Aus dem Institut für Virologie
der Universität zu Köln
Direktor: Universitätsprofessor Dr. med. F. Klein

Advancing single cell and multiplex cloning strategies to identify broadly neutralizing antibodies targeting infectious pathogens

Inaugural-Dissertation zur Erlangung der Doktorwürde
der Medizinischen Fakultät
der Universität zu Köln

vorgelegt von Nathalie Lehnen
aus Neuss

promoviert am 14. November 2022

Dekan: Universitätsprofessor Dr. med. G. R. Fink

1. Gutachter: Universitätsprofessor Dr. med. F. Klein

2. Gutachter: Universitätsprofessor Dr. rer. nat. H. Kashkar

Erklärung

Ich erkläre hiermit, dass ich die vorliegende Dissertationsschrift ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Bei der Auswahl und Auswertung des Materials sowie bei der Herstellung des Manuskriptes habe ich Unterstützungsleistungen von folgenden Personen erhalten:

Herr Prof. Dr. med. Florian Klein
Herr Dr. rer. nat. Christoph Kreer
Herr Dr. rer. nat. Matthias Döring

Weitere Personen waren an der Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich nicht die Hilfe einer Promotionsberaterin/eines Promotionsberaters in Anspruch genommen. Dritte haben von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertationsschrift stehen.

Die Dissertationsschrift wurde von mir bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Dr. Meryem Seda Ercanoglu führte die FACS Sorts durch. BG505-SOSIP.664+ Einzel-B Zellen wurden von Dr. med. Dr. nat. med. Philipp Schommers isoliert und zur Verfügung gestellt. Alle anderen, dieser Arbeit zu Grunde liegenden Experimente, sowie die Erhebung des zugrundeliegenden Datensatzes wurden nach entsprechender Anleitung durch Dr. Christoph Kreer von mir selbst durchgeführt. Dr. Matthias Döring entwickelte openPrimeR. Die Auswertung und Analyse der primer-template (PTP) Eigenschaften wurde von Matthias Döring

unter der Benutzung des openPrimer tools durchgeführt, graphisch visualisiert und aus der gemeinsamen Publikation übernommen (Unterpunkt 3.2.8, Unterpunkt 3.2.8.2., table 33, table 34, and figure 32). Die Taget-Funktionen zur Primer Optimierung wurden von Dr. Christoph Kreer formuliert und aus der gemeinsamen Publikation übernommen (Unterpunkt 3.2.4.2.).

Erklärung zur guten wissenschaftlichen Praxis:

Ich erkläre hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten (Amtliche Mitteilung der Universität zu Köln AM 132/2020) der Universität zu Köln gelesen habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen.

Köln, 08.09.2022

Unterschrift: _____

Acknowledgments

First of all, I would like to thank my supervisor of this doctoral thesis, Prof. Dr. med. Florian Klein for providing me the opportunity to work on this project in his laboratory and paving my way into science. Thank you for supporting me at any time during this process. You showed me how to be curious and ambitious for science every day.

I would like to express my deep appreciation to my tutor Dr. Christoph Kreer who taught me the most important steps from basics to high-end methods and reverse. Thank you for inviting me to your project and having trust in me and my work. You gave structure and a clear goal to this project.

Also, many thanks to Dr. Matthias Döring and Prof. Dr. rer nat. Nico Pfeiffer for the great co-operational work and scientific exchange.

I would like to thank the German Center for Infection Research (DZIF) for granting me a one year MD fellowship and research funding.

Thanks to all AG Klein members for being an outstanding, inspirating and team-spirited group. A special thanks to Dr. rer. medic. Daniela Weiland, for her organized, motivating and kind spirit in any tricky situation.

Many thanks to Simon Oehm for being that outstanding friend during our medical studies, residency and the process of this doctoral thesis- especially for proofreading the manuscript and your support in fighting the way through all the bureaucracy.

So much love and gratitude for you, Marc. Thank you for supporting me with encouragement and good spirits at any time during the writing process.

Finally, and most outstanding, I thank my parents and my brothers for giving me encouragement, enthusiasm and unconditional support ever since.

Table of contents

Table of contents

Table of contents

List of abbreviations

A	adenine
AID	activation induced deaminase
AIDS	acquired immunodeficiency syndrome
APC	antigen presenting cell or allophycocyanin
ART	antiretroviral therapy
BCR	B cell receptor
bNAb	broad neutralizing antibody
bp	base pairs
BSA	bovine serum albumin
C	cytosine
CD	cluster of differentiation
cDNA	complementary deoxyribonucleic acid
CDR1	complementary determining regions1
CDR2	complementary determining regions2
CDR3	complementary determining regions3
CD4bs	CD4 binding side
C$_H$	constant heavy chain domain
C$_L$	constant light chain domain
C$_\delta$	constant heavy chain of IgM
C$_\mu$	constant heavy chain of IgD
CSR	class switch recombination
DAPI	4',6-diamidino-2-phenylindole
dATP	deoxyadenosine triphosphate
ddH2O	double-distilled water
DMSO	Dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTPs	deoxynucleotide triphosphate
DTT	dithiothreitol
E. coli	Escherichia coli
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme-linked immunosorbent assay
EMA	European medicines agency
Fab	fragment antigen binding
FACS	fluorescence activated cell sorting
FBS	fetal bovine serum
Fc	crystallizable fragment
FCS-A	forward scatter
FDA	U.S. food and drug administration
FITC	fluorescein isothiocyanate
fw	forward
FWR1	framework region 1
FWR2	framework region 2
FWR3	framework region 3
G	guanin
GC-content	guanin/cytosine content
gDNA	genomic deoxyribonucleic acid
GFP	green fluorescent protein

List of abbreviations

HBSS	Hanks' Balanced Salt Solution
HCV	hepatitis C virus
HIV	human immunodeficiency virus
IgA	immunoglobulin A
IgD	immunoglobulin D
IgE	immunoglobulin E
IgG	immunoglobulin G
IGHD	immunoglobulin heavy chain diversity gene
IGHJ	immunoglobulin heavy chain joining gene
IGHV	immunoglobulin heavy chain variable gene
IgM	immunoglobulin M
IGKV	immunoglobulin kappa light chain variable region
IGLV	immunoglobulin lambda light chain variable region
ILP	integer linear program
IL-4	interleukin-4
IMGT	international ImMunoGeneTics information System
IQR	interquartile range
i$_x$	mismatch closest to the 3'end
LB-medium	lysogeny broth medium
L-part1	leader region 1
L-part2	leader region 2
LR	logistic regression (model)
mAbs	monoclonal antibodies
MACS	magnetic-activated cell separation
MD	medical doctor
Met	methionine
MgCl$_2$	magnesium chloride
MHC	major histocompatibility complex
MM	mismatch
mPCR	multiplex polymerase chain reaction
NGS	next generation sequencing
OFR	open reading frame
oPR	openPrimeR
PAMPs	pathogen-associated molecular patterns
PBMCs	peripheral blood mononuclear cells
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PTP	primer-template-pair
RACE	Rapid Amplification of cDNA Ends
rcf	relative centrifugal force
rev	reverse
RHP	random hexamer primer
RNA	ribonucleic acid
ROI	region of interest
RPMI	Roswell Park Memorial Institute Medium
RT	reverse transcription

List of abbreviations

SARS-CoV-2 severe acute respiratory syndrome coronavirus 2
SCO set cover optimization
SCP set cover problem
SHM somatic hypermutation
SSC side scatter

T thymine
TAE tris-acetate-EDTA
Taq DNA polymerase from Thermophilus aquaticus
TB terrific broth
TCR T cell receptor
Tm melting temperature
TR T cell receptor

UPM universal primer mix
UTR untranslated region

V_H/ VH variable heavy chain domain
V_L/ VL variable light chain domain

X_n Number of 3' hexamer mismatches

y_i amplification status

Δ**G** change in Gibbs free energy of annealing
Δ**G_f** change in free energy folding
Δ**G_s** change in Gibbs self-dimerization

Zusammenfassung

Advancing single cell and multiplex cloning strategies to identify broadly neutralizing antibodies targeting infectious pathogens

von Nathalie Lehnen

Aus dem Institut für Virologie der Universität zu Köln

Direktor: Universitätsprofessor Dr. med. Florian Klein

Die anhaltende Epidemie des humanen Immundefizienz-Virus-1 (HIV-1) sowie die aktuelle *severe acute respiratory syndrome coronavirus 2*[1] (SARS-CoV-2) Pandemie haben deutlich gemacht, dass schnelle und effiziente Methoden zum Nachweis und zur Analyse potenter neutralisierender Antikörper für die globale Prävention und Behandlung bestehender und künftiger Infektionskrankheiten von enormer Bedeutung sind. Obwohl neue Methoden für die Analyse des B-Zell-Repertoires und die Identifizierung und Produktion neutralisierender Antikörper in den letzten Jahren Gegenstand zahlreicher Untersuchungen waren, sind diese Prozesse nach wie vor sehr kosten- und arbeitsintensiv und stellen Wissenschaftler vor große methodische Herausforderungen. Einer der kritischsten Schritte bei der Antikörperisolierung ist die zweistufige multiplex Polymerasekettenreaktion (mPCR), insbesondere bei Vorliegen hoher somatischer Mutationsraten, welche während der Affinitätsreifung der Antikörper auftreten und die Amplifikation stören. Entscheidend für eine erfolgreiche mPCR ist hierbei das Primerdesign, da die Amplifikation von immunologischen *Templates* häufig durch *Primer-Template-Mismatches* oder ungünstige Primereigenschaften und -interaktionen beeinträchtigt wird.

Das Hauptziel dieser Arbeit war daher die Optimierung und Weiterentwicklung von mPCR-basierten Einzelzell- und Multiplex-Klonierungsstrategien zur Identifizierung von neutralisierenden Antikörpern gegen diverse Infektionserreger. Hierbei lag ein besonderer Fokus auf der Entwicklung eines optimalen Primersets.

Um ein standardisiertes und Effizienz-optimiertes Amplifikationsprotokoll zu etablieren, validierten und adaptierten wir bereits bestehende Konzepte. In diesem Rahmen stellten wir das R-basierte Primerdesign- und Auswertungstool „openPrimeR" vor, welches in einem Kooperationsprojekt mit dem Max-Planck-Institut Saarbrücken entwickelte wurde. Wir verwendeten openPrimeR zum Design von optimierten Primersets für die Amplifikation hochmutierter Antikörper. Um die Effizienz der openPrimeR-Designfunktion zuverlässig bewerten und rückkoppelnd im Programm anpassen zu können, entwickelten wir eine standardisierte *Immunoglobulin heavy chain variable region* (IGHV)-Genbibliothek, welche alle 53 humanen IGHV-Gensegmente repräsentiert. Die entsprechenden IGHV-Gensegmente wurden durch 5'-*rapid amplification of cDNA-ends polymerase chain reaction* (5′ RACE-PCR) aus *gepoolten*, naiven B-Zellen von 16 gesunden Spendern isoliert.
Anschließend erfolgten multiple PCR-Amplifikationsexperimente, sowohl auf der neu-etablierten IGHV-Genbibliothek als auch auf humanen naiven und antigen-spezifischen

---

[1] Nicht-übersetzbare englische Fachbegriffe werden im Folgenden *kursiv* gedruckt.

Einzel-B-Zellen. Hierbei erfolgte der direkte Vergleich von mehreren in openPrimeR entworfenen mit bereits etablierten Primersets. Das oPR(5)-IGHV Primerset zeigte im Vergleich zu etablierten Primersets eine überlegene Amplifikationsrate der IGHV-Genbibliothek und ist, insbesondere durch seine exklusive Bindung an die wenig mutierte Leader-Region des V-Gens, ein vielversprechender Kandidat für die Amplifikation hoch mutierter Antikörper. Darüber hinaus konnten wir einen großen *Taq*-PCR-basierten Datensatz generieren, der den Amplifikationsstatus von 2.820 PCRs (940 Triplikaten) enthält. Dieser wurde mit insgesamt 20 verschiedenen Primern auf allen Genen der IGHV-Genbibliothek erstellt und anschließend hinsichtlich physikalisch-chemischer Eigenschaften der jeweiligen *Primer-Template*-Paare und deren Einfluss auf den Amplifikationsstatus ausgewertet. Hierbei konnten die freie Energie des *Annealings* (ΔG), die absolute Anzahl der *Mismatches* innerhalb des 3'-Hexamers sowie die *Mismatch*-Position im Verhältnis zum 3'-Terminus als entscheidende Faktoren für den Amplifikationsstatus identifiziert werden.

Die Ergebnisse dieser Arbeit trugen wesentlich zur Entwicklung eines hocheffektiven Protokolls zur Identifizierung breit neutralisierender Antikörper bei, welches sich bereits bei der Isolation klinisch relevanter Antikörper-Kandidaten bewährt hat.

1 ABSTRACT

Multiplex polymerase chain reaction (mPCR) techniques have a versatile use in medicine, as they are implemented for the identification of pathogens and disease biomarkers and are often applied in genotyping. They additionally present a valuable tool for the amplification of lymphocyte receptors (such as antibodies). Fast and efficient methods for the detection and analysis of potent neutralizing antibodies are of tremendous value and are urgently needed for the global prevention and treatment of existing and upcoming infectious diseases, which has been demonstrated vividly in recent years by the epidemic of the human immunodeficiency virus-1 (HIV-1) and the current severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic. Although new methods for the analysis of the B cell repertoire and identification of neutralizing antibodies have been subject to many studies in recent years, these processes remain challenging and are highly cost- and labor-intensive. One of the most critical steps in antibody isolation is the two-step mPCR, especially in the presence of high somatic mutation rates, which occur during the antibodies' affinity maturation and disrupt amplification. Primer design is crucial for a successful mPCR, as amplification of immunological templates is often compromised by primer-template mismatches or unfavorable primer properties and interactions.

Thus, the main goal of this thesis was to optimize and further develop mPCR-based single-cell and multiplex cloning strategies for the identification of (broadly) neutralizing antibodies targeting infectious pathogens.

We intensively validated and adapted existing BCR amplification concepts to establish a standardized laboratory protocol with cost- and labor intensity optimized conditions. In this context, we introduced the R-based primer design and evaluation tool openPrimeR, which was developed in a collaborative project with the Max Planck Institute Saarbrücken. We used the openPrimeR's design function to generate promising primer sets capable of amplifying highly mutated antibodies. To enable the testing of the openPrimeR design function in an unbiased experiment setting and refeed gained information back to the program, we developed a standardized immunoglobulin heavy chain variable (IGHV) gene library which represents all 53 human functional IGHV gene segments. The corresponding IGHV gene segments were isolated by 5′-rapid amplification of cDNA-ends polymerase chain reaction (5′ RACE-PCR) from pooled naive B cells of 16 healthy donors.
We extensively tested several sets of de novo primers generated by the openPrimeR tool in direct comparison to well-established primer sets on both, antigen-experienced single human B cells and the IGHV gene library. The finally presented optimized oPR(5)-IGHV primer set showed favorable primer set properties, such as exclusive primer binding to the leader region of the V gene to ensure broad template coverage. The oPR(5)-IGHV demonstrated superior performance in amplifying the (germline) IGHV gene library in comparison to established primer sets and is a promising candidate for amplifying highly mutated antibodies. In addition, we were able to generate a large Taq-PCR-based dataset containing the amplification status of 2,820 PCRs (940 triplicates) using a total of 20 different primers on all genes of the IGHV gene library. The data set was analyzed with respect to physicochemical properties of the respective primer-template pairs (PTPs) and their influence on the amplification status. The analysis of our data set identified the free energy of the annealing (ΔG), the absolute number of mismatches present in the 3' hexamer and the mismatch position in relation to the 3' terminus as determining factors for the amplification status.

In summary, the results of this thesis have contributed significantly to the development of a highly effective protocol for the isolation of broadly neutralizing antibodies, which has already been proven successful in the detection of clinically relevant therapeutic antibody candidates.

2 INTRODUCTION

2.1 The human immune system and the advent of broad neutralizing antibodies

2.1.1 Human immunity and the value of immunological science

The human body is continuously surrounded by and exposed to an infinite number of foreign organisms by inhaling, digesting or experiencing direct skin and mucosal contact. Apparently, not every encounter leads to an infection or allergic reaction, which is highly determined by two factors: the organism's specific pathogenicity and the humans ability to activate its protection mechanisms.[1] The tremendous importance of a reliable defense system shows itself most impressively when mechanisms fail: a misguided or deficient immune response to pathogen encounter can cause severe, even deadly, infections and tumors, whereas an overactivity can lead to autoimmune diseases and allergic reactions.[1]

Until the beginning of the 20th century infectious diseases were the major causatives to death worldwide.[2] Even today, infectious diseases are one of the key contributors to human maladies and death, especially in areas with restricted access to a reliable healthcare system.[3] Among these, human immunodeficiency virus-1 (HIV-1) infections and acquired immunodeficiency syndrome (AIDS)-related deaths continue to be a global problem despite tremendous efforts to prevent transmission.[4] Just lately, the ongoing severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic has revealed the human fragility when it comes to highly contagious, rapidly spreading and potentially fatal pathogens.[5]

The origin of today's science of immunology is commonly attributed with Edward Jenner, who discovered in 1796 that cowpox induce protection against human smallpox, a disease which up until then often ended fatally. Jenner called this new procedure vaccination and described the inoculation of a person with an attenuated or weakened pathogen particle to induce protection and to ideally develop an immunological memory.[6] It is estimated that around 300 million people died of smallpox in the first three quarters of the 20th century, whereas due to a mass vaccination-based eradication program, not a single one has died from it since 1978. Not without reason, the success of vaccination is often considered the most important medical break-through in the past centuries.[7,8]

Over the last centuries, with the beginning of genetic engineering and here within an increasing insight into the human immunity, vaccination strategies continuously evolved until today, where we have a broad access to vaccinations against many infectious diseases. Nevertheless, some

organisms, especially viruses like HIV-1, pose a tricky challenge when it comes to preventing vaccinations or successful treatment, due to viral-specific escape mechanisms.[9]

Today, once more, the immense importance of valuable scientific research in the field of immunology for a better understanding of the human immune system, characterization of pathogens and the development of therapeutical strategies in the field of prevention (e.g., vaccines) and treatments has been demonstrated vividly by the SARS-CoV-2 pandemic and other worldwide heath crises.

2.1.2 The components of the immune system

To withstand the continuous exposure to pathogens, the human body possesses a communicating network of lymphoid organs, immune cells, humoral factors and cytokines, called the immune system. It comprises two interconnected main parts, the innate and the adaptive (also referred to as acquired) immunity to protect the body from pathogens and infections.[1]

The innate immunity serves as a first line of defense. It reacts immediately with several distinct mechanisms such as barrier function of the body's epithelia, recruitment and activation of neutrophils at the side of infection and tissue macrophages mediating cellular defense by phagocytosis and the release of chemokines.[10] Another part of the innate immunity is the complement system, which is a protein-based network that enhances (complements) the ability of antibodies and phagocytotic cells to clear the body from damaged cells or pathogens by promoting inflammation and cell lysis. Innate immunity highly relies on the recognition of conserved pathogen-associated molecular patterns (PAMPs), which are present on microbes but not on host cells.[11] Phagocytes bearing pattern-recognition receptors recognize broad patterns on microbes and present the processed pathogens to antigen-specific T cells.[1]

The innate immunity is crucial for every day's immediate defense against foreign organism, but it is evaded by certain pathogens and lacks the ability to provide a specific protection from reinfection. Therefore, once a pathogen is recognized and phagocytosed, a signaling-pathway activation of the adaptive immune system is induced.[1,12] The adaptive immune system on the other hand is characterized by a remarkable antigen specificity and memory due to the properties of its comprising cellular components, the B and T cells. A key feature of these lymphocytes is their expression of a unique antigen receptor with defined specificity. Thereby it provides the ability of a vast molecular recognition of antigens.[7]

Most of the mechanisms of the adaptive immune system take place once a B cell recognizes a foreign antigen. B cells can proliferate and differentiate into plasma cells without T cell

support when strong B cell receptor (BCR) stimulation is present. These T cell-independent immune responses mostly generate low affinity antibodies that typically do not give rise to B cell memory. The more common mechanism is the T cell-dependent humoral immune response, which takes place in a complex interaction between B and T cells. B cells internalize a recognized antigen and subsequently present peptide pieces on their surface via major histocompatibility complex (MHC) class II molecules. CD4+ T helper cells recognizing the same antigen as a B cell, usually through a different epitope, activate the cognate B cell by the release of various cytokines such as IL-4 and thereby promote B cell division. The B cell derives the ability to differentiate into a (long- or short-living) antibody producing plasma cell or a memory B cell.[13–16] In the germinal center, B cells that provide high-affinity antibodies and thereby capture a greater amount of antigen, are favored by a positive selection signal of the cognate T follicular helper cell.[17]
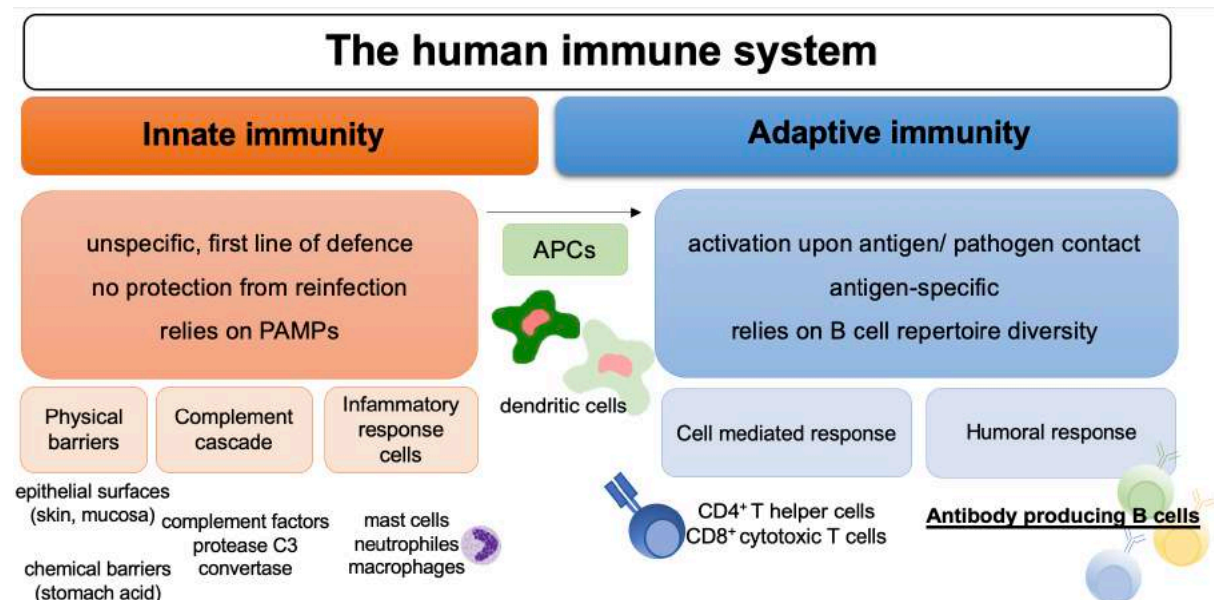
Figure 1: Components of the human immune system
This schematic figure gives an overview over the components of the human immune system and their contributing functions in the defense against pathogens as described above. PAMPs= pathogen-associated molecular patterns, APCs= antigen-presenting cells

2.1.3 Antibodies and their structure

Playing a key role in the adaptive immune system, antibodies (also referred to as immunoglobulins) are large Y-shaped proteins which are either expressed on a B cell to form the respective B cell receptor or are secreted into the plasma by plasma cells.
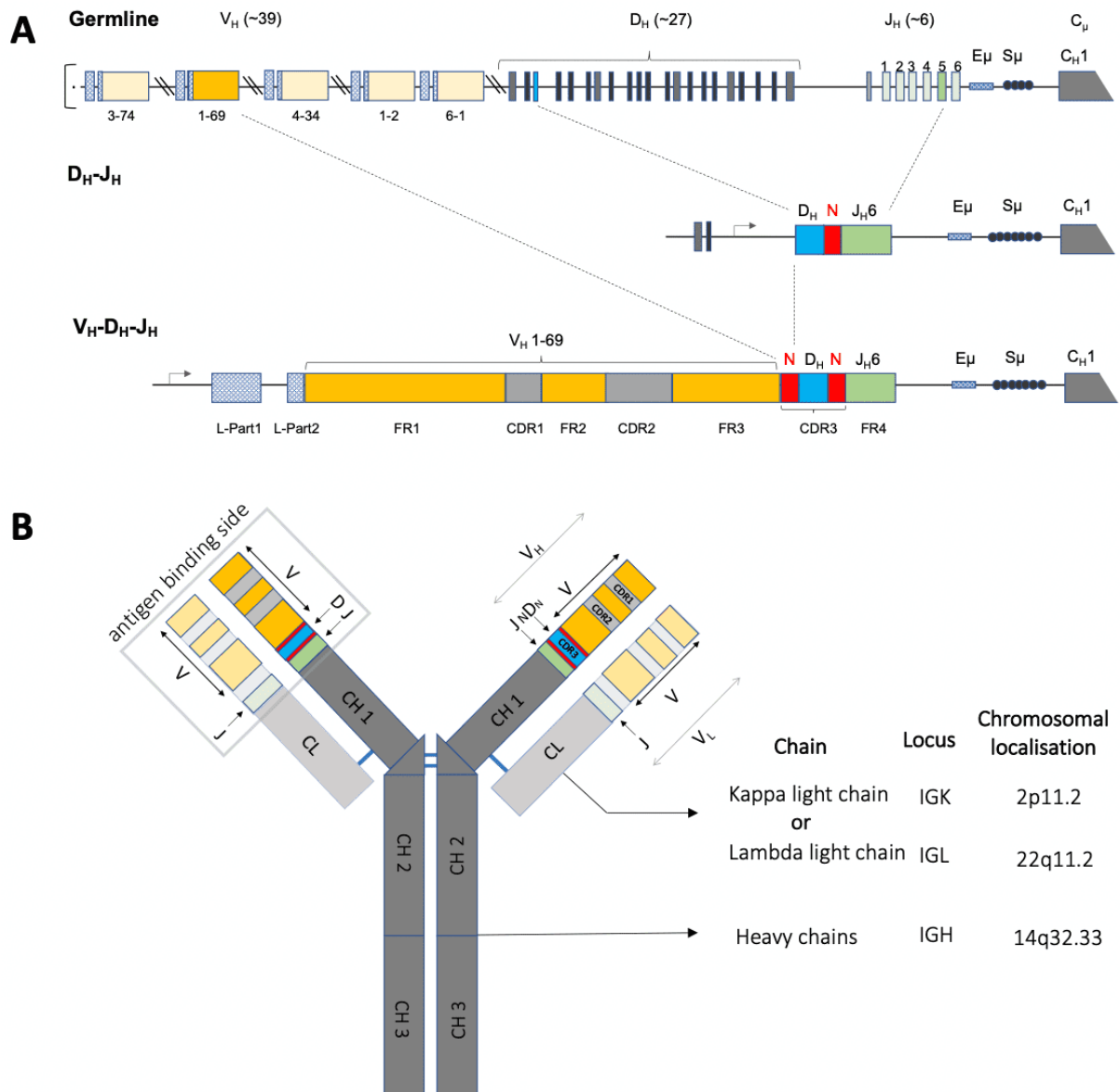
Figure 2: Schematic VDJ-rearrangement and antibody structure
(A) Key steps in antibody diversification via V-D-J-recombination: The primary antibody heavy chain repertoire is generated by variable (V), diversity (D) and joining (J) gene segment rearrangement on DNA level. Additionally, a random nontemplated addition of N-nucleotides (N) takes place. Numbers in brackets refer to the estimated number of human germline V_H, D_H and J_H segments. Eμ indicates the IgM intronic enhancer, Sμ indicates tandem repeats necessary for class-switch. Constant region (C_H) is only shown partially (by C_H1) in this figure.
(B) Schematic antibody structure on the example of an IgG1: Each antibody molecule consists of two identically heavy and two identically light chains (kappa or lambda). The chains are connected by interchain disulfide bridges (pictured by thin blue bars). This modified graphic is based on LeFranc et al. (2020)[18] and Georgiou et al. (2013)[19] and used with friendly permission of Nature Springer.

In the early B cell development in the bone marrow, antibodies are formed by somatic recombination of a large number of different immunoglobulin gene segments (**Figure 2A**).[20] Each antibody consists of two identical heavy chains and two identical, either kappa or lambda, lights chains, which are linked by disulfide bonds.

The light chain is formed by two domains, (a variable V_L and a constant domain C_L) whereas the heavy chain includes four or five domains (one V_H and three or four C_{H1-3} domains). Each variable domain (V_H and V_L) comprises four framework regions which form the structure providing beta sheet and support the three complementary determining regions. The complementary determining regions (CDR1, CDR2 and CDR3) present the hypervariable domains of the antibody.[18]

Most of the diversity of the naive human antibody repertoire is concentrated in the VDJ gene segment ligation of the heavy chain, also referred to as CDR3. The heavy chain CDR3 is the most diverse component of an antibody in terms of length and sequence divergence. It has therefore particular influence on the antibodies specificity.[21–23] Finally, when the variable domains (V_H and V_L) are paired together, their CDR loops form a conjoint hypervariable site at the tip of each antibody arm: the antigen-binding site (**Figure 2B**).[18]

Antibodies neutralize and clear pathogens by several processes: Neutralization takes place by coating the surface of a potentially pathogen (toxin, virus, or bacteria) to block its functionality. A further process is the so-called opsonization, where the variable (Fab) region of an antibody binds to the cell membrane of the pathogen. Monocytes and neutrophils recognize the constant (Fc) region of the antibody and subsequently aim to destruct the pathogen. Additionally, the activated Fc receptors on the phagocytes promote antigen presentation on MHC molecules and thereby support the T-cell response.[24] Neutralizing antibodies not only play a key role in antiviral immunity, they are also the correlate of many vaccines and therapeutic approaches.[25–28]

2.1.4 Repertoire division: IGHV, IGKV and IGLV families, genes, and alleles

According to the standardized classification and nomenclature of the immunoglobulins and T cell receptors by IMGT®, the human V gene segments of the heavy- and κ-light chains can be grouped into seven families respectively, whereas the V gene segments of the λ-light chain divide into eight families. The family members share a very high DNA sequence homology (at least 80%). This high amount of conserved sequence in each family implies an evolutionary pressure to conserve a certain V gene structure.[29,30] Each gene comprises several polymorphic variants which are characterized by mutations at nucleotide level and are called alleles. Alleles are identified by their core sequence of the V-, D- and J-regions and are titled in comparison to their reference sequence (*01).[31]

2.1.5   The diversity of the B cell repertoire

Each individual's potential antigen receptor repertoire is estimated to include approximately $5x10^{13}$ different antibodies (immunoglobulins).[32] The main contributors to this diversity are the variable domains of the heavy and light chains. In humans, their encoding genes are located in the IGH locus on chromosome 14 (14q32.33), the IGK locus an chromosome 2 (2p11.2), and the IGL locus on chromosome 22 (22q11.2) (see **Figure 2B).**[33] The IGH locus shows a dimension of a one megabyte locus and is among the most complex and variable regions of the human genome.[34]

The above described VDJ recombination plays a major role in the repertoire's diversification. As the rearrangement is imprecise, exonuclease induced trimming at the ends of the V, D, and J genes occurs as well as random addition of nucleotides at their junctions before the gene ligation (**Figure 2A**). The DNA recombination is irreversible and is unique to each cell.[35]
An additional diversity is created by the pairing of the variable domains of heavy and light chain to finally form the antigen recognition and binding site.[33] Later during the B cell development when encountering antigen, the rearranged V region undergoes a high rate of point mutations. This process is called somatic hypermutation (SHM).[35] SHM is catalyzed by the enzyme activation induced deaminase (AID) which enzymatically converts cytosine to uracil.[36] SHM mainly involves nucleotide base changes in the CDR regions. It optimizes the process of affinity maturation and therefore the antigen-epitope-binding by antibodies.[37]
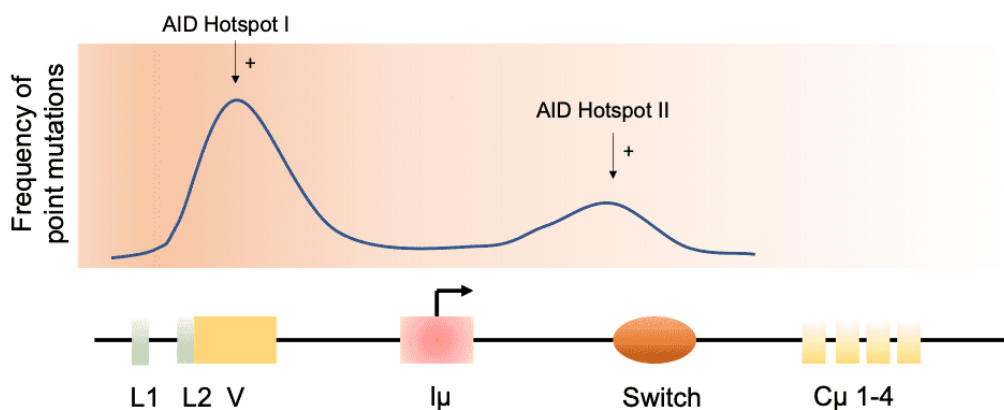
Figure 3: AID Hotspots
Driven by the activation-induced cytosine deaminase (AID), somatic hypermutation is characterized by base changes that occur throughout the variable (V) regions at a rate of ~1 $\times10^{-3}$ mutations per base pair per generation. AID activity hotspots concentrate in the V region (V) and downstream of the promoter (Iμ). Its activity is very limited in the constant region, where it promotes class switch. This modified graphic is based on Martin et al. (2002)[38], used with kind permission of Springer Nature.

The immunoglobulin's V gene is preceded by an exon encoding a leader peptide (L). The leader directs the antibody into the cell's secretory pathway. Its sequence is removed after

translation. The leader region exhibits a tenfold lower rate of AID induced mutations in comparison with the 5' region of the V gene (**Figure 3**).[39]

The potency of the natural antibody repertoire is further increased by class switching of the heavy chain. The heavy chain constant region (C_H) determines the antibodies isotype (antibody class) and is crucial for the antibody's diverse effector functions. The five main isotypes are IgM (µ), IgD (δ), IgG1-4 (γ), IgE (Ɛ) and IgA1-2 (α). The different constant regions in the heavy-chain locus are encoded in separate genes, which are located in juxtaposition to one another more downstream of the V region.[40] Initially, in naive B cells, only C_μ and C_δ are used. They are expressed along with the associated rearranged V region segments and produce a transmembrane IgM and IgD on the cell's surface.[41] In response to antigen stimulation the activated mature B cell can switch to the expression of a different downstream constant region gene. This occurs by a type of somatic rearrangement and is called class switch recombination (CSR) or isotype switch. Light chain C regions, other than the described heavy chain C regions, do not undergo class switch as they do not exhibit specific effector functions.[41]

Since the random recombination process in early B cell development, it is inevitable that autoreactive B cell receptors (antibodies that recognize self-antigens) emerge.[42] In the bone marrow and peripheral circulation, autoreactive B cells undergo a stringent selection by deletion or receptor editing to ensure that in (healthy) humans the occurrence of autoreactive antibodies is prevented.[43,44]

2.1.6 Broadly neutralizing antibodies

Neutralizing antibodies play a major role in the human antiviral immunity and most of today's vaccines rely on the principle of inducing neutralizing antibodies that mediate protection.[25–27] Among them, broadly neutralizing antibodies (bNAbs) are antibodies which are characterized by a broader, cross-clade and more efficient neutralizing activity in comparison to common neutralizing antibodies. In HIV-1, they usually show high levels of somatic mutations. bNAbs exert their neutralizing properties by binding to conserved epitopes of the respective pathogen, e.g., an envelope protein on a virus. Thereby they are able to bind multiple viral strains, even if these exhibited viral escape mechanisms.[45] Some HIV-1-specific bNAbs for example can bind to more than 95% of HIV-1-strains worldwide even at low concentration.[26] Other than the typical antibody responses, it can take years for HIV-1 bNAbs to develop within an individual. Clinical trials revealed, that 10-30% of HIV-1-infected persons develop serological activity to neutralize different viral isolates after two to four years of infection.[46] Only around 1% of these individuals, the so-called elite neutralizer, develop antibodies with exceptional cross-clade activity.[47]

The isolation of bNAbs from these people's blood was a huge obstacle until the advent of single-cell-based antibody cloning techniques. But still today, fast and efficient isolation methods for the identification and analysis of potent neutralizing antibodies are urgently needed to improve the global prevention and treatment of existing and upcoming infectious diseases.[45,48,49] The market of monoclonal antibodies (mAbs) represents of the fastest developing fields in modern biotechnology.[50] Over the last years, more than 90 new mAbs have been approved by the U.S. Food and Drug Administration (FDA) and European Medicines Agency (EMA) to provide treatment against a variety of diseases such as cancer, autoimmune or hematological disorders. Since the beginning of the COVID-19 pandemic, five anti-SARS-CoV-2 antibody products (Casirivimab and Imdevimab, Sotrovimab, Bamlanivimab and Etesevimab) were authorized for emergency use by the FDA or approved by the EMA.[51,52] *(Stand 06/2022)* Also, recently identified HIV-1-directed bNAbs demonstrated promising features and give hope for future HIV-1 vaccination and therapy 19pproaches.[53,45] Recent clinical trials already demonstrated suppression of viremia and delay of viral rebound after pausing antiretroviral therapy (ART) due to bNAb infusions.[54–56] In animal models, bNAbs showed high effectiveness in preventing HIV-1 infections.[57,58] Lately ongoing trials using the CD4bs-targeting bNAb VRC01 for passive immunization of humans are promising outlooks for a potent vaccination against HIV-1 in the future (ClinicalTrials.gov: NCT02716675, NCT02568215).

2.2 The value and challenges of polymerase chain reaction in the field of immunology

2.2.1 The basic principles of polymerase chain reaction

In 1985, the introduction of the polymerase chain reaction (PCR) by Kary B. Mullis revolutionized the way deoxyribonucleic acid (DNA) could be studied in detail due to a rapid production of billions of copies of a specific DNA fragment.[59] Today, multiprimer or multiplex PCR (mPCR) techniques allow the amplification of multiple target regions simultaneously. In the field of medicine they have a versatile use, as they are implemented for the identification of pathogens and disease biomarkers or are applied in genotyping. mPCR additionally presents a valuable tool for the amplification of lymphocyte receptors and antibodies. Furthermore in virology, PCR serves as a valid method for diagnosing viral infections, testing viral drug resistances and identify viral escape variants as well as their underlying processes.[60–62]

The basic components of a PCR include the cDNA template, specific primers, a reaction buffer, free nucleotides (dNTPs), the catalyzing enzyme named polymerase, salt and water (which is free from nuclease and DNA).

It is possible to copy a DNA fragment to nearly any extent within a three-stepped amplification cycle which consists of i) denaturation, ii) annealing and iii) elongation.[63]
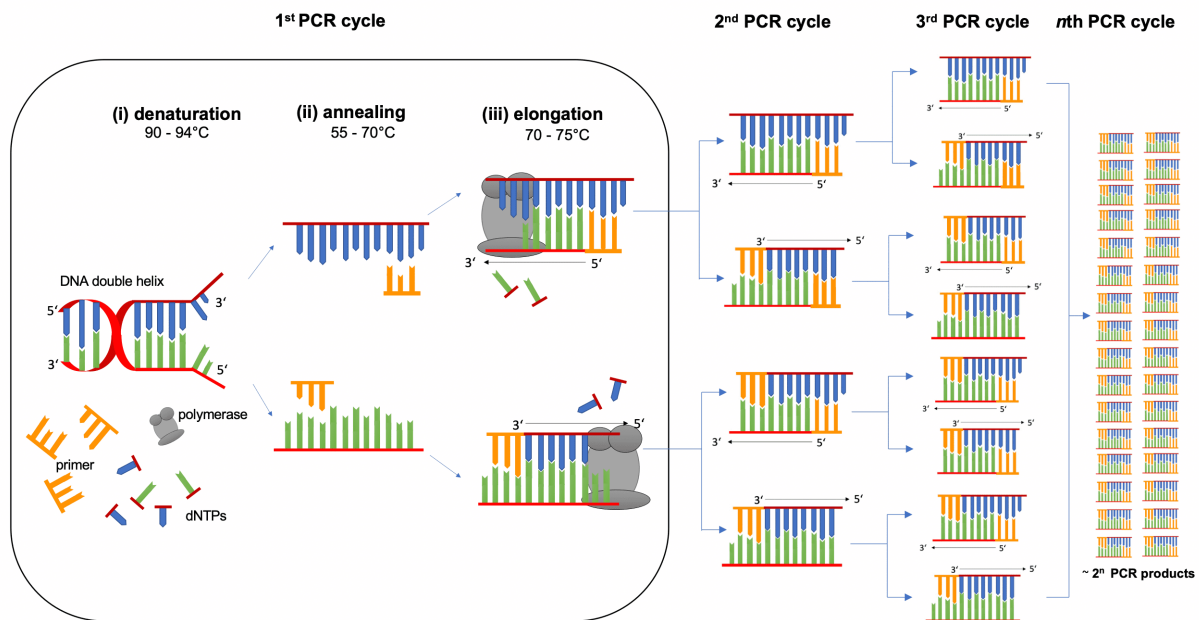
Figure 4: Schematic polymerase chain reaction steps
(modified illustration used with courtesy of Creative Commons Attribution- Share Alike 4.0. International License, https://creativecommons.org/licenses/by-sa/4.0/)

The i) denaturation takes place at a melting temperature around 90 to 94°C (depending on the templates guanin-cytosine (GC)-content) at which hydrogen bonds between the base pairs adenine (A) and thymine (T) and between cytosine (C) and guanine (G) break, resulting in two single stranded DNA molecules out of the double stranded DNA helix. The single stranded DNA serves as template for the following ii.) annealing. To proceed, the temperature is lowered between 55 and 70°C, depending on primer specific annealing temperature, to allow primers to bind (anneal) and form the primer-template heteroduplex.[63]

Primers are short single stranded DNA fragments with a defined sequence complementary to the target DNA and serve as starting points for the DNA extension. Once primer annealing took place, the temperature is elevated up to 70-75°C for the DNA polymerase to start synthesis. From 5' to 3' end (and vice versa) dNTPs are added complementary to the template strand. This process is called iii.) elongation. The three-stepped PCR cycle is repeated usually around 25 times on bulk complementary cDNA and up to 50 times on single cell cDNA, resulting in an exponential template amplification. A last temperature change to 72°C ends the reaction with the final incubation step **(see Figure 4)**.[63]

2.2.2 Primer design and its challenges

Primer design is crucial for a successful PCR. While it seems simple in theory, the reality of laboratory work teaches otherwise.

The primer-template-complementarity determines efficient primer annealing and is captured by change in Gibbs free energy of annealing (ΔG).[64] Mismatches, non-complementary base pairs, are unfavorable and affect ΔG. Especially mismatches in the six terminal (3') base pairs disrupt the polymerase binding and therefor lower the amplification status.[65,66] The closer to the 3' terminus a mismatch occurs, the higher its negative impact on amplification becomes.[67] Additionally, the type of mismatch effects the elongation efficacy: for example, a G (primer)- T (template) mismatch is less harmful for the amplification than a G (primer)- A (template) mismatch.[68] Repeats (di-nucleotides occurring repeatedly e.g. ATATAT) and runs (long runs of a single base, e.g. GGGG) should be strongly avoided. It is advisable to use guanines and cytosines in the last six nucleotides of a primer. The so-called GC clamp stabilizes the 3' region.[69] The GC content (the percentage of guanin und cytosine bases) should be between 40 and 60%. It is generally advisable to choose a primer length between 18 to 22 base pairs to ensure specificity but still allow easy binding.[70]

In multiplex or multiprimer PCRs even more aspects must be considered when designing a primer set. Primers do not bind exclusively to complementary cDNA but potentially bind to each other. These formations are called primer-dimer and might appear as by-products of a PCR when primer show a high degree of complementarity to each other. Primer-dimers can lower the amplification efficacy due to shortness in PCR reagents. Additionally, primers used in the same reaction should be carefully selected to have similar annealing temperatures. If the temperature range is too vast, some primers have difficulties to stay annealed while others have a higher likelihood of mismatch annealing. In both cases, amplification is affected, and an amplification bias might occur.[71] With an increasing template variation to be covered, the number of primers needed rises. Meanwhile the design of a multiprimer set becomes more and more complex.

2.2.3 The value and challenge of PCR on highly diverse templates, in particular the B cell repertoire

In the field of immunology, PCR based methods are of particular importance for many diagnostic and therapeutic pipelines but faces tremendous challenges due to the wide variety of templates and mutational burden that interfere with both, amplification and further sequence analysis. For instance, HIV-1and the hepatitis C virus (HCV) show a high extent of sequence diversity due to distinct error-prone replication processes which allow them to develop escape variants from immune response or treatment.[72] Sequence analysis of emerging viral escape variants is critical for detecting developed resistances to antiviral drugs and tailoring potentially successful therapy.[73,74]

Furthermore, B cell receptors (BCRs) and antibodies represent an intensively diverse template repertoire. They will be the main target of PCR analysis in this thesis. An in depth characterization of these receptors and their repertoire in the human body leads to a better understanding of immunological responses to pathogens and vaccination.[75–77] Moreover, it can provide a better understanding in autoimmune diseases and B cell malignancies. Furthermore, the deeper analysis of the B cell repertoire can lead to the detection of new mAbs.[78] Clinical application of bNAbs in infectious diseases are of high interest due to their features of cross-clade neutralizing, Fc-mediated effector functions, preferable pharmacokinetic profiles (e.g. in comparison with anti-viral drugs administered in HIV-1 treatment) and being potential candidates for new vaccines (see above).[79,80]

However, the amplification of the BCR presents a particular challenge for the design of multiplex PCR primers because the primers need to bind to the various segments of the immunoglobulin loci encoding for the heavy and either kappa or lambda light chain of each antibody.[81] A further challenge for PCR amplification is presented by a high degree of template variation due to somatic hypermutation, deletions and insertions which take place during B cell development.[75,45] Especially HIV-1 specific broadly neutralizing antibodies often show up to 30% sequence divergence compared with their germline precursor which leads to an unknown amount of template diversity und primer-template-missmatching.[42,82] The mutations leading to this divergence mostly occur in the CDRs and are critical for the neutralizing activity of such antibodies.[75,45] These areas should be strictly excluded from selection as primer binding side.

Further complicating matters, the number of primers in a multiplex PCR reaction should be restricted to a certain limit to avoid physicochemical problems such as primer self-dimerization, large ranges in primer melting-temperatures or unspecific and random amplification. Therefore, the theoretical idea of using one individual primer for each expected target would

lead to an unacceptable high amount of primers in the multiplex PCR on immunological templates.[60] The so-called "set cover problem" describes the trouble of identifying a minimal primer set with full template coverage. Naturally, the complexity of this problem increases rapidly with a rising template variation.[83]

For the immunological template of the heavy and light chains the described problems mainly account for the 5' primers (also called forward primers) as these need to bind to multiple possible sequences of the upstream leader region or the possibly highly mutated beginning of the variable region. It has been demonstrated that forward primers binding to the V gene's signal peptide encoding leader (L) region, favor the amplification of highly mutated variable V genes.[75,84]

The design of the 3' primer or so-called reverse primer on the other hand is comparatively less complex because its binding side can be set in the constant region of the respective chain. The constant region does not undergo somatic hypermutation and thus represents a constant binding region with a known stable sequence. Nonetheless, a different reverse primer needs to be designed not only for the different chains (heavy, kappa light and lambda light) but also for the different Ig isotypes which exhibit different constant region sequences (e.g., IgM, IgG).
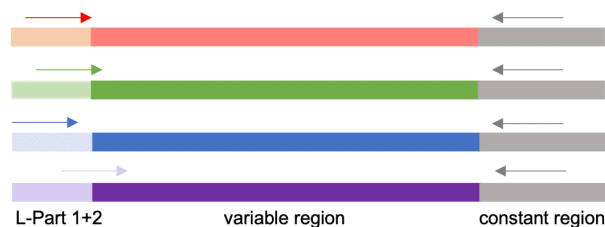
Figure 5: Schematic overview of primer binding sides for the amplification of the immunoglobulin variable region
Multicolored arrows indicating multiple different forward primers (each specific for another V-gene). Grey arrows indicating the one universal reverse primer binding in the constant (non-variable) region.

But amplification rate alone is not the only request for PCRs on immunological templates. An unbiased PCR amplification is of high importance in the analysis of the B cell repertoire when it comes to evaluating relative gene frequencies among the human population, in subgroups or in maladies and infections. Recent studies implicated that the DNA polymerase is one of the key driving factors of PCR bias.[85] Additionally, in multiplex PCRs, strongly differing efficacies of included primers can lead to a bias by disproportionate amplification of certain genes, which can disrupt studies on the B cell repertoire, e.g. the frequency of gene expression.[82]

2.2.4   Published primer sets to amplify V_H gene segments

In the past, the development of primer sets used for the amplification of human V_H gene segments has been studied intensively. Indeed, several primer sets have been shown to be successful candidates for single B cell cloning approaches to gain insights into the B cell repertoire and to isolate monoclonal antibodies targeting infectious diseases.[75,81,86-87]

Besides, Scheid et al. presented a primer set which shows promising features to especially identify highly mutated antibodies by setting the 5' primer binding side further upstream to the beginning of the respective leader region. Hereby, an increased discovery of immunoglobulins could be demonstrated.[75] However, there is the distinct disadvantage that the primer set consists of four individual primer sets, each targeting only genes from one or two different IGHV families, which requires multiple PCRs to be performed and results in high labor and cost requirements. Tiller et al. developed an efficient strategy that combines immunoglobulin gene repertoire analysis and immunoglobulin reactivity profiling at single cell level, using a primer set for semi-nested PCR with 2nd PCR primers containing direct restriction sides.[81] However, 2nd PCR primers were designed to bind in the FWR1, possibly causing the reversion of mutations in this area, which can be critical for the antibody's activity. Additionally, Ippolito et al. presented a primer set (priorly published by Lim et al.[88]) to analyze differences in the repertoire's V gene usage of humanized mice (using NOD-scid-IL2Rγ^{null} mice) and peripheral human blood B cells.[86] The primers were applied in a special touch-up PCR protocol[86] and demonstrated coverage near to 100% of all functional and putatively functional germline V genes in VBASE2, but no studies on primer performance with highly mutated V genes have yet been presented.[88] Finally, Tan et al. further investigated clonal characteristics of the paired infiltrating and circulating B lymphocyte repertoire in patients with primary biliary cholangitis using a minimal primer set without self-hybridizing but with a broad coverage of a heterogeneous set of V gene and C gene family sequences.[89] The region of interest in this study was the IGH-CDR3 region and primer binding was set to framework region 3 (FWR3). Therefore, a complete V gene amplification and analysis cannot be provided.
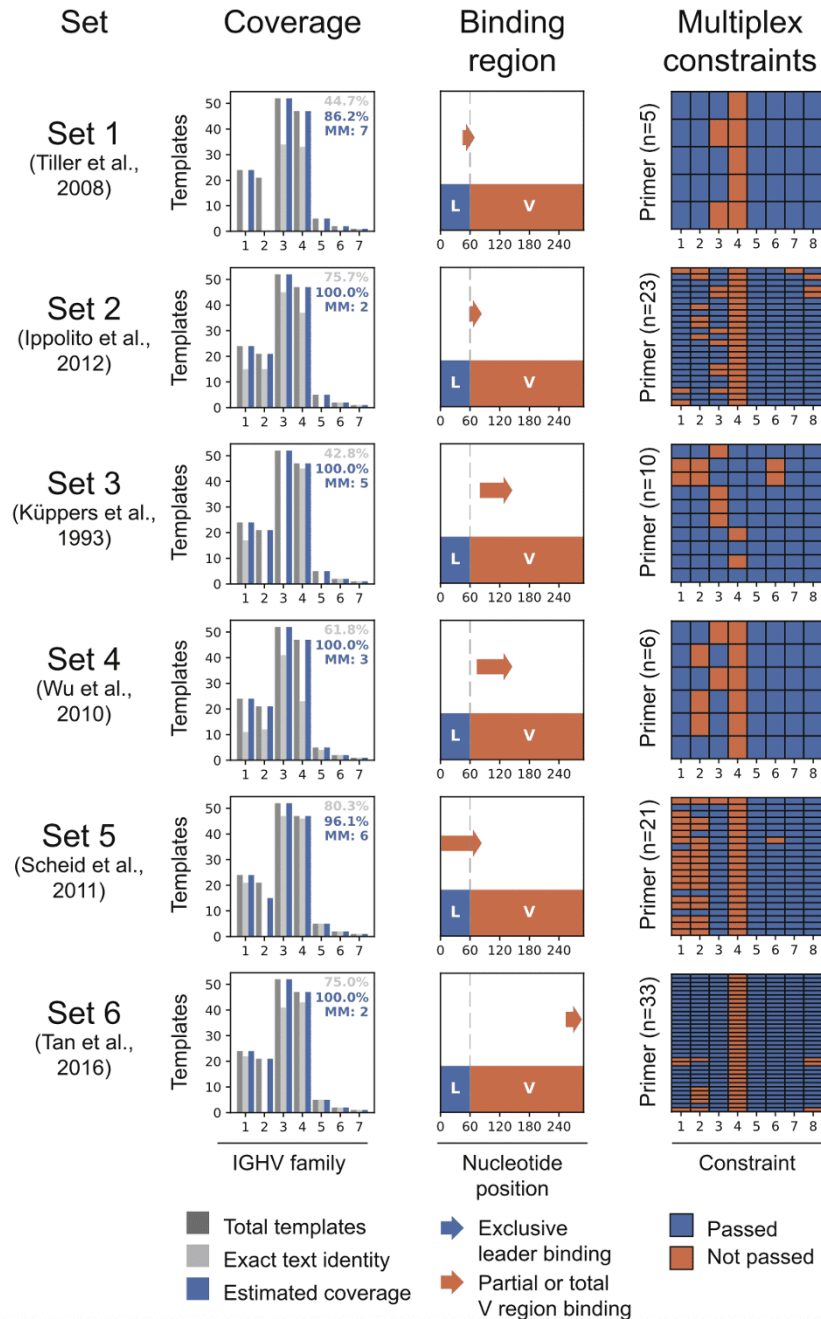
Figure 6: Evaluation of published primer sets using the primer design and evaluation tool openPrimeR

Representative primer sets that have been previously used to amplify VH gene segments (Set 1–6) were analyzed with openPrimeR. 152 VH gene segments (belonging to the 7 IGHV families) are indicated in dark grey bars. The template coverage of the respective primer set was determined by exact text identity (light grey). MM indicates the maximum number of 7 allowed mismatches, that is necessary to reach the depicted estimated coverage (blue). 5' and 3' outer most binding regions are shown for all primers as a composite arrow. Each primer was tested for eight multiplex constraints (1) Cross-dimers, (2) GC clamp, (3) GC ratio, (4) melting temperature deviation, (5) maximum number of nucleotide repeats, (6) maximum number of nucleotide runs, (7) coverage of at least 1 template, (8) and self-dimers). Constraints that were passed are colored in blue, those that were not passed are colored in orange. This figure and figure legend were published in 2020 in the Journal of Immunological Methods by Kreer et al.[90] (co-author N. Lehnen) and is used in approval with all co-authors and with courtesy of Creative Commons Attribution- Share Alike 4.0. International License, *https://creativecommons.org/licenses/by-sa/4.0/*

25

The overview presented in **Figure 6** by Kreer et al.[90] shows a detailed evaluation of the established primer sets described above. It reveals the limitations of the respective sets such as incomplete variable gene coverage, partial or total V gene binding or the labor-intensive requirement for multiple reactions.

In conclusion, a handful of specified primer sets for the amplification of human V_H genes are available, but none of them fulfilled all criteria for a reliable primer set, thus further detailed testing and primer development is required.

2.2.5 Published primer design tools for multiplex PCRs, their limitations, and the generation of openPrimeR

The simplest way to design a primer is to manually check the sequence of interest for possible primer binding sides and generate complementary single-stranded DNA fragments to serve as primers. Multi-sequence alignment can help to detect recurring or at least similar sequence segments for primer binding in multiple templates. However, as template amplification and multiplex PCRs become increasingly complex nowadays, multiplex primer design can hardly be done manually. Therefore, primer design tools are of great scientific interest. In the past, several primer design tools have been published to overcome the difficulties of multiplex primer design described in section 2.2.2 and 2.2.3. However, until now, none of these tools fulfilled our criteria for a reliable primer design tool. For the amplification of diverse immunological templates such as immunoglobulins or virial strains, primer tools should necessarily include a program to filter primers for preferrable properties, such as the GC content, GC clamp or the melting temperature. Furthermore, they should provide a feature to select a minimal set of primers that ensures full template coverage (set cover optimization (SCO)) and should be openly available. Additionally, a smart graphical user interface is helpful for a broad user application.

For example, the primer design tools GeneUp by Pesole et al.[91], G-Primer by Wang et al.[92], GreenSCPrimer by Jabado et al.[93], PAMP by Bashir et al.[94] and PriMux by Gardner et al. and Hysom et al.[95,96] provide a SCO, but out of those, only PriMux is currently available. However, PrimMux on the other hand, as many others, does not provide selection of the desired primer binding region, which is a necessary feature for escape variants analysis or mutation pattern profiling in diverse immunoglobulins. Another example is the Primer approximation multiplex PCR design algorithm by Bashir et al., which intends to generate multiplex primers for the detection of variable genomic lesions in cancer. It provides a versatile algorithm for primer design on genomic level, but it is not transferrable to the IGHV gene locus. Table 1 gives an overview of published multiplex primer design approaches. It demonstrates that all investigated

primer design tools lack at least one of the three major features. A reliable primer design tool for complex and highly mutated immunological templates is therefore urgently needed.

Table 1: Multiplex primer design approaches
This modified table was published in 2020 in the Journal of Immunological Methods by Kreer et al.[90] and is used in approval with all co-authors and with courtesy of Creative Commons Attribution- Share Alike 4.0. International License, *https://creativecommons.org/licenses/by-sa/4.0/*

Tool	Reference	Availability	GUI[a]	SCO[b]
OLIGO 7	(Rychlik, 2007)[97]	YES	YES	X
GeneFisher	(Giegerich et al., 1996)[98]	YES	YES	X
GeneUp	(Pesole et al., 1998)[91]	X	X	YES
CODEHOP	(Rose et al., 1998; Rose et al., 2003)[99,100]	YES	YES	X
DoPrimer	(Kämpke et al., 2001)[101]	X	X	X
HYDEN	(Linhart and Shamir, 2002)[102]	YES	X	X
PROBEMER	(Emrich et al., 2003)[103]	X	YES	X
MIPS	(Souvenir et al., 2003)[104]	X	X	X
Amplicon	(Jarman, 2004)[105]	YES	YES	X
G-PRIMER	(Wang et al., 2004)[92]	X	YES	YES
PDA-MS/UniQ	(Huang et al., 2005)[106]	X	X	YES
MuPlex	(Rachlin et al., 2005)[107]	X	YES	X
GreenSCPrimer	(Jabado et al., 2006)[93]	X	YES	YES
PrimerStation	(Yamada et al., 2006)[108]	YES	YES	X
MultiPrimer	(Lee et al., 2006)[109]	X	X	X
PRIMEGENS	(Srivastava and Xu, 2007)[110]	YES	YES	X
PAMP	(Bashir et al., 2007)[94]	X	X	YES
MPD	(Gardner et al., 2009)[96]	X	X	YES
FastPCR	(Kalendar et al., 2014)[111]	YES	YES	X
MPPrimer	(Shen et al., 2010)[112]	YES	YES	X
URPD	(Chuang et al., 2012)[113]	X	YES	X
PriMux	(Hysom et al., 2012; Gardner et al., 2014)[95,96]	YES	X	YES
PrimerMapper	(O'Halloran et al., 2016)[114]	YES	YES	X
Oli2Go	(Hendling et al., 2018)[115]	YES	YES	X
Ultioplex	(Yuan et al., 2021)	YES	YES	X
*openPrimeR***	*(Döring et al., 2019)*	YES**	YES**	YES**

a The software offers a graphical user interface (GIU) for interaction
b The software provides a set cover optimization (SCO)= automatic selection of minimal primer set that provides full template coverage
** openPrimeR is a newly computed primer design tool whose optimization process is based on the results of this thesis

To overcome the mentioned problems, the research group for Computational Biology (Prof. Dr. Nico Pfeifer) at the Max Planck Institute for Informatics, Saarland and the laboratory for Experimental Immunology (Prof. Dr. Florian Klein) at the Institute of Virology, University of Cologne, developed the primer design tool **openPrimeR** (http:// openprimer.mpi-inf.mpg.de/). openPrimeR is based on the programming language R, it is openly available and provides both, an evaluation and a primer design mode for multiplex PCR primers which target highly diverse templates. Both, primers and templates can be chosen from integrated data (immunoglobulins form the IMGT data base[18] or HIV reference sequence libraries of the Los Alamos National Laboratory HIV sequence database[116]) or can be uploaded individually in fasta format. The user can individually define a target region of interest (ROI).

The evaluation mode allows the user to determine characteristics of established or personal primer sets. The user can individually select of 12 different physicochemical properties (such as melting temperature, GC content, GC clamp) and five coverage conditions (such as number of mismatches or introduced stop codons) to be integrated. Additionally, the tool estimates the primers' performance in multiplex PCR experiments on the priorly selected templates. Results are presented in clearly arranged tables and graphs and allow a structured overview of multiplex properties and template coverage.

In the design mode, the program computes *de novo* multiplex primer sets for the amplification of priorly defined templates and region of interest (ROI). Primers that bind to each other (form dimers) or exhibit other undesired characteristics are excluded. Importantly, by the means of using either an integer linear program (ILP) or a greedy algorithm (for details see Döring et al., 2019, Scientific Reports)[60], the program addresses and ideally solves the SCO.

The optimization process of the tool's raw version, which was performed by extensive validation of pre-defined default and precast settings, presents one of the main parts of this thesis.

2.3   Aims of this thesis

Multiplex polymerase chain reaction (mPCR) techniques have a versatile use in medicine, as they are implemented for the identification of pathogens and disease biomarkers and are often applied in genotyping. They additionally present a valuable tool for the amplification of lymphocyte receptors (such as antibodies). Fast and efficient methods for the detection and analysis of potent neutralizing antibodies are of tremendous value and are urgently needed for the global prevention and treatment of existing and upcoming infectious diseases, which has been demonstrated vividly in the recent years by the epidemic of the human immunodeficiency virus-1 (HIV-1) and the current severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic. One of the major cruxes in the process of antibody identification is the generation of a reliable laboratory protocol, most importantly containing a powerful primer set for the PCR amplification of antibodies even with high mutation status.

Thus, the main goal of this thesis was to further develop single cell and multiplex cloning strategies that are able to identify potent neutralizing antibodies against infectious diseases. To this end, we aimed to establish a standardized and efficiency-optimized amplification protocol. In particular, the focus was set on the generation of a reliable primer testing instrument and the identification of a reliable primer set for the amplification of these antibodies even in highly mutated status.

In a collaboration with colleagues at the Max Planck Institute in Saarbrücken, we intended to optimize a newly developed primer design tool for the generation of primer sets which provide high coverage of diversly mutated immunoglobulin gene segments.

This work is based on the following hypotheses:

- The identification of broad neutralizing antibodies via polymerase chain reaction can be significantly improved by an optimized primer design (including a restricted primer binding region to the V gene's leader region and well-tuned physicochemical primer properties).
- Primer sets, designed using the openPrimeR tool, provide superior performance in amplifying highly mutated human IGHV genes and provide a broader coverage of the entire BCR repertoire compared to established primer sets.
- A synthetic immunoglobulin gene library can provide an unbiased template tool.
- Thermodynamic properties like the free energy of annealing, primer dimerization and the amount and position of mismatches are assumed to be the leading influencing factors of PCR amplification and should therefor be avoided in successful primer design

3   MATERIAL AND METHODS

3.1   Material

3.1.1   Electronic devices

Device Name	Manufacturer	Cat No.
Blue light transilluminator, ECX-F20	VWR	732-1359DE
Gel Doc XR+ Gel documentation system	Bio-Rad	5838
Electrophoresis power supply	Life Technologies	PS0091
Electrophoresis system owl A3-1	Thermo Fisher Scientific	A3-1
Eppendorf 5810R centrifuge, rotor S-4-104	Thermo Fisher Scientific	12836213
Bench top centrifuge 5424 R	Eppendorf	5404000010
BD FACS Aria Fusion	BD biosciences	na
Incubation shaker	Infors HT	na
NanoDrop One spectral photometer	Thermo Fisher Scientific	ND-ONE-W
QiaCube Connect	Qiagen	9002864
Thermomixer	Eppendorf	5384000012
Ultra-low-temperature freezer (−150°C)	Panasonic	MDF-C2156VAN
Ultra-low-temperature freezer (−80°C)	Panasonic	MDF-DU700VH
Veriti 96-well thermal cycler	Thermo Fisher Scientific	4375786
Vortex mixer	neoLab	D-8900

3.1.2   Consumables

Consumable Name	Manufacturer	Cat No.
10 ml serological pipette	Sarstedt	861254001
10 µl filter tips,	Sarstedt	70.1130.210
15 ml centrifugation tube	Sarstedt	62.554.502
200 µl filter tips	Sarstedt	70.760.213
25 ml serological pipette	Sarstedt	86.1685.00
5 ml falcon round-bottom polystyrene test tube with cell strainer snap cap	Corning	352235
5 ml serological pipette	Sarstedt	86.1253.00
5 ml tube with snap cap	VWR	525-0796
50 ml centrifugation tube	Sarstedt	62.547.254

(To be continued next page.)

Consumable Name	Manufacturer	Cat No.
50 ml serological pipette	Sarstedt	86.1256.001
96well PCR plate, semi skirted	PEQlab/VWR	81-35899
Accuspin Tubes Sterile, 50 ml Capacity	Sigma Aldrich	A2055-10EA
Adhesive PCR foil seal	VWR	PEQL82-0626-A
Amicon Ultra centrifugal filter units (30 kDa)	Millipore	Z717185
C-chip Neubauer improved	Carl Roth	T729.1
Cell culture flasks, 250 ml	Corning	CLS431407
MACS LS columns	Miltenyi Biotec	130-042-401
Multichannel pipette 0.5–10 µl	VWR	613-0884
NucleoSpin 96 PCR Clean-up Kit	Macherey Nagel	740658.4
NucleoSpin 96 Plasmid	Macherey Nagel	740625.4
NucleoSpin Gel and PCR Clean-up	Macherey Nagel	740609.250
Nunc 96 DeepWell plate, nontreated	Sigma Aldrich,	Z688738-32EA
Parafilm M	Carl Roth	H666.1
Petridish 92 × 16 mm	Sarstedt	82.1472.001
Preseparation filter (70 µM)	Miltenyi BioTec	130-095-823
Quadro MACS Separator	Miltenyi BioTec	130-090-976
Safe-Lock microcentrifuge tubes 1,5 ml	Sarstedt	0030120.086
Single-use reagent reservoir, 100 ml	Carl Roth	EKX3.1
Single-use reagent reservoir (25 ml)	Carl Roth	EKT7.1
Ultrafree-CL filter columns, 0.22 µm	Merck Millipore	UFC40GV0S

3.1.3 Reagents

Reagent /chemical /enzymes	Vendor	Cat no.
0.5 M EDTA	Thermo Fisher Scientific	AM9260G
6× DNA loading dye	Thermo Fisher Scientific	R0611
ACK Lysing Buffer	Thermo Fisher Scientific	A1049201
Agarose basic	PanReacAppliChem	A8963

(To be continued next page.)

Reagent /chemical /enzymes	Vendor	Cat no.
Cell Dissociation Buffer, enzyme-free, PBS	Thermo Fisher Scientific	13151014
CD19 microbeads, human	Miltenyi Biotec	130-050-301
dATP Solution (100mM)	ThermoFisher Scientific	R0141
DMSO	Sigma Aldrich	D2650
DNA Gel Loading Dye (6X)	ThermoFisher Scientific	R0611
dNTPs, 25 mM each	Thermo Fisher Scientific	R1122
Fetal Bovine Serum (FBS)	Sigma Aldrich	F4135
GeneRuler 1 kb Plus DNA ladder	Invitrogen	B00351
Hanks' Balanced Salt Solution (HBSS)	Sigma Aldrich	H6648
Histopaque-1077 Hybri-Max	Sigma	H8889
HotStarTaq DNA Polymerase (incl. Qiagen PCR Buffer)	Qiagen	203205
Nuclease-free water (not DEPC treated)	Thermo Fisher Scientific	AM9937
Platinum Taq Green Hot Start DNA Polymerase (incl. 10x Buffer, Mg, KB-Extender)	Thermo Fisher Scientific	11966018
Protein G Sepharose 4 Fast Flow	GE Life Sciences	17061805
Q5® Hot Start High-Fidelity 2X Master Mix	New, England Biolabs	M0494S
Random Hexamer Primer, 0.2 µg/µl	Thermo Fisher Scientific	S0142
Recombinant RNasin® Ribonuclease Inhibitor	Promega	N2515
RNaseOUT, 40 U/µl	Invitrogen	10777-019
Rnasin, 40 U/µl	Promega	N2515
RPMI (Roswell Park Memorial Institute) 1640 Medium	Life Technologies	11875-093
S.O.C Medium	Invitrogen	15544034
SuperScript™ III Reverse Transcriptase Kit	Thermo Fisher Scientific	18080044
SuperScript™ IV Reverse Transcriptase Kit	Thermo Fisher Scientific	18090050
SYBR Safe DNA gel strain	Invitrogen	S33102
Taq DNA polymerase	Quiagen	201203
Universal primer Mix (10x)	Clontech	634922

3 Material and Methods

3.1.4 Antibodies/ FACS staining

Antibody /staining	Vendor	Cat. No.
DAPI	Invitrogen	D1306
APC-H7 Mouse Anti-Human CD19	BD biosciences	560727
APC mouse anti-human IgG	BD Biosciences	550931
Alexa Fluor 700 mouse anti-human CD20	BD biosciences	560631
FITC mouse anti-human IgM	BD Biosciences	555782
hu IgD- PE-Cy7	BD Biosciences	561314
hu CD27-PerCPCy5.5	BD Biosciences	655429

3.1.5 Kits

Kit Name	Vendor	Cat. No.
IgG+ Memory B Cell Isolation Kit, human	Miltenyi Biotec	130-094-350
Rneasy Micro Kit	Qiagen	74004
SMARTer ®RACE 5'/3'-Kit	Clontech	634858, 634859
GeneJET Gel Extraction and DNA Cleanup Micro Kit	ThermoFisher Scientific	K0832
TOPO TA Cloning Kit, pCR4-TOPO vector, without competent cells	Invitrogen	450071

3.1.6 Cell lines

Competent DH5α Escherichia Coli, ThermoFisher scientific, Cat#18263012

3.1.7 Reagent setup

Tris-acetate-EDTA buffer (1x TAE buffer)

For 1x TAE buffer 242 g of Tris-(hydroxymethyl)-aminomethane were dissolved in 600 ml of distilled H_2O. 200 ml of 0.5 Molar (M) ethylenediaminetetraacetic acid (EDTA) (pH 8.0) and 57.1 ml of acetic acid were added. The buffer was filled to a final volume of 1 L with distilled H_2O. Finally, the TAE buffer was autoclaved. Storage at room temperature.

Fluorescence Activated Cell Sorting (FACS) buffer

FACS buffer was produced by adding 10 ml of 2% fetal bovine serum (FBS) and 2 ml of 0.5 M EDTA to 488 ml of 1x phosphate buffered saline (PBS). FACS buffer was stored at 4-8°C.

3 Material and Methods

Freezing medium (vol/vol)

Freezing medium consists of 90% FBS + 10% dimethyl sulfoxide (DMSO) and was produced daily fresh. Freezing medium was cooled to 4°C prior to usage.

Magnetic Activated Cell Separation (MACS) buffer

MACS buffer was produced by adding 200 µl of 0.5 M EDTA and 5 ml of 5% (wt/vol) sterile filtered bovine serum albumin (BSA) to 45 ml of PBS. MACS buffer was stored at 4-8 °C.

Ampicillin stock solution (50 mg/ml)

Ampicillin stock solution was produced by dissolving 0.5 g of ampicillin sodium salt in 10 ml of distilled H_2O. The solution was sterilized by filtering with a 0.22 µm Ultrafree-CL PVDF filter column. The store stock solution was stored at −20°C.

Agarose gel (wt/vol)

Agarose gels were prepared by adding 1 g (1%) or 2 g (2%) of agarose to 100 ml of TAE buffer. The solution was boiled until the agarose was completely dissolved. After cooling down to around 50°C, 5 µl of SYBR Safe DNA Gel Stain per 100 ml gel (1:20,000 dilution) were added and the gel was poured in a casting chamber.

Luria broth medium (LB medium)

LB medium was prepared by dissolving 10 g of tryptone, 5 g of yeast extract and 5 g of sodium chloride in 1 L of distilled water. Afterwards the medium was autoclaved and stored at 4–8°C.

LB-ampicillin agar plates

15 g of agar bacteriology grade were dissolved in 1 l of LB medium and sterilize by autoclaving. After cooling down to room temperature 1 ml of ampicillin stock solution (50µg/ml) was added. The solution was poured in Petri dishes and cooled down before storage at 4°C.

Pre-Terrific broth (TB) medium

12 g of tryptone, 24 g of yeast extract and 5 ml of glycerol were dissolved in 1 l of distilled water. Then the solution was autoclaved and store at 4°C.

TB salt buffer (10×)

125.4 g of K2HPO4 and 23.1 g of KH2PO4 were dissolved in 1 l of distilled water. The solution was autoclaved and stored at 4°C.

TB medium

3 Material and Methods

Terrific broth medium was prepared by adding 100 ml of TB salt buffer (10×) to 900 ml of preTB medium. TB medium was stored at 4°C.

3.1.8 Primers

The following table lists all primers used for single cell and bulk amplification as well as primers tested on the IGHV gene library. Previously published sets, here indicated by the first authors name, and newly computationally optimized multiplex primer sets from openPrimeR (oPR(1)-IGHV, oPR(3)-IGHV, oPR(4)-IGHV, oPR(5)-IGHV and oPR(6)-IGHV) were used. All primers were ordered from Thermo Fisher Scientific (formerly Life Technologies) under the brand of InvitorgenTM custom DNA oligos. All primers were diluted with nuclease-free water to a stock concentration of 50 μM. A primer volume ratio of 1:1 was used for primer mixes. Primers and primer mixes were stored at −20°C.

Melting temperatures ™ were calculated using the online Tm calculator from New England BioLabs (Version 1.13.0). Annealing temperature used for PCR with respective primer mixes were indicated in bold.

Table 2: List of primers

Set1 (Tiller_20'8 5' L-VH Mix)[81]		
Name	**Sequence**	**Tm**
5´ L-VH 1	ACAGGTGCCCACTCCCAGGTGCAG	70°C
5´ L-VH 3	AAGGTGTCCAGTGTGARGTGCAG*	60-62°C
5´ L-VH 4/6	CCCAGATGGGTCCTGTCCCAGGTGCAG	67°C
5´ L-VH 5	CAAGGAGTCTGTTCCGAGGTGCAG	58°C
*calculated annealing temperature for multiplex primer set		**(57°C)***

Set1 (Tiller_20'8 5' AgeI VH Mix)[81]		
Name	**Sequence**	**Tm**
5' AgeI VH1	CTGCA<u>ACCGGT</u>GTACATTCCCAGGTGCAGCTGGTGCAG	76°C
5' AgeI VH1/5	CTGCA<u>ACCGGT</u>GTACATTCCGAGGTGCAGCTGGTGCAG	75°C
5' AgeI VH3	CTGCA<u>ACCGGT</u>GTACATTCTGAGGTGCAGCTGGTGGAG	73°C
5' AgeI VH3-23	CTGCA<u>ACCGGT</u>GTACATTCTGAGGTGCAGCTGTTGGAG	72°C
5' AgeI VH4	CTGCA<u>ACCGGT</u>GTACATTCCCAGGTGCAGCTGCAGGAG	75°C
5' AgeI VH4-34	CTGCA<u>ACCGGT</u>GTACATTCCCAGGTGCAGCTACAGCAGTG	74°C
5' AgeI VH1-18	CTGCA<u>ACCGGT</u>GTACATTCCCAGGTTCAGCTGGTGCAG	74°C
5' AgeI VH1-24	CTGCA<u>ACCGGT</u>GTACATTCCCAGGTCCAGCTGGTACAG	73°C
5' AgeI VH3-33	CTGCA<u>ACCGGT</u>GTACATTCTCAGGTGCAGCTGGTGGAG	73°C
5' AgeI VH3-9	CTGCA<u>ACCGGT</u>GTACATTCTGAAGTGCAGCTGGTGGAG	72°C
5' AgeI VH4-39	CTGCA<u>ACCGGT</u>GTACATTCCCAGCTGCAGCTGCAGGAG	75°C
5' AgeI VH6-1	CTGCA<u>ACCGGT</u>GTACATTCCCAGGTACAGCTGCAGCAG	74°C
*calculated annealing temperature for multiplex primer set		**(57°C)***

Set 1(Tiller_2008)[81] reverse primers		
Name	**Sequence**	**Tm**
3´Cµ CH1 (γ)	GGAAGGTGTGCACGCCGCTGGTC	69°C
3´ IgG _internal	GTTCGGGGAAGTAGTCCTTGAC	58°C

(To be continued next page.)

Set2 (Ippolito_2012)[86,88]		
Ippolito_VH1-fw	CAGGTCCAGCTKGTRCAGTCTGG**	60-66°C
Ippolito_VH157-fw	CAGGTGCAGCTGGTGSARTCTGG**	64-67°C
Ippolito_VH2-fw	CAGRTCACCTTGAAGGAGTCTG**	55-58°C
Ippolito_VH3-fw	GAGGTGCAGCTGKTGGAGWCY**	60-64°C
Ippolito_VH4-fw	CAGGTGCAGCTGCAGGAGTCSG**	67°C
Ippolito_VH4-DP63-fwd	CAGGTGCAGCTACAGCAGTGGG	64°C
Ippolito_VH6-fw	CAGGTACAGCTGCAGCAGTCA	60°C
Ippolito_VH3N-fw	TCAACACAACGGTTCCCAGTTA	57°C
*calculated annealing temperature for multiplex primer set		**(63°C)***
** degenerated primer: K represents G or T; R represents A or G, S represents C or G, W represents A or T, Y represents C or T		

Set2 (Ippolito_2012)[86,88] reverse primers		
Ippolito_IgM_all_rev	GGTTGGGGCGGATGCACTCC	65°C
Ippolito_IgG_all_rev	SGATGGGCCCTTGGTGGARGC*	65-67°C
* degenerated primer: K represents G or T; R represents A or G, S represents C or G, W represents A or T, Y represents C or T		

oPR(1)-IGHV		
Name	**Sequence**	**Tm**
IGHV_Opt1_1_fw	AGGTGTCCAGTGTGAGGTGCA	62°C
IGHV_Opt1_2_fw	AGCTCCCAGATGGGTCCTGT	61°C
IGHV_Opt1_3_fw	ACTTTGTTCCACGCTCCTGCT	60°C
IGHV_Opt1_4_fw	TCCCAGGTTCAGCTGGTGCA	62°C
IGHV_Opt1_5_fw	TCCTCTTCTTGGTGGCAGCAG	60°C
IGHV_Opt1_6_fw	TCACAGGTACAGCTGCAGCAGT	62°C
IGHV_Opt1_7_fw	ATGGGGTCAACCGCCATCCT	62°C
IGHV_Opt1_8_fw	AGGTGTCCAGTGTCAGGTGCAG	63°C
IGHV_Opt1_9_fw	ACACACTCCTGCTGCTGACCAC	63°C
*calculated annealing temperature for multiplex primer set		**(54°C)***

(To be continued next page.)

oPR(3)-IGHV		
Name	**Sequence**	**Tm**
IGHV_Opt3_1_fw	AGTGTGAGGTGCAGCTGGTGGAGTC	66°C
IGHV_Opt3_2_fw	CACCTGTGGTTCTTCCTCCTCCTGGTG	66°C
IGHV_Opt3_3_fw	ATGGGTCCTGTCCCAGGTACAGCT	65°C
IGHV_Opt3_4_fw	TTTGGGCTGAGCTGGGTTTTCCTCGTT	66°C
IGHV_Opt3_5_fw	CTTTGTTCCACGCTCCTGCTGCTGAC	66°C
IGHV_Opt3_6_fw	GGGTCTTATCCCAGGTCACCTTGAAGGAGT	66°C
IGHV_Opt3_7_fw	ACTCCCAGGTTCAGCTGGTGCAGT	66°C
IGHV_Opt3_8_fw	CCTGGAGGATCCTCTTCTTGGTGGCA	65°C
IGHV_Opt3_9_fw	TGGACCTGGAGGTTCCTCTTTGTGGTG	65°C
IGHV_Opt3_10_fw	TTCTCCAAGGAGTCTGTGCCGAGGTG	65°C
IGHV_Opt3_11_fw	GGGGTGTCCTGTCACAGGTACAGCT	66°C
IGHV_Opt3_12_fw	ACAGGTGCCCACTCCCAAATGCAG	65°C
IGHV_Opt3_13_fw	TCCCCTCCACAGTGAGAGTCTGTGC	66°C
IGHV_Opt3_14_fw	AAAGCTGTCCAGTGTCAGGTGCAGTCTG	65°C
*calculated annealing temperature for multiplex primer set		**(59°C)***

oPR(4)-IGHV		
Name	**Sequence**	**Tm**
IGHV_Opt4_1_fw	CTGCTGGTGGCAGCTCCC	64°C
IGHV_Opt4_2_fw	TGGAGTTGGGGCTGAGCT	61°C
IGHV_Opt4_3_fw	CGCTGGGTTTTCCTTGTTGCTATTT	60°C
IGHV_Opt4_4_fw	GAGCTGGGTTCTCCTTGTTGC	61°C
IGHV_Opt4_5_fw	CACGCTCCTGCTGCTGAC	61°C
IGHV_Opt4_6_fw	TCCTCTTCTTGGTGGCAGCA	61°C
IGHV_Opt4_7_fw	TGGACTGGACCTGGAGGG	61°C
IGHV_Opt4_8_fw	GGTCAACCGCCATCCTCGC	64°C
IGHV_Opt4_9_fw	TCCTTCCTCATCTTCCTGCCG	61°C
IGHV_Opt4_10_fw	ACTTAAACCCAGGCTCCCCTCCA	65°C
IGHV_Opt4_11_fw	TGATTGCTGAGCTGTTCTGTGCT	60°C
*calculated annealing temperature for multiplex primer set		**(62°C)***

(To be continued next page.)

oPR(5)-IGHV		
Name	**Sequence**	**Tm**
IGHV _Opt5_1_fw	CACCTGTGGTTCTTCCTCCTCC	61°C
IGHV_Opt5_2_fw	CACCTGTGGTTCTTCCTCCTGC	61°C
IGHV_Opt5_3_fw	ATGGAGTTTGGGCTGAGCTGG	61°C
IGHV_Opt5_4_fw	ATGGAGTTGGGGCTGAGCTG	61°C
IGHV_Opt5_5_fw	TGGAGTTTTGGCTGAGCTGGG	61°C
IGHV_Opt5_6_fw	ACTTTGCTCCACGCTCCTGC	62°C
IGHV_Opt5_7_fw	ATGGACTGGACCTGGAGCATC	60°C
IGHV_Opt5_8_fw	ATGGACTGGACCTGGAGGTTCC	62°C
IGHV_Opt5_9_fw	ATGGACTGCACCTGGAGGATC	60°C
IGHV_Opt5_10_fw	ATGGACTGGACCTGGAGGGTCTTC	63°C
IGHV_Opt5_11_fw	TCTGTCTCCTTCCTCATCTTCCTGC	61°C
IGHV_Opt5_12_fw	GGACTGGATTTGGAGGGTCCTCTTC	62°C
IGHV_Opt5_13_fw	GCTCCGCTGGGTTTTCCTTG	60°C
IGHV_Opt5_14_fw	TGGGGTCAACCGCCATCC	62°C
IGHV_Opt5_15_fw	GGCCTCTCCACTTAAACCCAGG	61°C
IGHV_Opt5_16_fw	TGGACACACTTTGCTACACACTCC	60°C
*calculated annealing temperature for multiplex primer set		**(55°C)***

oPR(6)-IGHV		
Name	**Sequence**	**Tm**
IGHV _Opt6_1_fw	GTGGTTCTTCCTCCTCCTGGTG	61°C
IGHV_Opt6_2_fw	GGCTGAGCTGGGTTTTCCTTGTTG	62°C
IGHV_Opt6_3_fw	ATGGAGTTGGGGCTGAGCTG	61°C
IGHV_Opt6_4_fw	TGGACTGGACCTGGAGGATC	59°C
IGHV_Opt6_5_fw	GACATACTTTGTTCCACGCTCCTGC	61°C
*calculated annealing temperature for multiplex primer set		**(57°C)***

Ozawa_2006[117] reverse primers		
Name	**Sequence**	**Tm**
Ozawa_Cg-RT_rev	AGGTGTGCACGCCGCTGGTC	67°C
Ozawa_Cg-1st_rev	CGCCTGAGTTCCACGACACC	62°C
Ozawa_Cm-RT_rev	ATGGAGTCGGGAAGGAAGTC	57°C
Outer_kappa_rev	GGTGACTTCGCAGGCGTAG	68°C
Outer_lambda_1_rev	GCCGCGTACTTGTTGTTGC	68°C

M13 (pCR4-TOPO)		
Name	**Sequence**	**Tm**
M13_fw	GTAAAACGACGGCCAG	51°C
M13_rev	CAGGAAACAGCTATGAC	47°C

3.2   Methods

3.2.1   Preparation of single cell cDNA from human donors

3.2.1.1 Peripheral Blood Mononuclear Cell (PBMC) isolation

Blood samples from healthy individuals were collected at the blood bank of the University hospital Cologne in accordance with the requirements of the local Institutional Review Board (IRB; protocol number 16–054, University of Cologne, Cologne, Germany). Written informed consent was obtained from all individuals.

The presented method of peripheral blood mononuclear cell (PBMC) isolation separates the human blood cells by their density and facilitates the collecting of PBMCs without erythrocyte contamination. The whole blood was aliquoted in 50 ml falcon tubes, 2 mM EDTA were added and the blood was diluted in a ratio of 1:2 with Hanks' Balanced Salt Solution (HBSS). PBMC isolation was performed using Accuspin Tubes (50 ml). Accuspin Tubes were prepared by adding 15 ml Histopaque and centrifugated at 20°C and 800 relative centrifugal force (rcf) for one minute. Afterwards, a minimum of 10 ml and a maximum of 25 ml sample was added on top of the membrane followed by a centrifugation step that was performed at 20°C for 15 minutes at 900 rcf without brakes or acceleration. All further steps were performed on ice. PBMCs of each Accuspin Tube, visible as a dense band at the plasma/ Histopaque interface, were carefully collected by means of a pipette and transferred into separate 50 ml falcon tubes containing 25 ml cold Roswell Park Memorial Institute (RPMI)-Medium with 10% fetal bovine serum (FBS). Afterwards, the cells were centrifuged at 400 rcf for 25 minutes. The supernatant was discarded. Either direct proceeding with CD19 isolation was performed or PBMCs were cryopreserved with FBS and 10% DMSO and stored at the -80°C freezer for 24 hours before transferring to a -50°C freezer for long-term storage.

3.2.1.2 CD19 isolation from PBMCs with Miltenyi Microbeads

CD19+ cells were isolated from PBMCs with a positive selection kit via magnetic activated cell separation (MACS). According to the manufacturer's instructions, we used micro beads to magnetically label CD19+ cells and separate them with a specific column. However, the volume of the CD19 micro beads was reduced to 20% of the volume suggested by the manufacturer.

3.2.1.3 Fluorescence-activated cell sorting (FACS) for naive and antigen-experienced single cells

3.2.1.4 Cell preparation

CD19+ cells isolated from PBMCs were spun down and resuspended in 100 µl FACS buffer.

A first sort was performed with 3 x10^6 CD19+ PBMCs. Compensation controls were prepared with one unstained probe and respectively one probe for each single stain. 4',6-diamidino-2-phenylindole (DAPI) was used to stain fixed cells. Staining was performed with CD19 (APC-H7 Mouse Anti-Human CD19), CD20 (Alexa Fluor 700 mouse anti-human CD20), IgG-APC (APC mouse anti-human IgG), and IgM-FITC (FITC mouse anti-human IgM).

Table 3: Dilutions of antibodies and staining reagents used for single-cell sort

Antibody/ stain	Dilution
DAPI	1:100
APC-H7 Mouse Anti-Human CD19	na
APC mouse anti-human IgG	1:20
Alexa Fluor 700 mouse anti-human CD20	1:80
FITC mouse anti-human IgM	1:5

After 20 minutes incubation on ice under light-protected conditions the samples were washed again in 15 ml FACS buffer. Afterwards, the cells were spun down for 10 minutes at 400 g, the supernatant was discarded and the pellet was resuspended again in 2 ml FACS buffer. The strainer caps were wet with PBS before filtering the samples through. Then probes were gently vortexed and loaded onto the sorter, using the BD FACS Aria Fusion cell sorter.

3.2.1.5 Single Cell sorting

The BD FACS Aria Fusion was used for all cell sorts and analyses. Lymphocyte populations were identified in the FCS-A/SSC-A and single cells in the FCS-H/FCS-A. Gating was set to DAPI negative cells in FCS-A/DAPI. To identify B cells FCS-A/CD20 and FCS-A/CD19 cells were plotted, and gaiting was set to CD19+ and CD20+ cells. The final sorting gate was set up by displaying the bait protein against IgM or IgG and gaiting into the respectively IgM+ or IgG+ fraction was performed. Naive single B-cells, here defined as CD20+IgM+IgG-, and antigen experienced single B cells, defined as CD20+IgM-IgG+ cells, were sorted into 96-well PCR plates. Each well contained 4 µl of lysis buffer.

Table 4: Composition of lysis buffer

Component	Amount per well (µl)	Final concentration
Nuclease-free H_2O	3.1	
RNAsin (40 U/µl)	0.2	2 U/µl
RNase Out (40 U/µl)	0.1	1 U/µl
PBS (10x)	0.2	0.05x
DTT (100 mM)	0.4	1 mM
Sum	4	

Plates, which now contained ribonucleic acid (RNA) of the lysed single cells, were sealed with adhesive PCR foil and were directly frozen at -80°C.

A second single cell sort, for CD20+IgM-IgG+ single cells only, was performed with 1.75 x10^6 PBMCs. The same protocol, including the same staining and gaiting panel was used, except for reduction by CD19 APC-H7.
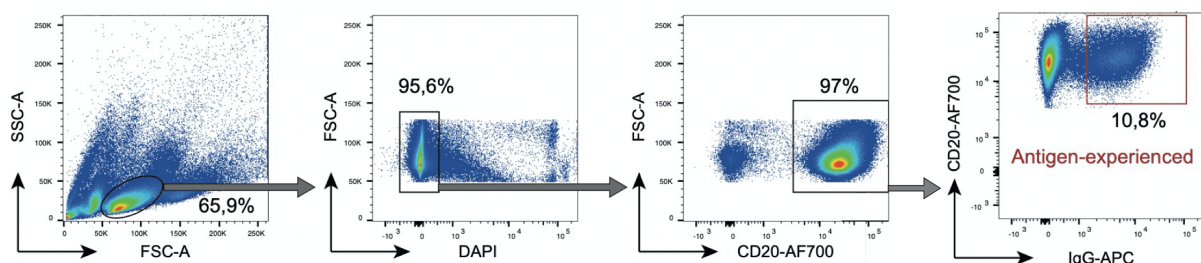
Figure 7: Gating strategy for single cell sorting of antigen-experienced B cells

3.2.1.6 Reverse Transcription (RT)

Total RNA from single cells was reverse transcribed into cDNA in the original 96-well sorting plate. The following steps were performed in a specifically prepared room for cDNA preparation, where no plasmid experiments take place. To exclude any kind of RNase or DNA contamination, all surfaces were wiped with RNase/DNA AWAY prior to the experiments. The plates were thawed on ice directly before the RT experiment.

Firstly, a Random Hexamer Primer Mix (RHP-mix) was prepared.

Table 5: Random Hexamer Primer (RHP) Mix

reagent	per plate*	per well	final conc.
Random Hexamer Primer (200 ng/µl)	75 µl	0.75 µl	21.43 ng /µl
NP-40 (10% (vol/vol))	50 µl	0.5 µl	0.71%(vol/vol)
RNaseOUT (40 U/µl)	15 µl	0.15 µl	0.86 U /µl
RNase free water	560 µl	5.60 µl	
Sum	700 µl	7 µl	

* calculated for 100 wells (4 extra wells) when using a multichannel pipette to avoid shortage in the last wells

7 µl of the RHP-mix were added to each well and carefully pipetted up and down. A 12-channel pipette was used to facilitate these steps. The plate was heated up to 65°C for one minute and afterwards placed on ice again for at least two minutes. Meanwhile the Reverse Transcription mix (RT mix) was prepared.

Table 6: Reverse Transcription (RT) Mix

reagent	per plate*	per well	final conc.
RNase free water	205 µl	2.05 µL	
5x RT Buffer	300 µl	3 µl	2.1x
DNTPs (25mM)	50 µl	0.5 µl	1.8 mM
DTT (100mM)	100 µl	1 µl	14 mM
RNasin (40 U/µl)	10 µl	0.1 µl	0.57 U /µl
RNaseOUT (40 U/µl)	10 µl	0.1 µl	0.57 U /µl
RT SuperScript IV** (200 U/µl)	250 µl	0.25 µl	7.14 U /µl
Sum	700 µl	7 µl	

* calculated for 100 wells (4 extra wells) when using a multichannel pipette to avoid shortage in the last wells
** Note: first experiments were performed using SuperScript III with the exact same protocol

7 µl of the RT mix was added to each well. Plates were covered with adhesive foil before briefly rinsing at 2000 rpm and incubating at room temperature for 10 to 15 minutes. RT was performed in a thermocycler first at 42°C for 10 minutes and then successively at 25°C for 10 minutes, at 50°C for 60 minutes and finally at 94°C for 5 minutes. Subsequently, the plates, which now contained the cDNA of a single B cell in each well, were frozen at -80°C or forwarded directly to PCR amplification (3.2.3).

3.2.2 Preparation of BG505$_{SOSIP.664}$+ single B cells

3.2.2.1 Sample collection and PBMC isolation

Blood samples from HIV-1-infected individuals were collected in accordance with the requirements of the local Institutional Review Board (IRB; protocol number 16-054, University of Cologne, Cologne, Germany). Written informed consent was obtained from all individuals.

PBMCs were isolated density-gradient centrifugation and stored at -150°C in 90% FBS and 10% DMSO until further use.

3.2.2.2 Single Cell Sort for HIV specific B Cells

The IgG+ Memory B Cell Isolation Kit (Miltenyi Biotec) was used to isolate B cells from PBMCs. Isolated cells were labeled with anti-human CD19-AF700 (BD), anti-human IgG-APC (BD), DAPI, and the respective HIV-1 Env bait. Therefor they were placed for 30 minutes on ice.

BG505SOSIP.664-Green fluorescent protein (-GFP) was used as HIV-1 Env bait. Env-reactive CD19+IgG+DAPI- single cells were sorted into 96-well plates containing 4 ml of lysis buffer as described above. Plates were stored at -80°C until further use. Afterwards, single cell cDNA synthesis was performed by reverse transcription as described above *(3.2.1.6)*.

Table 7: Dilution of fluorescently labeled antigen used for (HIV-1-specific) single-cell sort

Labeled antigen	Final concentration	Dye	Conjugation method	Estimated dye/ protein ratio
BG505-SOSIP	15 µg/ml	GFP	Fusion protein	1:1
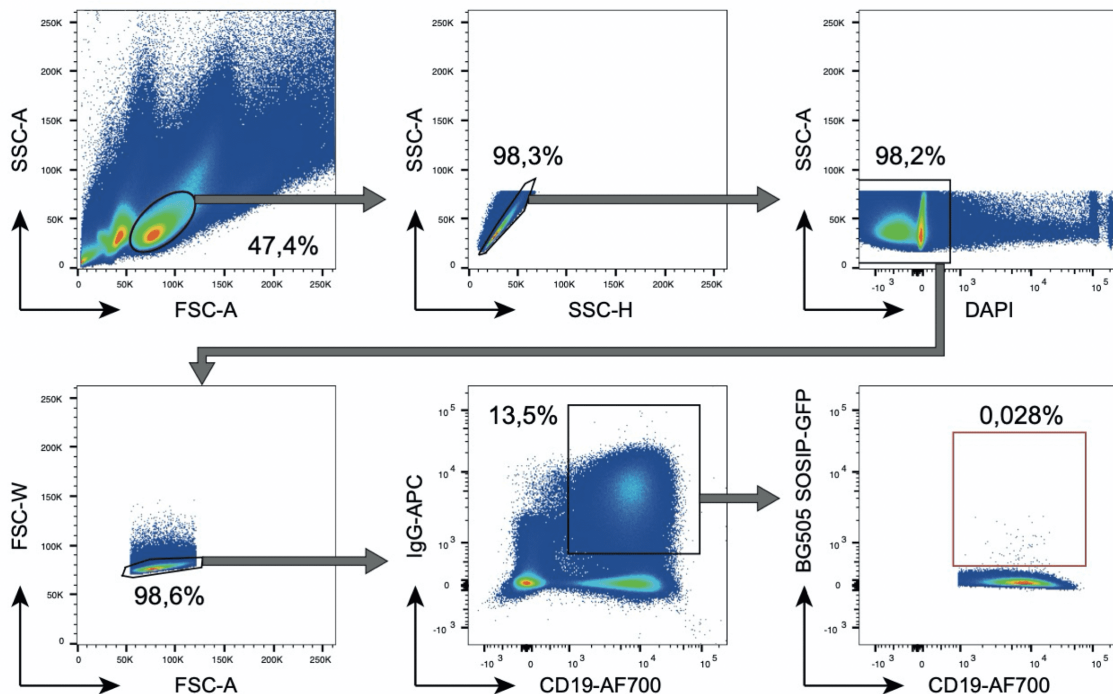
Figure 8: Gating strategy for HIV-1-specific B cells

3.2.3   Immunoglobulin gene amplification on single B cell level

3.2.3.1 PCR protocol for immunoglobulin gene amplification on single B cells

The target region of the single cell PCR experiments encompassed the heavy and light chain loci of the respective B cell. IgH, Igλ and Igκ V gene transcripts were amplified in independent experiments by nested or semi-nested PCR. If not indicated otherwise, 4 µl of single cell cDNA were used as template, allowing up to four different PCRs on one single B cell. Either PCR amplification for heavy, kappa and lambda chain or different primer sets on one chain were performed on each lysed single cell for comparison and analysis. First PCR reactions were

performed in 96-well plates in a total volume of 4 µl cDNA plus 38 µl 1st PCR-Mix per well. Different forward primer sets (Table 2) and corresponding reverse primers were tested under standardized conditions as followed.

Table 8: 1st PCR mix

reagent	per plate	per well	final conc.
Nuclease free water	3253 µl	32.53 µl	
10x Taq Buffer	448 µl	4.48 µl	1.18x
dNTP (25mM)	48 µl	0.48 µl	0.31 mM
5' primer (50mM)	15 µl	0.15 µl	0.2 µM
3' primer (50mM)	15 µl	0.15 µl	0.2 µM
HotstarTaq 5 U/µl)	21 µl	0.21µl	1.05 U/reac.
Sum	3800 µl	38 µl	

Table 9: 1st PCR cycler program

temperature	time	repetition
94°C	10 min	1x
94°C	30 sec	
X °C*	30 sec	50x
72°C	55 sec	
72°C	10 min	1x
4°C	hold	

*annealing temperature depending on used primer set (see Table 2).

Nested or semi-nested 2nd PCR reactions with gene-specific primer mixes were performed with 3.5 µl of unpurified first PCR product plus 38 µl 2nd PCR mix per reaction.

3 Material and Methods

Table 10: 2nd PCR Mix

reagent	per plate*	per well	final conc.
Nuclease free water	2428 µl	24.28 µl	
10x Taq Buffer	444 µl	4.44 µl	1,17x
loading buffer**	829 µl	8.29 µl	
dNTP (25mM)	47 µl	0.47 µl	0.31 mM
5' primer (50µM)	15 µl	0.15 µl	0.2 µM
3' primer (50µM)	15 µl	0.15 µl	0.2 µM
HotstarTaq	21 µl	0.21 µl	1.05 U/reac.
Sum	3800 µl	38 µl	

*calculated for 100 wells (4 extra wells) if using a multi-channel pipette to avoid shortage in the last wells

3 Material and Methods

Table 11: 2nd PCR cycler program

temperature	time	repetition
94°C	10 min	1x
94°C	30 sec	
X°C*	30 sec	50x
72°C	45 sec	
72°C	10 min	1x
4°C	hold	

*annealing temperature depending on used primer set (see Table 2).

First experiments were performed with the Qiagen HotstarTaq as described above. During our workflow we changed our standard polymerase to the PlatinumTaq DNA Polymerase by ThermoFisher and proceeded with the following changes in protocol: 1st PCR reactions were performed in 96-well plates in a total volume of 4 µl cDNA plus 38 µl 1st PCR mix (Platinum) per well.

Table 12: 1st PCR Mix (PlatinumTaq)

reagent	per plate*	per well	final conc.
Nuclease free water	2940 µl	29.4 µl	
10X Green PCR Buffer	410 µl	4.1 µl	1.1x
KB Extender (6%)	246 µl	2.46 µl	0.4%
50 mM MgCl2	123 µl	1.23 µl	1.6 mM
dNTPs (25 mM each)	33 µl	0.33 µl	0.22 µM
5' primer (50 µM)	16,5 µl	0.16 µl	0.22 µM
3' primer (50 µM)	16,5 µl	0.16 µl	0.22 µM
Platinum Taq (10 U/µl)	16,5 µl	0.16 µl	0.04 U/µL
Sum	3800 µl	38 µl	

*calculated for 100 wells (4 extra wells) when using a multichannel pipette to avoid shortage in the last wells

3 Material and Methods

Table 13: 1ˢᵗ PCR cycler program (PlatinumTaq)

temperature	time	repetition
94°C	2 min	1x
94°C	30 sec	
X °C*	30 sec	50x
72°C	55 sec	
4°C	hold	

*annealing temperature depending on used primer set (see Table 2).

For nested or semi-nested 2ⁿᵈ PCR 4 µl unpurified 1ˢᵗ PCR product and **19 µl** 2ⁿᵈ PCR mix (Platinum) were used.

Table 14: 2ⁿᵈ PCR Mix (PlatinumTaq)

reagent	per plate*	per well	final conc.
Nuclease free water	1477 µl	14.7 µl	
10X Green PCR Buffer	206 µl	2.05 µl	1.1x
KB Extender (6%)	123.5 µl	1.23 µl	0.4%
50 mM MgCl2	61.8 µl	0.62 µl	1.6 mM
dNTPs (25 mM each)	16.5 µl	0.16 µl	0.22 µM
5' primer (50 µM)	8.25 µl	0.08 µl	0.22 µM
3' primer (50 µM)	8.25 µl	0.08 µl	0.22 µM
Platinum Taq (10 U/µl)	8.25 µl	0.08 µl	0.04 U/µl
Sum	1900 µl	19 µl	

*calculated for 100 wells (4 extra wells) when using a multichannel pipette to avoid shortage in the last wells

Table 15: 2nd PCR cycler program (PlatinumTaq)

temperature	time	repetition
94°C	2 min	1x
94°C	30 sec	
X °C*	30 sec	50x
72°C	45 sec	
4°C	hold	

*annealing temperature depending on used primer set (see Table 2)

All 2nd PCR products were checked by agarose gel electrophoresis as follows. Bands were expected according to the used forward and reverse primer at 370-519 bp.

3.2.3.2 Agarose gel electrophoresis

The agarose gel electrophoresis was methodically used to separate DNA fragments by an electrical field according to their individual length. To obtain a 2% agarose gel with 6 rows à 50 lanes, 8 g agarose were boiled in 400 ml 1x TAE buffer for 2 minutes and carefully shook manually. The suspension was boiled until the agarose was fully dissolved and became totally clear. The suspension was cooled down to around 40°C before SYBR safe was added in a dilution of 1:20.0000. The suspension was shaken manually until the mixture was homogeneously colored. The gel was poured directly afterwards, avoiding air bubbles and cooled down until it solidified and turned milky dull. Then the gel was covered with 1x TAE Buffer and loaded with 8 μl of PCR products per lane. 6 μl GeneRuler 1 kb Plus DNA ladder were loaded in the first and last lane of each row as size indicator. The gels were usually run at 150 to 200 volt until the dye line was approximately 80% of the way down the gel. For visualization of the DNA fragments the BioRAD Gel DocTM XR+ Imaging system was used and gels were photographed for documentation and analysis.

3.2.3.3 Scoring of visualized PCR products

A visible band at the correct height was encoded with "1" or with the color green, meaning a positive coverage. No band or bands with incorrect length were encoded with "0" or colored grey/black. If triplicates or quadruplicates were performed, the cut off for "1" was set to at least 2 out of 3 correct bands or at least 3 out auf 4 correct bands.

3.2.3.4 Special PCR conditions for touch-up cycle PCR using primer Set2

Primer Set2 was designed for the bulk amplification of IGHV sequences in human and mice.[86,88] Amplification on single cell cDNA had not been described with this primer set in the literature before. We therefore adapted the cycler program recommended from the authors` protocol to single cell PCR amplification and extended the replication step from 20 to 42 cycles. In addition, we performed a semi-nested second PCR with a reverse primer (Ozawa_cg_rt) binding closer to the 5' end.

Table 16: 1st and 2nd single cell PCR cycler program for Set2

temperature	time	repetition
94°C	10 min	1x
92°C	1 min	
50°C	1 min	4x
72°C	1 min	
92°C	1 min	
55°C	1 min	4x
72°C	1 min	
92°C	1 min	
63°C	1 min	42x
92°C	1 min	
74°C	7 min	1x
4°C	hold	

3.2.4 OpenPrimeR

3.2.4.1 Computation of the OpenPrimeR design tool

OpenPrimeR is a newly constructed primer design tool, which intends to help researchers generate primer sets that are suitable for the amplification of diverse immunological templates and overcome difficulties presented by high mutational burden.[90]

It provides a clear and easy-to-use interface. It can be used to evaluate existing primer sets or design new primers for highly diverse templates. To generate an optimized primer set with a minimum of different primers and to provide full template coverage, the program uses a greedy algorithm or an integer linear program (for details see Döring et al., 2019, Scientific Reports).[60] The program supports the operating systems Linux, macOS, and Windows and is openly available at *http:// openprimer.mpi-inf.mpg.de/.*[60] For full program usage, the installation of the additional programs MELTING[118], ViennaRNA[119], and OligoArrayAux[120] is required.

openPrimeR provides both, (1) an evaluation mode to test available primer sets for their coverages on uploaded templates and analyze primer set properties, as well as (2) a design mode, in which *de novo* multiplex primer sets can be designed with full template coverage of a priorly set region of interest (ROI).
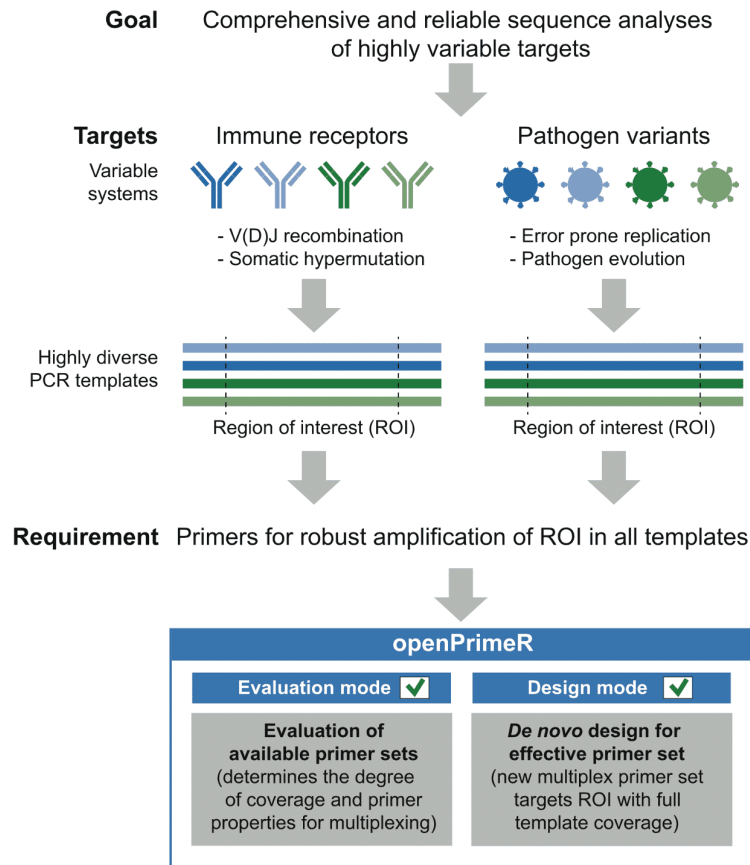
Figure 9: openPrimeR evaluation and design mode for primers targeting highly divers templates
openPrimeR provides both, a reliable evaluation mode for the evaluation of existing primers and a design mode to generate *de novo* primer sets that effectively amplify a ROI from a set of diverse templates, such as immunoglobulin receptors or pathogen variants. ROI= region of interest This figure and figure legend were published in the Journal of Immunological Methods (2020) by Kreer et al. and is used in approval with all co-authors. (Used with courtesy of Creative Commons Attribution- Share Alike 4.0. International License, *https://creativecommons.org/licenses/by-sa/4.0/*)

In both modes, a three stepped workflow is followed: (I.) Template definition: user can choose an immune template from the integrated database or upload individual templates in FASTA or CSV format. An individual region of interest can be defined manually. In evaluation mode, individual primers can be uploaded or be chosen from an integrated selection of published primer sets. (II.) Setting configuration: user can select and define permissible ranges for physicochemical property constraints, coverage conditions, and PCR settings (see **Table 17**) as individually needed. Depending on the selected mode, the final computation step is completed by either presenting an optimized primer set or a comprehensive analysis of the primer set of interest. openPrimeR provides results within comprehensive but clear graphs or tables. Results can be obtained directly in the program or be downloaded as CSV files.

The openPrimeR program was developed by Matthias Döring, M.Sc. in Bioinformatics, at the Max-Planck-Institute for Informatics, Saarbrücken. The experimental part of the program's adjustment, as well as the systematically process of in silico to in vitro and vice versa learning was performed in the laboratory of experimental immunology of Prof. Dr. Florian Klein. All *in vitro* experiments presented in this thesis were performed by Nathalie Lehnen (doctoral candidate), whereas computational optimization of the program was carried out by Matthias Döring and Christoph Kreer. Evaluation of PCR amplification status was performed by all five contributors: Matthias Döring, Seda Meryem Ercanoglu, Lutz Gieselmann, Christoph Kreer and Nathalie Lehnen.

Table 17: openPrimeR design mode options and range settings
* Annealing energy is determined with MELTING[118] and binding events are only counted if their binding energy is below the adjusted threshold.† Amplification efficiency is determined with DECIPHER[121] and binding events are only counted if their efficiency is within the adjusted range. (‡ The logistic regression model[60] was developed during the process of this thesis and was added after the generation of the oPR(5)-IGHV Set. It estimates a false positive rate for each binding event prediction and only events below the adjusted threshhold are counted)
# Free energy (ΔG) is calculated for secondary structures and dimerization events with OligoArrayAux[120] and ViennaRNA[119], respectively. Primers with free energies below the adjusted threshold are considered to violate the constraint. This table was published in the Journal of Immunological Methods (2020) by Kreer et al. and is used in approval with all co-authors. (Used with courtesy of Creative Commons Attribution- Share Alike 4.0. International License, *https://creativecommons.org/licenses/by-sa/4.0/*)

Options	Range Setting	Options	Range Setting
Template		**Constraints**	
Template source	preloaded/personal	Minimal number of individual targets	Off/ 1 to 100
Binding region	complete template	Primer length	Off/ 1 to 50
Primers		GC clamp	Off/ 0 to 10
Template sequence relationship	related/ divergent	Run length	Off/ 0 to 10
Target strand for design	sense/antisense/both	Repeat length	Off/ 0 to 10
Degeneracy	1 to 64	Melting temperature range	Off/ 40 to 90 °C
Coverage conditions		Melting temperature deviation	Off/ 0 to 15 °C
Allowed mismatches	0 to 20	Secondary structure ΔG#	Off/ -10 to 0 kcal/mol
Binding conditions		Specificity	Off/ 0 to 100%
Annealing energy*	Off/-20 to 0 kcal/mol	Self-dimerization ΔG#	Off/ -15 to 0 kcal/mol
Amplification efficiency†	Off/ 0.1- 100%	Cross-dimerization ΔG#	Off/ -15 to 0 kcal/mol
Coverage model‡	Off/ 0 to 100%	**PCR conditions**	
Forbidden number of 3' mismatches	0 (allowed) to 10	Polymerase	*Taq*/non-*Tac*
Codon design		Number of PCR cycles	1 to 100
Introduction of stop codons by mismatches	allowed/ forbidden	Na+ concentration	0 to 100 mM
Amino acid substitution by mismatch binding	allowed/ forbidden	Mg2+ concentration	0 to 100 mM
		Tris concentration	0 to 100 mM
		Primer concentration	0 to 1000 nM
		Template concentration	0.01 to 1000 nM
		Design options	
		Optimization algorithm	Greedy/ ILP
		Target coverage	0 to 100%

3.2.4.2 Optimization of primers in openPrimeR

The following target functions were formulated by C. Kreer and were published by Kreer et al. (Co-author N. Lehnen) in 2020 in the Journal of Immunological Methods.[90] They are used with courtesy of Creative Commons Attribution- Share Alike 4.0. International License, *https://creativecommons.org/licenses/by-sa/4.0/*

In openPrimeR the optimization of primers aims to select a minimal set of primers with the highest coverage ratio while taking into account given design constraints.

Two major selected optimization constraints were defined to design our primer sets. The first constraint is formulated for the maximal allowed difference in melting temperatures *(ΔTm ≥ 0)*. Primers are subdivided in subsets with minimal ranging melting temperatures [T1,T2] with T2 − T1 ≤ ΔTm. The SCO is addressed in each primer subset and the smallest primer set with the best coverage is selected. The second constraint is on the minimal allowed free energy of cross-dimerization *(ΔG$_{cd}^{min}$ ≤ 0 C)*. C. Kreer formulated a binary dimerization matrix to describe the prevention from cross dimerization as followed:

"Given a primer set P = {p$_1$, p$_2$, ..., p|P|} and let ΔG(p$_i$, p$_j$) indicate the free energy of dimerization for primers p$_i$ and p$_j$., the dimerization matrix D ∈ {0,1}|P|×|P| is defined by

$$D_{p_i,p_j} = \begin{cases} 1, \text{ if } \Delta G(p_i, p_j) < \Delta G_{cd}^{min} \\ 0, \text{otherwise} \end{cases}, \forall\ i, j \in \{1, 2,...,|\ P\ |\}$$

such that D$_{pi,pj}$ = 1 indicates that primers p$_i$ and p$_j$ dimerize according to ΔG$_{cd}^{min}$."[90]

Again, the program tries to solve the SCD by integrating a greedy algorithm or an integer linear program (ILP). In the case of applied greedy algorithm, the program selects a compatible primer with maximal coverage in each repetition round. Compatibility is given when for primer p$_i$ in combination to all selected primers p$_j$ Dpi, $_{pj}$ = 0. When a primer p$_i$ is added to the processing set, the primer's coverage events are registered and no longer need to be addressed by the following selection rounds of primers. If full coverage is achieved or no further compatible primer can be detected, the algorithm stops automatically.

The ILP used in the program relies on the branch-and-bound algorithm of lpsolve.[122] If the ILP is applied for primer design the following definition, formulated by C Kreer, is needed: "Let T = {t1, t2, ..., t|T|} indicate the set of templates. The indicator vector x ∈ {0,1}$^{|P|}$ shows whether p$_j$ is included in the optimal set or not by defining it as

$$x_j = \begin{cases} 1, & \text{if primer } p_j \text{ is selected} \\ 0, & \text{otherwise} \end{cases}, \forall j \in \{1, 2,...,| P |\}.$$

The coverage information is sum up in the coverage matrix $C \in \{0,1\}^{|T| \times |P|}$, which is defined by

$$C_{ij} = \begin{cases} 1, & \text{if } p_j \text{ covers } t_i \\ 0, & \text{otherwise} \end{cases}, \forall i \in \{1, 2,...,|T|\}, j \in \{1, 2,...,| P |\}.$$

The ILP is given by

$$\min \sum_{i=1}^{|P|} x_i$$

subject to

$$Cx \geq 1$$

$$x_i + x_j \leq 1 \; \forall \; p_i, p_j \in P \text{ where } D_{p_i, p_j} = 1. \qquad {}_{,, \; 90}$$

The formulated target function includes both, a side constraint to provide the coverage of each template by at least on primer and a constraint to restrict cross-dimerizing, while selecting a minimal set of primers.

3.2.5 Preparation of the IGHV gene library

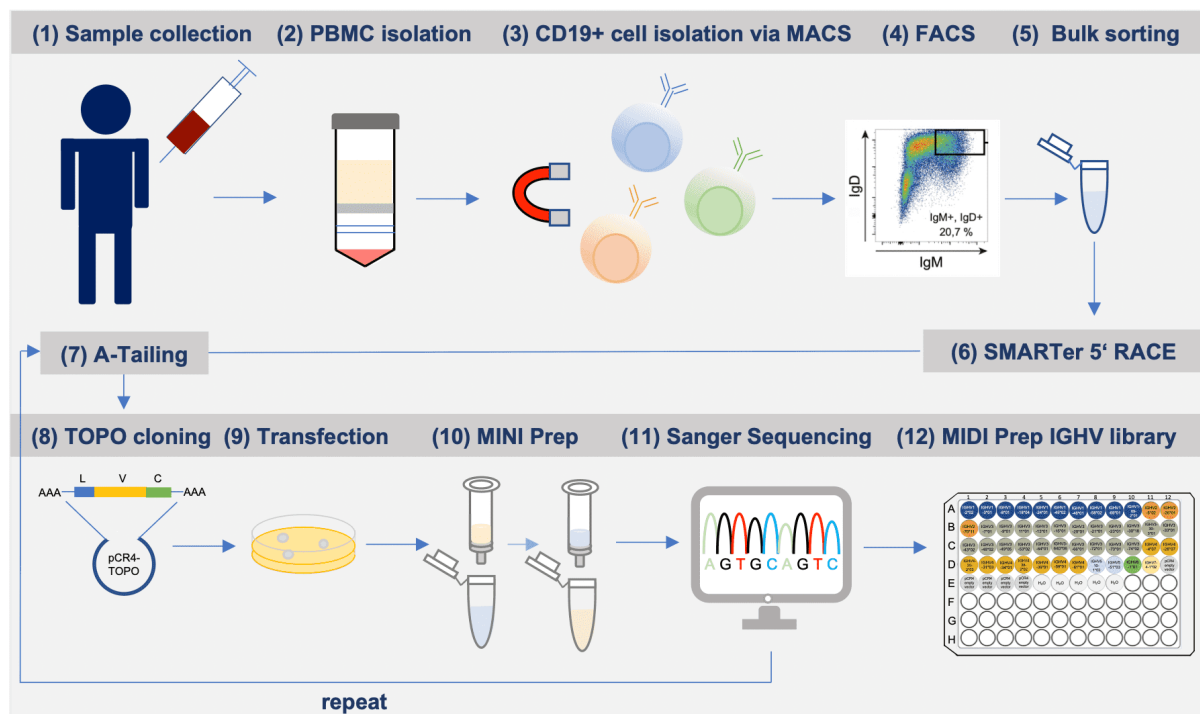### 3.2.5.1 Schematic Workflow of the IGHV gene library preparation

Figure 10: Schematic Workflow of the IGHV gene library preparation
(1) Whole blood donation from healthy individuals from the blood bank of the University hospital Cologne **(2)** Isolation of human peripheral blood mononuclear cells (PBMCs) by density gradient centrifugation **(3)** Isolation of CD19+ B cells via magnetic cell separation (MACS) **(4.+5.)** Bulk Fluorescence cell sorting of CD20+IgM+IgD+IgG- B cells **(6)** Using the Clontech SMARTer 5'RACE kit as an unbiased and fast cDNA transcription and PCR amplification method **(7)** Preparing the template for TOPO Cloning by adding adenine overhangs to the blunt PCR products **(8)** Inserting the PCR products into the pCR4-plasmid vector via TOPO Cloning **(9)** Transformation with competent DH4α E. coli. To check for successful transformation, colony-PCRs were performed on each selected clone (not shown) and it was only proceeded with step 10, if PCR results were positive **(10)** DNA preparation, using the NucleoSpin 96Plasmid kit from Macherey-Nagel to simultaneously perform 96 Mini preps **(11)** Sanger Sequencing outsourced to eurofins and GATC, Sequence annotation via IgBlast, Filtering sequences by fixed constraints **(12)** Completing the IGHV gene library with MIDI Preps of one of the 47 IGHV genes in each well. Repeating step (7) to (11) multiple times to cover at least one of each of the 47 IGHV genes.

3.2.5.2 Fluorescence-activated cell sorting (FACS)– Bulk sort for naive B cells (I)

For the preparation of the IGHV gene library PBMCs were isolated and cell preparation was performed according to the protocols described in section *3.2.1.1*. PBMCs were first positively selected for CD19+ cells by CD19 MicroBeads from MACS Miltenyi Biotec. Equal volumes of patient samples (#32-38 and #39-47) were pooled together. Staining was performed using DAPI, Alexa Fluor 700 mouse anti-human CD20, APC mouse anti-human IgG, human IgD-PE-Cy7, FITC mouse anti-human IgM.

Table 18: Dilutions of antibodies and staining reagents used for bulk sort (I)

Antibody/ stain	Dilution
DAPI	1:100
APC mouse anti-human IgG	1:20
Alexa Fluor 700 mouse anti-human CD20	1:20
FITC mouse anti-human IgM	1:5
hu IgD- PE-Cy7	1:20

$2,5 \times 10^5$ CD20+IgM+IgD+IgG- B cells were sorted. Cells were harvest by centrifugation and pellets were frozen at -80°C.

3.2.5.3 Fluorescence-activated cell sorting (FACS)– Bulk sort for naive B cells (II)

Again, PBMCs were first positively selected for CD19+ cells by CD19 MicroBeads from MACS Miltenyi Biotec. Equal volumes of patient samples (#165-172) were pooled together. Staining was performed using DAPI, Alexa Fluor 700 mouse anti-human CD20, APC mouse anti-human IgG, hu IgD- PE-Cy7, hu CD27-PerCPCy5.5 and FITC mouse anti-human IgM. hu CD27-PerCPCy5.5. CD27, known to be an additional marker on memory B cells, was added to the protocol for negative gaiting, due to a relatively high amount of mutated B cells in the first sort.

Table 19: Dilutions of antibodies and staining reagents used for bulk sort (II)

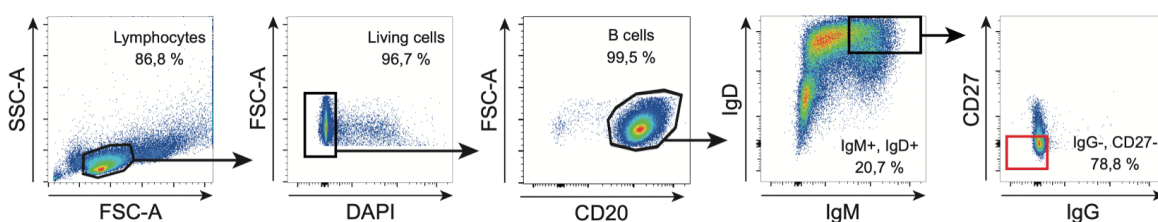| Antibody/ stain	Dilution
DAPI	1:100
APC mouse anti-human IgG	1:20
Alexa Fluor 700 mouse anti-human CD20	1:80
FITC mouse anti-human IgM	1:5
hu IgD- PE-Cy7	1:20
hu CD27-PerCPCy5.5	1:10

Figure 11: Gating strategies for bulk sorting: naive CD20+IgM+IgD+IgG-CD27- B cells

Again, about 2.5 ×10^5 CD20+IgM+IgD+IgG-CD27- B cells were sorted. Cells were harvest by centrifugation and pellets were frozen at -80°C.

3.2.5.4 Automated RNA Isolation on QiaCube with the RNeasy Micro Kit

Bulk sorted pellets were thawed, resuspended in 350 µl RLT lysis buffer and vortexed for 1 minute to homogenize. Then the RNA was purified via QiaCube according to the vendor`s standard protocol but without DNase I. The elution was carried out with 25 µl. Concentration and purity were determined by nano drop One spectral photometer from Thermo Fisher Scientific.

3.2.5.5 Bulk 5' Rapid amplification of cDNA ends (RACE) via SMARTer RACE

The Clonetech SMARTer ®RACE 5'/3'-Kit was used as an unbiased and fast cDNA-transcription combined with a PCR amplification method. cDNA was produced according to the manufacturer's instructions via Clontech. The protocol was performed as provided in the manual for mRNA isolation by QiaCube. Firstly, the buffer mix and the 5'- Race-Ready cDNA synthesis were prepared as follows:

Table 20: 5'-RACE Buffer Mix

Reagent	Volume (µl)
5X First-Strand Buffer	4.0
100 mM DTT	0.5
20 mM dNTPs	1.0
Total	**5.5**

3.2.5.1.1 5'-RACE-Ready cDNA preparation

As in the standard protocol, the oligo-dT reverse primer (5'-CDS Primer A) was used. 10 µl of RNA and 1 µl of 5'-CDS Primer A were combined in a separate microcentrifuge tube. Tubes were incubated at 72°C for 3 minutes, then cooled down to 42°C for 2 minutes. After cooling, the the tubes were spun briefly for 10 seconds at 14.000 x g to collect the contents at the bottom. For 5'-RACE cDNA synthesis reaction(s) 1 µl of the SMARTer II A Oligonucleotide was added per reaction. Master mix for all 5'- RACE-Ready cDNA synthesis reactions was prepared in the following order:

Table 21: 5'-RACE-Ready cDNA synthesis master mix

Reagent	Volume µl
Buffer Mix from Step 1	5.5
RNase Inhibitor (40 U/l)	0.5
SMARTScribe Reverse Transcriptase (100 U)	2
Total Volume	8

8 µl master mix were added to the denatured RNA, resulting in a total volume of 20 µl per cDNA synthesis reaction. Reverse transcription was performed at 42°C for 90 minutes and 70°C for 10 minutes. The first-strand cDNA synthesis product was diluted with the provided Tricine-EDTA Buffer.

3.2.5.1.2 5' RACE

5' RACE of cDNA was performed with the outermost reverse primers to ensure that PCR products contained all (reverse-) primer binding sides for planed primer experiments. For each chain (heavy, kappa and lambda) a different reverse primer was used, whereas the 10 X Universal Primer Mix (UPM) was used as forward primer for all PCR reactions. As reverse primers, we used Ozawa_Cm-RT_rev for IGHV amplification, Outer_Kappa_rev for IGKV amplification and Outer_lambda_1_rev for IGLV amplification, respectively.

Table 22: 5'-RACE cDNA amplification primers

Primer		sequence	Tm
10XU PM Mix	long primer	CTAATACGACTCACTATAGGGCAAGCAGTGGTAT CAACGCAGAGT	
	short primer	CTAATACGACTCACTATAGGGC	
Ozawa_Cm-RT_rev		ATGGAGTCGGGAAGGAAGTC	62°C
Outer_Kappa_rev		GGTGACTTCGCAGGCGTAG	68°C
Outer_lambda_1_rev		GCCGCGTACTTGTTGTTGC	68°C

3.2.5.6 Touch down PCR

From this step on, the whole protocol was repeated several times to obtain a solid amount of colony clones to be analyzed and to cover up (almost) any human germline IGHV gene.

3 Material and Methods

The original protocol provided by the manufacturer was modified by using the Q5 polymerase (instead of the provided SeqAmp DNA polymerase) and the reaction amount was downscaled to 30 µl (instead of 50µl).

Table 23: Q5 PCR Mix

Reagent	Volume (µl)
10x UPM	3
Reverse Primer (10 µM)	1.5
Q5 Mix	15
Template cDNA	1.5
ddH2O	9
sum	30

Table 24: Q5 PCR Touch Down PCR cycler Program

temperature	time	repetition
98°C	30 sec	1x
98°C	10 sec	5x
72°C	30 sec	
98°C	30 sec	5x
70°C	30 sec	
98°C	30 sec	
62°C /68°C*	30 sec	25x
72°C	30 sec	
72°C	5 min	1x
4°C	hold	

* melting temperature 62°C for IgH amplification, 68°C for kappa and lambda

30 µl of each PCR reaction were supplemented with 6 µl DNA gel loading dye (6x) and loaded on a 1% agarose gel together with GeneRuler 1 kb Plus DNA ladder as size indicator. PCR bands were visualized under ultraviolet (UV)-light and carefully cut out.

3.2.5.7 PCR Gel purification and DNA Clean-up

Before cloning, the purification of cut-out PCR DNA gel bands was performed with the GeneJET Gel Extraction and DNA Cleanup Micro Kit from ThermoFisher Scientific according to the manufactures protocol. 11 µl elution buffer was used for the elution step of which 1 µl

served for measurement of concentration via nano drop. Due to experienced low end-concentrations, all remaining 10 µl were used as template for the following A-tailing.

3.2.5.8 A-tailing

3´ Adenine-overhangs were added to the blunt PCR product to prepare the template for TOPO-Cloning.

Table 25: A-tailing Mix

reagent	volume	final conc.
Template	10 µl	varies
10x Buffer	2µl	1x
dATP (10 mM)	0.4 µl	0.2 mM
Qiagen Hotstartaq	0.2 µl	1 Unit (0,05U/µl)
Nuclease free water	7.4 µl	
Sum	20 µl	

Reagents were mixed and incubated for 15 minutes at 95°C and for 20 minutes at 70°C. Afterwards, the reaction was cooled down to 4°C.

3.2.5.9 TOPO-Cloning

The TOPO TA Cloning Kit for sequencing (without competent cells) from ThermoFisher provided an efficient strategy to insert the Taq-polymerase-amplified PCR product into a plasmid vector for further sequencing. The pCR™4-TOPO plasmid vector was used for all TOPO-cloning steps (plasmid map provided in the supplements (**Supplement Figure S7**). For the transformation competent DH5α E. coli cells were used. 4 µl of each A-tailing product was used per chain.

Table 26: TOPO Cloning Mix

Reagent	Volume
A-tailing product	4 µl
pCR4-TOPO Vector	1 µl
Salt	1 µl

Reagents were mixed and incubated at room temperature for 5 minutes. The resulting pCR4 TOPO constructs were transformed into the competent DH5α E. coli by adding 2-4 µl of the TOPO Cloning Mix to 50 µl bacteria. After 30 minutes on ice, the transformation was heat-inactivated at 42°C for 30 seconds. 250 µl SOC-Medium were added and the mixture was incubated one hour at 37°C while shaking at 500 rpm. Afterwards, the bacteria were pelleted

by centrifugation for 2 minutes at 1000 x g and were resuspended in 50 µl SOC-medium. All 50 µl were applied thinly and equally on previously prepared Luria broth (LB) agar plates (with ampicillin) and incubated over night at 37°C.

3.2.5.10 Colony PCR on E. coli colonies transformed with pCR4 vector

Colony PCRs were performed to check whether the transformation succeeded. To this end, M13_forward and M13_reverse primers, binding to the vector backbone, were used.

Table 27: Colony PCR master mix

reagent	per well	final conc.
Nuclease free water	19.75 µl	
10x Green Buffer w/o Mg	2.5 µl	1x
50 mM MgCl$_2$	0.75 µl	1.5 mM
25 mM dNTP mix	0.2 µl	0.2 mM
M13_fwd (50 µM)	0.1 µl	0.2 µM
M13_rev (50 µM)	0.1 µl	0.2 µM
KB Extender	1.5 µl	6%
Platinum Taq	0.1 µl	2 Units
Sum	25 µl	

96-well plates were prepared with 25 µl PCR mix in each well. LB Agar plates with ampicillin were prepared with a grid pattern containing at least 96 boxes. Clones were picked from the overnight culture plate with a sterile fine tip. The grid agar plate was inoculated and then the tip was dipped in the corresponding PCR mix to dispense the remaining bacteria for the PCR reaction.

Table 28: Colony PCR cycler program (Platinum Taq Polymerase)

temperature	time	repetition
94°C	5 min	1x
94°C	30 sec	
45°C	30 sec	25x
72°C	60 sec	
72°C	10 min	1x
4°C	hold	

The M13 primers bind in the backbone, close to each side of the insert, therefore a correct band was expected on the gel at around 830 bp.

Platinum Taq Green Hot Start DNA Polymerase allows a direct gel loading due to the 10x Green Buffer. Samples were loaded on a 2% agarose gel and ran at 150 to 200 volt until the dye line passed approximately 80% down the gel.

3.2.5.11 Culturing and DNA-preparation

Positive clones (detected by a correct band via gel electrophoresis) were collected from the previous grid pattern LB Agar plate. Each clone was cultured in 1.5 ml TB-medium at 37°C over night. Culture plates from the NucleoSpin 96Plasmid kit from Macherey-Nagel were used to culture 96 clones on each plate simultaneously (note that they are not automatically provided with the kid).

For the DNA preparation, the NucleoSpin 96Plasmid kit from Macherey-Nagel was used and the support protocol 'Using NucleoSpin 96 Plasmid with Benchtop centrifuge' was followed as instructed by the manufactures protocol, resulting in highly pure plasmid DNA ready for sequencing.

3.2.5.12 Sanger Sequencing and annotation

Sanger sequencing was outsourced to eurofins Genomics and GATC Biotech. All plasmids were (according to the manufacturers instruction) diluted to 15 µl at a concentration of 50 to100 ng/µl. All resulting sequences were investigated for their overall quality (manually and by phred score) and only high-quality sequencing results (allowing no phred score under 10) were used for further analysis and production.

The annotation software IgBLAST[123] was used to annotate heavy and light chain sequences. Leader regions were aligned and checked for mutations by hand, separately. Leader sequences were detected by counting mostly 57 base pairs (bp) upstream of the annotated V gene, obligatory starting with the base triplet AUG, coding for the amino acid start codon methionine (Met). Subsequently the assigned leader sequences were aligned with the published IMGT database germline leader sequence of the respective V gene allele, using the alignment tool Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/).

3.2.5.13 Source of the germline immunoglobulin sequences template: IMGT the International ImMunoGeneTics information System

The IMGT/LIGM-Database is a large-scale database of immunoglobulin (IG) and T cell receptor (TR) nucleotide sequences from different vertebrate species including humans. The IMGT/LIGM-database provides fully annotated germline and rearranged IG and TR gDNA and cDNA sequences.[33] It served as a global reference for the immunogenetic template in the openPrimer tool. For this project, the database was retrieved in April 2017. 147 IGHV, 64 IGKV, and 36 IGLV templates with complete L + V-region sequences were available at the time of access. The exact template was uploaded at that time in the openPrimeR tool and was used as template for the design of all oPR primer sets 5' RACE NGS data of naive B cells from our laboratory were used to supply the database to include at least one allele for each functional gene. Thereby, 5 IGHV, 2 IGKV, and 11 IGLV leader sequences were retrieved an added (for details see Kreer et al.[90]).

For the IGHV gene library the IMGT/ LIGM-DB was used as reference template. We filtered the database with the following criteria: species: human, chain type: IGHV, IGKV or IGLV, configuration type: rearranged germline, length: complete leader and V-gene sequence, functionality: functional or productive/ ORF

3.2.5.14 Missing genes: Cloning of ordered mini genes into pCR4 vector

For the preparation of the IGHV gene library 10 missing IGHV genes, so-called mini genes, short double stranded fragments, were ordered from eurofins genomics. The sequences were gathered as described above from IMGT/LIGM-database. All mutations were replaced by germline nucleotide sequence. We added 20 nucleotides at the 5' end (=5'UTR) and 112 nucleotides of the constant region at the 3' end to ensure that the chosen reverse primers can bind. Mini genes were ordered for the missing genes of IGHV1-45*02, IGHV3-13*01, IGHV1-58*02, IGHV3-20*01, IGHV3-72*01, IGHV3-73*01, IGHV4-28*07, IGHV5-10-1*03, IGHV4-38-2*02. The dried gene fragments were dissolved in 30 µl nuclease-free water for one hour. They were delivered in a concentration between 574,5 ng and 1262,7 ng per gene fragment. We proceeded with around 150 ng of each gene fragment for the A-tailing reaction and followed the cloning protocol as described (steps 3.2.5.8 - 3.2.5.12).

3.2.5.15 Master plate template

100 ready-to-use master plates were prepared with 47 wells containing one germline IGHV gene respectively, five wells containing the empty pCR4 backbone vector and five wells only containing nuclease free water for control. In each well 5 µl template in a concentration of 0.2 ng/µl (=1 ng per well) were aliquoted and stored in the -20°C freezer until using for PCR.

3.2.6 Preparation of the IGKV and IGLV gene library

To prepare a synthetic gene library for the light chain variable genes, steps 3.2.6.2. to 3.2.6.12. were performed analogously to the protocol for the IGHV gene library. Only in sub-step 3.2.6.5.2., the 5' RACE of cDNA was performed with either the Outer_Kappa_rev primer (GGTGACTTCGCAGGCGTAG, Tm 68°C) for IGKH or the Outer_lambda_1_rev primer (GCCGCGTACTTGTTGTTGC, Tm 68°C) for IGLV instead of the Ozawa_Cm-RT_rev for IGHV. IGKV and IGLV gene libraries were finalized and used for light chain primer evaluation after closure of this MD project (see Kreer et al.[90]).

3.2.7 PCR on IGHV gene library

Three different primer sets were tested independently on the IGHV gene library. One of the primer sets was specifically designed by openPrimeR for the amplification of highly mutated immunoglobulins and will be referred to as oPR(5)-IGHV. It consists of 16 forward primers. Primer Set1, published by Tiller et al. in 2008[81], contains four forward primers. Primer Set2, published by Ippolito et al. in 2012[86] contains eight forward primers. Set1 was originally designed for nested PCR. In our analysis, we only considered 1st PCR primers, assuming that the influence of 2nd PCR primers is negligible as templates that are not covered in the 1st PCR are unlikely to be amplified in 2nd PCR. PCRs were performed with each of the sets on all 47 templates in triplicates. In a separate experiment all single primers from the oPR(5)-IGHV primer set (16 primers) and all single primers from Set1 (four primers) were tested in triplicate measurements on the 47 IGHV gene templates, giving rise to 940 single primer triplicate measurements, later referred to as primer-template-pairs (PTP's). All PCRs were performed with the reverse primer Ippolito_IgM_all_rev.

PCR mastermix (Table 29) for each primer set or single primer was prepared and added on top of each well of the precast IGHV gene library plate. PCR cycles were run according to the Platinum Taq vendors instruction and adapted for individual primer annealing temperatures. The amplification status was analyzed by gel electrophesis on a 2% agarose gel with SYBR safe (1:20.000) loaded with 8 µl PCR product. Results were documented with the BioRAD Gel

3 Material and Methods

DocTM XR+ Imaging system. Bands were expected according to the used forward and reverse primer at around 370-519 bp and encoded with 1= amplified and 0= not amplified.

Table 29: PCR Master Mix on IGHV gene library

reagent	per plate*	per well	final conc.
RNase free water	929.25 µl	14.75 µl	
10x Green Buffer	157.5 µl	2.5 µl	1x
MgCl$_2$ (50 mM)	47.25 µl	0.75 µl	1.5 mM
dNTPs (25 mM)	12.6 µl	0.2 µl	0.2 mM
5' primer (50 µM)	6.3 µl	0.1 µl	0.2 µM
3' primer (50µM)	6.3 µl	0.1 µl	0.2 µM
KB Extender (6%)	94.5 µl	1.5 µl	0.45 %
Platinum Taq	6.3 µl	0.1 µl	2U/ rxn
Sum	1260 µl	20 µl	

*calculated for 63 wells (6 extra wells) if using a multi-channel pipette to avoid shortage in the last wells.

Table 30: Cycler program for PCR on IGHV gene library

temperature	time	repetition
94°C	2 min	1x
94°C	30 sec	
X°C*	30 sec	25x
72°C	55 sec	
72°C	10 min	1x
4°C	hold	

*annealing temperature depending on used primer set (55°C for oPR(5)-IGHV primer set, 57°C for Set1 and 63°C for Set2).

3.2.8 Evaluation of primer-template-pairs (PTPs) from novel *Tac* PCR data set

The amplification results on the IGHV gene library of all single primer PCRs from Set1 and the oPR(5)-IGHV set were collected to generate a novel *Tac* PCR data set. The evaluation of the primer-template-pairs was performed by M. Döring. The following evaluation method and terms were published in Scientific Reports 2019 by Döring et al[60] and are used with courtesy of Creative Commons Attribution- Share Alike 4.0. International License, *https://creativecommons.org/licenses/by-sa/4.0/*.

All considered primer-template-pairs (PTPs) $i \in \mathbb{N}$ were labeled $y_i \in \{Amplified, Unamplified\}$ based on their result analyzed via gel electrophoresis by five independent investigators. The investigators classified the amplification status of each PTP as either *Amplified* or *Unamplified* by inspecting the respective gels for visible bands at correct heights. The label of PTP i according to reviewer $j \in \{1, ..., 5\}$ was labeled *Amplified* if at least two out of three measurements were evaluated by the respective reviewer as *Amplified* ($y_{i,j}$ = *Amplified*). The number of times a respective PTP *I* was labeled as Amplified or Unamplified is indicated by the terms $n_{i,A} = |\{y_{i,j}|y_{i,j} = Amplified\}|$ and $n_{i,U} = |\{y_{i,j}|y_{i,j} = Unamplified\}|$. The term

$$y_i = \begin{cases} Amplified, \text{ if } n_{i,A} > n_{i,U} \\ Unamplified, \text{ otherwise} \end{cases}$$

ensures that PTPs were only labeled as *Amplified* if at least three out of five reviewers had labeled the PTP as *Amplified*.

The binding region of a primer was assumed with the minimum of mismatches. PTPs showing more than 12 mismatches were excluded, as an exact binding side could hardly be identified. Thereby, the data set was reduced to 908 (from 940) PTPs.

3.2.8.1 Evaluation of primer properties

The following properties were computed by openPrimeR to provide underlying characteristics of primers and PTPs within the data set. Selected properties of interest were the primer length, extent of GC clamp, GC ratio, melting temperature, free energy of secondary structure, self-dimerization and the position of primer-template-mismatch. The external tools MELTING[118], ViennaRNA[119] and OligoArrayAux[120] were used to calculate the ranging melting temperature or detect the appearance of secondary structures and self-dimerization.

3.2.8.2 Encoding features of the *Tac* data set

To investigate the impact of primer-template mismatches within the terminal 3' hexamer M. Döring implemented several encodings for this region as shown in Figure 12. M. Döring additionally formulated the mismatch vector and terms defining the 3' hexamer as followed. This method (including the coding system, mismatch vector and defining terms) was published in Scientific Reports 2019 by Döring et al[60] and is used with courtesy of Creative Commons Attribution- Share Alike 4.0. International License, *https://creativecommons.org/licenses/by-sa/4.0/*.
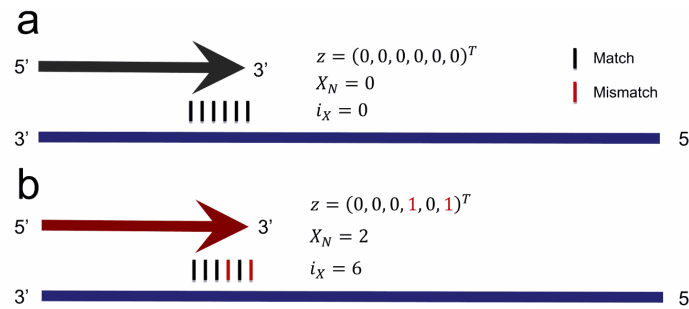
Figure 12: Examples for encoding mismatches within the 3'hexamer region.
Primers are indicated as arrows and templates are indicated as horizontal bars in blue. Arrowheads indicate the 3' hexamer region. Mismatches within the 3' hexamer are encoded via $z \in \{0,1\}^6$, $X_N \in \mathbb{N}_0$, and $i_X \in \{0,1,...,6\}$. While z uses a binary encoding to indicate the presence of mismatches within the 3' hexamer, X_N gives the total number of 3' hexamer mismatches, and i_X indicates the position of the 3' hexamer mismatch closest to the 3' terminus. (a) Absence of 3' terminal mismatches between primer and template. (b) Mismatches in the 3' hexamer at positions 4 and 6. (This figure and figure legend were published in Scientific Reports 2019 by Döring et al[60] and is used in approval with all co-authors and with courtesy of Creative Commons Attribution- Share Alike 4.0. International License, *https://creativecommons.org/licenses/by-sa/4.0/*)

To code the event of mismatches at any given position *(j)* in the 3' hexamer a binary encoding was used to formulate the mismatch feature vector $z \in \{0, 1\}^6$:

$$z_j = \begin{cases} 1, \text{ if there is a mismatch at position } j \text{ in the 3' hexamer} \\ 0, \text{ otherwise} \end{cases}.$$

The 3' hexamer position is identified by $j \in \{1, 2, ... 6\}$. For example, j=1 indicates the first position in the 3' hexamer, j=2 indicates the 2nd position and so on. To reflect the increasing effect of co-occurring mismatches in the 3' hexamer, the total number of mismatches is encoded as $X_N = \sum_j z_j$. As mismatch positions closest to the 3' terminus interfere with the primer efficacy to a greater extent,[66,124,125] an additional term to describe the 3' hexamer mismatch closest to the 3' terminus was encoded by setting

$$i_x = \begin{cases} \max_{j \in \{1,...,6\}} \{j \mid z_j = 1\}, \text{ if } X_N \neq 0 \\ 0, \text{ otherwise} \end{cases}.$$

69

Here, for example, $i_X = 0$ encodes for a primer without mismatch binding. A primer exhibiting mismatches at position 3 and 5 is encoded with $i_X = 5$.

3.2.9 Quantification and statistical analysis

Graphpad Prism (version 7.0b) was used for all statistical analysis. Unpaired t tests or two-sided Wilcoxon rank-sum test were used to calculate respective p-values. The significance level was set at 0.05.

4 RESULTS

4.1 Optimizing single B cell cloning strategies (Immunoglobulin gene amplification)

To study the B cell repertoire on single cell level and to identify neutralizing antibodies against pathogens, the quality and yield of the amplified variable heavy and light chains is crucial. To establish an optimized and standardized PCR protocol for the amplification of human B cells on single cell level, a multitude of PCR tests, focusing on different influencing factors, was run. Protocol settings were tested with different primer sets (former published sets and *de novo* designed primer sets by openPrimeR). For the following PCR amplification tests, freshly reverse transcribed cDNA from antigen-experienced single B cells was used.

4.1.1 Optimizing the amplification protocol

4.1.1.1 Analyzing the role of 1st and 2nd PCR on cDNA enrichment on single cell level

In immunoglobulin gene amplification on single cell level, nested or semi-nested PCR protocols are usually used to enrich the amplicon and to improve specificity and sensitivity of the respective reaction.
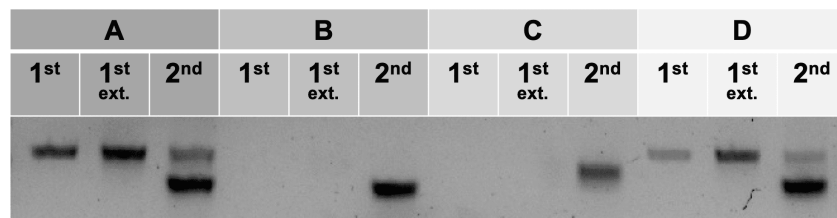
*1st ext.= twice repeated 1st PCR protocol

Figure 13: The role of 1st and 2nd PCR in cDNA enrichment on single cell level
Visualization via gel electrophoresis of 1st, 1st extended* and 2nd PCR products of PCR amplification of four different antigen-experienced single B cells (A-D). Primers used for semi-nested PCR: oPR(1)-IGHV set as forward primer set and Tiller_2008_3´CµCH1 in 1st and Tiller_2008_3´IgG _internal in 2nd PCR as reverse primers.

Figure 13 demonstrates the high impact of the 2nd PCR for the enrichment and thus the visualization and detection of amplified cDNA on single cell level. Two out of four V genes could only be visualized after 2nd PCR. An additional (extended) 1st PCR does not provide an amplification advantage, even if in (A) and (D) the signal in the 1st extended product seems to be slightly enriched compared to 1st alone.

4.1.1.2 Down scaling the reaction volume of the single cell amplification protocol

PCR amplification can be extremely cost-intensive with reagents accounting for the majority of costs. We therefore investigated on the possibility of downscaling the PCR reagent volume.
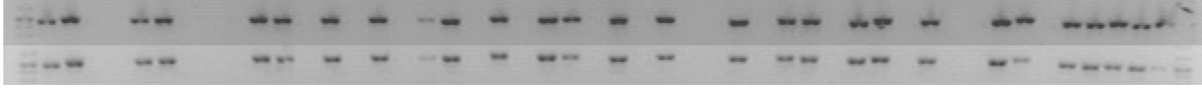

Figure 14: Evaluation of downscaling the reaction volume of PCR on single cell level (I.) Single cell IGHV amplification of 48 antigen-experienced single B cells in a nested PCR using primer Set1 with two different protocol settings: upper row shows the amplification status visualized via gel electrophoresis of the standard amplification protocol using 38 µl PCR-mix for 1st and 2nd PCR, lower row shows the amplification status using 38 µl PCR-mix for 1st PCR and 19 µl for 2nd PCR. First and last column contain DNA ladder.

Figure 14 demonstrates that the 2nd PCR volume could be reduced by 50% (from 38 µl to 19 µl) without losing amplification rate or visualization quality. The reduction of the 1st and 2nd PCR reagent volume by 50% on the other hand shows a significant decrease of the amplification status with only 8.33% (4 out of 48) positive amplifications in comparison to 47.92% (23 out of 48) as shown in Figure 15.

Figure 15: Evaluation of downscaling the reaction volume of PCR on single cell level (II.) Single cell IGHV amplification of 48 antigen-experienced single B cells in a nested PCR using primer Set1 with two different protocol settings: upper row shows the amplification status visualized via gel electrophoresis of the standard protocol using 38 µl PCR-mix for 1st and 2nd PCR, lower row shows the amplification status using 19 µl PCR-mix for 1st PCR and 2nd PCR. First and last column contain DNA ladder.

In conclusion, we adapted our standard single cell amplification protocol successfully by downscaling the 2nd PCR reagent volume to 19 µl.

4.1.1.3 Testing storage conditions

The single cell amplification protocol is based on a two-step PCR amplification. As not all 2nd PCRs can be run immediately after 1st PCRs, we investigated on how to store PCR plates after amplification.
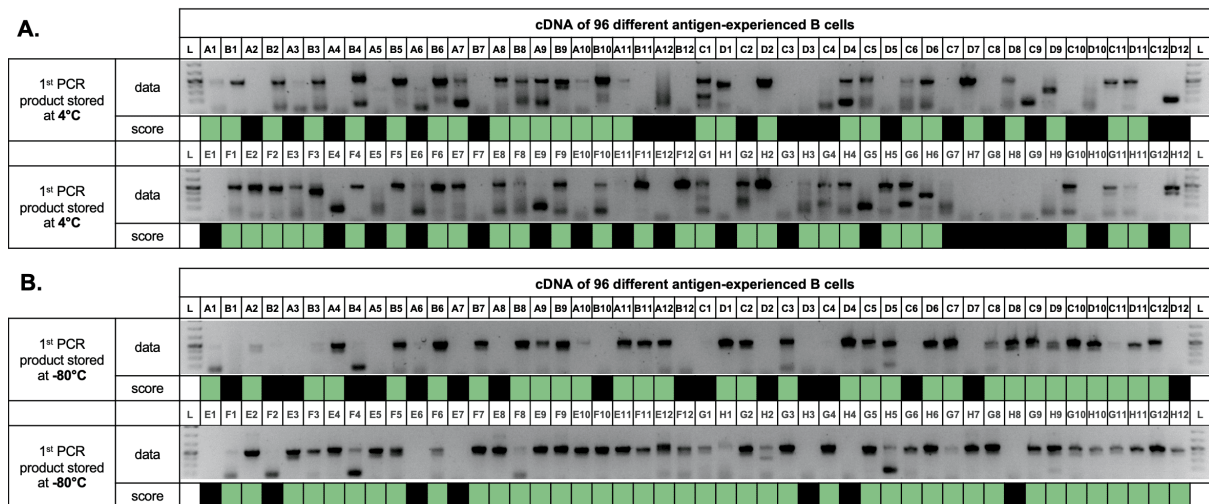
A.

B.

Figure 16: The storage temperature-depending quality and amplification status of 2nd PCR product

96-well plate single cell amplification of antigen-experienced B cells (using primer Set1 in semi-nested PCR) with A. 1st PCR product stored at 4°C and B. 1st PCR product stored at -80°C for 4 days. Scoring indicates "amplified" in green and "unamplified" in black. First and last lane contain DNA ladder.

Figure 16 shows a clear difference in quality and amplification status of 2nd PCR visualization via gel electrophoresis between storage of 1st PCR product at 4°C (A) and -80°C (B) for a period of 4 days. Whereas the amplification rate of the 96 antigen-experienced single B cells, amplified with primer Set1 in a semi-nested PCR, shows a 57.29% coverage (55 out of 96) when 1st PCR product was stored at 4°C, the identical 96 antigen-experienced single B cells (well A1-H12) amplified with the same primer set and protocol show a 79.17% coverage (76 out of 96) when stored at -80°C. Additionally, storing at 4°C leads to a higher amount of smear bands and unspecific amplification, visualized by bands at incorrect heights. This finding might implicate that storage at higher temperatures leads to unspecific reactions after finishing the cycler program. The lower amplification rate might thus be explained due to shortage in reagents, which were consumed during unspecific amplification processes. Therefore, we stored all 1st PCR products at -80°C or preferably proceeded directly with 2nd PCR amplification.

It is of note, that even if the overall amplification rate is higher in the -80°C storage plate, there are also 14 B cells which were only amplified in the 4°C storage plate, indicating that the PCR amplification on single cell level (even if performed with the same PCR protocol and primers) depends on multiple factors and is very susceptible to errors.

4.1.1.4 Comparison of amplification efficacies on single antigen-experienced B cells in dependence of different commercially available polymerases

The catalyzing enzyme polymerase is one of the key reagents in a PCR, especially on immunological templates, because the amplification status on the one hand and the potentially introduced bias on the other hand are highly dependent on the type and quality of polymerase. The enzyme is also one of the most cost intensive factors in a PCR. Thus, we tested the performance of three commercially available DNA polymerases on single human antigen-experienced B cells.

CD20+IgM-IgG+ mature B cells were sorted from PBMCs of healthy human doners. PCRs were performed according to the respective manufacture's manuals. Primer Set1 with its referring reverse primers was used in all amplifications. 192 PCRs on cDNA single cell level were performed using PolymeraseA (HotStarTaq, Qiagen), 144 reactions using PolymeraseB (Taq Polymerase, Qiagen) and 168 reactions using PolymeraseC (PlatinumTaq, Thermo Fisher Scientific). PCR results were classified as either "amplified" or "not amplified" according to a positive band at the correct height.
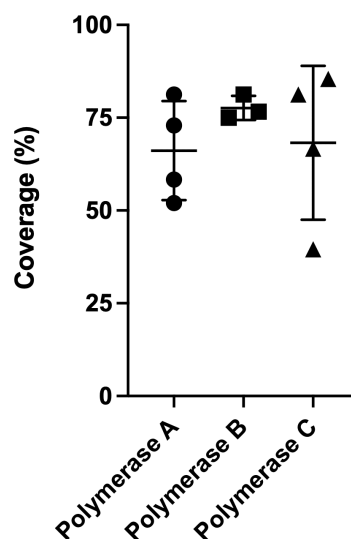

Figure 17: Comparison of amplification efficacies on single antigen-experienced B cells in dependence of three different commercially available polymerases
Each dot/ square/ triangle represents the amplification coverage of a complete 96-well plate, containing one single B cell in each well. Data are given as mean ± SD. SD = standard deviation

Comparing the mean coverage results of the three different polymerases in amplifying single human B cells, we found no significant difference between the three. We therefore decided to implement the less cost intense polymerase in our standard single cell amplification protocol, which in our case turned out to be Polymerase C.

Since this experiment only investigated on the amplification status via gel electrophoresis, no statement can be made on amplification biases or induced mutations by the different polymerases.

4.1.1.5 Testing efficacy and reproducibility of immunoglobulin gene amplification on single cell level in dependence of the used reverse transcriptase for cDNA production

To investigate how reliably a positive or negative PCR result on single cell level can be reproduced with our amplification protocol, the following PCR amplifications were performed in duplicates (with identical conditions on the same single B cell). The duplicate status was either set to amplified/amplified (green), unamplified/unamplified (yellow) or amplified/unamplified (grey), according to their visualization via gel electrophoresis (see Figure 19). cDNA of the 48-well single B cell plates A to F2 was produced by using the reverse transcriptase (RT) SuperScript III and plates G1 to I2 using RT SuperScript IV.

A) PCR conditions for each 48-well approach

	A	B	C	D1	D2	E1	E2	F1	F2	G1	G2	H1	H2	I1	I2
plate	#1			#2		#2		#3		#4		#5		#5	
rows	A-D	A-D	E-H	A-D	E-H	A-D	E-H	A-D	E-H	A-D	E-H	A-D	E-H	A-D	E-H
primer	Set1	oPR(5)	Set1	Set1	Set1	Set1	Set1	Set1	Set1	Set1	Set1	Set1	Set1	Set1	Set1
cDNA	4 µl	4 µl	4 µl	2 µl	2 µl	6 µl	6 µl	5,5, µl	5,5 µl	5,5 µl	5,5 µl	5,5 µl	5,5 µl	5,5 µl	5,5 µl
RT	Single cell reverse transcription with SuperScript III									Single cell reverse transcription with SuperScript IV					

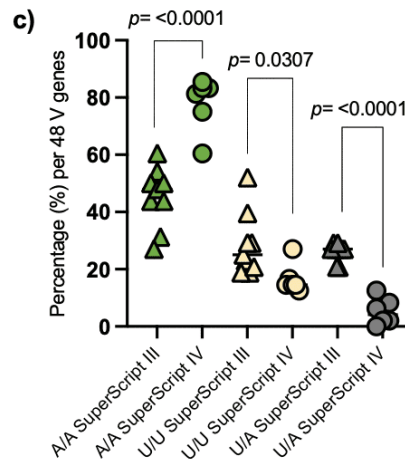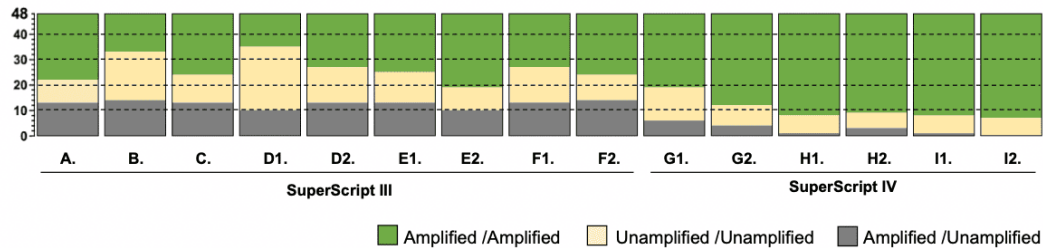**B) Amplification status of 48-well single B cell plates in duplicates**

Figure 18: Testing efficacy and reproducibility of immunoglobulin gene amplification on single cell level in dependence of the used reverse transcriptase for cDNA production
A) PCR conditions of each of the 15 (A-I2) 48-well plate PCR approaches. (RT= reverse transcriptase), **B)** Amplification status of the 15 (A-I2) 48-well single B cell plates in duplicates. Amplified/Amplified (green) for example indicates that both corresponding duplicate PCRs were set to amplified according to their visualization via gel electrophoresis. **C)** Percentage of Amplified/Amplified, Unamplified/Unamplified and Amplified/Unamplified status of each 48-well single B cell plates in dependance of the used reverse transcriptase (SuperScript III vs. IV). Each triangle/ dot represents the status percentage of one 48-well single B cell plate. A/A indicates the duplicate amplification status as Amplified/Amplified, U/U indicates Unamplified/Unamplified, and U/A indicates Unamplified/Amplified or vice versa. *P* values were calculated using unpaired t test

This experiment showed impressively that PCR amplification at the single cell level is subject to significant variations even under standardized conditions. In PCR approach A to F2, more than 20% (more than 10 out of 48) of the duplicates of each 48-well plate presented a lack of reproducibility of a positive or negative result. In contrast, in PCR approach G1 to I2 on average only 5.21 % of positive or negative results could not be reproduced in the respective duplicate. These two groups (experiments A to F2 vs. G1 to I2) mainly differed by the usage of RT for cDNA production from the respective single B cells. Figure 18C demonstrates a significant difference between percentage of failed reproducibility (grey triangles and dots) of PCR approaches on cDNA produced with SuperScript III and IV. Additionally, the percentage of a

reproduced amplification status is significantly higher in the SuperScript IV approaches (p=<0.0001), indicating a better overall amplification status on single cell level by using cDNA reverse transcribed by SuperScript IV. Therefore, we adapted our standard protocol of single cell reverse transcription by exchanging the RT SuperScript III by SuperScript IV.

Figure 19: Corresponding gel electrophoresis pictures of PCR duplicates on single B cells (48-well plates)

4 Results

4.1.2 Primer design and primer testing for an optimized B cell cloning strategy

4.1.2.1 Amplification efficacy of primer Set1, primer Set2 and oPR(1)-IGHV set on antigen-experienced single B cells from healthy donors

To investigate and compare the amplification efficacies of established primer sets used for the amplification of B cell receptors and the first *de novo* primer set designed by openPrimeR (oPR(1)-IGHV), the sets were tested on 96 different antigen-experienced single B cells (A1-H12) in direct comparison.
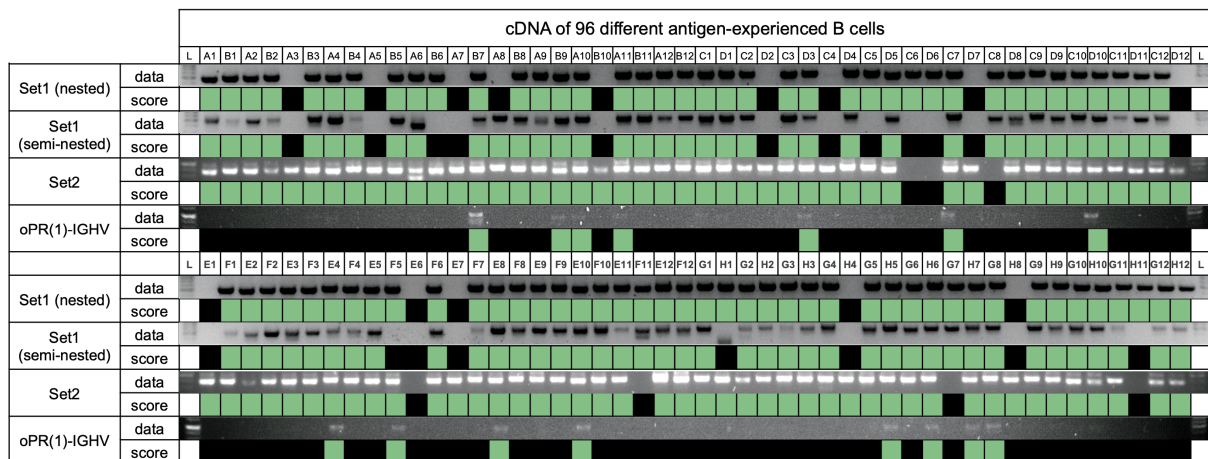

Figure 20: Visualization of PCR products from cDNA of 96 different antigen-experienced single B cells from healthy donors using Set1, Set2 and oPR(1)-IGHV
96 different antigen-experienced single B cells (A1-H12) were amplified using each of the following primer sets. Set1 was applied in both, a nested and a semi-nested PCR, Set2 and oPR(1)-IGHV were applied in semi-nested PCRs. PCR results were scored as either "amplified" (green) or "unamplified" (black) according to a positive (visual band at the correct height ~ 370-519 bp) or negative result. First and last lane contain DNA ladder (white)

Table 31: Primer overview

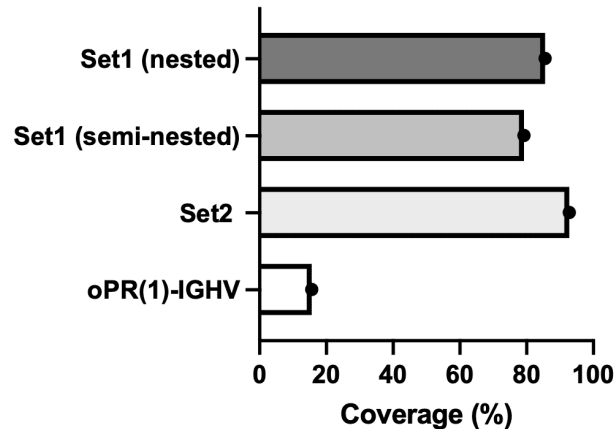|	forward primer mix 1st PCR	revers primer 1st PCR	forward primer mix 2nd PCR	revers primer 2nd PCR
A	Set1 (5' L-VH Mix)	3´ Cµ CH1	Set1 (5' AgeI VH Mix)	3´ IgG _internal
B	Set1 (5' L-VH Mix)	3´ Cµ CH1	Set1 (5' L-VH Mix)	3´ IgG _internal
C	Set2	Ozawa_Cg-RT_rev	Set2	Ippolito_IgG-all-rev
D	oPR(1)-IGHV	Ozawa_Cg-1st_rev	oPR(1)-IGHV	3´ IgG _internal

Figure 21: Coverage (%) of primer Set1 (nested and semi-nested PCR), Set2 and oPR(1)-IGHV set on 96 antigen experienced single B cells
96 different antigen-experienced single B cells (A1-H12) were amplified using each of the following primer sets. Set1 was applied in both, a nested and a semi-nested PCR, Set2 and oPR(1)-IGHV were applied in semi-nested PCRs. For respective gel pictures see Figure 20, for used primer mixes see Table 31.

Comparing the amplification status of 96 different antigen-experienced single B cells, Set2 showed the highest amplification rate at 92.78%. Set1 showed an 85.42% amplification rate applied in nested PCR and a 79.17% amplification rate applied in semi-nested PCR. In this case, changing PCR protocol from nested to semi-nested PCR led to more unspecific primer binding and unwanted PCR products (visualized on full gel electrophoresis pictures, not shown in the here presented shortened pictures). The amplification rate achieved with the oPR(1)-IGHV set is comparatively low at 15.63%.

Almost all 96 presented IGHV-genes were covered by either one of the three primer sets. Only one of the 96 genes could not be amplified at all (E6). Probably E6 exhibits a higher number of mutations in the primer binding regions and therefor amplification was compromised. Both Set1 and Set2 show a relatively high coverage but fail to cover all of the presented antigen-experienced IGHV-genes. Thus, none of the primer sets presented was a promising candidate for application to highly mutant templates. The oPR(1)-IGHV set showed very poor performance. In conclusion, primer design settings in the openPrimeR tool needed to be re-evaluated.

4.1.2.2 Overall coverage of the *de novo* primer sets oPR(4)-IGHV, oPR(5)-IGHV and oPR(6)-IGHV on antigen-experienced single B cells from healthy donors

Here we present three newly designed primer sets designed by the openPrimeR tool. For all sets, the primer binding region was limited to the leader region and primers binding further upstream in the V gene were strictly excluded. In addition, a maximum coverage of all annotated human germline IGHV genes, including their varying gene alleles, was intended to be achieved under the maximum formation of one allowed mismatch.

The oPR(4)-IGHV was designed with a low stringency constraints mode. The oPR(5)-IGHV was designed with a minimum secondary structure constraint while the oPR(6)-IGHV represents a "best-of-5" set, consisting of the 5 best primers (those covering most IGHV genes simultaneously) of a precursor set (not introduced here). Table 32 shows the individual primer set properties in comparison.

Table 32: Properties of the primer sets oPR(4)-IGHV, oPR(5)-IGHV and oPR(6)-IGHV
Numbers of primer binding region and binding position start and end refer to nucleotide position of leader plus V gene region, starting with 1 as first nucleotide of the leader region. Values shown in brackets indicate the range of the observed values, front numbers show the mean values of all primers of the set (MM= mismatch).

Primer Property	oPR(4)-IGHV	oPR(5)-IGHV	oPR(6)-IGHV
Allowed primer binding region	1 -47 (leader region only)	1 -47 (leader region only)	1 -47 (leader region only)
Allowed mismatches per primer	max. 1	max. 1	max. 1
Number of primers	11	16	5
Length	20.4 [18, 25]	21.8 [18, 25]	22.2 [20, 25]
Binding position start*	[2, 25]	[1, 12]	[1, 14]
Binding position end*	[19, 42]	[19, 33]	[20, 37]
GC clamp ratio*	0.59 [0.44, 0.72]	0.58 [0.5, 0.67]	0.57 [0.52, 0.6]
Melting temperature (Tm)	59.1 [57.1, 61.8]	59.5 [58.3, 62]	59.2 [57.6, 60.3]
Covered IGHV templates per primer (1MM)*	17.3 [1, 47]	16.13 [1, 44]	32.4 [18, 46]
Calculated IGHV coverage (1MM)*	147 of 155 (94.84%)	144 of 155 (92.90%)	133 of 155 (85.8%)

*computations and calculations were performed with the unfinished openPrimeR version in 11/2016

The three *de novo* primer sets and the well-established primer Set1 were each tested on 48 antigen-experienced B cells to compare their individual single cell amplification efficacy.
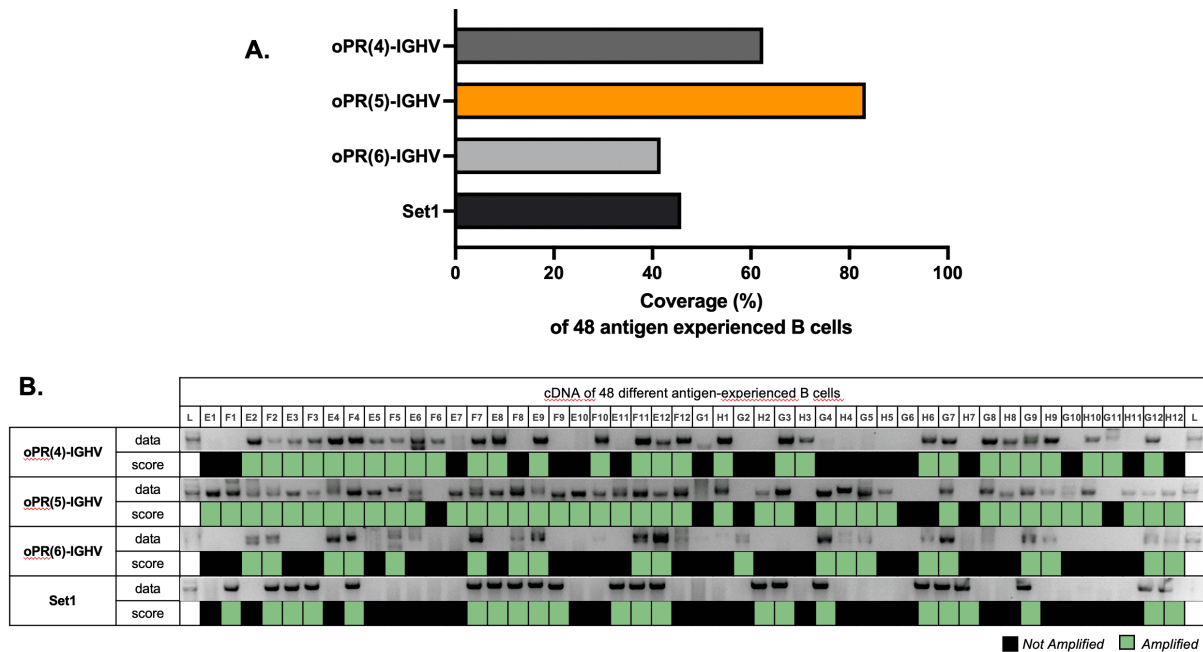

Figure 22: Coverage (%) of primer sets oPR(4)-IGHV, oPR(5)-IGHV, oPR(6)-IGHV and Set1 on 48 antigen-experienced single B cells A) Coverage (%) of oPR(4)-IGHV, oPR(5)-IGHV, oPR(6)-IGHV and Set1. **B)** Visualization via gel electrophoresis and scoring of amplification status of each of the 48 antigen-experienced single B cells (well E1-H12) by the respective primer set. Green indicates status as *Amplified*, black indicates status as *Not amplified.* First and last lane contain DNA ladder

Figure 22A shows the overall coverage of the respective primer set, visualized via graph bars. The oPR(5)-IGHV showed superior performance over the other sets with an overall coverage of 83.33%, followed by the oPR(4)-IGHV with an overall coverage of 62.5%. Lower coverages were detected with the oPR(6)-IGHV (41.66%) and Set1 (45.83%). The amplification status (*Amplified or Not Amplified*) of each of the 48 IGHV genes by the respective primer set are shown in Figure 22B. The oPR(5)-IGHV showed positive amplification of six IGHV genes which could not be detected by any of the other sets (E1, E7, E10, G10, H5, H11), while oPR(4)-IGHV covered 3 genes not amplified by the others (F6, H3 and G11) and oPR(6)-IGHV and Set1 covered each only one gene, not covered by the other sets (G2 and H7). Two single cell IGHV genes could not be amplified by any of the tested primer sets (G1 and G6).

In conclusion, we selected the oPR(5)-IGHV primer set as promising candidate for the following amplification experiments and optimizing processes.

4.1.2.3 Comparing amplification efficacy of primer Set1 and oPR(5)-IGHV on antigen-experienced single B cells from healthy donors and on highly mutated HIV1-specific neutralizing antibodies

Single-cell sorting of CD20+IgM-IgG+ single B cells from healthy donors was performed using the previous standard protocol (see 0). To potentially identify antibodies that account for potent neutralizing antibodies against HIV-1, single-cell sorting of Env-reactive B cells that bound to native-like BG505$_{SOSIP.664}$ was performed. BG505$_{SOSIP.664}$ binds to most of the known HIV-1-specific neutralizing antibodies but generally does not bind to antibodies that lack neutralizing activity.[126,127] These antibodies are known to exhibit a high rate of mutations.

To investigate the amplification efficacy of primer Set1 and the oPR(5)-IGHV on diverse templates, a multitude of 96-well plates containing cDNA of single B cells (either antigen-experienced single B cell from healthy donors or B cells bearing a HIV-1-specific antibody) were sorted and reverse transcribed. In total, Set1 was tested on 23 different 96-well plates (17 plates comprising antigen-experienced B cells and 6 plates comprising SOSIP-GFP positive B cells). oPR(5)-IGHV was tested on 14 different 96-well plates (8 plates comprising antigen-experienced B cells and 6 plates comprising SOSIP-GFP positive B cells), leading up to 3.552 PCRs to be evaluated for either *Amplified* or *Not Amplified* status.
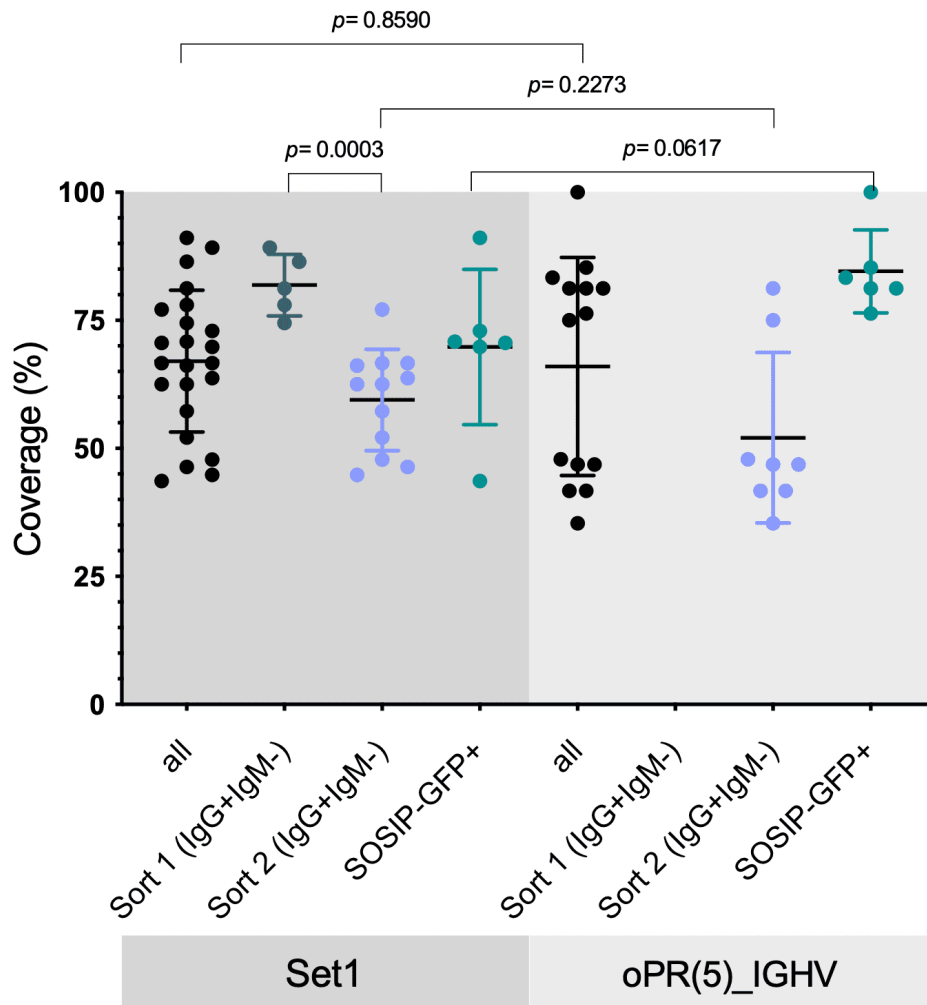
Figure 23: Analysis of PCR coverage of Set1 and oPR(5)-IGHV on antigen-experienced single B cells from healthy donors of two different single cell sorts and on highly mutated HIV-1-specific single B cells (stained with SOSIP-GFP)
Single cell Sort 1 and Sort 2 were performed with the exact same protocol for antigen-experienced single B cell sorting (CD20+IgM-IgG+). SOSIP-GFP single B cells are Env-reactive B cells that bound to native-like BG505$_{SOSIP.664}$. Each dot represents a whole 96-well plate amplified with the respective primer set (Set1 vs. oPR(5)-IGHV). Amplification coverage was evaluated via gel electrophoresis. P values were calculated by unpaired t-test.

Figure 23 demonstrates that no significant difference could be detected in the overall coverage of single human B cells (from healthy donors and HIV-1-specific B cells combined) between primer Set1 and oPR(5)-IGHV (p=0.8590). Likewise, there is no significant difference between the coverage of Set1 and the oPR(5)-IGHV on antigen-experienced B cells from Sort2 (p=0.2273) or SOSIP-GFP single B cells (p=0.0617) only. But, surprisingly, comparing the coverage of primer Set1 on antigen-experienced single B cell plates from Sort1 and Sort2, a significant difference in amplification rate between both sorts is detectable (p= 0.0003). This finding underlines the poor validity of primer comparison on different single B cells (mainly caused by an unknown sequence divergence of the BCR) and implies the need to develop an

instrument to perform unbiased primer tests in the context of immunological templates. Additionally, sorting quality needed to be re-evaluated.

4.2 Generating a synthetic immunoglobulin (heavy chain) gene library for unbiased primer testing and evaluation of the openPrimeR design tool

To address the problem of poor comparability of (primer) experiments on single cell level, caused by an unknown amount of sequence divergence, varying cDNA quality and a limited amount of cDNA and therefore limited test rounds, we developed the idea of generating a synthetic standard template-, a so-called immunoglobulin gene library. The idea behind the IGHV gene library was to generate a well-defined cDNA sequence template including one defined germline allele of all 53 functional IGHV gene segments.

The plasmid library was designed to i.) allow unbiased primer testing and evaluation of the openPrimeR design function, ii.) compare the in silico predicted coverage of different primer sets in the openPrimeR tool with the coverage measured *in vitro*, and iii.) examine which thermodynamic properties of the primer-template pairs (PTPs) influence the PCR amplification status. The synthetic germline IGHV library fully matches the sequence data uploaded in the openPrimeR program. Information gained by PCR experiments on the IGHV gene library was subsequently refed into the openPrimeR tool for further program optimization.

4.2.1 The IGHV gene library

The IMGT/LIGM-database served as sequence reference for the IGHV gene library. Data was retrieved in April 2017. The database was filtered with the following criteria: species: human, chain type: IGHV, configuration type: rearranged germline, length: complete leader and V-gene sequence, functionality: functional or productive/ ORF. This resulted in a total amount of 53 functional IGHV genes, which we intended to be represented in the gene library.

Upfront, we were able to cut down the 53 genes to 47 genes, due to complete or very high (at least 97.9%) sequence identity between certain V genes, including their corresponding leader sequence. Therefore, six genes were identified (namely IGHV1-69D, IGHV2-70D, IGHV3-23D, IGHV3-30-5, IGHV3-43D and IGHV4-30-4), whose sequences were represented within another gene and could be cut without loss of additional information in the context of our intended experiments (see Supplement Figure S1- Figure **S6**).

4 Results

The final library contained 10 IGHV1, three IGHV2, 21 IGHV3, nine IGHV4, two IGHV5 genes and one IGHV6 and IGHV7 gene, resulting in 47 different IGHV gene templates. Each well was prepared to contain 1 ng of template. Additionally, five wells containing the empty pCR4 vector backbone were included to rule out backbone binding as well as five wells with H_2O only as negative controls. Figure 24 shows a schematic example sheet of the final IGHV gene library.
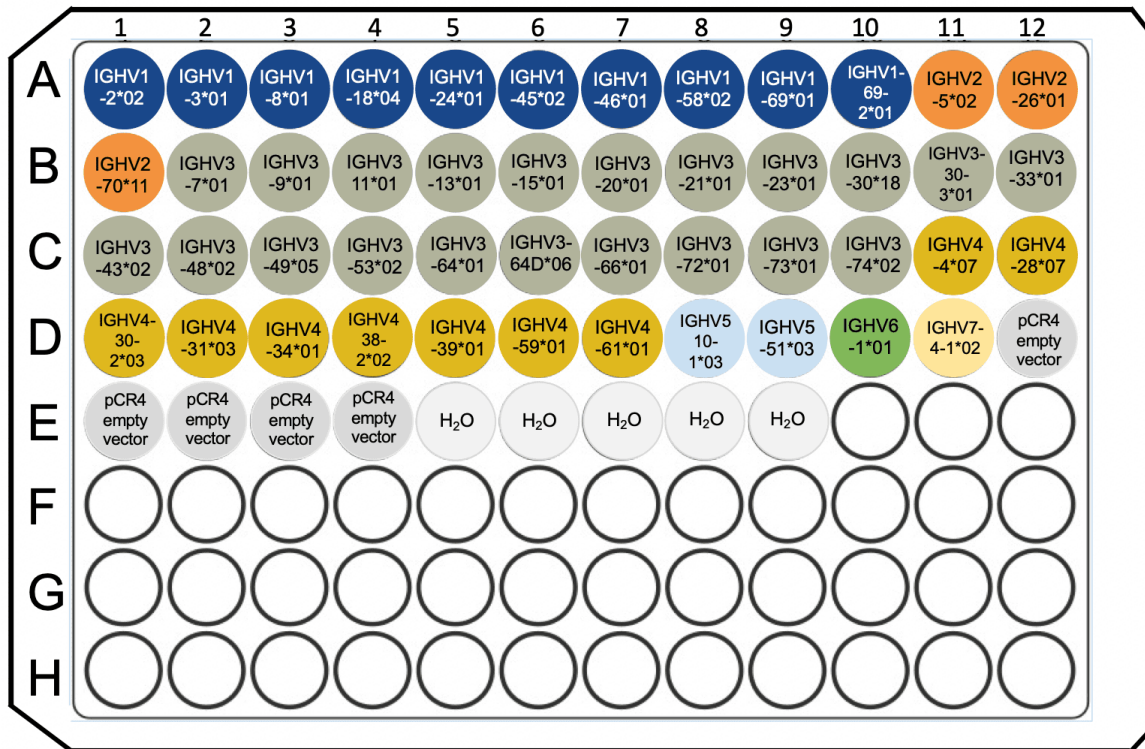

Figure 24: Example sheet of IGHV gene library plate

4.2.2 Analysis of the immunoglobulin sequencing data set gained during the preparation of the IGHV gene library

We isolated CD19+ cells from human PBMCs of 16 healthy donors and sorted CD20+IgM+IgD+IgG-CD27- B cells (naive B cells). The Clonetech SMARTer ®RACE 5'/3'-Kit was used as an unbiased and fast cDNA-transcription combined with a PCR amplification method. cDNA was produced according to the manufacturer's instructions via Clontech. Subsequently an A-tailing step was performed and leader plus V genes were cloned into pCR4 Topo-Vector (plasmid map provided in the supplements). During the IGHV gene library preparation 450 colony PCRs were performed on IGHV clones collected from the topo cloning colonies. Out of these 450 clones, 398 showed a positive PCR product (defined as sharp single band at approximately 830 bp), were subjected to DNA preparation and send for sequencing. Of these 398 mini-preps (plasmid DNA), we obtained 264 sequences that were annotated as

full-length IGHV genes at IMGT and passed our quality check (via phred score). The 264 heavy chain V gene sequences were filtered afterwards by the following criteria: full length sequence of the leader and the V gene, zero mutations in the leader, less than seven mutations in the whole V gene, first mutation not earlier than position 35 of the V gene and status (according to IMGT) must be functional or productive. This filter targets unmutated functional germline V genes (naive BCRs) with full length sequence annotation. 67 clones showed one or more mutations in the leader region, which lead to direct exclusion. 80 clones were excluded because they showed 7 or more mutations in their V gene. In 62 clones the first mutation in the V gene appeared earlier than position 35. 24 clones had to be excluded because their respective allele had no published annotated leader sequence or only a partial sequence in the IMGT database and could therefore not be analyzed correctly in the context of our filter.

In the end, 143 clones passed the filter and covered 37 of the 47 IGHV genes. The 10 missing genes were ordered as gene fragments to complete the gene library.

4.2.3   Naive B cell repertoire representation across 16 pooled healthy donors

Here, we present a comprehensive study of the functional IGHV family, gene and allele distribution over 264 IGHV sequences from 16 pooled donors.
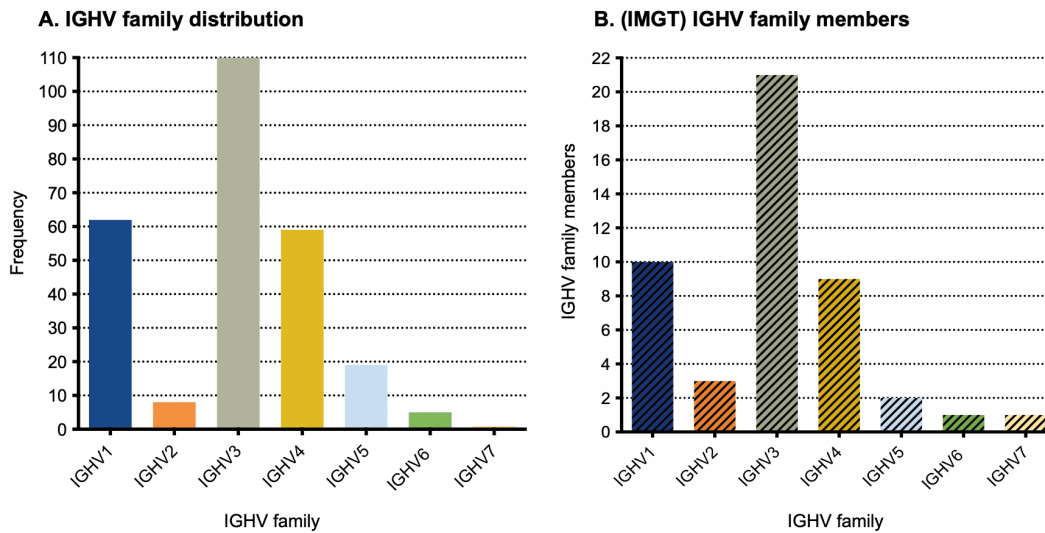

Figure 25: A. IGHV family distribution over 264 analyzed germline IGHV sequences from 16 pooled healthy donors
X-axis indicates the seven IGHV families, y-axis indicates the absolute frequency over the 264 analyzed sequences.
B. Total number of functional/ORF genes in each of the seven IGHV families according to the IMGT database (data status 04/2017)
x-axis indicates the seven IGHV families, y-axis indicates the total amount of genes belonging to each family

The size of each IGHV family is known to be variable, ranging from one functional/ORF gene in family 6 and 7 to 21 different genes in family 3 (see Figure 25B). The family distribution of the 264 detected IGHV genes (Figure 25A) resembles to a high extend to the general distribution in the IMGT database (Figure 25B). Only IGHV2 and IGHV5 show an inverted frequency regarding their proportional family members (IGHV2 library vs. IMGT 3.03% vs 6.38% and IGHV5 library vs. IMGT 7.20% vs. 4.26%). Additionally, IGHV family 7 is underrepresented over the population of our 16 donors (accounting for only 0.38% of all detected sequences while representing 2.13% of all annotated IMGT IGHV family members).

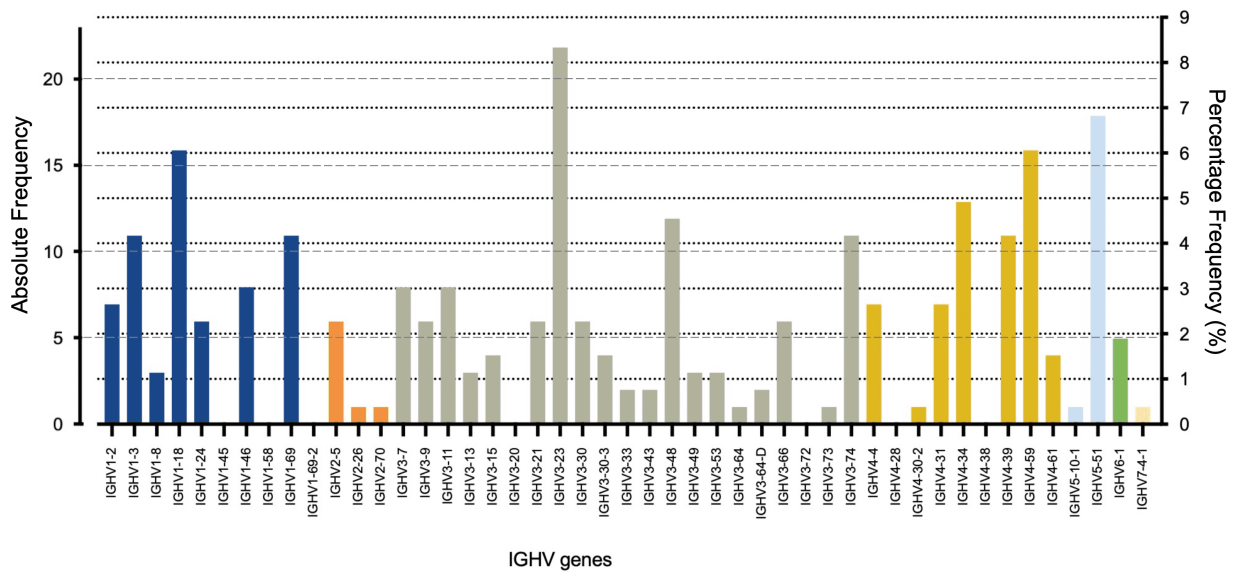Additionally, the sequence data was analyzed for IGHV gene and allele usage.

Figure 26: IGHV gene distribution over 264 analyzed sequences from 16 pooled healthy donors
X-axis indicates the 47 IGHV genes, left y-axis indicates the absolute gene frequency over the 264 analyzed sequences, right y-axis indicates the percentage frequency (%) of individual genes over the 264 analyzed sequences

We detected a clear predominance of certain genes and alleles among the analyzed pooled repertoires (see Figure 26 and Figure 27).

The four most frequently detected genes were IGHV1-18, representing 6.06%, IGHV3-23 representing 8.33%, IGHV4-59 representing 6.06% and IGHV5-51 representing 6.8% of all analyzed genes. We identified seven rare genes that were not represented at all, namely: IGHV1-45, IGHV1-58, IGHV1-69-2, IGHV3-20, IGHV3-72, IGHV4-28 and IGHV4-38. The biased usage of certain gene alleles among the 16 pooled individuals is demonstrated in Figure 27. In the allelic distribution a clear predominance of certain gene alleles was detectable. Additionally, it is noticeable that the majority of genes expressed exclusively one or at most two alleles, while IGHV3-11 was presented by four different alleles (*01, *04, *05 and *06) and IGHV1-69 was presented by five different alleles (*01, *02, *04, *06 and *10) among the 16 donors. 93 of the 160 IGHV gene alleles annotated by IMGT were not represented in this data set at all.
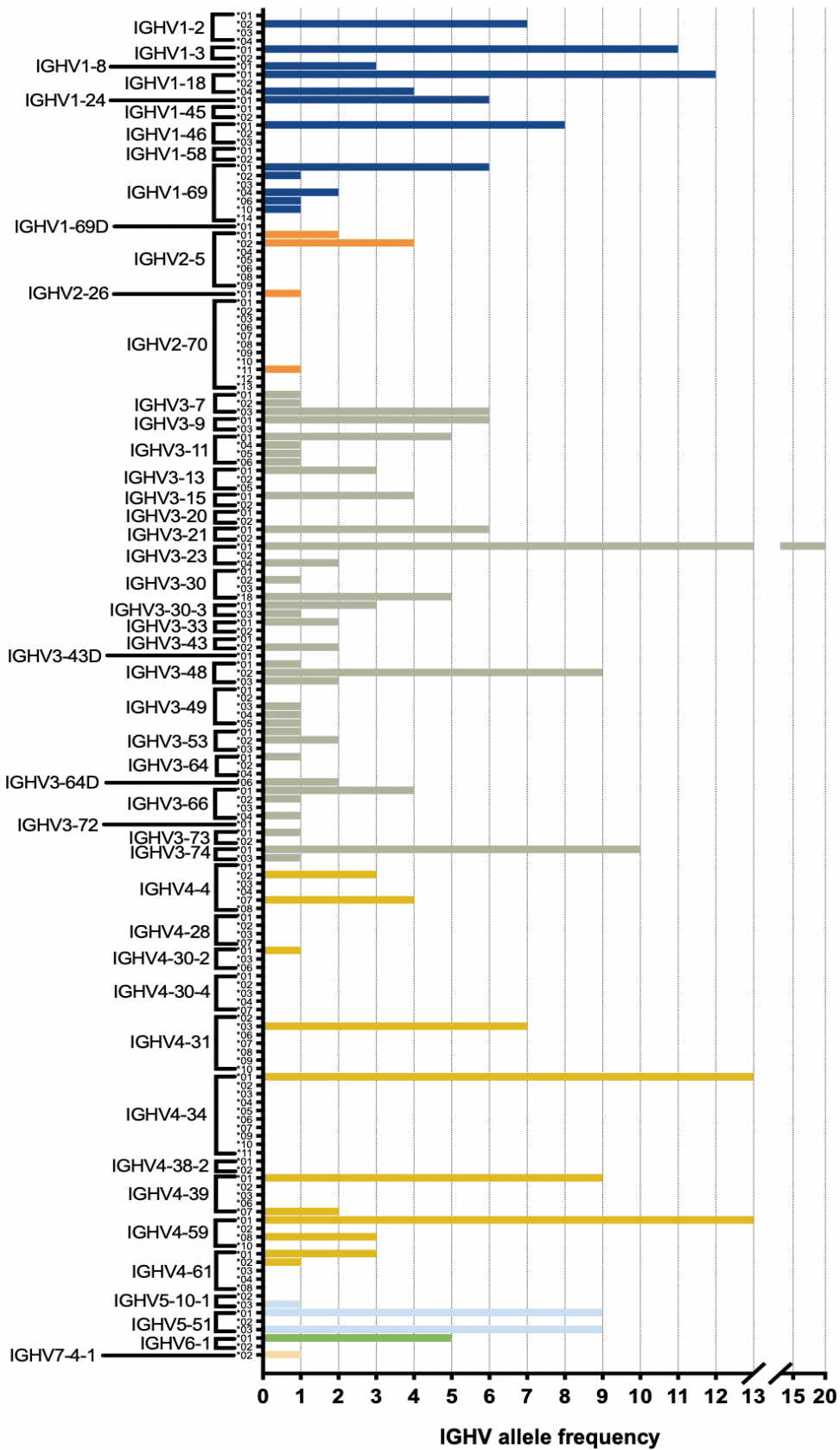
Figure 27: IGHV allelic distribution over 264 analyzed sequences from 16 pooled healthy donors
X-axis indicates the 160 different IGHV gene alleles, y-axis indicates the absolute allelic frequency over the 264 analyzed IGHV genes.

4.2.4 Sequence data gained during the preparation of synthetic IGKV and IGLV germline gene libraries

To allow unbiased primer testing on a plasmid library representing the kappa and lambda light chain, the workflow presented for IGHV (except for the choice of primer) was identically performed for both light chains. The light chain libraries were finalized and used for light chain primer evaluation after closure of this MD project (see Kreer et al.[90]) but BCR repertoire analysis of family, gene and allele distribution over 156 IGKV genes and 160 IGLV genes, which were cloned by N. Lehnen, are presented in the supplements (see Supplement Figure S8 – Figure S13).

4.2.5 CDR3 lengths of naive B cells from 16 pooled individuals

In addition to gene and allele usage, the BCRs specificity and binding capacity is critically dependent on the CDR3 region. Its length has profound effects on shape and structure of the CDR3 loop and differs in average between naive and antigen-experienced B cells.

260 out of 264 IGHV genes from our IGHV data set were analyzed using IgBlast to determine the respective CDR3 length. Four sequences were excluded due to incomplete annotation or low phred score in the CDR3 region. Figure 28 shows the frequency distribution of the CDR3 lengths over the 260 analyzed naive B cells with a mean length of 43.9 nucleotides and a standard deviation of 10.592.
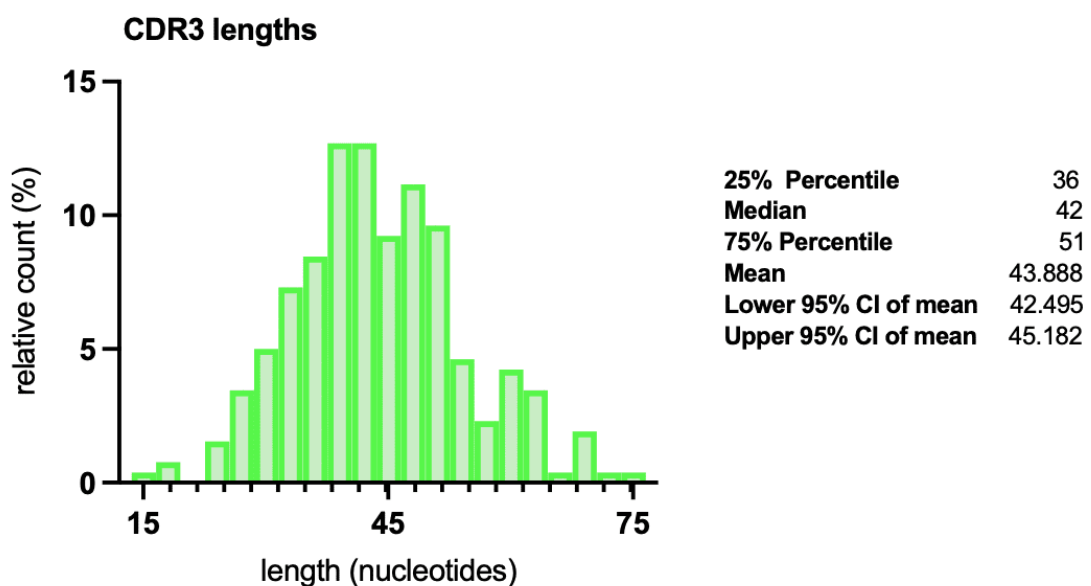

25% Percentile	36
Median	42
75% Percentile	51
Mean	43.888
Lower 95% CI of mean	42.495
Upper 95% CI of mean	45.182

Figure 28: Frequency distribution of CDR3 lengths of naive B cells from 16 pooled individuals
CDR3 regions and lengths were identified from 260 IGHV sequences of 16 pooled individuals using IgBlast[123]. A bin width of three was used for frequency distribution.

4.3 Evaluation of human IGHV repertoire amplification by *de novo* primer set oPR(5)-IGHV in comparison to established primer sets (Set1 and Set2), using a synthetic IGHV gene library

In order to evaluate the human IGHV gene coverage by the *de novo* primer set oPR(5)-IGHV in comparison to the well-established primer sets Set1 and Set2, PCR experiments using the respective primer mix were performed on the well-defined synthetic IGHV gene library. Four independent PCR experiments were performed with each of the primer sets. All mPCRs were performed with the reverse primer Ippolito_IgM_all_rev. Detailed results of all 564 PCRs, including gel electrophoresis images are presented in the supplements (see Supplement Figure S14).
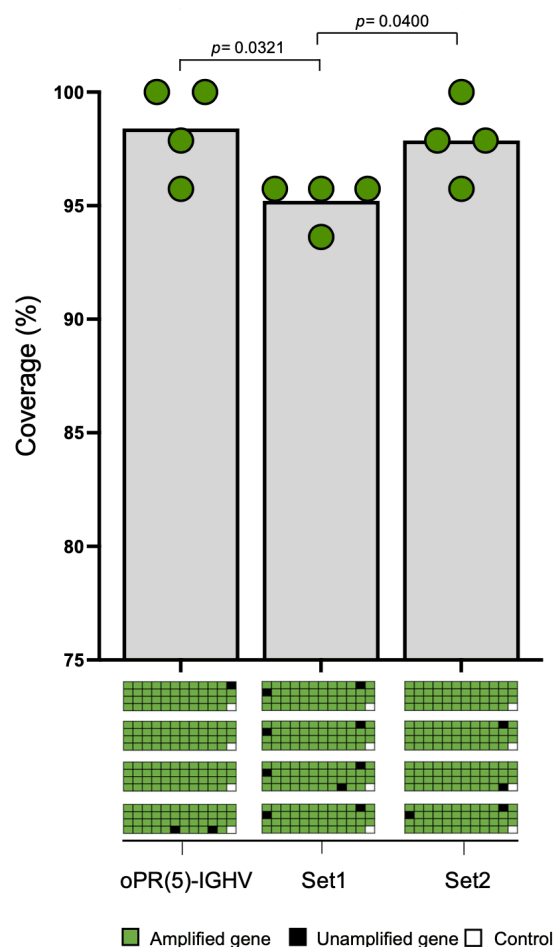

Figure 29: Validation oPR(5)-IGHV in comparison to established primer sets on the IGHV gene library
Comparison of the efficacy of the oPR(5)-IGHV primer mix to two published primer sets that have been used to isolate antibody sequences before, was performed in four independent PCR experiments on the IGHV gene library. The percentage coverage of each independent experiment is indicated as green dots. Bar graphs show mean coverage values over the four experiments. The results of each individual experiment is represented by multiwell plates underneath the graphs with green, black, and white squares indicating amplified, unamplified, and negative control wells, respectively. P-values for pairwise comparisons are indicated only below a significance level of 0.05. See also Supplement Figure S14 for corresponding gel electrophoresis pictures.

The oPR(5)-IGHV set showed a coverage of 100% in two out of four experiments and a mean overall coverage of 98.40%. Set1 showed a mean overall coverage of 95.21%. In four out of four experiments well A11, containing the germline IGHV2-26*01 gene and B1, containing the germline IGHV2-70*11 gene, were not amplified by Set1. Consistent to this finding, pre-evaluations in the openPrimeR evaluation mode (see Figure 6), could show that the primers from Set1 were, due to few sequence-complementarity, very unlikely to amplify genes from the IGHV family 2. Surprisingly, germline IGHV2-5*02 was covered by Set1 in all four experiments, suggesting that the respective primer introduces more than seven mutations or even stop codons, when amplifying this gene. Set2 showed a mean coverage of 97.87% with a random pattern of unamplified genes.

We could demonstrate that the *de novo* primer oPR(5)-IGHV has the capacity to possibly amplify all human VH immunoglobulins in germline status. oPR(5)-IGHV and Set2 could be demonstrated to show significantly superior amplification rates in pairwise comparison to Set1 (p= 0.0321 and p= 0.0400). It should be noted that additional assessment for bulk quantitative approaches would be required to eliminate or compensate for possible primer bias. Additionally, amplicons have not been sequenced to study possibly induced mutations by the individual primer sets at binding regions, which would be of special interest for planed antibody neutralization assays or in-depth repertoire studies.

4.4 Evaluation of the reproducibility of PCR experiments on the IGHV gene library by analysis of amplification status of 16 single primers from de novo primer set oPR(5)-IGHV on the 47 synthetic IGHV gene templates in triplicates

One of the ideas behind a standardized gene library was to create a well-defined PCR template, with as little template-provided bias as possible. As a proof of concept, the reproducibility of PCR experiments on the IGHV gene library was evaluated on 752 independent triplicate experiments using 16 single primers of the oPR(5)-IGHV primer set (combined with reverse primer Ippolito_IgM_all_rev).

	oPR(5)-IGHV_1	oPR(5)-IGHV_2	oPR(5)-IGHV_3	oPR(5)-IGHV_4	oPR(5)-IGHV_5	oPR(5)-IGHV_6	oPR(5)-IGHV_7	oPR(5)-IGHV_8	oPR(5)-IGHV_9	oPR(5)-IGHV_10	oPR(5)-IGHV_11	oPR(5)-IGHV_12	oPR(5)-IGHV_13	oPR(5)-IGHV_14	oPR(5)-IGHV_15	oPR(5)-IGHV_16	sum	Percentage of valid PTPs (%)
+ + +	11	10	26	28	20	3	12	18	11	10	1	7	19	2	0	3	181	24.59
+ + -	2	0	1	5	3	2	3	11	1	0	0	18	1	0	0	0	47	6.39
+ - -	1	1	0	1	2	0	5	4	1	2	0	11	0	0	0	0	28	3.80
- - -	33	36	20	13	22	26	27	14	34	35	46	11	27	45	47	44	480	65.22
invalid	0	0	0	0	0	16	0	0	0	0	0	0	0	0	0	0	16	(-)
sum	47	47	47	47	47	47	47	47	47	47	47	47	47	47	47	47	752	

Figure 30: Reproducibility of amplification status of 16 single primers of oPR(5)-IGHV on the IGHV gene library
16 single primers of the oPR(5)-IGHV set were tested in three independent PCR experiments on each of the 47 templates of the IGHV gene library. Amplification status of each experiment was evaluated as either amplified or unamplified according to their visualization via gel electrophoresis (see Supplement Figure S15). 16 triplicates were excluded (considered invalid) due to high amount of unspecific smear bands. Legend: (+) = amplified, (−) = not amplified. e.g. (++-) indicates two out of three PCR experiments with the respective primer showed a positive amplification status via gel electrophoresis.

Due to a high amount of unspecific smear bands, indicating contamination, 16 triplicates were excluded. Of note, all 16 invalid triplicates were detected with one primer, and only in one of three experiments. For respective gel electrophoresis pictures of each PCR see Supplement Figure S15). 89.81% (661 out of 736 valid PCRs) showed ether a positive or a negative PCR result in all three independent PCR experiments. Additionally, in 6.39% (47 out of 736 valid PCRs) a positive result could be reproduced in at least one more experiment.

We concluded that the IGHV gene library provides a highly satisfying template-tool for an unbiased amplification validation, if tests were run in triplicates. In Figure 31 the reproducibility of each amplification status of respective primer and IGHV gene template is visualized. Additionally, PCR experiments with four single primers of Set1 are included.

Figure 31: Visualization of reproducibility of amplification status of 20 single primers from primer set oPR(5)-IGHV and primer Set1 on 47 germline IGHV genes templates in triplicate experiments

16 single primers of the *de novo* oPR(5)-IGHV primer set and four single primers of the well-established primer set Set1 were tested in three independent PCR experiments each on the IGHV gene library. Amplification status of each individual experiment is colored dark green, light green, grey or black, indicating three out of three experiments as amplified, two out of three experiments as amplified, two out of three experiments as unamplified or all three experiments as unamplified, respectively. See also Supplement Figure S15 and Supplement Figure S16 for corresponding gel electrophoresis pictures.

4.5   Analysis of novel *Tac* PCR based IGHV data set regarding PTP properties and the influence of mismatch position and free energy of annealing on the amplification status

We generated a comprehensive *Taq* PCR-based IGHV data set comprising information on 940 different primer-template-pairs (PTPs). The following computational analysis of the physicochemical properties of each PTP (including the generation of Table33, 34 and Figure 32) was performed by Mathias Döring and results were pre-published in Scientific Reports in 2019 by Döring et al. (Co-author N. Lehnen) and are shown here with courtesy of Creative Commons Attribution- Share Alike 4.0. International License, *https://creativecommons.org/licenses/by-sa/4.0/.*

The data set includes the amplification status of the synthetic IGHV gene library by 20 different single primers from the *de novo* primer set oPR(5)-IGHV and the well-established primer Set1. Experiments were three times repeatetly performed for each combination primers and template, including 20 primers (16 primers of the oPR(5)-IGHV set and four primers from Set1) and 47 templates. A total of 2,820 PCRs (940 triplicates) were performed and evaluated by gel electrophoresis for ether amplified or unamplified status. In triplicate measurements we validated the status as *Amplified* when at least two out of three PCRs showed a positive result. Status evaluation was performed by five independent reviewers and PTPs were only titled positive if the majority of reviewers set the status to positive.

Gel electrophoresis images and evaluation status of all respective PCR experiments can be found in the supplements (see Supplement Figure S15 and Figure S16). PTPs with more than 12 mismatches were excluded since an exact primer binding side can hardly be determined. Hence, 908 of the 940 PTPs of the *Tac* PCR data set were selected to investigate which physicochemical properties exert the greatest influence on the amplification status. Therefore, openPrimeR was used to compute the physicochemical properties for each PTP. The most important features we investigated on were the free energy of annealing (ΔG) and three conditions related to the 3' terminus: the presence of mismatches within the 3' hexamer ($z \in \{0,1\}^6$), the total number of mismatches ($X_N \in \mathbb{N}_0$) and the position of the 3' hexamer mismatch closest to the 3' terminus ($i_X \in \{0,1,...,6\}$).

Property	Interpretation	oPR(5)-IGHV	Set1
ΔG	Free energy of annealing	[−4.9, −2.0]	[−8.6, −5.2]
i_x	Mismatch closest to 3′ end	[2, 6]	[0, 1]
X_N	Number of 3′ hexamer mismatches	[1, 3]	[0, 1]
$|GC|$	Extent of GC clamp	[1, 2]	[1, 1]
ΔG_f	Free energy of folding [kcal/mol]	[−1.53, −0.24]	[−1.24, −0.76]
ΔG_s	Free energy of self-dimerization [kcal/mol]	[−2.1, −0.7]	[−1.2, −0.8]
$y_i = Amplified$	Positive amplification status	217 of 720 (30.1%)	165 of 188 (87.8%)
$\sum_{xi} z_j$, $j = 1$	Number of mismatches at the start of the 3′ hexamer	271	25
$\sum_{xi} z_j$, $j = 2$	Number of mismatches at the 2nd position of the 3′ hexamer	226	4
$\sum_{xi} z_j$, $j = 3$	Number of mismatches at the 3rd position of the 3′ hexamer	272	31
$\sum_{xi} z_j$, $j = 4$	Number of mismatches at the 4th position of the 3′ hexamer	246	11
$\sum_{xi} z_j$, $j = 5$	Number of mismatches at the 5th position of the 3′ hexamer	308	12
$\sum_{xi} z_j$, $j = 6$	Number of mismatches at the 3′ terminal position	308	12

Table 33: Overview of the physicochemical properties of PTPs of the IGHV data set
908 PTPs of the IGHV data set were analyzed regarding their physicochemical properties (free energy of annealing (ΔG), mismatch closest to the 3' end (i_x), number of 3' hexamer mismatches (X_N), extent of GC clamp ($|GC|$), free energy of folding (ΔG_f), free energy of self-dimerization (ΔG_f), amplification status (y_i) and the number of mismatches at the respective 3' hexamer position $\sum_{xi} z_j$. Values shown in brackets indicate the inter-quartile range of the observed values. This (modified) table and table legend was published in Scientific Reports in 2019 by Döring et al. (Co-author N. Lehnen) and is used in approval with all co-authors and with courtesy of Creative Commons Attribution- Share Alike 4.0. International License, *https://creativecommons.org/licenses/by-sa/4.0/*

Table **33** gives an overview of the distribution of physicochemical properties of the IGHV data set, comparing the *de novo* oPR(5)-IGHV set and primer Set1. The primers from oPR(5)-IGHV and primer Set1 are characterized by contrasting rates of amplification and contrasting values of ΔG. 87.8% of PTPs (165 of 188) from Set1 were labeled *Amplified*, while only 30.1% (217 of 720) of PTPs from oPR(5)-IGHV were labeled *Amplified*. It is remarkable that the primers from Set1 consistently exhibited a higher rate of amplification than the primers from the oPR(5)-IGHV at any compared number of mismatches. Absolute numbers of mismatches must not be compared, due to a higher amount of primers (16 versus 4) and therefor a higher amount of total PTPs in the oPR(5)-IGHV set.

Additionally, ΔG of PTPs from the oPR(5)-IGHV primers showed a significantly higher inter-quartile range (IQR) at [−4.9, −2.0], in comparison to Set1 with a ΔG IQR of [−8.6, −5.2].

Bringing together the analyzed characteristics, Table 34 demonstrates the interaction between the number of primer-template mismatches, ΔG, and the amplification rate. In the overall data set, primers with up to 3 mismatches still showed a 100% amplification rate. It is remarkable that even primers binding with at most 6 mismatches still showed very high amplification rates at 83.33% in the overall data set and even 100% regarding Set1 alone.

Number of mismatches	i_x	ΔG [kcal/mol]	Amplification rate	Primer set
0	[0, 0]	[−16.616, −15.696]	100%	Overall
1	[0, 3]	[−14.353, −12.1]	100%	Overall
2	[0, 3]	[−12.0455, −9.656]	100%	Overall
3	[0, 4]	[−11.607, −7.9185]	100%	Overall
4	[2, 6]	[−10.796, −7.409]	92.31%	Overall
5	[0, 3]	[−7.047, −6.047]	88.89%	Overall
6	[0, 0]	[−8.603, −5.11325]	83.33%	Overall
7	[0, 3]	[−5.39, −4.212]	67.19%	Overall
8	[3, 6]	[−5.56075, −2.539]	34.04%	Overall
9	[4, 6]	[−3.5335, −2.1325]	23.08%	Overall
10	[4, 6]	[−4.09, −1.724]	18.02%	Overall
11	[4, 6]	[−3.74, −1.695]	10.53%	Overall
12	[6, 6]	[−2.624, −1.413]	3.75%	Overall
0	[0, 0]	[−16.07, −15.609]	100%	oPR(5)-IGHV
1	[0, 3]	[−13.283, −12.1]	100%	oPR(5)-IGHV
2	[0, 3.25]	[−11.94175, −9.656]	100%	oPR(5)-IGHV
3	[0, 4]	[−11.607, −7.66375]	100%	oPR(5)-IGHV
4	[2, 6]	[−10.974, −6.686]	90.91%	oPR(5)-IGHV
5	[2.5, 4.5]	[−8.36825, −6.4925]	75%	oPR(5)-IGHV
6	[3.25, 4]	[−4.4545, −2.9]	33.33%	oPR(5)-IGHV
7	[3, 6]	[−4.212, −2.539]	9.52%	oPR(5)-IGHV
8	[4, 6]	[−3.303, −2.06275]	18.06%	oPR(5)-IGHV
9	[5, 6]	[−3.0985, −2.0395]	13.51%	oPR(5)-IGHV
10	[5, 6]	[−3.393, −1.695]	11.26%	oPR(5)-IGHV
11	[5, 6]	[−3.351, −1.695]	4.2%	oPR(5)-IGHV
12	[6, 6]	[−2.608, −1.413]	2.6%	oPR(5)-IGHV
0	[0, 0]	[−20.79275, −16.616]	100%	Set1
1	[0, 2]	[−17.782, −14.045]	100%	Set1
2	[0, 0]	[−14.4805, −12.5605]	100%	Set1
3	[1, 1]	[−10.505, −10.505]	100%	Set1
4	[0.75, 2.25]	[−10.29475, −9.29225]	100%	Set1
5	[0, 0]	[−6.047, −6.047]	100%	Set1
6	[0, 0]	[−8.603, −5.208]	100%	Set1
7	[0, 0]	[−5.39, −5.208]	95.35%	Set1
8	[0, 0]	[−5.937, −3.95]	86.36%	Set1
9	[1, 6]	[−5.58, −2.89]	78.95%	Set1
10	[0, 3]	[−5.208, −2.956]	66.67%	Set1
11	[0, 2.25]	[−5.208, −2.8395]	64.29%	Set1
12	[4, 5.5]	[−2.6225, −1.9615]	33.33%	Set1

Table 34: Amplification rates in dependence of the number of PTP mismatches, position of 3' hexamer mismatch closest to the 3' terminus (i_x) and free energy of annealing (ΔG) The relationship between the number of total mismatches, the position of 3' hexamer mismatch closest to the 3' terminus (i_x), free energy of annealing (ΔG), and the amplification rate were analyzed for the overall set, and for oPR(5)-IGHV and Set1 individually. Numbers in brackets indicate value ranges. This modified table and table legend was pre-published in Scientific Reports 2019 by Döring et al. (Co-autor N. Lehnen) and is used in approval with all co-authors and with courtesy of Creative Commons Attribution- Share Alike 4.0. International License, *https://creativecommons.org/licenses/by-sa/4.0/*

The data set indicated coherence between the studied features: The lower the number of mismatches, the lower the inter-quartile ΔGs detected, and the higher amplification rate was

achieved. This pattern was demonstrated by the overall set and by both primer sets individually.
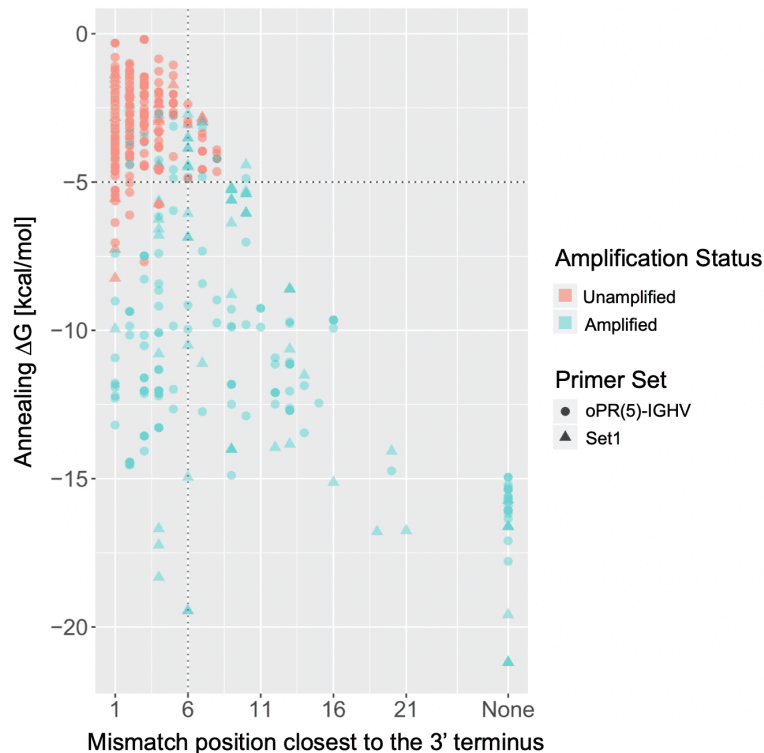

Figure 32: The impact of mismatch position and free energy of annealing ΔG on the amplification status
X-axis indicates the mismatch position closest to the primers 3' terminal end. Position 1 corresponds to last position of the primer and 3' end and e.g., position 6 corresponds the 6th last position. PTPs without mismatches are labeled *None*. PTPs of the oPR(5)-IGHV set are indicated as dots, PTPs of primer Set1 are indicated as triangles. *Amplified* PTPs are colored light blue, *Unamplified* PTPs are colored light red. The vertically dotted line marks the 3' hexamer, the horizontal dotted line marks free energy of -5 kcal/mol. Both lines indicate cut-offs for the amplification status. This modified figure and figure legend was published in Scientific Reports 2019 by Döring et al. (Co-author N. Lehnen) and is used in approval with all other co-authors and with courtesy of Creative Commons Attribution- Share Alike 4.0. International License, *https://creativecommons.org/licenses/by-sa/4.0/*

Figure 32 visualizes the 3' hexamer as critical region of primer mismatch for the amplification status: while mismatches outside of the 3' hexamer less frequently affected the amplification status, mismatches closer or within 3' terminus influence the amplification status to a great extent. Mismatches in the 3' hexamer are detected less frequently in *Amplified* PTPs than in *Unamplified* PTPs (X_N IQR of [0, 1] vs [2, 4]). This accounts especially for mismatches closer to the 3' hexamer (i_X IQR of [0, 3] vs [5, 6]). *Unamplified* PTP status is additionally associated with higher free energy of annealing. A cut-off can be set to $- 5$ kcal/mol, above which amplification becomes unlikely. Using a two-sided Wilcoxon rank-sum test, a significant difference between *Amplified* (n=382) and *Unamplified* (n=526) PTPs concerning ΔG (*p-value* <0.05) and i_X (*p-value* <0.05) was detected.

5   DISCUSSION

Potent neutralizing antibodies are of tremendous value and are urgently needed to combat global health crises caused by HIV-1, SARS-CoV-2 and other (potentially upcoming) infectious diseases. Additionally, studies of antigen-specific B cells and their respective BCR are needed to gain further insights into disease processes in autoimmunity,[128] and antitumor immune responses.[129] Therefor, reliable and efficient antibody and human B cell isolation methods are of great value.

Due to the B cell repertoire's origin from one of the most complex regions of the human genome[34] and its tremendous diversity gained during affinity maturation, especially in the variable region of the antibody's heavy and light chains[38], the isolation of antigen-specific antibodies presents one with great challenges. One of the most critical steps in antibody isolation is the two-step multiplex PCR amplification, especially in the context of high mutation frequencies. Here, mPCR requires complex primer design as PCR amplification is often compromised on immunological templates by primer-template mismatches or unfavorable primer properties and interactions. However, currently available primer design tools lack functionality for a SCO[100–103,105,107,110,111,130], are not openly available[91–94,96,101,103,104,113] or do not provide an easy-to-use graphical interface or workflow application, which allows defining a ROI and an individual primer binding site.[91,94,95,101,102,104,109]

This dissertation has intensively and profoundly been dedicated to further develop single cell and multiplex cloning strategies for the identification of (broadly) neutralizing antibodies against infectious diseases. To this end, we adapted existing concepts[42,75,81,131] to develop a standardized laboratory protocol with cost- and labor intensity optimized conditions as a prerequisite for a fast and efficient workflow. Additionally, we presented a newly developed primer design- and evaluation tool, called openPrimeR (as part of a co-operational work with the MPI Saarbrucken). For an unbiased evaluation of primer sets on defined immunological templates, we established a germline IGHV gene library from pooled naive B cells of 16 healthy individuals. In a continuous workflow of in-vitro testing of computationally designed primers on the gene library and in-silico evaluation in the openPrimeR tool, we optimized the tool's default- and constraint-settings for primer design. The finally presented oPR(5)-IGHV primer set demonstrated superior performance in amplifying the (germline) IGHV gene repertoire in comparison to established primer sets and is a promising candidate for the amplification of highly mutated antibodies. Finally, a *de-novo* Tac PCR data set, containing 940 evaluated PTPs, is presented and brought great insights into physicochemical properties of PCR amplification and primer design.

5 Discussion

5.1 Optimized adjustments of the laboratory protocol for the amplification of neutralizing antibodies from single human B cells

To implement an optimized and standardized protocol for the successful cloning of antibodies from single human B cells of defined origin, we adapted the priorly published protocol presented by von Boehmer et al.[131] from the Nussenzweig laboratory at the Rockefeller University in New York, which had been established and demonstrated initially by Wardemann et al.[42] and Tiller et al.[81] to be a well working protocol for the amplification of human B cell receptors. The complete cloning protocol consists of four critical steps, including (I) single B cell isolation by flow cytometry, (II) BCR heavy and light chains amplification by RT-PCR, (III) a cloning process into an expression plasmid and (IV) a transfection step to finally produce recombinant antibodies (IV).[42, 81,132] In this thesis, the focus was set to optimizing step (II), the BCR heavy and light chain's amplification by PCR.

To reduce labor intensity and speed up the simultaneous handling of large numbers of clones, all PCR purification and cloning reactions were performed in 96-well plates. Additionally, we could show, that downscaling the 2nd PCR reaction volume (38 µl to 19 µl, as demonstrated in section *4.1.1.2*) does not result in lower amplification rates or weaker gel bands, which allowed us to adapt the protocol by remarkably reducing the 2nd PCR's reagent volume. By down scaling the PCR reagent volume to half of the initial volume, we could archive a 67% cost reduction per cloned antibody in the later on established total protocol by Gieselmann et al. (*Co-author N. Lehnen*).[132]

Single cell amplification is a highly sensitive and error prone technique which can be influenced by a multitude of external factors. The catalyzing enzyme polymerase is one of the key reagents in a PCR, especially when applied on immunological templates, because the amplification status on the one hand and a potentially induced bias on the other hand, are highly dependent on the type and quality of polymerase.[68,85,133] The enzyme is also one of the most cost intensive factors in a polymerase chain reaction. We therefore tested the performance of three commercially available DNA polymerases on single human antigen-experienced B cells and found no significant difference in the amplification rates (see Figure 17). The protocol was therefore adapted by using the less cost intense polymerase, which in our case turned out to be PolymeraseC, the Platinum Taq Green Hot Start DNA Polymerase by Thermo Fisher Scientific. However, the interpretation of the results must not suggest that the choice of polymerase is irrelevant. The three DNA polymerases, which were compared here, had been preselected and were already characterized by high quality and specificity. Since the performed experiments only investigated on the amplification status via gel electrophoresis, no statement on amplification bias or mutations induced by the different

polymerases could be made. Future investigations herein may be particularly beneficial for studies of the B cell repertoire and accurate immunoglobulin nucleotide sequence amplification of potentially neutralizing antibodies.

Proceeding in the optimization process, we recommend for achieving a better amplification rate and less non-specific binding or smear bands, to proceed directly from the 1st to the 2nd PCR or, in case of a necessary interruption, to store 1st PCR plates at -80°C instead of 4°C or warmer (see Figure 16). Since RT-PCR on single-cell level is extremely susceptible to nucleotide contamination, it is highly recommendable to perform single-cell RT-PCRs only in workspaces, or even a designated room, where no other experiments on DNA templates (e.g., plasmids encoding antibody fragments) take place. PCR master mixes should exclusively be prepared inside a PCR cabinet, under regularly decontamination by UV irradiation and wiping with DNA decontaminating reagents. During all single cell experiments we held strictly to this method of operation and implemented it into our protocol. However, performing literally thousands of PCR reactions on single cell level, we found the amplification rates to be highly varying, even if applied on identically sorted B cells and using same protocol settings.

In section *4.1.1.5* we could demonstrate that the choice of the reverse transcribing enzyme in RNA RT highly influences the reproducibility of a positive amplification result and most importantly influences the overall amplification rate. We demonstrated, that using SuperScript IV instead of SuperScript III (both from ThermoFisher Scientific) significantly increases the overall amplification rate, as well as the likelihood of reproducing a positive PCR result in a duplicate measurement (for both p= <0.0001). It can be assumed that this observation is mainly explained by a higher cDNA yield and quality achieved by SuperScript IV as promoted by the manufacturer. Recently, Zucha et al.[134] corroborate these findings by comparing 11 RTases in quantitative RT qPCRs on single-cell and bulk templates and demonstrated that the SuperScript IV (and the Maxima H, also provided by ThermoFisher Scientific) show superiority over the other RTs, in terms of their ability to capture rare transcripts and their absolute reaction yield.[134] Consequently, we adapted the cDNA production protocol by exclusively using the RT SuperScript IV for the reverse transcription step.

By investigating on optimal reagent volume, storage conditions and highest performing enzyme usage (polymerase and reverse transcriptase), we were able to adjust the established single cell amplification protocols[42,81,131] to a less cost- and labor intensive and a more efficient protocol with higher reproducibility. Even if these findings and protocol adjustments are only subpoints in the whole process of antibody identification and production, they contributed to a

high extend to the further development of a protocol for the *"effective high-throughput isolation of fully human antibodies targeting infectious pathogens"* published by Gieselmann et al.[132]

5.2 openPrimeR for multiplex amplification of highly diverse templates

Successful mPCR-based single B cell cloning approaches are critically dependent on a reliable primer set, as a prerequisite of a broad, unbiased and efficient amplification.

In the past, several primer sets have contributed to the isolation of monoclonal antibodies or studies of the B cell repertoire.[75,81,86,87,135,136] Using the openPrimeR evaluation mode, Kreer et al.[90] (Co-author N. Lehnen) could show, that all of the investigated priorly established primer sets show calculated limitations, such as the missing coverage of all currently known IGHV genes,[81,75] providing only incomplete V gene segment amplification[81,86,135,87] or requiring multiple reactions and thus are highly cost- and labor-intensive[75] (see Figure 6). In this thesis, we intended to advance the PCR amplification by investigating on an optimized primer set for the identification of broad neutralizing antibodies.

Mutations that lead to a decrease in primers-template complementarity pose a particular challenge for the PCR amplification. While antigen-experienced B cells usually carry on 5-10% mutations within their BCR[137], bNAbs against HIV-1 often show less than 70% germline identity.[87,138] Previous PCR-based amplification approaches used sets of degenerated primers to overcome the difficulties of primer-template mismatch. However, due to non-specific or failed binding, 10-20% of all antibodies cannot be amplified using this method.[139–141] Additionally, degenerated primers bear a non-negligible risk of inducing or reverting mutations. The individual V gene segments mutations are crucial for the antibodies neutralizing activity and occur in both, the CDRs and FWRs of the antibody.[23] Primer binding to the heavy (and light-) chain leader region has been demonstrated to be more efficient in amplifying highly mutated V genes[75], which is comprehensible regarding the remarkably lower AID activity in this area.[39] Moreover, primer binding to the leader region prevents the induction or reversion of mutations in the FWR1, which can play a major role in the antibodies neutralizing activity.[23] Thus, we made exclusive leader binding a prerequisite for of all further primer sets designed with openPrimeR. Primers binding more downstream were directly excluded (see Supplement Figure S18).

We used openPrimeR to design a *de novo* optimized primer set, to facilitate the efficient amplification of highly mutated human immunoglobulin heavy chain gene segments. Primer sets were designed to cover the maximum amount of all included 152 IGHV gene alleles with a minimum number of primers (= considering the SCO). It is of note, that first experiments of

this thesis were performed at a very early stage of the openPrimeR tool's development. The presented primer sets oPR-IGHV(1) - oPR-IGHV(6) were designed in a iterative process of in-vitro testing of computationally designed primers and an in-silico adjustment of the tools starting conditions and filtering values (such as number of allowed mismatches, GC-content etc.) that were eventually chosen as default settings for openPrimeR (see supplement Figure S18). Within this optimization process, we were able to improve the initially weak primer performance of the oPR(1)-, oPR(4)- and oPR(6)-IGHV set (see Figure 20 and Figure 21) to finally present the highly satisfying primer set oPR(5)-IGHV. We could demonstrated that the oPR(5)-IGHV primer set showed equally overall coverage of antigen-experienced and HIV-1-specific B cells on single cell level compared to primer Set1. Additionally, the oPR(5)-IGHV primer set showed superior performance in covering the IGHV gene library compared to Set1 (Figure 29).

Primer Set1 is still broadly used in many mAbs isolation pipelines and multiple laboratory groups.[142–144] In the ongoing pandemic, new protocols to identify mAbs against SARS-CoV-2 relied on these primers.[145,146] However, the set lacks the ability to cover for all IGHV genes (see Figure 6 and Figure 29), which is an unfavorable feature for a primer set aiming to map the B cell repertoire and to detect mAbs against a broad variety of pathogens. Primer Set2 on the other hand presented high overall coverage rates in the amplification of murine and human IGHV genes in prior publications[86,88], as well as in the here presented experiments on single antigen-experienced B cells and on the IGHV gene library. However, this set bears the troubling disadvantage of primer binding to the FR1 (see Figure 6), which is a feature to be avoided urgently, as extensively discussed above.

The *de novo* primer set oPR(5)-IGHV provides both, an exclusive binding to the leader region, to capture antibodies with high mutational burden and a broad coverage of theoretically all known IGHV genes in germline status. To ensure the broad repertoire coverage, the oPR(5)-IGHV set remarkably includes some primers, that bind to one single IGHV gene (see supplement Figure S15), e.g. oPR(5)-IGHV_11), while the four primers of primer Set1 show all a relatively broad amplification spectrum among IGHV genes and families (see supplement Figure S16). Additionally, the oPR(5)-IGHV was demonstrated to amplify its target region with less PTP-mismatches in comparison to primer Set1 (Table 34). Again, this is a strength of the oPR(5)-IGHV set, as mismatch binding can lead to the introduction of unwanted sequence mutations and interfere analysis.

To conclude, we used openPrimeR to design an IGHV primer set that facilitates efficient amplification of highly mutated variable heavy chain sequences and could demonstrate its

efficiency on both, the gene library and on single cells. The oPR(5)-IGHV set is therefor a promising candidate for the precise amplification of antibody sequences independently of their mutational burden. Further testing on the isolation of broad and potent neutralizing antibodies against different pathogens would be needed. Specifically, it would be of interest to sequence the amplified variable region by the respective primers to identify germline divergence and thereby identify the primer set's specific capacity to amplify highly mutated V genes.

5.3 IGHV gene library – a powerful tool for standardized and unbiased primer testing and evaluation on human IGHV genes

The here presented IGHV gene library provides a powerful template tool for V gene-specific primer design experiments by including 47 representative heavy chain fragments from naive B cells. Each fragment includes a different functional complete heavy chain V gene, its respective full leader sequence and a partial constant region sequence, cloned into a pCR4-TOPO-vector backbone. This newly designed template represents the whole human germline IGHV repertoire and thereby provided ideal conditions for detailed studies of PTPs and allowed a differentiated learning on primer-properties influencing the amplification status.

As shown in Figure 23, the comparison of individual primer set efficacies on single B cell level can be very biased, especially in the context of antigen-experienced B cells, where primers must amplify V genes which bear an unknown amount of somatic hypermutation in the primer binding region and therefor influence the amplification in different manners. A structured sequencing of each amplified V gene would be needed to detect primer-template-complementarity and thus to draw valid conclusions about physicochemical properties influencing the reaction. The IGHV gene library therefor offered the opportunity to study precisely and under standardized conditions, which primer potentially amplifies which IGHV gene. Thereby, detailed studies of PTP properties influencing the amplification status could be made (see

Table 33 and Figure 32) and information was refed to the openPrimeR design tool. Additionally, the IGHV gene library can be easily manifolded via plasmid DNA preparation (Midi- or Maxiprep) and experiments can be performed in any desired amount of replication or experiment adjustment. The library minimizes PCR bias by providing a standardized plasmid amount of 1 ng plasmid template per well and a known template sequence.

By evaluating the amplification status of 16 individual primers on each of the 47 templates in triplicates, we could demonstrate a very good reproducibility: in 89.81% (661 out of 736 valid PCRs) a positive or negative amplification result was reproduced in all three experiments. In 96.19% a positive result was reproduced at least in two out auf three experiments or was

unamplified in all three experiments. This led to the conclusion that the IGHV gene library provides a highly satisfying template-tool for an unbiased validation of primer efficacies if tests were run in triplicates. Remaining fluctuations in reproducibility might be caused by other influencing variables, such as reagent- or thermocycler-dependent factors.

A similar methodology has been presented by Carlson et al.[82] who developed a synthetic T cell receptor (TCR) repertoire template, containing a mixture of equimolar DNA molecules encoding for all 56 possible combinations of the 14 V- and four J-segments of the TRC gamma annotated by IMGT. Using the synthetic template pool, Carlsen et al. were able to quantify the amount of each synthetic template pre- and post-multiplex PCR, which allowed them to adjust primer concentrations in mPCR approaches and led to the correction of uncovered bias.[82] To our best knowledge, we are first to provide a similar template for the human IGHV gene locus. Although the IGHV gene library presents a method for multiplex PCR assay on the human IGHV repertoire, it can possibly be adopted to other immune receptor loci, for example, IGKV and IGLV, (as started already during the process of this thesis (see Figure S8 to Figure S13) and finalized by C. Kreer et al. [90])), TCR-alpha or TCR-beta and many more. Therefore, our method presented here could potentially serve as a template protocol for an unbiased, quantitative multiplex PCR library preparation that could be useful to other researchers adopting it to their experiments on different gene loci.

5.4 The impact of mismatch position and ΔG on the amplification status.

By the evaluation of more 2,820 PCRs (940 triplicates) we presented a novel PCR data set, which provides the amplification status of all combinations of 20 primers and 47 IGHV templates. The amplification statuses included in our data set are based on the evaluation of gel electrophoresis imaging, which distinguishes them crucially from conventional PCR amplification studies, as these are usually based on qPCR.

Analyzing this data set with regard to the physicochemical properties of each PTP, we could define and specify established features governing the efficacy of PCR. For example, full complementarity between primer and template sequences was generally considered crucial for the amplification status,[147] but our data could specify that the total number of mismatches within PTPs can be higher than expected. It is remarkable, that in the analysis of the IGHV data set primers binding with up to 6 mismatches still show very high amplification rates at 83.33% in the overall data set and even 100% regarding Set1 alone. Most notably, we could demonstrate that PTPs labeled *Unamplified* via gel electrophoresis are most likely the result of primer-template conformations that exhibit a high free energy of annealing, an increased number of mismatches within the 3' hexamer and/or mismatches closer to the 3' terminus (see

Figure 32 and Table 34). Thus, we could conclude, that for future primer design not the absolute number of exhibited mismatches alone, but mismatch positions in relation to the 3' terminus and the free energy of annealing should be carefully considered.

The effects of 3' mismatches to PCR amplification have been subject to several published studies,[67, 68,124] but these studies, if at all, only superficially examined the relationship of 3' hexamer mismatch position and free energy of annealing. However, many studies highlight the type of nucleotide mismatch to influence amplification[67,124,148] a feature with we neglected in our analysis. Nonetheless, our data set provided great new insights into the physicochemical factors of PCR.

5.5 Naive B cell IGHV repertoire representation across 16 pooled healthy donors

Out of 450 processed colony clones cultivated during the preparation of the IGHV gene library (see Figure 10) we obtained 264 germline IGHV sequences, representing the naive B cell repertoire of 16 pooled healthy donors. Across this dataset, we analyzed the IGHV family, gene, and allele usage.

Despite the general tendency of pathogen-specific antibody responses to exhibit biased gene repertoires, it has been stated in the past, that a distinct pattern of IGHV family usage is conserved to a high extend across naive and memory B cell subsets and across individuals, as IGHV families 1, 3 and 4 are usually highly over-expressed relative to families 2, 5, 6 and 7.[86,149–152] Even though our sample size was comparatively much smaller than previous studies[153,150,154], we could demonstrate a similar pattern of IGHV family usage across the analyzed naive B cells of 16 pooled donors (see Figure 25A). Of note, this pattern generally parallels the genomic complexity of each IGHV family and might therefore be considered suggestive (see Figure 25B). Additionally, a clear predominance in the usage of certain IGHV genes was visible among the naive B cell subset of the 16 donors. Whereas seven rare genes were not represented at all, five genes (IGHV1-18, IGHV3-23, IGHV4-34, IGHV4-59 and IGHV5-51) were clearly overexpressed in comparison to most of the other genes. Quite similar patterns of IGHV gene usage in the naive B cell repertoire have been identified before in large scale studies.[154,150] For example, IGHV3-23 has been identified to be the most common V gene in naive B cell subsets[155,156], which conforms with our findings. IGHV3-23 was found to account for 8.33% of all V genes across the analyzed naive B cells of our donors (Mean x -2.13%).

Although the process of selecting gene segments for VDJ-recombination is known to be randomly, some genes are generally used more frequently than others. Common biases in the

gene segments usage have been demonstrated to be the result of a variety of mechanisms, including preferential recombination between certain V, D and J segments[153] and bias based on the distance between the V, D, and J segments.[157] Nonetheless, we also saw remarkable differences between our subset and other previously described high throuput studies, which revealed, that each individual has a unique IGHV genotype, and striking differences were generally seen between individuals.[150] It is of great importance to push forward studies on individual and subset V gene usages, as alterations of V gene family and gene homeostasis in the circulating B cell repertoire can be an expression of disease: e.g. a progressive decrease in V gene family VH1 and VH3 expression were found in plasma cells of HIV-1-infected patients[158,159] while the overexpression of certain IGHV genes accounts as prognostic marker for B cell malignancies, such as chronic lymphatic leukemia[160,161,] and follicular lymphoma[162], or are associated with diverse autoimmune processes.[163] However, statements about donor-specific differences in V gene usage cannot be made based on our data, as we used pooled cDNA from 16 donors. Finally, we did not perform paired analysis of VH- and VL sequences which is an important requirement when studying the individual V gene usage and detect complete antibodies.

The same applies to the analysis of the allele distribution. Again, we could demonstrate biased usage of certain gene alleles among the 16 pooled individuals. It was remarkable that most gene segments of our donors only presented single common alleles, with relatively rare allelic variants, while IGHV3-11 was presented by four and IGHV1-69 by five different alleles. As described by other repertoire studies before[150,164], copy number polymorphism can be suspected here, but again, it cannot be proven based on our data collection.

To sum up, for a deeper repertoire and immune response analysis, a donor-specific analysis with paired VH and VL information would have been needed. Also, studying the different representations of the V gene usage of naive in comparison to antigen-experienced B cell repertoires would be of further interest. As this information is missing in our data set, and the amount of analyzed sequences is relatively small in comparison to other studies, which easily included 100 times the quantity[150], we are aware of the limits of our repertoire data. Nonetheless, our findings underlined our assumption, that 5'RACE provides a versatile tool for unbiased gene amplification and any additional public information on subsets of the B cell repertoire can be helpful for other researchers.

5.6 Comparison with other methods for mAbs isolation and repertoire analysis

To advance the identification of mAbs targeting infectious pathogens, the focus of this thesis was set to optimizing the here presented bait protein-based cell sorting and mPCR-based single cell amplification method(s), which have been introduced by the Nussenzweig laboratory[131,132,165] and have previously proven successfully in isolating mAbs.[127,145,166,167] Nonetheless, in the past, other approaches and combinations of different techniques have been presented to study the B-cell repertoire and to isolate mAbs. For example, combinatorial phage display libraries have also been demonstrated to successfully and rapidly generate high amounts of different antibodies,[168–171] but they bring remarkable disadvantages, such as a random, and thereby potentially artificial, assembling of heavy and light chains, while exclusive and antigen-specific combinations might not be expressed at all. Finally, they provide no versatile tool for the study of B-cell repertoire representation due to multiple selection rounds by which rare but highly specific clones might get lost.

Other methods based on single-cell isolation often rely on cost- and labor-intensive B cell cultivation to screen (by the means of enzyme-linked immunosorbent assay (ELISA)) the supernatant for mAbs.[172–174] For example, a potent mAb targeting Influenza A hemagglutinins has been discovered by pre-screening of cell culture supernatant but required the cultivation of more than 100,000 B cells.[172] This underlines its inferiority to the method of antigen-specific single B cell sorting, which we chose. Besides of the B cell isolation, the BCR amplification by RT-PCR is most critical to a successful mAb identification.

To avoid the challenges of primer design covering all variable regions and exclude primer-induced bias, other approaches use 5'RACE (rapid amplification of 5' cDNA ends) using either oligo(dT) priming or a gene-specific primer.[175,176,137] Disadvantageous, this method requires multiple steps in between reverse transcription and PCR amplification (such as mRNA, degeneration, cDNA purification and polyA-tail addition) which makes it rather laborious. Additionally, due to small quantity of mRNA present in one single cell, conventional 5'RACE is more likely performed in bulk approaches[177] than on single cells, and thus does not provide pairing analysis of VH an VL sequences either.[117]

Single-cell RT-PCR on the other hand has been broadly used in BCR amplification but due to the relatively low mean efficacy of PCR amplification in priorly published B cell isolation pipelines, which were reported 50–74% PCR efficacy for heavy- and 40–65% for light-chain sequences[165,131,145], the main focus of this thesis was set to optimizing this critical step. As discussed in section 5.1., we firstly adjusted the protocol to be less cost- and labor-intensive.

Secondly, optimal primer design on immunological templates was intensively studied and let to the presentation of the oPR(5)-IGHV primer set, which demonstrated superior performance in IGHV amplification on the IGHV gene library (see section 4.3). After finalization of this thesis, further protocol advancements have been adapted, especially in the isolation, cloning and transfection steps. Finally, Gieselmann at al. *(Co-author N. Lehnen)* could propose a protocol, which yields heavy- and light-chain recoveries of >89%, which is a tremendous achievement and the prerequisite for a high-throughput capacity protocol.[132]

In the end, it is not possible to quantify the exact impact of the results and protocol advancements presented in this thesis on the described efficacy achievements, but they certainly build the cornerstone of the protocol improvement process, treating the most critical PCR amplification step.

While we successfully presented a high-throughput isolation pipeline on single cell level, in the future, next generation sequencing (NGS) approaches might be promising techniques to analyze even higher amounts of antibodies in shorter time.

NGS has already been used successfully in B cell repertoire analysis but these approaches lack the ability to provide paired information on heavy and light chain, obtained in individual single cell level analysis.[178] DeKosky et al. recently introduced a promising new technique that allows the subsequent RT-PCR generation of variable region heavy and light chain amplicons for NGS. Here, single B cells are sequestered into emulsion droplets which contain a lysis buffer and magnetic beads to capture the mRNA.[153] This technique has been demonstrated to enable the sequencing of paired VH-VL repertoires from more than 2×10^6 B cells per experiment. The technique could provide a 97% precision of pairing,[153] but unfortunately still entails a high loss of starting material and a biased mPCR step.

Further investigations are still needed to combine high-throughput NGS techniques with the precision of single cell RT-PCR, including a direct vector cloning protocol in the future.

5.7   Contribution to further publications and outlook

The results of this thesis contributed to a great extent to several subsequent research findings regarding single cell amplification techniques to identify and isolate highly mutated antibodies against infectious pathogens. In particular the work informed on the following aspects:

(i.) By learning from the here presented continuous *in vitro* testing of different primer sets on single human B cells and the specifically designed IGHV gene library, Matthias Döring and

5 Discussion

Christoph Kreer were able to train and adjust the restriction constraints in the openPrimeR design tool. In addition, Matthias Döring used the exclusive information provided by the IGHV data set of 940 PTPs to develop a "thermodynamic mismatch model (TMM)"[60] which predicts amplification of any given template by a specific primer. The model showed superior performance in comparison to priorly established models[179] and has already been proven to be a promising model for optimized primer design.[60] It can be retrieved freely via openPrimeR. The *Taq* PCR based IGHV data set presented within this thesis builds the fundament of the developed TMM. Therefore, it contributed significantly to the advancement of the openPrimeR design tool. These findings presented new insights into thermodynamic processes of PCR amplification, which will facilitate the design of future primer sets and might improve amplification and isolation outcomes on any (immunological) template.

(ii.) In a continuous loop of in vitro primer testing and in silico adjustments of starting conditions and filtering values, we were able to optimize the default settings in the openPrimeR tool. By evolving the openPrimeR design tool, the Klein Lab (Kreer et al.) was able to present an advanced version of the oPR(5)-IGHV primer set, which is published under the name oPR-IGHV and was complemented by corresponding primer sets for the light chain amplification: oPR-IGKV and oPR-IGLV primer sets (see supplement Table S4).[90] The new oPR-IGHV mix already demonstrated a significantly better performance of HIV-1-reactive B cells compared to the established primers of Set1 and Set2. Of note, C. Kreer et al. described, that the oPR-IGHV showed significantly higher amplification of V gene sequences with less than 70% V gene identity in comparison to by Set1 and Set2.[90] This is of particular value, as especially HIV-1 specific broadly neutralizing antibodies often show up to 30% sequence divergence compared with their germline precursor[42,82] and are therefore more likely to be identified by the oPR-IGHV primer set than by Set1 or Set2.[90] In conclusion, the openPrimeR tool demonstrated a highly satisfying primer design function for primer sets targeting highly divers template sequences, as they are needed in the identification of broad neutralizing antibodies. The tool may help many researchers conducting research on immunological templates overcome the obstacles encountered in designing primers for a variety of different templates.

(iii.) The protocol adjustments for the immunoglobulin gene amplification addressed in this thesis built the cornerstone for the development of a protocol for the *"effective high-throughput isolation of fully human antibodies targeting infectious pathogens" by Gieselmann et al*[132], which facilitates the production of hundreds of mAbs within less than two weeks. The validated protocol improvements of this thesis, such as the most efficient enzyme choices (reverse transcriptase and polymerase), down-scaling of reagent volumes, and many handling advices, as well as the newly developed oPR-primer sets, were implemented 1:1 in the final protocol

published by Gieselmann et al. (Co-author N. Lehnen) in nature protocols in 2021.[132] The Gieselmann protocol has been extensively validated and proven successful in the isolation of antigen-specific antibodies against multiple viruses, including HIV-1[127], human cytomegalovirus[180], Ebola virus[181] and SARS-CoV-2[166]. Moreover, the protocol may be applicable to studies of the BCR in a variety of diseases, including B cell malignancies such as lymphoma or autoimmune diseases.

An additional outstanding achievement under the usage of our protocol and the oPR-primer mixes was the detection and isolation of the broad neutralizing antibody (1-18), that efficiently targets the CD4 binding site of HIV-1 and was published by Schommers et al. in Cell in 2020.[127] In comparison to other potent CD4 binding site antibodies,[55,182,183] 1-18 restricts the development of viral escape and effectively suppresses HIV-1 in vivo.[127] The 1-18 bNAB is therefore a very promising candidate for future options of highly effective treatment and prevention of HIV-1 infection.

In summary, the results of this thesis have contributed significantly to the Klein laboratory's development of a highly effective protocol for the amplification of broad neutralizing antibodies, which has been proven successful in the detection of clinically relevant therapeutic candidates.

As vividly exemplified by the ongoing SARS-CoV-2 pandemic and the HIV-1 epidemic, efficient and globally available treatment strategies are (after primary prevention) of utmost importance. The isolation of monoclonal antibodies targeting these pathogens can provide promising options to fight these conditions.[51,127,] Fortunately, extensive research investigations are advancing isolation strategies each and every day and promising techniques such as high-throughput NGS, 5'RACE, and microencapsulation techniques are on the horizon. Until the implementation of these techniques into standardization, the here provided advanced single-cell and PCR based approach remains of tremendous value for research and clinical applications.

6 BIBLIOGRAPHY

1   Parkin J, Cohen B. An overview of the immune system. *Lancet* 2001;**357**(9270):1777-1789.

2   Dobson AP, Robin Carper E. Infectious diseases and human population history. *BioScience* 1996; **46**: 115–26.

3   WHO.int The Top 10 causes of death. December 9, 2020

4   Maartens G, Celum C, Lewin SR. HIV infection: Epidemiology, pathogenesis, treatment, and prevention. *Lancet* 2014; **384**: 258–71.

5   Fauci AS, Lane HC, Redfield RR. Covid-19 — Navigating the Uncharted. *N Engl J Med* 2020; **382:**1268-1269

6   Jenner E. An inquiry into the causes and effects of the variolæ vaccinæ, a disease discovered in some of the western counties of England, particularly Gloucestershire, and known by the name of the cow pox. London, Printed for the author, by S. Low, and sold by Law et. [Pdf] Retrieved from the Library of Congress, https://www.loc.gov/item/07036460/.

7   Seifert M, Küppers R. Human memory B cells. *Leukemia* 2016; **30**(12): 2283–92.

8   Oldstone MBA. Viruses, Plagues, and History Past, Present, and Future, 2nd edn. New York: Oxford University Press, 2020.

9   Diskin R, Klein F, Horwitz JA, et al. Restricting HIV-1 pathways for escape using rationally designed anti-HIV-1 antibodies. *J Exp Med* 2013 ;**210**(6):1235-1249.

10   Burrell CJ, Howard CR, Murphy FA. Fenner and White's Medical Virology (Fifth Edition), Chapter 5- The Innate Immunity. 2017. Ebook, published by *Academic Press, Elsevier*, London, United Kingdom

11   Corey L, Gilbert PB, Juraska M, et al. Two Randomized Trials of Neutralizing Antibodies to Prevent HIV-1 Acquisition. *N Engl J Med* 2021; 384: 1003–1014.

12   Gonźalez-Navajas JM, Gilbert PB, Raz E. The immediate protective response to microbial challenge. *Eur J Immunol* 2014;44**(9)**:2536-2549.

13   Budeus B, De Reynoso SS, Przekopowitz M, et al. Complexity of the human memory B-cell compartment is determined by the versatility of clonal diversification in germinal centers. *Proc Natl Acad Sci U S A* 2015;112(38):E5281-E5289.

14   Defrance T, Taillardet M, Genestier L. T cell-independent B cell memory. *Curr Opin Immunol* 2011;23(3):330-336.

15   Klein U, Dalla-Favera R. Germinal centres: Role in B-cell physiology and malignancy. *Nat Rev Immunol* 2008;8(1):22-33.

16   Jacob J, Kelsoe G. In situ studies of the primary immune response to (4-hydroxy-3-nitrophenyl)acetyl. II. A Common clonal origin for periarteriolar lymphoid sheath-associated foci and germinal centers. *J Exp Med* 1992;176(3):679-687.

17   Shulman Z, Gitlin S, Targ M, et al. T Follicular Helper Cell Dynamics in Germinal Centers. *Science*. 2013;341(6146):673-677.

18   Lefranc MP, Lefranc G. Immunoglobulins or antibodies: IMGT® bridging genes, structures and functions. *Biomedicines* 2020; 8(9):319.

19   Georgiou G, Ippolito GC, Beausang J. *et al.* The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat Biotechnol* 2014; **32**: 158–168.

20   Rajewsky N. Clonal selection and learning in the antibody system. *Nature* 1996; 381(6585):751-758

21   Xu JL, Davis MM. Diversity in the CDR3 Region of V H Is Sufficient for Most Antibody Specificities. Immunity 2000;13(1):37-45

22   Ippolito GC, Schelonka RL, Zemlin M, et al. Forced usage of positively charged amino acids in immunoglobulin CDR-H3 impairs B cell development and antibody production. *J Exp Med* 2006; 203(6): 1567–1578.

23   Klein F, Diskin R, Scheid JF, et al. Somatic mutations of the immunoglobulin framework are generally required for broad and potent HIV-1 neutralization. *Cell.* 2013;153(1):126-138.

24   Janeway CA Jr, Travers P, Walport M, et al. Immunobiology: The Immune System in Health and Disease. 5th edition. New York: Garland Science; 2001. Summary to Chapter 3. Available from: https://www.ncbi.nlm.nih.gov/books/NBK27116/

25   Burton DR. Antibodies, viruses and vaccines. Nat Rev Immunol 2002;**2**(9):706-713.

26   Plotkin SA. Correlates of protection induced by vaccination. *Clin Vaccine Immunol* 2010; 17(7): 1055–1065.

27   Du L, Yang Y, Zhang X. Neutralizing antibodies for the prevention and treatment of COVID-19. Cell Mol Immunol 2021; 18(10):2293-2306

28   Fan P, Chi X, Liu G, et al. Potent neutralizing monoclonal antibodies against Ebola virus isolated from vaccinated donors. *MAbs* 2020; 12(1): 1742457.

29   Schroeder HW, Hillson JL, Perlmutter RM. Structure and evolution of mammalian VH families. *Int Immunol.* 1990;**2**(1):41-50.

30   Rechavi G, Ram D, Glazer L, Zakut R, Givol D. Evolutionary aspects of immunoglobulin heavy chain variable region (V(H)) gene subgroups. *Proc Natl Acad Sci U S A* 1983;80(3):855-859.

31   Lefranc MP. WHO-IUIS nomenclature subcommittee for immunoglobulins and T cell receptors report. Immunogenetics 2007;**59**(12):899-902.

32   Lefranc MP. Immunoglobulin and T cell receptor genes: IMGT® and the birth and rise of immunoinformatics. *Front Immunol* 2014;**5**: 1–22.

33   Lefranc M. Immunoglobulins or Antibodies : IMGT ® Bridging Genes , Structures and Functions. *Biomedicines.* 2020 Sep; 8(9): 319.

34   Matsuda F, Ishii K, Bourvagnet P, et al. The complete nucleotide sequence of the human immunoglobulin heavy chain variable region locus. *J Exp Med* 1998;188(11):2151-2162.

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

36   Maul RW, Gearhart PJ. AID and Somatic Hypermutation. *Adv Immunol* 2010; 105: 159–191.

37   Pilzecker B, Jacobs H. Mutating for good: DNA damage responses during somatic hypermutation. *Front Immunol* 2019; 10: 1–13.

38   Martin A, Scharff MD. AID and mismatch repair in antibody diversification. *Nat Rev Immunol* 2002; **2**(8):605-614.

39   Rada C, González-Fernández A, Jarvis JM, Milstein C. The 5′ boundary of somatic hypermutation in a Vχ gene is in the leader intron. *Eur J Immunol* 1994;**24**(6):1453-1457.

40   Stavnezer J, Guikema JE, Schrader CE. Mechanism and regulation of class switch recombination. *Annu Rev Immunol* 2008; **26**:261-292.

41   Janeway CA, Travers P, Walport M, Shlomchik M. Chapter 4-15. The immunoglobulin heavy-chain isotypes are distinguished by the structure of their constant regions. In: Immunobiology: The Immune System in Health and Disease, 5th Edition. New York: Garland Science Publishing, 2001: 168–70.

42   Wardemann H, Yurasov S, Schaefer A, Young JW, Meffre E, Nussenzweig MC. Predominant autoantibody production by early human B cell precursors. *Science*. 2003; 301(5638):1374-1377.

43   Nemazee D. Receptor editing in lymphocyte development and central tolerance. *Nat Rev Immunol* 2006;**6**(10):728-740.

44   Yurasov S, Nussenzweig MC. Regulation of autoreactive antibodies. Curr Opin Rheumatol 2007;**19(**5):421-426.

45   Klein F, Mouquet H, Dosenovic P, Scheid JF, Scharf L, Nussenzweig MC. Antibodies in HIV-1 vaccine development and therapy. *Science*. 2013;**341**(6151):1199-1204.

46   Doria-Rose NA, Klein RM, Manion MM, et al. Frequency and Phenotype of Human Immunodeficiency Virus Envelope-Specific B Cells from Patients with Broadly Cross-Neutralizing Antibodies. *J Virol* 2009;**83(**1):188-199.

47   Simek MD, Rida W, Priddy FH, et al. Human Immunodeficiency Virus Type 1 Elite Neutralizers: Individuals with Broad and Potent Neutralizing Activity Identified by Using a High-Throughput Neutralization Assay together with an Analytical Selection Algorithm. *J Virol* 2009;**83**(14):7337-7348.

48   Kaur K, Sullivan M, Wilson PC. Targeting B cell responses in universal influenza vaccine design. *Trends Immunol*. 2011;**32**(11):524-531.

49   Gray ES, Madiga MC, Hermanus T, et al. The neutralization breadth of HIV-1 develops incrementally over four years and is associated with CD4+ T cell decline and high viral load during acute infection. *J Virol*. 2011;**85**(10):4828-4840.

50   Grilo AL, Mantalaris A. The Increasingly Human and Profitable Monoclonal Antibody Market. *Trends Biotechnol*. 2019;**37**(1):9-16.

51   Kaplon H, Chenoweth A, Crescioli S, Reichert JM. Antibodies to watch in 2022. *MAbs*. 2022;**14**(1):2014296.

52   COVID-19 Treatment Guidelines Panel. Coronavirus Disease 2019 (COVID-19) Treatment Guidelines. National Institutes of Health. Available at https://www.covid19treatmentguidelines.nih.gov/. Accessed 02/02/2022.

53   Corey L, Gilbert PB, Juraska M, et al. Two Randomized Trials of Neutralizing Antibodies to Prevent HIV-1 Acquisition. N Engl J Med 2021; 384: 1003–14.

54   Lynch RM, Boritz E, Coates EE, et al. Virologic effects of broadly neutralizing antibody VRC01 administration during chronic HIV-1 infection. *Sci Transl Med*. 2015;**7**(319):319ra206.

55   Caskey M, Klein F, Lorenzi JC, et al. Viraemia suppressed in HIV-1-infected humans by broadly neutralizing antibody 3BNC117 [published correction appears in Nature. 2016 Jul 28;**535**(7613):580]. *Nature*. 2015;**522**(7557):487-491.

56   Caskey M, Schoofs T, Gruell H, et al. Antibody 10-1074 suppresses viremia in HIV-1-infected individuals. *Nat Med*. 2017;**23**(2):185-191.

57   Gautam R, Nishimura Y, Pegu A, et al. A single injection of anti-HIV-1 antibodies protects against repeated SHIV challenges. *Nature*. 2016;**533**(7601):105-109.

58   Shingai M, Donau OK, Plishka RJ, et al. Passive transfer of modest titers of potent and broadly neutralizing anti-HIV monoclonal antibodies block SHIV infection in macaques. *J Exp Med* 22 September 2014; 211 (10): 2061–2074.

59   Mullis KB. The unusual origin of the polymerase chain reaction. *Sci Am*. 1990;**262**(4):56-65.

60   Döring M, Kreer C, Lehnen N. *et al.* Modeling the Amplification of Immunoglobulins through Machine Learning on Sequence-Specific Features. *Sci Rep* **9**, 10748 (2019).

61   N. Beerenwinkel *et al*., "Geno2pheno: interpreting genotypic HIV drug resistance tests," in *IEEE Intelligent Systems*, vol. 16, no. 6, pp. 35-41, Nov.-Dec. 2001

62   Filchakova O, Dossym D, Ilyas A, Kuanysheva T, Abdizhamil A, Bukasov R. Review of COVID-19 testing and diagnostic methods. *Talanta*. 2022;**244**:123409.

63   AMC Technical Brief. PCR-the polymerase chain reaction. Anal Methods 2014; **6**: 333–6.

64   Sipos R, Székely AJ, Palatinszky M, Révész S, Márialigeti K, Nikolausz M. Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targetting bacterial community analysis. *FEMS Microbiol Ecol*. 2007;60(2):341-350.

65   Whiley DM, Sloots TP. Sequence variation in primer targets affects the accuracy of viral quantitative PCR. *J Clin Virol*. 2005;34(2):104-107.

66   Klein D, Leutenegger CM, Bahula C, et al. Influence of preassay and sequence variations on viral load determination by a multiplex real-time reverse transcriptase-polymerase chain reaction for feline immunodeficiency virus. *J Acquir Immune Defic Syndr*. 2001;**26**(1):8-20.

67   Kwok S, Kellogg DE, McKinney N, et al. Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies. *Nucleic Acids Res*. 1990;**18**(4):999-1005.

68   Huang MM, Arnheim N, Goodman MF. Extension of base mispairs by Taq DNA polymerase: implications for single nucleotide discrimination in PCR. *Nucleic Acids Res*. 1992;20(17):4567-4573.

69   Lorenz TC. Polymerase chain reaction: basic protocol plus troubleshooting and optimization strategies. *J Vis Exp*. 2012;(63):e3998.

70   Yuryev A. *et al*. Predicting the success of primer extension genotyping assays using statistical modeling. *Nucleic Acids Res.* 2002; **30**, 131e–131.

71   Markoulatos P, Siafakas N, Moncany M. Multiplex polymerase chain reaction: a practical approach. *J Clin Lab Anal*. 2002;**16**(1):47-51.

72   Abram ME, Ferris AL, Shao W, Alvord WG, Hughes SH. Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. *J Virol*. 2010;84(19):9864-9878.

73   Taylor BS, Sobieszczyk ME, McCutchan FE, Hammer SM. The challenge of HIV-1 subtype diversity [published correction appears in N Engl J Med. 2008 Oct 30;359(18):1972] [published correction appears in N Engl J Med. 2008;**359**(18):1965-6]. *N Engl J Med*. 2008;**358**(15):1590-1602.

74   Bailey JR, Barnes E, Cox AL. Approaches, Progress, and Challenges to Hepatitis C Vaccine Development. *Gastroenterology*. 2019;156(2):418-430.

75   Scheid JF, Mouquet H, Ueberheide B, et al. Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science*. 2011;**333**(6049):1633-1637.

76   Muellenbeck MF, Ueberheide B, Amulic B, et al. Atypical and classical memory B cells produce Plasmodium falciparum neutralizing antibodies. *J Exp Med*. 2013;**210**(2):389-399.

77   Galson JD, Clutterbuck EA, Trück J, et al. BCR repertoire sequencing: different patterns of B-cell activation after two Meningococcal vaccines. *Immunol Cell Biol*. 2015;**93**(10):885-895.

78   Pappas L, Foglierini M, Piccoli L, et al. Rapid development of broadly influenza neutralizing antibodies through redundant mutations. *Nature*. 2014;516(7531):418-422.

79   Keizer RJ, Huitema AD, Schellens JH, Beijnen JH. Clinical pharmacokinetics of therapeutic monoclonal antibodies. *Clin Pharmacokinet*. 2010;49(8):493-507.

80   Asokan M, Dias J, Liu C, et al. Fc-mediated effector function contributes to the in vivo antiviral effect of an HIV neutralizing antibody. *Proc Natl Acad Sci U S A*. 2020;**117**(31):18754-18763.

81   Tiller T, Meffre E, Yurasov S, Tsuiji M, Nussenzweig MC, Wardemann H. Efficient generation of monoclonal antibodies from single human B cells by single cell RT-PCR and expression vector cloning [published correction appears in J Immunol Methods. 2008 May 20;334(1-2):142]. *J Immunol Methods*. 2008;**329**(1-2):112-124.

82   Carlson, C., Emerson, R., Sherwood, A. *et al.* Using synthetic templates to design an unbiased multiplex PCR assay. *Nat Commun* **4**, 2680 (2013).

83   Feige U. A Threshold of ln n for Approximating Set Cover. *J ACM* 1998; 45: 634–652.

84   Freund NT, Scheid JF, Mouquet H, Nussenzweig MC. Amplification of highly mutated human Ig lambda light chains from an HIV-1 infected patient. *J Immunol Methods*. 2015;418:61-65.

85   Pan, W., Byrne-Steele, M., Wang, C. *et al.* DNA polymerase preference determines PCR priming efficiency. BMC Biotechnol 2014, **14** (10)

86   Ippolito GC, Hoi KH, Reddy ST, et al. Antibody repertoires in humanized NOD-scid-IL2Rγnull mice and human B cells reveals human-like diversification and tolerance checkpoints in the mouse. PLoS One 2012; **7**(4), e35497

87   Wu YC, Kipling D, Leong HS, Martin V, Ademokun AA, Dunn-Walters DK. High-throughput immunoglobulin repertoire analysis distinguishes between human IgM memory and switched memory B-cell populations. *Blood*. 2010;116(7):1070-1078.

88   Lim TS, Mollova S, Rubelt F, et al. V-gene amplification revisited - An optimised procedure for amplification of rearranged human antibody genes of different isotypes. *N Biotechnol*. 2010;**27**(2):108-117.

89   Tan YG, Wang XF, Zhang M, et al. Clonal characteristics of paired infiltrating and circulating B lymphocyte repertoire in patients with primary biliary cholangitis. *Liver Int*. 2018;**38**(3):542-552.

90   Kreer C, Döring M, Lehnen N, et al. openPrimeR for multiplex amplification of highly diverse templates. *J Immunol Methods* 2020; 480: 112752.

91   Pesole G, Liuni S, Grillo G, et al. GeneUp: a program to select short PCR primer pairs that occur in multiple members of sequence lists. *Biotechniques*. 1998;**25(**1):112-123.

92   Wang J, Li KB, Sung WK. G-PRIMER: greedy algorithm for selecting minimal primer set. *Bioinformatics*. 2004;**20**(15):2473-2475.

93   Jabado OJ, Palacios G, Kapoor V, et al. Greene SCPrimer: A rapid comprehensive tool for designing degenerate primers from multiple sequence alignments. *Nucleic Acids Res* 2006; 34(22), 6605–11.

94   Bashir A, Liu YT, Raphael BJ, Carson D, Bafna V. Optimization of primer design for the detection of variable genomic lesions in cancer. *Bioinformatics* 2007; **23**(21): 2807–2815.

95   Hysom DA, Naraghi Arani P, Elsheikh M, Carrillo AC, Williams PL, Gardner SN. Skip the alignment: Degenerate, multiplex primer and probe design using k-mer matching instead of alignments. PLoS One 2012; **7**(4): e34560.

96   Gardner SN, Jaing CJ, Elsheikh MM, Peña J, Hysom DA, Borucki MK. Multiplex degenerate primer design for targeted whole genome amplification of many viral genomes. *Adv Bioinformatics*. 2014;2014:101894.

97   Rychlik W. OLIGO 7 primer analysis software. *Methods Mol Biol*. 2007;402:35-60.

98   Giegerich R, Meyer F, Schleiermacher C. GeneFisher--software support for the detection of postulated genes. *Proc Int Conf Intell Syst Mol Biol*. 1996;4:68-77.

99   Rose TM, Schultz ER, Henikoff JG, Pietrokovski S, McCallum CM, Henikoff S. Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. *Nucleic Acids Res*. 1998;26(7):1628-1635.

100   Rose TM, Henikoff JG, Henikoff S. CODEHOP (COnsensus-DEgenerate Hybrid Oligonucleotide Primer) PCR primer design. *Nucleic Acids Res*. 2003;**31**(13):3763-3766.

101   Kämpke T, Kieninger M, Mecklenburg M. Efficient primer design algorithms. *Bioinformatics*. 2001;17(3):214-225.

102   Linhart C, Shamir R. The degenerate primer design problem*. Bioinformatics* 2002; 18: 172–81.

103   Emrich SJ, Lowe M, Delcher AL. PROBEmer: A web-based software tool for selecting optimal DNA oligos. *Nucleic Acids Res* 2003; **31**(13): 3746–3750.

104   Souvenir R, Buhler J, Stormo G, Zhang W. Selecting Degenerate Multiplex PCR Primers. In: Benson, G., Page, R.D.M. (eds) Algorithms in Bioinformatics. 2003, WABI 2003. Lecture Notes in Computer Science(), vol 2812. Springer, Berlin, Heidelberg

105   Jarman SN. Amplicon: Software for designing PCR primers on aligned DNA sequences. *Bioinformatics* 2004; **20**(10): 1644–5.

106   Huang YC, Chang CF, Chan CH, et al. Integrated minimum-set primers and unique probe design algorithms for differential detection on symptom-related pathogens. *Bioinformatics*. 2005;**21**(24):4330-4337.

107   Rachlin J, Ding C, Cantor C, Kasif S. MuPlex: Multi-objective multiplex PCR assay design. *Nucleic Acids Res* 2005; 33(2): 544–7.

108   Yamada T, Soma H, Morishita S. PrimerStation: a highly specific multiplex genomic PCR primer design server for the human genome. *Nucleic Acids Res*. 2006;34(Web Server issue):W665-W669.

109   Lee C, Wu JS, Shiue YL, Liang HL. MultiPrimer: software for multiplex primer design. *Appl Bioinformatics*. 2006;**5**(2):99-109.

110   Srivastava GP, Xu D. Genome-scale probe and primer design with PRIMEGENS. *Methods Mol Biol*. 2007;402:159-176.

111   Kalendar R, Lee D, Schulman AH. FastPCR software for PCR, in silico PCR, and oligonucleotide assembly and analysis. *Methods Mol Biol*. 2014;1116:271-302.

112   Shen Z, Qu W, Wang W, et al. MPprimer: A program for reliable multiplex PCR primer design. *BMC Bioinformatics* 2010; 11(143)

113   Chuang LY, Cheng YH, Yang CH. URPD: A specific product primer design tool. *BMC Res Notes* 2012; **5**(306)

114   O'Halloran DM. PrimerMapper: High throughput primer design and graphical assembly for PCR and SNP detection. *Sci Rep* 2016; 6: 20631,1–10.

115   Hendling M, Pabinger S, Peters K, Wolff N, Conzemius R, Barišic I. Oli2go: an automated multiplex oligonucleotide design tool. *Nucleic Acids Res*. 2018;46(W1):W252-W256.

116   Kuiken C, Korber B, Shafer RW. HIV sequence databases. *AIDS Rev*. 2003;**5**(1):52-61.

117   Ozawa T, Kishi H, Muraguchi A. Amplification and analysis of cDNA generated from a single cell by 5'-RACE: application to isolation of antibody heavy and light chain variable gene sequences from single B cells. *Biotechniques*. 2006;40(4)469–78.

118   Le Novère N. MELTING, computing the melting temperature of nucleic acid duplex. *Bioinformatics* 2002; **17**(12): 1226–1227.

119   Lorenz R, Bernhardt SH, Honer Zu Siederdissen C, et al. ViennaRNA Package 2.0. Algorithm Mol Biol 2011; 6(26)

120   Markham NR, Zuker M. UNAFold: software for nucleic acid folding and hybridization. *Methods Mol Biol*. 2008;453:3-31.

121   Wright ES, Yilmaz LS, Ram S, Gasser JM, Harrington GW, Noguera DR. Exploiting extension bias in polymerase chain reaction to improve primer specificity in ensembles of nearly identical DNA templates. *Environ Microbiol*. 2014;**16**(5):1354-1365.

122   Berkelaar, M., Eikland, K., Notebaert, P., 2004. Lp Solve: Open Source (Mixed-Integer) Linear Programming System.

123   Ye J, Ma N, Madden TL, Ostell JM. IgBLAST: an immunoglobulin variable domain sequence analysis tool. Nucleic Acids Res 2013; **41**(W1): 34–40.

124   Stadhouders R, Pas SD, Anber J, Voermans J, Mes TH, Schutten M. The effect of primer-template mismatches on the detection and quantification of nucleic acids using the 5' nuclease assay. *J Mol Diagn*. 2010;**12**(1):109-117.

125   Ayyadevara S, Thaden JJ, Shmookler Reis RJ. Discrimination of primer 3'-nucleotide mismatch by taq DNA polymerase during polymerase chain reaction. *Anal Biochem*. 2000;**284**(1):11-18.

126   Sliepen K, Medina-Ramírez M, Yasmeen A, Moore JP, Klasse PJ, Sanders RW. Binding of inferred germline precursors of broadly neutralizing HIV-1 antibodies to native-like envelope trimers. *Virology*. 2015;486:116-120.

127   Schommers P, Gruell H, Abernathy ME, et al. Restriction of HIV-1 Escape by a Highly Broad and Potent Neutralizing Antibody. *Cell*. 2020;180(3):471-489.e22.

128   Marston B, Palanichamy A, Anolik JH. B cells in the pathogenesis and treatment of rheumatoid arthritis. *Curr Opin Rheumatol*. 2010;22(3):307-315.

129   Sharonov GV, Serebrovskaya EO, Yuzhakova DV, Britanova OV, Chudakov DM. B cells, plasma cells and antibody repertoires in the tumour microenvironment. *Nat Rev Immunol*. 2020;**20**(5):294-307.

130   Giegerich R, Meyer F, Schleiermacher C. GeneFisher--software support for the detection of postulated genes. *Proc Int Conf Intell Syst Mol Biol* 1996; **4**: 68–77.

131   von Boehmer L, Liu C, Ackerman S, et al. Sequencing and cloning of antigen-specific antibodies from mouse memory B cells. *Nat Protoc*. 2016;**11**(10):1908-1923.

132   Gieselmann L, Kreer C, Ercanoglu MS, et al. Effective high-throughput isolation of fully human antibodies targeting infectious pathogens. *Nat Protoc* 2021; 16: 3639–71.

133   Mayorov VI, Rogozin IB, Adkison LR, Gearhart PJ. DNA polymerase eta contributes to strand bias of mutations of A versus T in immunoglobulin genes. *J Immunol*. 2005;**174**(12):7781-7786.

134   Zucha D, Androvic P, Kubista M, Valihrach L. Performance Comparison of Reverse Transcriptases for Single-Cell Studies. *Clin Chem*. 2020;**66**(1):217-228.

135   Küppers R, Zhao M, Hansmann ML, Rajewsky K. Tracing B cell development in human germinal centres by molecular analysis of single cells picked from histological sections. *EMBO J*. 1993;**12**(13):4955-4967.

136   Tan Y, Wang Y, Zhang M, et al. Clonal Characteristics of Circulating B Lymphocyte Repertoire in Primary Biliary Cholangitis. *J Immunol* 2016; **197**(5): 1609–1620.

137   He L, Sok D, Azadnia P, et al. Toward a more accurate view of human B-cell repertoire by next-generation sequencing, unbiased repertoire capture and single-molecule barcoding. Sci Rep 2014; 4: 21–3.

138   Scheid JF, Mouquet H, Feldhahn N, et al. A method for identification of HIV gp140 binding memory B cells in human blood. *J Immunol Methods*. 2009;**343**(2):65-67.

139   Wang Z, Raifu M, Howard M, et al. Universal PCR amplification of mouse immunoglobulin gene variable regions: the design of degenerate primers and an assessment of the effect of DNA polymerase 3' to 5' exonuclease activity. *J Immunol Methods*. 2000;**233**(1-2):167-177.

140   Krebber A, Bornhauser S, Burmester J, et al. Reliable cloning of functional antibody variable domains from hybridomas and spleen cell repertoires employing a reengineered phage display system. *J Immunol Methods*. 1997;**201**(1):35-55.

141   Koren S, Kosmac M, Colja Venturini A, Montanic S, Curin Serbec V. Antibody variable-region sequencing as a method for hybridoma cell-line authentication. *Appl Microbiol Biotechnol*. 2008;**78**(6):1071-1078.

142   Doria-Rose NA, Schramm CA, Gorman J, et al. Developmental pathway for potent V1V2-directed HIV-neutralizing antibodies. *Nature*. 2014;**509**(7498):55-62.

143   Zhou Y, Liu Z, Wang Z, et al. Single-Cell Sorting of HBsAg-Binding Memory B Cells from Human Peripheral Blood Mononuclear Cells and Antibody Cloning. *STAR Protoc*. 2020;**1**(3):100129.

144   Guselnikov SV, Belovezhets TN, Kulemzin SV, Gorchakov AA, Taranin AV. A simple way to increase recovery of the expressed *VH* and *VL* genes in single-sorted human B cells. *Biotechniques*. 2019;**67**(4):184-187.

145   Rogers TF, Zhao F, Huang D, et al. Isolation of potent SARS-CoV-2 neutralizing antibodies and protection from disease in a small animal model. *Science*. 2020;**369**(6506):956-963.

146   Zhou Y, Liu Z, Li S, et al. Enhancement versus neutralization by SARS-CoV-2 antibodies from a convalescent donor associates with distinct epitopes on the RBD. *Cell Rep*. 2021;**34**(5):108699.

147   Cha RS, Thilly WG. Specificity, efficiency, and fidelity of PCR. *PCR Methods Appl*. 1993;**3**(3):S18-S29.

148   Wu JH, Hong PY, Liu WT. Quantitative effects of position and type of single mismatch on single base primer extension. *J Microbiol Methods*. 2009;**77**(3):267-275.

149   Glanville J, Kuo TC, von Büdingen HC, et al. Naive antibody gene-segment frequencies are heritable and unaltered by chronic lymphocyte ablation. *Proc Natl Acad Sci U S A*. 2011;**108(**50):20066-20071.

150   Boyd SD, Gaëta BA, Jackson KJ, et al. Individual variation in the germline Ig gene repertoire inferred from variable region gene rearrangements. *J Immunol*. 2010;**184**(12):6986-6992.

151   Tian C, Luskin GK, Dischert KM, Higginbotham JN, Shepherd BE, Crowe JE. Immunodominance of the V H 1–46 Antibody Gene Segment in the Primary Repertoire of Human Rotavirus-Specific B Cells Is Reduced in the Memory Compartment through Somatic Mutation of Nondominant Clones . *J Immunol* 1 March 2008; **180**(5): 3279–3288.

152   Gorny MK, Wang XH, Williams C, et al. Preferential use of the VH5-51 gene segment by the human immune response to code for antibodies against the V3 domain of HIV-1. *Mol Immunol*. 2009;**46**(5):917-926.

153   DeKosky, B., Kojima, T., Rodin, A. *et al.* In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire. *Nat Med* **21**, 86–91.

154   Rubelt F, Bolen CR, McGuire HM, et al. Individual heritable differences result in unique cell lymphocyte receptor repertoires of naïve and antigen-experienced cells. *Nat Commun*. 2016;**7**:11112.

155   Briney BS, Willis JR, McKinney BA, Crowe JE Jr. High-throughput antibody sequencing reveals genetic evidence of global regulation of the naïve and memory repertoires that extends across individuals. *Genes Immun*. 2012;**13**(6):469-473.

156   Milner EC, Hufnagle WO, Glas AM, Suzuki I, Alexander C. Polymorphism and utilization of human VH Genes. *Ann N Y Acad Sci*. 1995;**764**:50-61.

157   Hansen TØ, Lange AB, Barington T. Sterile DJH rearrangements reveal that distance between gene segments on the human Ig H chain locus influences their ability to rearrange. *J Immunol*. 2015;**194**(3):973-982.

158   David D, Demaison C, Bani L, Thèze J. Progressive decrease in VH3 gene family expression in plasma cells of HIV-infected patients. *Int Immunol*. 1996;**8**(8):1329-1333.

159   Scamurra RW, Miller DJ, Dahl L, et al. Impact of HIV-1 infection on VH3 gene repertoire of naive human B cells. *J Immunol*. 2000;**164**(10):5482-5491.

160   Tobin G, Thunberg U, Johnson A, et al. Chronic lymphocytic leukemias utilizing the VH3-21 gene display highly restricted Vλ2-14 gene use and homologous CDR3s: Implicating recognition of a common antigen epitope. *Blood* 2003; **101**(12): 4952–7.

161   Ghia P, Stamatopoulos K, Belessi C, et al. Geographic patterns and pathogenetic implications of IGHV gene usage in chronic lymphocytic leukemia: The lesson of the IGHV3-21 gene. *Blood* 2005; **105**(4): 1678–85.

162   Berget E, Molven A, Løkeland T, Helgeland L, Vintermyr OK. IGHV gene usage and mutational status in follicular lymphoma: Correlations with prognosis and patient age. *Leuk Res*. 2015;**39**(7):702-708.

163   Bashford-Rogers RJM, Bergamaschi L, McKinney EF, et al. Analysis of the B cell receptor repertoire in six immune-mediated diseases. *Nature*. 2019;574(7776):122-126.

164   Mikocziova I, Gidoni M, Lindeman I, et al. Polymorphisms in human immunoglobulin heavy chain variable genes and their upstream regions. *Nucleic Acids Res*. 2020;**48**(10):5499-5510.

165   Smith K, Garman L, Wrammert J, et al. Rapid generation of fully human monoclonal antibodies specific to a vaccinating antigen. *Nat Protoc* 2009; **4**: 372–84.

166   Kreer C, Zehner M, Weber T, et al. Longitudinal Isolation of Potent Near-Germline SARS-CoV-2-Neutralizing Antibodies from COVID-19 Patients [published correction appears in Cell. 2020;**182**(6):1663-1673]. *Cell*. 2020;**182**(4):843-854.e12.

167   Joyce MG, Wheatley AK, Thomas PV, et al. Vaccine-Induced Antibodies that Neutralize Group 1 and Group 2 Influenza A Viruses. *Cell*. 2016;**166**(3):609-623.

168   Alfaleh MA, Alsaab HO, Mahmoud AB, et al. Phage Display Derived Monoclonal Antibodies: From Bench to Bedside. *Front Immunol* 2020; **11**:1986

169   Huse WD, Sastry L, Iverson SA, et al. Generation of a large combinatorial library of the immunoglobulin repertoire in phage lambda. *Science*. 1989;**246**(4935):1275-1281.

170   Fox S, Leitch AE, Duffin R, Haslett C, Rossi AG. Neutrophil apoptosis: relevance to the innate immune response and inflammatory disease. *J Innate Immun*. 2010;**2**(3):216-227.

171   McCafferty J, Griffiths AD, Winter G, Chiswell DJ. Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* 1990; 348: 552–4.

172   Corti D, Voss J, Gamblin SJ, et al. A neutralizing antibody selected from plasma cells that binds to group 1 and group 2 influenza A hemagglutinins. *Science*. 2011;**333**(6044):850-856.

173   Xiao X, Chen Y, Varkey R, et al. A novel antibody discovery platform identifies anti-influenza A broadly neutralizing antibodies from human memory B cells. *MAbs*. 2016;**8**(5):916-927.

174   Wiesner M, Zentz C, Mayr C, et al. Conditional immortalisation of human B cells by CD40 Ligation. *PLoS One* 2008; **1** e1464.

175   Ruberti F, Cattaneo A, Bradbury A. The use of the RACE method to clone hybridoma cDNA when V region primers fail. *J Immunol Methods*. 1994;**173**(1):33-39.

176   Doenecke A, Winnacker EL, Hallek M. Rapid amplification of cDNA ends (RACE) improves the PCR-based isolation of immunoglobulin variable region genes from murine and human lymphoma cells and cell lines. *Leukemia*. 1997;**11**(10):1787-1792.

177   Huang C, Stollar BD. Construction of representative immunoglobulin variable region cDNA libraries from human peripheral blood lymphocytes without in vitro stimulation. *J Immunol Methods*. 1991;**141**(2):227-236.

178   Reddy ST, Ge X, Miklos AE, et al. Monoclonal antibodies isolated without screening by analyzing the variable-gene repertoire of plasma cells. *Nat Biotechnol*. 2010;**28**(9):965-969.

179   Döring M, Kreer C, Lehnen N, Klein F, Pfeifer N. Modeling the Amplification of Immunoglobulins through Machine Learning on Sequence-Specific Features. *Sci Rep* 2019; 9: 10748.

180   Theobald SJ, Kreer C, Khailaie S, et al. Repertoire characterization and validation of gB-specific human IgGs directly cloned from humanized mice vaccinated with dendritic cells and protected against HCMV. 2020 *Plos Pathogens* **16**(7):e1008560.

181   Ehrhardt SA, Zehner M, Krähling V, et al. Polyclonal and convergent antibody response to Ebola virus vaccine rVSV-ZEBOV. *Nat Med*. 2019;**25**(10):1589-1600.

182   Wu X, Yang ZY, Li Y, et al. Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science*. 2010;**329**(5993):856-861.

183   Lynch RM, Wong P, Tran L, et al. HIV-1 fitness cost associated with escape from the VRC01 class of CD4 binding site neutralizing antibodies. *J Virol*. 2015;**89**(8):4201-4213.

184   Kreer C, Döring M, Lehnen N, et al. openPrimeR for multiplex amplification of highly diverse templates. *J Immunol Methods*. 2020;480:112752.

7 Supplements

7.1 List of figures

7 Supplements

7.2 List of tables

7.3   List of supplement figures

7.4 List of supplement tables

7.5 Supplement material

S1: Sequence identity between six human functional IGHV genes from the IMGT database

Six of the annotated 53 human functional IGHV genes (including their corresponding leader region) from the IMGT database *(standpoint 07/2016)* show complete or very high (at least 97.9%) sequence identity to another IGHV gene. Alignment was performed with the multisequence alignment tool Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/).

Gene Headers	Alignement	Position
L22582|IGHV1-69*01|Homo KC713934|IGHV1-69D*01|Homo	atggactggacctggaggttcctctttgtggtggcagcagctacaggtgtccagtcccag atggactggacctggaggttcctctttgtggtggcagcagctacaggtgtccagtcccag **	60 60
L22582|IGHV1-69*01|Homo KC713934|IGHV1-69D*01|Homo	gtgcagctggtgcagtctggggctgaggtgaagaagcctgggtcctcggtgaaggtctcc gtgcagctggtgcagtctggggctgaggtgaagaagcctgggtcctcggtgaaggtctcc **	120 120
L22582|IGHV1-69*01|Homo KC713934|IGHV1-69D*01|Homo	tgcaaggcttctggaggcaccttcagcagctatgctatcagctgggtgcgacaggcccct tgcaaggcttctggaggcaccttcagcagctatgctatcagctgggtgcgacaggcccct **	180 180
L22582|IGHV1-69*01|Homo KC713934|IGHV1-69D*01|Homo	ggacaagggcttgagtggatgggagggatcatccctatctttggtacagcaaactacgca ggacaagggcttgagtggatgggagggatcatccctatctttggtacagcaaactacgca **	240 240
L22582|IGHV1-69*01|Homo KC713934|IGHV1-69D*01|Homo	cagaagttccagggcagagtcacgattaccgcggacgaatccacgagcacagcctacatg cagaagttccagggcagagtcacgattaccgcggacgaatccacgagcacagcctacatg **	300 300
L22582|IGHV1-69*01|Homo KC713934|IGHV1-69D*01|Homo	gagctgagcagcctgagatctgaggacacggccgtgtattactgtgcgagaga gagctgagcagcctgagatctgaggacacggccgtgtattactgtgcgaga-- **	353 351

Figure S1: Sequence alignment of leader and variable region of IGHV1-69*01 and IGHV1-69D*01

IGHV1-69*01 and IGHV1-69D*01 have 100% leader and V gene identity, except for two additional nucleotides at the sequence end of IGHV1-69*01

Legend: Star (*) indicates correct alignment, colon (:) indicates the exchange of adenine with thymine or vice versa, single point (●) indicates exchange of adenine with cytosine or vice versa or exchange of adenine with guanine and vice versa. Space character ( ) indicates no alignment at all due to missing nucleic acid or exchange of guanine with cytosine or cytosine with thymine and vice versa. Bar (-) indicates no alignment due to sequence ending.

Gene Headers	Alignement	Position
X92238|IGHV2-70*03|Homo KC713935|IGHV2-70D*04|Homo	atggacatactttgttccacgctcctgctactgactgtcccgtcctgggtcttatcccag atggacatactttgttccacgctcctgctactgactgtcccgtcctgggtcttatcccag **	60 60
X92238|IGHV2-70*03|Homo KC713935|IGHV2-70D*04|Homo	gtcaccttgaaggagtctggtcctgcgctggtgaaacccacacagaccctcacactgacc gtcaccttgaaggagtctggtcctgcgctggtgaaacccacacagaccctcacactgacc **	120 120
X92238|IGHV2-70*03|Homo KC713935|IGHV2-70D*04|Homo	tgcaccttctctgggttctcactcagcactagtggaatgcgtgtgagctggatccgtcag tgcaccttctctgggttctcactcagcactagtggaatgcgtgtgagctggatccgtcag **	180 180
X92238|IGHV2-70*03|Homo KC713935|IGHV2-70D*04|Homo	cccccagggaaggccctggagtggcttgcacgcattgattgggatgatgataaattctac cccccagggaaggccctggagtggcttgcacgcattgattgggatgatgataaattctac **	240 240
X92238|IGHV2-70*03|Homo KC713935|IGHV2-70D*04|Homo	agcacatctctgaagaccaggctcaccatctccaaggacacctccaaaaaccaggtggtc agcacatctctgaagaccaggctcaccatctccaaggacacctccaaaaaccaggtggtc **	300 300
X92238|IGHV2-70*03|Homo KC713935|IGHV2-70D*04|Homo	cttacaatgaccaacatggaccctgtggacacggccgtgtattactg----------- cttacaatgaccaacatggaccctgtggacacagccacgtattactgtgtgcacggatac ***.***. *********	347 358

Figure S2: Sequence alignment of leader and variable region of IGHV2-70*03 and IGHV2-70D*04

IGHV2-70*03 and IGHV2-70D*04 have 100% leader identity and 100% V gene identity until position 333.

Legend: Star (*) indicates correct alignment, colon (:) indicates the exchange of adenine with thymine or vice versa, single point (•) indicates exchange of adenine with cytosine or vice versa or exchange of adenine with guanine and vice versa. Space character ( ) indicates no alignment at all due to missing nucleic acid or exchange of guanine with cytosine or cytosine with thymine and vice versa. Bar (-) indicates no alignement due to sequence ending

Gene Headers	Alignement	Position
M35415|IGHV3-23*02|Homo AC244492|IGHV3-23D*01|Homo	atggagtttgggctgagctggcttttttcttgtggctattttaaaaggtgtccagtgtgag atggagtttgggctgagctggcttttttcttgtggctattttaaaaggtgtccagtgtgag **	60 60
M35415|IGHV3-23*02|Homo AC244492|IGHV3-23D*01|Homo	gtgcagctgttggagtctgggggaggcttggtacagcctggggggtccctgagactctcc gtgcagctgttggagtctgggggaggcttggtacagcctggggggtccctgagactctcc **	120 120
M35415|IGHV3-23*02|Homo AC244492|IGHV3-23D*01|Homo	tgtgcagcctctggattcacctttagcagctatgccatgagctgggtccgccaggctcca tgtgcagcctctggattcacctttagcagctatgccatgagctgggtccgccaggctcca **	180 180
M35415|IGHV3-23*02|Homo AC244492|IGHV3-23D*01|Homo	gggaaggggctggagtgggtctcagctattagtggtagtggtggtagcacatactacgga gggaaggggctggagtgggtctcagctattagtggtagtggtggtagcacatactacgca ** *	240 240
M35415|IGHV3-23*02|Homo AC244492|IGHV3-23D*01|Homo	gactccgtgaagggccggttcaccatctcaagagacaattccaagaacacgctgtatctg gactccgtgaagggccggttcaccatctccagagacaattccaagaacacgctgtatctg *****************************.******************************	300 300
M35415|IGHV3-23*02|Homo AC244492|IGHV3-23D*01|Homo	caaatgaacagcctgagagccgaggacacggccgtatattactgtgcgaaaga caaatgaacagcctgagagccgaggacacggccgtatattactgtgcgaaaga ***	353 353

Figure S3: Sequence alignment of leader and variable region of IGHV3-23*02 and IGHV3-23D*01

IGHV3-23*02 and IGHV3-23D*01 have 99.3% sequence identity, with first non-conformity at position 239

Legend: Star (*) indicates correct alignment, colon (:) indicates the exchange of adenine with thymine or vice versa, single point (•) indicates exchange of adenine with cytosine or vice versa or exchange of adenine with guanine and vice versa. Space character ( ) indicates no alignment at all due to missing nucleic acid or exchange of guanine with cytosine or cytosine with thymine and vice versa. Bar (-) indicates no alignement due to sequence ending

Gene Headers	Alignement	Position
X92214|IGHV3-30*18|Homo AC244456|IGHV3-30-5*01|Homo	atggagtttgggctgagctgggttttcctcgttgctcttttaagaggtgtccagtgtcag atggagtttgggctgagctgggttttcctcgttgctcttttaagaggtgtccagtgtcag **	60 60
X92214|IGHV3-30*18|Homo AC244456|IGHV3-30-5*01|Homo	gtgcagctggtggagtctggggggaggcgtggtccagcctgggaggtccctgagactctcc gtgcagctggtggagtctggggggaggcgtggtccagcctgggaggtccctgagactctcc **	120 120
X92214|IGHV3-30*18|Homo AC244456|IGHV3-30-5*01|Homo	tgtgcagcctctggattcaccttcagtagctatggcatgcactgggtccgccaggctcca tgtgcagcctctggattcaccttcagtagctatggcatgcactgggtccgccaggctcca **	180 180
X92214|IGHV3-30*18|Homo AC244456|IGHV3-30-5*01|Homo	ggcaaggggctggagtgggtggcagttatatcatatgatggaagtaataaatactatgca ggcaaggggctggagtgggtggcagttatatcatatgatggaagtaataaatactatgca **	240 240
X92214|IGHV3-30*18|Homo AC244456|IGHV3-30-5*01|Homo	gactccgtgaagggccgattcaccatctccagagacaattccaagaacacgctgtatctg gactccgtgaagggccgattcaccatctccagagacaattccaagaacacgctgtatctg **	300 300
X92214|IGHV3-30*18|Homo AC244456|IGHV3-30-5*01|Homo	caaatgaacagcctgagagctgaggacacggctgtgtattactgtgcgaaaga caaatgaacagcctgagagctgaggacacggctgtgtattactgtgcgaaaga ***	353 353

Figure S4: Sequence alignment of leader and variable region of IGHV3-30*18 and IGHV3-30-5*01

IGHV3-30*18 and IGHV3-30-5*01 have 100% leader and V gene identity

Legend: Star (*) indicates correct alignment, colon (:) indicates the exchange of adenine with thymine or vice versa, single point (•) indicates exchange of adenine with cytosine or vice versa or exchange of adenine with guanine and vice versa. Space character ( ) indicates no alignment at all due to missing nucleic acid or exchange of guanine with cytosine or cytosine with thymine and vice versa. Bar (-) indicates no alignment due to sequence ending

Gene Headers	Alignement	Position
M99672|IGHV3-43*01|Homo KC713950|IGHV3-43D*01|Homo	atggagtttggactgagctgggttttccttgttgctattttaaaaggtgtccagtgtgaa atggagtttggactgagctgggttttccttgttgctattttaaaaggtgtccagtgtgaa **	60 60
M99672|IGHV3-43*01|Homo KC713950|IGHV3-43D*01|Homo	gtgcagctggtggagtctggggggagtcgtggtacagcctggggggtccctgagactctcc gtgcagctggtggagtctggggggagtcgtggtacagcctggggggtccctgagactctcc **	120 120
M99672|IGHV3-43*01|Homo KC713950|IGHV3-43D*01|Homo	tgtgcagcctctggattcacctttgatgattataccatgcactgggtccgtcaagctccg tgtgcagcctctggattcacctttgatgattatgccatgcactgggtccgtcaagctccg *********************************.**************************	180 180
M99672|IGHV3-43*01|Homo KC713950|IGHV3-43D*01|Homo	gggaagggtctggagtgggtctctcttattagttgggatggtggtagcacatactatgca gggaagggtctggagtgggtctctcttattagttgggatggtggtagcacctactatgca **.*********	240 240
M99672|IGHV3-43*01|Homo KC713950|IGHV3-43D*01|Homo	gactctgtgaagggccgattcaccatctccagagacaacagcaaaaactccctgtatctg gactctgtgaagggtcgattcaccatctccagagacaacagcaaaaactccctgtatctg ************** ***	300 300
M99672|IGHV3-43*01|Homo KC713950|IGHV3-43D*01|Homo	caaatgaacagtctgagaactgaggacaccgccttgtattactgtgcaaaagata caaatgaacagtctgagagctgaggacaccgccttgtattactgtgcaaaagata ******************.***********************************	355 355

Figure S5: Sequence alignment of leader and variable region of IGHV3-43*01 and IGHV3-43D*01

IGHV3-43*01 and IGHV3-43D*01 have 100% leader identity and 98.9% V gene identity, with first non-conformity at position 154

Legend: Star (*) indicates correct alignment, colon (:) indicates the exchange of adenine with thymine or vice versa, single point (•) indicates exchange of adenine with cytosine or vice versa or exchange of adenine with guanine and vice versa. Space character ( ) indicates no alignment at all due to missing nucleic acid or exchange of guanine with cytosine or cytosine with thymine and vice versa. Bar (-) indicates no alignment due to sequence ending.

Gene Headers	Alignement	Position
Z14237|IGHV4-31*03|Homo	atgaaacacctgtggttcttcctcctgctggtggcagctcccagatgggtcctgtcccag	60
Z14238|IGHV4-30-4*01|Homo	atgaaacacctgtggttcttcctcctgctggtggcagctcccagatgggtcctgtcccag	60
	**	
Z14237|IGHV4-31*03|Homo	gtgcagctgcaggagtcgggcccaggactggtgaagccttcacagaccctgtccctcacc	120
Z14238|IGHV4-30-4*01|Homo	gtgcagctgcaggagtcgggcccaggactggtgaagccttcacagaccctgtccctcacc	120
	**	
Z14237|IGHV4-31*03|Homo	tgcactgtctctggtggctccatcagcagtggtggttactactggagctggatccgccag	180
Z14238|IGHV4-30-4*01|Homo	tgcactgtctctggtggctccatcagcagtggtgattactactggagttggatccgccag	180
	******************************.************* ***************	
Z14237|IGHV4-31*03|Homo	cacccaggggaagggcctggagtggattgggtacatctattacagtgggagcacctactac	240
Z14238|IGHV4-30-4*01|Homo	cccccaggggaagggcctggagtggattgggtacatctattacagtgggagcacctactac	240
	*.**	
Z14237|IGHV4-31*03|Homo	aacccgtccctcaagagtcgagttaccatatcagtagacacgtctaagaaccagttctcc	300
Z14238|IGHV4-30-4*01|Homo	aacccgtccctcaagagtcgagttaccatatcagtagacacgtccaagaaccagttctcc	300
	*** **************	
Z14237|IGHV4-31*03|Homo	ctgaagctgagctctgtgactgccgcggacacggccgtgtattactgtgcgagaga 356	
Z14238|IGHV4-30-4*01|Homo	ctgaagctgagctctgtgactgccgcagacacggccgtgtattactgtgccagaga 356	
	****************************.*********************** *****	

Figure S6: Sequence alignment of leader and variable region of IGHV4-31*03 and IGHV4-30-4*01

IGHV4-31*03 and IGHV4-30-4*01 have 100% leader identity and 97.9% V gene identity, with first non-conformity at position 155

Legend: Star (*) indicates correct alignment, colon (:) indicates the exchange of adenine with thymine or vice versa, single point (•) indicates exchange of adenine with cytosine or vice versa or exchange of adenine with guanine and vice versa. Space character ( ) indicates no alignment at all due to missing nucleic acid or exchange of guanine with cytosine or cytosine with thymine and vice versa. Bar (-) indicates no alignement due to sequence ending
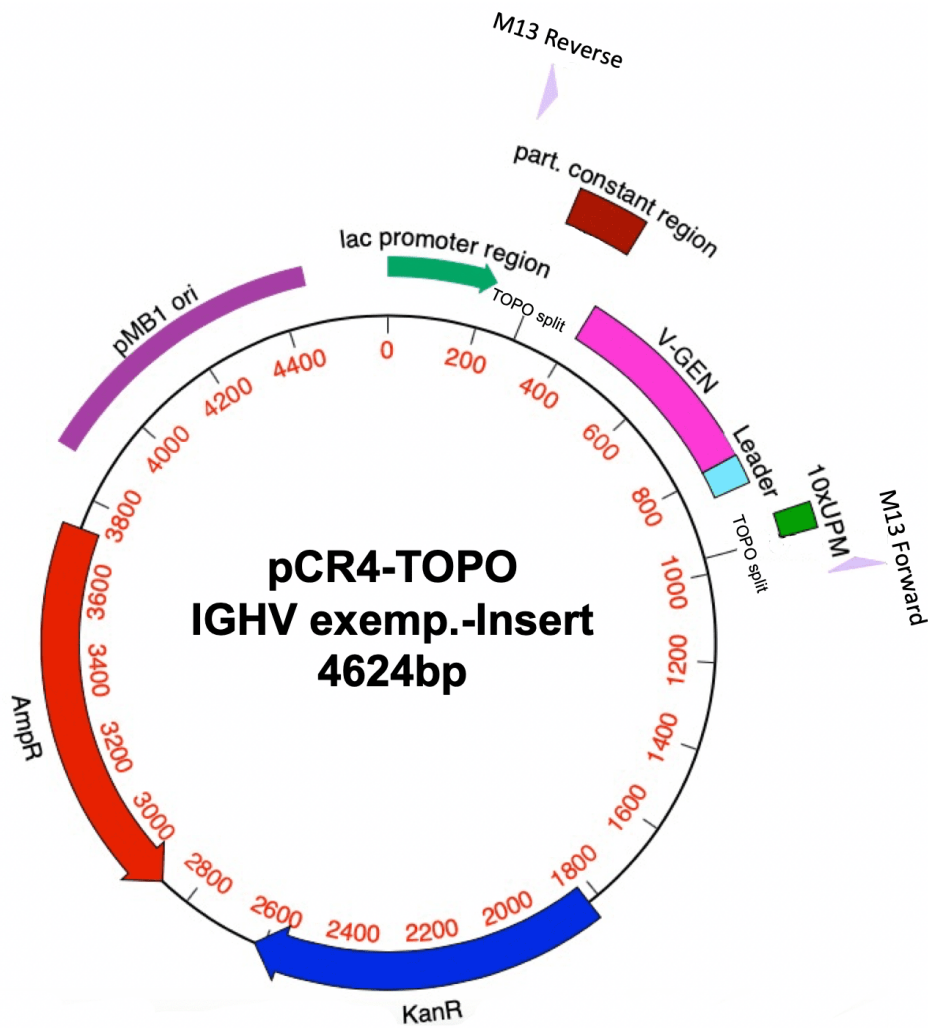
Figure S7: Exemplary plasmid map of an IGHV plus leader plus partial constant region sequence insert cloned into the pCR^{TM}4-TOPO vector

The exemplary insert of the germline IGHV1-2*01 here pictured as turquoise bar (Leader region), pink bar (V region) and brown bar (partial constant region). AmpR pictured as red arrow indicates ampicillin resistance, purple triangles indicate the M13 forward and M13 reverse primer binding sides. Plasmid maps were generated with MacVector.

Table S1: Family, gene, and allele count of 264 IGHV sequences from IGHV gene library preparation

Allele	Seq.	Allele	Seq.	Allele	Seq.	Gene	Seq.	Family	Seq.
IGHV1-2*01	0	IGHV3-13*01	3	IGHV4-28*03	0	IGHV1-2	7	IGHV1	62
IGHV1-2*02	7	IGHV3-13*02	0	IGHV4-28*07	0	IGHV1-3	11	IGHV2	8
IGHV1-2*03	0	IGHV3-13*05	0	IGHV4-30-2*01	1	IGHV1-8	3	IGHV3	110
IGHV1-2*04	0	IGHV3-15*01	4	IGHV4-30-2*03	0	IGHV1-18	16	IGHV4	59
IGHV1-3*01	11	IGHV3-15*02	0	IGHV4-30-2*06	0	IGHV1-24	6	IGHV5	19
IGHV1-3*02	0	IGHV3-20*01	0	IGHV4-30-4*01	0	IGHV1-45	0	IGHV6	5
IGHV1-8*01	3	IGHV3-20*02	0	IGHV4-30-4*02	0	IGHV1-46	8	IGHV7	1
IGHV1-18*01	12	IGHV3-21*01	6	IGHV4-30-4*03 = IGHV4-31*03	0	IGHV1-58	0		
IGHV1-18*02	0	IGHV3-21*02	0	IGHV4-30-4*04	0	IGHV1-69	11		
IGHV1-18*04	4	IGHV3-23*01	20	IGHV4-30-4*07	0	IGHV1-69-2	0		
IGHV1-24*01	6	IGHV3-23*02	0	IGHV4-31*02	0				
IGHV1-45*01	0	IGHV3-23*04	2	IGHV4-31*03	7	IGHV2-5	6		
IGHV1-45*02	0	IGHV3-30*01	0	IGHV4-31*06	0	IGHV2-26	1		
IGHV1-46*01	8	IGHV3-30*02	1	IGHV4-31*07	0	IGHV2-70	1		
IGHV1-46*02	0	IGHV3-30*03	0	IGHV4-31*08	0				
IGHV1-46*03	0	IGHV3-30*18	5	IGHV4-31*09	0	IGHV3-7	8		
IGHV1-58*01	0	IGHV3-30-3*01	3	IGHV4-31*10	0	IGHV3-9	6		
IGHV1-58*02	0	IGHV3-30-3*03	1	IGHV4-34*01	13	IGHV3-11	8		
IGHV1-69*01	6	IGHV3-33*01	2	IGHV4-34*02	0	IGHV3-13	3		
IGHV1-69*02	1	IGHV3-33*02	0	IGHV4-34*03	0	IGHV3-15	4		
IGHV1-69*03	0	IGHV3-43*01	0	IGHV4-34*04	0	IGHV3-20	0		
IGHV1-69*04	2	IGHV3-43*02	2	IGHV4-34*05	0	IGHV3-21	6		
IGHV1-69*06	1	IGHV3-43D*01= IGHV3-43*01	0	IGHV4-34*06	0	IGHV3-23	22		
IGHV1-69*10	1	IGHV3-48*01	1	IGHV4-34*07	0	IGHV3-30	6		
IGHV1-69*14	0	IGHV3-48*02	9	IGHV4-34*09	0	IGHV3-30-3	4		
IGHV1-69-2*01	0	IGHV3-48*03	2	IGHV4-34*10	0	IGHV3-33	2		
		IGHV3-49*01	0	IGHV4-34*11	0	IGHV3-43	2		
IGHV2-5*01	2	IGHV3-49*02	0	IGHV4-38-2*01	0	IGHV3-48	12		
IGHV2-5*02	4	IGHV3-49*03	1	IGHV4-38-2*02	0	IGHV3-49	3		
IGHV2-5*04	0	IGHV3-49*04	1	IGHV4-39*01	9	IGHV3-53	3		
IGHV2-5*05	0	IGHV3-49*05	1	IGHV4-39*02	0	IGHV3-64	1		
IGHV2-5*06	0	IGHV3-53*01	1	IGHV4-39*03	0	IGHV3-64-D	2		
IGHV2-5*08	0	IGHV3-53*02	2	IGHV4-39*06	0	IGHV3-66	6		
IGHV2-5*09	0	IGHV3-53*03	0	IGHV4-39*07	2	IGHV3-72	0		
IGHV2-26*01	1	IGHV3-64*01	1	IGHV4-59*01	13	IGHV3-73	1		
IGHV2-70*01	0	IGHV3-64*02	0	IGHV4-59*02	0	IGHV3-74	11		
IGHV2-70*02	0	IGHV3-64*04	0	IGHV4-59*08	3				
IGHV2-70*03	0	IGHV3-64D*06	2	IGHV4-59*10	0	IGHV4-4	7		
IGHV2-70*06	0	IGHV3-66*01	4	IGHV4-61*01	3	IGHV4-28	0		
IGHV2-70*07	0	IGHV3-66*02	1	IGHV4-61*02	1	IGHV4-30-2	1		
IGHV2-70*08	0	IGHV3-66*03	0	IGHV4-61*03	0	IGHV4-31	7		
IGHV2-70*09	0	IGHV3-66*04	1	IGHV4-61*04	0	IGHV4-34	13		
IGHV2-70*10	0	IGHV3-72*01	0	IGHV4-61*08	0	IGHV4-38	0		
IGHV2-70*11	1	IGHV3-73*01	1			IGHV4-39	11		
IGHV2-70*12	0	IGHV3-73*02	0	IGHV5-10-1*02	0	IGHV4-59	16		
IGHV2-70*13	0	IGHV3-74*01	10	IGHV5-10-1*03	1	IGHV4-61	4		
		IGHV3-74*03	1	IGHV5-51*01	9				
IGHV3-7*01	1			IGHV5-51*02	0	IGHV5-10-1	1		
IGHV3-7*02	1	IGHV4-4*01	0	IGHV5-51*03	9	IGHV5-51	18		
IGHV3-7*03	6	IGHV4-4*02	3						
IGHV3-9*01	6	IGHV4-4*03	0	IGHV6-1*01	5	IGHV6-1	5		
IGHV3-9*03	0	IGHV4-4*04	0	IGHV6-1*02	0				
IGHV3-11*01	5	IGHV4-4*07	4			IGHV7-4-1	1		
IGHV3-11*04	1	IGHV4-4*08	0	IGHV7-4-1*02	1				
IGHV3-11*05	1	IGHV4-28*01	0						
IGHV3-11*06	1	IGHV4-28*02	0						

Table S2: Family, gene, and allele count of 156 IGKV sequences from IGKV library preparation

Allele	Seq.
IGKV1-5*01	0
IGKV1-5*02	0
IGKV1-5*03	7
IGKV1-6*01	0
IGKV1-6*02	0
IGKV1-8*01	1
IGKV1-9*01	1
IGKV1-12*01	6
IGKV1-12*02	1
IGKV1-13*02	2
IGKV1-16*01	0
IGKV1-16*02	2
IGKV1-17*01	2
IGKV1-17*02	0
IGKV1-17*03	0
IGKV1-27*01	2
IGKV1-33*01	1
IGKV1-39*01	15
IGKV1D-8*01	0
IGKV1D-8*02	0
IGKV1D-8*03	0
IGKV1D-12*01	3
IGKV1D-12*02	0
IGKV1D-13*01	0
IGKV1D-13*02	0
IGKV1D-16*01	0
IGKV1D-16*02	0
IGKV1D-33*01	0
IGKV1D-39*01	0
IGKV1D-43*01	0
IGKV1-NL1*01	0

Allele	Seq.
IGKV2-24*01	0
IGKV2-28*01	9
IGKV2-29*02	0
IGKV2-30*01	2
IGKV2-30*02	1
IGKV2-40*01	0
IGKV2D-26*01	0
IGKV2D-26*03	0
IGKV2D-28*01	1
IGKV2D-29*01	2
IGKV2D-29*02	2
IGKV2D-30*01	0
IGKV2D-40*01	0

Allele	Seq.
IGKV3-11*01	24
IGKV3-11*02	0
IGKV3-15*01	19
IGKV3-20*01	35
IGKV3D-7*01	0
IGKV3D-11*01	0
IGKV3D-11*02	0
IGKV3D-11*03	0
IGKV3D-15*01	1
IGKV3D-15*02	0
IGKV3D-15*03	0
IGKV3D-20*01	2
IGKV3D-20*02	1

Allele	Seq.
IGKV4-1*01	13

Allele	Seq.
IGKV5-2*01	0

Allele	Seq.
IGKV6-21*01	1
IGKV6-21*02	0
IGKV6D-21*01	0
IGKV6D-21*02	0

Gene	Seq.
IGKV1-5	7
IGKV1-6	0
IGKV1-8	1
IGKV1-9	1
IGKV1-12	7
IGKV1-13	2
IGKV1-16	2
IGKV1-17	2
IGKV1-27	2
IGKV1-33	1
IGKV1-39	15
IGKV1D-8	0
IGKV1D-12	3
IGKV1D-13	0
IGKV1D-16	0
IGKV1D-33	0
IGKV1D-39	0
IGKV1D-43	0
IGKV1-NL1	0

Gene	Seq.
IGKV2-24	0
IGKV2-28	9
IGKV2-29	0
IGKV2-30	3
IGKV2-40	0
IGKV2D-26	0
IGKV2D-28	1
IGKV2D-29	4
IGKV2D-30	0
IGKV2D-40	0

Gene	Seq.
IGKV3-11	24
IGKV3-15	19
IGKV3-20	35
IGKV3D-7	0
IGKV3D-11	0
IGKV3D-15	1
IGKV3D-20	3

Gene	Seq.
IGKV4-1	13

Gene	Seq.
IGKV5-2	0

Gene	Seq.
IGKV6-21	1
IGKV6D-21	0

Family	Seq.
IGKV1	43
IGKV2	17
IGKV3	82
IGKV4	13
IGKV5	0
IGKV6	1

Table S3: Family, gene, and allele count of 160 IGLV sequences from IGLV library preparation

Allele	Seq.
IGLV1-31	0
IGLV1-36*01	1
IGLV1-40*01	15
IGLV1-40*02	0
IGLV1-44*01	8
IGLV1-47*01	3
IGLV1-47*02	1
IGLV1-51*01	1
IGLV1-51*02	1
IGLV2-5*01	2
IGLV2-8*01	9
IGLV2-8*02	0
IGLV2-11*01	8
IGLV2-11*02	0
IGLV2-14*01	54
IGLV2-14*02	0
IGLV2-18*02	2
IGLV2-18*03	0
IGLV2-18*04	0
IGLV2-23*01	5
IGLV2-23*02	6
IGLV2-23*03	2
IGLV3-1*01	12
IGLV3-9*02	0
IGLV3-9*03	0
IGLV3-10*01	3
IGLV3-10*02	0
IGLV3-12*02	0
IGLV3-16*01	0
IGLV3-19*01	1
IGLV3-21*01	3

Allele	Seq.
IGLV3-21*02	3
IGLV3-21*03	2
IGLV3-22*02	0
IGLV3-25*02	2
IGLV3-25*03	0
IGLV3-27*01	0
IGLV4-3*01	1
IGLV4-60*02	0
IGLV4-69*01	4
IGLV4-69*02 P	0
IGLV5-37*01	3
IGLV5-39*02	0
IGLV5-45*02	1
IGLV5-45*03	2
IGLV5-45*04	0
IGLV5-52*01	0
IGLV6-57*02	1
IGLV7-43*01	0
IGLV7-46*01	2
IGLV7-46*02	0
IGLV8-61*02	0
IGLV9-49*02	0
IGLV9-49*03	1
IGLV10-54*01	1
IGLV10-54*02	0
IGLV10-54*03	0

Gene	Seq.
IGLV1-31	0
IGLV1-36	1
1GLV1-40	15
IGLV1-44	8
IGLV1-47	4
IGLV1-51	2
IGLV2-5	2
IGLV2-8	9
IGLV2-11	8
IGLV2-14	54
IGLV2-18	2
IGLV2-23	13
IGLV3-1	12
IGLV3-9	0
IGLV3-10	3
IGLV3-12	0
IGLV3-16	0
IGLV3-19	1
IGLV3-21	8
IGLV3-22	0
IGLV3-25	2
IGLV3-27	0
IGLV4-3	1
IGLV4-60	0
IGLV4-69	4
IGLV5-37	3
IGLV5-39	0
IGLV5-45	3
IGLV5-52	0
IGLV6-57	1
IGLV7-43	0
IGLV7-46	2
IGLV8-61	0
IGLV9-49	1
IGLV10-54	1

Family	Seq.
IGLV1	30
IGLV2	88
IGLV3	26
IGLV4	5
IGLV5	6
IGLV6	1
IGLV7	2
IGLV8	0
IGLV9	1
IGLV10	1

IGKV family distribution
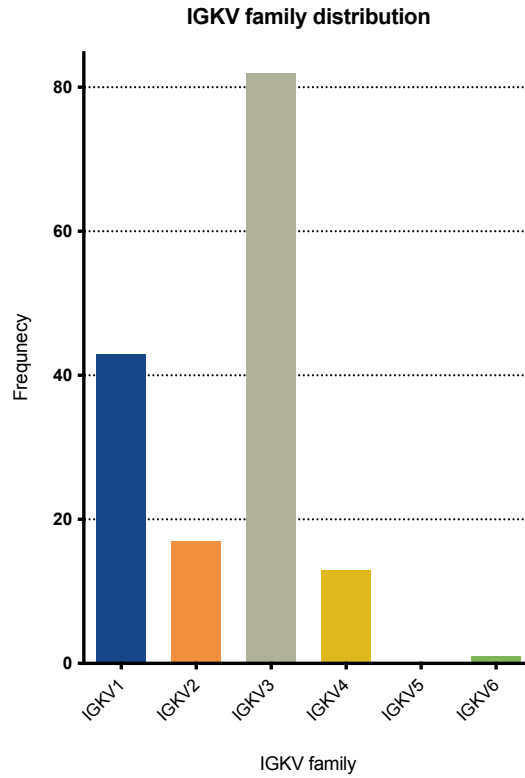
Figure S8: IGKV family distribution over 156 analyzed IGKV germline sequences (Visualization of data from Table S2)
X-axis indicates the six IGKV families, Y-axis indicates the absolute frequency over the 156 analyzed sequences from 16 pooled healthy individuals.
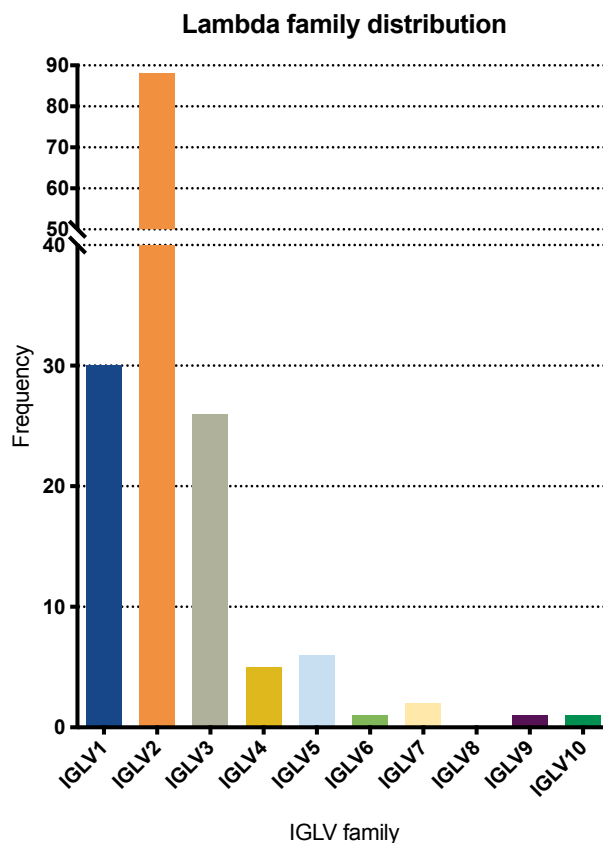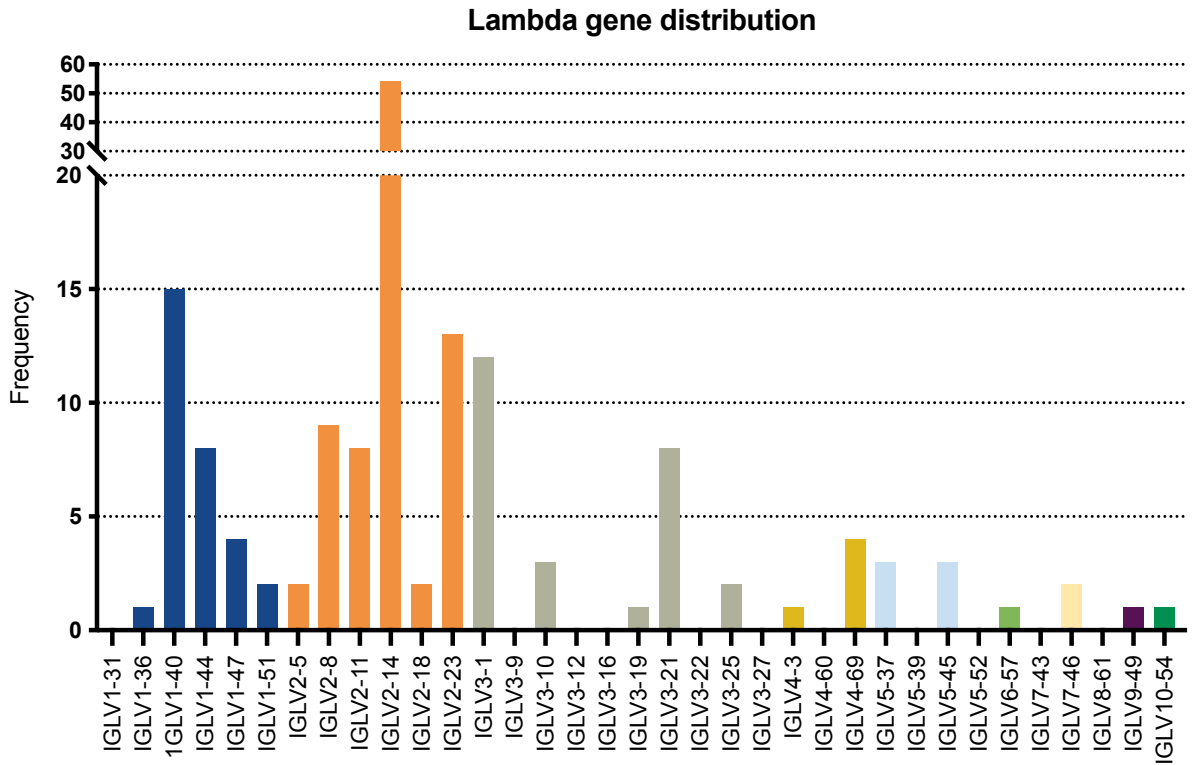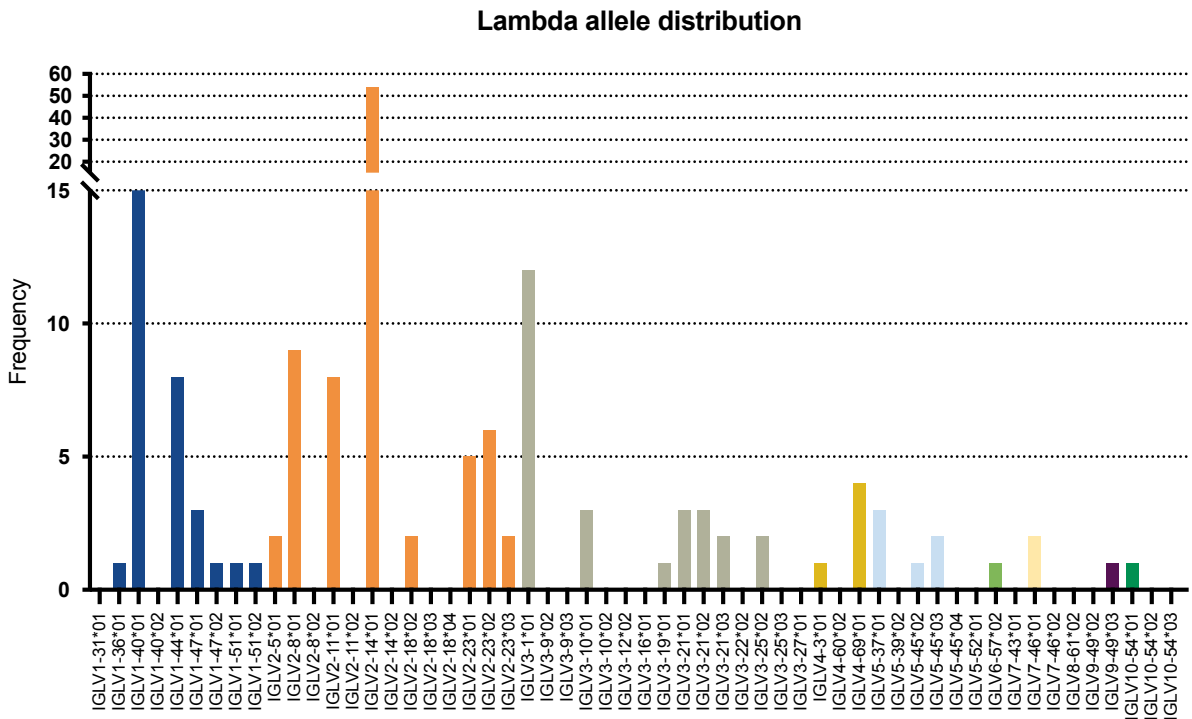
**IGKV gene distribution**
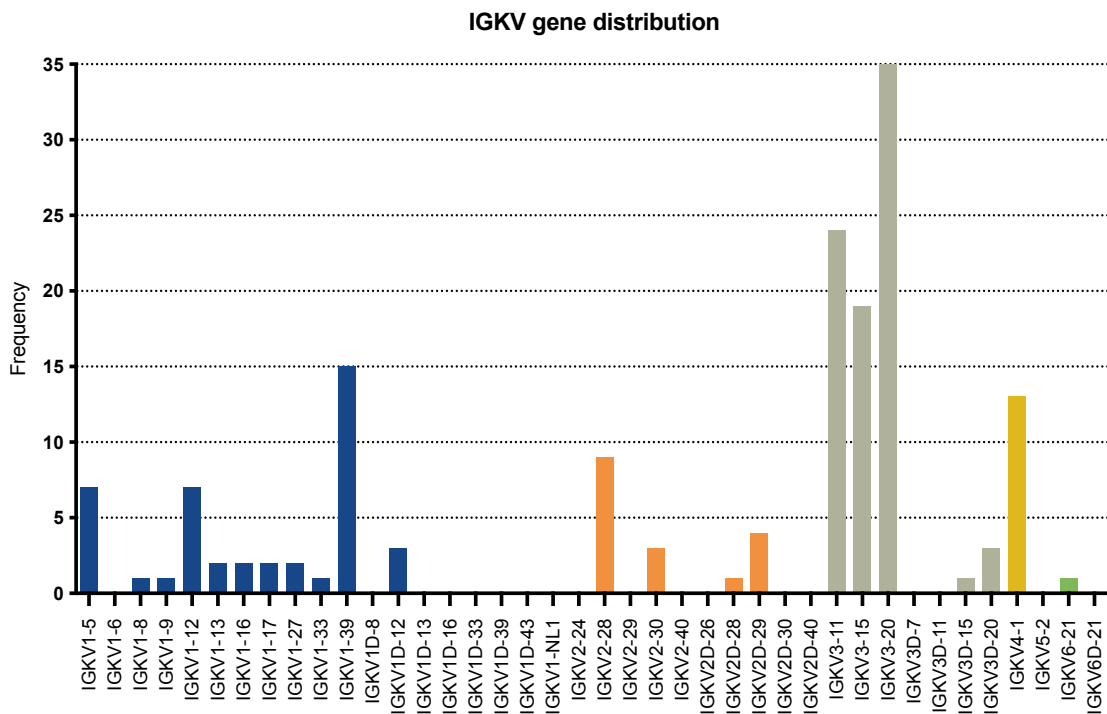
Figure S9: IGKV gene distribution over the 156 analyzed IGKV germline sequences (Visualization of data from Table S2)
X-axis indicates the 40 IGKV genes, Y-axis indicates the absolute frequency over the 156 analyzed sequences from 16 pooled healthy individuals.

IGKV allele distribution
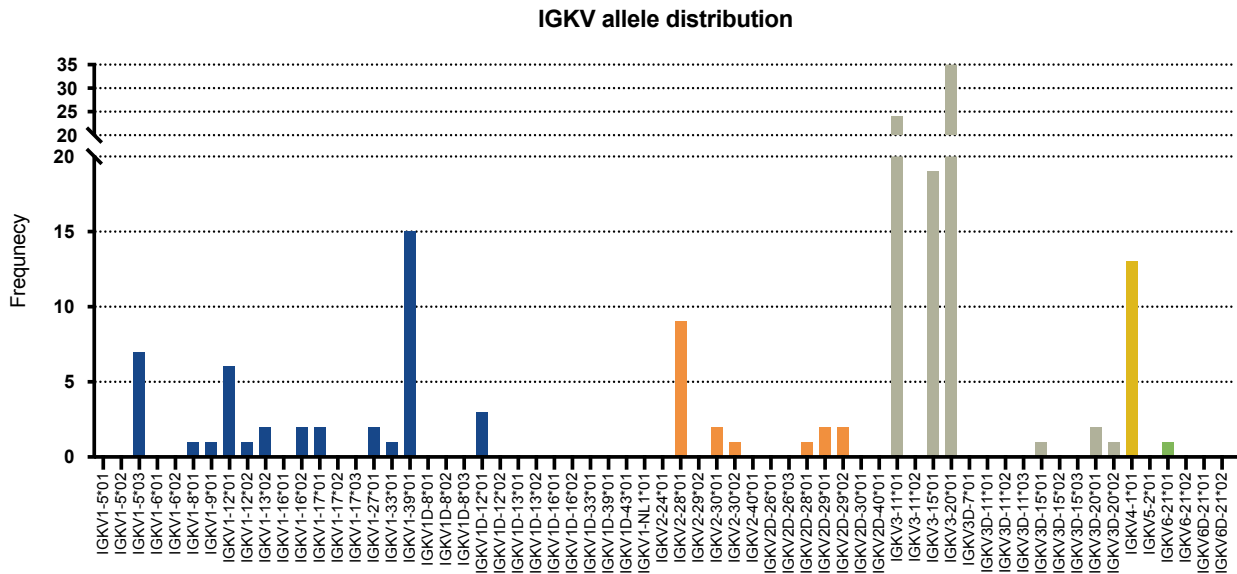
Figure S10: IGKV allele distribution over the 156 analyzed IGKV germline sequences (Visualization of data from Table S2)
X-axis indicates the 63 IGKV alleles, y-axis indicates the absolute frequency over the 156 analyzed sequences from 16 pooled healthy individuals.

Lambda family distribution

Figure S11: IGLV family distribution over 160 analyzed IGLV germline sequences (Visualization of data shown in Table S3)
X-axis indicates the ten IGLV families, y-axis indicates the absolute frequency over the 160 analyzed sequences from 16 pooled healthy individuals.

Lambda gene distribution

Figure S12: IGLV gene distribution over 160 analyzed IGLV germline sequences (Visualization of data shown in Table S3)
X-axis indicates the 35 IGLV genes, y-axis indicates the absolute frequency over the 160 analyzed sequences from 16 pooled healthy individuals.

Lambda allele distribution

Figure S13: IGLV gene distribution over 160 analyzed IGLV germline sequences (Visualization of data shown in Table S3)
X-axis indicates the 35 IGLV genes, Y-axis indicates the absolute frequency over the 160 analyzed sequences from 16 pooled healthy individuals.
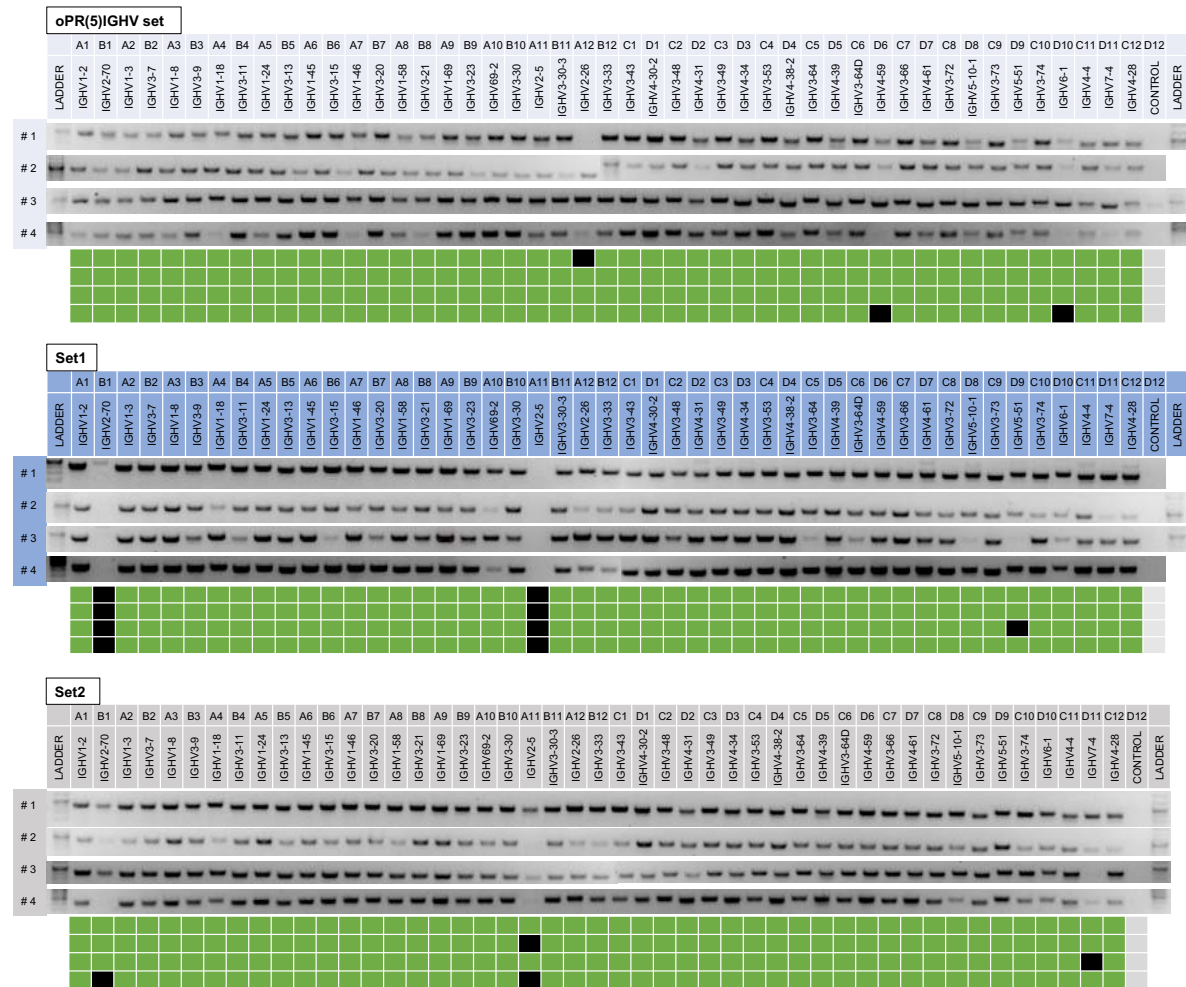
Figure S14: Validation of primer set amplification status on IGHV gene library
Primer set experiments in quadruples on the IGHV gene library. oPR(5)-IGHV set is shown in the upper row, Set1 in the middle and Set2 in the lower row. Amplification status is visualized under each gel picture and ether validated as "amplified", indicated in green, or "unamplified", indicated in black. Each last lane presents the negative control, containing the empty pCR4 vector backbone.
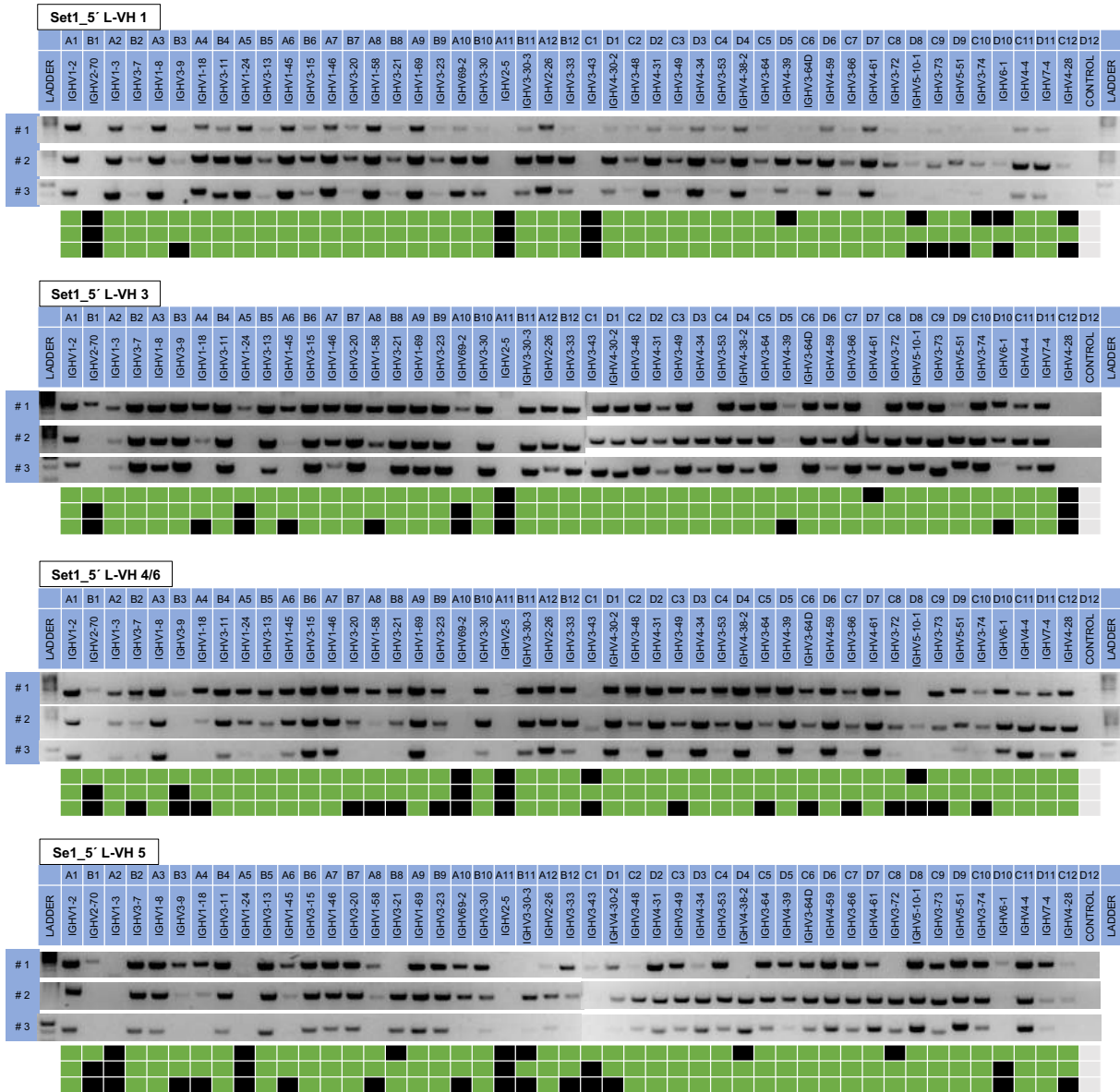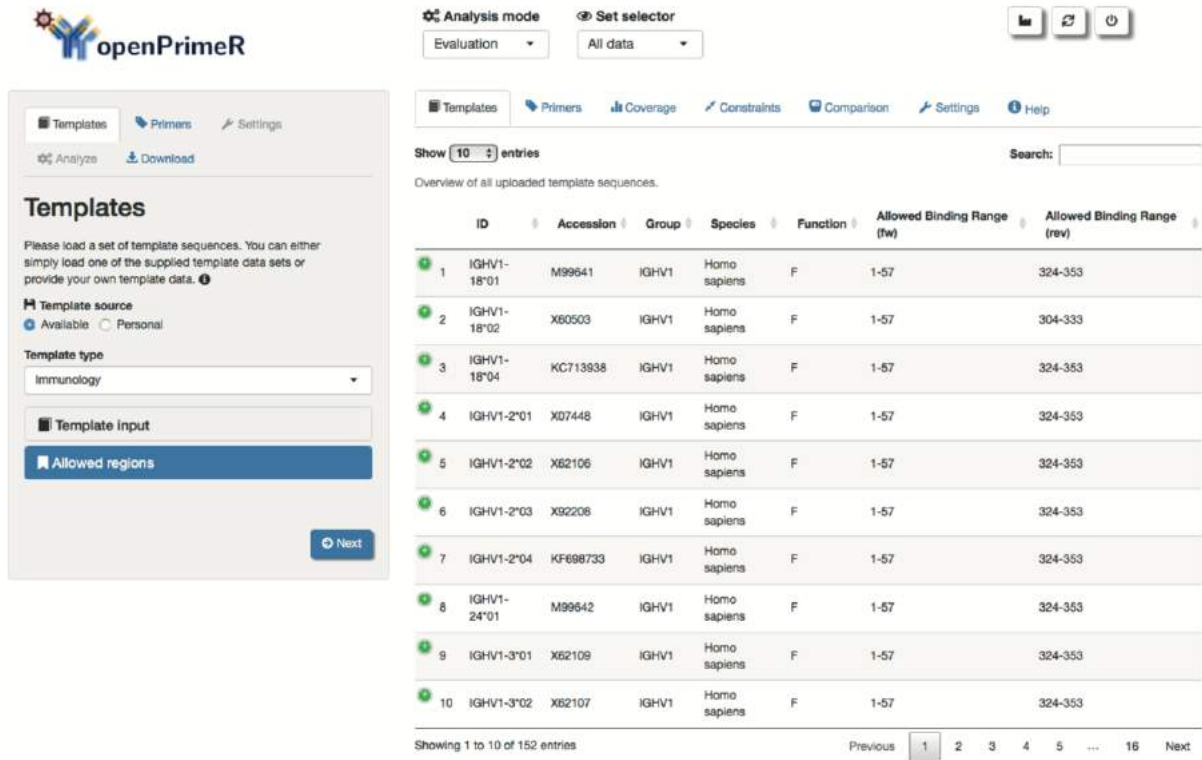
7 Supplements

oPR(5)IGHV_1

oPR(5)IGHV_2

oPR(5)IGHV_3
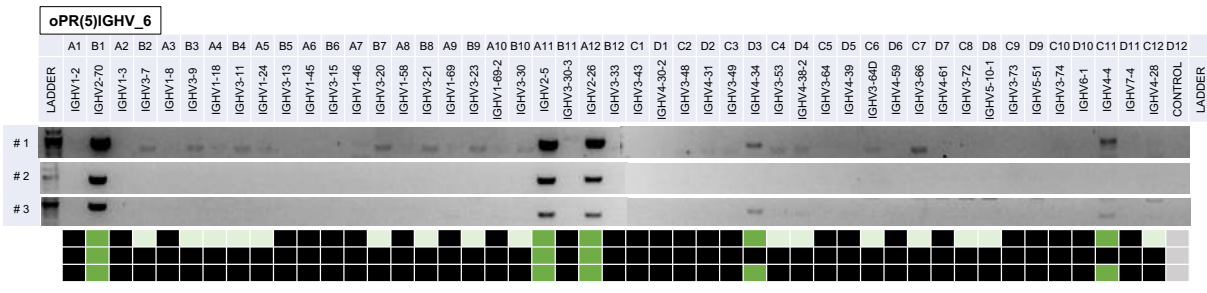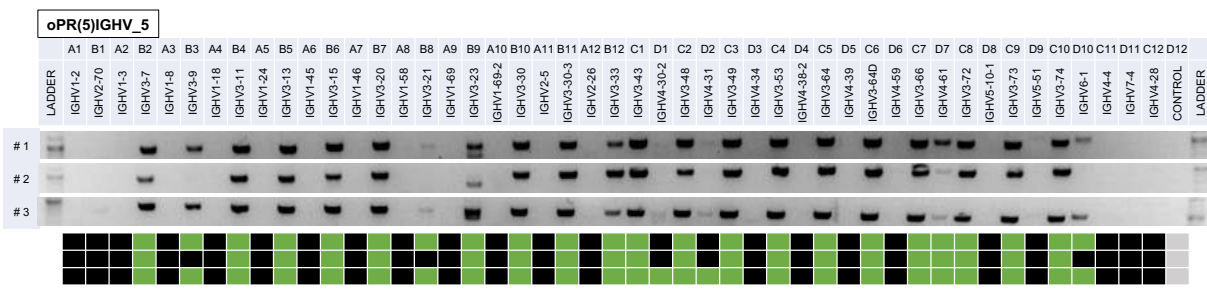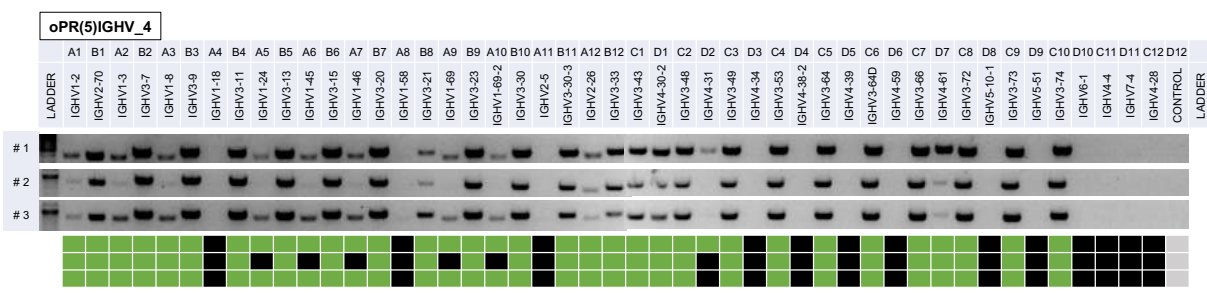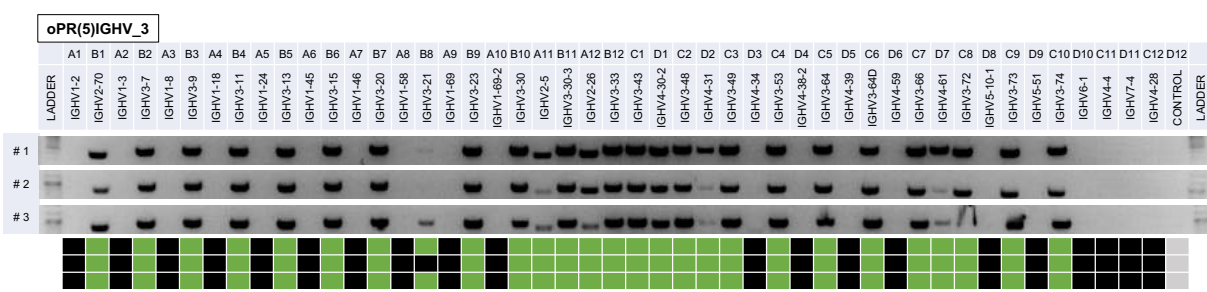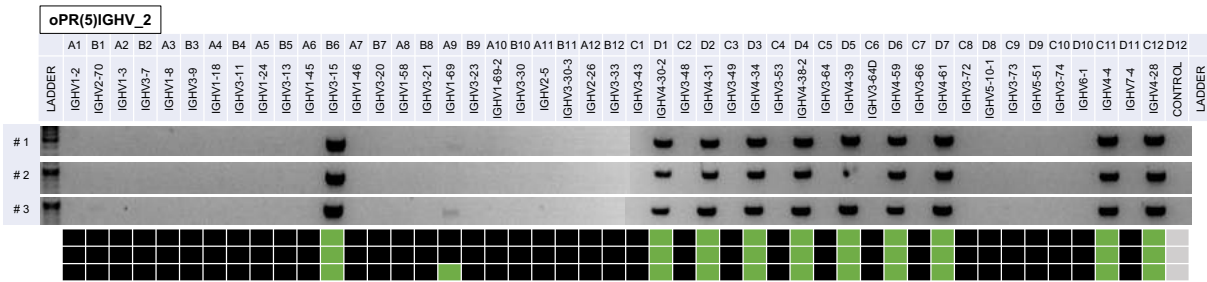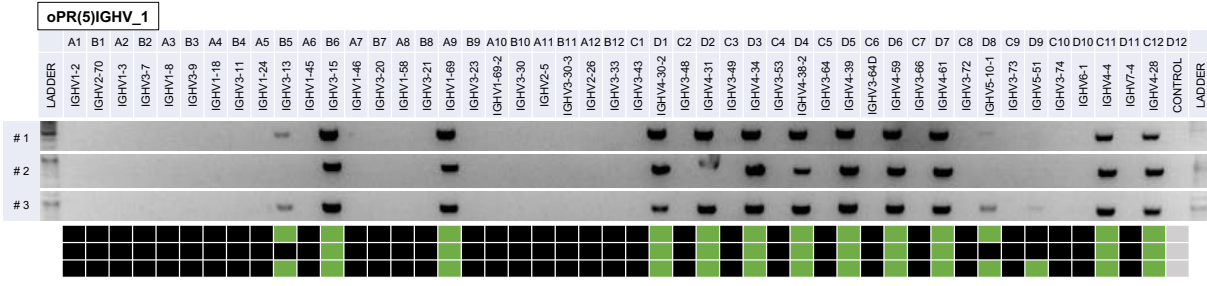
oPR(5)IGHV_4

oPR(5)IGHV_5

oPR(5)IGHV_6

(to be continued next page)
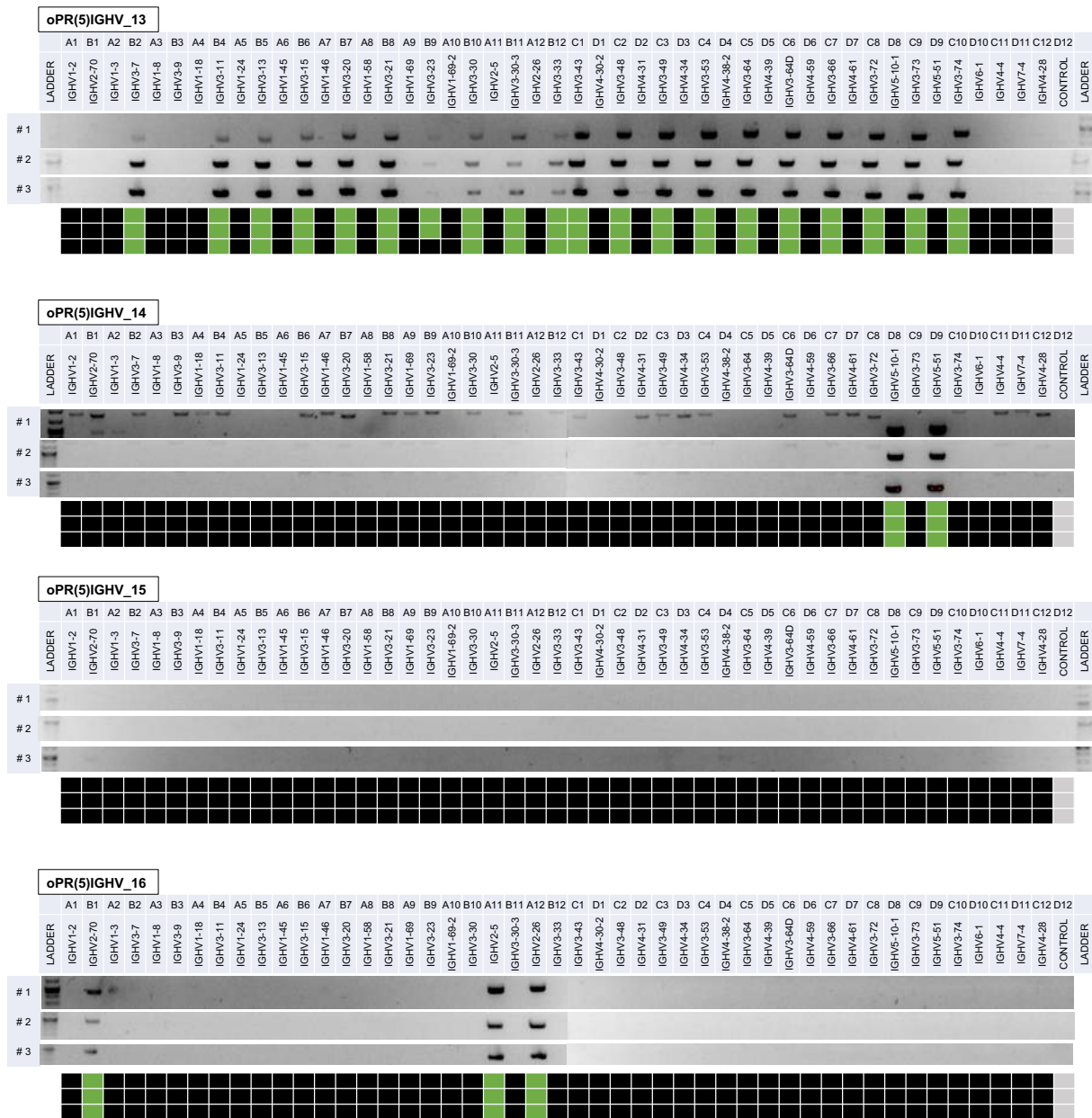
Figure S15: Validation of oPR(5)-IGHV single primer amplification status on IGHV gene library

oPR(5)-IGHV single primers tested in triplicates on the IGHV gene library. Header indicates each of the 16 single primers (oPR(5)-IGHV_1 – oPR(5)-IGHV_16). Amplification status is visualized under each gel pictures and either validated as "amplified", indicated in green, or "unamplified", indicated in black. Each last lane presents the negative control, containing the empty pCR4 vector backbone.

Figure S16: Validation of Set1 single primer amplification status on IGHV gene library
Single primer from Set1 tested in triplicates on the IGHV gene library. Header indicates each of the four single primers (Tiller_5'-L-VH1 – Tiller_5'L-VH5). Amplification status is visualized under each gel pictures and ether validated as "amplified", indicated in green or "unamplified", indicated in black.

Figure S17: openPrimeR graphical user interface
openPrimeR provides a shiny application, which can be used for interactive usage. The left panel provides the input interfaces and guides the user stepwise through the evaluation or design mode. The right panel offers several tabs for output display. This graphic has been published by Kreer et al. in the Journal of Immunological Methods in 2020 and was used in approval with all co-authors and with courtesy of Creative Commons Attribution- Share Alike 4.0. International License, *https://creativecommons.org/licenses/by-sa/4.0/*
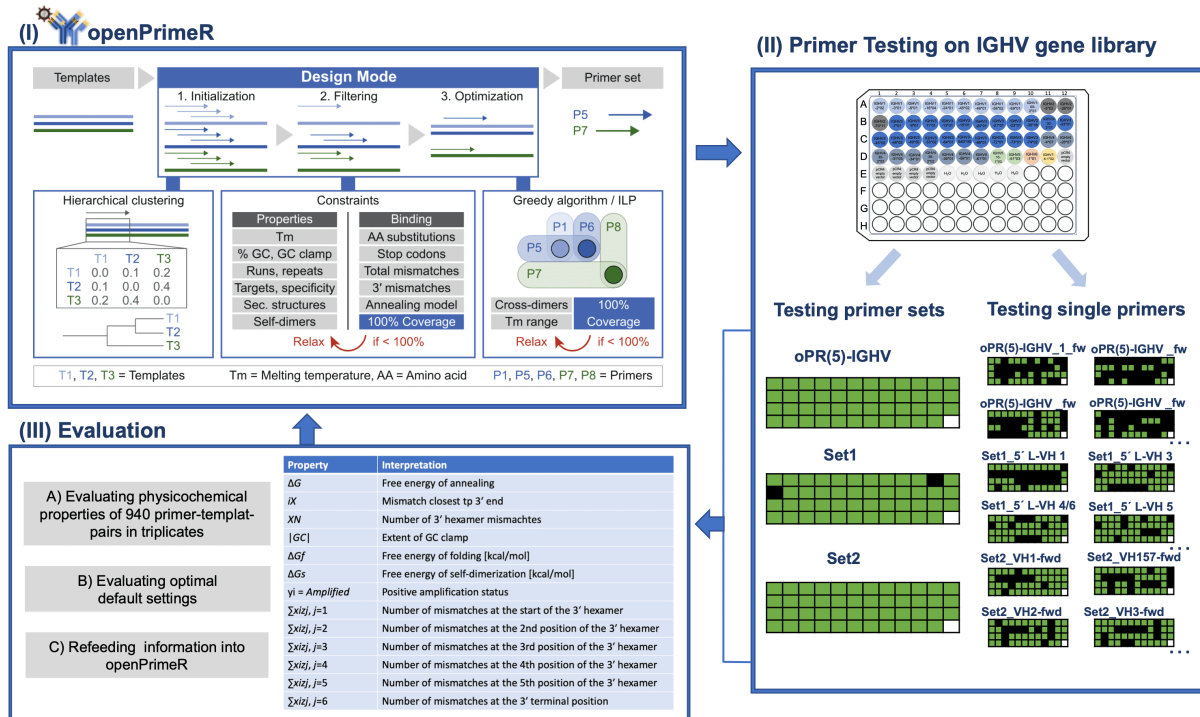
Figure S18: Continuous workflow of openPrimeR optimization

(I.) openPrimeR workflow for de novo primer design. Templates are uploaded into openPrimeR and a target region can be selected. An initial set of primers of specified length is either constructed by extracting template subsequences (non-degenerate primers) or by aligning and clustering of template subsequences (de-generate primers). Next, coverage is determined and primers are filtered according to user-defined constraints on their physicochemical properties. Finally, either a greedy algorithm or an integer linear program (ILP) finds an approximate (Greedy: P1, P6, P7) or exact (ILP: P5 and P7) minimal set of primers that covers all templates. If the coverage level is not achieved, constraints can automatically be relaxed to user-defined boundaries. (II.) Primers designed using openPrimeR and established primer sets (Set1, Set2) were tested on the IGHV gene library either using the complete primer sets or each single primer individually. (III.) Amplification status either set to *Amplified* or *Unamplified* and PTP specific physicochemical properties were evaluated. Concluding information was refed to the openPrimeR tool (e.g. by implementing the newly developed TMM). This modified graphic (I.) has been published by Kreer et al. in the Journal of Immunological Methods in 2020 and was used in approval with all co-authors and with courtesy of Creative Commons Attribution- Share Alike 4.0. International License, *https://creativecommons.org/licenses/by-sa/4.0/*

Table S4: openPrimeR optimized primer sets oPR-IGHV, oPR-IGKV and oPR-IGLV for the amplification of antibodies heavy and light chains (Kreer et al. 2020, Journal of Immunological Methods)[184].

openPrimeR IGHV set (oPR-IGHV)		
Name	**Sequence**	**Length**
oPR-IGHV-1_fw	ATGGACTGGACCTGGAGCATCC	22
oPR-IGHV-2_fw	ATGGACTGGACCTGGAGGATCCTC	24
oPR-IGHV-3_fw	ATGGACTGGACCTGGAGGGTCTTC	24
oPR-IGHV-4_fw	ATGGACTGGATTTGGAGGGTCCTCTTC	27
oPR-IGHV-5_fw	ATGGACACACTTTGCTACACACTCCTGC	28
oPR-IGHV-6_fw	ACTTTGCTCCACGCTCCTGC	20
oPR-IGHV-7_fw	GGCTGAGCTGGGTTTTCCTTGTTG	24
oPR-IGHV-8_fw	GGCTCCGCTGGGTTTTCCTTGTTG	24
oPR-IGHV-9_fw	CACCTGTGGTTCTTCCTCCTGCTG	24
oPR-IGHV-10_fw	ATGAAACACCTGTGGTTCTTCCTCCTCC	28
oPR-IGHV-11_fw	ACATCTGTGGTTCTTCCTTCTCCTGGTG	28
oPR-IGHV-12_fw	GCCTCTCCACTTAAACCCAGGCTC	24
oPR-IGHV-13_fw	ATGTCTGTCTCCTTCCTCATCTTCCTGC	28
oPR-IGHV-14_fw	ATGGAGTTGGGGCTGAGCTGG	21
oPR-IGHV-15_fw	ATGGGGTCAACCGCCATCCTC	21
openPrimeR IGKV set (oPR-IGKV)		
Name	**Sequence**	**Length**
oPR-IGKV-1_fw	ATGAGGCTCCTTGCTCAGCTTCTGG	25
oPR-IGKV-2_fw	ATGGAAGCCCCAGCTCAGCTTC	22
oPR-IGKV-3_fw	CCCAGCTCAGCTTCTCTTCCTCCTG	25
oPR-IGKV-4_fw	TGGTGTTGCAGACCCAGGTCTTCATTTC	28
oPR-IGKV-5_fw	GTCCCAGGTTCACCTCCTCAGCTTC	25
oPR-IGKV-6_fw	GCCATCACAACTCATTGGGTTTCTGCTG	28
oPR-IGKV-7_fw	TCCCTGCTCAGCTCCTGGG	19
oPR-IGKV-8_fw	CCTGGGACTCCTGCTGCTCTG	21
openPrimeR IGHV set (oPR-IGLV)		
Name	**Sequence**	**Length**
oPR-IGLV-1_fw	CCCTGGGTCATGCTCCTCCTGAAATC	26
oPR-IGLV-2_fw	CTCTGCTGCTCCTCACTCTCCTCAC	25
oPR-IGLV-3_fw	ATGGCATGGATCCCTCTCTTCCTCG	25
oPR-IGLV-4_fw	CCTCTCTGGCTCACTCTCCTCACTC	25
oPR-IGLV-5_fw	ACACTCCTGCTCCCACTCCTCAAC	24
oPR-IGLV-6_fw	ATGGCCTGGATCCCTCTACTTCTCC	25
oPR-IGLV-7_fw	ATGGCCTGGGTCTCCTTCTACC	22
oPR-IGLV-8_fw	ATGGCCTGGACTCCTCTCTTTCTGTTC	27
oPR-IGLV-9_fw	ATGGCCTGGATGATGCTTCTCCTC	24
oPR-IGLV-10_fw	GTCCCCTCTCTTCCTCACCCTCATC	25
oPR-IGLV-11_fw	CTCCTCGCTCACTGCACAGG	20
oPR-IGLV-12_fw	CCTCTCCTCCTCACCCTCCTC	21
oPR-IGLV-13_fw	CTCCTCCTCACCCTCCTCACTC	22
oPR-IGLV-14_fw	ATGGCCTGGACCCCTCTCC	19
oPR-IGLV-15_fw	ATGGCCTGGACCCCACTCC	19

8   Pre-publication of results

1. Kreer, C., Döring, M., **_Lehnen, N.,_** Ercanoglu, M.S., Gieselmann, L., Luca, D., Jain, K., Schommers, P., Pfeifer, N., Klein, F. openPrimeR for multiplex amplification of highly diverse templates. Journal of Immunological Methods 2020; **480**. DOI:10.1016/j.jim.2020.112752.

2. Döring, M., Kreer, C., **_Lehnen, N.,_** Klein, F., Pfeifer, N. Modeling the Amplification of Immunoglobulins through Machine Learning on Sequence-Specific Features. *Scientific Reports* 2019; **9**: 10748.

3. Gieselmann, L., Kreer, C., Ercanoglu, M.S., **_Lehnen, N.,_** Zehner, M., Schommers, S., Potthoff, J., Gruell, H., Klein, F., Effective high-throughput isolation of fully human antibodies targeting infectious pathogens. Nature Protocols 2021; **16:**3639-3671. doi:10.1038/s41596-021-00554-w.