View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Publications from Karolinska Institutet

From the Department of Microbiology, Tumor and Cell Biology
Karolinska Institutet, Stockholm, Sweden

DEFINITION OF IMMUNOGLOBULIN GERMLINE GENES BY NEXT GENERATION SEQUENCING FOR STUDIES OF ANTIGEN-SPECIFIC B CELL RESPONSES

Néstor Vázquez Bernat

Karolinska Institutet

Stockholm 2019

All previously published papers were reproduced with permission from the publisher.
Cover art by Silvia Bernat Senante
Published by Karolinska Institutet.
Printed by Eprint AB
© Néstor Vázquez Bernat, 2019
ISBN 978-91-7831-625-0

DEFINITION OF IMMUNOGLOBULIN GERMLINE GENES BY NEXT GENERATION SEQUENCING FOR STUDIES OF ANTIGEN-SPECIFIC B CELL RESPONSES

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Néstor Vázquez Bernat

Principal Supervisor:
Professor Gunilla B. Karlsson Hedestam
Karolinska Institutet
Department of Microbiology, Tumor and Cell Biology

Co-supervisor:
Dr. Martin M. Corcoran
Karolinska Institutet
Department of Microbiology, Tumor and Cell Biology

Opponent:
Associate Professor Scott D. Boyd
Stanford University
School of Medicine

Examination Board:
Associate Professor Caroline Grönwall
Karolinska Institutet
Center for Molecular Medicine

Professor Lennart Svensson
Linköping University
Department of Clinical and Experimental Medicine

Dr. Stephen Malin
Karolinska Institutet
Department of Medicine, Solna

“Nanos gigantum humeris insidentes” — attributed to Bernard of Chartres

ABSTRACT

Immunoglobulins play a critical role in the adaptive immune system, existing as cell surface-expressed B cell receptors and secreted antibodies. Circulating antibodies are the main correlate of protective immunity for most vaccines. An improved understanding of the germline genes that rearrange to encode the vast repertoire of antibodies is therefore of central interest. Despite this, current databases of immunoglobulin germline gene variation are incomplete, both for humans and research animal models, limiting studies of antigen-specific B cell responses.

In **Paper I**, we developed a computational tool, IgDiscover, which infers germline immunoglobulin V alleles from the repertoire of expressed antibodies in a given individual. We validated IgDiscover for the identification of human, mouse and rhesus macaque IGHV alleles and described novel IGHV alleles in all three species. Our results highlighted a high degree of inter-individual allelic diversity in rhesus macaques. In **Paper II**, we optimized and compared two major immunoglobulin library production methods based on 5′RACE and 5′multiplex PCR, respectively. We observed that, despite 5′RACE being unbiased in terms of amplification and having the advantage of not requiring 5′ end IGHV genomic information, current limitations on high-throughput sequence read length resulted in the 5′ multiplex method delivering a higher quality output due to its shorter amplicon size. In **Paper III**, we inferred germline immunoglobulin alleles in 45 macaques from four sub-populations of the two most common species used in biomedical research, rhesus and cynomolgus macaques. We confirmed and extended our observations concerning high inter-individual diversity, demonstrating that it was highest among Indonesian cynomolgus macaques and lowest among Mauritian cynomolgus macaques in the sub-populations studied. We compiled comprehensive IGHV, D and J allele databases and used several methods to independently validate novel alleles.

In conclusion, the work presented in this thesis establishes a road map to generate individualized immunoglobulin germline gene databases from diverse species, even if genomic immunoglobulin loci information is limited. This thesis also examines the advantages and disadvantages of commonly used next generation sequencing library preparation methods. Finally, it reports novel inferred immunoglobulin alleles in humans and macaques and illustrates a high degree of inter-individual immunoglobulin allelic diversity in primates, underlining the utility of generating individualized immunoglobulin databases for studies of immune repertoires.

POPULAR SCIENCE SUMMARY

English:

Antibodies are protein molecules produced by B cells in our bodies. They are present in all vertebrates where their primary role is to help fight infections. The two ends of an antibody have different functions: one end binds the pathogen (antigen) while the other end recruits other cells of the immune system to help control the infection. Antibodies are generated from different sets of genes, which are combined in a semi-randomized fashion to produce a great number of binding structures able to recognize a broad spectrum of antigens. When B cells encounter the same antigen several times, they accumulate mutations in their antibody genes. Some mutations are favorable and result in antibodies that bind the specific antigen better. B cells with improved antibodies survive and persist over time, offering long-lived immunity, while the other B cells are eliminated. By studying the antibodies generated after an infection or vaccination, researchers can better understand host immune responses to specific diseases and design strategies to develop protective vaccines.

Antibody genes are quite variable between individuals and the specific alleles (gene variants) in an individual (human or animal) can affect the capacity to combat a certain disease. These genes, which in total are several hundred, are present in unusually complex regions of the genome that are difficult to characterize with standard sequencing techniques. As a result, the current databases of antibody gene variants are incomplete. Therefore, we developed an alternative way to define known and novel antibody alleles within an individual, which is both faster and more economic, using a program called IgDiscover. This method allows for the generation of antibody gene/allele databases from many individuals in a short period of time. In this thesis, I have applied IgDiscover to study antibody genes in both humans and medically relevant animal models. The results presented thus improve our understanding of host antibody genetics, which will help guide future efforts to develop effective vaccines.

The papers included in this thesis describe: I) the process of discovering and defining immunoglobulin genes and alleles using IgDiscover; II) the development and evaluation of different protocols to study antibody genes using high-throughput sequencing; and III) the use of IgDiscover to define antibody germline genes in four sub-populations of macaques frequently used in vaccine studies.

Svenska:

Antikroppar är proteinmolekyler som produceras av B celler i våra kroppar. De hjälper oss att försvara oss mot infektioner, är evolutionärt konserverade och finns hos alla ryggradsdjur. Antikroppar består av två delar med olika funktioner: den ena delen binder främmande ämnen och den andra signalerar till immunsystemet för att förstärka svaret. Antikroppar produceras från gener som satts samman på ett kombinatoriskt sätt vilket ger upphov till ett stort antal bindningsspecificiteter. När en B cell aktiveras av samma främmande struktur flera gånger börjar mutationer ansamlas i generna. De B celler som har mest effektiva antikroppar selekteras och vissa av dessa ger livslång immunitet. Genom att studera antikroppar som genererats som svar mot infektioner eller vaccinationer kan forskare öka sin förståelse för immunsystemet och använda kunskapen för utveckling av nya vaccin.

Antikroppsgener varierar mycket och de genvarianter (alleler) en individ har kan påverka dess känslighet mot vissa sjukdomar. Generna (flera hundra separata segment) finns i regioner av genomet som är mycket komplexa, vilket medför att traditionella sekvenseringsmetoder är ineffektiva. Tillgängliga databaser över antikroppsgener är därför ofullständiga. Vi utvecklade en alternativ metod för att karaktärisera antikroppsgener som är både billigare och mer effektiv, vilken vi kallar IgDiscover. In den här avhandlingen har jag använt IgDiscover för att studera antikroppsgener i både människa och medicinskt relevanta djurmodeller. Kunskapen ger detaljerad information som kan användas till att utveckla bättre vaccin och behandlingar.

Arbeten som ingår i avhandlingen beskriver: I) IgDiscover processen för att upptäcka antikroppsgenerna och alleler; II) optimering av protokoll för hög-kvalitativ antikroppsrepertoaranalys; och III) tillämpning av IgDiscover på fyra populationer av primatsläktet makak, en djurmodell som ofta används i immunologiska studier.

Català:

Els anticossos són molècules proteiques produïdes pels limfòcits B en els nostres cossos. Estan presents en tots els vertebrats i ens ajuden a combatre les infeccions. Els dos costats de l'anticòs tenen diferents funcions: un s'uneix al patogen (antigen) i l'altre recluta cèl·lules del sistema immunitari per controlar la infecció. Els anticossos es generen a partir de diferents conjunts de gens que es combinen de forma semialeatòria per a produir una gran quantitat d'estructures que reconeixen un ampli espectre d'antígens. Quan els limfòcits B troben un antigen diverses vegades, acumulen mutacions als gens dels anticossos. Algunes d'aquestes mutacions són favorables, donant anticossos que s'uneixen millor a l'antigen. Els limfòcits B amb anticossos millorats sobreviuen i persisteixen al cap del temps, proporcionant immunitat a llarg termini, mentre que els altres limfòcits són eliminats. Estudiant els anticossos generats després d'una infecció o vacunació, els investigadors poden comprendre millor les respostes immunitàries de l'hoste contra malalties concretes i dissenyar estratègies per desenvolupar vacunes protectores.

Els gens dels anticossos són força variables entre individus, i els al·lels (variants d'un gen) dels anticossos que posseeix un individu (humà o animal) poden afectar la capacitat d'aquest per a combatre una malaltia determinada. Els gens, que en total són uns centenars, es troben en una àrea del genoma inusualment complexa i són difícils de caracteritzar amb els mètodes tradicionals de seqüenciació. Com a resultat, les bases de dades de variants de gens d'anticossos actuals són incompletes. Per això, nosaltres hem desenvolupat una forma alternativa de definir els al·lels dels gens d'anticossos coneguts i nous en un individu, que és alhora més ràpida i econòmica, utilitzant un programa anomenat IgDiscover. Aquest mètode permet generar bases de dades per a molts individus en molt poc temps. En aquesta tesi, he aplicat IgDiscover per estudiar els gens dels anticossos en humans i animals de laboratori mèdicament rellevants. Els resultats milloren el nostre coneixement de genètica d'anticossos, el qual ajudarà a guiar esforços futurs per crear vacunes efectives.

Els articles científics inclosos en aquesta tesi descriuen: I) el procés de descobrir i definir gens i al·lels utilitzant IgDiscover; II) el desenvolupament i avaluació de diferents protocols per estudiar els gens dels anticossos mitjançant seqüenciació massiva; i III) l'ús de IgDiscover per definir els gens dels anticossos en quatre subpoblacions de macacos utilitzats amb freqüència en estudis de vacunes.

Castellano:

Los anticuerpos son moléculas proteicas producidas por los linfocitos B en nuestros cuerpos. Están presentes en todos los vertebrados y nos ayudan a combatir las infecciones. Los dos lados del anticuerpo tienen diferentes funciones: uno se une al patógeno (antígeno) y el otro recluta células del sistema inmunitario para controlar la infección. Los anticuerpos se generan a partir de diferentes conjuntos de genes que se combinan de forma semialeatoria para producir una gran cantidad de estructuras que reconocen un amplio espectro de antígenos. Cuando los linfocitos B encuentran un antígeno varias veces, acumulan mutaciones en los genes de los anticuerpos. Algunas de estas mutaciones son favorables, dando anticuerpos que se unen mejor al antígeno. Los linfocitos B con anticuerpos mejorados sobreviven y persisten al cabo del tiempo, proporcionando inmunidad a largo plazo, mientras que los otros linfocitos son eliminados. Estudiando los anticuerpos generados después de una infección o vacunación, los investigadores pueden comprender mejor las respuestas inmunitarias del huésped contra enfermedades concretas y diseñar estrategias para desarrollar vacunas protectoras.

Los genes de los anticuerpos son bastante variables entre individuos, y los alelos (variantes de un gen) de los anticuerpos que posee un individuo (humano o animal) pueden afectar la capacidad de este para combatir una dolencia determinada. Los genes, que en total son unos centenares, se encuentran en un área del genoma inusualmente compleja y son difíciles de caracterizar con los métodos tradicionales de secuenciación. Como resultado, las bases de datos de variantes de genes de anticuerpos actuales son incompletas. Por esto, nosotros hemos desarrollado una forma alternativa de definir los alelos de los genes de anticuerpos conocidos y nuevos en un individuo, que es a la vez más rápida y económica, utilizando un programa denominado IgDiscover. Este método permite generar bases de datos para muchos individuos en muy poco tiempo. En esta tesis, he aplicado IgDiscover para estudiar los genes de los anticuerpos en humanos y animales de laboratorio médicamente relevantes. Los resultados mejoran nuestro conocimiento de genética de anticuerpos, el cual ayudará a guiar esfuerzos futuros para crear vacunas efectivas.

Los artículos científicos incluidos en esta tesis describen: I) el proceso de descubrir y definir genes y alelos utilizando IgDiscover; II) el desarrollo y evaluación de diferentes protocolos para estudiar los genes de los anticuerpos mediante secuenciación de alto rendimiento; y III) el uso de IgDiscover para definir los genes de los anticuerpos en cuatro subpoblaciones de macacos utilizados con frecuencia en estudios de vacunas.

LIST OF SCIENTIFIC PAPERS

I. Martin M. Corcoran, Ganesh E. Phad, **Néstor Vázquez Bernat**, Christiane Stahl-Hennig, Noriyuki Sumida, Mats A.A. Persson, Marcel Martin, and Gunilla B. Karlsson Hedestam. "*Production of individualized V gene databases reveals high levels of immunoglobulin genetic diversity*." Nature Communications – 2016 Dec; 7:13642.

II. **Néstor Vázquez Bernat**, Martin M. Corcoran, Uta Hardt, Mateusz Kaduk, Ganesh E. Phad, Marcel Martin and Gunilla B. Karlsson Hedestam. "*High-quality library preparation for NGS-based immunoglobulin germline gene inference and repertoire expression analysis.*" Frontiers in Immunology – 2019 Apr; 10:660.

III. **Néstor Vázquez Bernat**, Izabela Nowak, Mateusz Kaduk, David Spencer, Nancy Haigwood, Pauline Maisonasse, Roger Le Grand, Martin M. Corcoran and Gunilla B. Karlsson Hedestam. "*High inter-individual diversity among rhesus and cynomolgus macaques revealed by immunoglobulin germline gene inference.*" Manuscript.

PUBLICATIONS NOT INCLUDED IN THIS THESIS

I. Ganesh E. Phad, **Néstor Vázquez Berna**t, Yu Feng, Jidnyasa Ingale, Paola Andrea Martinez Murillo, Sijy O'Dell, Yuxing Li, John R. Mascola, Christopher Sundling, Richard T. Wyatt and Gunilla B. Karlsson Hedestam. "*Diverse antibody genetic and recognition properties revealed following HIV-1 envelope glycoprotein immunization.*" Journal of Immunology – 2015 Jun; 194(12):5903- 14.

II. Navid Madani, Amy M. Princiotto, David Easterhoff, Todd Bradley, Kan Luo, Wilton B. Williams, Hua-Xin Liao, M. Anthony Moody, Ganesh E. Phad, **Néstor Vázquez Bernat**, Bruno Melillo, Sampa Santra, Amos B. Smith, III, Gunilla B. Karlsson Hedestam, Barton Haynes, and Joseph Sodroski. "*Antibodies elicited by multiple envelope glycoprotein immunogens in primates neutralize primary human immunodeficiency viruses (HIV-1) sensitized by CD4-mimetic compounds.*" Journal of Virology – 2016 Apr; 90(10):5031-5046.

III. Paola Martinez-Murillo, Karen Tran, Javier Guenaga, Gustaf Lindgren, Monika Àdori, Yu Feng, Ganesh E. Phad, **Néstor Vázquez Bernat**, Shridhar Bale, Jidnyasa Ingale, Viktoriya Dubrovskaya, Sijy O'Dell, Lotta Pramanik, Mats Spångberg, Martin Corcoran, Karin Loré, John R. Mascola, Richard T. Wyatt, Gunilla B. Karlsson Hedestam. "*Particulate array of well-ordered HIV clade C Env trimers elicits neutralizing antibodies that display a unique V2 cap approach.*" Immunity – 2017 May; 46(5):804- 817.e7.

IV. Ganesh E. Phad, Pradeepa Pushparaj, Karen Tran, Viktoriya Dubrovskaya, Monika Àdori, Paola Martinez-Murillo, **Néstor Vázquez Bernat**, Suruchi Singh, Gilman Dionne, Sijy O'Dell, Komal Bhullar, Sanjana Narang, Chiara Sorini, Eduardo J. Villablanca, Christopher Sundling, Ben Murrell, John R. Mascola, Lawrence Shapiro, Marie Pancera, Marcel Martin, Martin Corcoran, Richard T. Wyatt, and Gunilla B. Karlsson Hedestam. "*Extensive dissemination and intra-clonal maturation of HIV Env vaccine-induced B cell responses.*" Journal of Experimental Medicine – 2019 Nov.

V. Viktoriya Dubrovskaya, Karen Tran, Gabriel Ozorowski, Javier Guenaga, Richard Wilson, Shridhar Bale, Christopher A. Cottrell, Hannah L. Turner, Gemma Seabright, Sijy O'Dell, Jonathan L. Torres, Lifei Yang, Yu Feng, Daniel P. Leaman, **Néstor Vázquez Bernat**, Tyler Liban, Mark Louder, Krisha McKee, Robert T. Bailer, Arlette Movsesyan, Nicole A. Doria-Rose, Marie Pancera, Gunilla B. Karlsson Hedestam, Michael B. Zwick, Max Crispin, John R. Mascola, Andrew B. Ward and Richard T. Wyatt. "*Vaccination with glycan-modified HIV NFL envelope trimer-liposomes elicits broadly neutralizing antibodies to multiple sites of vulnerability.*" Immunity – 2019 Nov.

CONTENTS

1 B Cells and Humoral Immunity 1
- 1.1 Introduction to Immunoglobulins 1
- 1.2 Immunoglobulin Formation 1
- 1.3 VDJ Recombination 2
- 1.4 B Cell Activation and Proliferation 3

2 B Cell Responses 7
- 2.1 Antigen-Specific B Cell Responses in Vaccination 7
- 2.2 Functional Studies of Antibody Responses 7

3 Repertoire Sequencing 9
- 3.1 Sequencing Technologies 9
- 3.2 Applications 10

4 Immunoglobulin Genetics and Polymorphism 13
- 4.1 Immunoglobulin Loci and Gene Databases 13
- 4.2 Germline Gene Discovery 14
- 4.3 Animal Models 15

5 Aims 19

6 Results and Discussion 21
- 6.1 Germline Inference (Paper I) 21
- 6.2 Library Preparation (Paper II) 23
- 6.3 Macaque Immunoglobulin Germlines (Paper III) 24

7 Concluding Remarks and Future Perspectives 27

8 Acknowledgements 29

9 References 35

LIST OF ABBREVIATIONS

AAALAC	Association for Assessment and Accreditation of Laboratory Animal Care
Ab	Antibody
AFL	Astrid Fagreaus Laboratory
AID	Activation-induced cytidine deaminase
AIRR-Seq	Adaptive immune receptor repertoire sequencing
AIRRc	AIRR Community
APC	Antigen-presenting cell
BCR	B cell receptor
bNAb	Broadly neutralizing antibody
bp	Base pair
CAIRR	CEDAR-AIRR
CDR3	Complementarity-determining region 3
CEDAR	Center for Expanded Data Annotation and Retrieval
CNV	Copy number variation
D	Diversity
DB	Database
FDC	Follicular dendritic cells
GC	Germinal center
HC	Heavy chain
HIV	Human immunodeficiency virus
HTS	High-throughput sequencing
IARC	Inferred Allele Review Committee
Ig	Immunoglobulin
IgH	Immunoglobulin heavy chain
IGHD	Immunoglobulin heavy chain D
IGHJ	Immunoglobulin heavy chain J
IGHV	Immunoglobulin heavy chain V
IgK	Immunoglobulin kappa chain
IGKV	Immunoglobulin kappa chain V
IgL	Immunoglobulin lambda chain
IGLV	Immunoglobulin lambda chain V
IMGT	International ImMunoGeneTics Information System

J	Joining
LC	Light chain
mAb	Monoclonal antibody
MHC	Major histocompatibility complex
MiAIRR	Minimal Information about AIRR
MPSS	Massively parallel signature sequencing
MTPX	Multiplex
NGC	Novel gene candidate
NGS	Next-generation sequencing
NHEJ	Non-homologous end joining
OGRDB	Open Germline Receptor Database
ORF	Open reading frame
RACE	Rapid amplification of cDNA ends
RAG	Recombination activating gene
Rep-seq	Immune repertoire sequencing
RSS	Recombination signaling sequences
SHM	Somatic hypermutation
SIV	Simian immunodeficiency virus
SMRT	Single-molecule real-time
TdT	Terminal deoxynucleotidyl transferase
Tfh	Follicular helper T cell
UMI	Unique molecular identifiers
UTR	Untranslated region
V	Variable

1 B CELLS AND HUMORAL IMMUNITY

1.1 INTRODUCTION TO IMMUNOGLOBULINS

The production of immunoglobulins (Igs) (reviewed in (Schroeder and Cavacini 2010)) is a key defense mechanism in the immune system. Antibodies (Abs, secreted forms of Igs) help combat diseases by targeting antigens with high specificity and by promoting effector functions that help in clearing infections.

Two key processes confer a high degree of specificity to Igs; first, the generation of an extremely diverse set of Igs capable of recognizing an extensive array of antigenic epitopes, and second, the incorporation of mutations and selection of the highest affinity antibodies, which occurs during the process of affinity maturation. Surface-bound Igs (B cell receptors, BCRs) are generated during B cell development through combinatorial assembly of gene segments of the heavy chain (HC) locus (IgH) and the kappa and lambda light chain (LC) loci (IgK and IgL). In each B cell, one HC and one LC randomly pair to generate a repertoire of highly diverse BCRs. Following productive rearrangements and pairing, the BCR is expressed on the cell surface, after which the B cell is subjected to different steps of negative selection to remove potentially self-reactive cells.

When B cells become activated by protein antigens in the context of an immune response, they enter a germinal center (GC) reaction where they receive activating signals from cognate CD4+ T cells. This in turn switches on a process known as somatic hypermutation (SHM), which generates BCRs with decreased, equal or increased affinity for the antigen. Only B cells exhibiting improved affinity BCRs are positively selected in the GC, leading to the generation of antigen-specific memory B cells and plasma cells. The work described in this thesis focuses on the variable (V), diversity (D) and joining (J) gene segments, which are the building blocks of the antigen-binding regions of antibodies. In the interest of space, I am not describing the constant (C) genes or discussing antibody effector function in this thesis frame as this is a separate topic of investigation.

1.2 IMMUNOGLOBULIN FORMATION

As mentioned, Igs are formed by the pairing of a HC with a LC, either kappa or lambda. HCs are encoded by the semi-random joining of V, D and J gene segment, while LCs are formed by joining a V gene segment with a J gene segment (Tonegawa 1983). These processes occur in a highly coordinated manner (Alt, Blackwell, and Yancopoulos 1987) (**Fig 1**). First, a HC D (IGHD) and a HC J (IGHJ) segment from one of the maternal or paternal derived IgH loci are recombined. If this yields a stop codon or an out-of-frame sequence, then a IGHD-IGHJ rearrangement will be attempted in the IgH locus on the other chromosome. Following successful IGHD-IGHJ recombination, HC V (IGHV) rearrangement occurs the same way. Once a productive HC is generated, the LC will be rearranged. Rearrangement of kappa VJ (IGKV,J) genes will be attempted first, and if this fails, lambda VJ (IGLV,J) genes will be

used. This sequential process ensures that only one HC and one LC will be expressed in each B cell, which is referred to as allelic exclusion (Nussenzweig et al. 1987) (depicted in **Fig. 1**).

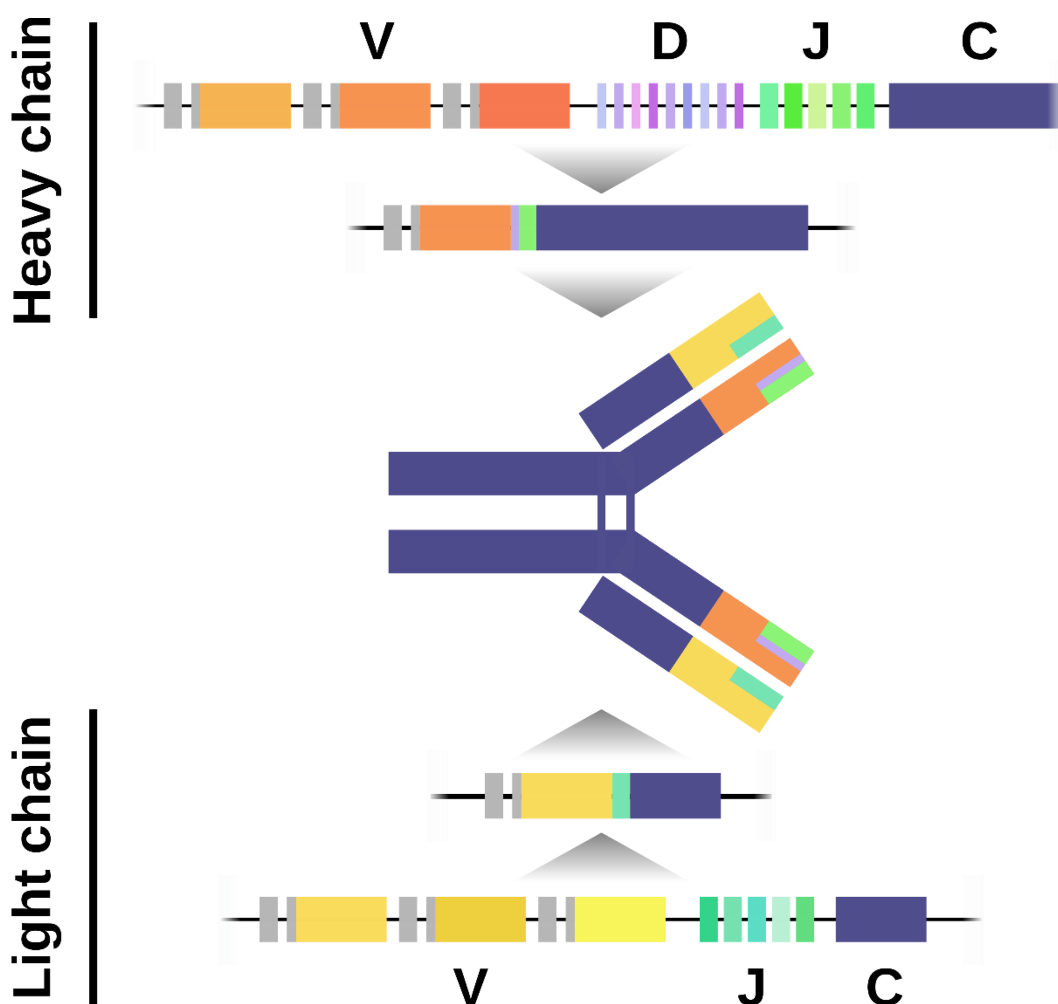

Figure 1. Rearrangement of the V(D)J genes of HC and LC to form the Ig protein.

1.3 VDJ RECOMBINATION

V(D)J rearrangements (reviewed in (Van Gent and Van Der Burg 2007)) require sequences called recombination signaling sequences (RSS) that flank each V, D and J gene segment and are recognized by the recombination-activating gene enzymes, RAG-1 and RAG-2 (reviewed in (Jung and Alt 2004; Dudley et al. 2005)). These sequences contain a repetitive nonamer and heptamer sequence separated by a spacer of either 12 or 23 nucleotides. Once recognized, these sequences align with each other, resulting in the formation of double-stranded breaks and hairpins (reviewed in (Schatz and Swanson 2011)). Both hairpins are then cleaved and ligated in a process termed non-homologous end joining (NHEJ). Recruitment of the Artemis protein complex then opens the hairpins in an imprecise manner, which can generate palindromic overhangs. Once the ends are aligned, terminal deoxynucleotidyl transferase

(TdT) adds non-templated semi-random nucleotides (Gauss and Lieber 1996) to both 3′ ends (Yancopoulos et al. 1984; Benedict et al. 2000). Finally, exonucleases remove non-matching nucleotides at both ends and polymerases add missing ones, and the two sequences are subsequently merged. This procedure can easily produce frame shifts or stop codons, which would cause termination of the rearrangement in that locus and initiation of the same procedure on the other Ig locus. This process creates highly polymorphic HC VDJ and LC VJ junctions, which are located in the complementarity-determining region 3 (CDR3) of the antibody molecule The CDR3 thus provides an "identity tag", which together with the VDJ gene usage can be used to identify clonally related sequences derived from the same B cell precursor.

B cells require the formation of the HC and LC sequences at different steps of their development to receive positive survival signals; at the pro-B and pre-B cell stage, respectively. B cells that do not display a functional BCR are eliminated. B cells expressing functional BCRs are also tested for self-reactivity in a process called negative selection. If the affinity to self-antigen is too high, the cell may reactivate the RAG enzymes to try to produce an alternative LC in a process termed receptor editing (Tiegs, Russell, and Nemazee 1993; Gay et al. 1993). If this is not successful, the B cell will undergo apoptosis to prevent the potential formation of autoantibodies. Once the BCR is formed and the cell has survived negative selection, the "fate" of the B cell is sealed and the Ig genes will not be further modified – except if it is selected into a GC reaction to undergo affinity maturation as described below.

1.4 B CELL ACTIVATION AND PROLIFERATION

Once a naïve B cell encounters an antigen (usually in one of the secondary lymphoid organs) and receives a positive signal through the BCR, it may proliferate and differentiate into a short-lived plasma cell (Coutinho and Möller 1973; Jacobs and Morrison 1975) (**Fig. 2**). Such rapid responses are T cell-independent and help mount a first-line response against invading pathogens (reviewed in (Fagarasan and Honjo 2000)). Short-lived plasma cell responses are important to control infections in the first week(s) before a more potent T-dependent antibody response is elicited (reviewed in (Nutt et al. 2015)).

In parallel, B cells internalize and process protein antigens in order to present pathogen-derived peptides to T cells via major histocompatibility complex II (MHC class II) molecules, to mount the T cell-dependent response. Interactions between B cells and cognate T cells seed GCs in B cell follicles of secondary lymphoid organs. The GCs are readily visible by optical microscopy around 10-14 days after antigenic challenge (reviewed in (MacLennan 1994; Victora and Nussenzweig 2012). In the GC, B cells first locate to the dark zone (**Fig. 2**), where they proliferate and the enzyme activation-induced cytidine deaminase (AID) introduces random deamination of cytosine residues in the actively transcribed rearranged Ig gene, generating uracil bases. This process is the aforementioned SHM, a hallmark of the GC

reaction ((Muramatsu et al. 2000) and reviewed in (Larson and Maizels 2004; Teng and Papavasiliou 2007; Milstein and Neuberger 1996)). Highly error-prone B cell polymerases are then recruited to fill the gaps created when uracil residues are removed, with mutations being generated at a much higher frequency than in other parts of the genome (Bachl, Ertongur, and Jungnickel 2006). Mutations that lead to unproductive BCR expression initiate apoptosis of the B cell. Only B cells that express a functional BCR migrate to the light zone of the GC, where selection occurs. The light zone is characterized by the presence of abundant follicular dendritic cells (FDC), an antigen-presenting cell (APC) of non-hematopoietic origin that presents intact protein antigen to B cells for their subsequent uptake and presentation of MHC class II-restricted peptides to follicular CD4+ T helper (Tfh) cells (reviewed in (Allen and Cyster 2008)) (**Fig. 2**). FDCs are known to present intact antigen on their surface for prolonged periods (Chen et al. 1978) via the complement system (Papamichail et al. 1975; Klaus and Humphrey 1977) and their expression of Fc receptors (Yoshida, van den Berg, and Dijkstra 1993), which capture antibody-antigen complexes (reviewed in (Kranich and Krautler 2016)). It is believed that competition for antigen presented by FDCs dictates which B cells take up and present more antigen to Tfh cells, which in turn provide pro-survival signals to cognate B cells (Victora et al. 2010). This process allows B cells that encode BCRs with improved antigen affinity to be successful in the GC and emerge as long-lived memory B cells and plasma cells (reviewed in (D. M. Tarlinton 2008)).

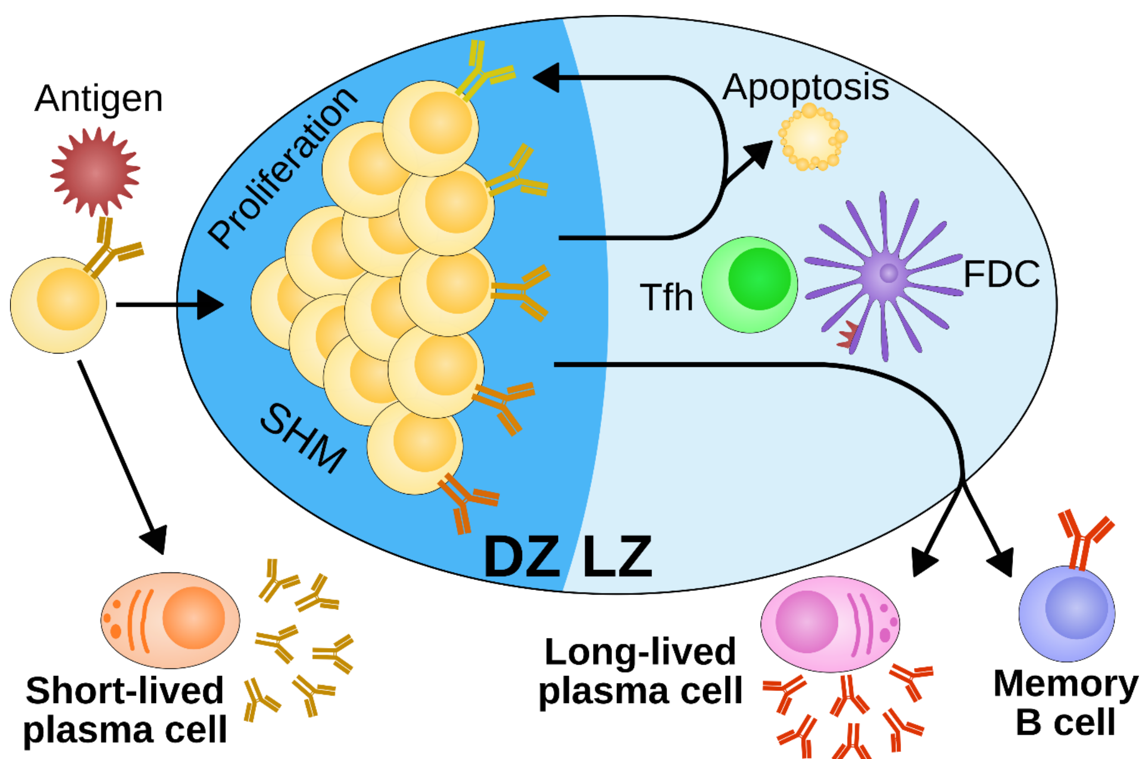

Figure 2. B cell activation and proliferation outside and inside the GC.

Before exiting the GC, affinity-matured B cells undergo class switch recombination to acquire the effector function that best combats a given antigen. This involves switching from IgM/IgD isotype to IgA, IgE or IgG, depending on the type of signals the B cells receive from T cells. AID also participates in this process, but instead of generating random deaminations, like in SHM, AID is recruited to specific sites preceding the different constant region genes. The exact mechanism is not well understood. A possible solution could be that AID causes double-strand breaks in this region due to high concentration of deaminations (reviewed in (Durandy 2003)), or that it is involved in merging the double strand breaks generated (Zan and Casali 2008).

B cells that exit the GC differentiate into either long-lived plasma cells, which home to the bone marrow and produce antibodies for long periods of time, or memory B cells, which circulate between the periphery and secondary lymphoid organs and are poised to generate rapid recall responses if re-exposed to the same or a similar antigen (**Fig. 2**). The mechanisms that dictate memory cell and plasma cell fates are not known. At least two models have been proposed (reviewed in (D. Tarlinton 2012)): either daughter cells of a given B cell clone proceed down the two different pathways through a stochastic process (Duffy et al. 2012), or that progeny of a given clone arise through asymmetric cell division where one daughter cell differentiates into a plasma cell and the other daughter cell remains a memory B cell (Barnett et al. 2012). In either case, a mechanism that retains the clonal lineage, while allowing both the generation of plasma cells and the maintenance of slowly dividing stem cell-like memory B cell pool must exist (Luckey et al. 2006). The mechanisms underlying these processes are highly relevant for the development of protective vaccines and this remains an important area of research.

2 B CELL RESPONSES

2.1 ANTIGEN-SPECIFIC B CELL RESPONSES IN VACCINATION

Vaccination is a prophylactic approach that aims to induce a level of immunity that prepares the host to respond efficiently when exposed to the real infection, such that disease is prevented. Vaccines have improved greatly since the early engraftments reported in China (Needham 2000) and Edward Jenner's smallpox variolation in 1796 (reviewed in (Stern and Markel 2005)). The development of vaccines represented a major breakthrough in healthcare, exemplified by the smallpox eradication (Foege, Millar, and Henderson 1975). Vaccines can be classified into two categories, whole pathogen vaccines (inactivated or attenuated) and component/purified antigens, including recombinant subunit vaccines. Whereas live attenuated vaccines are highly effective after a single immunization, due to their prolonged antigen expression and induction of innate immunity, inactivated and subunit vaccines usually need multiple immunizations. Purified subunit vaccines also need an adjuvant to boost the response, since they lack components that activate the innate immune system (Duan and Mukherjee 2016).

The majority of licensed vaccines were developed empirically, but vaccine research is now a major branch of basic immunology. For pathogens that cause chronic infections, such as human immunodeficiency virus (HIV)-1, subunit vaccines represent the only acceptable option due to the severity of the disease they could generate if attenuation or inactivation failed. So far, the elicitation of cross-reactive antibody responses against HIV-1 by subunit vaccination has proven elusive (Mascola and Montefiori 2010). Broadly neutralizing Abs (bNAbs) are induced in a subset of chronically HIV-1-infected individuals ((Sather et al. 2009; Gray et al. 2011), and reviewed in (Kwong, Mascola, and Nabel 2013)). A major focus in the field of HIV-1 vaccine research is to mimic these processes in the context of subunit vaccination (reviewed in (Burton and Mascola 2015; van Haaren, van den Kerkhof, and van Gils 2017; Kwong and Mascola 2018)). Two very recent studies reported long-awaited encouraging results when vaccine-induced bNAb responses were induced in rhesus macaques (Kong et al. 2019) and in rabbits (Dubrovskaya et al. 2019). Although other immune cells play important roles in the induction of vaccine-induced responses (e.g. innate immune cells and T cells, reviewed in (Pulendran and Ahmed 2011)), Abs are the main correlate of protection for most licensed vaccines (reviewed in (Plotkin 2008)).

2.2 FUNCTIONAL STUDIES OF ANTIBODY RESPONSES

Since the immune system produces an extremely broad array of Igs with very diverse specificities, the response to any given antigen is also diverse. The antibody response is usually studied at the polyclonal level, often by examining antigen-specific reactivity in serum. There are several methods to study polyclonal antibody responses (neutralization, ELISA or B cell ELISpot assays), which represent the standard ways to evaluate most vaccine

candidates. One disadvantage of these tests is that the results represent an average of a very diverse set of antibodies. For a more comprehensive understanding, the isolation of antigen-specific monoclonal antibodies (mAbs) provides possibilities for in-depth examination of the elicited response. Once a mAb is isolated, its genetic and functional properties can be determined. Furthermore, structural studies can reveal how an Ab interacts with its target antigen, both in the context of infection (Wrammert et al. 2008; Wu et al. 2010; Scheid et al. 2011; L. Huang, Lange, and Zhang 2014; Kallewaard et al. 2016; J. Huang et al. 2016; Cox et al. 2016; Murugan et al. 2018) and immunization (Sundling et al. 2012; Y. Li et al. 2013; Navis et al. 2014; Phad et al. 2015; Martinez-Murillo et al. 2017; K. Smith et al. 2013; Henry et al. 2019). However, the isolation of mAbs is low-throughput, which means that only a fraction of the elicited antibodies can be studied. Therefore, there is a great need for the development of higher throughput methods that can be applied to study highly polyclonal Ab repertoires.

3 REPERTOIRE SEQUENCING

3.1 SEQUENCING TECHNOLOGIES

With the development of next-generation sequencing (NGS), also called high-throughput sequencing (HTS), immunologists quickly devised methods to study adaptive immune receptors, either B cell (Weinstein et al. 2009; Glanville et al. 2009; Boyd et al. 2009; Arnaout et al. 2011) or T cell repertoires (Freeman et al. 2009; Robins et al. 2009; 2010).

In contrast to the conventional sequencing approaches used previously (Sanger and Coulson 1975; Maxam and Gilbert 1977), NGS allows for much improved coverage of these highly diverse repertoires. One of the first methods reported was based on pyrosequencing (Nyrén, Pettersson, and Uhlén 1993), which allows for the real-time reading of DNA synthesis (Ronaghi et al. 1996). A limitation was that the signal emitted by a single DNA molecule is too weak to parallelize into many reactions. This roadblock was bypassed by the invention of bridge amplification (Illumina (Fedurco et al. 2006)) and emulsion amplification (Roche/454 (Margulies et al. 2005)) (**Fig. 3A**). An alternative is single-molecule real-time (SMRT) sequencing, in which a single DNA molecule is measured per well, using either bright fluorophores excited with a laser (PacBio (Korlach et al. 2010) and reviewed in (Rhoads and Au 2015)) or electrical currents being affected by the differential blockage of a nanopore (Nanopore (Branton et al. 2009)) (**Fig. 3B**). These methods, sometimes called third generation sequencing, generate a higher intensity signal from a single sequence and thus can forego the need for amplification, simplifying sample preparation for genome sequencing.

Although these technologies have seen many improvements over the years, there has not been a clearly superior prevailing method, as was the case previously for Sanger sequencing. Each technology possesses different strengths and caters better to specific applications. As of the writing of this thesis, Illumina and other multi-sequence technologies have relatively low estimated error rates (0,01%) (Pfeiffer et al. 2018) and high-throughput (depending on instrument and read length). Illumina, specifically, has the lowest price per base sequenced. These technologies are well suited to amplicon sequencing for lengths below 600 base pairs (bp). The technology can also be applied to shotgun sequencing of the genome (explained below), but the short-read length makes repetitive areas of the genome a challenge to assemble, even at very high read depth. On the other hand, SMRT from PacBio, allows for extremely long subreads in the tens of kb, with the caveat of a 10-15% error rate in those reads. Pacbio fixes the quality scores by circular sequencing, which creates a trade-off between read length and quality. This makes the method specifically well suited for genome sequencing (with the ability to resolve repetitive regions that are intractable with short reads), although PacBio sequencing can also be applied to intermediate size amplicons (700-5000 bp) with a higher cost per read than Illumina.

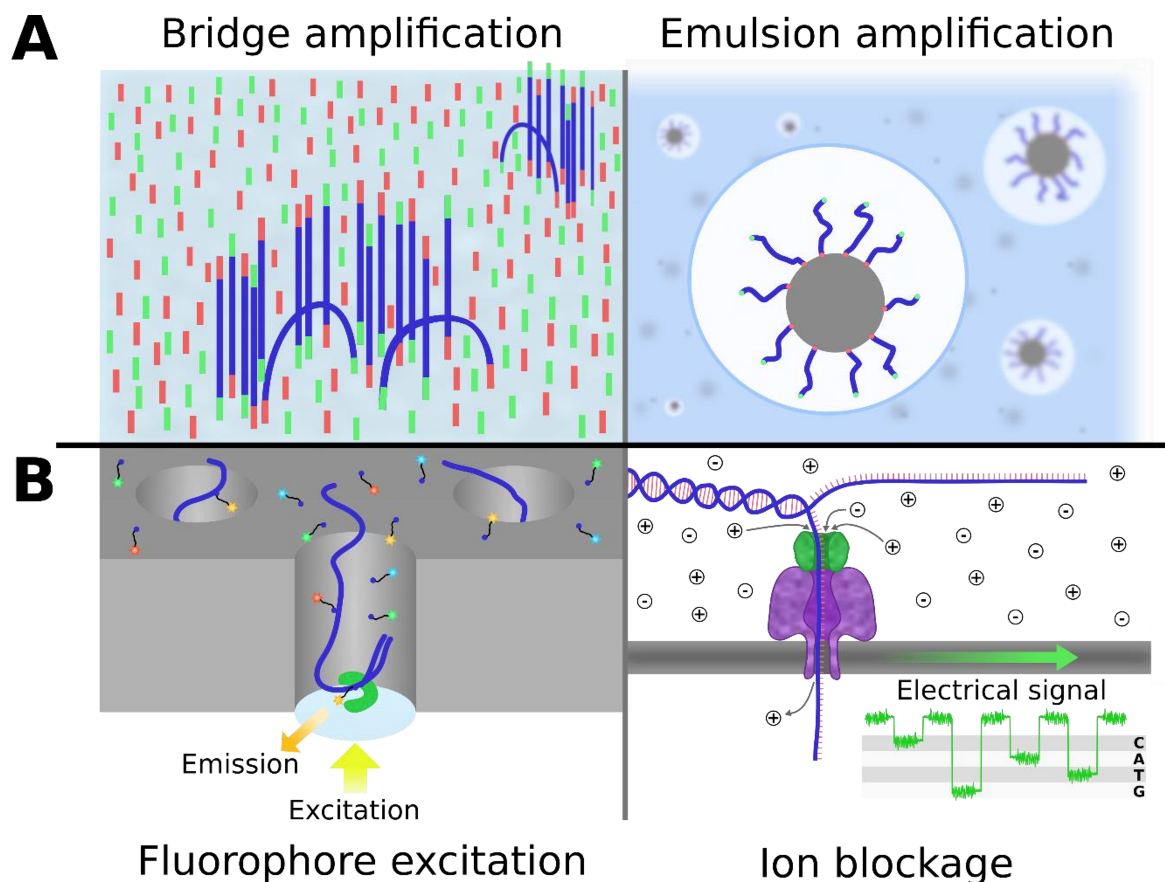

Figure 3. Main high-throughput sequencing technologies. A) Amplification methods to increase the signal with bridge amplification (Illumina) and emulsion amplification (Roche). B) Single molecule sequencing methods either fluorophores (PacBio) or electrical sensor of ion blockage (Nanopore).

3.2 APPLICATIONS

The advent of improved sequencing technologies allowed for the exploration of Ig repertoires at an unprecedented level of detail (reviewed in (Georgiou et al. 2014; Robinson 2014; Calis and Rosenberg 2014; Wardemann and Busse 2017; Davis and Boyd 2019)). Sequence analysis of immune repertoires is commonly referred to as immune repertoire sequencing (Rep-seq) or adaptive immune receptor repertoire sequencing (AIRR-seq). However, these methods also require high quality bioinformatic analyses to filter and organize the data. So far, computational methods to study immune repertoires have focused on repertoire diversity, architecture, convergence and antibody evolution (reviewed in (Miho et al. 2018)). These features of the repertoire offer information about how individuals and populations respond to infection and vaccination (Jackson et al. 2014; C. Wang et al. 2015; Hoehn et al. 2016; Waltari et al. 2018; Hoehn et al. 2019), thus providing new opportunities for diagnosis and treatment. Specific population groups may also be interrogated, such as the elderly (C. Wang et al. 2014; De Bourcy et al. 2017; Hoehn et al. 2019) or individuals with immune dysregulations (Bashford-Rogers et al. 2013; Stamatopoulos et al. 2017). One of the most striking discoveries generated by these studies is the level of convergence of repertoires from different individuals, or different animals, when infected or vaccinated ((Jackson et al. 2014;

Galson, Trück, Fowler, Clutterbuck, et al. 2015; Trück et al. 2015; Greiff et al. 2017) and reviewed in (Collins and Jackson 2018)). These so-called public repertoires or clonotypes appear to be linked to immunogenicity and protection (Trück et al. 2015), and could possibly be used in diagnostics or for the development of new immunogens.

As for all new technologies, it is important to understand potential sources of error when analyzing NGS data (reviewed in (Baum, Venturi, and Price 2012)). High-quality library construction methods are central for optimal downstream analysis, and method reliability is another field of study (Menzel et al. 2014). The analysis of the Rep-seq data without proper consideration of experimental design can lead to unreliable conclusions (Uduman et al. 2014), and close contact between experimentalists and analysts is strongly desired (Hoehn et al. 2016). The use of unique molecular identifiers (UMI, also referred to as barcodes) is important to allow for sequencing error correction and potential PCR amplification biases (Kivioja et al. 2012).

The rapid adoption of Rep-seq applications in the immunology field meant that the sequencing technology preceded the analysis tools and standards required to generate reproducible, robust and shareable data. In 2015, the Adaptive Immune Receptor Repertoire Community (AIRRc) was formed to address these issues (Breden et al. 2017). This community has already generated the Minimal Information about Adaptive Immune Receptor Repertoire (MiAIRR) data requirements to facilitate data comparison from different sources (Rubelt et al. 2017), created a pipeline based on the Center for Expanded Data Annotation and Retrieval (CEDAR) technology (CAIRR) for web-based metadata submission compliant with MiAIRR (Bukhari et al. 2018), and established the Inferred Allele Review Committee (IARC) for examining novel inferred sequences from Rep-seq data (Ohlin et al. 2019). The continued work of this community will be important for the Rep-seq field.

The HCs and LCs of a functional antibody are produced from different mRNA molecules. The availability of methods to perform Rep-seq of the linked chains could facilitate the identification of clonal lineages, despite very recent reports indicating LC might not be necessary (Zhou and Kleinstein 2019), and would allow mAb isolation from such data. Researchers have attempted to resolve this shortcoming by stochastic linkage in well distributions (for T cells (Howie et al. 2015)), physical linkage of the chains (Dekosky et al. 2013; DeKosky et al. 2015; Mcdaniel et al. 2016), and single cell well-barcoding (Busse et al. 2014; Lu et al. 2014). These methods focus exclusively on receptor identification. Another approach is to obtain this information bioinformatically from single-cell transcriptomic data (Upadhyay et al. 2018; Lindeman et al. 2018; Rizzetto et al. 2018; Afik, Raulet, and Yosef 2019), which can provide additional information on the "parental" B cell. The throughput and complexity of each of these methods vary, so the particular application might define which one is more suitable.

Rep-seq data, as mentioned, can provide an overview of the expressed repertoire at any given time, but the specificity and functionality of the response remains elusive. This has created an interest in predicting antibody structure from paired HC and LC sequencing data to deduce

function (DeKosky et al. 2016). A vast array of methods have been created to predict, in-silico, the antibody structure ((Sircar, Kim, and Gray 2009; Klausen et al. 2015; Leem et al. 2016; Lepore et al. 2017; Lapidoth et al. 2019) and reviewed in (Kovaltsuk et al. 2017)). These computational methods can be applied to a much greater number of sequences than experimental methods, and in some cases, can provide predictions of Ab-antigen interactions (Weitzner et al. 2017). Despite these advantages, it is worth noticing that these methods either require, or can benefit from, experimental data (Sela-Culang, Ofran, and Peters 2015; Weitzner et al. 2017). Often, they are based on previously published structural data, which can be limited, especially in different animal models (Weitzner et al. 2017). For these reasons, they are not a complete alternative to pairing Rep-seq data with mAb isolation, which is ultimately more reliable in phenotype (Jiang et al. 2013; J. Lee et al. 2016; Phad et al. 2019). Novel expression methods are being created, which greatly increase the throughput of mAb expression and link it with Rep-seq data (B. Wang et al. 2018).

Both for studies of mAbs and Rep-seq data, a comprehensive knowledge of the Ig loci and germline gene segments is required. Specifically, a complete database (DB) of germline V(D)J genes for the species being studied is required. Despite the conservation of the 5′ untranslated region (UTR) and leader sequences of V genes, mixtures of many different primers are needed for unbiased amplification of antibody sequences (Chiang et al. 1989; Larrick et al. 1989). Comprehensive germline gene DBs are also necessary for correct allele assignment, which is especially important for Ab evolution/lineage tracing. Library preparation methods and sequencing quality affect those assignments and tools have been developed to decide which sequences can be assigned with confidence (B. Zhang et al. 2015). The level of SHM has been reported to be a requirement for bNAbs against HIV-1 (Klein et al. 2013; Garces et al. 2015; Bonsignori et al. 2016), and their calculation also depends on a correct allele assignment. One study investigated whether germline allele content differed between individuals producing bNAbs (Scheepers et al. 2015). The most interesting aspect of this study was the discovery of multiple novel alleles in individuals enrolled in the HIV+ cohort, suggesting that further work is needed to fully resolve this. It was also suggested that the individual germline content shapes the naive repertoire, gene usage and CDR3 length more strongly than the specific response to an antigen (C. Wang et al. 2015). Moreover, although the capacity to mount a response against a specific antigen is not generally germline gene-dependent, certain classes of bNAbs are germline-restricted, such as the VRC01 class of antibodies (against HIV-1) using VH1-2 specific alleles (Scharf et al. 2013), and IGHV1-69 gene restricted bNAbs (against influenza) (Avnir et al. 2016). Further studies with the individual germline DBs are required to draw a stronger conclusion. As researchers move into lineage targeting for vaccine candidate epitopes, the frequency of specific alleles in the population will be important to consider. Taken together, increased knowledge of Ig genetics is extremely useful for studies of antigen-specific B cell responses.

4 IMMUNOGLOBULIN GENETICS AND POLYMORPHISM

4.1 IMMUNOGLOBULIN LOCI AND GENE DATABASES

Shotgun sequencing (Staden 1979; Anderson 1981; Gardner et al. 1981) is the method of choice for assembling large portions of the genome from shorter reads. This approach was used for the assembly of the first human genomes (Venter et al. 2001; Lander et al. 2001), as well as for many other species (Waterston et al. 2002). It consists of breaking down longer DNA stretches into short segments, sequencing them, and subsequently assembling them into contigs computationally to recapitulate the sequenced area. This approach, which is very useful for most of the genome, encounters issues in the Ig loci, since the repetitive structure characteristic of these regions makes the assembly highly error-prone, even at high sequencing depth (reviewed in (Watson and Breden 2012)). Therefore, the Ig loci are less well characterized compared to other genomic regions.

In addition to the large number of gene segments from which antibodies are assembled, there are frequent genetic polymorphisms in the Ig genes, resulting in substantial allelic diversity in the human population ((Willems van Dijk et al. 1989; van Dijk, Sasso, and Milner 1991; E H Sasso, Van Dijk, and Milner 1990; Weng et al. 1992; Boyd et al. 2010) and reviewed in (Watson, Glanville, and Marasco 2017)), much of which remains to be defined in populations from different parts of the world ((Weng et al. 1992; Eric H. Sasso, Buckner, and Suzuki 1995) and reviewed in (Watson, Glanville, and Marasco 2017)). For these reasons, even though the first Ig loci maps were generated in the mid-1990s (Cook et al. 1994; Tomlinson et al. 1994), and more complete sequences were made available soon thereafter (Matsuda et al. 1998), the Ig gene DBs are not complete to this day.

The reference DB for Ig genes is the International ImMunoGeneTics Information System® (IMGT) (http://www.imgt.org), founded in 1989 (reviewed in (Lefranc et al. 2015)). To date, the germline gene DB (Gene-DB (Giudicelli, Chaume, and Lefranc 2005)) of IMGT contains 280 IGHV, 30 IGHD, and 13 IGHJ IGH alleles denoted as functional for humans, and 103 IGHV and 14 IGHD pseudogenes and open reading frames (ORF)). Since a major proportion of the studies that defined these alleles were performed in humans of Caucasian background, the Gene-DB is biased towards alleles present in this population. Additional alleles are still being discovered (Ohm-Laursen, Larsen, and Barington 2005; Romo-González et al. 2005; Y. Wang et al. 2008; Boyd et al. 2010; Y. Wang et al. 2011; Watson et al. 2013; Scheepers et al. 2015). Moreover, many of the early IGHV alleles originally reported have been put into question (C. E. H. Lee et al. 2006; Y. Wang et al. 2008). Further studies are needed to make the reference germline Ig gene DBs comprehensive and error-free.

Another common feature of the Ig loci is that there is considerable gene copy number variation (CNV) between individuals, involving deletions and/or duplications, sometimes of relatively large regions (Pramanik et al. 2011; Watson et al. 2013; Luo, Yu, and Song 2016). CNV adds another layer of complexity and suggests a certain level of redundancy between individual IGHV genes. Although the full extent of CNV is not known, software to identify

CNVs from NGS genomic data exist (Luo, Yu, and Song 2016). These approaches, though, still rely on the availability of genomic Ig loci information and could benefit from more comprehensive allelic DBs.

4.2 GERMLINE GENE DISCOVERY

The initial steps of any genetic analysis of Igs, whether from single B cell isolations or bulk Rep-seq, is assignment to a reference DB (Calis and Rosenberg 2014). This step defines gene/allele usage and forms the basis for SHM calculations. As mentioned, the current public Ig germline gene DBs are incomplete because the Ig loci are difficult to sequence with conventional genome sequencing approaches due to the presence of highly repetitive segments, as well as high allelic and structural variation (reviewed in (Watson and Breden 2012)). Ongoing efforts are focused on obtaining higher resolution haplotype sequence information over these loci to facilitate annotation and resolve CNVs (Watson et al. 2013). Recent advances in SMRT might eventually overcome these difficulties by obtaining much longer reads to assemble, but the technology is costly, has a quality/length tradeoff, and is not widespread (reviewed in (Rhoads, Fai Au 2015)). To this end, the pairing of SMRT with 2nd generation sequencing methods has been tested to improve the quality in long reads of SMRT (Mahmoud et al. 2019). In addition, the Ig loci contain a high number of pseudogenes and other non-expressed genes, making it challenging to identify functional genes used in rearranged and expressed BCRs.

Ig gene usage in functional repertoires can only be obtained from Rep-seq. Even if the number of V(D)J gene segments produce an extensive number of possible rearrangement combinations, early studies uncovered biases in the gene usage ((Suzuki et al. 1995; Rao et al. 1999) and reviewed in (Jackson et al. 2013)), which has been reported to be consistent amongst different individuals (Boyd et al. 2010), and is affected by CNVs of the gene segment (Glanville et al. 2011). It has also been reported that gene family usage is rather stable over time in the Ig repertoire, and infections only cause a temporary skewing of the repertoire (Van Dijk-Härd and Lundkvist 2002), likely due to the expansion of short-lived antigen-specific plasma cells. Comprehensive knowledge of allelic and structural variation in the human population (and in commonly used animal models) coupled with information about the frequency with which different V(D)J genes are used in naïve B cell repertoires, would be highly useful for immunological studies.

Rep-seq analysis offers new opportunities to obtain germline data linked to gene/allele usage. Recently developed computational tools offer the capacity to infer complete genotypes, including novel alleles from expressed repertoires (Gadala-Maria et al. 2015; Corcoran et al. 2016; W. Zhang et al. 2016; Ralph and Matsen 2019). These tools and others applied to Rep-seq data have led to the discovery of increasing numbers of novel Ig alleles (Boyd et al. 2010; Gadala-Maria et al. 2015; Galson, Trück, Fowler, Münz, et al. 2015; Corcoran et al. 2016; Kirik et al. 2017; Vázquez Bernat et al. 2019; Ralph and Matsen 2019). Furthermore, Rep-

seq V(D)J data can be used for haplotype analysis (Kidd et al. 2012; Kirik et al. 2017; Peres et al. 2019), which helps elucidate the structure of the Ig loci (Gidoni et al. 2019). Moreover, although it was reported that the human Ig DBs for LC are more comprehensive (Collins et al. 2008), we and others have reported many novel alleles, suggesting that more work is required also for the LC germline allele DBs (Watson et al. 2015; Vázquez Bernat et al. 2019). Interestingly, researchers have found evidence of the importance of population-specific LC alleles in immune susceptibilities (Feeney, Atkinson 1996) and further studies of Ig LC allelic variation can be expected to uncover additional medically relevant correlates.

These advances in germline allele inference have led the AIRR community to create the IARC committee for the evidence-evaluation of novel inferred alleles (Ohlin et al. 2019). This committee created the Open Germline Receptor DB (OGRDB) (Lees et al. 2019), where novel alleles can be submitted for the committee's consideration. Novel alleles identified by existing and new computational tools, and the meticulous analysis of the OGRDB, should lead to the generation of comprehensive DBs of Ig genes in the near future, resolving a long-standing shortcoming in the immunology field.

4.3 ANIMAL MODELS

Laboratory mice are one of the most important animal models in biomedical research, due to the ease by which they can be manipulated and their inbred nature. To date, the germline gene DB (Giudicelli, Chaume, and Lefranc 2005) of IMGT contains 325 IGHV, 32 IGHD, and 8 IGHJ functional alleles for mice, and 81 IGHV, 6 IGHD and 1 IGHJ pseudogenes and ORFs. These reference alleles consist mainly of alleles from the BALB/c and C57BL/6 strains, which share few alleles between them (Collins et al. 2015). A recent study also found little overlap between the available reference DB and inbred wild-derived species, with many novel alleles identified in the latter (Watson et al. 2019). This suggests that Ig germline gene DBs for mice are also incomplete and further studies are required, especially on strains other than BALB/c and C57BL/6.

The IMGT DB contains even less information for other small animal models, with only 39 IGHV, 10 IGHD and 11 IGHJ alleles listed for rabbits and no information listed for guinea pigs or ferrets, which are also frequently used in research. Focused efforts to develop germline Ig DBs also for these species and others are, therefore, needed.

Small animal models are not suited for all research questions. In some cases, non-human primates, usually macaques, are required. Macaques offer a high degree of genetic similarity to humans (94% average), as well as similar anatomy and cell types. Also, several cell surface markers used to delineate immune cells are highly conserved between humans and macaques, meaning that many reagents used to phenotype cells by flow cytometry are cross-reactive. For these reasons, macaques, and specifically rhesus and cynomolgus macaques (reviewed in (Roos and Zinner 2015)), are commonly used in biomedical studies. An additional reason to use macaques is that they are susceptible to certain viral infections that small animals are not

(reviewed in (Estes, Wong, and Brenchley 2018)). For example, for studies of HIV-1, closely related simian immunodeficiency virus (SIV) (Chakrabarti et al. 1987), or viruses that are chimeras of SIV and HIV-1 (so called SHIVs (J. Li et al. 1992)), have shown utility as challenge models in macaques. In 2012, our group described a high degree of homology in Ig genes between macaques and humans, with a similar V gene family distribution (Sundling et al. 2012). Studies by others suggested that despite differences in the specific germline Ig genes between macaques and humans, their responses converge with similar characteristics in IgG class-switched responses (Vigdorovich et al. 2016), underlining their utility in immunogenicity studies. Currently, there is considerable ongoing effort in the field to obtain a better understanding of macaque Ig genes and alleles, as described below.

For rhesus macaques used in medical research the two most common origins are India and China. India imposed an export ban on rhesus macaques in the 1970s. However, prior to this, animals were exported to the 7 large National Primate Centers in the USA (NPRCresearch.org), each of which houses its own self-contained breeding colony. Thus, while USA scientists primarily use Indian origin rhesus macaques, much of the rest of the world uses Chinese origin rhesus macaques imported from breeding facilities in China. In my thesis work, I mainly analyzed samples from animals maintained in the Astrid Fagreaus Laboratory (AFL), an AAALAC (Association for Assessment and Accreditation of Laboratory Animal Care)-accredited research facility at Karolinska Institutet equipped to handle work with macaques and other animal species. The AFL does not have its own macaque breeding colony and imports animals upon request - usually Chinese origin macaques. I have also collaborated with two European primate facilities from which we obtained samples (Göttingen in Germany and CEA in France) and with two US primate facilities, the Oregon National Primate Research Center outside Portland (ohsu.edu/onprc) and the Yerkes National Primate Research Center in Atlanta (yerkes.emory.edu/) from which we analyzed samples or shared protocols.

Aside from Indian and Chinese origin rhesus macaques, several other macaque species are used in biomedical research, the most common being cynomolgus macaques. Cynomolgus macaques are distributed across Southeast Asia, spanning Myanmar, Thailand, Cambodia, Vietnam, Indonesia and the Philippines (**Fig. 4**). There is no physical natural barrier between the habitat of cynomolgus and rhesus macaques, meaning that there is likely a significant genetic overlap between these two types and introgression between the species was reported (Bonhomme et al. 2009). An interesting sub-group are the Mauritian cynomolgus macaques, a population of specific interest in biomedical research. These macaques are said to have been imported to the Mauritian island during the 16th century (reviewed in (Sussman and Tattersall 1986)). The small number of animals imported (estimated less than 20 individuals from Java or Indochina (Bonhomme et al. 2008; Osada et al. 2015)) gave rise to a large colony, today estimated to be around 50,000 animals. Previous genetic analyses of these animals suggest a lower degree of inter-individual diversity compared to Malaysian cynomolgus macaques (Osada et al. 2015), consistent with a small founder population. Their decreased genetic diversity has been mainly studied in the context of the MHC (Leuchte et al. 2004; Krebs et

al. 2005; Budde et al. 2010; Blancher et al. 2012), where it is estimated that Chinese-breed cynomolgus macaques (Vietnamese, Cambodian and Indonesian origin) display a ten-fold higher level of diversity compared to their Mauritian counterparts (Karl et al. 2017). In this thesis, I collaborated with scientists at the CEA in Paris where Mauritian cynomolgus macaques are housed and used in vaccine trials (periscope-project.eu/consortium/cea/).

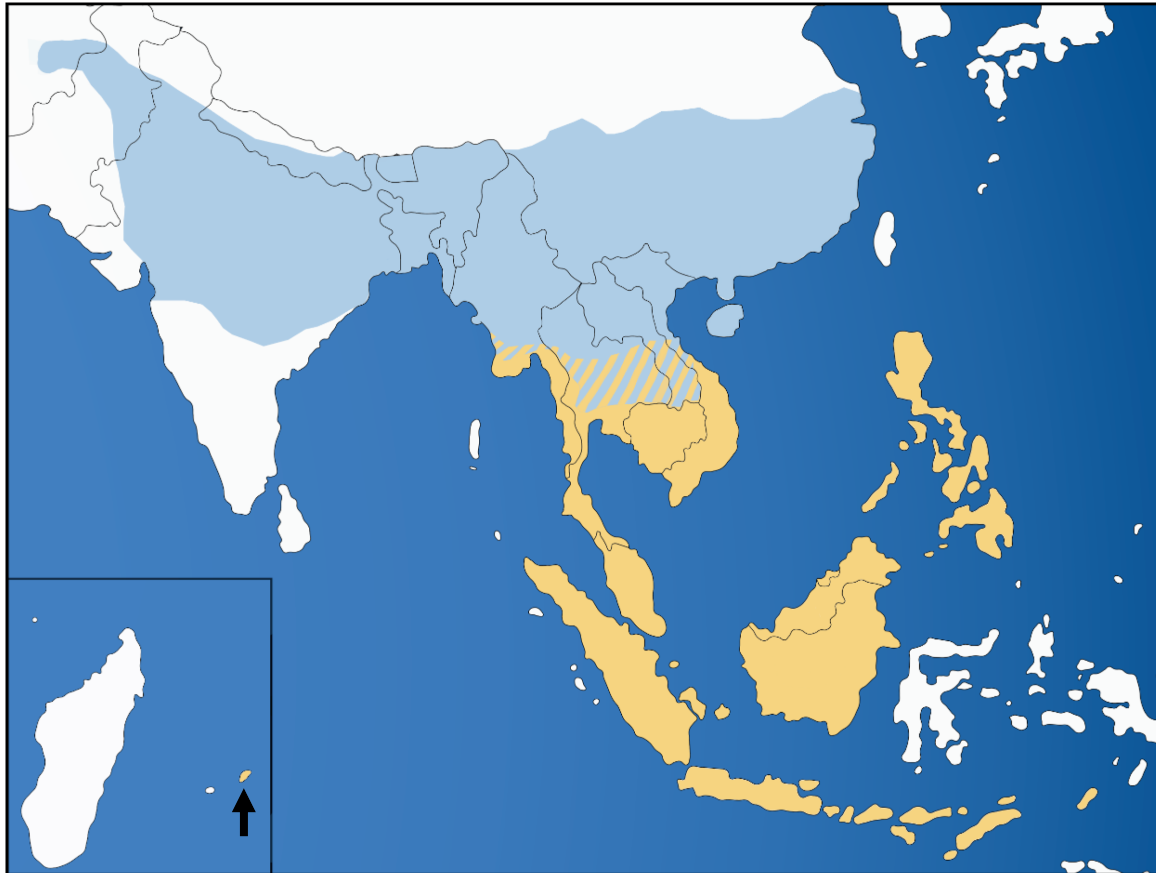

Figure 4. Approximate distribution of rhesus (blue) and cynomolgus (orange) macaques and their overlap (striped) (based on (Street et al. 2007)). Insert in the bottom-left corner which includes the island of Mauritius, located near Madagascar.

The first rhesus macaque genome was published in 2007 (Gibbs et al. 2007), while the cynomolgus genome was reported in 2011-2012 (Ebeling et al. 2011; Yan et al. 2011; Higashino et al. 2012). To date, the Ig germline gene DB (Gene-DB (Giudicelli, Chaume, and Lefranc 2005)) of IMGT contains 19 IGHV, 24 IGHD and 7 IGHJ functional alleles for rhesus macaques, and 62 IGHV, 25 IGHD and 6 IGHJ functional alleles for cynomolgus macaques. These numbers are very low when compared to humans and mice, suggesting that the reference DBs are incomplete. Indeed, many studies have reported a much greater number of Ig alleles (Francica et al. 2015; Corcoran et al. 2016; Ramesh et al. 2017; Rosenfeld et al. 2019; W. Zhang et al. 2019; Kong et al. 2019; Cirelli et al. 2019; Sundling et al. 2012). The high levels of intra-species variation seen in macaque IGHV, compared to human, coincide

with variability in other immune loci (Otting et al. 2005) and other regions of the genome ((Fawcett et al. 2011; Yuan et al. 2012) and reviewed in (Rogers and Gibbs 2014)). Macaque genetic variability, more generally, needs to be further studied and incorporated into guidelines for biomedical research and breeding facilities (Haus et al. 2014).

Despite efforts to obtain higher quality Ig loci genomic data and gene annotations from these macaque species (Zimin et al. 2014; Osada et al. 2015; Yu et al. 2016; Ramesh et al. 2017; Cirelli et al. 2019), the previously described challenges associated with assembling sequence reads over these loci, and the inter-individual macaque diversity, make the currently available DBs incomplete. Moreover, the Southeast-Asian cynomolgus macaques are widely distributed and have differentially intermixed with rhesus macaques, making their nominal origin uninformative and genetic background controls necessary (Osada et al. 2010). Altogether, this underlines the importance of applying new inference tools to generate improved Ig germline allele DBs for animal models - especially for rhesus and cynomolgus macaques - in combination with genomic sequencing to obtain more reliable results when using these species in immunological studies.

5 AIMS

The specific aims for the individual papers were:

Paper I: To develop a computational approach for individualized Ig germline allele inference from expressed antibody repertoires, to identify the full allelic content from an individual, including novel alleles.

Paper II: To determine the advantages and limitations of the two most common library preparation methods, 5′ rapid amplification of cDNA ends (5′RACE) and 5′ multiplex amplification (5′MTPX), by comparing the sequence quality and potential biases affecting the output data.

Paper III: To investigate Ig germline allele diversity and overlap between groups of Chinese and Indian rhesus macaques, and Indonesian and Mauritian cynomolgus macaques, using germline allele inference, and to generate comprehensive DBs of well-validated IGHV, D and J alleles for these species.

6 RESULTS AND DISCUSSION

6.1 GERMLINE INFERENCE (PAPER I)

To address the insufficiencies of the current Ig germline DBs, our group developed a computational tool, IgDiscover, which allows rapid identification of known and novel Ig V alleles from expressed antibody repertoires of humans and research animals. It analyses full-length rearranged VDJ sequences from IgM libraries, or VJ sequences from IgK and IgL libraries, sequenced on Illumina's MiSeq platform. From this analysis, the IgDiscover algorithm generates personalized V allele DBs and identifies novel alleles present in each individual.

The algorithm of IgDiscover can be divided into four main steps: assignment, discovery, filtering, and replacement. Briefly, after the reads are merged (PEAR (J. Zhang et al. 2014)) and preprocessed, they are assigned to an initial DB with IgBlast (Ye et al. 2013). The algorithm then locates sequences in each of the assignments that cluster aside from the reference gene segment and generates novel allele candidates from the consensus of those sequences, removing random PCR and sequencing error (**Fig. 5**). These candidate alleles are then filtered to discern real germline alleles from spurious sequences by counting the number of individual rearrangements they are found to be associated with. IgDiscover computes all the CDR3s, Ds and Js associated with the V sequence in a cluster, as well as the frequency of the most common CDR3 length. A real germline allele is expected to be found in multiple independent rearrangements with different Ds and Js and different non-templated regions, yielding different CDR3s with a distribution of lengths. Candidate alleles that lack evidence of multiple rearrangements are discarded (**Fig. 5**). Once a new V DB is generated with the sequences that passed the filter, IgDiscover substitutes the initial DB for the new DB and, in case the initial DB was too limited, iterates the process several times to infer all unique alleles present in each individual.

We evaluated IgDiscover using mouse, human and rhesus macaque IgM repertoires and demonstrated that, not only could the software recover alleles intentionally removed from the input DB, but it also inferred several previously undescribed alleles in all three species. Rhesus macaques displayed a high degree of IGHV allelic diversity; the combined IGHV DB obtained from five Chinese rhesus macaque had 240 unique alleles, of which 30 were found in the Indian rhesus macaque DB. To validate alleles, we designed primers for targeted PCR of genomic DNA and Sanger sequenced 42 allele segments that matched 100% those obtained by IgDiscover inference, of which 34 were novel.

During the development of IgDiscover, we tested whether it could be used with a minimal initial DB, such as one with just one allele from each of the 7 IGHV families, or with an Ig DB from another species. This means that this approach can be used to construct germline antibody gene DBs from animal models with little or no prior information about the Ig loci. The IgDiscover tool could therefore be used to generate comprehensive DBs for multiple research animal species, including those commonly used in infection and immunization

studies, such as the guinea pig (Kennedy et al. 2019; Bernstein et al. 2019; Evseenko et al. 2019; Abhishek et al. 2018), rabbits (McCoy et al. 2016; Tran et al. 2019), ferrets (Martina et al. 2003; De Jonge et al. 2016), camelids (McCoy et al. 2012; Forsman et al. 2008), and several primates (Geisbert, Strong, and Feldmann 2015), with very limited Ig DBs. Thus, inference tools such as IgDiscover will help bridge the knowledge gap and offer an efficient way to define germline alleles in a variety of species.

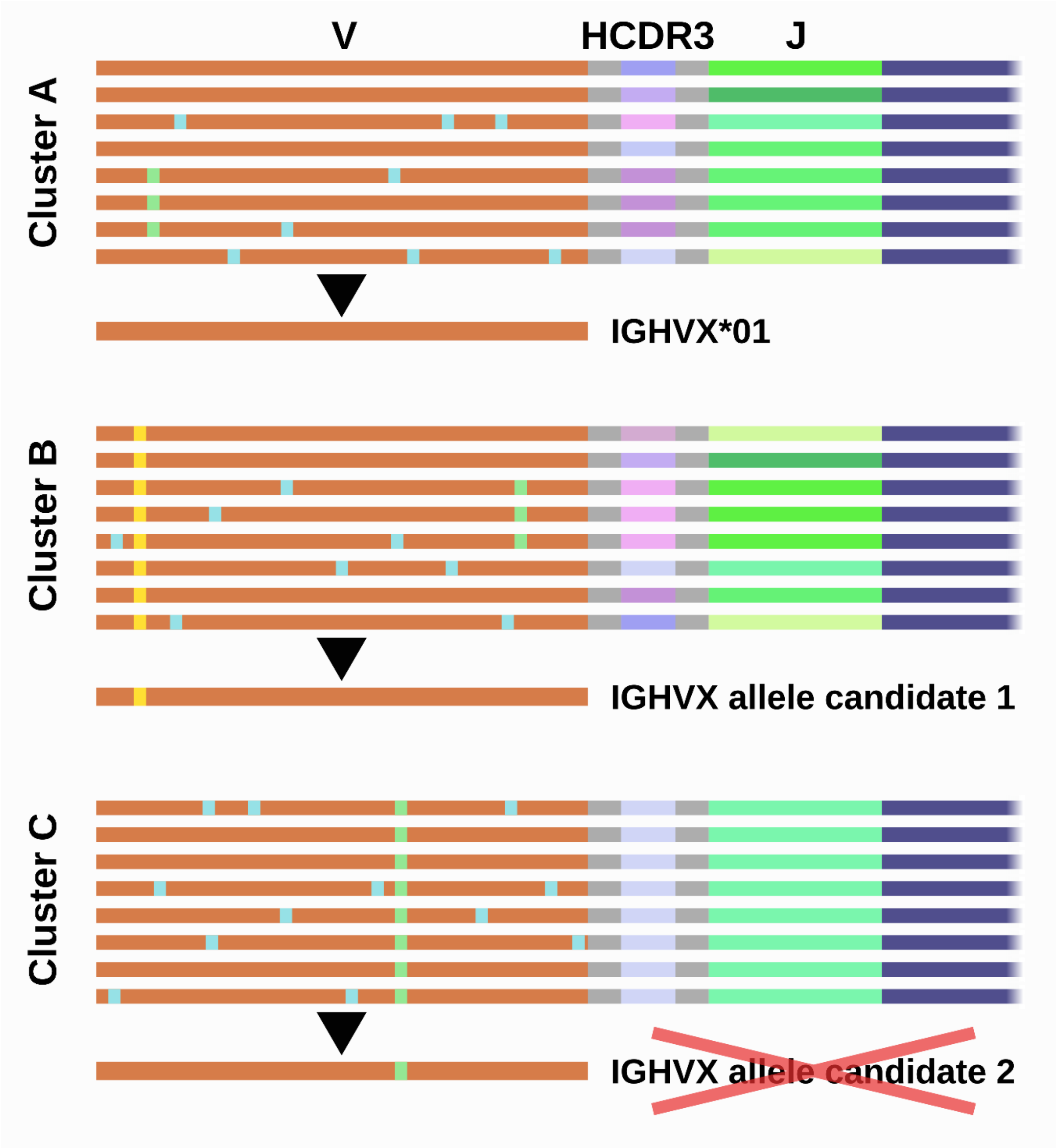

Figure 5. Schematic of the consensus building and filtering of alleles in IgDiscover in three kinds of sequencing clusters: known allele (A), real novel allele (B) and false novel allele (C). Over the V segment there are represented single nucleotide polymorphisms in yellow, SHM in green and PCR and sequencing errors in light blue.

6.2 LIBRARY PREPARATION (PAPER II)

Both Ig germline gene inference and Rep-seq analyses rely on robust library preparation methods for sequencing. The most common techniques used to produce the amplicons are 5′RACE (Frohman, Dush, and Martin 1988) and 5′MTPX (Chamberlain et al. 1988). A main objective during my doctoral training was to develop improved library production protocols for HC and LC Rep-seq analysis for both PCR methods. The improved 5′RACE method included a novel approach to shorten the total amplicon length by 20-25 bp via the inclusion of Illumina's Read1 sequence as universal forward primer in the template switch reaction. The 5′MTPX protocol included sets of primers for the human heavy, kappa and lambda chains, which were designed based on the information obtained from the 5′RACE libraries, and located in the upstream leader region of the V gene segments to obtain coverage of the full-length V(D)J sequence. Both protocols included UMIs to correct for potential biases in the template amplification and instrument-introduced errors.

For this paper, I produced HC libraries from six human volunteers, which I sequenced using the Illumina MiSeq platform and analyzed with IgDiscover. The analysis demonstrated that the 5′RACE amplicons were on average 82±2 bp longer than the 5′MTPX amplicons. 5′RACE is often considered to be advantageous to avoid the potential primer biases generated by 5′MTPX (Baum, Venturi, and Price 2012; He et al. 2014). In our paper, we found that with 5′RACE, the amplicon length for some of the VDJ sequences exceeded the limit of the MiSeq 2x300 bp V3 kit. Particularly, we observed a potential bias against IGHV3 family genes, since their 5′UTRs are the longest. This was reflected as a lower representation of IGHV3-using VDJ sequences in the data obtained by 5′RACE compared to the data from the 5′MTPX libraries, which could specifically affect VDJ sequences with longer HCDR3s.

Furthermore, we found a decreased percentage of sequences matching the inferred germline alleles in 5′RACE, despite the fact that the libraries were generated from the same starting mRNA. We observed that the increased amplicon length for 5′RACE caused the inclusion of more low-quality bp from the end of each sequencing read, which could not be corrected due to the smaller overlap between the merged reads in 5′RACE. Overall, we estimated a 20% decline in error-free sequences in 5′RACE, assuming per-read error is approximately Poisson distributed and averaging across all reads in each dataset. In all six subjects, the number of IGHV germline alleles inferred was lower for 5′RACE libraries. In brief, we identified, on average, 11 alleles more per individual with the 5′MTPX method. These findings demonstrated that with the current sequencing technologies, the 5′MTPX method is superior.

We also tested the initial mRNA template amount in the 5′MTPX PCR and found that less than 200 ng of mRNA yielded fewer HCDR3s, HCDR3/UMI% and Ds per inferred allele. Finally, I produced IGHV, IGKV and IGLV libraries from one human individual, starting with 200 ng of mRNA and using the 5′MTPX method. We analyzed the sequenced libraries with IgDiscover to produce heavy, kappa and lambda individualized Ig DBs. We identified 55 IGHV, 37 IGKV and 40 IGLV alleles, of which three IGHV, one IGKV and six IGLV were novel. Interestingly, it was previously reported that the IMGT's LC DB (IGKV) is more

comprehensive than HC (Collins et al. 2008). However, our results from just one Caucasian individual seem to indicate otherwise, illustrating that further studies in LC Ig germline genes should be conducted.

An important consideration is that IgDiscover utilizes a consensus building mechanism; thus, spurious sequencing and PCR errors are excluded from the inferred output. The higher error frequency found in the 5′RACE libraries could have a greater impact on lineage tracing analyses or studies of repertoire diversity. Our results indicate that the library preparation method, primer design and amount of template used all have effects on the efficiency of germline gene inference. These parameters therefore need to be taken into consideration.

6.3 MACAQUE IMMUNOGLOBULIN GERMLINES (PAPER III)

In paper III, the objective was to investigate the IGHV germline genes in rhesus and cynomolgus macaques from four sub-populations, Indian and Chinese origin rhesus macaques, and Indonesian and Mauritian origin cynomolgus macaques. I first produced 5′RACE libraries from eight macaques (two per sub-population) and employed IgDiscover to obtain their allelic upstream sequences. I then designed two sets of 5′MPTX primers, placed in the 5′UTR and the leader region respectively. I produced one library with each of the sets of 15 Indian rhesus, 12 Chinese rhesus, 12 Mauritian cynomolgus and 6 Indonesian cynomolgus, and sequenced the libraries with Illumina’s MiSeq 2x300 bp V3 kits (90 libraries in total). I analyzed the libraries using IgDiscover, filtered the output and obtained individualized DBs for all 45 macaques.

For this study, we used as the initial DB a recently published set of 66 IGHV genes and 103 IGHV alleles obtained from full genome sequencing of one Indian rhesus macaque (Cirelli et al. 2019). From the 45 macaques, IgDiscover re-identified exact matches for 54 genes and 63 alleles described in the initial DB. In our initial analysis, we discovered IGHV gene segments that clustered separately from the reference DBs genes, suggesting that there may be other genes in addition to the 66 IGHV reported in Cirelli et. al. (2019). We therefore incorporated 26 of those sequences in the input DB of the analysis under the ID NGC (novel gene candidate) to avoid allelic exclusion. The “allelic ratio” filter employed by IgDiscover, removes alleles expressed below an adjustable ratio (0.12 default) of the highest expressed allele for that gene, which could erroneously remove mis-assigned alleles if some genes are not present in the initial DB.

IgDiscover also uses the number of IGHD and IGHJ associated to a given IGHV segment to filter out false positives. Therefore, to identify as many Ds and Js as possible we designed primers encompassing the 37 IGHD and 7 IGHJ genes in the initial DB in BLAT (UCSC genome browser (Kent 2002)). We then PCR-amplified the genomic DNA from the macaques using these primers, and sequences were cloned into plasmids and Sanger sequenced. This yielded 19 IGHD and 6 IGHJ additional novel alleles, which we incorporated into the input

DB to improve the performance of IgDiscover. These Ds and Js were shown to be expressed after the analysis.

We obtained a total of 1307 IGHV alleles, of which 641 were found in more than one animal, which serves as a means of validation (Collins et al. 2015; Scheepers et al. 2015). We found 445 IGHV alleles in more than one animal that were not previously reported (297 for cynomolgus macaques and 221 for rhesus macaques). Furthermore, 42 of the alleles we found only in one animal were reported in previous studies. We designed genomic primers encompassing several of the gene segments. These were amplified, cloned and Sanger sequenced. This further validated 77 alleles (26 of which were found only in one animal). We produced DBs for rhesus and cynomolgus macaque Ig alleles, which were validated by at least one of the methods described above (cross-validation in more than one animal, targeted PCR and genomic sequencing, or previously reported). The full rhesus macaque DB contained 461 IGHV, 52 IGHD and 15 IGHJ alleles and the cynomolgus macaques DB contained 416 IGHV, 50 IGHD and 15 IGHJ alleles.

In the DBs generated for the 45 animals, the rhesus and cynomolgus macaque DBs shared 168 IGHV alleles, which is 26.1% of the 641 alleles present in more than one macaque. Indian and Chinese rhesus shared 172 IGHV alleles (42.4%), and Mauritian and Indonesian cynomolgus macaques shared 66 IGHV alleles (16.4%). We analyzed the diversity by the number of unique average alleles found per macaque sequenced and the average percentage of alleles found in just one macaque, and found that the Indonesian cynomolgus were the most diverse group and the Mauritian cynomolgus the least, with Chinese rhesus macaques being slightly more diverse than the Indian rhesus macaques. These data agree with previous reports of cynomolgus macaques from Mauritius having a less diverse genome (Osada et al. 2015), MHC genes (Karl et al. 2017), and mitochondrial DNA (D. G. Smith, Mcdonough, and George 2007) compared to their Southeast-Asian counterpart.

This work extended our knowledge about macaque Ig allele diversity and offered compelling evidence for the need to use individualized approaches if the goal is to understand the ontogeny and maturation of antibody repertoires in macaques. For genetic studies of antigen-specific B cell responses (Rep-seq analysis or mAb isolation), there is a need to obtain individualized macaque Ig gene DB to correctly assign antibody sequences and accurately calculate SHM.

7 CONCLUDING REMARKS AND FUTURE PERSPECTIVES

The work described in this thesis constitutes a strategy for generating Ig germline DBs for any animal species, regardless of the Ig genomic information available. We demonstrated how libraries can be generated with only constant region genomic information using 5′RACE, to produce 5′ Ig V sequence information that can be used for 5′MTPX primer design, which in turn yields libraries well within the capacity of the Illumina MiSeq platform. Knowledge of the total number of V genes in humans and mice makes it abundantly clear that the DBs for research animals such as macaques and many small animal species are incomplete. The inference process described in this thesis, using the publicly available IgDiscover computational tool, can help create such DBs in a faster and more economic manner than what can be achieved using genomic sequencing methods.

Ig germline gene inference is not a substitute for long-read genomic sequencing technologies like SMRT. The latter can provide coordinates and distribution of genes, discern what are genes and alleles, and identify placement of duplications and deletions events. A shortcoming of genomic sequencing approaches is the challenge of knowing which alleles are functional or not, especially given that the Ig loci harbors large numbers of pseudogenes and non-expressed alleles that appear in genomic data but have no impact in the expressed, rearranged repertoire. Thus, germline inference using Rep-seq data and genomic sequencing can be considered complementary approaches. The combination of these two approaches will be important for the immunology field in the future.

Germline gene identification and assignment is critical for antibody lineage studies since misidentification of the parental germline alleles not only result in inaccurate definition of clonal relatedness and overestimations of SHM, but also alter phylogenetic trees of the antibody evolution. This highlights the importance of individualized DB approaches in species with high genetic diversity. The extremely high level of diversity found in macaques makes these findings and the experimental approaches taken here highly relevant to the vaccine field. Further studies are necessary to generate comprehensive DBs for other species commonly used in biomedical research.

8 ACKNOWLEDGEMENTS

First and foremost, I need to thank my main supervisor **Gunilla Karlsson Hedestam**. From the day we met, you treated me already as a part of the team, and when I asked to continue in the group as a PhD student, you did not hesitate to give me a chance. Your love for science, drive and passion are unparalleled, and despite the multiple challenges of the modern day group leader, you continue to put a tremendous amount of effort in making the group the best place possible. You have ears for everyone, and there is never an issue too small to bring up to you. You also taught me to be critical of myself and the data, and to always look for a better angle or other explanations. Going in to show you the results was always a challenge. You are strong-willed and like to give input in any facet of a manuscript/presentation but have always good suggestions and an open mind for good arguments. Thanks for the opportunity and your patience.

My co-supervisor, **Martin Corcoran**, you showed me what research should be like. Your creative and borderline mad scientist ideas often paid off big time, and you always had one more test to do. I must admit that my impatient being sometimes got a little exasperated by your never-ending experimental suggestions, but there was no experiment or result that did not improve our projects afterwards. Time for questions always existed even if you were busier than I have ever been, and your dark and witty humor lightened up any boring or annoying time.

My mentor, **Fredrik Lanner**, who I should have met many, many more times, since every time we had a talk you provided me with wise and encouraging words.

Ben Murrell, thanks for being the chairperson and for all your patience with my analysis questions. I know it was hard for you that I did not learn to code in your beloved Julia, but you were always extremely helpful. We never had a discussion about science where I did not learn much more than what I came asking for.

To my colleagues at GKH lab, you guys were the bedrock of this PhD experience. **Monika**, you are an unbelievable colleague to have, cool under pressure, always smiling and with positive words. I wish I could have learned a lot more from you in these years, especially on how to be so optimistic and happy. To **Sanjana** for caring so much, sharing food and getting me nervous at the wrong times, and **Pradee** for your kindness and goodwill and for never listening to any of my unrequested advice. **Uta** for the discussions about data analysis and for all the times you tried to change my vocabulary, demeanor, noisy feet or noisy chewing. **Xaquin**, for joining the group and picking up the macaque project. **Mateusz**, for being so different and funny and for the many monologues we shared together. Your “out of this world” humor always cracked me up. **Marco**, for being the “yes” man even more than you should, running faster than you could, and not agreeing with me on almost anything serious. You were a big help when I needed it the most and a much better person than you give yourself credit for. Sometimes all the help one needs is to see a friend running ahead. Finally, **Sharesta**, my PhD

buddy, you were here since week two of my life in the group and never ceased to be an unbelievable friend and a supporting colleague. We have been through thick and thin together. Always inappropriate when necessary and with little filter. For editing 95% of the texts I have written in the lab, even these acknowledgements, so my inability to write coherent texts does not see the light of day. The word "impossible" has no meaning for you, and I hope some of your unmatched optimism and strong will rubbed off on me, for I definitely feel like a better person after having met you. Words are certainly not enough to express my gratitude to you.

To the students I had the honor to supervise, since they all taught me more than I taught them. **Komal** and **Laura** for being my student guinea pigs and having so much patience with my long and often weird explanations. **Izabela**, for being the help I so desperately needed during the last year of my PhD, and being crazy enough to stay in the lab after six months with me as supervisor. I think I could not have gotten a better help and support. Good to have someone to talk and share some interesting movies with while it felt that chaos was all around me.

To the "extended GKH lab constellation", **Gerry**, **Marc** and **Leo** for showing passion for science and research. **Ainhoa** por quejarnos siempre del mundo entero y por esforzarte ne sonreir. **Ben**, for the good times we remember and the ones we do not (and might be better that way). **Lifeng**, for your kindness, generosity and good heart, and for showing me how much one can do against all odds. An endless supply of delicious tea and good times around a ping-pong table. To **Jonathan** for showing how "cool" scientists can be and how to correctly pronounce data. **Junjie**, for our noodle search in Riga and for not checking in the liquor. **Julian** for your not-so-funny but interesting facts and always being up for a good discussion. **Chris**, for your charm and your accent, and for your fabulous country of origin. We never had as many scotch tastings as I wanted, but every one of them was a blast. **Leo,** for being the most Australian German ever, our deep and shallow conversations and all the coffee breaks. Your capacity to be cool and sociable under almost any circumstance is amazing and I hope I learned a thing or two from you. You forced me to dance way above my level, appreciated the real me and were not afraid to set me straight.

The people that have been in the lab in the past, **Marjon** for her cynical view on academia, **Bastian** for asking more questions than you answered and being a source of endless smiles and positive comments. **Paola** for showing me how to fight for what you want, **Elina** for encouraging me to learn Swedish, making me look like I did when needed, and putting me in my place when it was necessary, and **Ganesh** for your weird table tennis service. **Lotta** for always bringing good humor and all the effort to any endeavor, and **Gabriel** for making science look easy. **Martina** you had the best protocols ever conceived and the most experience with protein purification that I have ever seen. **Vika** your humor and bold attitude were refreshing and helpful in hard times.

I especially want to thank all the co-authors, which I have not previously mentioned, for their collaboration in making the publications included in this thesis possible: **Christiane**, **Nori**, **Mats**, **Marcel**, **David**, **Nancy**, **Pauline**, and **Roger**.

KI long term staff, **Patrick** for the morning classes and trusting me to lead them, **Tina** for your dedication and commitment, y **Martin R.** por tu buen rollo y las conversaciones de futbol y politica. **Åsa** for being the best student administrator ever, helping me with my Swedish and always having a moment and **Gesan** for your concern for the well-being of the students. **Kristina** for your hard work and patience with me and other students.

To the DSA folks with our endless meetings, even though the changes we enacted were small, the company while trying to enact them was great. **Benedek**, for being a stickler to the rules until those rules bother you and generally being my nemesis. Always chuckling and concerned for others, you got your better medal in the end and it was definitely well deserved. **Iulia**, for being our perfect dictator, firm but caring. **Leif**, for trying to control the meeting and mostly failing to contain Southern tempers. **Qiaoli**, for your commitment to perfection in a messy world. You cannot always make everyone happy, but you always tried nonetheless. **Eva**, for being the embodiment of the relentless social fight and showing me the most impact is done without talking much or loudly. **Mirco**, your passion for bureaucracy is incredible, and your incapacity to not weigh in in every subject was a cruel but fair reflection of myself. You are kind and nice and tolerated way too much from me. **Sebastian,** for bike support and sharing the vegan experience for a month. **Yildiz**, for your optimism while overworked, mindful positive editing and brilliant counterpoints. **Ayla**, your positive attitude and fighting spirit were an inspiration. **Susanne**, for always getting me off balance and being your radiant self. **Hannes,** for never being "appropriate" and ventilating the sauna properly. **Henna**, for your capacity to listen and your empathy. **Chenhong**, for laughing at everything all the time.

All my other KI friends, **Silke**, for your cheerfulness and German charm. **Mitch**, for showing me to be humble with my designs and movies. **Vanessa**, for sharing my eco-concerns and your good humor. **Mariana**, por ser muito muito legal. **Leonie**, for your kindness and generous laughter. **Anton**, for your smile, the help with my disastrous Swedish and the medical and personal advice. **Thomas** for your crazy attitude, and **Amanda** because partying with you is dangerous. **Joanna** for being an excellent role model as a president and always being cool. **Milind** for our discussions and your charm. **Shady**, for being so uniquely nice all the time and asking "how are you" and really meaning it. **Jaime** for never having had time to really meeting you but anyways knowing how great you are. **Carina**, for your unlimited encouraging words. **Tatianna**, per acompanyar me en l'aventura en la distancia i per aquells retrobaments de 20 segons de passada queixant-nos del PhD i el món. **JP**, for being the most generous and easy-going person I have ever met and making every encounter fun (maybe too much). **Sandra**, per la manera tan positiva que tens de prendre't la vida i la teva energia. **Álvaro** y **Rafa**, por vuestro salero y gracia. **Lamberto** y **Patricia**, por vuestra naturalidad y buen rollo. **Luisma**, por ser un motivao, un viciao y hablar por los codos, y **Gema** por aguantar mis bromas pesadas sobre Darwin. **Carles**, per les trobades a la feina, al carrer, a l'avió...i el teu bon rollo. **Astrud**, for making photography look easy. **Lil**, pal, for conquering your bouldering fears, always being fun and razor-sharp. **Aurelie**, for telling me my Martini was awful. **Martin S.**, "der Übermensch", perfect friend to push my physical boundaries and challenge my beliefs. Amazed that you tolerated my constant poking and incessant "today I don't feel it" attitude. I

feel blessed to have encountered you. Finally to my roomies, **Nati** and **Jojo**, who basically adopted me and never let me feel lonely, because without your help, patience, support, craziness, hugs, food, drinks, parties, shared tears and smiles I would not have survived. I owe you more than I can ever repay.

Non-KI folk too! **Emelie** for your brilliant laughter and your combative feminism and social views. **Lluis**, **Miriam** i **Guim** per els tes a mitja tarda, les passejades i ferme recordar la terra. **Rocio** por tu buen humor y tu buen rollo. Podría aprender mucho de ti, si supiera aprender. **Marc P.** perquè estem massa d'acord en tot i haver-nos conegut fa dos dies, i perquè vas tenir paciència amb els meus hobbies de disseny i programació. Ets un crack! **Andrew** for generously teaching me photography and beer culture and how to harass a harasser.

International friends who despite being far away, sustained me when I needed it and were an all-present source of joy, especially when we did meet. **Vinnie**, perché senza il tuo e "savoir faire" italiano puro e perfetto la mia vita sarebbe sicuramente più desolante e il mio guardaroba molto meno elegante. **Jessica**, per tot l'esforç que poses a la nostra amistat fins i tot quan jo fallo, ets molt gran, i per ser el teu segon amic després del **Marc S.**, qui deu ser la persona més alegre que he conegut mai. **Maialen**, porque detrás de lo arisca que eres se esconde mucho. **Paco**, porque al lado de la definición de buena persona en el diccionario no hay otra foto que la tuya. Nunca llegaré a ser tan bueno y generoso como tu, pero pienso intentarlo. **Xavi**, **Robert**, **Edu** i **Edgar** per les festes de nadal junts i els records. **Steven**, for your slow mornings, happy with everything attitude and crazy energy only surpassed by **Greta,** la hormiga atómica and word machinegun who turns any occasion into a party. **Julia**, for our serious conversation topics and our foolish ones. **Camila**, because life was indeed sweet enough. **Idoia**, perquè ens veiem molt poc però ens sentim molt aprop. **Albert**, per les converses, les discussions, les bromes, els silencis (tot i poquets) i les històries viscudes (que trigues vint minuts a explicar).

Moltes gràcies **Sara**, de tot cor. No hauria vingut ni a Suècia si no fos per tu i en aquests 5 anys i mig al nord he après moltes coses de tu, d'entre elles a saber quan ser absolutament optimista és bo i a mai creure que quelcom és impossible. M'has acompanyat en la major part d'aquesta aventura al nord i mai hauria pogut triar millor companya de viatge. Vull fer les gràcies extensives a tota la teva família, i en particular a la **Carmen**, el **Lluís** i la **Quica** qui em van cuidar i tractar com a part de la família des de el primer minut.

Finalment, res hauria estat possible sense la meva família a la qual estimo molt encara que sigui des de la distància. **Tieta**, amb les teves abraçades d'os i les marqueses de xocolata, **Edgar,** el meu primo de zumosol, amb el waterbasket, copichuelas i les pales a la platja, **Estela**, la mejor compañera de canasta, **Aran** el terremoto i astrònom en potencia y **Telma** la que em moro de ganes de conèixer. **Sergi**, per ser un buenazo sempre, **Patri** por centrarnos al primo y "en verdad" ser una gran incorporación a la familia, **Andrea** pel teu sentit de la moda que sempre m'omple l'armari al nadal i les teves abraçades i petons. **Tere**, por malcriarme todo lo que puedes, **Sebas** pel teu sentit de l'humor, **Carlos** por tus ganas de discutir y **Abuela** por tu corazón tan grande y el amor que me has dado. **Avia** i **Avi**, que em veu criar i cuidar tota la vida. Mai entendré com persones que van passar per tant podien tenir tota aquesta bondat i

amor a dins. Us trobo a faltar cada dia i hauria donat el que fos per tenir-vos encara aquí. **Nil**, gràcies per ensenyar-me a ensenyar i a tenir més paciència (encara que no ho sembli). Sé que no sempre he estat el millor germà gran, però sempre té estimat i he sentit la teva estima. No podria haver demanat un germà petit més gran en tots els sentits. **Pares**, potser sí que els pins són massa alts i que és la confitura el que importa i no el pot. **Papa**, tienes tus defectos como cualquier ser humano, pero algunos son virtudes disfrazadas. Gracias por enseñarme a defenderme, pelear y a ser tozudo como nadie, y que a veces ser un pesimista te ayuda a estar preparado. Gracias por darme tanto. **Mama**, tot i que tothom sempre em diu que sóc com el papa, són les parts en les quals he sortit a tu que la gent valora més. M'has donat la vida, l'amor, el costat artístic, la humilitat (de vegades massa) i tot el que sempre he necessitat i més i no hi ha prou paraules per agrair-ho tot.

THANK YOU ALL

9 REFERENCES

Abhishek, B Kumar, Anjay, A K Mishra, C Prakash, A Priyadarshini, and M Rawat. 2018. “Immunization with Salmonella Abortusequi Phage Lysate Protects Guinea Pig against the Virulent Challenge of SAE-742.” *Biologicals : Journal of the International Association of Biological Standardization* 56 (November): 24–28. https://doi.org/10.1016/j.biologicals.2018.08.006.

Afik, Shaked, Gabriel Raulet, and Nir Yosef. 2019. “Reconstructing B-Cell Receptor Sequences from Short-Read Single-Cell RNA Sequencing with BRAPeS.” *Life Science Alliance* 2 (4): e201900371. https://doi.org/10.26508/lsa.201900371.

Allen, Christopher D.C., and Jason G. Cyster. 2008. “Follicular Dendritic Cell Networks of Primary Follicles and Germinal Centers: Phenotype and Function.” *Seminars in Immunology*. https://doi.org/10.1016/j.smim.2007.12.001.

Alt, Frederick W., T. Keith Blackwell, and George D. Yancopoulos. 1987. “Development of the Primary Antibody Repertoire.” *Science* 238 (4830): 1079–87. https://doi.org/10.1126/science.3317825.

Anderson, Stephen. 1981. “Shotgun DNA Sequencing Using Cloned DNase I-Generated Fragments.” *Nucleic Acids Research* 9 (13): 3015–27. https://doi.org/10.1093/nar/9.13.3015.

Arnaout, Ramy, William Lee, Patrick Cahill, Tracey Honan, Todd Sparrow, Michael Weiand, Chad Nusbaum, Klaus Rajewsky, and Sergei B. Koralov. 2011. “High-Resolution Description of Antibody Heavy-Chain Repertoires in Humans.” *PLoS ONE* 6 (8). https://doi.org/10.1371/journal.pone.0022365.

Avnir, Yuval, Corey T. Watson, Jacob Glanville, Eric C. Peterson, Aimee S. Tallarico, Andrew S. Bennett, Kun Qin, et al. 2016. “IGHV1-69 Polymorphism Modulates Anti-Influenza Antibody Repertoires, Correlates with IGHV Utilization Shifts and Varies by Ethnicity.” *Scientific Reports* 6 (February). https://doi.org/10.1038/srep20842.

Bachl, Jürgen, Isin Ertongur, and Berit Jungnickel. 2006. “Involvement of Rad18 in Somatic Hypermutation.” *Proceedings of the National Academy of Sciences of the United States of America* 103 (32): 12081–86. https://doi.org/10.1073/pnas.0605146103.

Barnett, Burton E., Maria L. Ciocca, Radhika Goenka, Lisa G. Barnett, Junmin Wu, Terri M. Laufer, Janis K. Burkhardt, Michael P. Cancro, and Steven L. Reiner. 2012. “Asymmetric B Cell Division in the Germinal Center Reaction.” *Science* 335 (6066): 342–44. https://doi.org/10.1126/science.1213495.

Bashford-Rogers, Rachael J.M., Anne L. Palser, Brian J. Huntly, Richard Rance, George S. Vassiliou, George A. Follows, and Paul Kellam. 2013. “Network Properties Derived from Deep Sequencing of Human B-Cell Receptor Repertoires Delineate b-Cell Populations.” *Genome Research* 23 (11): 1874–84. https://doi.org/10.1101/gr.154815.113.

Baum, Paul D., Vanessa Venturi, and David A. Price. 2012. “Wrestling with the Repertoire: The Promise and Perils of next Generation Sequencing for Antigen Receptors.” *European Journal of Immunology*. https://doi.org/10.1002/eji.201242999.

Benedict, Cindy L., Susan Gilfillan, To-Ha Thai, and John F. Kearney. 2000. “Terminal Deoxynucleotidyl Transferase and Repertoire Development.” *Immunological Reviews* 175 (1): 150–57. https://doi.org/10.1111/j.1600-065X.2000.imr017518.x.

Bernstein, David I, Rhonda D Cardin, Derek A Pullum, Fernando J Bravo, Konstantin G Kousoulas, and David A Dixon. 2019. “Duration of Protection from Live Attenuated vs. Sub Unit HSV-2 Vaccines in the Guinea Pig Model of Genital Herpes: Reassessing Efficacy Using Endpoints from Clinical Trials.” *PloS One* 14 (3): e0213401. https://doi.org/10.1371/journal.pone.0213401.

Blancher, Antoine, Alice Aarnink, Nicolas Savy, and Naoyuki Takahata. 2012. “Use of Cumulative Poisson Probability Distribution as an Estimator of the Recombination Rate in an Expanding Population: Example of the Macaca Fascicularis Major Histocompatibility Complex.” *G3: Genes, Genomes, Genetics* 2 (1): 123–30. https://doi.org/10.1534/g3.111.001248.

Bonhomme, Maxime, Antoine Blancher, Sergi Cuartero, Lounès Chikhi, and Brigitte Crouau-Roy. 2008. “Origin and Number of Founders in an Introduced Insular Primate: Estimation from Nuclear Genetic Data.” *Molecular Ecology* 17 (4): 1009–19. https://doi.org/10.1111/j.1365-294X.2007.03645.x.

Bonhomme, Maxime, Sergi Cuartero, Antoine Blancher, and Brigitte Crouau-Roy. 2009. “Assessing Natural Introgression in 2 Biomedical Model Species, the Rhesus Macaque (Macaca Mulatta) and the Long-Tailed Macaque (Macaca Fascicularis).” *Journal of Heredity* 100 (2): 158–69. https://doi.org/10.1093/jhered/esn093.

Bonsignori, Mattia, Tongqing Zhou, Zizhang Sheng, Lei Chen, Feng Gao, M Gordon Joyce, Gabriel Ozorowski, et al. 2016. “Maturation Pathway from Germline to Broad HIV-1 Neutralizer of a CD4-Mimic Antibody.” *Cell* 165 (2): 449–63. https://doi.org/10.1016/j.cell.2016.02.022.

Bourcy, Charles F.A. De, Cesar J. Lopez Angel, Christopher Vollmers, Cornelia L. Dekker, Mark M. Davis, and Stephen R. Quake. 2017. “Phylogenetic Analysis of the Human Antibody Repertoire Reveals Quantitative Signatures of Immune Senescence and Aging.” *Proceedings of the National Academy of Sciences of the United States of America* 114 (5): 1105–10. https://doi.org/10.1073/pnas.1617959114.

Boyd, Scott D., Bruno A. Gaëta, Katherine J. Jackson, Andrew Z. Fire, Eleanor L. Marshall, Jason D. Merker, Jay M. Maniar, et al. 2010. “Individual Variation in the Germline Ig Gene Repertoire Inferred from Variable Region Gene Rearrangements.” *The Journal of Immunology* 184 (12): 6986–92. https://doi.org/10.4049/jimmunol.1000445.

Boyd, Scott D., Eleanor L. Marshall, Jason D. Merker, Jay M. Maniar, Lyndon N. Zhang, Bita Sahaf, Carol D. Jones, et al. 2009. “Measurement and Clinical Monitoring of Human Lymphocyte Clonality by Massively Parallel V-D-J Pyrosequencing.” *Science Translational Medicine* 1 (12). https://doi.org/10.1126/scitranslmed.3000540.

Branton, Daniel, David W. Deamer, Andre Marziali, Hagan Bayley, Steven A. Benner, Thomas Butler, Massimiliano Di Ventra, et al. 2009. “The Potential and Challenges of Nanopore Sequencing.” In *Nanoscience and Technology: A Collection of Reviews from Nature Journals*, 261–68. World Scientific Publishing Co. https://doi.org/10.1142/9789814287005_0027.

Breden, Felix, Eline T. Luning Prak, Bjoern Peters, Florian Rubelt, Chaim A. Schramm, Christian E. Busse, Jason A. Vander Heiden, et al. 2017. “Reproducibility and Reuse of Adaptive Immune Receptor Repertoire Data.” *Frontiers in Immunology* 8 (NOV). https://doi.org/10.3389/fimmu.2017.01418.

Budde, Melisa L., Roger W. Wiseman, Julie A. Karl, Bozena Hanczaruk, Birgitte B. Simen, and David H. O’Connor. 2010. “Characterization of Mauritian Cynomolgus Macaque Major Histocompatibility Complex Class I Haplotypes by High-Resolution Pyrosequencing.” *Immunogenetics* 62 (11–12): 773–80. https://doi.org/10.1007/s00251-010-0481-9.

Bukhari, Syed Ahmad Chan, Martin J. O’Connor, Marcos Martínez-Romero, Attila L. Egyedi, Debra Willrett, John Graybeal, Mark A. Musen, Florian Rubelt, Kei Hoi Cheung, and Steven H. Kleinstein. 2018. “The CAIRR Pipeline for Submitting Standards-Compliant B and T Cell Receptor Repertoire Sequencing Studies to the National Center for Biotechnology Information Repositories.” *Frontiers in Immunology* 9 (AUG). https://doi.org/10.3389/fimmu.2018.01877.

Burton, Dennis R., and John R. Mascola. 2015. “Antibody Responses to Envelope Glycoproteins in HIV-1 Infection.” *Nature Immunology*. Nature Publishing Group. https://doi.org/10.1038/ni.3158.

Busse, Christian E., Irina Czogiel, Peter Braun, Peter F. Arndt, and Hedda Wardemann. 2014. “Single-Cell Based High-Throughput Sequencing of Full-Length Immunoglobulin Heavy and Light Chain Genes.” *European Journal of Immunology* 44 (2): 597–603. https://doi.org/10.1002/eji.201343917.

Calis, Jorg J A, and Brad R Rosenberg. 2014. “Characterizing Immune Repertoires by High Throughput Sequencing: Strategies and Applications.” *Trends in Immunology* 35 (12): 581–90. https://doi.org/10.1016/j.it.2014.09.004.

Chakrabarti, Lisa, Mireille Guyader, Marc Alizon, Muthiah D. Daniel, Ronald C. Desrosiers, Pierre Tiollais, and Pierre Soigo. 1987. “Sequence of Simian Immunodeficiency Virus from Macaque and Its Relationship to Other Human and Simian Retroviruses.” *Nature* 328 (6130): 543–47. https://doi.org/10.1038/328543a0.

Chamberlain, J S, R A Gibbs, J E Ranier, P N Nguyen, and C T Caskey. 1988. “Deletion Screening of the Duchenne Muscular Dystrophy Locus via Multiplex DNA Amplification.” *Nucleic Acids Research* 16 (23): 11141–56. https://doi.org/10.1093/nar/16.23.11141.

Chen, Lei L., Andrea M. Frank, Judy C. Adams, and Ralph M. Steinman. 1978. “Distribution of Horseradish Peroxidase (HRP)- Anti-HRP Immune Complexes in Mouse Spleen with Special Reference to Follicular Dendritic Cells.” *Journal of Cell Biology* 79 (1): 184–99.

https://doi.org/10.1083/jcb.79.1.184.
Chiang, Y L, R Sheng-Dong, M A Brow, and J W Larrick. 1989. "Direct CDNA Cloning of the Rearranged Immunoglobulin Variable Region." *BioTechniques* 7 (4): 360–66. http://www.ncbi.nlm.nih.gov/pubmed/2629848.
Cirelli, Kimberly M, Diane G Carnathan, Bartek Nogal, Jacob T Martin, Oscar L Rodriguez, Amit A Upadhyay, Chiamaka A Enemuo, et al. 2019. "Slow Delivery Immunization Enhances HIV Neutralizing Antibody and Germinal Center Responses via Modulation of Immunodominance." *Cell* 177 (5): 1153-1171.e28. https://doi.org/10.1016/j.cell.2019.04.012.
Collins, Andrew M., and Katherine J.L. Jackson. 2018. "On Being the Right Size: Antibody Repertoire Formation in the Mouse and Human." *Immunogenetics*. Springer Verlag. https://doi.org/10.1007/s00251-017-1049-8.
Collins, Andrew M., Yan Wang, Krishna M. Roskin, Christopher P. Marquis, and Katherine J.L. Jackson. 2015. "The Mouse Antibody Heavy Chain Repertoire Is Germline-Focused and Highly Variable between Inbred Strains." *Philosophical Transactions of the Royal Society B: Biological Sciences* 370 (1676). https://doi.org/10.1098/rstb.2014.0236.
Collins, Andrew M, Yan Wang, Viveka Singh, Phillip Yu, Katherine J Jackson, and William A Sewell. 2008. "The Reported Germline Repertoire of Human Immunoglobulin Kappa Chain Genes Is Relatively Complete and Accurate." *Immunogenetics* 60 (11): 669–76. https://doi.org/10.1007/s00251-008-0325-z.
Cook, G. P., I. M. Tomlinson, G. Walter, H. Riethman, N. P. Carter, L. Buluwela, G. Winter, and T. H. Rabbitts. 1994. "A Map of the Human Immunoglobulin V(H) Locus Completed by Analysis of the Telomeric Region of Chromosome 14q." *Nature Genetics* 7 (2): 162–68. https://doi.org/10.1038/ng0694-162.
Corcoran, Martin M., Ganesh E. Phad, Néstor Vázquez Bernat, Christiane Stahl-Hennig, Noriyuki Sumida, Mats A.A. Persson, Marcel Martin, and Gunilla B.Karlsson Karlsson Hedestam. 2016. "Production of Individualized V Gene Databases Reveals High Levels of Immunoglobulin Genetic Diversity." *Nature Communications* 7 (December): 13642. https://doi.org/10.1038/ncomms13642.
Coutinho, Antonio, and Göran Möller. 1973. "B Cell Mitogenic Properties of Thymus-Independent Antigens." *Nature New Biology* 245 (140): 12–14. https://doi.org/10.1038/newbio245012a0.
Cox, Kara S, Aimin Tang, Zhifeng Chen, Melanie S Horton, Hao Yan, Xin-Min Wang, Sheri A Dubey, et al. 2016. "Rapid Isolation of Dengue-Neutralizing Antibodies from Single Cell-Sorted Human Antigen-Specific Memory B-Cell Cultures." *MAbs* 8 (1): 129–40. https://doi.org/10.1080/19420862.2015.1109757.
Davis, Mark M., and Scott D. Boyd. 2019. "Recent Progress in the Analysis of AβT Cell and B Cell Receptor Repertoires." *Current Opinion in Immunology* 59 (August): 109–14. https://doi.org/10.1016/j.coi.2019.05.012.
Dekosky, Brandon J., Gregory C. Ippolito, Ryan P. Deschner, Jason J. Lavinder, Yariv Wine, Brandon M. Rawlings, Navin Varadarajan, et al. 2013. "High-Throughput Sequencing of the Paired Human Immunoglobulin Heavy and Light Chain Repertoire." *Nature Biotechnology* 31 (2): 166–69. https://doi.org/10.1038/nbt.2492.
DeKosky, Brandon J., Takaaki Kojima, Alexa Rodin, Wissam Charab, Gregory C. Ippolito, Andrew D. Ellington, and George Georgiou. 2015. "In-Depth Determination and Analysis of the Human Paired Heavy- and Light-Chain Antibody Repertoire." *Nature Medicine* 21 (1): 86–91. https://doi.org/10.1038/nm.3743.
DeKosky, Brandon J., Oana I. Lungu, Daechan Park, Erik L. Johnson, Wissam Charab, Constantine Chrysostomou, Daisuke Kuroda, et al. 2016. "Large-Scale Sequence and Structural Comparisons of Human Naive and Antigen-Experienced Antibody Repertoires." *Proceedings of the National Academy of Sciences of the United States of America* 113 (19): E2636–45. https://doi.org/10.1073/pnas.1525510113.
Dijk-Härd, Iris Van, and Inger Lundkvist. 2002. "Long-Term Kinetics of Adult Human Antibody Repertoires." *Immunology* 107 (1): 136–44. https://doi.org/10.1046/j.1365-2567.2002.01466.x.
Dijk, K W van, E H Sasso, and E C Milner. 1991. "Polymorphism of the Human Ig VH4 Gene Family." *Journal of Immunology (Baltimore, Md. : 1950)* 146 (10): 3646–51. http://www.ncbi.nlm.nih.gov/pubmed/1673987.
Duan, Linna, and Eric Mukherjee. 2016. "Focus: Microbiome: Janeway's Immunobiology, Ninth Edition." *The Yale Journal of Biology and Medicine* 89 (3): 424.

Dubrovskaya, Viktoriya, Karen Tran, Gabriel Ozorowski, Javier Guenaga, Richard Wilson, Shridhar Bale, Christopher A. Cottrell, et al. 2019. “Vaccination with Glycan-Modified HIV NFL Envelope Trimer-Liposomes Elicits Broadly Neutralizing Antibodies to Multiple Sites of Vulnerability.” *Immunity*, November. https://doi.org/10.1016/j.immuni.2019.10.008.
Dudley, Darryll D., Jayanta Chaudhuri, Craig H. Bassing, and Frederick W. Alt. 2005. “Mechanism and Control of V(D)J Recombination versus Class Switch Recombination: Similarities and Differences.” *Advances in Immunology* 86: 43–112. https://doi.org/10.1016/S0065-2776(04)86002-4.
Duffy, Ken R, Cameron J Wellard, John F Markham, Jie H S Zhou, Ross Holmberg, Edwin D Hawkins, Jhagvaral Hasbold, Mark R Dowling, and Philip D Hodgkin. 2012. “Activation-Induced B Cell Fates Are Selected by Intracellular Stochastic Competition.” *Science (New York, N.Y.)* 335 (6066): 338–41. https://doi.org/10.1126/science.1213230.
Durandy, Anne. 2003. “Activation-Induced Cytidine Deaminase: A Dual Role in Class-Switch Recombination and Somatic Hypermutation.” *European Journal of Immunology*. https://doi.org/10.1002/eji.200324133.
Ebeling, Martin, Erich Küng, Angela See, Clemens Broger, Guido Steiner, Marco Berrera, Tobias Heckel, et al. 2011. “Genome-Based Analysis of the Nonhuman Primate Macaca Fascicularis as a Model for Drug Safety Assessment.” *Genome Research* 21 (10): 1746–56. https://doi.org/10.1101/gr.123117.111.
Estes, Jacob D, Scott W Wong, and Jason M Brenchley. 2018. “Nonhuman Primate Models of Human Viral Infections.” *Nature Reviews. Immunology* 18 (6): 390–404. https://doi.org/10.1038/s41577-018-0005-7.
Evseenko, Vasily A, Natalia P Kolosova, Anderi S Gudymo, Semion V Maltsev, Julia A Bulanovich, Nataliya I Goncharova, Natalia V Danilchenko, Vasiliy Y Marchenko, Ivan M Susloparov, and Alexander B Ryzhikov. 2019. “Intranasal Immunization of Guinea Pig with Trivalent Influenza Antigen Adjuvanted by Cyclamen Europaeum Tubers Extract.” *Archives of Virology* 164 (1): 243–47. https://doi.org/10.1007/s00705-018-4023-3.
Fagarasan, S., and T. Honjo. 2000. “T-Independent Immune Response: New Aspects of B Cell Biology.” *Science*. https://doi.org/10.1126/science.290.5489.89.
Fawcett, Gloria L., Muthuswamy Raveendran, David R. Deiros, David Chen, Fuli Yu, Ronald A. Harris, Yanru Ren, et al. 2011. “Characterization of Single-Nucleotide Variation in Indian-Origin Rhesus Macaques (Macaca Mulatta).” *BMC Genomics* 12 (June). https://doi.org/10.1186/1471-2164-12-311.
Fedurco, Milan, Anthony Romieu, Scott Williams, Isabelle Lawrence, and Gerardo Turcatti. 2006. “BTA, a Novel Reagent for DNA Attachment on Glass and Efficient Generation of Solid-Phase Amplified DNA Colonies.” *Nucleic Acids Research* 34 (3). https://doi.org/10.1093/nar/gnj023.
Foege, W H, J D Millar, and D A Henderson. 1975. “Smallpox Eradication in West and Central Africa.” *Bulletin of the World Health Organization* 52 (2): 209–22. http://www.ncbi.nlm.nih.gov/pubmed/1083309.
Forsman, Anna, Els Beirnaert, Marlén M I Aasa-Chapman, Bart Hoorelbeke, Karolin Hijazi, Willie Koh, Vanessa Tack, et al. 2008. “Llama Antibody Fragments with Cross-Subtype Human Immunodeficiency Virus Type 1 (HIV-1)-Neutralizing Properties and High Affinity for HIV-1 Gp120.” *Journal of Virology* 82 (24): 12069–81. https://doi.org/10.1128/JVI.01379-08.
Francica, Joseph R, Zizhang Sheng, Zhenhai Zhang, Yoshiaki Nishimura, Masashi Shingai, Akshaya Ramesh, Brandon F Keele, et al. 2015. “Analysis of Immunoglobulin Transcripts and Hypermutation Following SHIV(AD8) Infection and Protein-plus-Adjuvant Immunization.” *Nature Communications* 6 (April): 6565. https://doi.org/10.1038/ncomms7565.
Freeman, J. Douglas, René L. Warren, John R. Webb, Brad H. Nelson, and Robert A. Holt. 2009. “Profiling the T-Cell Receptor Beta-Chain Repertoire by Massively Parallel Sequencing.” *Genome Research* 19 (10): 1817–24. https://doi.org/10.1101/gr.092924.109.
Frohman, M. A., M. K. Dush, and G. R. Martin. 1988. “Rapid Production of Full-Length CDNAs from Rare Transcripts: Amplification Using a Single Gene-Specific Oligonucleotide Primer.” *Proceedings of the National Academy of Sciences of the United States of America* 85 (23): 8998–9002. https://doi.org/10.1073/pnas.85.23.8998.
Gadala-Maria, Daniel, Gur Yaari, Mohamed Uduman, and Steven H. Kleinstein. 2015. “Automated Analysis of High-Throughput B-Cell Sequencing Data Reveals a High Frequency of Novel Immunoglobulin V Gene Segment Alleles.” *Proceedings of the National Academy of Sciences of*

the United States of America 112 (8): E862–70. https://doi.org/10.1073/pnas.1417683112.
Galson, Jacob D., Johannes Trück, Anna Fowler, Elizabeth A. Clutterbuck, Márton Münz, Vincenzo Cerundolo, Claudia Reinhard, et al. 2015. "Analysis of B Cell Repertoire Dynamics Following Hepatitis B Vaccination in Humans, and Enrichment of Vaccine-Specific Antibody Sequences." *EBioMedicine* 2 (12): 2070–79. https://doi.org/10.1016/j.ebiom.2015.11.034.
Galson, Jacob D., Johannes Trück, Anna Fowler, Márton Münz, Vincenzo Cerundolo, Andrew J. Pollard, Gerton Lunter, and Dominic F. Kelly. 2015. "In-Depth Assessment of within-Individual and Inter-Individual Variation in the B Cell Receptor Repertoire." *Frontiers in Immunology* 6 (OCT). https://doi.org/10.3389/fimmu.2015.00531.
Garces, Fernando, Jeong Hyun Lee, Natalia de Val, Alba Torrents de la Pena, Leopold Kong, Cristina Puchades, Yuanzi Hua, et al. 2015. "Affinity Maturation of a Potent Family of HIV Antibodies Is Primarily Focused on Accommodating or Avoiding Glycans." *Immunity* 43 (6): 1053–63. https://doi.org/10.1016/j.immuni.2015.11.007.
Gardner, Richard C., Alan J. Howarth, Peter Hahn, Marianne Brown-Luedi, Robert J. Shepherd, and Joachim Messing. 1981. "The Complete Nucleotide Sequence of an Infectious Clone of Cauliflower Mosaic Virus by M13mp7 Shotgun Sequencing." *Nucleic Acids Research* 9 (12): 2871–88. https://doi.org/10.1093/nar/9.12.2871.
Gauss, G H, and M R Lieber. 1996. "Mechanistic Constraints on Diversity in Human V(D)J Recombination." *Molecular and Cellular Biology* 16 (1): 258–69. https://doi.org/10.1128/mcb.16.1.258.
Gay, Denise, Thomas Saunders, Sally Camper, and Martin Weigert. 1993. "Receptor Editing: An Approach by Autoreactive B Cells to Escape Tolerance." *Journal of Experimental Medicine* 177 (4): 999–1008. https://doi.org/10.1084/jem.177.4.999.
Geisbert, Thomas W., James E. Strong, and Heinz Feldmann. 2015. "Considerations in the Use of Nonhuman Primate Models of Ebola Virus and Marburg Virus Infection." *Journal of Infectious Diseases*. Oxford University Press. https://doi.org/10.1093/infdis/jiv284.
Gent, D. C. Van, and M. Van Der Burg. 2007. "Non-Homologous End-Joining, a Sticky Affair." *Oncogene*. https://doi.org/10.1038/sj.onc.1210871.
Georgiou, George, Gregory C. Ippolito, John Beausang, Christian E. Busse, Hedda Wardemann, and Stephen R. Quake. 2014. "The Promise and Challenge of High-Throughput Sequencing of the Antibody Repertoire." *Nature Biotechnology*. https://doi.org/10.1038/nbt.2782.
Gibbs, Richard A., Jeffrey Rogers, Michael G. Katze, Roger Bumgarner, George M. Weinstock, Elaine R. Mardis, Karin A. Remington, et al. 2007. "Evolutionary and Biomedical Insights from the Rhesus Macaque Genome." *Science*. https://doi.org/10.1126/science.1139247.
Gidoni, Moriah, Omri Snir, Ayelet Peres, Pazit Polak, Ida Lindeman, Ivana Mikocziova, Vikas Kumar Sarna, et al. 2019. "Mosaic Deletion Patterns of the Human Antibody Heavy Chain Gene Locus Shown by Bayesian Haplotyping." *Nature Communications* 10 (1). https://doi.org/10.1038/s41467-019-08489-3.
Giudicelli, Véronique, Denys Chaume, and Marie Paule Lefranc. 2005. "IMGT/GENE-DB: A Comprehensive Database for Human and Mouse Immunoglobulin and T Cell Receptor Genes." *Nucleic Acids Research* 33 (DATABASE ISS.). https://doi.org/10.1093/nar/gki010.
Glanville, Jacob, Tracy C. Kuo, H. Christian Von Büdingen, Lin Guey, Jan Berka, Purnima D. Sundar, Gabriella Huerta, et al. 2011. "Naive Antibody Gene-Segment Frequencies Are Heritable and Unaltered by Chronic Lymphocyte Ablation." *Proceedings of the National Academy of Sciences of the United States of America* 108 (50): 20066–71. https://doi.org/10.1073/pnas.1107498108.
Glanville, Jacob, Wenwu Zhai, Jan Berka, Dilduz Telman, Gabriella Huerta, Gautam R Mehta, Irene Ni, et al. 2009. "Precise Determination of the Diversity of a Combinatorial Antibody Library Gives Insight into the Human Immunoglobulin Repertoire." *Proceedings of the National Academy of Sciences of the United States of America* 106 (48): 20216–21. https://doi.org/10.1073/pnas.0909775106.
Gray, E. S., M. C. Madiga, T. Hermanus, P. L. Moore, C. K. Wibmer, N. L. Tumba, L. Werner, et al. 2011. "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." *Journal of Virology* 85 (10): 4828–40. https://doi.org/10.1128/jvi.00198-11.
Greiff, Victor, Ulrike Menzel, Enkelejda Miho, Cédric Weber, René Riedel, Skylar Cook, Atijeh Valai, et al. 2017. "Systems Analysis Reveals High Genetic and Antigen-Driven

Predetermination of Antibody Repertoires throughout B Cell Development." *Cell Reports* 19 (7): 1467–78. https://doi.org/10.1016/j.celrep.2017.04.054.
Haaren, Marlies M. van, Tom L.G.M. van den Kerkhof, and Marit J. van Gils. 2017. "Natural Infection as a Blueprint for Rational HIV Vaccine Design." *Human Vaccines and Immunotherapeutics*. Taylor and Francis Inc. https://doi.org/10.1080/21645515.2016.1232785.
Haus, Tanja, Betsy Ferguson, Jeffrey Rogers, Gaby Doxiadis, Ulrich Certa, Nicola J. Rose, Robert Teepe, Gerhard F. Weinbauer, and Christian Roos. 2014. "Genome Typing of Nonhuman Primate Models: Implications for Biomedical Research." *Trends in Genetics*. Elsevier Ltd. https://doi.org/10.1016/j.tig.2014.05.004.
He, Linling, Devin Sok, Parisa Azadnia, Jessica Hsueh, Elise Landais, Melissa Simek, Wayne C. Koff, Pascal Poignard, Dennis R. Burton, and Jiang Zhu. 2014. "Toward a More Accurate View of Human B-Cell Repertoire by next-Generation Sequencing, Unbiased Repertoire Capture and Single-Molecule Barcoding." *Scientific Reports* 4 (October). https://doi.org/10.1038/srep06778.
Henry, Carole, Anna-Karin E. Palm, Henry A. Utset, Min Huang, Irvin Y. Ho, Nai-Ying Zheng, Theresa Fitzgerald, et al. 2019. "Monoclonal Antibody Responses after Recombinant HA Vaccine versus Subunit Inactivated Influenza Virus Vaccine: A Comparative Study." *Journal of Virology*, August. https://doi.org/10.1128/jvi.01150-19.
Higashino, Atsunori, Ryuichi Sakate, Yosuke Kameoka, Ichiro Takahashi, Makoto Hirata, Reiko Tanuma, Tohru Masui, Yasuhiro Yasutomi, and Naoki Osada. 2012. "Whole-Genome Sequencing and Analysis of the Malaysian Cynomolgus Macaque (Macaca Fascicularis) Genome." *Genome Biology* 13 (7). https://doi.org/10.1186/gb-2012-13-7-r58.
Hoehn, Kenneth B., Anna Fowler, Gerton Lunter, and Oliver G. Pybus. 2016. "The Diversity and Molecular Evolution of B-Cell Receptors during Infection." *Molecular Biology and Evolution* 33 (5): 1147–57. https://doi.org/10.1093/molbev/msw015.
Hoehn, Kenneth B., Jason A. Vander Heiden, Julian Q. Zhou, Gerton Lunter, Oliver G. Pybus, and Steven H. Kleinstein. 2019. "Repertoire-Wide Phylogenetic Models of B Cell Molecular Evolution Reveal Evolutionary Signatures of Aging and Vaccination." *Proceedings of the National Academy of Sciences*, October, 201906020. https://doi.org/10.1073/pnas.1906020116.
Howie, Bryan, Anna M. Sherwood, Ashley D. Berkebile, Jan Berka, Ryan O. Emerson, David W. Williamson, Ilan Kirsch, et al. 2015. "High-Throughput Pairing of T Cell Receptor α and β Sequences." *Science Translational Medicine* 7 (301). https://doi.org/10.1126/scitranslmed.aac5624.
Huang, Jinghe, Byong H Kang, Elise Ishida, Tongqing Zhou, Trevor Griesman, Zizhang Sheng, Fan Wu, et al. 2016. "Identification of a CD4-Binding-Site Antibody to HIV That Evolved Near-Pan Neutralization Breadth." *Immunity* 45 (5): 1108–21. https://doi.org/10.1016/j.immuni.2016.10.027.
Huang, Lin, Miles D Lange, and Zhixin Zhang. 2014. "VH Replacement Footprint Analyzer-I, a Java-Based Computer Program for Analyses of Immunoglobulin Heavy Chain Genes and Potential VH Replacement Products in Human and Mouse." *Frontiers in Immunology* 5 (February): 40. https://doi.org/10.3389/fimmu.2014.00040.
Jackson, Katherine J. L., Marie J. Kidd, Yan Wang, and Andrew M. Collins. 2013. "The Shape of the Lymphocyte Receptor Repertoire: Lessons from the B Cell Receptor." *Frontiers in Immunology*. https://doi.org/10.3389/fimmu.2013.00263.
Jackson, Katherine J.L., Yi Liu, Krishna M. Roskin, Jacob Glanville, Ramona A. Hoh, Katie Seo, Eleanor L. Marshall, et al. 2014. "Human Responses to Influenza Vaccination Show Seroconversion Signatures and Convergent Antibody Rearrangements." *Cell Host and Microbe* 16 (1): 105–14. https://doi.org/10.1016/j.chom.2014.05.013.
Jacobs, D M, and D C Morrison. 1975. "Stimulation of a T-Independent Primary Anti-Hapten Response in Vitro by TNP-Lipopolysaccharide (TNP-LPS)." *Journal of Immunology (Baltimore, Md. : 1950)* 114 (1 Pt 2): 360–64. http://www.ncbi.nlm.nih.gov/pubmed/1090653.
Jiang, Ning, Jiankui He, Joshua A. Weinstein, Lolita Penland, Sanae Sasaki, Xiao Song He, Cornelia L. Dekker, et al. 2013. "Lineage Structure of the Human Antibody Repertoire in Response to Influenza Vaccination." *Science Translational Medicine* 5 (171). https://doi.org/10.1126/scitranslmed.3004794.
Jonge, JØrgen De, Irina Isakova-Sivak, Harry Van Dijken, Sanne Spijkers, Justin Mouthaan, Rineke De Jong, Tatiana Smolonogina, Paul Roholl, and Larisa Rudenko. 2016. "H7N9 Live Attenuated Influenza Vaccine Is Highly Immunogenic, Prevents Virus Replication, and Protects against

Severe Bronchopneumonia in Ferrets." *Molecular Therapy* 24 (5): 991–1002. https://doi.org/10.1038/mt.2016.23.
Jung, David, and Frederick W Alt. 2004. "Unraveling V(D)J Recombination; Insights into Gene Regulation." *Cell* 116 (2): 299–311. https://doi.org/10.1016/s0092-8674(04)00039-x.
Kallewaard, Nicole L, Davide Corti, Patrick J Collins, Ursula Neu, Josephine M McAuliffe, Ebony Benjamin, Leslie Wachter-Rosati, et al. 2016. "Structure and Function Analysis of an Antibody Recognizing All Influenza A Subtypes." *Cell* 166 (3): 596–608. https://doi.org/10.1016/j.cell.2016.05.073.
Karl, Julie A., Michael E. Graham, Roger W. Wiseman, Katelyn E. Heimbruch, Samantha M. Gieger, Gaby G.M. Doxiadis, Ronald E. Bontrop, and David H. O'Connor. 2017. "Major Histocompatibility Complex Haplotyping and Long-Amplicon Allele Discovery in Cynomolgus Macaques from Chinese Breeding Facilities." *Immunogenetics* 69 (4): 211–29. https://doi.org/10.1007/s00251-017-0969-7.
Kennedy, E M, S D Dowall, F J Salguero, P Yeates, M Aram, and R Hewson. 2019. "A Vaccine Based on Recombinant Modified Vaccinia Ankara Containing the Nucleoprotein from Lassa Virus Protects against Disease Progression in a Guinea Pig Model." *Vaccine* 37 (36): 5404–13. https://doi.org/10.1016/j.vaccine.2019.07.023.
Kent, W James. 2002. "BLAT-The BLAST-Like Alignment Tool Resource 656 Genome Research," 656–64. https://doi.org/10.1101/gr.229202.
Kidd, Marie J., Zhiliang Chen, Yan Wang, Katherine J. Jackson, Lyndon Zhang, Scott D. Boyd, Andrew Z. Fire, Mark M. Tanaka, Bruno A. Gaëta, and Andrew M. Collins. 2012. "The Inference of Phased Haplotypes for the Immunoglobulin H Chain V Region Gene Loci by Analysis of VDJ Gene Rearrangements." *The Journal of Immunology* 188 (3): 1333–40. https://doi.org/10.4049/jimmunol.1102097.
Kirik, Ufuk, Lennart Greiff, Fredrik Levander, and Mats Ohlin. 2017. "Parallel Antibody Germline Gene and Haplotype Analyses Support the Validity of Immunoglobulin Germline Gene Inference and Discovery." *Molecular Immunology* 87 (July): 12–22. https://doi.org/10.1016/j.molimm.2017.03.012.
Kivioja, Teemu, Anna Vähärautio, Kasper Karlsson, Martin Bonke, Martin Enge, Sten Linnarsson, and Jussi Taipale. 2012. "Counting Absolute Numbers of Molecules Using Unique Molecular Identifiers." *Nature Methods* 9 (1): 72–74. https://doi.org/10.1038/nmeth.1778.
Klaus, G G, and J H Humphrey. 1977. "The Generation of Memory Cells. I. The Role of C3 in the Generation of B Memory Cells." *Immunology* 33 (1): 31–40. http://www.ncbi.nlm.nih.gov/pubmed/301502.
Klausen, Michael Schantz, Mads Valdemar Anderson, Martin Closter Jespersen, Morten Nielsen, and Paolo Marcatili. 2015. "LYRA, a Webserver for Lymphocyte Receptor Structural Modeling." *Nucleic Acids Research* 43 (W1): W349–55. https://doi.org/10.1093/nar/gkv535.
Klein, Florian, Ron Diskin, Johannes F Scheid, Christian Gaebler, Hugo Mouquet, Ivelin S Georgiev, Marie Pancera, et al. 2013. "Somatic Mutations of the Immunoglobulin Framework Are Generally Required for Broad and Potent HIV-1 Neutralization." *Cell* 153 (1): 126–38. https://doi.org/10.1016/j.cell.2013.03.018.
Kong, Rui, Hongying Duan, Zizhang Sheng, Kai Xu, Priyamvada Acharya, Xuejun Chen, Cheng Cheng, et al. 2019. "Antibody Lineages with Vaccine-Induced Antigen-Binding Hotspots Develop Broad HIV Neutralization." *Cell* 178 (3): 567-584.e19. https://doi.org/10.1016/j.cell.2019.06.030.
Korlach, Jonas, Keith P. Bjornson, Bidhan P. Chaudhuri, Ronald L. Cicero, Benjamin A. Flusberg, Jeremy J. Gray, David Holden, Ravi Saxena, Jeffrey Wegener, and Stephen W. Turner. 2010. "Real-Time DNA Sequencing from Single Polymerase Molecules." *Methods in Enzymology* 472: 431–55. https://doi.org/10.1126/science.1162986.
Kovaltsuk, Aleksandr, Konrad Krawczyk, Jacob D. Galson, Dominic F. Kelly, Charlotte M. Deane, and Johannes Trück. 2017. "How B-Cell Receptor Repertoire Sequencing Can Be Enriched with Structural Antibody Data." *Frontiers in Immunology*. Frontiers Media S.A. https://doi.org/10.3389/fimmu.2017.01753.
Kranich, Jan, and Nike Julia Krautler. 2016. "How Follicular Dendritic Cells Shape the B-Cell Antigenome." *Frontiers in Immunology*. Frontiers Research Foundation. https://doi.org/10.3389/fimmu.2016.00225.
Krebs, Kendall C., ZheYuan Jin, Richard Rudersdorf, Austin L. Hughes, and David H. O'Connor.

2005. “Unusually High Frequency MHC Class I Alleles in Mauritian Origin Cynomolgus Macaques.” *The Journal of Immunology* 175 (8): 5230–39. https://doi.org/10.4049/jimmunol.175.8.5230.

Kwong, Peter D., and John R. Mascola. 2018. “HIV-1 Vaccines Based on Antibody Identification, B Cell Ontogeny, and Epitope Structure.” *Immunity* 48 (5): 855–71. https://doi.org/10.1016/j.immuni.2018.04.029.

Kwong, Peter D., John R. Mascola, and Gary J. Nabel. 2013. “Broadly Neutralizing Antibodies and the Search for an HIV-1 Vaccine: The End of the Beginning.” *Nature Reviews Immunology*. https://doi.org/10.1038/nri3516.

Lander, Eric S., Lauren M. Linton, Bruce Birren, Chad Nusbaum, Michael C. Zody, Jennifer Baldwin, Keri Devon, et al. 2001. “Initial Sequencing and Analysis of the Human Genome.” *Nature* 409 (6822): 860–921. https://doi.org/10.1038/35057062.

Lapidoth, Gideon, Jake Parker, Jaime Prilusky, and Sarel J. Fleishman. 2019. “AbPredict 2: A Server for Accurate and Unstrained Structure Prediction of Antibody Variable Domains.” *Bioinformatics* 35 (9): 1591–93. https://doi.org/10.1093/bioinformatics/bty822.

Larrick, J W, L Danielsson, C A Brenner, M Abrahamson, K E Fry, and C A Borrebaeck. 1989. “Rapid Cloning of Rearranged Immunoglobulin Genes from Human Hybridoma Cells Using Mixed Primers and the Polymerase Chain Reaction.” *Biochemical and Biophysical Research Communications* 160 (3): 1250–56. https://doi.org/10.1016/s0006-291x(89)80138-x.

Larson, Erik D., and Nancy Maizels. 2004. “Transcription-Couple Mutagenesis by the DNA Deaminase AID.” *Genome Biology*. https://doi.org/10.1186/gb-2004-5-3-211.

Lee, C. E. H., B. Gaëta, H. R. Malming, M. E. Bain, W. A. Sewell, and A. M. Collins. 2006. “Reconsidering the Human Immunoglobulin Heavy-Chain Locus: 1. An Evaluation of the Expressed Human IGHD Gene Repertoire.” *Immunogenetics* 57 (12): 917–25. https://doi.org/10.1007/s00251-005-0062-5.

Lee, Jiwon, Daniel R Boutz, Veronika Chromikova, M Gordon Joyce, Christopher Vollmers, Kwanyee Leung, Andrew P Horton, et al. 2016. “Molecular-Level Analysis of the Serum Antibody Repertoire in Young Adults before and after Seasonal Influenza Vaccination.” *Nature Medicine* 22 (12): 1456–64. https://doi.org/10.1038/nm.4224.

Leem, Jinwoo, James Dunbar, Guy Georges, Jiye Shi, and Charlotte M. Deane. 2016. “ABodyBuilder: Automated Antibody Structure Prediction with Data–Driven Accuracy Estimation.” *MAbs* 8 (7): 1259–68. https://doi.org/10.1080/19420862.2016.1205773.

Lees, William, Christian E Busse, Martin Corcoran, Mats Ohlin, Cathrine Scheepers, Frederick A Matsen, Gur Yaari, Corey T Watson, Andrew Collins, and Adrian J Shepherd. 2019. “OGRDB: A Reference Database of Inferred Immune Receptor Genes.” *Nucleic Acids Research*, September. https://doi.org/10.1093/nar/gkz822.

Lefranc, Marie Paule, Véronique Giudicelli, Patrice Duroux, Joumana Jabado-Michaloud, Géraldine Folch, Safa Aouinti, Emilie Carillon, et al. 2015. “IMGT R, the International ImMunoGeneTics Information System R 25 Years On.” *Nucleic Acids Research* 43 (D1): D413–22. https://doi.org/10.1093/nar/gku1056.

Lepore, Rosalba, Pier P. Olimpieri, Mario A. Messih, and Anna Tramontano. 2017. “PIGSPro: Prediction of ImmunoGlobulin Structures V2.” *Nucleic Acids Research* 45 (W1): W17–23. https://doi.org/10.1093/nar/gkx334.

Leuchte, N., N. Berry, B. Köhler, N. Almond, R. LeGrand, R. Thorstensson, F. Titti, and U. Sauermann. 2004. “MhcDRB-Sequences from Cynomolgus Macaques (Macaca Fascicularis) of Different Origin.” *Tissue Antigens* 63 (6): 529–37. https://doi.org/10.1111/j.0001-2815.2004.0222.x.

Li, John, Carol I. Lord, William Haseltine, Norman L. Letvin, and Joseph Sodroski. 1992. “Infection of Cynomolgus Monkeys with a Chimeric Hiv-l/Siv<inf>mac</Inf> Virus That Expresses the HIV-1 Envelope Glycoproteins.” *Journal of Acquired Immune Deficiency Syndromes* 5 (7): 639–46.

Li, Yang, Jaclyn L. Myers, David L. Bostick, Colleen B. Sullivan, Jonathan Madara, Susanne L. Linderman, Qin Liu, et al. 2013. “Immune History Shapes Specificity of Pandemic H1N1 Influenza Antibody Responses.” *Journal of Experimental Medicine* 210 (8): 1493–1500. https://doi.org/10.1084/jem.20130212.

Lindeman, Ida, Guy Emerton, Lira Mamanova, Omri Snir, Krzysztof Polanski, Shuo Wang Qiao, Ludvig M. Sollid, Sarah A. Teichmann, and Michael J.T. Stubbington. 2018. “BraCeR: B-Cell-

Receptor Reconstruction and Clonality Inference from Single-Cell RNA-Seq." *Nature Methods*. Nature Publishing Group. https://doi.org/10.1038/s41592-018-0082-3.
Lu, Daniel R., Yann Chong Tan, Sarah Kongpachith, Xiaoyong Cai, Emily A. Stein, Tamsin M. Lindstrom, Jeremy Sokolove, and William H. Robinson. 2014. "Identifying Functional Anti-Staphylococcus Aureus Antibodies by Sequencing Antibody Repertoires of Patient Plasmablasts." *Clinical Immunology* 152 (1–2): 77–89. https://doi.org/10.1016/j.clim.2014.02.010.
Luckey, Chance John, Deepta Bhattacharya, Ananda W Goldrath, Irving L Weissman, Christophe Benoist, and Diane Mathis. 2006. "Memory T and Memory B Cells Share a Transcriptional Program of Self-Renewal with Long-Term Hematopoietic Stem Cells." *Proceedings of the National Academy of Sciences of the United States of America* 103 (9): 3304–9. https://doi.org/10.1073/pnas.0511137103.
Luo, Shishi, Jane A. Yu, and Yun S. Song. 2016. "Estimating Copy Number and Allelic Variation at the Immunoglobulin Heavy Chain Locus Using Short Reads." *PLoS Computational Biology* 12 (9). https://doi.org/10.1371/journal.pcbi.1005117.
MacLennan, Ian C. M. 1994. "Germinal Centers." *Annual Review of Immunology* 12 (1): 117–39. https://doi.org/10.1146/annurev.iy.12.040194.001001.
Mahmoud, Medhat, Marek Zywicki, Tomasz Twardowski, and Wojciech M. Karlowski. 2019. "Efficiency of PacBio Long Read Correction by 2nd Generation Illumina Sequencing." *Genomics* 111 (1): 43–49. https://doi.org/10.1016/j.ygeno.2017.12.011.
Margulies, Marcel, Michael Egholm, William E. Altman, Said Attiya, Joel S. Bader, Lisa A. Bemben, Jan Berka, et al. 2005. "Genome Sequencing in Microfabricated High-Density Picolitre Reactors." *Nature* 437 (7057): 376–80. https://doi.org/10.1038/nature03959.
Martina, Byron E. E., Bart L. Haagmans, Thijs Kuiken, Ron A. M. Fouchier, Guus F. Rimmelzwaan, Geert van Amerongen, J. S. Malik Peiris, Wilina Lim, and Albert D. M. E. Osterhaus. 2003. "SARS Virus Infection of Cats and Ferrets." *Nature* 425 (6961): 915–915. https://doi.org/10.1038/425915a.
Martinez-Murillo, Paola, Karen Tran, Javier Guenaga, Gustaf Lindgren, Monika Àdori, Yu Feng, Ganesh E Phad, et al. 2017. "Particulate Array of Well-Ordered HIV Clade C Env Trimers Elicits Neutralizing Antibodies That Display a Unique V2 Cap Approach." *Immunity* 46 (5): 804-817.e7. https://doi.org/10.1016/j.immuni.2017.04.021.
Mascola, John R., and David C. Montefiori. 2010. "The Role of Antibodies in HIV Vaccines." *Annual Review of Immunology* 28 (1): 413–44. https://doi.org/10.1146/annurev-immunol-030409-101256.
Matsuda, Fumihiko, Kazuo Ishii, Patrice Bourvagnet, Kei Ichi Kuma, Hidenori Hayashida, Takashi Miyata, and Tasuku Honjo. 1998. "The Complete Nucleotide Sequence of the Human Immunoglobulin Heavy Chain Variable Region Locus." *Journal of Experimental Medicine* 188 (11): 2151–62. https://doi.org/10.1084/jem.188.11.2151.
Maxam, A. M., and W. Gilbert. 1977. "A New Method for Sequencing DNA." *Proceedings of the National Academy of Sciences of the United States of America* 74 (2): 560–64. https://doi.org/10.1073/pnas.74.2.560.
McCoy, Laura E., Marit J. van Gils, Gabriel Ozorowski, Terrence Messmer, Bryan Briney, James E. Voss, Daniel W. Kulp, et al. 2016. "Holes in the Glycan Shield of the Native HIV Envelope Are a Target of Trimer-Elicited Neutralizing Antibodies." *Cell Reports* 16 (9): 2327–38. https://doi.org/10.1016/j.celrep.2016.07.074.
McCoy, Laura E., Anna Forsman Quigley, Nika M. Strokappe, Bianca Bulmer-Thomas, Michael S. Seaman, Daniella Mortier, Lucy Rutten, et al. 2012. "Potent and Broad Neutralization of HIV-1 by a Llama Antibody Elicited by Immunization." *The Journal of Experimental Medicine* 209 (6): 1091–1103. https://doi.org/10.1084/jem.20112655.
Mcdaniel, Jonathan R., Brandon J. DeKosky, Hidetaka Tanno, Andrew D. Ellington, and George Georgiou. 2016. "Ultra-High-Throughput Sequencing of the Immune Receptor Repertoire from Millions of Lymphocytes." *Nature Protocols* 11 (3): 429–42. https://doi.org/10.1038/nprot.2016.024.
Menzel, Ulrike, Victor Greiff, Tarik A. Khan, Ulrike Haessler, Ina Hellmann, Simon Friedensohn, Skylar C. Cook, Mark Pogson, and Sai T. Reddy. 2014. "Comprehensive Evaluation and Optimization of Amplicon Library Preparation Methods for High-Throughput Antibody Sequencing." *PLoS ONE* 9 (5). https://doi.org/10.1371/journal.pone.0096727.

Miho, Enkelejda, Alexander Yermanos, Cédric R. Weber, Christoph T. Berger, Sai T. Reddy, and Victor Greiff. 2018. “Computational Strategies for Dissecting the High-Dimensional Complexity of Adaptive Immune Repertoires.” *Frontiers in Immunology* 9 (FEB). https://doi.org/10.3389/fimmu.2018.00224.
Milstein, César, and Michael S. Neuberger. 1996. “Maturation of the Immune Response.” In , 451–85. https://doi.org/10.1016/S0065-3233(08)60494-5.
Muramatsu, Masamichi, Kazuo Kinoshita, Sidonia Fagarasan, Shuichi Yamada, Yoichi Shinkai, and Tasuku Honjo. 2000. “Class Switch Recombination and Hypermutation Require Activation-Induced Cytidine Deaminase (AID), a Potential RNA Editing Enzyme.” *Cell* 102 (5): 553–63. https://doi.org/10.1016/S0092-8674(00)00078-7.
Murugan, Rajagopal, Lisa Buchauer, Gianna Triller, Cornelia Kreschel, Giulia Costa, Gemma Pidelaserra Martí, Katharina Imkeller, et al. 2018. “Clonal Selection Drives Protective Memory B Cell Responses in Controlled Human Malaria Infection.” *Science Immunology* 3 (20). https://doi.org/10.1126/sciimmunol.aap8029.
Navis, Marjon, Karen Tran, Shridhar Bale, Ganesh E. Phad, Javier Guenaga, Richard Wilson, Martina Soldemo, et al. 2014. “HIV-1 Receptor Binding Site-Directed Antibodies Using a VH1-2 Gene Segment Orthologue Are Activated by Env Trimer Immunization.” *PLoS Pathogens* 10 (8). https://doi.org/10.1371/journal.ppat.1004337.
Needham, Joseph. 2000. *Science and Civilisation in China Medicine by Joseph Needham. Volume 6. Biology and Biological Technology. Part 6. Medicine*. Cambridge: Cambridge Univ. Pr. file:///D:/gigapedia/Needh0521632625.pdf.
Nussenzweig, Michel C., Albert C. Shaw, Eric Sinn, David B. Danner, Kevin L. Holmes, Herbert C. Morse, and Philip Leder. 1987. “Allelic Exclusion in Transgenic Mice That Express the Membrane Form of Immunoglobulin μ.” *Science* 236 (4803): 816–19. https://doi.org/10.1126/science.3107126.
Nutt, Stephen L, Philip D Hodgkin, David M Tarlinton, and Lynn M Corcoran. 2015. “The Generation of Antibody-Secreting Plasma Cells.” *Nature Reviews. Immunology* 15 (3): 160–71. https://doi.org/10.1038/nri3795.
Nyrén, Pål, Bertil Pettersson, and Mathias Uhlén. 1993. “Solid Phase DNA Minisequencing by an Enzymatic Luminometric Inorganic Pyrophosphate Detection Assay.” *Analytical Biochemistry* 208 (1): 171–75. https://doi.org/10.1006/abio.1993.1024.
Ohlin, Mats, Cathrine Scheepers, Martin Corcoran, William D. Lees, Christian E. Busse, Davide Bagnara, Linnea Thörnqvist, et al. 2019. “Inferred Allelic Variants of Immunoglobulin Receptor Genes: A System for Their Evaluation, Documentation, and Naming.” *Frontiers in Immunology*. Frontiers Media S.A. https://doi.org/10.3389/fimmu.2019.00435.
Ohm-Laursen, Line, Stine Rosenkilde Larsen, and Torben Barington. 2005. “Identification of Two New Alleles, IGHV3-23*04 and IGHJ6*04, and the Complete Sequence of the IGHV3-h Pseudogene in the Human Immunoglobulin Locus and Their Prevalences in Danish Caucasians.” *Immunogenetics* 57 (9): 621–27. https://doi.org/10.1007/s00251-005-0035-8.
Osada, Naoki, Nilmini Hettiarachchi, Isaac Adeyemi Babarinde, Naruya Saitou, and Antoine Blancher. 2015. “Whole-Genome Sequencing of Six Mauritian Cynomolgus Macaques (Macaca Fascicularis) Reveals a Genome-Wide Pattern of Polymorphisms under Extreme Population Bottleneck.” *Genome Biology and Evolution* 7 (3): 821–30. https://doi.org/10.1093/gbe/evv033.
Osada, Naoki, Yasuhiro Uno, Katsuhiko Mineta, Yosuke Kameoka, Ichiro Takahashi, and Keiji Terao. 2010. “Ancient Genome-Wide Admixture Extends beyond the Current Hybrid Zone between Macaca Fascicularis and M. Mulatta.” *Molecular Ecology* 19 (14): 2884–95. https://doi.org/10.1111/j.1365-294X.2010.04687.x.
Otting, Nel, Corrine M C Heijmans, Riet C. Noort, Natasja G. De Groot, Gaby G M Doxiadis, Jon J. Van Rood, David I. Watkins, and Ronald E. Bontrop. 2005. “Unparalleled Complexity of the MHC Class I Region in Rhesus Macaques.” *Proceedings of the National Academy of Sciences of the United States of America* 102 (5): 1626–31. https://doi.org/10.1073/pnas.0409084102.
Papamichail, M, C Gutierrez, P Embling, P Johnson, E J Holborow, and M B Pepys. 1975. “Complement Dependence of Localisation of Aggregated IgG in Germinal Centres.” *Scandinavian Journal of Immunology* 4 (4): 343–47. https://doi.org/10.1111/j.1365-3083.1975.tb02635.x.
Peres, Ayelet, Moriah Gidoni, Pazit Polak, and Gur Yaari. 2019. “RAbHIT: R Antibody Haplotype Inference Tool.” *Bioinformatics*, June. https://doi.org/10.1093/bioinformatics/btz481.

Pfeiffer, Franziska, Carsten Gröber, Michael Blank, Kristian Händler, Marc Beyer, Joachim L. Schultze, and Günter Mayer. 2018. "Systematic Evaluation of Error Rates and Causes in Short Samples in Next-Generation Sequencing." *Scientific Reports* 8 (1). https://doi.org/10.1038/s41598-018-29325-6.
Phad, Ganesh E., Pradeepa Pushparaj, Karen Tran, Viktoriya Dubrovskaya, Monika Àdori, Paola Martinez-Murillo, Néstor Vázquez Bernat, et al. 2019. "Extensive Dissemination and Intraclonal Maturation of HIV Env Vaccine-Induced B Cell Responses." *The Journal of Experimental Medicine*, November, jem.20191155. https://doi.org/10.1084/jem.20191155.
Phad, Ganesh E., Néstor Vázquez Bernat, Yu Feng, Jidnyasa Ingale, Paola Andrea Martinez Murillo, Sijy O'Dell, Yuxing Li, et al. 2015. "Diverse Antibody Genetic and Recognition Properties Revealed Following HIV-1 Envelope Glycoprotein Immunization." *The Journal of Immunology* 194 (12): 5903–14. https://doi.org/10.4049/jimmunol.1500122.
Plotkin, Stanley A. 2008. "Vaccines: Correlates of Vaccine-Induced Immunity." *Clinical Infectious Diseases* 47 (3): 401–9. https://doi.org/10.1086/589862.
Pramanik, Sreemanta, Xiangfeng Cui, Hui Yun Wang, Nyam Osor Chimge, Guohong Hu, Li Shen, Richeng Gao, and Honghua Li. 2011. "Segmental Duplication as One of the Driving Forces Underlying the Diversity of the Human Immunoglobulin Heavy Chain Variable Gene Region." *BMC Genomics* 12 (January). https://doi.org/10.1186/1471-2164-12-78.
Pulendran, Bali, and Rafi Ahmed. 2011. "Immunological Mechanisms of Vaccination." *Nature Immunology*. https://doi.org/10.1038/ni.2039.
Ralph, Duncan K., and Frederick A. Matsen. 2019. "Per-Sample Immunoglobulin Germline Inference from B Cell Receptor Deep Sequencing Data." *PLOS Computational Biology* 15 (7): e1007133. https://doi.org/10.1371/journal.pcbi.1007133.
Ramesh, Akshaya, Sam Darko, Axin Hua, Glenn Overman, Amy Ransier, Joseph R. Francica, Ashley Trama, et al. 2017. "Structure and Diversity of the Rhesus Macaque Immunoglobulin Loci through Multiple de Novo Genome Assemblies." *Frontiers in Immunology* 8 (OCT). https://doi.org/10.3389/fimmu.2017.01407.
Rao, S P, J M Riggs, D F Friedman, M S Scully, T W LeBien, and L E Silberstein. 1999. "Biased VH Gene Usage in Early Lineage Human B Cells: Evidence for Preferential Ig Gene Rearrangement in the Absence of Selection." *Journal of Immunology (Baltimore, Md. : 1950)* 163 (5): 2732–40. http://www.ncbi.nlm.nih.gov/pubmed/10453015.
Rhoads, Anthony, and Kin Fai Au. 2015. "PacBio Sequencing and Its Applications." *Genomics, Proteomics and Bioinformatics*. Beijing Genomics Institute. https://doi.org/10.1016/j.gpb.2015.08.002.
Rizzetto, Simone, David N.P. Koppstein, Jerome Samir, Mandeep Singh, Joanne H. Reed, Curtis H. Cai, Andrew R. Lloyd, Auda A. Eltahla, Christopher C. Goodnow, and Fabio Luciani. 2018. "B-Cell Receptor Reconstruction from Single-Cell RNA-Seq with VDJPuzzle." *Bioinformatics (Oxford, England)* 34 (16): 2846–47. https://doi.org/10.1093/bioinformatics/bty203.
Robins, Harlan S., Paulo V. Campregher, Santosh K. Srivastava, Abigail Wacher, Cameron J. Turtle, Orsalem Kahsai, Stanley R. Riddell, Edus H. Warren, and Christopher S. Carlson. 2009. "Comprehensive Assessment of T-Cell Receptor β-Chain Diversity in Aβ T Cells." *Blood* 114 (19): 4099–4107. https://doi.org/10.1182/blood-2009-04-217604.
Robins, Harlan S., Santosh K. Srivastava, Paulo V. Campregher, Cameron J. Turtle, Jessica Andriesen, Stanley R. Riddell, Christopher S. Carlson, and Edus H. Warren. 2010. "Overlap and Effective Size of the Human CD8+ T Cell Receptor Repertoire." *Science Translational Medicine* 2 (47). https://doi.org/10.1126/scitranslmed.3001442.
Robinson, William H. 2014. "Sequencing the Functional Antibody Repertoire." *Nature Reviews Rheumatology* 11 (3): 171–82. https://doi.org/10.1038/nrrheum.2014.220.
Rogers, Jeffrey, and Richard A. Gibbs. 2014. "Comparative Primate Genomics: Emerging Patterns of Genome Content and Dynamics." *Nature Reviews Genetics*. Nature Publishing Group. https://doi.org/10.1038/nrg3707.
Romo-González, Tania, Jorge Morales-Montor, Mauricio Rodríguez-Dorantes, and Enrique Vargas-Madrazo. 2005. "Novel Substitution Polymorphisms of Human Immunoglobulin VH Genes in Mexicans." *Human Immunology* 66 (6): 731–39. https://doi.org/10.1016/j.humimm.2005.03.002.
Ronaghi, Mostafa, Samer Karamohamed, Bertil Pettersson, Mathias Uhlén, and Pål Nyrén. 1996. "Real-Time DNA Sequencing Using Detection of Pyrophosphate Release." *Analytical Biochemistry* 242 (1): 84–89. https://doi.org/10.1006/abio.1996.0432.

Roos, Christian, and Dietmar Zinner. 2015. “Diversity and Evolutionary History of Macaques with Special Focus on Macaca Mulatta and Macaca Fascicularis.” In *The Nonhuman Primate in Nonclinical Drug Development and Safety Assessment*, 3–16. Elsevier Inc. https://doi.org/10.1016/B978-0-12-417144-2.00001-9.

Rosenfeld, Ronit, Anat Zvi, Eitan Winter, Ronen Hope, Ofir Israeli, Ohad Mazor, and Gur Yaari. 2019. “A Primer Set for Comprehensive Amplification of V-Genes from Rhesus Macaque Origin Based on Repertoire Sequencing.” *Journal of Immunological Methods* 465 (February): 67–71. https://doi.org/10.1016/j.jim.2018.11.011.

Rubelt, Florian, Christian E. Busse, Syed Ahmad Chan Bukhari, Jean Philippe Bürckert, Encarnita Mariotti-Ferrandiz, Lindsay G. Cowell, Corey T. Watson, et al. 2017. “Adaptive Immune Receptor Repertoire Community Recommendations for Sharing Immune-Repertoire Sequencing Data.” *Nature Immunology*. Nature Publishing Group. https://doi.org/10.1038/ni.3873.

Sanger, F., and A. R. Coulson. 1975. “A Rapid Method for Determining Sequences in DNA by Primed Synthesis with DNA Polymerase.” *Journal of Molecular Biology* 94 (3). https://doi.org/10.1016/0022-2836(75)90213-2.

Sasso, E H, K W Van Dijk, and E C Milner. 1990. “Prevalence and Polymorphism of Human VH3 Genes.” *Journal of Immunology (Baltimore, Md. : 1950)* 145 (8): 2751–57. http://www.ncbi.nlm.nih.gov/pubmed/1976703.

Sasso, Eric H., Jane Hoyt Buckner, and Lucy A. Suzuki. 1995. “Ethnic Differences in Polymorphism of an Immunoglobulin VH3 Gene.” *Journal of Clinical Investigation* 96 (3): 1591–1600. https://doi.org/10.1172/JCI118198.

Sather, D. N., J. Armann, L. K. Ching, A. Mavrantoni, G. Sellhorn, Z. Caldwell, X. Yu, et al. 2009. “Factors Associated with the Development of Cross-Reactive Neutralizing Antibodies during Human Immunodeficiency Virus Type 1 Infection.” *Journal of Virology* 83 (2): 757–69. https://doi.org/10.1128/jvi.02036-08.

Scharf, Louise, Anthony P. West, Han Gao, Terri Lee, Johannes F. Scheid, Michel C. Nussenzweig, Pamela J. Bjorkman, and Ron Diskin. 2013. “Structural Basis for HIV-1 Gp120 Recognition by a Germ-Line Version of a Broadly Neutralizing Antibody.” *Proceedings of the National Academy of Sciences of the United States of America* 110 (15): 6049–54. https://doi.org/10.1073/pnas.1303682110.

Schatz, David G, and Patrick C Swanson. 2011. “V(D)J Recombination: Mechanisms of Initiation.” *Annual Review of Genetics* 45: 167–202. https://doi.org/10.1146/annurev-genet-110410-132552.

Scheepers, Cathrine, Ram K. Shrestha, Bronwen E. Lambson, Katherine J. L. Jackson, Imogen A. Wright, Dshanta Naicker, Mark Goosen, et al. 2015. “Ability To Develop Broadly Neutralizing HIV-1 Antibodies Is Not Restricted by the Germline Ig Gene Repertoire.” *The Journal of Immunology* 194 (9): 4371–78. https://doi.org/10.4049/jimmunol.1500118.

Scheid, JF, Hugo Mouquet, Beatrix Ueberheide, and Ron Diskin. 2011. “Sequence and Structural Convergence of Broad and Potent HIV Antibodies That Mimic CD4 Binding.” *Science* 333 (September): 1633–37. http://www.sciencemag.org/content/333/6049/1633.short.

Schroeder, Harry W, and Lisa Cavacini. 2010. “Structure and Function of Immunoglobulins.” *Journal of Allergy and Clinical Immunology* 125 (2, Supplement 2): S41–52. https://doi.org/https://doi.org/10.1016/j.jaci.2009.09.046.

Sela-Culang, Inbal, Yanay Ofran, and Bjoern Peters. 2015. “Antibody Specific Epitope Prediction-Emergence of a New Paradigm.” *Current Opinion in Virology* 11 (April): 98–102. https://doi.org/10.1016/j.coviro.2015.03.012.

Sircar, Aroop, Eric T. Kim, and Jeffrey J. Gray. 2009. “RosettaAntibody: Antibody Variable Region Homology Modeling Server.” *Nucleic Acids Research* 37 (SUPPL. 2). https://doi.org/10.1093/nar/gkp387.

Smith, David Glenn, John W. Mcdonough, and Debra A. George. 2007. “Mitochondrial DNA Variation within and among Regional Populations of Longtail Macaques (Macaca Fascicularis) in Relation to Other Species of the Fascicularis Group of Macaques.” *American Journal of Primatology* 69 (2): 182–98. https://doi.org/10.1002/ajp.20337.

Smith, Kenneth, Jennifer J Muther, Angie L Duke, Emily McKee, Nai-Ying Zheng, Patrick C Wilson, and Judith A James. 2013. “Fully Human Monoclonal Antibodies from Antibody Secreting Cells after Vaccination with Pneumovax®23 Are Serotype Specific and Facilitate Opsonophagocytosis.” *Immunobiology* 218 (5): 745–54. https://doi.org/10.1016/j.imbio.2012.08.278.

Staden, R. 1979. “A Strategy of DNA Sequencing Employing Computer Programs.” *Nucleic Acids Research* 6 (7): 2601–10. https://doi.org/10.1093/nar/6.7.2601.
Stamatopoulos, B, A Timbs, D Bruce, T Smith, R Clifford, P Robbe, A Burns, et al. 2017. “Targeted Deep Sequencing Reveals Clinically Relevant Subclonal IgHV Rearrangements in Chronic Lymphocytic Leukemia.” *Leukemia* 31 (4): 837–45. https://doi.org/10.1038/leu.2016.307.
Stern, Alexandra Minna, and Howard Markel. 2005. “The History of Vaccines and Immunization: Familiar Patterns, New Challenges - If We Could Match the Enormous Scientific Strides of the Twentieth Century with the Political and Economic Investments of the Nineteenth, the World’s Citizens Might Be Much Healthier.” *Health Affairs*. https://doi.org/10.1377/hlthaff.24.3.611.
Street, Summer L., Randall C. Kyes, Richard Grant, and Betsy Ferguson. 2007. “Single Nucleotide Polymorphisms (SNPs) Are Highly Conserved in Rhesus (Macaca Mulatta) and Cynomolgus (Macaca Fascicularis) Macaques.” *BMC Genomics* 8 (December). https://doi.org/10.1186/1471-2164-8-480.
Sundling, Christopher, Yuxing Li, Nick Huynh, Christian Poulsen, Richard Wilson, Sijy O’Dell, Yu Feng, John R. Mascola, Richard T. Wyatt, and Gunilla B. Karlsson Hedestam. 2012. “High-Resolution Definition of Vaccine-Elicited B Cell Responses against the HIV Primary Receptor Binding Site.” *Science Translational Medicine* 4 (142): 142ra96. https://doi.org/10.1126/scitranslmed.3003752.
Sussman, Robert W., and Ian Tattersall. 1986. “Distribution, Abundance, and Putative Ecological Strategy of Macaca Fascicularis on the Island of Mauritius, Southwestern Indian Ocean.” *Folia Primatologica* 46 (1): 28–43. https://doi.org/10.1159/000156234.
Suzuki, I, L Pfister, A Glas, C Nottenburg, and E C Milner. 1995. “Representation of Rearranged VH Gene Segments in the Human Adult Antibody Repertoire.” *Journal of Immunology (Baltimore, Md. : 1950)* 154 (8): 3902–11. http://www.ncbi.nlm.nih.gov/pubmed/7706729.
Tarlinton, David. 2012. “B-Cell Differentiation: Instructive One Day, Stochastic the Next.” *Current Biology : CB* 22 (7): R235-7. https://doi.org/10.1016/j.cub.2012.02.045.
Tarlinton, David M. 2008. “Evolution in Miniature: Selection, Survival and Distribution of Antigen Reactive Cells in the Germinal Centre.” *Immunology and Cell Biology* 86 (2): 133–38. https://doi.org/10.1038/sj.icb.7100148.
Teng, Grace, and F Nina Papavasiliou. 2007. “Immunoglobulin Somatic Hypermutation.” *Annual Review of Genetics* 41: 107–20. https://doi.org/10.1146/annurev.genet.41.110306.130340.
Tiegs, Susan L., David M. Russell, and David Nemazee. 1993. “Receptor Editing in Self-Reactive Bone Marrow B Cells.” *Journal of Experimental Medicine* 177 (4): 1009–20. https://doi.org/10.1084/jem.177.4.1009.
Tomlinson, M. Ian, Graham P. Cook, P. Carter Nigel, Ramnath Elaswarapu, Sarah Smith, Gerald Walter, Lakjaya Buluwela, Terence H. Rabbitts, and Greg Winter. 1994. “Human Immunoglobulin Vh and D Segments on Chromosomes 15q11.2 and 16p11.2.” *Human Molecular Genetics* 3 (6): 853–60. https://doi.org/10.1093/hmg/3.6.853.
Tonegawa, Susumu. 1983. “Somatic Generation of Antibody Diversity.” *Nature* 302 (5909): 575–81. https://doi.org/10.1038/302575a0.
Tran, Vuvi G, Arundhathi Venkatasubramaniam, Rajan P Adhikari, Subramaniam Krishnan, Xing Wang, Vien T M Le, Hoan N Le, et al. 2019. “Efficacy of Active Immunization With Attenuated α-Hemolysin and Panton-Valentine Leukocidin in a Rabbit Model of Staphylococcus Aureus Necrotizing Pneumonia.” *The Journal of Infectious Diseases*, August. https://doi.org/10.1093/infdis/jiz437.
Trück, Johannes, Maheshi N. Ramasamy, Jacob D. Galson, Richard Rance, Julian Parkhill, Gerton Lunter, Andrew J. Pollard, and Dominic F. Kelly. 2015. “Identification of Antigen-Specific B Cell Receptor Sequences Using Public Repertoire Analysis.” *The Journal of Immunology* 194 (1): 252–61. https://doi.org/10.4049/jimmunol.1401405.
Uduman, Mohamed, Mark J. Shlomchik, Francois Vigneault, George M. Church, and Steven H. Kleinstein. 2014. “Integrating B Cell Lineage Information into Statistical Tests for Detecting Selection in Ig Sequences.” *The Journal of Immunology* 192 (3): 867–74. https://doi.org/10.4049/jimmunol.1301551.
Upadhyay, Amit A., Robert C. Kauffman, Amber N. Wolabaugh, Alice Cho, Nirav B. Patel, Samantha M. Reiss, Colin Havenar-Daughton, et al. 2018. “BALDR: A Computational Pipeline for Paired Heavy and Light Chain Immunoglobulin Reconstruction in Single-Cell RNA-Seq Data.” *Genome Medicine* 10 (1). https://doi.org/10.1186/s13073-018-0528-3.

Vázquez Bernat, Néstor, Martin Corcoran, Uta Hardt, Mateusz Kaduk, Ganesh E. Phad, Marcel Martin, and Gunilla B. Karlsson Hedestam. 2019. "High-Quality Library Preparation for NGS-Based Immunoglobulin Germline Gene Inference and Repertoire Expression Analysis." *Frontiers in Immunology* 10 (April). https://doi.org/10.3389/fimmu.2019.00660.
Venter, Craig. J., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, et al. 2001. "The Sequence of the Human Genome." *Science* 291 (5507): 1304–51. https://doi.org/10.1126/science.1058040.
Victora, Gabriel D, and Michel C Nussenzweig. 2012. "Germinal Centers." *Annual Review of Immunology* 30 (January): 429–57. https://doi.org/10.1146/annurev-immunol-020711-075032.
Victora, Gabriel D, Tanja A Schwickert, David R Fooksman, Alice O Kamphorst, Michael Meyer-Hermann, Michael L Dustin, and Michel C Nussenzweig. 2010. "Germinal Center Dynamics Revealed by Multiphoton Microscopy with a Photoactivatable Fluorescent Reporter." *Cell* 143 (4): 592–605. https://doi.org/10.1016/j.cell.2010.10.032.
Vigdorovich, Vladimir, Brian G Oliver, Sara Carbonetti, Nicholas Dambrauskas, Miles D Lange, Christina Yacoob, Will Leahy, Jonathan Callahan, Leonidas Stamatatos, and D Noah Sather. 2016. "Repertoire Comparison of the B-Cell Receptor-Encoding Loci in Humans and Rhesus Macaques by next-Generation Sequencing." *Clinical & Translational Immunology* 5 (7): e93. https://doi.org/10.1038/cti.2016.42.
Waltari, Eric, Manxue Jia, Caroline S. Jiang, Hong Lu, Jing Huang, Cristina Fernandez, Andrés Finzi, et al. 2018. "5' Rapid Amplification of CDNA Ends and Illumina MiSeq Reveals B Cell Receptor Features in Healthy Adults, Adults with Chronic HIV-1 Infection, Cord Blood, and Humanized Mice." *Frontiers in Immunology* 9 (MAR). https://doi.org/10.3389/fimmu.2018.00628.
Wang, Bo, Brandon J DeKosky, Morgan R Timm, Jiwon Lee, Erica Normandin, John Misasi, Rui Kong, et al. 2018. "Functional Interrogation and Mining of Natively Paired Human VH:VL Antibody Repertoires." *Nature Biotechnology* 36 (2): 152–55. https://doi.org/10.1038/nbt.4052.
Wang, Chen, Yi Liu, Mary M. Cavanagh, Sabine Le Saux, Qian Qi, Krishna M. Roskin, Timothy J. Looney, et al. 2015. "B-Cell Repertoire Responses to Varicella-Zoster Vaccination in Human Identical Twins." *Proceedings of the National Academy of Sciences* 112 (2): 500–505. https://doi.org/10.1073/pnas.1415875112.
Wang, Chen, Yi Liu, Lan T. Xu, Katherine J. L. Jackson, Krishna M. Roskin, Tho D. Pham, Jonathan Laserson, et al. 2014. "Effects of Aging, Cytomegalovirus Infection, and EBV Infection on Human B Cell Repertoires." *The Journal of Immunology* 192 (2): 603–11. https://doi.org/10.4049/jimmunol.1301384.
Wang, Yan, Katherine J. Jackson, Bruno Gäeta, William Pomat, Peter Siba, William A. Sewell, and Andrew M. Collins. 2011. "Genomic Screening by 454 Pyrosequencing Identifies a New Human IGHV Gene and Sixteen Other New IGHV Allelic Variants." *Immunogenetics* 63 (5): 259–65. https://doi.org/10.1007/s00251-010-0510-8.
Wang, Yan, Katherine J L Jackson, William A. Sewell, and Andrew M. Collins. 2008. "Many Human Immunoglobulin Heavy-Chain IGHV Gene Polymorphisms Have Been Reported in Error." *Immunology and Cell Biology* 86 (2): 111–15. https://doi.org/10.1038/sj.icb.7100144.
Wardemann, Hedda, and Christian E. Busse. 2017. "Novel Approaches to Analyze Immunoglobulin Repertoires." *Trends in Immunology*. Elsevier Ltd. https://doi.org/10.1016/j.it.2017.05.003.
Waterston, Robert H., Kerstin Lindblad-Toh, Ewan Birney, Jane Rogers, Josep F. Abril, Pankaj Agarwal, Richa Agarwala, et al. 2002. "Initial Sequencing and Comparative Analysis of the Mouse Genome." *Nature* 420 (6915): 520–62. https://doi.org/10.1038/nature01262.
Watson, Corey T., and Felix Breden. 2012. "The Immunoglobulin Heavy Chain Locus: Genetic Variation, Missing Data, and Implications for Human Disease." *Genes and Immunity*. https://doi.org/10.1038/gene.2012.12.
Watson, Corey T., Karyn M. Steinberg, John Huddleston, Rene L. Warren, Maika Malig, Jacqueline Schein, A. Jeremy Willsey, et al. 2013. "Complete Haplotype Sequence of the Human Immunoglobulin Heavy-Chain Variable, Diversity, and Joining Genes and Characterization of Allelic and Copy-Number Variation." *American Journal of Human Genetics* 92 (4): 530–46. https://doi.org/10.1016/j.ajhg.2013.03.004.
Watson, Corey T., Karyn Meltz Steinberg, Tina A. Graves, Rene L. Warren, Maika Malig, Jacqueline Schein, Richard K. Wilson, Robert A. Holt, Evan E Eichler, and Felix Breden. 2015. "Sequencing of the Human IG Light Chain Loci from a Hydatidiform Mole BAC Library

Reveals Locus-Specific Signatures of Genetic Diversity.” *Genes and Immunity* 16 (1): 24–34. https://doi.org/10.1038/gene.2014.56.
Watson, Corey T, Jacob Glanville, and Wayne A Marasco. 2017. “The Individual and Population Genetics of Antibody Immunity.” *Trends in Immunology* 38 (7): 459–70. https://doi.org/10.1016/j.it.2017.04.003.
Watson, Corey T, Justin T Kos, William S Gibson, Leah Newman, Gintaras Deikus, Christian E Busse, Melissa L Smith, Katherine JL Jackson, and Andrew M Collins. 2019. “ A Comparison of Immunoglobulin IGHV , IGHD and IGHJ Genes in Wild-derived and Classical Inbred Mouse Strains .” *Immunology & Cell Biology*, October. https://doi.org/10.1111/imcb.12288.
Weinstein, Joshua A, Ning Jiang, Richard A White, Daniel S Fisher, and Stephen R Quake. 2009. “High-Throughput Sequencing of the Zebrafish Antibody Repertoire.” *Science (New York, N.Y.)* 324 (5928): 807–10. https://doi.org/10.1126/science.1170020.
Weitzner, Brian D., Jeliazko R. Jeliazkov, Sergey Lyskov, Nicholas Marze, Daisuke Kuroda, Rahel Frick, Jared Adolf-Bryfogle, Naireeta Biswas, Roland L. Dunbrack, and Jeffrey J. Gray. 2017. “Modeling and Docking of Antibody Structures with Rosetta.” *Nature Protocols* 12 (2): 401–16. https://doi.org/10.1038/nprot.2016.180.
Weng, Nan-Ping -P, James G. Snyder, Li-Yuan -Y Yu-Lee, and Donald M. Marcus. 1992. “Polymorphism of Human Immunoglobulin VH4 Germ-line Genes.” *European Journal of Immunology* 22 (4): 1075–82. https://doi.org/10.1002/eji.1830220430.
Willems van Dijk, K, H W Schroeder, R M Perlmutter, and E C Milner. 1989. “Heterogeneity in the Human Ig VH Locus.” *Journal of Immunology (Baltimore, Md. : 1950)* 142 (7): 2547–54. http://www.ncbi.nlm.nih.gov/pubmed/2494263.
Wrammert, Jens, Kenneth Smith, Joe Miller, William A Langley, Kenneth Kokko, Christian Larsen, Nai-Ying Zheng, et al. 2008. “Rapid Cloning of High-Affinity Human Monoclonal Antibodies against Influenza Virus.” *Nature* 453 (7195): 667–71. https://doi.org/10.1038/nature06890.
Wu, Xueling, Zhi-yong Yang, Yuxing Li, Carl-magnus Hogerkorp, R William, Michael S Seaman, Tongqing Zhou, et al. 2010. “Rational Design of Envelope Identifies Broadly Neutralizing Human Monoclonal Antibodies to HIV-1.” *Science* 329 (5993): 856–61. https://doi.org/10.1126/science.1187659.Rational.
Yan, Guangmei, Guojie Zhang, Xiaodong Fang, Yanfeng Zhang, Cai Li, Fei Ling, David N. Cooper, et al. 2011. “Genome Sequencing and Comparison of Two Nonhuman Primate Animal Models, the Cynomolgus and Chinese Rhesus Macaques.” *Nature Biotechnology* 29 (11): 1019–23. https://doi.org/10.1038/nbt.1992.
Yancopoulos, George D., Stephen V. Desiderio, Michael Paskind, John F. Kearney, David Baltimore, and Frederick W. Alt. 1984. “Preferential Utilization of the Most JH-Proximal VH Gene Segments in Pre-B-Cell Lines.” *Nature* 311 (5988): 727–33. https://doi.org/10.1038/311727a0.
Ye, Jian, Ning Ma, Thomas L Madden, and James M Ostell. 2013. “IgBLAST: An Immunoglobulin Variable Domain Sequence Analysis Tool.” *Nucleic Acids Research* 41 (Web Server issue): W34-40. https://doi.org/10.1093/nar/gkt382.
Yoshida, K, T K van den Berg, and C D Dijkstra. 1993. “Two Functionally Different Follicular Dendritic Cells in Secondary Lymphoid Follicles of Mouse Spleen, as Revealed by CR1/2 and FcR Gamma II-Mediated Immune-Complex Trapping.” *Immunology* 80 (1): 34–39. http://www.ncbi.nlm.nih.gov/pubmed/8244461.
Yu, Guo Yun, Suzanne Mate, Karla Garcia, Michael D. Ward, Ernst Brueggemann, Matthew Hall, Tara Kenny, Mariano Sanchez-Lockhart, Marie Paule Lefranc, and Gustavo Palacios. 2016. “Cynomolgus Macaque (Macaca Fascicularis) Immunoglobulin Heavy Chain Locus Description.” *Immunogenetics* 68 (6–7): 417–28. https://doi.org/10.1007/s00251-016-0921-2.
Yuan, Qiaoping, Zhifeng Zhou, Stephen G. Lindell, J. D. Higley, Betsy Ferguson, Robert C. Thompson, Juan F. Lopez, et al. 2012. “The Rhesus Macaque Is Three Times as Diverse but More Closely Equivalent in Damaging Coding Variation as Compared to the Human.” *BMC Genetics* 13 (June). https://doi.org/10.1186/1471-2156-13-52.
Zan, Hong, and Paolo Casali. 2008. “AID- and Ung-Dependent Generation of Staggered Double-Strand DNA Breaks in Immunoglobulin Class Switch DNA Recombination: A Post-Cleavage Role for AID.” *Molecular Immunology* 46 (1): 45–61. https://doi.org/10.1016/j.molimm.2008.07.003.
Zhang, Bochao, Wenzhao Meng, Eline T Luning Prak, and Uri Hershberg. 2015. “Discrimination of Germline V Genes at Different Sequencing Lengths and Mutational Burdens: A New Tool for

Identifying and Evaluating the Reliability of V Gene Assignment.” *Journal of Immunological Methods* 427 (December): 105–16. https://doi.org/10.1016/j.jim.2015.10.009.
Zhang, Jiajie, Kassian Kobert, Tomáš Flouri, and Alexandros Stamatakis. 2014. “PEAR: A Fast and Accurate Illumina Paired-End ReAd MergeR.” *Bioinformatics* 30 (5): 614–20. https://doi.org/10.1093/bioinformatics/btt593.
Zhang, Wei, Xinyue Li, Longlong Wang, Jianxiang Deng, Liya Lin, Lei Tian, Jinghua Wu, et al. 2019. “Identification of Variable and Joining Germline Genes and Alleles for Rhesus Macaque from B Cell Receptor Repertoires.” *The Journal of Immunology* 202 (5): 1612–22. https://doi.org/10.4049/jimmunol.1800342.
Zhang, Wei, I. Ming Wang, Changxi Wang, Liya Lin, Xianghua Chai, Jinghua Wu, Andrew J. Bett, Govindarajan Dhanasekaran, Danilo R. Casimiro, and Xiao Liu. 2016. “IMPre: An Accurate and Efficient Software for Prediction of T- and B-Cell Receptor Germline Genes and Alleles from Rearranged Repertoire Data.” *Frontiers in Immunology* 7 (NOV). https://doi.org/10.3389/fimmu.2016.00457.
Zhou, Julian Q., and Steven H. Kleinstein. 2019. “Cutting Edge: Ig H Chains Are Sufficient to Determine Most B Cell Clonal Relationships.” *The Journal of Immunology* 203 (7): 1687–92. https://doi.org/10.4049/jimmunol.1900666.
Zimin, Aleksey V., Adam S. Cornish, Mnirnal D. Maudhoo, Robert M. Gibbs, Xiongfei Zhang, Sanjit Pandey, Daniel T. Meehan, et al. 2014. “A New Rhesus Macaque Assembly and Annotation for Next-Generation Sequencing Analyses.” *Biology Direct* 9 (1): 20. https://doi.org/10.1186/1745-6150-9-20.