ARH-77 cell microsomes spike-in revealed the presence of almost all VH family (VH1/7, VH3, VH4, VH5 and VH6) and VK family (VK1, VK2/4/6, VK3 and VK5) gene families with IgM, IgG and IgK isotypes (Figure 14). The PCR result using only B cell microsomes revealed the presence of VH1/7, VH3, VH4, VH5 and VH6 gene families with IgM isotype, while only the VH6 gene family was detected with IgG

isotype primer. Also, all VK gene families (VK1, VK2/4/6, VK3, VK5) were present in the assembled DNA (Figure 15).

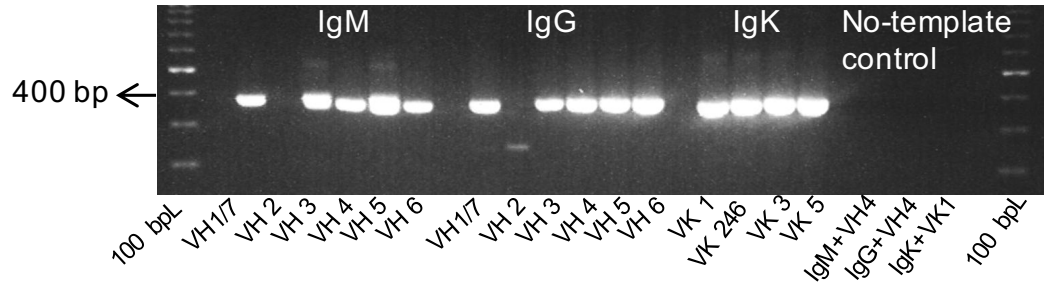

Figure 14: Analytical PCR showing the presence of the different VH and VK gene families in the emulsion-assembled DNA from B cell microsomes spiked with 10% ARH-77 microsomes: Nested PCR amplified products with IgM/IgG heavy chain isotypes (Figure 8, lanes 10) were used as template to perform analytical PCR with primers amplifying the different VH and VK gene families. Fragments of ~370 bp resulted with both IgM and IgG nested primers combined with primers for different VH gene families. No band for the VH2 gene family was detected for both IgM and IgG isotypes. Products of ~370 bp resulted from kappa nested primers combined with different VK family primers, and no bands were detected in the no-template control.

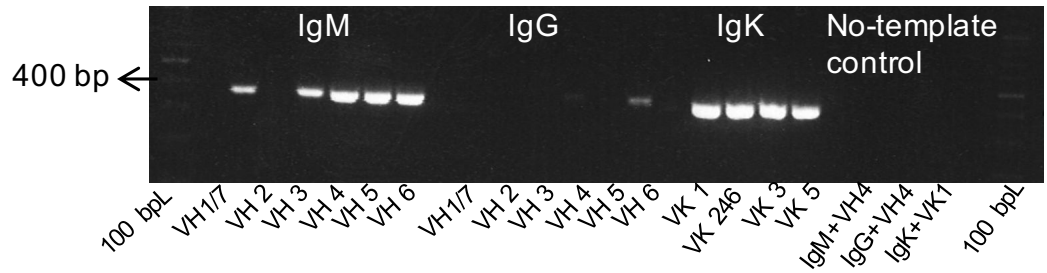

Figure 15: Analytical PCR showing VH and VK gene families in the emulsion-assembled DNA from B cell microsomes: Nested PCR amplified products with IgM/IgG heavy chain isotypes (Figure 9, lanes 10) were used as a template and analytical PCR was performed with primers for the different VH and VK gene families. Fragments of ~370 bp resulted with IgM isotype for all VH gene families except for the VH2 family, and only VH6 family amplification was detected with IgG isotype primer. Products of ~370 bp resulted from kappa nested primer combined with different VK family primers. Also, no bands were detected for the no-template control.

4.1.4.4 Next generation sequencing library preparation and MiSeq sequencing

Next, we verified the concept of using rER microsomes to retain native HC-LC pairings from individual B cells by preparing Illumina TruSeq DNA libraries and performing MiSeq 2x250 bases paired-end sequencing using the MiSeq sequencing platform. MiSeq sequencing libraries were prepared (Chapter 3.3.1) from gel-purified IgM/IgG nested PCR products of both open- and emulsion-assembled samples (B cell microsomes spiked with 10% ARH-77 cells, and only B cell microsomes) following the Illumina TruSeq DNA sample preparation guide. We multiplexed the open and emulsion samples together, loaded them onto one flow cell and performed Illumina MiSeq 2x250 bases paired-end sequencing of the resulting antibody repertoire libraries.

4.1.5 Bioinformatic analysis of paired antibody repertoire sequencing data obtained from MiSeq sequencing

MiSeq 2x250 bases paired-end sequencing on antibody repertoires resulted in ~3.1 million raw sequencing reads from the emulsion PCR assembly, and ~3.8 million reads from the open PCR assembly samples. Briefly, the raw sequencing reads were quality-filtered for minimum Phred quality scores of 20 over 50 percent of nucleotides to ensure high quality reads in the CDR3-containing regions, followed by annotation of HC (IgM and IgG) and LC (IgK) sequences using the constant region primer sequences. Next, the annotated reads from both open and emulsion assembly samples were aligned to the human Ig germline genes (IMGT annotation) and clustered using the MiXCR software[109] to determine the number of unique paired clonotypes. Using our bioinformatics pipeline (Figure 16), we retained ~75% of paired reads with V(D)J and C isotype information from both the open assembly and emulsion assembly samples. The aligned sequencing data was further corrected for PCR and sequencing errors using the MiXCR software, and the CDR-H3 nucleotide sequences were extracted and clustered to determine the number of unique B cell clonotypes in the analyzed antibody repertoires (Table 4). The extracted clonotype information was converted into a human-readable text file

(Figure 17). We then created a pairing statistics file which shows the frequencies and pairing combinations of the different VH and VK gene families in both the open and emulsion assembly samples. The pairing statistics data (Figure 18a – 18d) shows the usage of different VH and VK genes between the open and emulsion assembly samples, which also correlate with the number of unique B cell clonotypes obtained from MiXCR analysis. This data suggests correct amplification of HC-LC pairs in the emulsion assembly sample while the amplification of HC-LC pairs is random in the open assembly sample, as expected.

Next, to determine the native HC-LC pairing accuracies using our method, we identified and extracted the known HC- and LC-CDR3 sequences from the ARH-77 spike-in cells in both the open and emulsion assembly samples. We determined the HC-LC pairing frequencies by comparing the total number of HC-CDR3 reads paired with the correct known LC-CDR3 reads in both the open (Figure 19a) and emulsion assembly samples (Figure 19b). We observed that the ARH-77 positive control spike-in demonstrated a high pairing accuracy of >88% in the emulsion samples (Fig 19b). By considering one nucleotide mismatch due to sequencing errors or nucleotide mismatches that arose during PCR amplification, the HC-LC pairing accuracy of ARH-77 cells was 17.6-fold above the background in the emulsion assembly sample, while the HC-LC pairing accuracy was only 1.8-fold above the background with the open assembly sample, as expected. This result provided us with the proof-of-principle for our paired immunoglobulin repertoire sequencing technique and suggested that the endogenous HC-LC pairing information from single B cells can be obtained with high pairing accuracy.

Sample type	**Raw sequencing reads**	**Paired reads with isotype information**	**V(D)J annoted read pairs from MiXCR analysis**	**Final clonotype count**
Open assembly	3,812,979	2,900,083	2,786,236	2,173
Emulsion assembly	3,162,883	2,394,241	2,366,173	899

Table 4. Summary of sequencing results obtained from Illumina MiSeq 2x250 bp paired-end sequencing

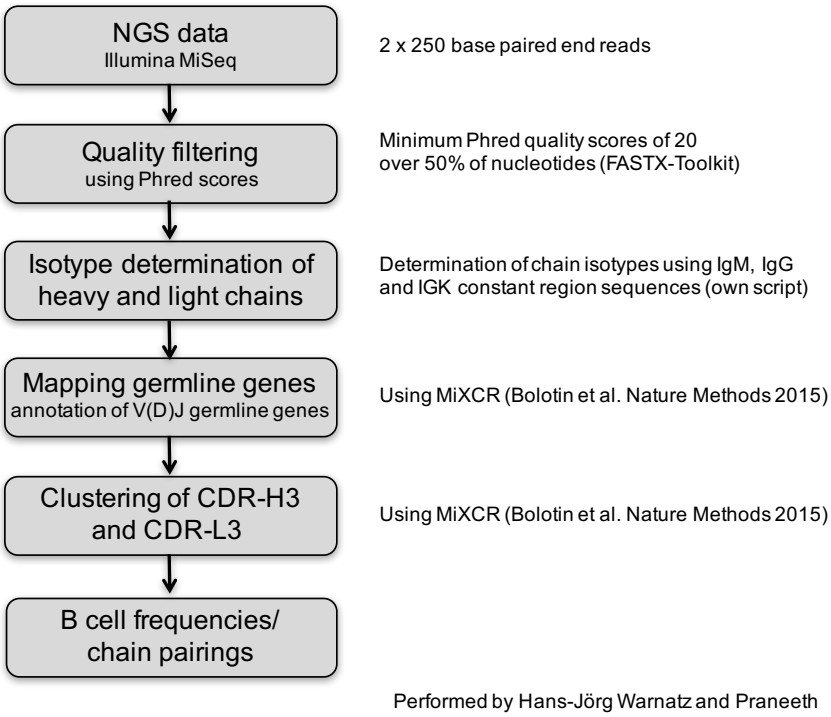

Figure 16: Bioinformatics pipeline used for analysis of paired antibody repertoire sequencing data: Illumina MiSeq 2x250 bases paired-end sequencing data was obtained from the MiSeq sequencing platform and the raw sequencing data was analyzed using our bioinformatics pipeline. The bioinformatics script was written and executed by Dr. Hans-Jörg Warnatz, while paired repertoire sequencing analysis was performed by me and Dr. Hans-Jörg Warnatz.

	A	B	C	D	E	F
1	**Clone count**	**AA. seq. CDR3**	**Best V hit**	**Best D hit**	**Best J hit**	**Best C hit**
2	1973787	CARGLLRGGWNDVDYYYGMDVW	IGHV4-34*00	IGHD1-1*00	IGHJ6*00	IGHGP*00
3	94724	CASVRLVREYNDYW	IGHV3-48*00	IGHD6-19*00	IGHJ4*00	IGHM*00
4	65748	CAAPSMTDW	IGHV4-39*00	IGHD3-22*00	IGHJ4*00	IGHM*00
5	56309	CAKDHPIMGATRTDLW	IGHV1-46*00	IGHD1-26*00	IGHJ5*00	IGHM*00
6	42376	CARDPLGGYDYW	IGHV4-34*00	IGHD5-12*00	IGHJ4*00	IGHM*00
7	37271	CARDLLRGGWNDVDYYYGMDVW	IGHV4-34*00	IGHD1-1*00	IGHJ6*00	IGHGP*00
8	20515	CARNVAAAGTSYFDYW	IGHV4-59*00	IGHD6-13*00	IGHJ4*00	IGHM*00
9	12185	CARGLLRGAGT_VDYYYGMDVW	IGHV4-34*00	IGHD1-1*00	IGHJ6*00	IGHGP*00
10	9772	CARGLLRGGLE_DVDYYYGMDVW	IGHV4-34*00	IGHD1-1*00	IGHJ6*00	IGHGP*00
11	716	CAKDPPIMGATRTDLW	IGHV1-46*00	IGHD1-26*00	IGHJ5*00	IGHM*00
12	650	GASVRLVREYNDYW	IGHV3-48*00	IGHD6-19*00	IGHJ4*00	IGHM*00
13	474	CASVRLVRR_DYYYGMDVW	IGHV3-48*00	IGHD6-19*00	IGHJ6*00	IGHGP*00
14	439	WAKDHPIMGATRTDLW	IGHV1-46*00	IGHD1-26*00	IGHJ5*00	IGHM*00
15	398	CARGTAMIPYYYGMDVW	IGHV4-59*00	IGHD5-5*00	IGHJ6*00	IGHM*00
16	336	CARASSGGAGT_VDYYYGMDVW	IGHV4-34*00	IGHD1-1*00	IGHJ6*00	IGHGP*00
17	325	WARDPLGGYDYW	IGHV4-34*00	IGHD5-12*00	IGHJ4*00	IGHM*00
18	329	CARGLLRGGWN_RGLLLWYGRL	IGHV4-34*00	IGHD1-1*00	IGHJ6*00	IGHGP*00
19	302	CAKDHPIMGAPRTDLW	IGHV1-46*00	IGHD1-26*00	IGHJ5*00	IGHM*00
20	267	CAKDRPIMGATRTDLW	IGHV1-46*00	IGHD1-26*00	IGHJ5*00	IGHM*00
21	231	CAAPAMTDW	IGHV4-39*00	IGHD2-2*00	IGHJ5*00	IGHM*00
22	191	CASPSMTDW	IGHV4-39*00	IGHD3-22*00	IGHJ5*00	IGHM*00

Figure 17: Screenshot of top clonotypes obtained from MiXCR analysis for the emulsion assembly sample

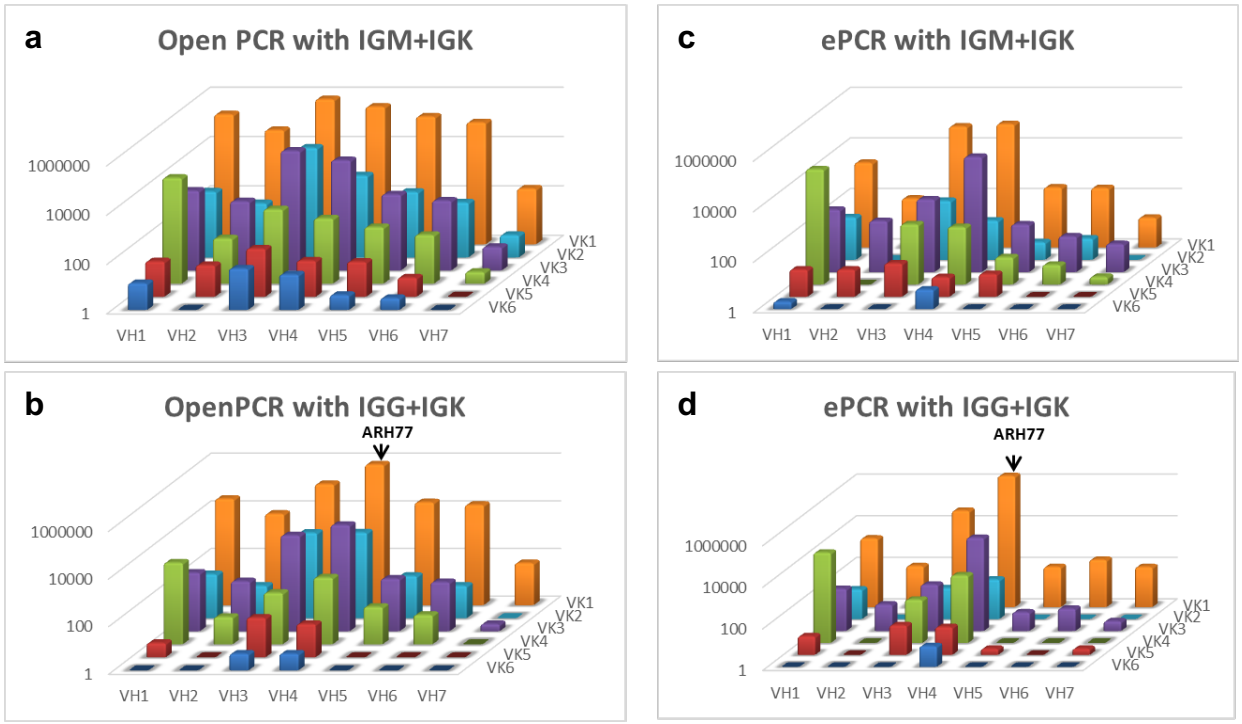

Figure 18. HC-LC pairing frequencies in open and emulsion samples: Distribution of different VH and VK gene families observed in total paired HC-LC gene repertoires in open PCR (a, b) and in emulsion PCR (c, d).

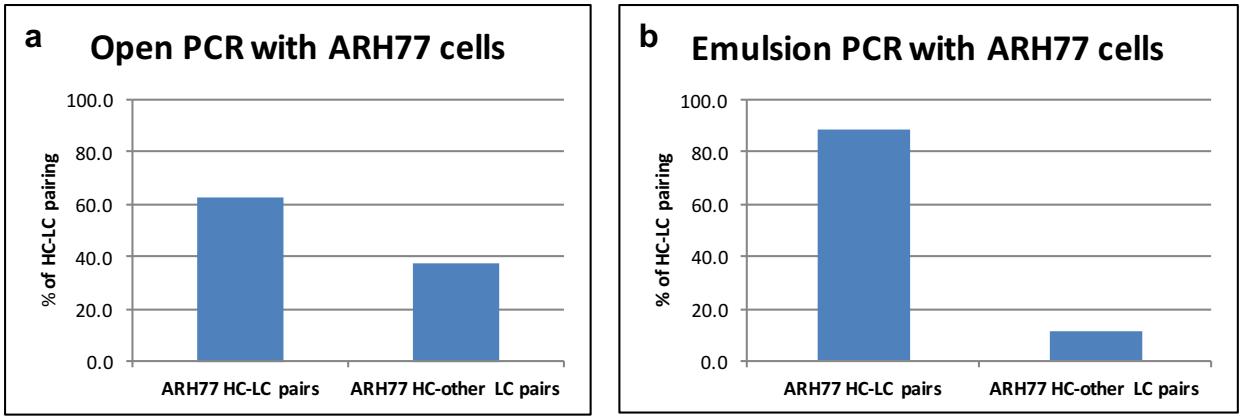

Figure 19. Comparison of HC-LC pairing accuracy of ARH-77 cells in open PCR and emulsion PCR: The known HC- and LC-CDR3 sequences of the ARH-77 positive control spike-in were determined from the sequences obtained from open and emulsion PCR samples. Correct HC-LC pairing accuracy of ~65% was observed with ARH77 cells in the open PCR sample (a) while the pairing accuracy was >88% in the emulsion PCR sample (b).

4.2 MiSeq 2x250 bases paired repertoire sequencing using optimized protocol

Our first MiSeq 2x250 bases paired-end sequencing results have provided the proof-of-concept to preserve native HC-LC pairs from each single B cell with a pairing accuracy of >88%. However, the total number of B cell clones (899 B cell clones) identified using our method might inadequately represent the analyzed human B cell repertoire. In order to achieve a much higher estimate of analyzed human B cell repertoires, I focused my next experiments towards troubleshooting the previously observed difficulties with PCR amplifications and optimize our method to retain native antibody HC-LC pairs from single B cells. Here are some of the important observations made, and how I solved them to optimize our method:

1) We hypothesized that one of the oil phase components (e.g. Triton X-100, Chapter 3.2.2) might have affected the stability of emulsions during high denaturation temperatures of RT-PCR assembly reactions. To overcome this, I have set up the emulsion reaction using oil components from the commercially available Micellula DNA Emulsion and Purification Kit (Chapter 3.2.4), where the emulsion components are stable at temperatures up to 100˚C.

2) Previous low yields of amplified IgM/IgG products (Figure 12 and 13) could be due to enzyme denaturation at the interface between the hydrophilic and hydrophobic phases within the water-in-oil emulsions used for clonal PCR amplifications. We found that addition of two-fold higher amounts of enzyme in the emulsion reactions resulted in enhanced amplification and higher yields.

3) The RT-PCR assembly results in a complex mixture of HC-LC assembled DNA, also containing GC-rich templates. Another possible explanation for previous low yields of amplified products could be that nested PCR amplification is hampered by GC-rich templates through the formation of stable secondary structures [138]. To overcome this possible problem, longer denaturation, annealing and extension times were used during nested PCR amplification (Chapter 3.2.5), which resulted in higher yields of amplified HC-LC assembly products.

4) Reduction in the number of nested PCR cycles eliminated previously observed faint background signals in negative controls.

Having optimized all the above-mentioned steps using leukocyte RNA and ARH-77 microsomes, I applied the enhanced protocol (Chapter 3.2.4) to study vaccination-induced changes in $CD19^+$ B cell repertoires before and 7 days after Tetanus Toxoid/Diphtheria Toxoid (Td) booster immunization.

4.2.1 Paired HC-LC amplification using $CD19^+$ B cells of a healthy individual from before and 7 days after Td booster vaccination

Freshly derived human peripheral whole blood samples were obtained from a healthy individual both before and 7 days after Td booster vaccination from in.vent Diagnostica GmbH. Informed consent from the individual was obtained according to the accompanying ethical vote (Freiburg Ethics Commission International, Feki code 011/1763). $CD19^+$ B cells from the whole blood samples were isolated using the pluriBead positive non-magnetic cell isolation method (Chapter 3.1.3), and rER microsomes from 1.5 million B cells were prepared as previously described (Chapter 4.1.4, Figure 11). As a positive control for endogenous HC-LC pairing, microsomes from frozen ARH-77 cells were prepared, and 0.5% (v/v) of ARH-77 microsomes were spiked to $CD19^+$ B cell-derived microsomes (Figure 20). The resulting rER microsomes together with RT-PCR mix were passed into water-in-oil emulsion droplets and were amplified in two separate reactions with primers specific for IgM and IgG, respectively (Chapter 3.2.4). The rER microsome-associated antibody HC and LC mRNAs were reverse transcribed with primers specific for HC and LC genes, assembled by VH/VK OE primers, followed by PCR amplification within each emulsion droplet. The emulsion-assembled products were extracted and loaded onto an agarose gel, and assembled DNAs of ~740 bp were size-selected by gel extraction. The DNAs from the size-selected products were recovered using the Zymo Gel DNA Purification Kit, and nested PCR amplification was performed with different cycle numbers (28, 31 and 34; Chapter 3.2.5) to determine the best cycle numbers. Nested PCR-amplified products of ~710 bp were detected in both open- and emulsion-assembled DNA, with IgM and IgG isotype-specific primers, for both

before and after Td booster immunization samples (Figure 21a and 21b). Next, we verified the presence of all VH and VK gene families in the emulsion-assembled DNA from both before and after Td booster immunization by performing an analytical PCR (Chapter 3.2.6) with only 15 amplification cycles. A band of ~370 bp was detected for all the VH and VK gene families in IgM-IgK and IgG-IgK nested PCR products of both before and after Td booster vaccination samples (Figure 22a and 22b).

After emulsion OE RT-PCR and subsequent nested PCR, Illumina TruSeq DNA libraries were prepared from IgM and IgG assembly products obtained after 28 cycles of nested PCR amplification, from both before and after Td booster immunization samples. The resulting antibody repertoire libraries from B cells were multiplexed together, loaded onto a flow cell, and Illumina MiSeq 2x250 bases paired-end sequencing was performed on the Illumina MiSeq sequencing platform.

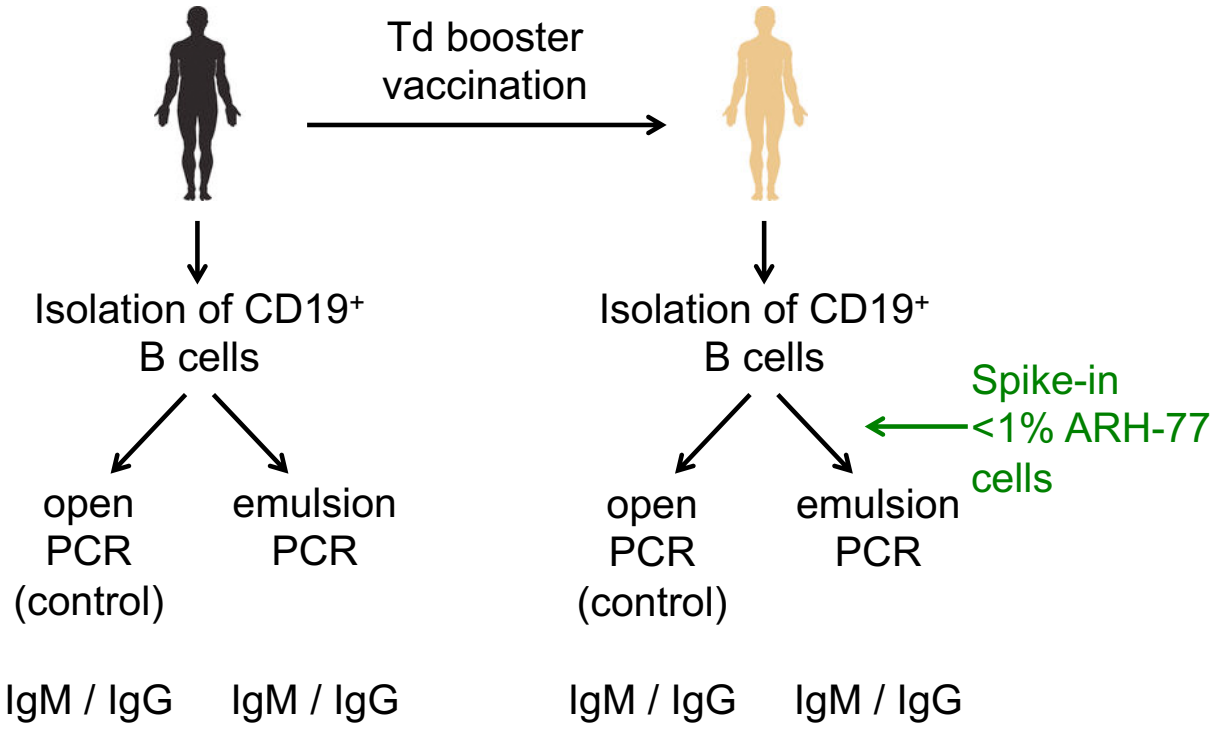

Figure 20: Experimental setup for analysis of TT vaccine responses. CD19+ B cells were isolated from peripheral human whole blood samples from both before and 7 days after Td booster vaccination: An open PCR (without emulsion) with primers specific for IgM and IgG isotypes was performed as a control to show the random pairing of HC-LC for both before and after Td booster vaccination samples. Emulsion PCR was performed with primers specific for IgM and IgG isotypes for both before and after Td booster vaccination samples to obtain the native HC-LC pairings from individual B cells.

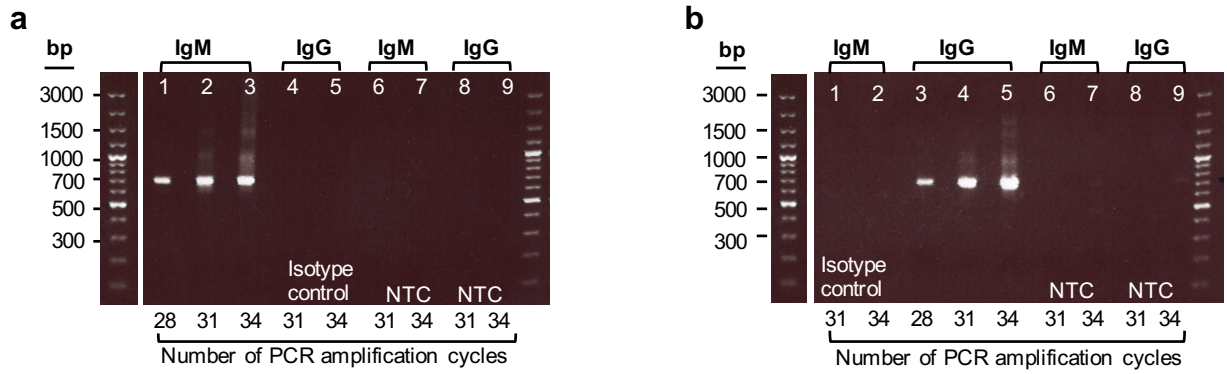

Figure 21: Determination of the optimal nested PCR cycle number: (a) IgM-IgK emulsion-assembled DNAs from 7 days after Td booster vaccination samples were used as a template for nested PCR amplification with 3 different cycle numbers ranging from 28-34 cycles. PCR amplification with IgM nested primers yielded a single band of ~710 bp, demonstrating the HC-LC assembled DNA (lanes 1,2 and 3). No background signals were observed with IgG isotype control (lane 4 and 5) and no-template control (NTC) amplified with isotype specific nested primers (lanes 6-9) run at maximum cycle number. (b) Nested PCR amplification using IgG-IgK emulsion assembled DNA from post Td booster vaccination sample.

Note: Similar results were obtained for nested PCR amplification using IgM-IgK and IgG-IgK emulsion assembled DNA from before vaccination samples.

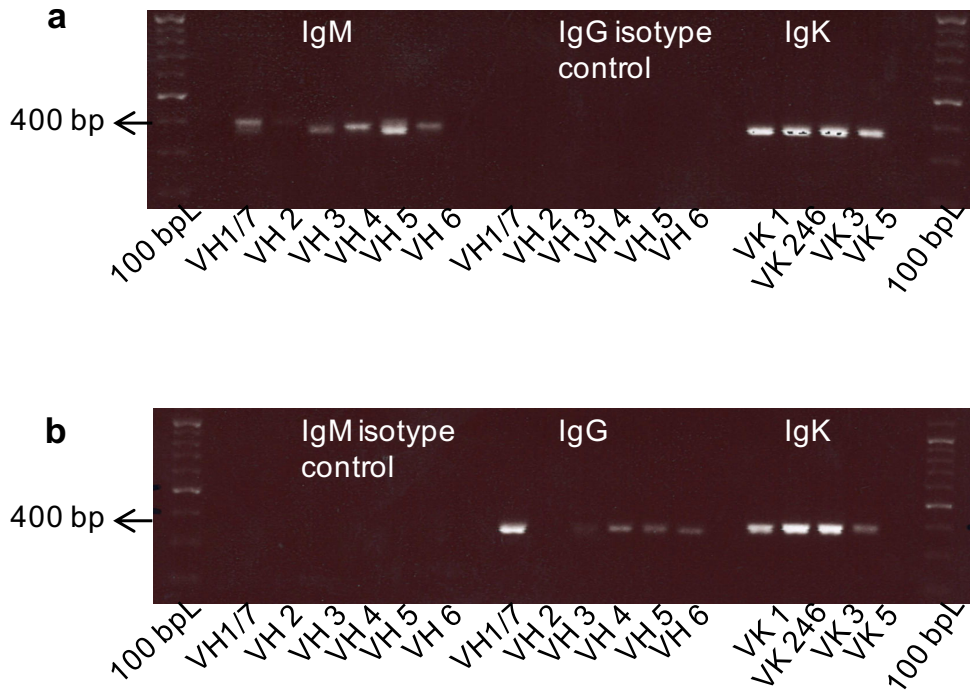

Figure 22: Analytical PCR showing VH and VK gene families in emulsion assembled DNA from Td booster vaccination samples: Nested PCR products with IgM-IgK (Figure 21a, lane 1) and IgG-IgK (Figure 21b, lane 3) were used as templates and analytical PCR with 15 cycles of amplification was performed with primers for the different VH and VK gene families. Fragments of ~370 bp resulted with IgM (Figure 22a) and IgG (Figure 22b) isotype primers for all VH gene families. Correct fragments of ~370 bp resulted from Kappa nested primers combined with different VK family primers (Figure 22a and 22b). No bands were detected with IgM- (Figure 18b) and IgG (Figure 18a) -isotype controls.

Note: Similar results were obtained for analytical PCR amplification using IgM-IgK and IgG-IgK emulsion assembled DNA from before vaccination samples.

4.2.2 Bioinformatics analysis of paired HC-LC sequencing data

MiSeq 2x250 bases paired-end sequencing resulted in raw sequencing reads of ~5.9 million and ~5.2 million reads with IgM and IgG isotypes, respectively, from emulsion assembly products of before vaccination samples (Table 5). Also, ~5.3 million reads and ~5 million reads were obtained for IgM and IgG isotypes, respectively, from emulsion assembly products of post Td booster vaccination samples. The sequencing reads from the IgM and IgG assembly products of both before and after vaccination samples were analyzed using our bioinformatics pipeline (Figure 16) as

described above. From MiXCR analysis, we retained ~81% and ~85% of paired reads with V(D)J and C isotype information, for both IgM and IgG isotypes, respectively, from before vaccination sample (Table 5). Post Td booster vaccination, we retained ~75% and ~83% HC-LC paired reads with IgM and IgG isotype information, respectively. The CDR-H3 nucleotide sequences from the annotated reads were extracted and clustered together, and clonotype information was exported into a human-readable text file (Figure 13). Next, we filtered out HC-CDR3 amino acid sequences containing duplicates, frameshifts and stop codons to determine the number of paired CDR3 clones (requiring ≥2 reads per pair) in the analyzed B cell population. We obtained 2,246 and 5,124 HC-LC pairs for IgM and IgG, respectively, from before vaccination samples. The post-Td booster immunization resulted in 4,032 and 3,028 HC-LC pairs with IgM and IgG isotypes, respectively (Table 5). Among these clonotypes, we identified 213 (IgM) and 129 (IgG) HC-CDR3 clones that were present in both before and after Td booster vaccination samples. We noticed that the top 10 HC-LC pairs of IgM and IgG isotypes, from both before and after vaccination, constituted 57% and 49% (with IgM isotype) and 61% and 76% (with IgG isotype) of the total aligned reads, respectively. This observation demonstrates a clonotype distribution that is skewed towards the most frequent HC-LC pairs. Subsequently, we extracted paired HC-LC sequences of ARH-77 cells from the sequencing data to determine the pairing accuracies using our method. The ARH-77 spike-in HC-LC pairing demonstrated low pairing accuracies potentially as a result of the freeze-thaw process. Nevertheless, the reads from the ARH-77 cell line demonstrated preferential pairing of the known HC with the correct corresponding LC (Figure 23).

Immunization	**Pre-immunization (day 0)**		**Post-Td booster immunization (day 7)**	
Approximate number of fresh CD19+ B cells used for preparation of microsomes	1.5 million		1.5 million	
Isotype	IgM	IgG	IgM	IgG
Raw sequencing read pairs (HC-LC) obtained from MiSeq 2x250 bases paired end sequencing	5,860,554	5,185,375	5,356,219	4,979,748
Total number of aligned read pairs (HC-LC)	5,238,212	4,411,684	4,647,787	4,332,934
Unique HC-LC pairs with ≥1 read per pair	5,858	19,090	9,609	11,119
HC-LC pairs with ≥2 reads per pair	2,200	4,841	4,031	2,872

Table 5: Overview of pre- and post Td booster immunization sequencing results

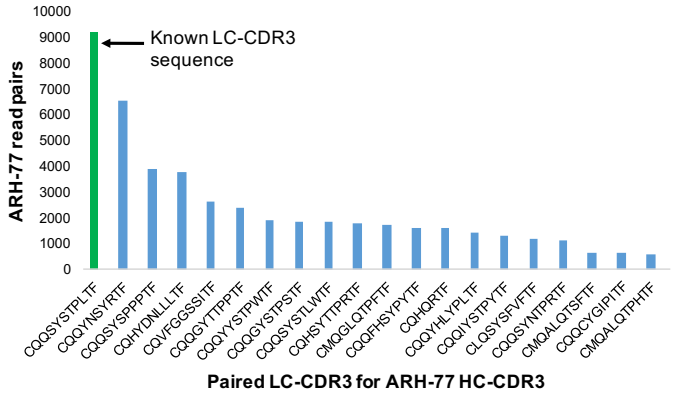

Figure 23: Light chain pairing distribution for ARH-77 cell spike-in post Td-booster immunization sample: The ARH-77 HC-CDR3 sequence was preferentially paired to the correct LC-CDR3 sequence (CQQSYSTPLTF, shown in green), with some mispairing to other light chains (shown in blue).

Next, we generated heat maps using ggplot2 (a plotting software for R) showing the paired HC-LC pairing frequencies of different germline families from all aligned reads (Figure 24a-d). We observed that some VH-VK pairings (e.g. VH3-VK1, VH4-VK1, and VH4-VK3) were highly frequent (up to 78% of total reads) for both IgM (Figure 24a) and IgG (Figure 24c) isotypes in before vaccination sample. In contrast, after Td booster vaccination, other pairings, such as VH1-VK1 (20%) and VH1-VK2 (55%) were mostly observed, followed by VH3-VK1 (47% in IgM and 8% in IgG) and VH4-VK1 (7% in IgM and 15% in IgG) pairings, in both IgM (Figure 24b) and IgG (Figure 24d) isotypes, respectively. These observations demonstrate significant changes in the Ig repertoire VH gene family usage and expansion of certain B cell clones in response to antigen stimulation[91,139-142]. Using our method, we also identified certain HC-LC pairs (e.g. VH7-VK5 and VH7-VK6) that are generally observed at lower frequencies (Figure 24a-d), as reported in previous studies[124,143] illustrating the sensitivity of our method in detecting rare clonal pairs.

Finally, we extended our paired repertoire sequencing analysis to identify previously reported TT antigen-specific B cell receptor sequences from our sequencing data. We compared the IgG HC-CDR3 amino acid sequences obtained from after vaccination sample with TT-specific HC-CDR3 sequences reported in prior studies[124,144 139,145,146]. We found two previously reported HC-CDR3 sequences in our data set (CARQADNWFDPW[144] and CATGRTLDYW[139]).

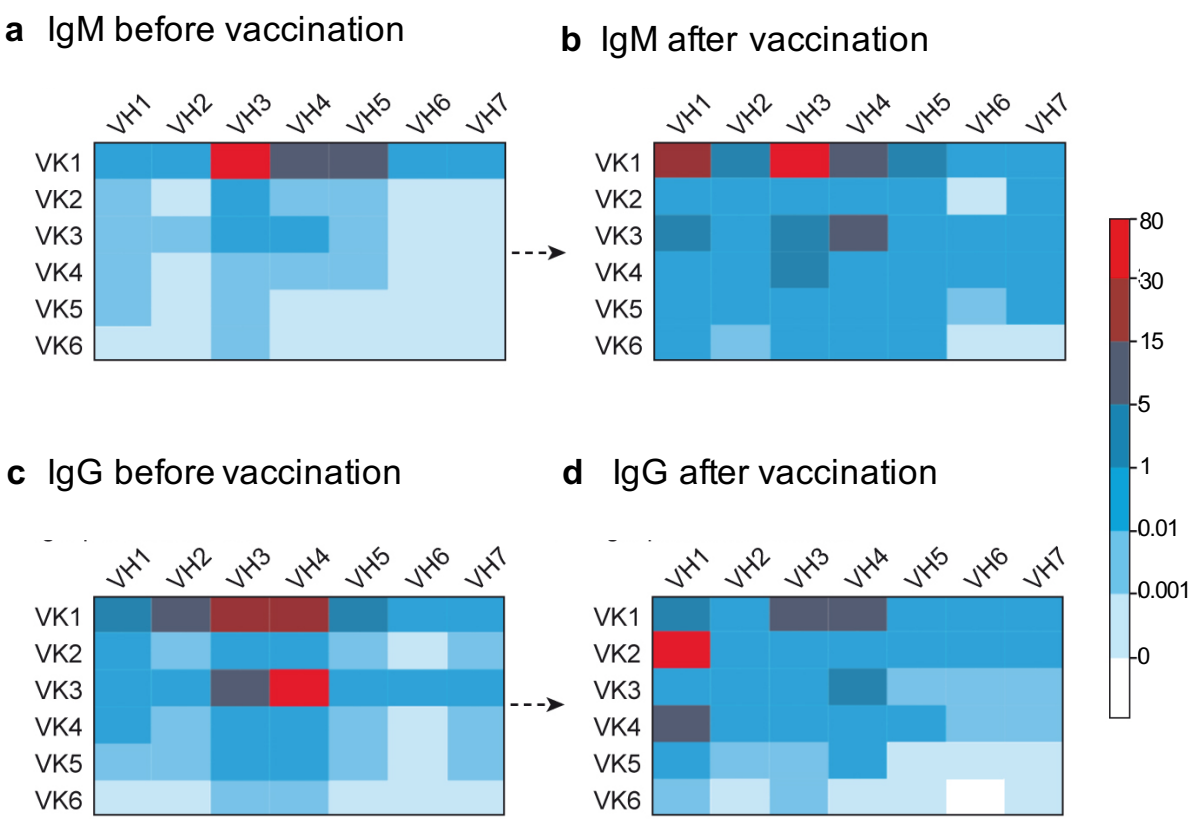

Figure 24: VH-VK pairing frequencies of different germline gene families observed in before and after Td booster vaccination samples: (a) IgM HC-LC pairing frequencies in the CD19^{+} B cell repertoire from before vaccination sample (total aligned reads: 5,238,212; final clonotype count: 2,246). (b) IgM HC-LC pairing frequencies in the CD19^{+} B cell repertoire after Td booster vaccination (total aligned reads: 4,647,787; final clonotype count: 4,032). (c) IgG HC-LC pairing frequencies in the CD19^{+} B cell repertoire from before vaccination sample spiked with 0.1% ARH-77 microsomes (total aligned reads: 4,411,684; final clonotype count: 5,124). (d) IgG HC-LC pairing frequencies in the CD19^{+} B cell repertoire spiked with 0.1% ARH-77 microsomes from after Td booster vaccination sample (total aligned reads: 4,332,934; final clonotype count: 3,028). Colors indicate percentage of reads for indicated VH-VK pairings among all reads in the analyzed B cell repertoire.

4.3 Antibody Cloning, expression and binding studies

To demonstrate a potential application of our paired antibody repertoire sequencing technique, we performed antibody cloning, expression and binding studies using ELISA assays for the discovery of TT antigen-specific human monoclonal antibodies.

The current MiSeq sequencing platform generates paired antibody repertoire sequencing data that is limited to 2x250/2x300 bases, which comprise the most informative FR2-CDR2-FR3-CDR3-FR4 region and Ig constant region. Although this information is sufficient to determine the native HC-LC pairs, it is not adequate to clone the entire native antibody pairs into respective expression vectors for production of monoclonal antibodies. To overcome this limitation, we designed a two-step PCR strategy which allows cloning of full-length HC and LC pairs into respective expression vectors containing IgG and IgK constant regions, respectively.

In our design, we selected fourteen TT vaccine-induced HC-LC pairs (likely originating from clonally expanded $CD19^+$ B cells post immunization), including one previously reported TT antigen-specific HC-CDR3 sequence (CARQADNWFDPW), from the IgG post Td booster vaccination sequencing data. The full-length HC-LC gene transcripts of selected clonotypes were amplified using total RNA isolated from frozen $CD19^+$ B cells of post Td booster vaccination sample using our two-step PCR strategy (Chapter 3.4.1-4, Figure 25). In the first step, RT-PCR amplification was performed using forward primers that anneal to the respective VH/VK gene family FR1 region including a BssHII restriction digestion overhang at the 5' end, and a reverse primer specific for the respective HC/LC CDR3 and 18 bases of FR4 sequence. Next, a second PCR amplification was performed using DNA from the first step, forward primers that anneal to the BssHII restriction site, and reverse primers annealing to 18 nucleotides of FR4 sequence possessing remaining FR4 region with a NheI/HindIII restriction digestion overhangs (Figure 26). The total RNA used for the above experiments was isolated from frozen $CD19^+$ B cells, which demonstrated low viability due to freeze-thawing; thus certain HC and LC pairs of interest could not be amplified[146].

Next, the resulting HC and LC amplicons were cloned into IgG (pCMV-CD40-4IE3-HC) and IgK (pCMV-CD40-4IE3-LC) expression vectors, respectively and were transformed into *E. coli* Top10 chemically competent cells. The transformed colonies were verified for HC and LC inserts by Sanger sequencing, and the plasmid DNA was purified from the colonies that showed the closest match to the consensus sequence. For recombinant mAb production, the HC- and corresponding LC-encoding plasmids were co-transfected into HEK 293T cells using the

polyethyleneimine transfection method. Post transfection, IgG-containing cell culture supernatants were harvested on day 4 and indirect ELISA using day 4 cell culture supernatants on plates coated with TT antigen was performed for the discovery of TT antigen-specific mAbs. We identified three TT antigen-specific mAbs, named Top1, Top2, Top3 and Top4 (Figure 27). Intriguingly, the Top2, Top3, and Top4 HC-LC pairs were also identified in the before vaccination sample, albeit at much lower frequencies (<0.1% of total reads), suggesting clonal expansion of these pre-existing clonotypes after antigen exposure. However, a previously identified TT antigen-specific HC-CDR3 (CARQADNWFDPW) clonotype did not bind to TT antigen using our method (Figure 27). This result demonstrates that over 28% of the selected clonal pairs identified using this method bind to TT antigen, thus validating the application of this technique for discovery and production of mAbs using native HC-LC pairing information from single B cells.

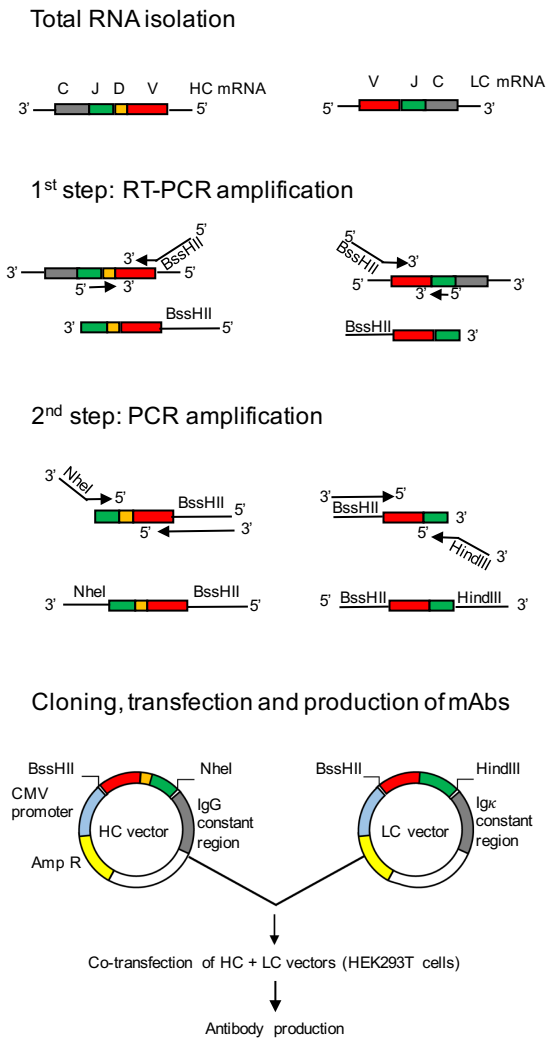

Figure 25: Schematic illustration of the two-step PCR strategy used for cloning antibody HC and LC into respective expression vectors for mAb production: (a) Total RNA was isolated from frozen $CD19^+$ B cells obtained after Td booster vaccination. (b) Amplification of selected HC-LC pairs was performed by RT-PCR with forward primers annealing to the respective VH/VK gene family with a 5' BssHII restriction digestion overhang, and reverse primers annealing to partial CDR3 (D-region) and 18 bases of FR4 sequence (J-region). (c) PCR amplification was performed using forward primers with restriction digestion sites (BssHII) and 3 nucleotides of the respective V-region, and reverse primers encoding complete FR4 sequences with a restriction site at the 3' end (NheI/HindIII). Primers used in the first step RT-PCR and second step PCR amplifications are listed in Appendix table 2. (d) The HC and LC sequences are cloned into IgG and IgK expression vectors, respectively and HC and the corresponding LC plasmids were co-transfected into HEK293T cells for expression of recombinant monoclonal antibodies.

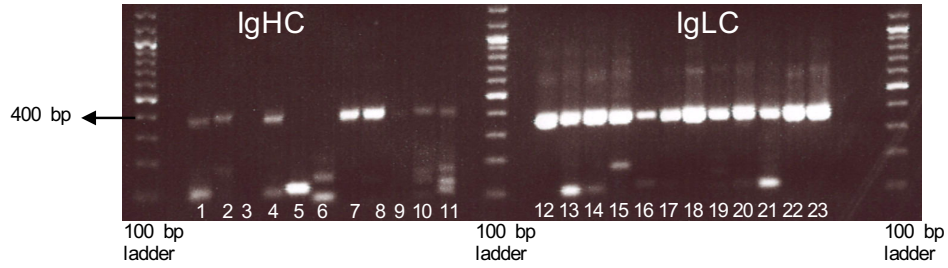

Figure 26: Two-step PCR amplification results of selected fourteen HC-LC pairs: The selected fourteen most abundant HC-LC pairs were amplified using a two-step PCR strategy as demonstrated in Figure 20. The two-step PCR results in HC-amplified products of ~380 bp while the LC-amplified products are ~340 bp. It should be noted that the total RNA used for the above experiments was isolated from frozen $CD19^{+}$ B cells, which demonstrated low viability due to freeze-thawing; thus certain HC and LC pairs of interest could not be amplified.

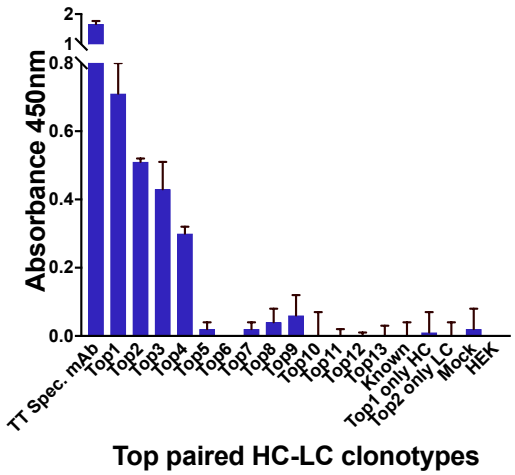

Figure 27: Characterization of TT-specific mAbs identified by indirect ELISA: Transfected HEK 293T cells supernatants containing IgG antibodies were harvested on day 4 from fourteen most abundant clones and were analyzed by indirect ELISA using plates coated with TT antigen. Four clones named Top1, Top2, Top3 and Top4 resulted in a positive signal with indirect ELISA, demonstrating specific binding of these clones to TT antigen. The commercially obtained TT-specific mAb used as a positive control for this assay resulted in a strong signal, while the negative controls (Top1 only LC, Top3 only HC, Mock and HEK293T cells supernatants) resulted in negligible signals. The X-axis represents the clone number or positive and negative controls, while the Y-axis shows the absorbance values at 450 nm. All ELISA experiments were performed in triplicates.

4.4 Brief summary of results

The results from this project demonstrate the development and validation of a novel and simple strategy to retain native paired antibody heavy and light chain sequences from single B cells, and its application for the discovery of mAbs. Our results show that rER microsome-associated mRNAs can be used to retain native antibody HC-LC pairs from single B cells, processing millions of cells per experiment, with a high pairing accuracy of >88%. By applying this method to study IgM and IgG immunoglobulin repertoire changes between pre- and post-Td booster vaccination samples, we identified two previously reported TT antigen-specific HC-CDR3 sequences and discovered four novel TT antigen-specific mAbs. This result demonstrates the application of our method for the discovery of mAbs. In conclusion, our method overcomes previously reported difficulties in paired Ig repertoire sequencing.

Chapter 5: Discussion

Each antibody producing B cell contains a pair of heavy chain (HC) and light chains (LC). The discovery of antibody gene rearrangements as a fundamental mechanism for generation of immune repertoire diversity demonstrated that B and T cells of the adaptive immune system in higher organisms can recognize and bind to enormous number of antigens and confer protection against diseases[34,147,148]. Preliminary studies on Ig repertoires using Sanger sequencing method[72] have initiated the measurement, interpretation, and understanding of antibody diversity using relatively very few B cells (around 300 cells), thus providing a glimpse into antibody repertoires[73], PMID:[49,149,150]. However, these studies were often limited by low throughput and involved laborious efforts to characterize B cell repertoires using a larger number of B cells (e.g. $>10^6$ cells per experiment).

The advancements in high-throughput DNA sequencing technology platforms and dramatic drop in sequencing costs[151], [152,153] has enabled researchers to profile heterogeneous Ig repertoires at an unprecedented depth, providing greater insights into humoral immune responses. The first in-depth antibody repertoire sequencing using next generation sequencing (NGS) technologies was reported in 2009[85]. Since then, NGS has revolutionized immunology research by providing greater insights into the complexity of antibody repertoires[71,154,155], bypassing *in vitro* screening for the production of recombinant monoclonal antibodies (mAbs)[76,93]. However, initial studies used total RNA isolated from bulk B cell populations and mostly focused on obtaining V(D)J sequence information of antibody HC. Although this approach is substantial for some applications[76,93], bulk lysis and extraction of total RNA from B cells results in loss of the native light chain pairing for the individual B cells. Thus, sequencing of native antibody heavy and light chain pairs (HC-LC) from each individual B cell remained a major challenge for a long time. Also, the pairing of heavy chains with diverse light chains remained challenging and laborious (e.g. phage display technology) to select antigen-specific HC-LC pairs for cloning and recombinant expression of monoclonal antibodies.

Very recently, single-cell based paired immune receptor repertoire sequencing technologies were reported to retain native antibody HC-LC pairs or T cell

receptor alpha (TCRα) and beta (TCRβ) chain pairs. For example, (i) DeKosky *et al*. captured 68,000 single B cells in 125 picoliter polydimethylsiloxane (PDMS) microwell slides containing poly(dT) magnetic beads, followed by cell lysis and mRNA capture on the beads[124]. Subsequently, emulsion-based linkage RT-PCR was performed to physically link and amplify antibody HC and LC mRNAs captured on the poly(dT) beads from single cells, followed by sequencing of the paired HC-LC amplicons on the MiSeq sequencing platform[124]. (ii) In another approach, Busse *et al.* have developed a two-dimensional cell barcoding strategy using a forward and reverse primer matrix[116]. In their approach, >46,000 single B cells were sorted into 384-well plates, cell-specific barcodes were added using V gene-barcoded forward primers and C gene-barcoded reverse primers. PCR amplification and sequencing of single barcoded antibody sequences was carried out using the Roche 454 pyrosequencing technology[116] (iii) In another study, Turchaninova *et al*. reported a cell-based emulsion RT-PCR method to retain native TCRα-β chain pairing using $8x10^6$ peripheral blood mononuclear cells (PBMCs)[122]. In this approach, single cells encapsulated in emulsion droplets were lysed by heating to 65˚C, and the released TCRα and TCRβ mRNAs were reverse transcribed, physically linked using overlap extension primers, and amplified within each droplet. Subsequently, emulsion-assembled samples were extracted, selectively amplified, and were sequenced using the Illumina MiSeq 2x150 bases paired-end sequencing platform. While this approach used $>8x10^6$ PBMCs, only very few TCRα-β pairs were identified (>700 pairs).

Although the above-mentioned approaches could successfully identify native antibody HC-LC pairs or TCRα-β chain pairs, these studies were limited by low cell number ($<10^5$ cells)[116, 124]or low-throughput (~700 pairs)[122] and required fabrication of custom chips[124]. In this project, we bypassed the above-mentioned drawbacks in single-cell paired repertoire sequencing methods and developed a simple, cost-efficient and innovative approach to sequence antibody repertoires from millions of B cells in one experiment.

In this study, we first isolated $CD19^+$ B cell populations from freshly derived human peripheral whole blood samples, as blood is the most easily accessible source for B cells in humans[52, 156]. We then established a protocol to obtain maximal yields of rER microsome-associated mRNAs from individual B cells while preserving the native

antibody HC and LC pairing information. Using our approach, enriched rER microsomes can be prepared in three steps, using differential centrifugation and a table-top centrifuge at only 20,800 g, bypassing conventional tedious ultracentrifugation steps at 100,000 g[157-159].

The next challenge for us was to know if the enriched rER microsome-associated mRNAs could be used to link and amplify native immunoglobulin HC and LC from each original B cell. To verify this, we spiked 10% of ARH-77 microsomes (expressing known HC and LC) to B cell-derived microsomes (corresponding to ~1 million cells). We then performed emulsion OE RT-PCR amplification to generate native HC-LC pairs, followed by nested PCR amplification and Illumina MiSeq 2x250 bases sequencing. From our first sequencing results, we identified 899 unique HC-LC pairs (Table 4), and the ARH-77 spike-in control demonstrated a high pairing accuracy of >88% (Figure 19b). These results demonstrated a proof-of-principle of our method to retain native antibody HC-LC pairs from individual B cells with high pairing accuracy. Thus, by obtaining these results, our single-cell paired repertoire sequencing method overcomes previously reported difficulties, such as low cell numbers, cell entrapment and cell lysis in the emulsion droplets, and RNA degradation during PCR reactions[116,121,124]. Further, removal of genomic DNA and non-secretory mRNA during microsome preparation results in the reduction of PCR artifacts due to off-target amplification. However, a moderate throughput (899 HC-LC pairs) was obtained after analysis of one million B cells. We hypothesized that this moderate throughput may arise due to technical pitfalls observed in our workflow (Figure 7), for example (i) Inhibition of emulsion RT-PCR amplification reactions due to relatively higher concentrations of non-ionic surfactants such as Triton X-100 or Tween 20[160,161] (ii) RT-PCR enzyme denaturation at the interphase of hydrophilic and hydrophobic components during emulsion OE RT-PCR[122] (iii) hampering of nested PCR amplification by GC-rich templates due to formation of secondary structures[138]. We then circumvented these potential technical difficulties by (i) preparing highly stable water-in-oil emulsions using the commercially available micellula DNA Emulsion and Purification Kit, (ii) addition of twice the amount of RT-PCR enzyme to the emulsion reaction and, (iii) increasing the denaturation, annealing and extension times for nested PCR amplification.

Next, we optimized our workflow to retain a much higher throughput of paired antibody repertoires and applied it to understand vaccination-induced changes in human $CD19^+$ B cell repertoires. We processed 1.5 million $CD19^+$ B cells isolated from human peripheral whole blood both before (day 0) and 7 days after Td booster vaccination. We then prepared rER microsomes and performed OE RT-PCR followed by subsequent nested PCR amplification and Illumina MiSeq 2x250 bases paired-end sequencing. Using our approach, we identified 2,246 and 5,124 antibody HC-LC pairs from the before vaccination sample (day 0) for IgM and IgG isotypes, respectively. From the post Td booster vaccination sample, we identified 4,031 and 3,028 antibody HC-LC pairs for the IgM and IgG isotypes, respectively. This demonstrated that higher throughput of paired antibody repertoires could be obtained using our optimized method, in comparison to the early reported studies on paired immune receptor repertoire sequencing technologies[116,121,122,124]. The whole process from B cell isolation to identifying native antibody pairs spans only four days, including high-throughput sequencing on the Illumina MiSeq sequencing platform. In contrast to another established single-cell RT-PCR methods which are expensive, laborious and time-consuming[162], our method is cost-effective and can be performed in any laboratory using minimal laboratory equipment (that is, a high-speed table-top centrifuge, a magnetic stirrer and a thermocycler).

Memory B cells are formed within germinal centers during primary infections and provide enhanced and robust antibody response during secondary infections and immunizations[156,163,164]. The human peripheral whole blood contains a majority of antigen-specific plasmablasts while memory B cells are present at very low frequencies (e.g. 1 in 2,500-100,000 memory B cells)[145,165]. Using our approach, we identified 213 (IgM) and 129 (IgG) HC-CDR3 clonotypes that are persistent both before and after Td booster vaccination. This demonstrates the application of our method to track and identify pre-existing B cells, possibly from the antigen-specific memory B cell compartment[91,139,144]. We also observed that 3 out of 4 novel TT-specific HC-CDR3 antibody sequences (Top2, Top3, Top4, Figure 27) identified were also found in the pre-vaccination sample, but at very low frequencies (<0.1% of total reads). This further confirms the sensitivity of our method for identifying and tracking the clonal expansion of antigen-specific B cells that are mostly arising from memory B cells circulating in the peripheral human whole blood[91,139]. This also

verifies the application of our method for discovery of antigen-specific monoclonal antibodies in a high-throughput manner[52,91,93,124,166].

Antibody repertoire studies in response to vaccination (PMID: 2300757, PMID:, PMCID: [52,167] provide valuable information on antigen-specific antibody sequences with known specificities for diseases such as Influenza[76,118,124], and Tetanus[79,124,144,146]Thus, the identification of known antigen-specific antibody sequences from previously reported studies would be beneficial to study immune responses generated for specific vaccine antigens, and to track known antibody sequences for vaccine development[91]. In our study, we identified two known TT antigen-specific HC-CDR3 sequences (CARQADNWFDPW and CATGRTLDYW)[144] from the IgG post Td-booster vaccination repertoire, simply by comparing the HC-CDR3 sequences from post-vaccination to previously published TT antigen-specific sequences. This observation suggests the suitability of our method to identify known antigen-specific antibody sequences to vaccine antigens, diseases, and/or for tracking B cell maturation to identify broadly neutralizing antibodies[91].

Immunoglobulin expression levels (i.e. Ig mRNA transcripts per cell) in different B cell subsets vary dramatically at different levels. For example, plasma B cells contain up to 100-fold higher Ig mRNA transcript levels compared to naive B cells[91,168]. Hence, quantitative assessment of mRNA transcripts in the analyzed B cell repertoire would be difficult. In our sequencing data, we observed that the top 10 paired HC-LC sequences attributed for 49-76% of the total aligned reads from both IgM and IgG repertoires of pre- and post-vaccination samples, respectively. This observation indicated a skewed view on the sequenced B cell antibody repertoires due to different Ig mRNA expression levels from different B cell subsets. Thus, our method reflects Ig mRNA amounts and associated antibody protein levels, rather than absolute B cell numbers. We also noticed a biased amplification of certain V-gene segments[102,114] using primers for the VK1 and VK3 LC gene families (Figure 20a-d) in our sequencing data. Hence, the reported antibody HC-LC pairs inadequately represent the actual B cell frequencies. However, a more accurate B cell number and frequency using our method could be obtained by reducing PCR amplification biases by adjusting the primer concentrations and limiting the amplification cycles.

If quantification of the antibody repertoires is the primary focus, more accurate B cell frequencies could be achieved by using unique molecular identifiers (UMIs) to reduce PCR amplification biases during nested PCR combined with computational approaches. Also, sorting B cell subsets of interest (e.g. memory B cells and plasma cells) based on cell surface markers[169,170] before generating sequencing libraries could provide a reliable estimate of analyzed antibody repertoires[171].

As discussed above, until very recently, native antibody HC-LC pairs could be identified from single cells using up to $<10^5$ B cells. However, within the last two years, two emulsion-based paired antibody HC-LC repertoire sequencing techniques were reported which retains native antibody HC-LC pairs from single cells, using millions of B cells in one experiment[125]. For example, DeKosky *et al*. have constructed a flow-focusing apparatus to capture single cells into emulsion droplets containing lysis buffer and magnetic beads to capture the mRNA transcripts [125, 172]. The HC and LC mRNAs were reverse transcribed and physically linked within each emulsion droplet by linkage PCR, followed by subsequent nested PCR amplification and Illumina MiSeq sequencing. This approach was used to retain native HC-LC pairings from >2 million B cells in one experiment. Although substantial, the construction of a flow-focusing apparatus is a complex and laborious process and requires a dedicated person for operation and monitoring. In another approach, Briggs *et al.* used microfluidic devices to pass B cells and unique barcoded beads with lysis buffer into individual emulsion droplets (AbVitro technology). Cells encapsulated within each individual droplet were lysed, HC and LC mRNA transcripts were reverse transcribed and template switch-tagged by short adaptor sequence. Subsequently, emulsion PCR-amplified products were recovered and sequenced using the Illumina MiSeq sequencing platform. In this study, they processed 3 million B cells in one experiment to retain native HC-LC pairs. Although this technology is very powerful, it is limited by the use of complex and expensive microfluidic devices which are not affordable for laboratories with limited resources.

In summary, the method described in this thesis offers a straightforward approach to retain native antibody HC-LC pairs at single-cell level, processing millions of B cells per experiment. This method does not require the physical separation of B cells

using a flow cytometer, flow-focusing apparatus or complex and expensive microfluidic devices [116, 122, 124, 125, 172]. In contrast to the existing paired immune repertoire sequencing technologies, our method offers a cost-effective way to retain paired immune repertoire sequences from single cells using only common laboratory equipment such as a table-top centrifuge, a magnetic stirrer, and a thermocycler. Thus, this technology can make paired Ig repertoire sequencing widely applicable for laboratories with limited resources. The whole process from B cell isolation to sequencing on Illumina MiSeq platform, and analyzing native antibody HC-LC pairs can be performed by a single researcher within four days. Validation of antigen-specific antibodies can be carried out within two weeks' time, after acquisition of sequencing data. In conclusion, the presented method can be exploited in applications ranging from monoclonal antibody discovery, vaccine profiling and autoimmune diseases to cancer immunotherapies. In particular, this technology could be valuable to assess and defy fast-spreading pandemics (e.g. 2014 Ebola outbreak) by enabling rapid generation of therapeutic monoclonal antibodies. We believe this method can be widely used to accelerate translational research investigating several aspects of immune biology.

Bibliography

1 Eisele, N. A. & Anderson, D. M. Host Defense and the Airway Epithelium: Frontline Responses That Protect against Bacterial Invasion and Pneumonia. *Journal of pathogens* 2011, 249802, doi:10.4061/2011/249802 (2011).

2 Yatim, K. M. & Lakkis, F. G. A brief journey through the immune system. *Clinical journal of the American Society of Nephrology : CJASN* 10, 1274 1281, doi:10.2215/CJN.10031014 (2015).

3 Turvey, S. E. & Broide, D. H. Innate immunity. *The Journal of allergy and clinical immunology* 125, S24-32, doi:10.1016/j.jaci.2009.07.016 (2010).

4 Medzhitov, R. & Janeway, C., Jr. Innate immunity. *The New England journal of medicine* 343, 338-344, doi:10.1056/NEJM200008033430506 (2000).

5 Espinosa, V. & Rivera, A. First Line of Defense: Innate Cell-Mediated Control of Pulmonary Aspergillosis. *Frontiers in microbiology* 7, 272, doi:10.3389/fmicb.2016.00272 (2016).

6 Akira, S., Uematsu, S. & Takeuchi, O. Pathogen recognition and innate immunity. *Cell* 124, 783-801, doi:10.1016/j.cell.2006.02.015 (2006).

7 van der Vaart, M., Spaink, H. P. & Meijer, A. H. Pathogen recognition and activation of the innate immune response in zebrafish. *Advances in hematology* 2012, 159807, doi:10.1155/2012/159807 (2012).

8 Iwasaki, A. & Medzhitov, R. Control of adaptive immunity by the innate immune system. *Nature immunology* 16, 343-353, doi:10.1038/ni.3123 (2015).

9 Cooper, M. D. & Alder, M. N. The evolution of adaptive immune systems. *Cell* 124, 815-822, doi:10.1016/j.cell.2006.02.001 (2006).

10 Flajnik, M. F. & Kasahara, M. Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Nature reviews. Genetics* 11, 47-59, doi:10.1038/nrg2703 (2010).

11 Dempsey, P. W., Vaidya, S. A. & Cheng, G. The art of war: Innate and adaptive immune responses. *Cellular and molecular life sciences : CMLS* 60, 2604-2621, doi:10.1007/s00018-003-3180-y (2003).

12 Boehm, T. Design principles of adaptive immune systems. *Nat Rev Immunol* 11, 307-317, doi:10.1038/nri2944 (2011).

13 Vitetta, E. S. *et al.* Memory B and T cells. *Annual review of immunology* 9, 193-217, doi:10.1146/annurev.iy.09.040191.001205 (1991).

14 Shlomchik, M. J. & Weisel, F. Germinal center selection and the development of memory B and plasma cells. *Immunological reviews* 247, 52-63, doi:10.1111/j.1600-065X.2012.01124.x (2012).

15 Medzhitov, R. Recognition of microorganisms and activation of the immune response. *Nature* 449, 819-826, doi:10.1038/nature06246 (2007).

16 Lindenmann, J. Origin of the terms 'antibody' and 'antigen'. *Scandinavian journal of immunology* 19, 281-285 (1984).

17 Jerne, N. K. The Natural-Selection Theory of Antibody Formation. *Proc Natl Acad Sci U S A* 41, 849-857 (1955).

18 Silverstein, A. M. The curious case of the 1960 Nobel Prize to Burnet and Medawar. *Immunology* 147, 269-274, doi:10.1111/imm.12558 (2016).

19 Tonegawa, S. Nobel lecture in physiology or medicine--1987. Somatic generation of immune diversity. *In vitro cellular & developmental biology : journal of the Tissue Culture Association* 24, 253-265 (1988).

20 Jerne, N. K. The generative grammar of the immune system. *The EMBO journal* 4, 847-852 (1985).

21 Milstein, C. From the structure of antibodies to the diversification of the immune response. *The EMBO journal* 4, 1083-1092 (1985).

22 Kohler, G. Derivation and diversification of monoclonal antibodies. *The EMBO journal* 4, 1359-1365 (1985).

23 Seda, V. & Mraz, M. B-cell receptor signalling and its crosstalk with other pathways in normal and malignant cells. *European journal of haematology* 94, 193-205, doi:10.1111/ejh.12427 (2015).

24 Borghesi, L. & Milcarek, C. From B cell to plasma cell: regulation of V(D)J recombination and antibody secretion. *Immunologic research* 36, 27-32, doi:10.1385/IR:36:1:27 (2006).

25 Parker, D. C. T cell-dependent B cell activation. *Annual review of immunology* 11, 331-360, doi:10.1146/annurev.iy.11.040193.001555 (1993).

26 Maverakis, E. *et al.* Glycans in the immune system and The Altered Glycan Theory of Autoimmunity: a critical review. *Journal of autoimmunity* 57, 1-13, doi:10.1016/j.jaut.2014.12.002 (2015).

27 Murphy, R. M. *et al.* Size and structure of antigen-antibody complexes.

Electron microscopy and light scattering studies. *Biophysical journal* 54, 45 56, doi:10.1016/S0006-3495(88)82929-1 (1988).

28 Sternberger, L. A., Hardy, P. H., Jr., Cuculis, J. J. & Meyer, H. G. The unlabeled antibody enzyme method of immunohistochemistry: preparation and properties of soluble antigen-antibody complex (horseradish peroxidase-antihorseradish peroxidase) and its use in identification of spirochetes. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* 18, 315-333, doi:10.1177/18.5.315 (1970).

29 Manser, T., Wysocki, L. J., Margolies, M. N. & Gefter, M. L. Evolution of antibody variable region structure during the immune response. *Immunological reviews* 96, 141-162 (1987).

30 Anderson, R. Manipulation of cell surface macromolecules by flaviviruses. *Advances in virus research* 59, 229-274 (2003).

31 Lefranc, M. P. Nomenclature of the human immunoglobulin kappa (IGK) genes. *Experimental and clinical immunogenetics* 18, 161-174, doi:49195 (2001).

32 Lefranc, M. P. Nomenclature of the human immunoglobulin lambda (IGL) genes. *Experimental and clinical immunogenetics* 18, 242-254, doi:49203 (2001).

33 Watson, C. T. & Breden, F. The immunoglobulin heavy chain locus: genetic variation, missing data, and implications for human disease. *Genes Immun* 13, 363-373, doi:10.1038/gene.2012.12 (2012).

34 Dreyer, W. J. & Bennett, J. C. The molecular basis of antibody formation: a paradox. *Proc Natl Acad Sci U S A* 54, 864-869 (1965).

35 Tonegawa, S. Somatic generation of antibody diversity. *Nature* 302, 575-581 (1983).

36 Pascual, V. & Capra, J. D. Human immunoglobulin heavy-chain variable region genes: organization, polymorphism, and expression. *Advances in immunology* 49, 1-74 (1991).

37 Tallmadge, R. L., Tseng, C. T., King, R. A. & Felippe, M. J. Developmental progression of equine immunoglobulin heavy chain variable region diversity. *Developmental and comparative immunology* 41, 33-43, doi:10.1016/j.dci.2013.03.020 (2013).

38 Sedrak, P., Hsu, K. & Mohan, C. Molecular signatures of anti-nuclear

antibodies--contribution of heavy chain framework residues. *Molecular immunology* 40, 491-499 (2003).

39 Ill, C. R. *et al.* Design and construction of a hybrid immunoglobulin domain with properties of both heavy and light chain variable regions. *Protein engineering* 10, 949-957 (1997).

40 Haidar, J. N. *et al.* A universal combinatorial design of antibody framework to graft distinct CDR sequences: a bioinformatics approach. *Proteins* 80, 896 912, doi:10.1002/prot.23246 (2012).

41 Sela-Culang, I., Kunik, V. & Ofran, Y. The structural basis of antibody-antigen recognition. *Front Immunol* 4, 302, doi:10.3389/fimmu.2013.00302 (2013).

42 Al-Lazikani, B., Lesk, A. M. & Chothia, C. Standard conformations for the canonical structures of immunoglobulins. *J Mol Biol* 273, 927-948, doi:10.1006/jmbi.1997.1354 (1997).

43 North, B., Lehmann, A. & Dunbrack, R. L., Jr. A new clustering of antibody CDR loop conformations. *J Mol Biol* 406, 228-256, doi:10.1016/j.jmb.2010.10.030 (2011).

44 Shirai, H., Kidera, A. & Nakamura, H. H3-rules: identification of CDR-H3 structures in antibodies. *FEBS letters* 455, 188-197 (1999).

45 Schroeder, H. W., Jr. & Cavacini, L. Structure and function of immunoglobulins. *The Journal of allergy and clinical immunology* 125, S41-52, doi:10.1016/j.jaci.2009.09.046 (2010).

46 Matsuda, F. *et al.* The complete nucleotide sequence of the human immunoglobulin heavy chain variable region locus. *The Journal of experimental medicine* 188, 2151-2162 (1998).

47 Pallares, N., Lefebvre, S., Contet, V., Matsuda, F. & Lefranc, M. P. The human immunoglobulin heavy variable genes. *Experimental and clinical immunogenetics* 16, 36-60, doi:19095 (1999).

48 Ruiz, M., Pallares, N., Contet, V., Barbi, V. & Lefranc, M. P. The human immunoglobulin heavy diversity (IGHD) and joining (IGHJ) segments. *Experimental and clinical immunogenetics* 16, 173-184, doi:19109 (1999).

49 Barbie, V. & Lefranc, M. P. The human immunoglobulin kappa variable (IGKV) genes and joining (IGKJ) segments. *Experimental and clinical immunogenetics* 15, 171-183 (1998).

50 Pallares, N., Frippiat, J. P., Giudicelli, V. & Lefranc, M. P. The human immunoglobulin lambda variable (IGLV) genes and joining (IGLJ) segments. *Experimental and clinical immunogenetics* 15, 8-18 (1998).

51 Bassing, C. H., Swat, W. & Alt, F. W. The mechanism and regulation of chromosomal V(D)J recombination. *Cell* 109 Suppl, S45-55 (2002).

52 Georgiou, G. *et al.* The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat Biotechnol* 32, 158-168, doi:10.1038/nbt.2782 (2014).

53 Wardemann, H. & Busse, C. E. Novel Approaches to Analyze Immunoglobulin Repertoires. *Trends Immunol* 38, 471-482, doi:10.1016/j.it.2017.05.003 (2017).

54 van Gent, D. C., Ramsden, D. A. & Gellert, M. The RAG1 and RAG2 proteins establish the 12/23 rule in V(D)J recombination. *Cell* 85, 107-113 (1996).

55 Hiom, K. & Gellert, M. Assembly of a 12/23 paired signal complex: a critical control point in V(D)J recombination. *Molecular cell* 1, 1011-1019 (1998).

56 Ma, Y., Lu, H., Schwarz, K. & Lieber, M. R. Repair of double-strand DNA breaks by the human nonhomologous DNA end joining pathway: the iterative processing model. *Cell cycle* 4, 1193-1200, doi:10.4161/cc.4.9.1977 (2005).

57 Mombaerts, P. *et al.* RAG-1-deficient mice have no mature B and T lymphocytes. *Cell* 68, 869-877 (1992).

58 Hess, J. *et al.* Induction of pre-B cell proliferation after de novo synthesis of the pre-B cell receptor. *Proc Natl Acad Sci U S A* 98, 1745-1750, doi:10.1073/pnas.041492098 (2001).

59 Schatz, D. G., Oettinger, M. A. & Schlissel, M. S. V(D)J recombination: molecular biology and regulation. *Annual review of immunology* 10, 359-383, doi:10.1146/annurev.iy.10.040192.002043 (1992).

60 Mansilla-Soto, J. & Cortes, P. VDJ recombination: Artemis and its in vivo role in hairpin opening. *The Journal of experimental medicine* 197, 543-547 (2003).

61 Rose, N. R. Molecular mimicry and clonal deletion: A fresh look. *Journal of theoretical biology* 375, 71-76, doi:10.1016/j.jtbi.2014.08.034 (2015).

62 Pelanda, R. & Torres, R. M. Central B-cell tolerance: where selection begins. *Cold Spring Harbor perspectives in biology* 4, a007146, doi:10.1101/cshperspect.a007146 (2012).

63 Su, T. T. & Rawlings, D. J. Transitional B lymphocyte subsets operate as distinct checkpoints in murine splenic B cell development. *J Immunol* 168, 2101-2110 (2002).

64 Giraud, M. *et al.* Fast multiclonal clusterization of V(D)J recombinations from high-throughput sequencing. *BMC genomics* 15, 409, doi:10.1186/1471-2164 15-409 (2014).

65 Maul, R. W. & Gearhart, P. J. AID and somatic hypermutation. *Advances in immunology* 105, 159-191, doi:10.1016/S0065-2776(10)05006-6 (2010).

66 Stavnezer, J. & Amemiya, C. T. Evolution of isotype switching. *Seminars in immunology* 16, 257-275, doi:10.1016/j.smim.2004.08.005 (2004).

67 De Silva, N. S. & Klein, U. Dynamics of B cells in germinal centres. *Nat Rev Immunol* 15, 137-148, doi:10.1038/nri3804 (2015).

68 Awdeh, Z. L., Williamson, A. R. & Askonas, B. A. Isoelectric focusing in polyacrylamide gel and its application to immunoglobulins. *Nature* 219, 66-67 (1968).

69 Pannetier, C. *et al.* The sizes of the CDR3 hypervariable regions of the murine T-cell receptor beta chains vary as a function of the recombined germ-line segments. *Proc Natl Acad Sci U S A* 90, 4319-4323 (1993).

70 Pannetier, C., Even, J. & Kourilsky, P. T-cell repertoire diversity and clonal expansions in normal and clinical samples. *Immunology today* 16, 176-181 (1995).

71 Ademokun, A. *et al.* Vaccination-induced changes in human B-cell repertoire and pneumococcal IgM and IgA antibody at different ages. *Aging Cell* 10, 922-930, doi:10.1111/j.1474-9726.2011.00732.x (2011).

72 Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74, 5463-5467 (1977).

73 Seidman, J. G. *et al.* Multiple related immunoglobulin variable-region genes identified by cloning and sequence analysis. *Proc Natl Acad Sci U S A* 75, 3881- 3885 (1978).

74 Wu, T. T. & Kabat, E. A. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *The Journal of experimental medicine* 132, 211-250 (1970).

75 Weitkamp, J. H. *et al.* Infant and adult human B cell responses to rotavirus share common immunodominant variable gene repertoires. *J Immunol* 171, 4680-4688 (2003).

76 Wrammert, J. *et al.* Rapid cloning of high-affinity human monoclonal antibodies against influenza virus. *Nature* 453, 667-671, doi:10.1038/nature06890 (2008).

77 Tiller, T. *et al.* Efficient generation of monoclonal antibodies from single human B cells by single cell RT-PCR and expression vector cloning. *J Immunol Methods* 329, 112-124, doi:10.1016/j.jim.2007.09.017 (2008).

78 Sant'Angelo, D. B. *et al.* A molecular map of T cell development. *Immunity* 9, 179-186 (1998).

79 Poulsen, T. R., Jensen, A., Haurum, J. S. & Andersen, P. S. Limits for antibody affinity maturation and repertoire diversification in hypervaccinated humans. *J Immunol* 187, 4229-4235, doi:10.4049/jimmunol.1000928 (2011).

80 Casrouge, A. *et al.* Size estimate of the alpha beta TCR repertoire of naive mouse splenocytes. *J Immunol* 164, 5782-5787 (2000).

81 Arstila, T. P. *et al.* A direct estimate of the human alphabeta T cell receptor diversity. *Science* 286, 958-961 (1999).

82 Glenn, T. C. Field guide to next-generation DNA sequencers. *Molecular ecology resources* 11, 759-769, doi:10.1111/j.1755-0998.2011.03024.x (2011).

83 Warren, E. H., Matsen, F. A. t. & Chou, J. High-throughput sequencing of B- and T-lymphocyte antigen receptors in hematology. *Blood* 122, 19-22, doi:10.1182/blood-2013-03-453142 (2013).

84 Boyd, S. D. *et al.* Measurement and clinical monitoring of human lymphocyte clonality by massively parallel VDJ pyrosequencing. *Sci Transl Med* 1, 12ra23 (2009).

85 Weinstein, J. A., Jiang, N., White, R. A., 3rd, Fisher, D. S. & Quake, S. R. High- throughput sequencing of the zebrafish antibody repertoire. *Science* 324, 807-810, doi:10.1126/science.1170020 (2009).

86 Loman, N. J. *et al.* Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol* 30, 434-439, doi:10.1038/nbt.2198 (2012).

87 Bashford-Rogers, R. J. *et al.* Capturing needles in haystacks: a comparison of B-cell receptor sequencing methods. *BMC immunology* 15, 29, doi:10.1186/s12865-014-0029-0 (2014).

88 Glanville, J. *et al.* Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. *Proc Natl Acad Sci U S A* 106, 20216-20221, doi:10.1073/pnas.0909775106 (2009).

89 Glanville, J. *et al.* Naive antibody gene-segment frequencies are heritable and unaltered by chronic lymphocyte ablation. *Proc Natl Acad Sci U S A* 108, 20066-20071, doi:10.1073/pnas.1107498108 (2011).

90 Schoettler, N., Ni, D. & Weigert, M. B cell receptor light chain repertoires show signs of selection with differences between groups of healthy individuals and SLE patients. *Molecular immunology* 51, 273-282, doi:10.1016/j.molimm.2012.03.028 (2012).

91 Galson, J. D., Pollard, A. J., Truck, J. & Kelly, D. F. Studying the antibody repertoire after vaccination: practical applications. *Trends Immunol* 35, 319-331, doi:10.1016/j.it.2014.04.005 (2014).

92 Zhu, J. *et al.* De novo identification of VRC01 class HIV-1-neutralizing antibodies by next-generation sequencing of B-cell transcripts. *Proc Natl Acad Sci U S A* 110, E4088-4097, doi:10.1073/pnas.1306262110 (2013).

93 Reddy, S. T. *et al.* Monoclonal antibodies isolated without screening by analyzing the variable-gene repertoire of plasma cells. *Nat Biotechnol* 28, 965-969, doi:10.1038/nbt.1673 (2010).

94 Menzel, U. *et al.* Comprehensive evaluation and optimization of amplicon library preparation methods for high-throughput antibody sequencing. *PloS one* 9, e96727, doi:10.1371/journal.pone.0096727 (2014).

95 Carlson, C. S. *et al.* Using synthetic templates to design an unbiased multiplex PCR assay. *Nature communications* 4, 2680, doi:10.1038/ncomms3680 (2013).

96 Ellefson, J. W. *et al.* Synthetic evolutionary origin of a proofreading reverse transcriptase. *Science* 352, 1590-1593, doi:10.1126/science.aaf5409 (2016).

97 Shugay, M. *et al.* Towards error-free profiling of immune repertoires. *Nat Methods* 11, 653-655, doi:10.1038/nmeth.2960 (2014).

98 Wang, B. *et al.* Corrigendum: Facile Discovery of a Diverse Panel of Anti-Ebola Virus Antibodies by Immune Repertoire Mining. *Sci Rep* 6, 27229, doi:10.1038/srep27229 (2016).

99 Best, K., Oakes, T., Heather, J. M., Shawe-Taylor, J. & Chain, B. Computational analysis of stochastic heterogeneity in PCR amplification efficiency revealed by single molecule barcoding. *Sci Rep* 5, 14629, doi:10.1038/srep14629 (2015).

100 Friedensohn, S., Khan, T. A. & Reddy, S. T. Advanced Methodologies in High- Throughput Sequencing of Immune Repertoires. *Trends in biotechnology* 35, 203-214, doi:10.1016/j.tibtech.2016.09.010 (2017).

101 Prabakaran, P., Streaker, E., Chen, W. & Dimitrov, D. S. 454 antibody sequencing - error characterization and correction. *BMC research notes* 4, 404, doi:10.1186/1756-0500-4-404 (2011).

102 Benichou, J., Ben-Hamo, R., Louzoun, Y. & Efroni, S. Rep-Seq: uncovering the immunological repertoire through next-generation sequencing. *Immunology* 135, 183-191, doi:10.1111/j.1365-2567.2011.03527.x (2012).

103 Kircher, M., Heyn, P. & Kelso, J. Addressing challenges in the production and analysis of illumina sequencing data. *BMC genomics* 12, 382, doi:10.1186/1471- 2164-12-382 (2011).

104 Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome research* 8, 186-194 (1998).

105 Bokulich, N. A. *et al.* Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat Methods* 10, 57-59, doi:10.1038/nmeth.2276 (2013).

106 Yaari, G. & Kleinstein, S. H. Practical guidelines for B-cell receptor repertoire sequencing analysis. *Genome Med* 7, 121, doi:10.1186/s13073-015-0243-2 (2015).

107 Safonova, Y. *et al.* IgRepertoireConstructor: a novel algorithm for antibody repertoire construction and immunoproteogenomics analysis. *Bioinformatics* 31, i53-61, doi:10.1093/bioinformatics/btv238 (2015).

108 Kuchenbecker, L. *et al.* IMSEQ--a fast and error aware approach to immunogenetic sequence analysis. *Bioinformatics* 31, 2963-2971, doi:10.1093/bioinformatics/btv309 (2015).

109 Bolotin, D. A. *et al.* MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods* 12, 380-381, doi:10.1038/nmeth.3364 (2015).

110 Kinde, I., Wu, J., Papadopoulos, N., Kinzler, K. W. & Vogelstein, B. Detection and quantification of rare mutations with massively parallel sequencing. *Proc Natl Acad Sci U S A* 108, 9530-9535, doi:10.1073/pnas.1105422108 (2011).

111 Jabara, C. B., Jones, C. D., Roach, J., Anderson, J. A. & Swanstrom, R. Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. *Proc Natl Acad Sci U S A* 108, 20166-20171, doi:10.1073/pnas.1110064108 (2011).

112 Kivioja, T. *et al.* Counting absolute numbers of molecules using unique molecular identifiers. *Nat Methods* 9, 72-74, doi:10.1038/nmeth.1778 (2011).

113 Kou, R. *et al.* Benefits and Challenges with Applying Unique Molecular Identifiers in Next Generation Sequencing to Detect Low Frequency Mutations. *PloS one* 11, e0146638, doi:10.1371/journal.pone.0146638 (2016).

114 Khan, T. A. *et al.* Accurate and predictive antibody repertoire profiling by molecular amplification fingerprinting. *Sci Adv* 2, e1501371, doi:10.1126/sciadv.1501371 (2016).

115 Rubelt, F. *et al.* Onset of immune senescence defined by unbiased pyrosequencing of human immunoglobulin mRNA repertoires. *PloS one* 7, e49774, doi:10.1371/journal.pone.0049774 (2012).

116 Busse, C. E., Czogiel, I., Braun, P., Arndt, P. F. & Wardemann, H. Single-cell based high-throughput sequencing of full-length immunoglobulin heavy and light chain genes. *Eur J Immunol* 44, 597-603, doi:10.1002/eji.201343917 (2014).

117 Murugan, R., Imkeller, K., Busse, C. E. & Wardemann, H. Direct high-throughput amplification and sequencing of immunoglobulin genes from single human B cells. *Eur J Immunol* 45, 2698-2700, doi:10.1002/eji.201545526 (2015).

118 Tan, Y. C. *et al.* High-throughput sequencing of natively paired antibody chains provides evidence for original antigenic sin shaping the antibody response to influenza vaccination. *Clinical immunology* 151, 55-65, doi:10.1016/j.clim.2013.12.008 (2014).

119 Marcus, J. S., Anderson, W. F. & Quake, S. R. Microfluidic single-cell mRNA isolation and analysis. *Analytical chemistry* 78, 3084-3089, doi:10.1021/ac0519460 (2006).

120 Toriello, N. M. *et al.* Integrated microfluidic bioprocessor for single-cell gene expression analysis. *Proc Natl Acad Sci U S A* 105, 20173-20178, doi:10.1073/pnas.0806355106 (2008).

121 White, A. K. *et al.* High-throughput microfluidic single-cell RT-qPCR. *Proc Natl Acad Sci U S A* 108, 13999-14004, doi:10.1073/pnas.1019446108 (2011).

122 Turchaninova, M. A. *et al.* Pairing of T-cell receptor chains via emulsion PCR. *Eur J Immunol* 43, 2507-2515, doi:10.1002/eji.201343453 (2013).

123 Williams, R. *et al.* Amplification of complex gene libraries by emulsion PCR. *Nat Methods* 3, 545-550, doi:10.1038/nmeth896 (2006).

124 DeKosky, B. J. *et al.* High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nat Biotechnol* 31, 166-169, doi:10.1038/nbt.2492 (2013).

125 DeKosky, B. J. *et al.* In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire. *Nat Med* 21, 86-91, doi:10.1038/nm.3743 (2015).

126 Abas, L. & Luschnig, C. Maximum yields of microsomal-type membranes from small amounts of plant material without requiring ultracentrifugation. *Anal Biochem* 401, 217-227, doi:10.1016/j.ab.2010.02.030 (2010).

127 Stephens, S. B., Dodd, R. D., Lerner, R. S., Pyhtila, B. M. & Nicchitta, C. V. Analysis of mRNA partitioning between the cytosol and endoplasmic reticulum compartments of mammalian cells. *Methods in molecular biology* 419, 197-214, doi:10.1007/978-1-59745-033-1_14 (2008).

128 Stuart-Keil, K. G., Hohnstock, A. M., Drees, K. P., Herrick, J. B. & Madsen, E. L. Plasmids responsible for horizontal transfer of naphthalene catabolism genes between bacteria at a coal tar-contaminated site are homologous to pDTG1 from pseudomonas putida NCIB 9816-4. *Applied and environmental microbiology* 64, 3633-3640 (1998).

129 Devynck, M. A., Pernollet, M. G., Matthews, P. G. & Meyer, P. [Demonstration of specific receptors for angiotensin III in rat adrenal glands]. *Comptes rendus hebdomadaires des seances de l'Academie des sciences. Serie D: Sciences naturelles* 284, 1293-1296 (1977).

130 Briney, B. S., Willis, J. R., Hicar, M. D., Thomas, J. W., 2nd & Crowe, J. E., Jr. Frequency and genetic characterization of V(DD)J recombinants in the human peripheral blood antibody repertoire. *Immunology* 137, 56-64, doi:10.1111/j.1365- 2567.2012.03605.x (2012).

131 Ge, Q. *et al.* Emulsion PCR-based method to detect Y chromosome microdeletions. *Anal Biochem* 367, 173-178, doi:10.1016/j.ab.2007.05.008 (2007).

132 Schutze, T. *et al.* A streamlined protocol for emulsion polymerase chain reaction and subsequent purification. *Anal Biochem* 410, 155-157, doi:10.1016/j.ab.2010.11.029 (2011).

133 Giudicelli, V., Chaume, D. & Lefranc, M. P. IMGT/V-QUEST, an integrated software program for immunoglobulin and T cell receptor V-J and V-D-J rearrangement analysis. *Nucleic Acids Res* 32, W435-440, doi:10.1093/nar/gkh412 (2004).

134 Caro, L. G. & Palade, G. E. Protein Synthesis, Storage, and Discharge in the Pancreatic Exocrine Cell. An Autoradiographic Study. *The Journal of cell biology* 20, 473-495 (1964).

135 Hermesh, O. & Jansen, R. P. Take the (RN)A-train: localization of mRNA to the endoplasmic reticulum. *Biochimica et biophysica acta* 1833, 2519-2525, doi:10.1016/j.bbamcr.2013.01.013 (2013).

136 Schneider-Poetsch, T. *et al.* Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin. *Nat Chem Biol* 6, 209-217, doi:10.1038/nchembio.304 (2010).

137 Kraut-Cohen, J. & Gerst, J. E. Addressing mRNAs to the ER: cis sequences act up! *Trends in biochemical sciences* 35, 459-469, doi:10.1016/j.tibs.2010.02.006 (2010).

138 Oyola, S. O. *et al.* Optimizing Illumina next-generation sequencing library preparation for extremely AT-biased genomes. *BMC genomics* 13, 1, doi:10.1186/1471-2164-13-1 (2012).

139 Truck, J. *et al.* Identification of antigen-specific B cell receptor sequences using public repertoire analysis. *J Immunol* 194, 252-261, doi:10.4049/jimmunol.1401405 (2015).

140 Galson, J. D. *et al.* BCR repertoire sequencing: different patterns of B-cell activation after two Meningococcal vaccines. *Immunol Cell Biol* 93, 885-895, doi:10.1038/icb.2015.57 (2015).

141 Parameswaran, P. *et al.* Convergent antibody signatures in human dengue. *Cell Host Microbe* 13, 691-700, doi:10.1016/j.chom.2013.05.008 (2013).

142 Jiang, N. *et al.* Lineage structure of the human antibody repertoire in response to influenza vaccination. *Sci Transl Med* 5, 171ra119, doi:10.1126/scitranslmed.3004794 (2013).

143 Demaison, C. *et al.* Analysis of human VH gene repertoire expression in peripheral CD19+ B cells. *Immunogenetics* 42, 342-352 (1995).

144 Frolich, D. *et al.* Secondary immunization generates clonally related antigen-specific plasma cells and memory B cells. *J Immunol* 185, 3103-3110, doi:10.4049/jimmunol.1000911 (2010).

145 Franz, B., May, K. F., Jr., Dranoff, G. & Wucherpfennig, K. Ex vivo characterization and isolation of rare memory B cells with antigen tetramers. *Blood* 118, 348-357, doi:10.1182/blood-2011-03-341917 (2011).

146 Lavinder, J. J. *et al.* Identification and characterization of the constituent human serum antibodies elicited by vaccination. *Proc Natl Acad Sci U S A* 111, 2259- 2264, doi:10.1073/pnas.1317793111 (2014).

147 Hood, L. & Talmage, D. W. Mechanism of antibody diversity: germ line basis for variability. *Science* 168, 325-334 (1970).

148 Leder, P. *et al.* The organization and diversity of immunoglobulin genes. *Proc Natl Acad Sci U S A* 71, 5109-5115 (1974).

149 Ehlich, A., Martin, V., Muller, W. & Rajewsky, K. Analysis of the B-cell progenitor compartment at the level of single cells. *Current biology : CB* 4, 573-583 (1994).

150 Klein, U., Rajewsky, K. & Kuppers, R. Human immunoglobulin (Ig)M+IgD+ peripheral blood B cells expressing the CD27 cell surface antigen carry somatically mutated variable region genes: CD27 as a general marker for somatically mutated (memory) B cells. *The Journal of experimental medicine* 188, 1679-1689 (1998).

151 Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nat Biotechnol* 26, 1135- 1145, doi:10.1038/nbt1486 (2008).

152 Mardis, E. R. Next-generation sequencing platforms. *Annual review of analytical chemistry* 6, 287-303, doi:10.1146/annurev-anchem-062012-092628 (2013).

153 Quail, M. A. *et al.* A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC genomics* 13, 341, doi:10.1186/1471-2164-13-341 (2012).

154 Arnaout, R. *et al.* High-resolution description of antibody heavy-chain repertoires in humans. *PloS one* 6, e22365, doi:10.1371/journal.pone.0022365 (2011).

155 Lim, T. S. *et al.* V-gene amplification revisited - An optimised procedure for amplification of rearranged human antibody genes of different isotypes. *N Biotechnol* 27, 108-117, doi:10.1016/j.nbt.2010.01.001 (2010).

156 Robinson, W. H. Sequencing the functional antibody repertoire--diagnostic and therapeutic discovery. *Nat Rev Rheumatol* 11, 171-182, doi:10.1038/nrrheum.2014.220 (2015).

157 Graham, J. M. Preparation of crude subcellular fractions by differential centrifugation. *TheScientificWorldJournal* 2, 1638-1642, doi:10.1100/tsw.2002.851 (2002).

158 Morin, F., Tay, S. & Simpkins, H. A comparative study of the molecular structures of the plasma membranes and the smooth and the rough endoplasmic-reticulum membranes from rat liver. *The Biochemical journal* 129, 781-788 (1972).

159 Sukhodub, A. L. & Burchell, A. Preparation of intact microsomes from cultured mammalian H4IIE cells. *Journal of pharmacological and toxicological methods* 52, 330-334, doi:10.1016/j.vascn.2005.04.016 (2005).

160 Weyant, R. S., Edmonds, P. & Swaminathan, B. Effect of ionic and nonionic detergents on the Taq polymerase. *BioTechniques* 9, 308-309 (1990).

161 Schrader, C., Schielke, A., Ellerbroek, L. & Johne, R. PCR inhibitors - occurrence, properties and removal. *Journal of applied microbiology* 113, 1014- 1026, doi:10.1111/j.1365-2672.2012.05384.x (2012).

162 Smith, K. *et al.* Rapid generation of fully human monoclonal antibodies specific to a vaccinating antigen. *Nat Protoc* 4, 372-384, doi:10.1038/nprot.2009.3 (2009).

163 Lang, M. L. How do natural killer T cells help B cells? *Expert review of vaccines* 8, 1109-1121, doi:10.1586/erv.09.56 (2009).

164 Airoldi, I. *et al.* Heterogeneous expression of interleukin-18 and its receptor in B-cell lymphoproliferative disorders deriving from naive, germinal center, and memory B lymphocytes. *Clinical cancer research : an official journal of the American Association for Cancer Research* 10, 144-154 (2004).

165 McHeyzer-Williams, M., Okitsu, S., Wang, N. & McHeyzer-Williams, L. Molecular programming of B cell memory. *Nat Rev Immunol* 12, 24-34, doi:10.1038/nri3128 (2011).

166 Hou, D., Chen, C., Seely, E. J., Chen, S. & Song, Y. High-Throughput Sequencing-Based Immune Repertoire Study during Infectious Disease. *Front Immunol* 7, 336, doi:10.3389/fimmu.2016.00336 (2016).

167 Galson, J. D. *et al.* Analysis of B Cell Repertoire Dynamics Following Hepatitis B Vaccination in Humans, and Enrichment of Vaccine-specific Antibody Sequences. *EBioMedicine* 2, 2070-2079, doi:10.1016/j.ebiom.2015.11.034 (2015).

168 Klein, U., Kuppers, R. & Rajewsky, K. Evidence for a large compartment of IgM- expressing memory B cells in humans. *Blood* 89, 1288-1298 (1997).

169 Mroczek, E. S. *et al.* Differences in the composition of the human antibody repertoire by B cell subsets in the blood. *Front Immunol* 5, 96, doi:10.3389/fimmu.2014.00096 (2014).

170 Jackson, S. M., Wilson, P. C., James, J. A. & Capra, J. D. Human B cell subsets. *Advances in immunology* 98, 151-224, doi:10.1016/S0065-2776(08)00405-7 (2008).

171 Boyd, S. D. & Joshi, S. A. High-Throughput DNA Sequencing Analysis of Antibody Repertoires. *Microbiology spectrum* 2, doi:10.1128/microbiolspec.AID- 0017-2014 (2014).

172 McDaniel, J. R., DeKosky, B. J., Tanno, H., Ellington, A. D. & Georgiou, G. Ultra- high-throughput sequencing of the immune receptor repertoire from millions of lymphocytes. *Nat Protoc* 11, 429-442, doi:10.1038/nprot.2016.024 (2016).

Appendix

6.1 HC and Kappa LC primer sequences used for overlap extension RT-PCR and subsequent nested PCR amplifications.

All primers used in thesis were obtained from *meta***bio**n, Germany. The isotype specific reverse primers are denoted as “rev”, while VH/VK overlap extended forward primers are denoted as “fwd”. The isotype specific nested PCR are denoted as “nested”.

Working concentration (nM)	**Primer name**	**Primer sequence (5’-3’)**
375	IgM_RT_rev	CACAGGAGACGAGGGGGAAA
375	IgG_RT_rev	AGGGYGCCAGGGGGAAGAC
375	IgK_RT_rev	GATGAAGACAGATGGTGCAG
17.5	Overlap_VH1/7_fwd	tcgtgcctatatccttactgactctgcGGCCTCAGTGAAGGTCTCCTGCAAG
17.5	Overlap_VH2_fwd	tcgtgcctatatccttactgactctgcGTCTGGTCCTACGCTGGTGAAACCC
17.5	Overlap_VH3_fwd	tcgtgcctatatccttactgactctgcCTGGGGGGTCCCTGAGACTCTCCTG
17.5	Overlap_VH4_fwd	tcgtgcctatatccttactgactctgcCTTCGGAGACCCTGTCCCTCACCTG
17.5	Overlap_VH5_fwd	tcgtgcctatatccttactgactctgcCGGGGAGTCTCTGAAGATCTCCTGT
17.5	Overlap_VH6_fwd	tcgtgcctatatccttactgactctgcTCGCAGACCCTCTCACTCACCTGTG
7	Overlap_VK1_fwd	gcagagtcagtaaggatataggcacgaGACATCCRGDTGACCCAGTCTCC
7	Overlap_VK2/4/6_fwd	gcagagtcagtaaggatataggcacgaGGGATATTGTGMTGACYCAGWCTCC
7	Overlap_VK3_fwd	gcagagtcagtaaggatataggcacgaGGAGAAATTGTRWTGACRCAGTCTCC
7	Overlap_VK5_fwd	gcagagtcagtaaggatataggcacgaGCAGAAACGACACTCACGCAGTCTC
400	IgM_nested	GGTTGGGGCGGATGCACTCC
400	IgG_nested	SGATGGGCCCTTGGTGGARGC
400	IgK_nested	CAGATGGTGCAGCCACAGTTC

Appendix Table 1: Overlap-extension RT-PCR and subsequent nested PCR primers

6.2 Primers used in first-step RT-PCR and second step PCR amplification for amplification and cloning of fourteen selected clonotypes

First-step RT-PCR HC and LC V-region forward primers with restriction digestion sites are denoted by “FR1_Fwd” and CDR3 clonotype-specific reverse primers are denoted by “CDR3_Rev”. Primers used in second step PCR amplification are denoted by “RD_FR1_Fw” for forward primers and “RD_FR4_Rev” for reverse primers.

Primer name	**Primer sequence (5’-3’)**
HC_VH1_FR1_Fwd	ATCTAAGCGCGCACTCCCAGGTCCAGCTKGTRCAGTCTGG
HC_VH3_FR1_Fwd	ATCTAAGCGCGCACTCCGAGGTGCAGCTGKTGGAGWCY
HC_VH4_FR1_Fwd	ATCTAAGCGCGCACTCCCAGGTGCAGCTGCAGGAGTCSG
LC_VK1_FR1_Fwd	ATATAAGCGCGCACTCCGACATCCRGDTGACCCAGTCTCC
LC_VK246_FR1_Fwd	ATATAAGCGCGCACTCCGATATTGTGMTGACBCAGWCTCC
LC_VK3_FR1_Fwd	ATATAAGCGCGCACTCCGAAATTGTRWTGACRCAGTCTCC
LC_VK5_FR1_Fwd	ATATAAGCGCGCACTCCGAAACGACACTCACGCAGTCTC
HC_Top1_CDR3_Rev	GACCAGGGTTCCCTGGCCCCAGTAGTCAAAGACATAGTTAATGAGCCACTGCTGTCTCGCTTTCGC
HC_Top2_CDR3_Rev	GACCATTGTCCCTTGGCCCCAGATATCAAGAGCATAACCATAATTGTATCCCCTAG
HC_Top3_CDR3_Rev	GACCAGGGTTCCCTGGCCCCAGTAATCCAACGTCCTCCCACTCGCACA
HC_Top4_CDR3_Rev	GACCAGGGTTCCCTGGCCCCAGTAGTCAAAGTACTTCTCCCAATTCAGCTCTCTCGC
HC_Top5_CDR3_Rev	CCAGATATCAAAAGCATTCCAGCTGGTACTATAGCCGACATTCCTATCTCTCGC
HC_Top6_CDR3_Rev	GACCATTGTCCCTTGGCCCCAGAGATCGAAAGGAACGAGAGCAGCTGGAACAACTAAAATATCGCGAGC CCCCCCTCT
HC_Top7_CDR3_Rev	GACCAGGTTTCCCTGGCCCCAGGGGTCGAACCAGTTCAGGGCGTCACGAATCGCTTTAGTTTCTCTCGC
HC_Top8_CDR3_Rev	GACCAGGGTTCCCTGGCCCCAGTAGTCAAAGTATCCGGCTCCGTGAACCATAATGGTATCTCTCGC
HC_Top9_CDR3_Rev	GACCAGGGTGCCACGGCCCCAGAGAGCGAAGTACCAGTAGGGGATGAGATAACCATTCAAAATATCGTA ATTGAT
HC_Top10_CDR3_Rev	GACCATTGTCCCTTGGCCCCAGAGATCGAAAGGAACGAGAGCAGCTGGAACACCTAAAATATCGCGAGT CC
HC_Top11_CDR3_Rev	GACCAGACTTCCCTGGCCCCAGTGGTCAGAGTAGCCGACTCCCCTAACCATACGGGTATCTCTCGC
HC_Top12_CDR3_Rev	GACCAGGGTTCCCTGGCCCCAGAAGGAGTAGTAACCAGAAGTATTATAGTAATGAGTTGC
HC_Top13_CDR3_Rev	GACCACTGTCCCTTGGCCCCAGACATCAAAAGCATTCCAGCTGGTACTATAGCCGACATTCCTATCTCTC GC
HC_Top14_CDR3_Rev	GACCAGGGTTCCCTGGCCCCAGTAGTCAAAACTAGGTCGTCCCATAGCTGTATTCGGCCGACTTCTCGC
Known_HC_CDR3_Rev	GACCAGGGTTCCCTGGCTCCAGGGGTCGAACCAGTTGTCAGCCTGTCTC
LC_Top1_CDR3_Rev	CACCTTGGTCCCTCCGCCGAAAGTGAGCGGGAGACTGTTAGCCTGTTG
LC_Top2_CDR3_Rev	CACTTTGGTCCCAGGGCCGAAAGTGAAAGGAGTTTGCAGACCTTGCAT
LC_Top3_CDR3_Rev	CACCTTGGTCCCTTGGCCGAACGTCCGATAACTATTATACTGTTGGCA
LC_Top4_CDR3_Rev	CACCTTGGTCCCAGGGCCGAAAGTGAATCGGGTGTAACTCTGTTG
LC_Top5_CDR3_Rev	CACCTTGGTCCCTTGGCCGAACGTTCGAGAATAACTATTATACTGTTG
LC_Top6_CDR3_Rev	CAGTCGTGTCCCTTGGCCGAAGGTGATCGGAATACCATAACATTGCTG
LC_Top7_CDR3_Rev	CACCTTGGTCCCCCCGCCGAAAGTGAAGACGAAACTGTAACTCTGTAG

LC_Top8_CDR3_Rev	CACCTTGGTCCCTCCGCCGAAAGTGAGCGGGGTACTGTAACTCTGTTG
LC_Top9_CDR3_Rev	CACTTTGGTCCCAGGGCCGAAAGTGAACCTCGACTGTTG
LC_Top10_CDR3_Rev	CAGTCGTGTCCCTTGGCCGAAGGTGATCGGAATACCATAACATTGCTG
LC_Top11_CDR3_Rev	CACCTTGGTCCCTCCGCCGAAAGTGAGCGGGAAACTGTTAGCCTGCTG
LC_Top12_CDR3_Rev	CAGCTTGGTCCCCTGGCCAAAAGTGTACAGGGGACTGGAAGTCTGTTG
LC_Top13_CDR3_Rev	CACCTTGGTCCCTTGGCCGAACGTTCGAGAATAACTATTATACTGTTG
LC_Top14_CDR3_Rev	CACCTTGGTCCCTCCGCCGAAAGTGAGCGGGGTACTGTAACTCTGTTG
Known_LC_CDR3_Rev	CAGCTTGGTCCCCTGGCCAAAAGTTAGAGGGAAATCATCATGTTGTAG
HC_VH1_RD_FR1_Fwd	ATCTAAGCGCGCACTCCCAG
HC_VH3_RD_FR1_Fwd	ATCTAAGCGCGCACTCCGAG
LC_VK1_RD_FR1_Fwd	ATATAAGCGCGCACTCCGAH
LC_VK3_RD_FR1_Fwd	ATATAAGCGCGCACTCCGAA
HC_Top2_RD_FR4_Rev	AAATTTGCTAGCGCTCGAGTGAAGAGACGGTGACCATTGTCCCTTGGCCCC
HC_Top1_Top3_Top4_ Top8_Top12_RD_FR4_Rev	AAATTTGCTAGCGCTCKAGGAGGAGACGGWGACCAGGGTTCCCTGGCC
HC_Top5_RD_FR4_Rev	AAATTTGCTAGCGCTCGAGGAAGAGACGGTGACCACTGTCCCTTGGCCCCAGATATCAAAAGCATT
HC_Top6_Top10_FR4_ RD_Rev	AAATTTGCTAGCGCTCGAGGAAGAGATGGTGACCATTGTCCCTTGGCC
HC_Top7_RD_FR4_Rev	AAATTTGCTAGCGCTCGAGGAGGAGACGGTGACCAGGTTTCCCTGGCC
HC_Top9_RD_FR4_Rev	AAATTTGCTAGCGCTCGAGGAGGAGACAGCGACCAGGGTGCCACGGCC
HC_Top11_RD_FR4_Rev	AAATTTGCTAGCGCTCGAGGAGGAGACGGTGACCAGACTTCCCTGGCC
HC_Top13_RD_FR4_Rev	AAATTTGCTAGCGCTCGAGGAAGAGACGGTGACCACTGTCCCTTGGCC
HC_Top14_RD_FR4_Rev	AAATTTGCTAGCGCTCGAGGAGGAGACGGTGACCAGGGTTCCCTGGCC
HC_Known_RD_FR4_Rev	AAATTTGCTAGCGCTCGAGGAGGAGACGGTGACCAGGGTTCCCTGGCT
LC_Top2_RD_FR4_Rev	ATATTAAAGCTTAAGCTTTTTGATATCCACTTTGGTCCCAGGGCC
LC_Top3_Top5_Top13_ RD_FR4_Rev	ATTATAAAGCTTTTTGATTTCCACCTTGGTCCCTTGGCC
LC_Top1_Top8_Top11_ Top14_RD_FR4_Rev	ATTATAAAGCTTTTTGATCTCCACCTTGGTCCCTCCGCC
LC_Top4_RD_FR4_Rev	ATTATAAAGCTTTTTGATATCCACCTTGGTCCCAGGGCC
LC_Top6_RD_FR4_Rev	ATTATAAAGCTTTTTAATCTCCAGTCGTGTCCCTTGGCC
LC_Top7_RD_FR4_Rev	ATTATAAAGCTTTCTGATCTCCACCTTGGTCCCCCCGCC
LC_Top9_RD_FR4_Rev	ATTATAAAGCTTTTTGATATCCACTTTGGTCCCAGGGCC
LC_Top10_RD_FR4_Rev	ATTATAAAGCTTTTTAATCTCCAGTCGTGTCCCTTGGCC
LC_Top12_RD_FR4_Rev	ATTATAAAGCTTTTTGATCTCCASCTTGGTCCCCTGGCC
LC_known_RD_FR4_Rev	ATTATAAAGCTTTTTGATGTCCAGCTTGGTCCCCTGGCC

Appendix table 2: Primer sequences for first-step RT-PCR and second-step PCR amplifications

6.3 Shell script used for retracting Top1 and Top2 LC-CDR3 amino acid paired to a given HC-CDR3 sequence in the analyzed paired antibody HC-LC sequencing data

```
#!/bin/sh
# Usage: sh lg_HC-LC_for_loop.sh INPUT_FILE1 INPUT_FILE2 OUTPUT_DIRECTORY
INPUT_FILE1=$1;
INPUT_FILE2=$2;
OUTPUT_DIRECTORY=$3;
#echo "determine the frequency of all HC CDR3 sequences...";
#echo " cut column 4 (HC_CDR3_A.A) from the HC_CDR3_clonotypes table obtained from MiXCR software...";
#cut -f4 $INPUT_FILE2 > ${OUTPUT_DIRECTORY}/HC_CDR3.txt;
#Reference to remove the first line
#echo "$(tail -n +2 file.txt)" > file.txt
#echo "Remove the header from the extracted HC CDR3 clonotypes from input file 2...";
#echo "$(tail -n +2 ${OUTPUT_DIRECTORY}/HC_CDR3.txt)" > ${OUTPUT_DIRECTORY}/IgM_HC_CDR3.txt
#echo "cut HC CDR3 sequence...";
#sort -nr ${OUTPUT_DIRECTORY}/frequency_all_lines.txt | perl -ne '$_=~ s/^[0-9 \s]+//gi; print ' > ${OUTPUT_DIRECTORY}/HC_CDR3_sequences_sort.txt;
#if false
#then
#HC_CDR3_COUNT=$(wc -l ${OUTPUT_DIRECTORY}/HC_CDR3_CLONOTYPES.txt | awk '{print $1}')
#echo "Count the number of lines in the file HC_CDR3_CLONOTYPES.txt - $count1"
#echo $HC_CDR3_COUNT
#if false
#then
# Lines as variable count
#echo ${`wc -l ${INPUT_FILE1} | grep -oh "[0-9]*"`};
var=`wc -l ${INPUT_FILE1} | perl -ne '$_=~ s/^\s+//gi; $_=~ s/\s+.*$//gi; print $_."\n"' | head -1` ;
#grep -oh "[\d]*"`
echo "$var"
for ((a=1; a<=$var ; a++))
do
#echo "$a";
#Put the top HC CDR3 sequence from the derived HC_CDR3_CLONOTYPES.txt file into a variable CDR3_SEQUENCE
#CDR3_SEQUENCE=$(head -${a} ${INPUT_FILE1} | tail -1);
CDR3_SEQUENCE=`head -${a} ${INPUT_FILE1} | tail -1`
#echo "$CDR3_SEQUENCE";
#Check whether the field 7 of paired alignment file is same as field 4 from the HC_CDR3_clonotypes table from MiXCR and extract the paired HC-LC sequences into top${a}.txt file in the output directory
awk -v CDR3=$CDR3_SEQUENCE '($7 == CDR3) {print}' $INPUT_FILE2 > ${OUTPUT_DIRECTORY}/top${a}.txt;
#echo "$CDR3";
#Count the number of lines with HC CDR3 sequences from the top${a}.txt files from the output directory
TOP_HC_CDR3=$( cat ${OUTPUT_DIRECTORY}/top${a}.txt | wc -l);
#echo "$TOP_HC_CDR3";
#echo "determine the number for top LC CDR3 sequence...";
#cut field 14 (LC CDR3 A.A sequence) from the paired HC-LC sequences in top${a}.txt file, sort them according to the unique CDR3 sequence in the reverse order and get the count of the top1 LC CDR3 and top2 LC CDR3 sequence
```

```
#cut -f14 ${OUTPUT_DIRECTORY}/top${a}.txt | sort | uniq > ${OUTPUT_DIRECTORY}/LC_CDR3/top${a}.txt
#python /Users/Praneeth/Desktop/grep.py -a
# Remove column 1 (sequence ID) from the extracted files in the output directory and preint all the lines into seperate output directory.
#awk '{ $1="";print}' top6.txt > top6_output.txt
TOP1_LC_VJ=$( cut -f9,11,14 ${OUTPUT_DIRECTORY}/top${a}.txt | sort | uniq -c | sort -nr | perl -ne '$_=~ s/^[0-9 \s]+//gi; print ' | head -1);
#echo "$TOP1_LC_VJ";
TOP1_LC_CDR3=$( cut -f14 ${OUTPUT_DIRECTORY}/top${a}.txt | sort | uniq -c | sort -nr | perl -ne '$_=~ s/^\s+//gi; $_=~ s/\s+.*$//gi; print $_."\n"' | head -1 );
#echo "$TOP1_LC_CDR3";
TOP1_LC_CDR3_SEQUENCE=$( cut -f14 ${OUTPUT_DIRECTORY}/top${a}.txt | sort | uniq -c | sort -nr | perl -ne '$_=~ s/^[0-9 \s]+//gi; print ' | head -1);
#echo "$TOP1_LC_CDR3_SEQUENCE";
#awk -v TOP1_LC_CDR3_LINES=$TOP1_LC_CDR3_SEQUENCE '($14 == CDR3){print}'${OUTPUT_DIRECTORY}/top${a}.txt > ${OUTPUT_DIRECTORY}/top${a}_LC_CDR3.txt;
#echo "calculate percentage of correct pairing of HC-LC sequences...";
TOP1_HC_LC_PAIRING=`echo $TOP1_LC_CDR3*100/$TOP_HC_CDR3 | bc -l`;
#echo $TOP1_HC_LC_PAIRING;
#echo "calculating the percentage of other HC-LC pairs...";
#VARIABLE=`echo $TOP_HC_CDR3-$TOP1_LC_CDR3 | bc -l`;
#echo $VARIABLE;
#TOP1_MISPAIRING=`echo $VARIABLE*100/$TOP_HC_CDR3 | bc -l`;
#echo $TOP1_MISPAIRING;
#echo "determine the count and CDR3 sequence for top2 LC CDR3...";
TOP2_LC_VJ=$( cut -f9,11,14 ${OUTPUT_DIRECTORY}/top${a}.txt | sort | uniq -c | sort -nr | perl -ne '$_=~ s/^[0-9 \s]+//gi; print ' | head -2 |tail -1);
#echo "$TOP2_LC_VJ";
TOP2_LC_CDR3=$( cut -f14 ${OUTPUT_DIRECTORY}/top${a}.txt | sort | uniq -c | sort -nr | perl -ne '$_=~ s/^\s+//gi; $_=~ s/\s+.*$//gi; print $_."\n"' | head -2 |tail -1 );
#echo "$TOP2_LC_CDR3";
TOP2_LC_CDR3_SEQUENCE=$( cut -f14 ${OUTPUT_DIRECTORY}/top${a}.txt | sort | uniq -c | sort -nr | perl -ne '$_=~ s/^[0-9 \s]+//gi; print ' | head -2 | tail -1);
#echo "$TOP2_LC_CDR3_SEQUENCE";
#echo "calculate percentage of correct pairing of HC-LC sequences...";
TOP2_HC_LC_PAIRING=`echo $TOP2_LC_CDR3*100/$TOP_HC_CDR3 | bc -l`;
#echo $TOP2_HC_LC_PAIRING;
#echo "calculating the percentage of other HC-LC pairs...";
#TOP2_VARIABLE=`echo $TOP_HC_CDR3-$TOP2_LC_CDR3 | bc -l`;
#echo $TOP2_VARIABLE;
#TOP2_MISPAIRING=`echo $TOP2_VARIABLE*100/$TOP_HC_CDR3 | bc -l`;
#echo $TOP2_MISPAIRING;
#echo "determine the fold difference between top1 and top2 LC CDR3 sequences to the paired HC CDR3 sequence"

FOLD_DIFFERENCE=`echo $TOP1_HC_LC_PAIRING/$TOP2_HC_LC_PAIRING | bc -l`;
#echo FOLD_DIFFERENCE;
#echo -e $CDR3_SEQUENCE"\t"$TOP1_LC_CDR3_SEQUENCE"\t"$TOP2_LC_CDR3_SEQUENCE"\t"$TOP_HC_CDR3"\t"$TOP1_LC_CDR3"\t"$TOP2_LC_CDR3"\t"$TOP1_HC_LC_PAIRING"\t"$TOP2_HC_LC_PAIRING"\t"$FOLD_DIFFERENCE >> ${OUTPUT_DIRECTORY}/CDR3_results.txt;
```

```
echo-e
$CDR3_SEQUENCE"\t"$TOP1_LC_CDR3_SEQUENCE"\t"$TOP1_LC_VJ"\t"$TOP2_LC_CDR3_SEQUENCE"\t"$TOP2_LC_
VJ"\t"$TOP_HC_CDR3"\t"$TOP1_LC_CDR3"\t"$TOP2_LC_CDR3"\t"$TOP1_HC_LC_PAIRING"\t"$TOP2_HC_LC_PAIRING"\t
"$FOLD_DIFFERENCE
done;
#echo "done.";
#echo "";
exit;
#
#fi
```

6.4 R script used for generating heat maps for analysis of immune repertoire changes in before and after Td booster vaccination samples

```
data<- read.delim2(file = '~/Desktop/heatmaps/Ig_combined_sequences/IgM_prevac_combined_analysis.csv',header = T,sep =
        ',')
data<-as.matrix(data)
data.num<- matrix(as.numeric(data),nrow = nrow(data),byrow = F)
colnames(data.num)<- colnames(data)
rownames(data.num)<- rownames(data)
#max.data.num <- max(data.num)
#data.num <- data.num / max.data.num

data.num[(data.num == 0)] <- 0
data.num[(data.num <= 0.001) & (data.num > 0)] <- 0.001
data.num[(data.num <= 0.01) & (data.num > 0.001)] <- 0.01
data.num[(data.num <= 1) & (data.num > 0.01)] <- 1
data.num[(data.num <= 5) & (data.num > 1)] <- 5
data.num[(data.num <= 15) & (data.num > 5)] <- 15
data.num[(data.num <= 30) & (data.num > 15)] <- 30
data.num[(data.num <= 100) & (data.num > 30)] <- 100


#data.num <- data.num * max.data.num

#install.packages('ggplot2')
library(ggplot2)

library(reshape2)

########################################
########################################

data.melt = melt(data.num)
colnames(data.melt)<- c('time', 'concentration', 'value')

require(grid)

pals<- c('white', 'deepskyblue',
    'deepskyblue2','deepskyblue3','deepskyblue4','indianred4', 'red')
percentage <- c(0, 0.0001, 0.001, 0.01, 0.05, 0.15, 0.3, 1)

pdf(file = '~/Desktop/heatmap.pdf',width = 7, height=4)

ggplot(data.melt, aes(concentration, time, fill = value)) +
 geom_tile(aes(fill = value)) +
 scale_fill_gradientn(colours=pals, na.value = "white",guide =
          'legend', name="to modify",values=percentage) +
 scale_x_discrete(limits=colnames(data.num)) +
 scale_y_discrete(limits=rownames(data.num)) +
 ggtitle("TITLE") +
 theme(axis.text.x = element_text(size=10, angle=45, hjust = 1, vjust
              = 1), axis.text.y = element_text(size=10), legend.key.height =
    unit(1,"cm"), legend.key.width = unit(1, "cm"))
dev.off
```

Publications

1) Devulapally PR, Jörg Bürger, Thorsten Mielke, Zoltán Konthur, Hans Lehrach, Marie-Laure Yaspo, Jörn Glöklar, Hans-Jörg Warnatz. Simple paired heavy- and light –chain antibody repertoire sequencing using endoplasmic reticulum microsomes. *(Manuscript in press)*