Analysis of the human B cell repertoire following vaccination

Jacob Galson
Department of Paediatrics
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy (DPhil)
Trinity Term 2016

Analysis of the human B cell repertoire following vaccination

Jacob Galson, St Cross College. DPhil in Paediatrics, Trinity Term 2016

Abstract

A diverse B cell repertoire is essential for recognition and response to infectious and vaccine antigens. High-throughput sequencing of B cell receptor (BCR) genes can now be used to study the B cell repertoire at a depth which reflects its true diversity. As a relatively new technology, there is little information on the structure of the repertoire at baseline, whether antigen-specific changes can be detected from the total repertoire following antigen stimulus, and what the potential clinical applications of BCR repertoire sequencing are.

In this thesis, robust laboratory and bioinformatic techniques were developed for studying the BCR repertoire. These were then applied to healthy participants to assess inter- and intra-individual variation in the repertoire. Hepatitis B vaccination was then used as a model system to determine how the repertoire responded to both primary and booster vaccination. Tracking repertoire dynamics following booster vaccination identified the presence of time-limited changes in the total BCR repertoire following stimulation. Cell sorting and sequencing of vaccine-specific cells in addition to sequencing the total repertoire allowed deconvolution of the vaccine-specific response from background repertoire fluctuations. Studying the response to primary vaccination showed the same time-limited changes, and revealed that a surprising number of the B cells activated appear to be derived from memory cells, and activated by the vaccine in a cross-reactive manner. More specific applications of BCR repertoire sequencing were then investigated in the context of meningococcal and influenza vaccine studies. These were able to distinguish the different cell subsets activated in response to meningococcal polysaccharide and conjugate vaccines, and shed light on how the AS03 adjuvant increases pandemic influenza vaccine immunogenicity.

In summary, presented in this thesis are some of the first BCR repertoire data following Hepatitis B, meningococcal and influenza vaccination. These data have increased our fundamental understanding of the BCR repertoire, and how this responds to vaccination. Insights from these data raise promise for the application of this technology to clinical settings for vaccine evaluation, disease diagnostics and monitoring, and therapeutic antibody discovery, and have provided a foundation for many further studies.

Declaration

I herewith declare that I have produced this thesis without the prohibited assistance of third parties and without making use of aids other than those specified in the text and acknowledgements section; notions taken over directly or indirectly from other sources have been identified as such. This thesis has not previously been presented in identical or similar form to any other English or foreign examination board.

The thesis work was conducted from October 2012 to May 2016 under the supervision of Dr. Dominic Kelly and Professor Gerton Lunter at The University of Oxford.

Jacob Galson
University of Oxford, May 2016

Acknowledgements

I would like to thank my two academic supervisors, Dr. Dominic Kelly, and Prof. Gerton Lunter, as well as Prof. Andrew Pollard for giving me the opportunity to carry out this project, and for their advice and support throughout my DPhil. In particular, I thank Dominic for his continual guidance and inspiration, and his patience in replying to over 1,000 emails from me over the last 4 years. In addition, I would like to thank everyone at the Oxford Vaccine Group who was involved with study organisation, sample collection, laboratory processing, and helpful discussions. I would also like to thank my industrial supervisor, Dr Robbert van der Most, for giving me the opportunity to go over to Belgium to work at GSK and experience research in an industrial setting. Finally, I would also like to thank all of the study participants who have made this work possible.

Funding for this DPhil was provided by the Biotechnology and Biological Sciences Research Council, and GlaxoSmithKline in the form of an iCASE studentship. Further study funding was provided by the National Institute for Health Research Oxford Biomedical Research Council.

In addition, I would like to acknowledge specific contributions from the following:

- Dr. Dominic Kelly, Dr. Johannes Truck and Prof. Andrew Pollard for assistance with designing the Hepatitis B vaccine study
- Prof. Andrew Pollard, Dr. Matthew Snape, Dr. Dominic Kelly, Dr. Alastair McGregor and Dr. Maheshi Ramasamy for designing and carrying out the meningococcal vaccine study
- All of the doctors and nurses at the Oxford Vaccine Group for carrying out study screening, and blood sampling
- Administrative staff at the Oxford Vaccine Group for study organisation
- The Microbiology laboratory at the John Radcliffe Hospital for carrying out anti-HBs testing
- The High Throughput Genomics group at the Wellcome Trust Centre for Human Genetics (subsidized by grant 090532/Z/09Z) for carrying out Illumina sequencing
- Mr. Craig Waugh and Dr. Elizabeth Clutterbuck for help with FACS for the HepB and ACWY studies
- Prof. Gerton Lunter, Dr. Anna Fowler, Dr. Johannes Truck, and Dr. Daniel O'Connor for help with bioinformatics
- Dr. Marton Munz for writing the sequence clustering script
- Dr. Johannes Truck, Dr. Dominic Kelly, and Dr. Elizabeth Clutterbuck for helpful discussions on everything B cell related
- Ms. Laurie Dumont for assistance with carrying out sample preparation for the influenza study

- Mr. Philippe Auquier, '*mon parrain*', for showing me the ropes at GSK, and improving my French

On a personal note, I would like to thank my family for all of their encouragement and support which has enabled me to get this far. I would also like to thank Laurie Holt for all of her support.

Publications

The following publications relate to work completed whilst undertaking this thesis.

Original research

- J. D. Galson, J. Trück, D. F. Kelly. & R. van der Most **Investigating the effect of AS03 adjuvant on the plasma cell repertoire following pH1N1 influenza vaccination.** - submitted
- J. D. Galson, J. Trück, E. Clutterbuck, A. Fowler, V. Cerundolo, A. J. Pollard, G. Lunter & D. F. Kelly. **B cell repertoire dynamics after sequential Hepatitis B vaccination, and evidence for cross-reactive B cell activation.** *Genome Medicine*, 2016, 8(68):1-13
- A. Llibre, C. Lopez-Macias, T. Marafioti, H. Mehta, A. Partridge, C. Kanzig, F. Rivellese, J. D. Galson, L. J. Walker, P. Milne, R. E. Phillips, D. Kelly, G. J. Freeman, M. E. El Sheikh, P. Klenerman & C. B. Willberg. **LLT1 and CD161 expression in human germinal centers promotes B cell activation and CXCR4 downregulation.** *Journal of Immunology*, 2016, 196(5):2085-2094
- J. D. Galson, J. Trück, A. Fowler, E. Clutterbuck, M. Münz, V. Cerundolo, C. Reinhard, R. van der Most, A. J. Pollard, G. Lunter & D. F. Kelly. **Analysis of B cell repertoire dynamics following Hepatitis B vaccination in humans, and enrichment of vaccine-specific antibody sequences.** *EBioMedicine*, 2015, 2(12):2070-2079
- J. D. Galson, J. Trück, A. Fowler, M. Münz, V. Cerundolo, A. J. Pollard, G. Lunter & D. F. Kelly. **In-depth assessment of within-individual and inter-individual variation in the B cell receptor repertoire.** *Frontiers in Immunology*, 2015, 6(531):1-13
- J. D. Galson, E. Clutterbuck, J. Trück, M. N. Ramasamy, M. Münz, A. Fowler, V. Cerundolo, A. J. Pollard, G. Lunter & D. F. Kelly. **BCR repertoire sequencing: different patterns of B-cell activation after two Meningococcal vaccines.** *Immunology and Cell Biology*, 2015, May:1-11
- J. Trück, M. N. Ramasamy, J. D. Galson, R. Rance, J. Parkhill, G. Lunter, A. J. Pollard & D. F. Kelly. **Identification of antigen-specific B cell receptor sequences using public repertoire analysis.** *Journal of Immunology*, 2014, 194(1):252-61

Review articles

- J. D. Galson, D. F. Kelly & J. Trück. **Identification of antigen-specific B cell receptor sequences from the total B cell repertoire.** *Critical Reviews in Immunology*, 2015, 35(6):463-78
- J. D. Galson, A. J. Pollard, J. Trück & D. F. Kelly. **Studying the antibody repertoire after vaccination: practical applications.** *Trends in Immunology*, 2014, 35(7):319-31

- J. D. Galson & D. O'Connor. **Immunogenicity.** *eLS. John Wiley & Sons Ltd* , July 2015

Clinical trial protocols

- D. F. Kelly, J. D. Galson, J. Trück, E. Kemp & A. J. Pollard. **Hepatitis B immunisation: A two-part study investigating antigen-specific B cell receptors.** *ClinicalTrials.gov Bethesda (MD): National Library of Medicine (US)*, 2013, NLM Identifier: NCT01821547

Contents

List of Figures

List of Tables

Glossary

AA	Amino acid
AID	Activation-induced cytidine deaminase
APC	Antigen-presenting cell
BCR	B cell receptor
bp	Base pair
CDR	Complimentarity-determining region
CMV	Cytomegalovirus
CSR	Class-switch recombination
DC	Dendritic cell
EBV	Epstein-barr virus
Env	Human immunodeficiency virus envelope glycoprotein
FACS	Fluorescence-activated cell sorting
FDC	Follicular dendritic cell
FMO	Fluorescence minus one control
FO	Follicular origin
GC	Germinal center
HA	Haemagglutinin
HAI	Haemagglutination inhibition assay
HBsAg	Hepatitis B surface antigen
Hib	*Haemophilus influenzae* type B
Ig	Immunoglobulin
LPS	Lipopolysaccharide
mAb	Monoclonal antibody
MACS	Magnetic-activated cell sorting
MHC	Major histocompatibility complex
MZ	Marginal zone
NGS	Next-generation sequencing
nt	Nucleotide
PAMP	Pathogen associated molecular pattern
PBMC	Peripheral blood mononuclear cell
PBS	Phosphate buffered saline
PC	Plasma cell
PCR	Polymerase chain reaction
PRR	Pattern regognition receptor
RAG	Recombination activating gene
RSS	Recombination signal sequence
SEM	Standard error of the mean
T_{FH}	Follicular helper T cell
TCR	T cell receptor
TD	T cell-dependent
TdT	Terminal dideoxynucleotidyl transferase
TI	T cell-independent
TIV	Trivalent influenza vaccine
TLR	Toll-like receptor
TT	Tetanus toxoid
UMI	Unique molecular identifier

1

Introduction

1.1 The role of B cells in the immune system

The immune system comprises a complex set of interacting cells and molecules, such as cytokines, that provide protection against infectious disease. This system can broadly be split into innate immunity, which provides a generalised and rapid first line of defence against an invading pathogen, and adaptive immunity which provides specific and long-lasting protection against the pathogen. B cells are primarily a component of the adaptive immune system, but evidence is also emerging for their function in innate immunity as well (1). The clinical importance of B cells is demonstrated by the life-threatening disorders associated with primary immunodeficiencies resulting from defects in the production and function of B cells (2).

1.1.1 Structure and function of immunoglobulin

The primary immune function of B cells is the production of antigen-specific immunoglobulin (Ig). This immunoglobulin is present on the surface of B cells as the B cell receptor (BCR), where it participates in the process of B cell activation through specific recognition of antigen, and is also secreted in the form of antibody. This antibody may either bind to and directly neutralize the pathogen or pathogenic product (secreted molecules important for pathogenesis, such as toxins), opsonise the pathogen to enhance phagocytosis, or fix terminal complement leading to bactericidal activity (3). To achieve this functionality, the antibody molecule consists

of two highly variable antigen-binding regions, as well as a constant region which modulates the immune activity.

The basic structure of the Ig consists of two identical heavy (IgH) and light (IgL) polypeptide chains, joined by disulphide bonds (4) (Figure 1.1). There are two classes of light chain - κ and λ - which are coded for by different chromosomal loci. The N-terminus of each IgH and IgL chain forms the variable region (denoted as V_H and V_L respectively), while the C-terminus forms the constant region. The variable regions are generated by the rearrangement of different germline variable (V), diversity (D; IgH only) and joining (J) gene segments (Section 1.2.1). Variability is not spread evenly through the V_H and V_L regions, but concentrated in three hypervariable domains, found in both the heavy and light chains, denoted HV1, HV2 and HV3 (5). The segments that adjoin these hypervariable regions are the less variable framework regions, FWR1, FWR2, FWR3 and FWR4. After pairing of the heavy and light variable domains, the hypervariable regions of the heavy and light chains come together to create a single hypervariable site, which is responsible for binding the antigen. As it is these hypervariable regions that are responsible for antigen-binding, they are also known as the complementarity-determining regions, CDR1, CDR2 and CDR3. These can form a variety of different structures to allow binding to antigens with different structures (6). CDR1 and CDR2 are within the V gene segment, while CDR3 (the most variable) is formed at the junction of the V, D and J gene segments.

The constant region of the Ig is at the C-terminus, and defines its isotype. The constant region mediates the effector response of the antibody, by binding to Fc receptors on effector cells, and by activating other immune pathways such as complement. There are five different human Ig isotypes, including 9 isotype subclasses in total (IgA1, IgA2 IgG1, IgG2, IgG3, IgG4 IgD, IgE and IgM), each defined by the constant regions, and with varying effector functions (Table 1.1). A B cell may switch from expressing antibodies of a certain isotype by changing

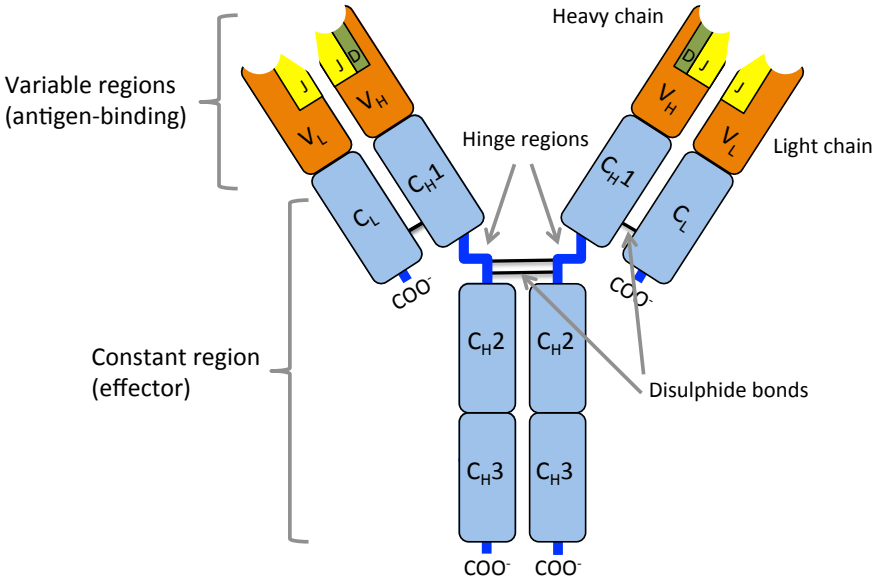

Figure 1.1: The structure of the immunoglobulin molecule - Each immunoglobulin molecule consists of two paired heavy (H) and light (L) chains joined by disulphide bonds, which forms two variable antigen-binding regions (V), and a constant effector region (C). Adapted from (7).

the constant region of the IgH gene, while leaving the variable region the same, in a process known as class-switch recombination (CSR).

Table 1.1: Properties of Ig isotypes - Adapted from (7, 8). +++ = Very Strong ; ++ = Strong ; + = Moderate ; - = None

Isotype	Serum (%)	Structure	Half-life (days)	Complement activation	FcR binding	Function
IgA1	13	Monomer, Dimer	6	+	+	Mucosal response
IgA2	2	Monomer, Dimer	6	+	+	Mucosal response
IgG1	50	Monomer	23	++	+	Secondary response
IgG2	17	Monomer	23	+	-	Secondary response
IgG3	5	Monomer	7	+++	+	Secondary response
IgG4	4	Monomer	23	-	-	Regulation
IgM	10	Pentamer	5	+++	-	Initial response
IgD	<0.5	Monomer	3	-	-	B cell development
IgE	<0.01	Monomer	0.5	-	+	Allergy

IgD is the first isotype to be expressed by naive B cells, and although IgD antibodies have no known effector functions, it is thought that membrane-bound IgD is used for the regulation of B cell development (9). IgM is also expressed on naive B cells, and is the first antibody to be produced during an immune response. IgM antibodies therefore often have low affinity for the antigen, but they are able to form pentamers, giving a high binding avidity. During an immune response, there may then be class-switching of the B cells from expressing IgM to expressing

either IgG, IgA or IgE antibodies. IgG is the most common subclass present in the serum, and produced during both antiviral and antibacterial immune responses. The four subclasses have subtly different properties (Table 1.1), and response to protein antigens is generally mediated by IgG1 and IgG3, while the response to polysaccharide antigens is mediated by IgG2 and IgG4 (10). Although IgA is present in the serum (as a monomer), its primarily role is as a mucosal antibody (as a dimer) where it is able to neutralise toxins, and prevent invasion of viruses and bacteria at mucosal surfaces (11). IgE is the least abundant antibody, and produced in response to parasitic worm infections, and during allergic reactions. IgE is able to strongly activate mast cells and granulocytes to stimulate an inflammatory response (8).

1.1.2 B cells in innate immunity

The most basic innate immune mechanisms are the anatomical barriers which prevent initial invasion by pathogens. Epithelial surfaces are not only highly impermeable, but also contain chemical and biological barriers. The respiratory tract contains a bactericidal mucus layer that traps pathogens for subsequent removal by ciliary activity. In the gastrointestinal tract, there is not only a bactericidal environment, but also a community of natural gut micro-organisms, which may outcompete pathogenic organisms. The efficacy of these barriers is aided by the presence of both monoreactive (high-affinity binding to a single target antigen) and polyreactive (lower-affinity binding to a broad range of antigens) mucosal antibodies, which provide protection against a broad range of potential pathogens (12). While the monoreactive antibodies tend to be derived from previous pathogen-specific immune responses, the polyreactive antibodies are thought to be natural antibodies derived from innate-like B cells (discussed below), and recognise commonly found pathogenic targets (1).

Providing a pathogen does evade the first line of defences to infect the body, there is then the potential for recognition of the pathogen by tissue-resident cells of the innate immune system (macrophages and dendritic cells (DCs)). The principle of this recognition is based upon

using a limited number of pattern recognition receptors (PRRs), such as the Toll-Like Receptors (TLRs) to recognise certain conserved pathogen-associated molecular patterns (PAMPs) (13). Activation of these PRRs results in an inflammatory response, and cytokine release, which activates the circulating cells of the innate immune system (monocytes, granulocytes and natural killer cells). Innate immune cells such as DCs may also internalise the pathogen, degrade it, and then present the antigens on their surface in association with a major hisotcompatibility complex (MHC) II. These antigen-presenting cells (APCs), are then involved in the activation of T cells to mediate the adaptive immune response. It is worth noting that B cells can also function as primary APCs, but are only efficient for the antigen to which their BCR is specific (14). As it is the innate immune cells which are responsible for activating the B cells of the adaptive immune system, stronger innate activation will generally result in an increased B cell response - a property which can be exploited during vaccine design by the addition of innate immune adjuvants (15). There is also positive feedback from the activation of B cells by the innate cells, as the antibody produced by the B cells further activates the innate cells by binding to their Fc receptors.

In addition to the cellular mechanisms of innate immunity, there are a number of circulating proteins collectively known as the complement system, which enhance pathogen destruction and clearance once activated. The complement system can be activated both as part of the innate immune response (alternative pathway, and mannose-binding lectin pathway), and the adaptive immune response (classical pathway). In the classical pathway, it is antibody binding to the pathogen that activates the complement system. Experiments with *Streptococcus pneumoniae* have illustrated that a functioning classical pathway is actually essential for complement-mediated phagocytosis (16).

The final role of B cells in the innate immune system is in the production of natural antibodies. The presence of these antibodies was first demonstrated in germ-free mice, who

appear to maintain normal serum IgM concentrations, despite the absence of any specific pathogenic stimulus (17). In addition, some B cells can be activated by binding of their PRRs to PAMPs independently of their BCR specificity, as has been shown with the activation of B cells in response to bacterial cell membrane lipopolysaccharide through recognition by Toll-like receptor (TLR) 4 (18). These natural antibodies are generally of low-affinity, but have high polyspecificty, and are thus effective against a range of pathogens. Natural antibody production is primarily from B1 cells, and marginal zone B cells, indicating that these function as innate-like B cell subsets. The most abundant natural antibody in humans is anti-Gal, which recognises the α-galactosyl epitope, and comprises approximately 1% of the serum antibody of each isotype (19). The α-galactosyl epitope is highly prevalent on the polysaccharides of bacterial cell walls, which is why this antibody may be so abundant (20). Interestingly, the α-galactosyl epitope is also present on the cells of most non-primate mammals, so they do not produce any anti-Gal antibody (21).

1.1.3 B cells in adaptive immunity

Adaptive immunity is comprised of two systems: humoral immunity (mediated by B cells), and cellular immunity (mediated by T cells). Rather than recognising general PAMPs, the cell surface BCRs and T cell receptors (TCR) recognise specific antigens on the pathogen. An individual B or T cell produces a BCR or TCR with a single specificity. In order to facilitate recognition of any possible antigen that may enter the body, each individual cell has the potential to produce a receptor with a different specificity. The result is a vast anticipatory repertoire of B and T cells with different receptor variants, termed the BCR repertoire and TCR repertoire. The human BCR repertoire has the theoretical potential to produce up to 10^{11} variants (22), although experimental measurement suggests that it is actually in the region of 10^6 - 10^7 variants in a single individual (23, 24, 25). A fundamental difference between the BCR and TCR is that while the TCR only recognises antigens presented on APCs in the context of

MHC I or MHC II complexes, the BCR recognises antigens in their unprocessed, native form either in solution or on the cell surface. BCRs and TCRs may therefore recognise different epitopes in their response to the same antigen. TCR epitopes are small linear proteins which bind MHC molecules, whereas BCR epitopes can take a greater diversity of forms. These BCR epitopes may be related to the tertiary, or even quaternary structure of the antigens (26), as well as small linear sections of the protein (27).

Once a specific B cell is activated during an immune response, it proliferates, and may differentiate into antibody-secreting plasma cells (PCs). There are two distinct classes of T cell, which are named according to their expression of either the CD4 or CD8 cell surface marker. Activated CD4 T cells (T helper cells) function as B cell co-activators, and CD8 T cells (T killer cells) mediate cellular immunity by destroying infected cells. Both B and T cells also have the ability to differentiate into memory cells, thus priming the system for a more rapid response upon re-encounter with the same pathogen.

1.1.4 B cell subsets

The contribution of B cells to different aspects of immunity is carried out by different subsets of B cells (Table 1.2). However, it should be noted that the exact role of these subsets in different responses, and fine delineation of further subsets is the subject of ongoing research (1, 28, 29, 30). Antigen-naive human B cells can broadly be split into two groups: B1 and B2 cells.

B1 cells are formed during foetal development, and cannot be renewed, so decrease with age (33, 34). B1 cells are considered to be innate B cells as they are rapidly activated by PAMPs rather than through their BCR, and are the predominant source of natural IgM and IgA secretion (18, 34). Such antibody is polyreactive, and able to recognise repetitive structures, so is important in the initial response to encapsulated bacteria. B1 cells have a similar phenotype

Table 1.2: B cell subsets - Properties, and functions of the major B cell subsets found in humans

	B1	**B2**			**Terminally differentiated**	
		Naive		**Memory**		
		Marginal Zone	**Follicular**		**Short-lived PCs**	**Long-lived PCs**
Development:	Foetal liver	Bone marrow	Bone marrow	Lymphoid organs	Lymphoid organs	Lymphoid organs
Main location:	Peritoneal and pleural cavities	Marginal zone of the spleen	Lymphoid follicles of spleen and lymph nodes	Lymphoid follicles of spleen and lymph nodes	Lymphoid organs	Bone marrow
Activation mechanism:	PAMPs	TI antigens	TD antigens	TI/TD antigens	-	-
Function:	Rapid polyclonal response. No memory or class-switching	Rapid antigen-specific response. Limited memory or class-switching	Antigen-specific response. Generates memory and class-switches	Rapid antigen-experienced response	Pathogen clearance	Maintains serological memory
Immune category:	Innate	Innate/ Adaptive	Adaptive	Adaptive	Adaptive	Adaptive
Surface markers:	CD19+, CD20+, CD27+, CD43+, CD70-	CD19+, CD20+, CD27+, IgM+, IgD+	CD19+, CD20+, CD27-, IgM+, IgD+	CD19+, CD20+, CD27+, IgD-	CD19+, CD20-, CD27+, CD38+, CD138+	CD19-, CD20-, CD27+, CD38+, CD138+

PC; Plasma cell, TI; T-independent, TD; T-dependent
Surface markers derived from (31, 32, 33)

to pre-PCs, which may explain their ability to be rapidly activated to secrete antibody (30), but they do not have the ability to form long-lived PCs or memory cells.

B2 cells are the largest population of B cells, are continually generated in the bone marrow, and circulate through the blood and lymphoid tissues. Follicular origin (FO) and marginal zone (MZ) B cells both arise from this B2 lineage, and enable detection of a range of T-independent and T-dependent antigens (Section 1.2.2). It is only B2 cells which are able to undergo CSR and produce a high-affinity memory response, and it is thus these B cells that are considered the main mediators of adaptive immunity. For B2 cells, following activation there may be formation of memory cells, and PCs. PCs can further be subdivided into both short-lived populations which have high levels of antibody secretion to clear infection, and long-lived populations which have sustained lower levels of antibody secretion and migrate to

the bone marrow to maintain serological memory (35). The population of long-lived PCs in the bone marrow therefore represents the infection history of an individual, and has been shown to persist for over 40 years (32).

The exact nature of MZ B cells is debated; they have the same developmental pathway as the FO B cells, but have more similar functionality to B1 cells. It is likely that they are able to function as both innate-like B1 cells, and adaptive FO B2 cells, and thus blur the boundaries that are conventionally ascribed to these two immune pathways (1). For example, many MZ B cells express polyreactive BCRs, and have high numbers of TLRs, enabling rapid activation by the same repetitive antigens on encapsulated bacteria that activate B1 cells (36). However, unlike B1 cells, the MZ cells are also able to differentiate to produce high-affinity memory cells, thus giving sustained, and more specific protection (37).

1.2 The generation of B cell diversity

An effective adaptive immune response relies on the generation of a vast BCR repertoire that is able to recognise any potential pathogen that may infect the body. The first mechanism for generating diversity occurs during B cell development with the rearrangement of the V, D and J germline gene segments to form the complete Ig gene. The second mechanism occurs following antigen encounter through somatic mutation of the BCR followed by selection for antigen binding in the germinal centre (GC).

1.2.1 B cell development

VDJ recombination

B cells develop from hematopoietic stem cells in the bone marrow, and once mature, migrate through the peripheral blood to the lymphoid organs. In immature B cells, there is not a functional Ig gene, but instead the Ig heavy and light chain loci are comprised of multiple different V, D (heavy chain only) and J gene segments. During B cell development, there

is site-specific recombination by recombination activating genes (RAG) 1 and 2, to delete the intervening DNA between one of each of the gene segments to leave the functional Ig gene (Figure 1.2) (38, 39). RAG1 and RAG2 recognise recombination signal sequences (RSS) flanking the gene segments to catalyse the reaction (40). RSSs consist of a conserved heptamer and nonamer sequence split by either a 12 or 23 base pair (bp) spacer. Recombination can only occur between a pair of RSSs with dissimilar spacers, which prevents recombination of incorrect gene segments to each other (39). On the IgH locus, the V and J gene segments cannot directly join, as they are both flanked by the 23 bp spacer. The D segments have the 12 bp spacer, thus allowing the V-D, and D-J recombination. On the κ IgL locus, the V segments are flanked by the 23 bp spacer, and the J segments by the 12 bp spacer, while the converse is true for the λ IgL locus, thus allowing direct V-J recombination.

The IgH locus is located on chromosome 14, and consists of approximately 45 functional V segments, $\sim$27 functional D segments, and 6 functional J segments in addition to the constant region gene segments for the nine different isotype subclasses (41). The IgL κ and λ loci are on chromosome 2 (containing $\sim$76 V segments and $\sim$5 J segments) and 22 (containing $\sim$74 V segments and $\sim$7 J segments) respectively (42, 43). This segmental organisation means that just for IgH, there is the potential to produce 7,290 unique VDJ recombinants. Factoring in IgL diversity, this then rises to more than 10^9 unique variants. It should be noted, however, that the exact number of functional gene segments differs between individuals due to heterozygosity at the Ig gene loci, and this will cause minor differences in repertoire diversity between individuals (44).

The process of VDJ recombination occurs in an ordered and regulated fashion. IgH recombination occurs before IgL, and starts in the early pro-B cells by the joining of the D and J segments on both chromosomes. V to DJ joining then occurs on a single chromosome, and only if this rearrangement is subsequently found to be non-productive, does it proceed

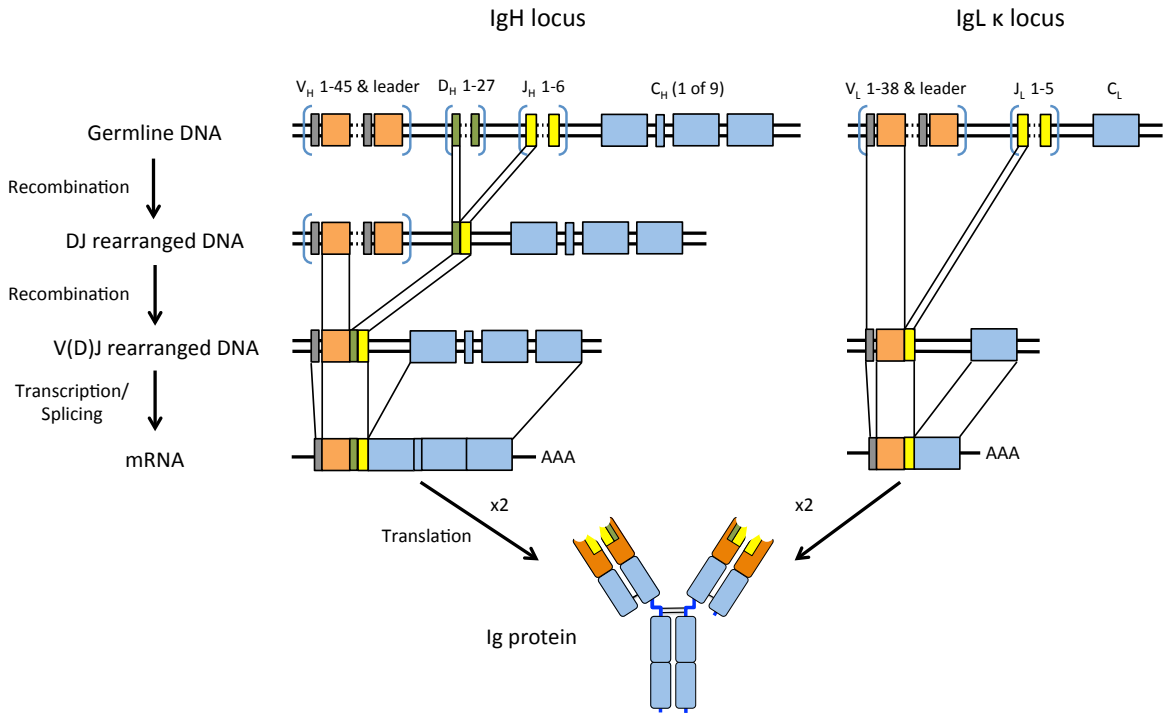

Figure 1.2: Stages of VDJ recombination - There is an ordered process of IgH VDJ recombination and IgL VJ recombination to form a functional Ig gene. One each of multiple potential V, D and J genes can be joined, giving vast potential for combinatorial diversity. Each V segment is associated with a leader sequence (in grey), which can direct the protein into the cells secretary pathways. The constant domain illustrated here consists of four separate exons to the V domain exon. These are all spliced together during RNA processing. The IgL locus illustrated here is the κ locus. Adapted from (7).

on the second chromosome to form a pre-B cell. Pre-B cells are selected for functional IgH by expression with a surrogate light chain (45). If there is successful activation of this pre-B cell receptor, IgL rearrangement proceeds; this occurs first at the κ locus, and if unsuccessful moves to the λ locus, giving a 65:35% κ:λ ratio in mature B cells in humans (46). Once the functional BCR is formed to give immature B cells, there is then a negative selection process to remove self-reactive B cells before they leave the bone marrow as mature B cells (47). B cells that bind to multivalent self antigens on the surface of cells in the bone marrow undergo apoptosis, and B cells that bind to soluble monovalent antigens do survive and leave the bone marrow, but become anergic, and unable to respond to antigen (47). The proportion of B cells removed due to self-reactivity is not well documented, but it is likely to be large, as the size of the BCR repertoire measured using phage display libraries (where there is no removal of

self-reactive or non-productive rearrangements) is approximately 10,000 times greater than the measurements based on human sampling (22, 23, 24, 25).

Junctional diversification

As well as the combinatorial diversity introduced during VDJ recombination, further diversity is introduced to the variable regions at the junctions of the V, D and J segments (39). During the recombination process, short palindromic sequences (P-nucleotides) are added from asymmetric opening of DNA hairpins which are present at the junctions to aid pairing (48). Additionally, up to 20 random single nucleotides (N-nucleotides) may be added at the junctions by the action of the terminal dideoxynucleotidyl transferase (TdT) enzyme (49). Further junctional diversity then arises from the deletion of some of these nucleotides by exonuclease activity (49, 50). Although this nucleotide addition and deletion can vastly increase diversity, approximately two thirds of Ig genes have non-productive out of frame rearrangements, so it is costly to the cells (51).

Receptor editing and V_H replacement

The VDJ recombination process has the potential to produce an enormous diversity of Ig genes, but it comes at a cost of producing a large number of non-productive rearrangements, or auto-reactive rearrangements which may subsequently lead to B cell deletion. In order to retain some of these B cells, the RAG-mediated recombination systems remain active in early B cells, and may confer improved receptor functionality through either receptor editing or V_H replacement (52, 53). In receptor editing, BCR signalling in auto-reactive B cells initiates RAG activity, and continued IgL recombination. This causes a new 5' V segment to recombine with a new 3' J segment to excise the existing VJ join and create a new one (52).

Receptor editing of the IgH gene cannot occur, as the 5' V segments and 3' J segments both have a 23-bp RSS spacer, so cannot recombine. However, IgH V segments were found

to contain cryptic RSS sites embedded within the 3' end of the segment (54). This cryptic RSS can recombine with a 23-bp RSS from a 5' V segment, thus allowing replacement of the V segment with a new one. A small portion of the initial V segment remains that is 3' of the cryptic RSS, allowing the occurence of V_H replacement to be monitored, and showing that it is an important mechanism in repertoire formation and diversification (55).

1.2.2 B cell activation

B cells require multiple signals in order to become fully activated. Activation induces B cell expansion, and differentiation into effector memory and PCs, thus leading to changes in the B cell repertoire. In addition, further diversity can be introduced during this process by high levels of somatic hypermutation of the Ig genes followed by selective survival of those B cells with the greatest affinity for the antigen (affinity maturation), and also changes in the isotype class of the Ig gene (CSR).

Initial B cell activation occurs when an antigen specifically binds to the BCR. Secondary activation can then be divided into T cell independent (TI) and T cell dependent (TD) scenarios based on the requirement for T cell involvement (Figure 1.3). There are important immunological distinctions between the outcomes of these pathways. The majority of antigens activate B cells in a TD manner, resulting in high-affinity class-switched responses, and the induction of memory. TI activation typically results in lower-affinity IgM responses, and does not induce long-term memory, and is thus often associated with the innate function of B cells (Section 1.1.2).

T-independent B cell activation

TI B cell activation can further be split into TI-1 or TI-2 pathways depending on the source of the secondary activation step. In the TI-1 pathway, secondary B cell activation comes from recognition of polyclonal B cell stimulants by cell surface receptors such as TLRs (Table 1.4).

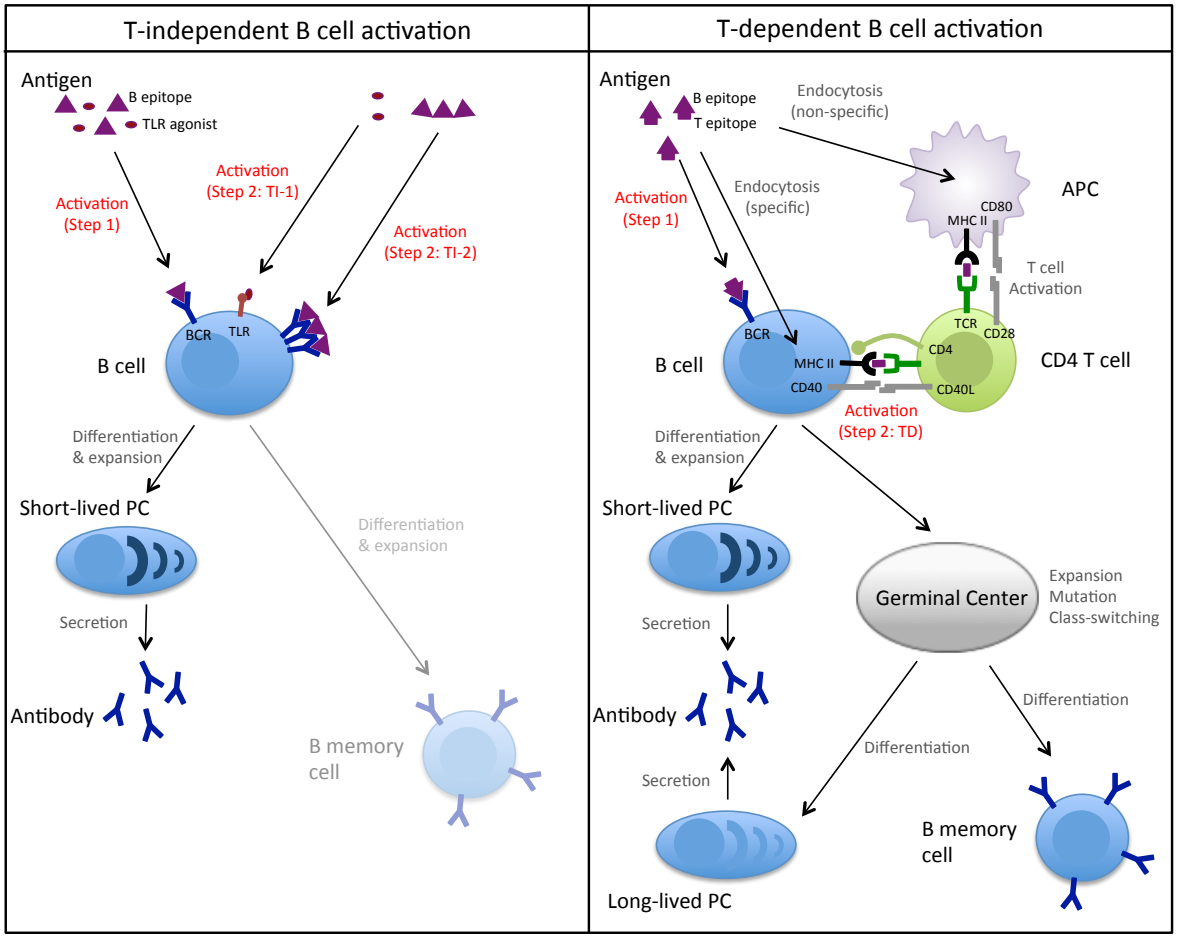

Figure 1.3: TI and TD B cell activation - In both TI and TD activation, B cells are initially activated by BCR binding to antigen. In TI activation, the secondary activation step may arise from activation of TLRs (TI-1), or by extensive cross-linking of the BCR (TI-2). Following TI activation, there is terminal differentiation into short-lived PCs, although evidence is starting to emerge for germinal centre independent differentiation into memory cells as well. In TD activation, the secondary activation step comes from interaction of the B cell with a CD4 T cell. This results in both the production of terminally differentiated short-lived PCs, and germinal centre formation, and subsequent long-lived PC, and memory cell generation.

Such antigens are generally from microbial origin, and include some lipopolysaccharides (LPS), microbial CpG DNA, viral RNA, and polymeric proteins. For example, LPS is able to stimulate B cell activation via TLR4 activation (56). The antigens responsible for the primary and secondary activation steps may be independent - thus, a B cell may be initially activated in an antigen-specific manner by BCR signalling, then be further activated in a non-specific manner by TLR activation. In the TI-2 pathway, secondary activation comes instead from repeated BCR activation and cross-linking by antigens with multiple repeating epitopes (57).

Clinically important examples of TI-2 antigens include the bacterial capsule polysaccharides of *Streptococcus pneumoniae, Haemophilus influenzae* type b, and *Neisseria meningitidis*, which have highly repetitive structures to achieve efficient B cell activation.

Table 1.4: Characteristics of TI and TD B cell activation - Adapted from (58)

Activation	Antigens	Responding B cells	Secondary activation	GC formation	Memory response
TI-1	Polysaccharides	B1, MZ	TLR signalling	No	Limited
TI-2	LPS, CpG, lipopolysaccharides, RNA	B1, MZ	BCR crosslinking	No	Limited
TD	Protein	FO	CD4 T cells	Yes	Yes

TI activation of naive cells arises more rapidly than TD responses, and generally utilizes the more innate-like B1 and MZ B cell subsets. Following TI activation, there is terminal differentiation into short-lived PCs. It has long been thought that such responses do not result in the generation of high-affinity class-switched memory cells and long-lived PCs, but evidence is starting to emerge for the presence of TI memory responses that are phenotypically distinct from TD responses (59). Adoptive transfer of B1 B cells from donor mice who have cleared *B. hermsii* infection conferred protection in naive $Rag^{-/-}$ mice, indicating that at least in this system, even B1 cells may be able to generate some memory (60). The TI memory response, and lifespan of the memory cells appears to relate to the degree of B cell activation, with joint TI-1 and TI-2 activation likely to be necessary (61).

T-dependent B cell activation

In TD activation, the secondary activation step comes from the interaction of a B cell with a CD4 T helper cell that has been activated by the same antigen. CD4 T cells are initially activated by interaction of their TCR with antigen presented on the surface of an APC in association with an MHC class II molecule. B cells are also able to internalise the antigen, and function as APCs, and it is the interaction of the TCR and CD4 with the antigen-MHC

II complex on the B cell, in conjunction with the CD40/CD40 ligand co-receptors that results in TD B cell activation (62). TD antigens are usually proteins, as they must contain both T and B cell epitopes; these epitopes need not be identical, but they must be physically linked.

Following B cell activation, there is rapid proliferation of the cells, with some B cells differentiating in the extrafollicular foci directly into PCs to produce IgM antibody (63). These PCs are generally short-lived with a half-life of approximately three days (35). Alternatively, activated B cells may migrate down chemokine gradients into the follicular area of the secondary lymphoid organs, and initiate a GC reaction. Initiation of the GC is further aided by a subset of CD4 T helper cells known as follicular helper T cells (T_{FH}). T_{FH} express the CXCR5 homing marker allowing them to migrate to the B cell follicles, and have increased levels of CD40 ligand expression, making them highly effective at B cell activation. The GCs then provide a specialised microenvironment to stimulate affinity maturation, CSR, and the production of long-lived PCs and memory cells (Figure 1.4). In antigen-naive hosts, GCs take several days to become established, but in primed hosts, they may form within hours (64). GCs are typically present for 3-4 weeks following antigen exposure, but in the presence of persisting antigen, they may remain active for many months after formation (65).

Within the GC, the B cells produce activation-induced cytidine deaminase (AID), which causes somatic hypermutation. AID deaminates cytosine (C) to uracil (U) in the genomic DNA. During cell division, U then pairs with adenine (A) rather than guanine (G). This leads to C to thymine (T), or G to A single nucleotide conversions. Hypermutation is limited to only the IgH and IgL genes of B cells (67). The mutated B cells are then subject to a selection process based on antigen-binding affinity (affinity maturation). This selection is mediated by the ability of the B cell to uptake antigen from the follicular dendritic cells (FDC), and present it to the T_{FH} cells within the GC to receive further activation signals (68). Since the FWRs provide the structural backbone of the BCRs, fewer mutations are selected for in the FWRs

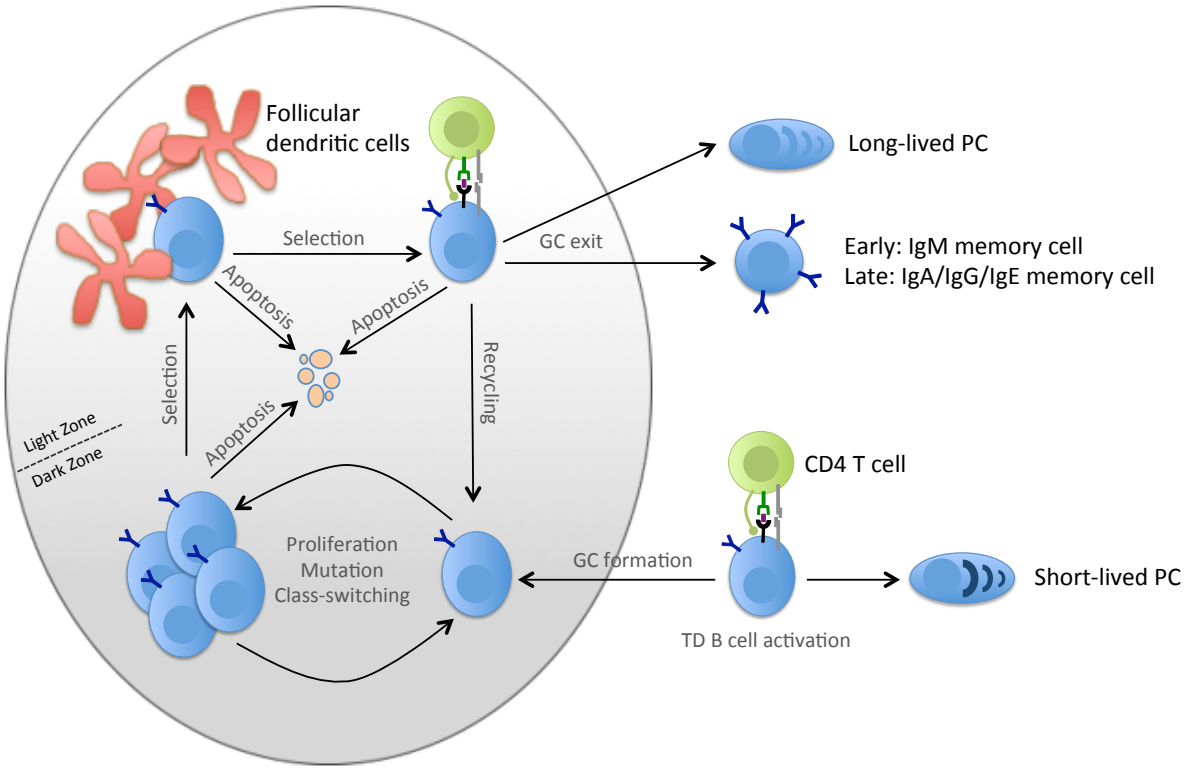

Figure 1.4: The germinal center reaction - Upon TD B cell activation, some B cells directly differentiate into short-lived PCs, and the remainder form GCs. Within the GC, B cells proliferate, and produce AID. AID stimulates both hypermutation, and class-switching. Mutated B cells are then selected for antigen-binding affinity by their ability to uptake antigen from FDCs, and present it to T cells. Positively selected B cells differentiate into long-lived PCs, or memory cells and leave the GC. Adapted from (66)

compared to the CDRs where the antigen-binding residues are located. It is estimated that the level of hypermutation in GC emigrant B cells is >4% per nt, or >10% per amino acid (AA), in the CDRs, thus giving significant potential for increasing B cell diversity (69). Selected B cells have three possible fates: some remain in the GC and undergo further cycles of proliferation, mutation and selection, while the remainder are exported from the GC and differentiate into either long-lived PCs or memory B cells (70). The signals controlling differentiation into either a PC or memory cell are not well understood, but could also be related to BCR affinity. It has been observed that PCs emigrating from the GC have higher affinity BCRs than the emigrating memory cells, indicating that increased CD4 T cell interaction may be required for PC differentiation (71).

As well as being involved in affinity maturation, it is also the activity of AID which stim-

ulates CSR (Figure 1.5). Following VDJ recombination of IgH, both IgM and IgD isotypes can be expressed by alternative splicing. CSR involves deletion of the IgM and IgD constant region DNA segments, thus allowing expression of the downstream constant regions instead (IgG/IgA/IgE) (72). On the 5' side of each constant region segment is a tandem repeat unit known as a switch region. AID causes mutation in these switch regions, followed by the introduction of double-strand breaks by mismatch repair proteins. These double-strand breaks then facilitate end-joining recombination between two switch regions, and the excision of the intervening DNA (73).

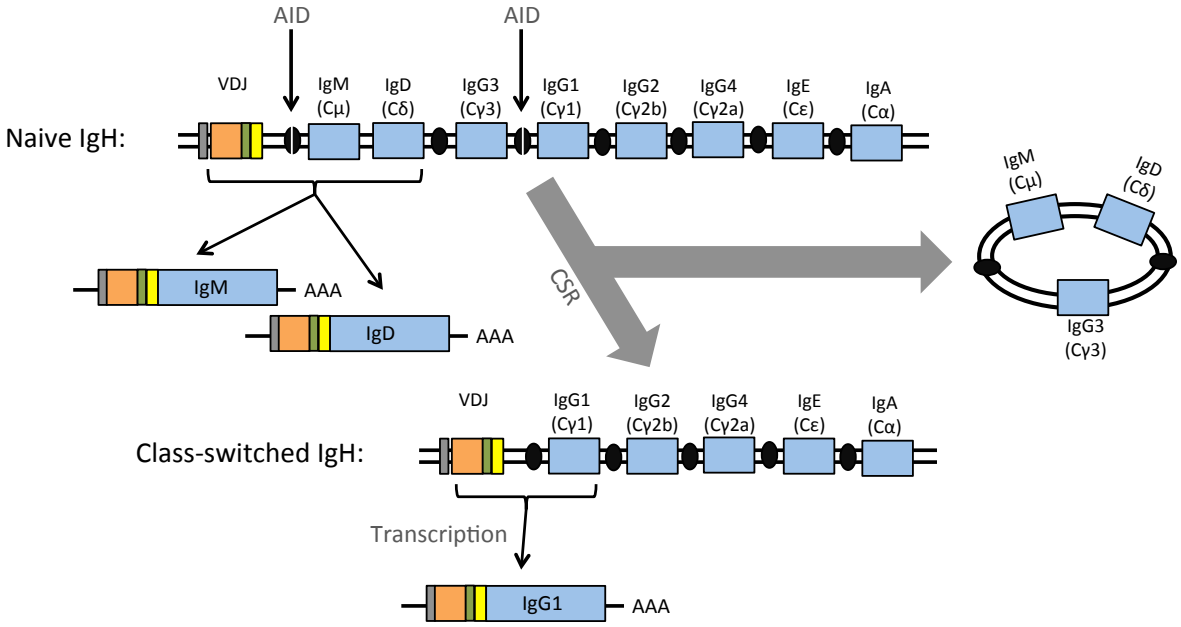

Figure 1.5: Mechanism of class-switch recombination - In naive B cells, IgM and IgD exons are both transcribed by alternate splicing. Following B cell activation, AID introduces mutations into the switch regions between the IgH constant region exons causing the formation of double-strand breaks. Rejoining of the breaks causes excision of intervening DNA, allowing transcription of a downstream constant region exon (IgG1 in the example shown). Adapted from (73)

Memory B cell activation

So, B cell diversity is introduced both during B cell development, and during B cell activation. While the naive B cell repertoire must be diverse enough to recognise any potential invading pathogen, this recognition may be of low affinity. The additional diversification that occurs during B cell activation is therefore required to give high affinity recognition. If an antigen has

already been encountered, resulting in the production of memory B cells, then these memory B cells represent a pre-diversified pool of cells able to respond upon subsequent encounter with the same antigen. The memory B cells may directly differentiate into PCs, and can also generate new GC reactions. This leads to a more rapid, and higher-affinity response upon secondary encounter with the same antigen, thus giving improved protection. This improved protection upon antigen re-encounter is one feature of B cells which can be exploited by vaccination to confer protection against disease.

1.3 Exploiting B cells through vaccination

Since the demonstration of the first vaccine by Edward Jenner in 1796, vaccination has grown to become one of the most important public health interventions worldwide. Where effective vaccines exist, they have been able to greatly reduce disease, preventing millions of deaths each year (74). With the exception of the Bacille Calmette-Guérin vaccine, the primary correlate of vaccine-induced protection for almost all licensed vaccines is the presence of vaccine antigen-specific functional antibody in the blood. Therefore a better understanding of B cell biology and its relationship to antibody production is an important area for vaccine development.

1.3.1 Vaccine mode of action

The principle of vaccination is to use an attenuated or killed pathogen, or a subunit of the pathogen to induce an adaptive immune response without causing disease. This adaptive immune response will then protect against disease if there is future exposure to that pathogen (Figure 1.6). Long-term protection arises from the production of long-lived PCs and memory B cells specific to the pathogen during the immune response, and the subsequent maintenance of pathogen-specific antibody in the serum and mucosal secretions (75). Short-lived PCs produce large amounts of antibody rapidly after antigen exposure, while long-lived PCs migrate to the bone marrow where they provide long-term secretion of smaller quantities of antibody

(76). Serum antibody levels persist for variable periods of time following primary vaccination, ranging from more than >50 years in the case of smallpox (77), to just a few years in the case of Hepatitis B (HepB) (78). The reasons behind these differences are not well understood, but mean that for vaccines with short-lived protection, such as HepB, repeat doses are required to replenish serum antibody levels.

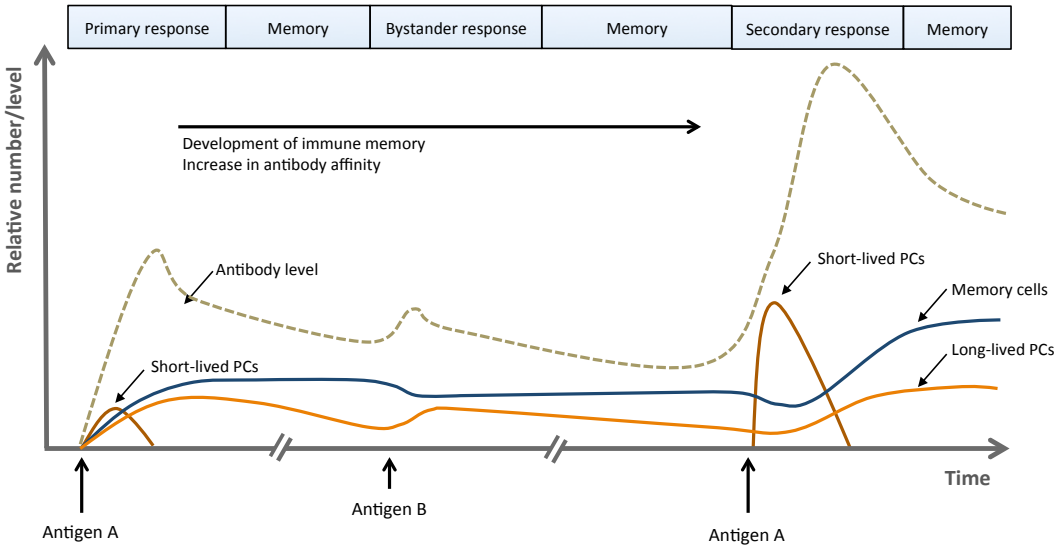

Figure 1.6: Primary and secondary response to an antigen - Primary antigen exposure to antigen 'A' activates antigen-specific B cells to produce short-lived and long-lived PCs, and memory cells. Antibody produced by the PCs gives protective immunity. Subsequent exposure to a different antigen, called here antigen 'B', causes bystander re-activation of small number of memory cells specific to antigen 'A' , which can replenish the long-lived PC population. Secondary re-exposure to antigen 'A' causes rapid reactivation of memory B cells and further increase of the long-lived PC and memory cell populations. As the secondary response utilises a greater pool of antigen-specific cells, which have already gone through a degree of affinity maturation, it is larger and more effective than the primary response.

As well as long-lived PCs, protection is given by the production of memory cells. Memory cells circulate through the antigen-draining sites of secondary lymphoid tissues, and are able to become rapidly activated by the presence of the antigen to both produce short-lived PCs, and form new GCs (64). Furthermore, memory cells will have already been through affinity maturation and may be class-switched, so will produce antibodies of higher affinity for the antigen than are produced during the primary response (79). In the case of tetanus toxoid (TT) vaccination, the limits of affinity maturation after repeat vaccination have been studied. The limit was not reached after a primary immunisation course of four vaccine doses, but was

reached after two subsequent booster doses. After this point, repeated vaccination increased just the number of cells rather than their affinity (80). In fact, even when serum antibody levels have declined to non-protective levels, protection may still be maintained by the presence of these memory cells (78), but this is highly dependent on the type of infection. Where there is a long incubation period, such as in HepB infection, there may be time for the memory cells to become activated and produce antibody in time to clear the infection. Where invasion is more rapid, such as in the case of *Haemophilus influenzae* type b (Hib) infection, persistence of serum antibody is required for protection (81).

As well as antigen-specific memory B cell activation, there is also evidence for small numbers of memory B cells being polyclonally stimulated during the response to non-specific antigens (bystander stimulation), allowing maintenance of serological memory in the absence of re-stimulation with the same antigen (82) (see Figure 1.6 response to antigen 'B').

1.3.2 Measuring vaccine immunogenicity

The development of effective vaccines against specific pathogens relies on understanding the B cell response to the vaccine antigen. B cell responses to vaccination are conventionally assessed by investigating the antigen-specific antibody found in the peripheral blood serum after vaccination; antibody levels have then been used as a correlate of immunity (83). Antigen-specific antibody levels generated in response to a vaccine are generally assessed by an enzyme-linked immunosorbant assay (ELISA). It is becoming apparent though, that for some antigens, or in some populations, the quantity of antibody produced cannot reliably be used to determine risk of infection (84), and immune protection has been observed in the absence of detectable antibody levels in HepB vaccinated individuals (78). Looking at the functional ability of the antibody by measuring their neutralization or opsonophagocytic capacity can give a more informative idea of protective capacity of the antibody (85), but studying antibody alone can not give insight into the underlying biology of the protective response to vaccination.

Understanding what determines the quantity, persistence and quality of antibody production requires investigation at the cellular level of the B cell subsets responsible for producing the antibody, and how they are stimulated in response to vaccination. Persistence of antibody in the serum is determined by the combination of short-lived PCs, long-lived PCs, as well as memory cells which are used to renew the PC population (82). Quality and persistence of antibody may be dependent on the B cell subsets that are recruited in the immune response, and whether this is in a TD or TI manner (86, 87). The overall vaccine response is then also influenced by multiple host genetic and environmental factors, including previous antigen exposures, which leads to great inter-individual variability (88). Much of our current knowledge of the cellular response to vaccines comes from studying animal models, where bone marrow and lymphoid tissues can easily be obtained to directly study GCs and resident populations of PCs and memory B cells. In humans, a less invasive approach must be used, so circulating B cells in peripheral blood are studied, and immunological mechanisms inferred from observations of changes in the frequency and nature of different B cell populations.

Enzyme-linked immunospot (ELISpot) assays have proved successful in detecting the presence of antigen-specific B cells in the peripheral blood. Cell culture methods can also be used in conjunction with ELISpot to differentiate between PCs, and memory B cells (89). These techniques have been used to define the kinetics of the antigen-specific memory cell and PC response to vaccination, showing that there is a distinct peak of PCs detected in the peripheral blood around day 10 following primary immunisation, and day 7 following booster immunisation (90, 91). Memory cells also start to be found in the peripheral blood at the same time as the PCs, but continue to increase in number for up to a month following vaccination, and persist for longer (90, 91). These studies suggest that PC frequency following immunisation relates to the concurrent increase in serum antibody levels, but does not predict antibody persistence. In some cases, it has been possible to relate frequency of memory cells in the peripheral blood

after vaccination to the persistent serum antibody response, but this appears to be dependent on the specific vaccinating antigen (92).

Flow cytometry can be used for finer delineation of the B cell subsets activated in response to vaccination, and can give some additional immunological insight. For example, flow cytometry has been used to study the cellular response to both pneumococcal conjugate and plain polysaccharide vaccines in humans (93). While an immediate serum antibody response is seen after both vaccinations, the plain polysaccharide vaccine caused a depletion of the antigen-specific memory B cell pool, and B1 B cell pool (93), which could be a cause of the hyporesponsiveness seen with repeat plain polysaccharide vaccination (94), and indicative of a TI response. Conversely, the conjugate vaccine caused an increase in the antigen-specific memory B cell pool, suggestive of a TD response that could give more persistent protection (93).

Studying peripheral blood gene expression can also give high level insight into both the B cell subsets used, and other immune pathways employed in the response to certain vaccines. For example, measuring gene expression patterns following influenza vaccination revealed a common but transient signature associated with PCs produced by the vaccine, which related to robust antibody responses (95). A meta-analysis of gene expression studies using five different vaccines has further revealed that there appear to be distinct B cell signatures associated with polysaccharide vaccines, conjugate vaccines, and viral vaccines, some of which can be related to antibody response (96).

While the cellular studies described have given great insight into some of the underlying mechanisms of effective B cell responses to certain vaccines, there are many aspects of the B cell response to vaccination that are not easily answered using these techniques. For example, it is not possible to study the diversity of the antibody response to specific antigens or epitopes, and see how this impacts immunogenicity. Also, while global changes in the frequencies of

cell populations can be determined, it is not possible to directly examine the specific B cells which are activated, and how they evolve over time, undergoing mutation, class-switching, and differentiation to form different B cell subsets. It has also proved challenging to definitively relate the cellular response to the antibody response, as it appears that not all B cells activated by vaccines contribute to the serum antibody repertoire (97).

1.4 Analysing the B cell repertoire

A deeper understanding of the B cell response can potentially be obtained by studying the BCR gene sequences of the B cells underlying the production of protective antibody. Knowing the sequences allows responses to be investigated at the individual B cell level rather than just at the population level, and allows tracking of the B cell clones over time, and between different B cell subsets. Furthermore, having the actual sequence allows monoclonal antibody generation, and fine characterisation of the individual antibodies. It is only in recent years that advances in next-generation sequencing (NGS) have allowed in-depth measurement of this vast system at a level that reflects its great diversity. In this section, the advances in the technology used for studying the B cell repertoire are discussed, and the biological insight this has given is reviewed. In addition, the use of both vaccines as a tool with which to investigate the B cell repertoire, and the use of B cell repertoire analysis as a tool for vaccinology are discussed.

1.4.1 Methods for studying the B cell repertoire

Low-resolution methods

The earliest studies of the B cell repertoire used isoelectric focusing of antibodies on polyacrylamide gels (98), resolving the antibodies into patterns of discrete bands based on their isoelectric pH values. CDR3 size spectratyping, a polymerase chain reaction (PCR) and electrophoresis-based method for determining CDR3 diversity based on nucleotide length distribution, could also give some insight into the B cell repertoire diversity (99). The advent of

Sanger sequencing allowed the exact nucleotide sequences coding for specific BCRs to be determined, albeit in small numbers (100). Lymphocytes can be isolated after vaccination, followed by production of immortalized cell lines. Rearranged BCR DNA can then be amplified from these cell lines, and sequenced (101, 102). Additionally, cloning amplified DNA into expression vectors allows functional characterization of sequences for antigen-specificity (103). Combining this technique with fluorescence-activated single-cell sorting (FACS) has also allowed more precise definition of the B cell subsets being studied.

Such methods have been used to investigate small numbers of monoclonal antibodies (mAbs) generated against a variety of antigens, and although limited by cell numbers, we now know some antigen-specific sequences used in response to vaccines against influenza (104, 105), tetanus (80, 106), *Haemophilus influenzae* type b (Hib) (101, 102, 107) and HepB (108, 109, 110, 111) among others (112). These studies have shown that the B cell response to an antigen with a simple biochemical structure, such as Hib polysaccharide, appears markedly oligoclonal, with elevated V_H3-23 gene usage, and a conserved 'GYGMD' CDR3 AA motif dominating the repertoire in different individuals (101, 102, 107, 113, 114). Repertoire diversity is also restricted for protein antigens (e.g. TT, influenza haemagglutinin (HA)), but the expanded clones differ more between individuals than for Hib polysaccharide (80, 104, 106). Even in a single individual, there appears to be little similarity in the response after repeated TT vaccination, with only one third of the clones sequenced being shared between vaccination events (80).

High-throughput sequencing

The use of Sanger sequencing for characterization of single BCR genes from small numbers of cells is highly robust, as there is a high signal-to-noise ratio, low error rates, and a long read length. For characterizing whole repertoires of BCR genes from large numbers of cells, Sanger sequencing is too labour intensive, but the advent of NGS enabled the simultaneous

sequencing of BCR genes from millions of cells, thus giving insight into a much larger sample of the repertoire (Figure 1.7). However, NGS platforms have higher error rates and shorter read lengths compared to Sanger sequencing, which must be taken into account during experimental design. There are a number of different NGS platforms currently on the market which have been used for BCR sequencing, each utilizing different template amplification strategies, sequencing chemistries, and detection methods (115). This leads to different types and degrees of sequence error (indels and substitutions) in the output data from each platform as well as different limitations such as read length, depth of sequencing (number of sequences that can be sequenced simultaneously), and cost (Table 1.6).

Table 1.6: Comparison of NGS platforms - Error information derived from Bolotin *et al.* (116), and costs for 454, Illumina and Ion Torrent platforms derived from Loman *et al.* (117). Other information obtained from manufacturers websites, accessed on April 2015.

Platform	Mechanism	Error rate/type	Cost/Mb	Length	Number
454	Pyrosequencing	1.4% of reads homopolymer associated indels	~$31	500 bp	10^5 - 10^6
Illumina (MiSeq)	Dye-terminator sequencing	3.2% of reads random substitutions	~$0.5	2 x 300 bp	$>10^7$
Ion Torrent (314 chip)	Semiconductor sequencing	1.2% of reads homopolymer associated indels	~$23	400 bp	10^5 - 10^6
Nanopore (MinIon)	Nanopore sequencing	5-30% of reads indels and substitutions	~$0.004	Up to 10 kbp	$\sim 10^6$

In addition to the sequencing method, different templates (gDNA (23, 118), vs. mRNA (25, 119)), and PCR amplification strategies can be used. When analysing gDNA, sequence abundance should correspond to cell abundance, however, this is not the case when analysing mRNA, because different B cell subsets have different Ig expression levels (120). An advantage of working with mRNA is that isotype information can be obtained, which is not possible with gDNA due to the presence of an intron between the constant and variable regions (41). To generate sufficient DNA for sequencing, BCR-specific PCR amplification, or cDNA synthesis followed by PCR amplification in the case of an RNA template, must be conducted. Either a multiplex PCR method, or 5' rapid amplification of cDNA ends (5' RACE) may be used, and

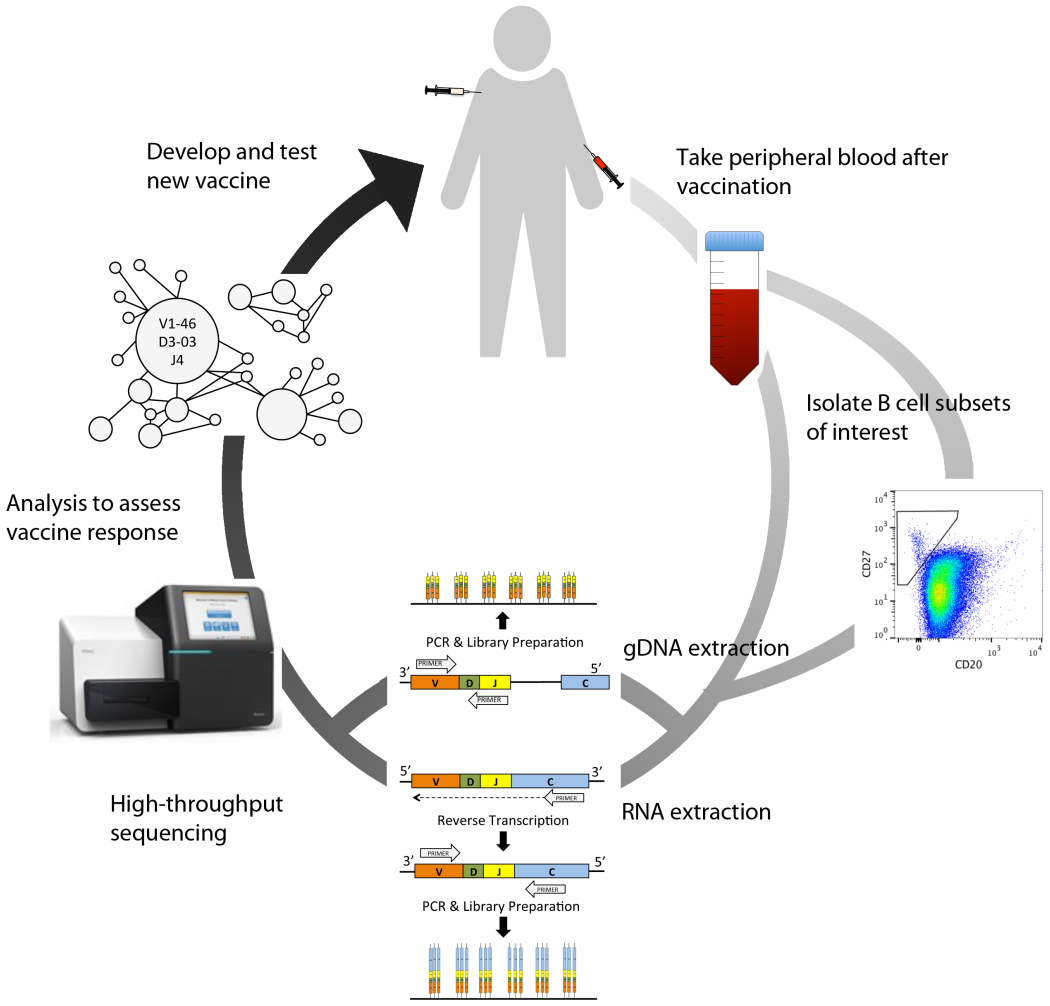

Figure 1.7: Overview of BCR repertoire sequencing - For most BCR repertoire studies in humans, peripheral blood is used as the source of B cells. In certain instances, magnetic or fluorescence cell sorting can be used to isolate specific B cell subsets of interest. Either gDNA, or mRNA are then then extracted from the cells for use as a template in PCR and library preparation. NGS is then conducted, and a variety of bioinformatic methods used to analyse the data.

numerous primer sets have been published (25, 121, 122).

Once sequence data have been obtained, there are a number of processing steps that must be carried out before the BCR repertoire can be accurately reconstructed, and specific analyses carried out. These can broadly be split into sequence pre-processing, annotation, and clonotyping.

Pre-processing

Regardless of the specific methods used, both PCR and sequencing introduce error into the

datasets, and PCR may also introduce bias due to preferential amplification of certain templates. These technical errors in the sequence data must therefore be addressed during preprocessing of the sequence data. As BCR sequences undergo somatic hypermutation, it is not possible to simply align the sequences to germline and define any differences as erroneous, so more sophisticated computational or experimental error correction methods must be used (Figure 1.8).

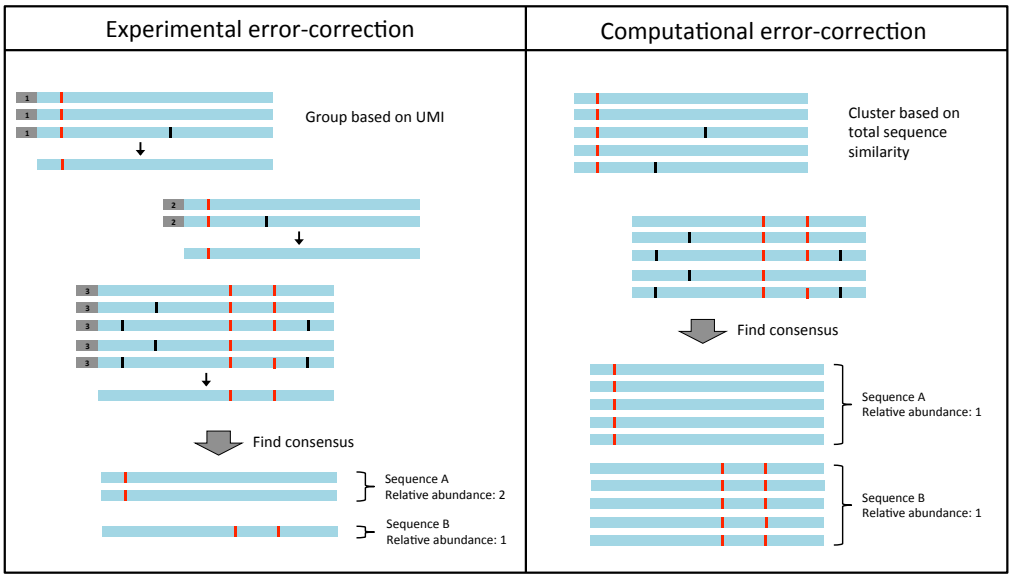

Figure 1.8: Approaches for error-correction in BCR repertoire data - For experimental error correction, UMIs are introduced during reverse transcription, and then consensus sequences generated from reads with the same UMI. For computational error correction, reads are clustered based on sequence similarity, and clusters represented by the concensus sequence. Grey bars represent UMI's, red bars represent deviations from germline due to somatic hypermutation, and black bars represent deviations from germline due to sequencing or PCR errors.

For computational error correction, likely erroneous sequences can first be removed from the dataset using a sequence quality based filtering approach (typically based on Phred score) (123). Related sequences which are assumed to have arisen from the same starting RNA molecule, but differ due to PCR or sequencing errors are then clustered together based on an empirically determined similarity threshold, and collapsed to give the consensus (124). Experimental error correction can be carried out by the use of unique molecular identifier (UMI) tags. These are highly diverse nucleotide tags which are appended to the sequence during reverse transcription, or second strand cDNA synthesis (25). Consensus sequences can

then be generated from amplicons with the same UMI to remove both error, and amplification bias (125). While use of UMI's does reduce error, high sequencing depth is required to ensure there are multiple reads for each UMI, and it is also the case that errors can be incorporated into the UMIs themselves, thus leading to erroneous sequences.

Annotation

Annotation of sequences involves aligning them to a germline database; this alignment involves the identification of the V, D and J gene segment precursors to the sequence as well as the boundaries between the segments, and the location of the CDRs. Such alignment is necessary to determine where somatic hypermutation has occurred in the sequence. There have been many tools developed for VDJ sequence alignment (126, 127, 128, 129, 130), with trade-offs between speed, accuracy, flexibility in VDJ germline database used, and ease of use. Inaccuracies in the annotation can occur if sequences are highly mutated making it difficult to determine their germline precursor, and also from an incompleteness of the germline databases used (44, 131).

Clonotyping

Once individual sequences have been annotated, the population structure of the total repertoire can be determined. This typically involves clustering of related sequences into clonal groupings, where each clonal grouping contains sequences that are derived from the same common ancestor (132, 133). Typically, sequences are clustered based on V, D and J segment use and CDR3 sequence identity. Varying sequence similarity thresholds have been used for clustering, ranging from 80-100% homology (133). Once clonal groupings have been determined, lineage trees can be constructed from the sequences, to determine the ancestral relationships between them (134).

Specific analyses

Once the BCR repertoire data have been processed, specific analyses can be conducted. Immunological status can be measured by assessing features such as the presence of clonal expan-

sions, and total repertoire diversity (135). Specific repertoire features such as average CDR3 length, biases in V, D and J segment use, or mutation levels can be further used to compare repertoires from different cohorts, over time, or in different cell or tissue samples. Analysing patterns of somatic hypermutation can be used to indicate selective pressures on certain clones, and be related to GC activity (136). If multiple samples are taken from the same individual either from different compartments, or over time, the movement of clones can be tracked between the different samples (124, 137). Analysis of the overlap in the repertoire between different individuals (called the convergent, or public repertoire) can also be conducted, to asses potential convergent B cell evolution in response to common antigen stimuli. Overlap can be measured based on specific shared sequences between repertoires, or by clustering together data from multiple samples, and analysing overlap of cluster members (133).

1.4.2 Insights from studying the B cell repertoire

Despite the relative immaturity of high-throughput BCR repertoire sequencing methods, analysis of repertoire data has still given significant novel insights into the structure of the repertoire, and how this responds to antigen stimulation.

Structure of the human B cell repertoire

The first in-depth BCR repertoire study was conducted in zebrafish; this is a useful model system due to its small number of antibody-producing B cells (∼300,000), which is approximately five orders of magnitude lower than in humans (44). This study gave the first insights into the entire BCR repertoire of an organism, showing that zebrafish use only 50-86% of all possible VDJ combinations, that there is a bias in which VDJ segments are used, and that this bias is similar between individual zebrafish. It is not possible to sequence every B cell in humans, so representative samples (generally derived from peripheral blood) are taken. The first human studies sought to discover frequencies of VDJ segment usage, the similarity in VDJ

segment usage and CDR3 AA sequences between individuals, and to estimate the total size of the BCR repertoire (22, 23, 24, 44, 138, 139, 140, 141). Size estimates are difficult because samples taken are a small representation of the entire repertoire, but sequencing independent replicate libraries in a capture-recapture analysis indicates a minimum bound of 10^6 unique V_H rearrangements (23). This estimate neglects size differences in the repertoire of different B cell subsets, so there are likely to be a greater number of unique naive B cells, but fewer unique memory cells than this (25). Comparing the repertoire across individuals showed that VDJ segments were used in unequal frequencies in the repertoire, but that their pattern of usage was similar between individuals, indicating inherent biases in the VDJ recombination process, and preferential use of core gene segments (23, 24, 44). Despite similar VDJ usage, there was limited overlap in the CDR3 repertoires between the two individuals studied by Arnaout *et al.* (24).

The structure of the B cell repertoire has also been investigated as a function of age, to see if it can help to explain the phenomenon of immunosenescence - a decrease in immune function in the elderly (142). In a comparison of 27 individuals between the ages of 20 and 89, it was found that elderly individuals had longer CDR3 regions, greater persistence of large clones, and increased mutational loads compared to younger individuals (143). Furthermore, this study documented both cytomegalovirus (CMV), and epstein-barr virus (EBV) infection status. CMV infection resulted in increased V gene mutation in the IgG and IgM repertoire, while EBV infection resulted in an increased number of persistent clonal groups in the repertoire. As the rate of chronic viral infection increases with age, the effect of these two variables on the repertoire will be tightly linked. Interestingly, sex was not found to effect the repertoire in this study, despite the presence of other known sex-specific effects on the immune system (144). The effect of age on the BCR repertoire response to vaccination has also been investigated in a small number of individuals in two separate studies (99, 119). Jiang *et al.* studied the BCR

repertoire before and after influenza vaccination in children (8-17 years), young adults (18-30 years) and the elderly (70-100 years) (119). As there were only four individuals in the elderly age group, statistical comparisons were not performed, but this study did show that in two of the elderly individuals, the repertoire was highly clonal and had a greater mutational load compared to the younger individuals. In another study, the isotype-specific antibody repertoire was studied after simultaneous administration with pneumococcal and influenza vaccination in six younger (19-45 years) and six older (70-89 years) individuals (99). There was a slight increase in IgG mutation in the older age group, but the most striking differences were in the IgA and IgM repertoires, which displayed slower clonal expansion as well as less mutation, and longer CDR3 regions in the elderly.

Further studies have used FACS to isolate different B cell subsets, and show differences in the repertoire between them (122, 141, 145). To determine whether IgM memory cells were early emigrants from a TD GC response (and so related to IgG memory cells), or were instead formed after stimulation with TI antigens, the repertoire in IgM memory and IgG memory cells was compared. Wu *et al.* and Briney *et al.* both observed differences in VDJ segment composition in the IgM and IgG memory repertoires (122, 141). Wu *et al.* also found the IgM repertoire to contain fewer negatively charged amino acids, and greater levels of tyrosine, as well as having a lower hydrophobicity and aliphatic index than the IgG repertoire, supporting the notion that IgG and IgM memory cells could comprise populations with distinct TD or TI origins, which have been formed through responses to different antigenic stimuli (although it cannot be ruled out that IgG and IgM memory cells have the same origins, but are subject to distinct regulatory mechanisms) (122). Mroczek *et al.* extended these studies by analysing the repertoire of 8 different B cell subsets (immature IgM, transitional IgM, mature IgM, memory IgD+ IgM, memory IgD- IgM, PC IgM, memory IgD- IgG and PC IgG) isolated from a single individual (145), and found differences in both VDJ segment composition, and CDR3 AA

composition of all of them. Results of the studies from Mroczek *et al.* and Wu *et al.* were highly consistent, with both observing multiple increases in IGHV1 family genes and IGHV3-23 in IgM compared to IgG memory cells.

Changes in the B cell repertoire following antigen stimulation

Vaccination can be used as a controlled method to study the immune response in humans. The BCR repertoire following vaccination has been examined using NGS as part of nine different studies (Table 1.8). These studies have been able to broadly demonstrate minor changes in VDJ segment usage, and in the size and diversity of the different B cell lineages after vaccination (119, 146, 147). By comparing NGS data to sequences from known antigen-specific cells, it appears that the total repertoire data obtained by NGS includes BCR sequences that are targeted against vaccine antigens (97, 119, 148, 149, 150). These antigen-specific sequences appear to be enriched in the shared (public) repertoire after vaccination (149, 150). Such sequences may also be tracked between successive vaccination events, and may be used to demonstrate memory recall (25). The statistical likelihood of an identical BCR sequence being generated in independent recombination events is so low that if the same sequence is detected in successive samples, it likely represents a derivative from the same B cell clone (151).

Vaccine studies have also been able to demonstrate quantitative differences in the reaction of the BCR repertoire in response to different influenza vaccines (25, 119). They have shown that, as well as being a confounding factor in the absence of immune stimulation, age also affects how the repertoire responds to vaccination (119, 146). Finally, studies on identical twins have shown that whilst VDJ segment usage in naive cells appears to be genetically determined, in the vaccine-activated repertoire, there is very little overlap in the specific clones used between twins, indicating that there are also random or individual-specific effects influencing the specific antigen response (119, 152).

Table 1.8: Studies analysing the B cell repertoire following vaccination

Vaccine	Cells used	Methodology	Key findings	Ref.
Influenza				
TIV	Memory B cells 14 days after vaccination (1 participant).	Single-cell VH:VL linkage PCR (RNA template) and 2x250 bp Illumina sequencing.	Validated accuracy of VH:VL pairings identified using their high-throughput method. Identified 240 putatively influenza-specific CDR-H3:CDR-L3 pairings.	(148)
TIV or LAIV	PBMCs on the day of vaccination, and on day 7 and 28 after two vaccinations given a year apart (28 participants).	VH-specific reverse transcription and 2nd strand synthesis, with barcodes, followed by PCR (RNA template). Custom 100 x 120 bp Illumina sequencing protocol.	Showed different repertoire dynamics after TIV and LAIV vaccination. TIV induced a stronger response, with more abundant IgG lineages than LAIV. Some sequences present after the first vaccine are also found after the second vaccine - hypothesised that these are from recalled memory B cells.	(25)
TIV or LAIV	Naive B cells, and PCs, on the day of vaccination and on day 7/8 and day 28 after vaccination (17 participants, three age groups).	VH-specific multiplex PCR (RNA template), and Roche 454 sequencing.	The influenza-specific BCR repertoire in older individuals was more clonal, and had a greater mutational load than the repertoire in younger individuals. In twins, the mutational load of the IgM repertoire was similar, but diverged for the IgG repertoire, indicating that the naive repertoire is more influenced by individual genetics, but the memory repertoire is more influenced by environmental stimuli.	(119)
TIV	PBMCs on 18 timepoints around vaccination (3 participants).	VH-specific multiplex PCR (RNA template), and Roche 454 sequencing.	V and J segment usage differs between individuals, and is conserved within individuals over time. There is clonal expansion and contraction of certain clones in response to the vaccine with different participants exhibiting different dynamics. There are a small number of highly mutated, persistent clones found within all individuals, potentially corresponding to long-lived B cell memory or indicative of chronic infection.	(147)
TIV	PBMCs on the day of vaccination, and on days 7 and 21 after vaccination. PCs on day 7 after vaccination (27 participants).	VH-specific multiplex PCR (gDNA template), and Roche 454 sequencing.	Clonal sequence expansion correlated with the serum antibody response. The expanded sequences were enriched for influenza-specificity. Found common V gene use in response to vaccination in all participants.	(149)
Tetanus				
TT	PCs 7 days after vaccination (1 participant).	High-throughput single-cell VH:VL linkage PCR (RNA template) and 2x250 bp Illumina sequencing.	Identified 86 putatively TT-specific CDR-H3:CDR-L3 pairings. Cloning ten of these into HEK293K cells followed by competitive ELISA of the antibodies produced showed them to be TT-specific.	(148)
TT	Total PCs, memory B cells, and antigen-specific PCs 7 days and 3 months after vaccination (2 participants).	VH and VL-specific multiplex PCR (RNA template), and Roche 454 sequencing in addition to proteomic analysis of TT-specific serum antibodies.	Analysed the serum antibody repertoire by using the VH sequence database to interpret results from high-resolution liquid chromatography tandom mass spectrometry of the serum antibodies. Showed that 5% of the PC clonotypes identified by sequencing at day 7 could subsequently also be detected in the serological response 9 months after vaccination.	(97)
Varicella-zoster virus				

Continued on next page

Vaccine	Cells used	Methodology	Key findings	Ref.
VZV	PBMCs on the day of vaccination, and on days 8, 14 and 28 after vaccination (4 pairs of identical twins).	VH-specific multiplex PCR (gDNA and RNA template) and Roche 454 sequencing.	VDJ segment use was more similar within than between twin pairs, and this was more pronounced for IgM than IgG sequences. Twin pairs had greater numbers of convergent, mutated sequences that non twin pairs, indicating some similarity in the memory repertoire. However, the response to vaccination at the sequence levels was unique to each individual regardless of twin status.	(152)
Multiple				
TIV & PPV23	PBMCs on the day of vaccination and on day 7 and 28 after vaccination (14 participants, two age groups).	Semi-nested isotype and VH-specific multiplex PCR (RNA template), and Roche 454 sequencing.	The repertoire changed at day 7 post-vaccination, but returns to a baseline-like state after 28 days. Comparing the repertoire in young and elderly individuals after vaccination indicated clonal expansion is delayed in older individuals. Showed age-related differences in IgA and IgM repertoire dynamics.	(146)
Hib-MenC-TT	B cells on the day of vaccination, and B cells or PCs 7 days after vaccination (5 participants).	VH-specific multiplex PCR (RNA template), and Roche 454 sequencing.	Searched for convergent sequences after vaccination, and found that these were enriched for specificity towards the vaccine antigens. The number of Hib-specific sequences found 7 days after vaccination correlated with anti-Hib avidity index (a correlate of protection) 28 days after vaccination.	(150)

TIV; Trivalent inactivated influenza vaccine, LAIV; live attenuated influenza vaccine, PPV23; 23-valent pneumococcal polysaccharide vaccine, PBMCs; peripheral blood mononuclear cells

1.4.3 The application of B cell repertoire analysis to vaccination

Whilst vaccination offers a useful tool with which to study the basic biology of the BCR repertoire, BCR repertoire sequencing technology also has potential applications in multiple areas of vaccine development and testing. This includes determining the properties of the specific B cells used in the response to different vaccines, predicting vaccine safety, and guiding the development of more effective vaccines by increasing our understanding of B cell and GC immunology. Many of these applications have yet to be fully realised, and will be further explored throughout the results sections of this thesis.

Generating mAbs

Perhaps the most well defined use of BCR sequencing after vaccination is in the generation of mAbs. mAbs can be used for passive immunisation against certain diseases (153), and have important research applications. After vaccination, PCs (or antigen-enriched cells) can be iso-

lated from peripheral blood by FACS and sequenced to identify the sequences of the antibodies that are produced in response to the antigen. Generally, the most abundant sequences are then cloned into an expression vector for functional characterization. The requirement of both V_H and V_L sequences to be obtained for mAb generation resulted in initial studies using a single-cell sorting approach followed by V_H and V_L sequencing of individual cells (104, 154). As methods are developed for linking the V_H:V_L repertoires, it will be possible to identify mAbs in a more high-throughput manner (as demonstrated by Reddy *et al.* (155)). Despite the exact method used, such studies have demonstrated 6-76% (80, 154, 155, 156) of cell lines created produce antigen-specific antibodies, resulting in a significant decrease in the time taken to generate effective mAbs compared to the conventional methods of screening large combinatorial libraries.

Tracking known sequences

There are numerous antibody sequences in the literature that have a known specificity (103). Incorporating BCR repertoire sequencing into studies of vaccines, autoimmunity and infectious diseases will aid in the creation of increasingly comprehensive databases containing antibody sequences with defined specificity and function (Figure 1.9). In theory, if these databases contained sequences known to be specific for a certain vaccine antigen, we could search for these previously described sequences in the post-immunisation repertoire during new studies of the vaccine antigen, and infer the degree to which a protective immune response has been generated. Current data from a limited range of antigens suggest that a vaccine response is likely to be complex, and in-depth studies of the BCR repertoire will be required to fully understand the relationship between NGS data and immunogenicity, as discussed later. Although there are currently few known sequences for auto-reactive antibodies, the identification of known auto-reactive sequences in the post-vaccination repertoire, would be of interest in terms of signalling the potential for certain vaccines to increase the risk of autoimmune disease. However, the

large number of self epitopes, and the potential diversity of auto-reactive sequences that could target these epitopes make it unlikely that all potential auto-reactive sequences could ever be identified. A recent study of the antibody repertoire in MS patients identified expansion of some sequences potentially related to the disease (157), and as such sequences are discovered, it would be beneficial to routinely test for them in vaccine trials as a precautionary measure.

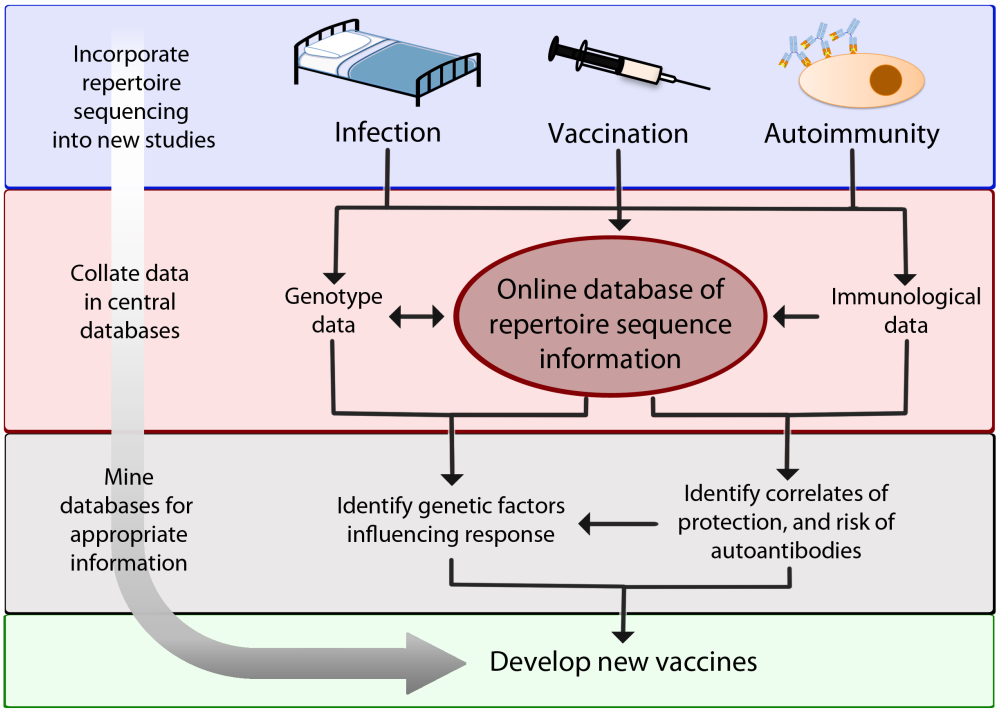

Figure 1.9: Proposed method for uncovering applications of BCR sequencing - Proposed flow of information from repertoire sequencing projects to aid in the generation of data with practical uses. For such data to be suitable for meta-analysis, it is important that standardised protocols and file formats are used.

Considering a specific example of the utility of tracking known sequences for vaccine development, in the case of HIV, neutralizing antibodies have been discovered against the envelope glycoprotein (Env) in HIV-infected individuals, but stimulating formation of these antibodies by vaccination is difficult, as their germline precursors have poor affinity for Env (158, 159). In some HIV infected individuals, it appears that neutralizing antibodies may be formed from mutated, cross-reactive precursors that were initially generated during previous non-HIV infections, and then further mutate to become specific for Env (160). It has been observed in

one study, that the unmutated precursor of an effective anti-HIV antibody was able to bind to the founder HIV virus. Sequencing the viral Env in parallel to the BCR genes at multiple timepoints after virus transmission indicated that BCR maturation was preceded by viral evolution, suggesting that such co-evolution is a potential method for driving formation of anti-Env neutralizing antibodies (161). It seems for HIV, that an effective vaccine requires prior identification of the correct germline precursor BCR sequence. The germline precursor could then be stimulated using an alternate immunogen to Env, and successive immunogens then administered to guide B cell maturation to form the desired final antibody (162). The success of this work is being driven by the ability to identify low-abundance BCR sequences in high-depth BCR sequencing experiments (159).

Investigating the breadth of the antibody response

As well as identifying specific sequences, NGS data can give insight into the breadth of the BCR repertoire (a measure of the diversity, in terms of the number and abundance of different sequence clones (163)) used in an immune response. For some vaccines, the antigenic site targeted is highly variable (HIV-1 Env/Plasmodium vivax Duffy-binding protein), or undergoes seasonal change (influenza HA); a broader response can potentially confer protection against more antigenic variants. BCR repertoire sequencing has been instrumental in demonstrating that use of a TLR agonist in an oil-in-water adjuvanted malaria vaccine formulation is able to broaden the antibody response in mice. Use of the TLR agonist in the vaccine led to improved antigen neutralization, and efficacy against more varied malarial strains (164). Similar observations have been made using oil-in-water adjuvanted influenza vaccines, which induced broadly cross-reactive responses against the HA head, providing some protection against drifted strains of the virus (165).

While it seems reasonable to suggest that the breadth of the BCR repertoire is a measure of ability to neutralize diverse antigenic strains, it is also possible that single antibodies can have

broadly neutralizing activity by targeting conserved regions on otherwise variable antigens. This appears to be the case with influenza HA, where broad neutralization of multiple variants can arise from a clonal antibody response targeting the conserved stem region or a diverse response targeting the variable head region (166). Stem-binding antibodies are not found after seasonal influenza vaccination, but of the 28 HA-binding mAbs generated from eight subjects after 2009 pandemic H1N1 influenza vaccination, Li *et al.* found that three of these were able to bind the stem region and elicit cross-reactive protection (166). The authors hypothesize that as the influenza response is predominately driven by memory cells, and the pandemic virus has a highly divergent HA head compared to seasonal variants, only the cross-reactive memory cells would be stimulated. This is a contested issue though, as a study by Wrammert *et al.* in which over 50 influenza-specific mAbs were generated from PC sequences after influenza vaccination, showed that they exhibited the highest affinity for the current circulating strain of influenza (104). The conflicting nature of these studies indicate that the high-throughput nature of NGS data would be better able to give insight into the true breadth of the response after seasonal influenza vaccination. Using conventional immunological assays in parallel with BCR repertoire sequencing would also be useful to improve our understanding of how repertoire data relates to functional measures of cross-reactivity.

Understanding immunological mechanisms of vaccination

The majority of vaccines and immunisation schedules were initially developed empirically, rather than through a detailed understanding of the immunological mechanisms of vaccination. Data generated from BCR repertoire sequencing both after vaccination, and in populations without vaccine challenge, is starting to give great insight into the underlying immunology of vaccination. For vaccines such as hepatitis B and malaria RTS,S, where different dose schedules are available, the schedule used affects vaccine immunogenicity (167, 168). Using BCR sequencing to improve understanding of the GC reaction and BCR maturation pathways could

aid the design of optimal dose schedules. Booster doses of a vaccine are used to increase the number and affinity of antigen-specific antibodies. Using surface plasmon resonance to analyse the antibody repertoire after repeated doses of TT booster vaccination in two previously immunised individuals, the limits of antibody affinity maturation were determined to be reached after just two booster doses for this antigen (80). Sequencing PCs produced after these vaccinations indicated that the limit of somatic hypermutation was also reached after two doses. A third dose of TT did not increase either the affinity of the anti-TT antibodies, or the diversity of the repertoire (80). Repeat doses of vaccine given close together may increase the magnitude of the antibody response by re-stimulating GCs, or by stimulating new GCs to form, resulting in the production of new PCs. Sorting and sequencing antigen-specific B cells present in the peripheral blood after vaccination may allow capture of recent GC emigrants. Observing the relative numbers of these cells, and their mutational load, could give insight into proliferation and affinity maturation within the GC after repeat vaccination. It would also be instructive to understand the interaction between dose interval and antibody maturation. There is some evidence from RTS,S vaccination that increasing the dose interval increases the time available for SHM to occur, thus leading to higher affinity antibodies being produced (168). Dose schedules could then be altered accordingly to optimise responses, and prevent unnecessary vaccinations.

Measuring vaccine immunogenicity

Developing new vaccines, and predicting efficacy of available vaccines in new populations, requires measuring correlates of immunity. Correlates are generally measures of the immune response, such as antibody concentration, or function (e.g. serum bactericidal assay for group C *N. meningitides* or HA inhibition assay for influenza), which relate to vaccine-induced protection against infection. For many vaccines, such correlates are often unavailable, or hard to measure and standardise between laboratories (83). Sequencing the BCR repertoire after vaccination can provide a detailed dissection of the vaccine-induced B cells underlying the an-

tibody response, and a potential application of this is in deriving novel correlates of protection. Although repertoire data can be used to identify known antigen-specific sequences, and the breadth of the response, it remains to be demonstrated whether it can be used as a valid standalone measure of immunogenicity. A critical first step is to distinguish the vaccine-specific repertoire from the total repertoire. The vaccine-specific repertoire (in terms of number and abundance of sequences) would then need to be correlated with a known indicator of vaccine response, or with protection from infection. Other aspects that must be considered for a good correlate of immunity are that the signal detected is not transient, and that it is not confounded by other individual factors such as genotype, antigen exposure history, and age. Signal persistence in BCR repertoire sequencing studies has not been well investigated. It is clear that perturbations can be seen on days 7, 8 and 14 after vaccination (Table 1.8), and some sequences are found to persist in individuals for at least a year, and are identifiably re-stimulated upon further vaccination (25). It is less clear to what degree such perturbations in the total repertoire reflect perturbations in the antigen-specific repertoire, and are related to vaccine immunogenicity.

1.4.4 Challenges of B cell repertoire analysis

Understanding of the BCR repertoire has increased enormously over the past decade, due mainly to improved sequencing technology. There is an increasingly comprehensive view of the structure of the baseline BCR repertoire, and how this differs between individuals and B cell subsets. An appreciation of how vaccine antigens dynamically restructure the BCR repertoire, and how certain adjuvants affect this process is also emerging. However, as it is still a relatively new technology, there remain a number of challenges that must be overcome to uncover the full potential of this technology.

Creating common protocols and terminology

In order to allow useful public databases to be created, and meta-analyses to be conducted (Figure 1.9), there must be standardisation of BCR repertoire sequencing. At each step in the process (Figure 1.7), different laboratories have adopted different practices, making comparison of results problematic (169, 170). The use of different primers and PCR protocols potentially introduces different degrees of PCR amplification bias, and error into the resultant dataset. The different sequencing platforms used (Table 1.6) also result in different length reads, types of sequencing error, and depth of sequencing (116). Also, some protocols analyse genomic DNA (gDNA) (23, 118), whereas others analyse RNA (25, 119).

Once data are generated, there is also a lack of standardised analysis pipeline. Dealing with large sequence datasets requires the implementation of bioinformatic frameworks for data storage and analysis. Although there are several publicly available tools for BCR sequence analysis (171), most laboratories use custom bioinformatics pipelines, making it difficult to repeat analyses even when working from the same dataset. Furthermore, there are discrepancies in certain definitions pertaining to BCR data, making both interpretation and repetition of analysis challenging. For example, much analysis focuses on sequence clusters which should all theoretically derive from clonally related B cells. However, the published definitions of a sequence clone range from just sharing the same V, D and J gene segment usage (141), to having the same V, D and J segment usage in addition to having highly similar CDR3 sequences (25). Another example is the definition of the actual CDR3 region boundaries, which are commonly based on the intervening sequence between fixed points in the V and J gene segments. Some laboratories base this on the conserved cysteine residue at Kabat position 92 (IMGT position 104) in the V segment, and the conserved tryptophan residue at Kabat position 103 (IMGT position 118) in the J segment (150), while other laboratories consider a slightly shorter region within this, where there is no longer homology with the V and J segments (118).

Understanding the 'normal' repertoire, and technological limitations

Whilst the advent of BCR sequencing technology spurred a large number of studies of the repertoire in a state of disease, or after an intervention, there is relatively little quality control data, or baseline data. It is not possible to sequence the entire repertoire in humans, so a representative sample is normally taken. How well such samples accurately represent the total repertoire, and how repeatable the measures are is not well understood. Repeatability studies have been conducted in mice (172), but not exhaustively in humans, who have a much larger repertoire (22). Not having a large source of healthy control data makes it difficult to assess whether the repertoire of patient cohorts can be considered normal or not. Furthermore, there are also no studies assessing the repertoire of healthy individuals over time to see how much variation there is in the absence of any intervention. This means that when analysing repertoire time-courses in response to vaccination, it is difficult to deconvolute changes in the repertoire that are likely caused by the vaccine, from the background immune fluctuations. The lack of such control data hampers the development of clinical applications of the technology, where measures must be highly repeatable, and well validated.

Deconvoluting the antigen-specific and total repertoire

The ability to distinguish between the total BCR repertoire, and antigen-specific BCR repertoire is key for any attempt to reliably utilize NGS data for elucidation of specific B cell responses. Whereas NGS methods are well suited to studying perturbations in the total sampled repertoire, to date the low-resolution methods have provided the most insight into the antigen-specific repertoire (Section 1.4.1). To assess the full diversity of the antigen-specific repertoire, it is necessary to have methods for identification of antigen-specific sequences from total NGS repertoire data.

One way to identify such sequences is to search for those that have a degree of identity to previously described sequences, a technique that Zhu *et al.* used for de-novo identification

of new VRC01 antibodies (from a new donor) against HIV (173). Cross-donor phylogenetic analysis using known VRC01 sequences, and sequences from the new donor, was used to identify novel VRC01 V_H sequences based on evolutionary similarity. V_L sequences were identified based on the presence of a five AA motif present in the previously described sequences, and when reconstituted with the V_H genes were able to neutralize HIV-1 with varying potency.

An a priori method for the identification of antigen-specific sequences from total repertoire data is by analysis of the BCR sequences that are shared between multiple individuals following recent exposure (through infection, immunisation or autoimmunity) to a common antigen - this is termed the 'convergent', or 'public' repertoire. The extent to which an individual's B cell repertoire for an antigen is shared with other individuals remains uncertain and is a key question to answer in order to determine the utility of 'convergent' repertoire analysis. Such convergence has been seen in the case of dengue fever, where a convergent CDR3 signature (predominantly 'ARLDYYYYYGMDL') is enriched for during acute disease compared to at convalescence, or in healthy controls (118). Vaccine studies are an ideal tool for investigating the degree to which convergent sequences are likely to be antigen-specific (Figure 1.10). The ability to control the timing of antigen administration and sampling in vaccine studies increases our ability to sample the repertoire at timepoints when there is maximum egress of antigen-specific cells into the sampled compartment. Generally, an antigen-specific PC burst is seen on days 4-10 in the peripheral blood following vaccination (with different kinetics for a primary vs. secondary response (90, 91, 95)). In response to both influenza (149), and Hib-MenC-TT vaccination (150), the day 7 convergent repertoire does appear to be enriched for vaccine-specific sequences. The majority of NGS studies have focused on sampling around day 7, where PCs are most likely to peak, yet with sufficient depth of sequencing it may also be possible to detect convergent repertoires related to antigen-specific memory cell populations present at later time points as well (106).

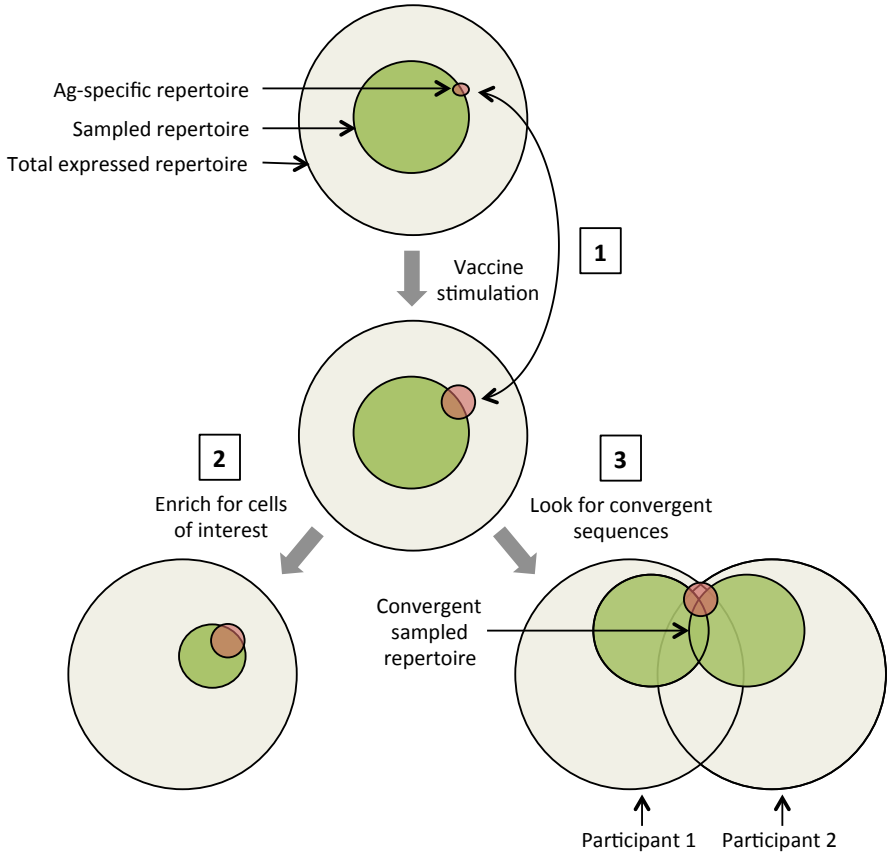

Figure 1.10: Using vaccines to investigate the antigen-specific antibody repertoire - Vaccination will cause an increase in the number of antigen-specific cells in the repertoire responding to the vaccine. There are then three ways in which antigen-specific sequences can be enriched for in the sampled repertoire, either physically or during analysis. (**1**) Identification of sequences that have a relative increase, and persist over time in the sampled repertoire after vaccination. (**2**) Physical enrichment using FACS to isolate PCs or antigen-specific cells for analysis, so that the sampled repertoire contains a greater proportion of antigen-specific sequences. (**3**) Identification of sequences shared by more than one individual (convergent repertoire), after stimulation with the same antigen. Using a combination of these three methods will likely give the greatest enrichment of antigen-specific sequences for subsequent analysis.

Linking V_H and V_L repertoires

An initial limitation of NGS for characterisation of the BCR repertoire was that either the V_H or V_L repertoire could be determined independently, but the pairing between these was lost. As the heavy chain is the most important in determining antigen-binding specificity, this is usually sequenced on its own. However when a sequence is found of suspected importance, it is not easy to validate the specificity of this sequence by producing a mAb. If both V_H and V_L repertoires are sequenced, the pairing can be roughly reconstituted using phylogenetic matching. Here, phylogenies are constructed independently from the V_H and V_L data, and where there are

similar architectures, it is inferred that the similar branches in the two phylogenies are derived from the same B cell, and thus the sequences from them are matched. This method has been used to pair V_H and V_L sequences from variants of an anti-HIV broadly neutralizing antibody; the authors reconstituted both matched and unmatched V_H:V_L pairs, and show that their matching method led to significantly reduced autoreactivity, which they suggest supports the validity of this matching approach (173).

Microfluidics approaches are enabling the development of more accurate sequencing of natively paired V_H and V_L repertoires. Dekosky *et al.* have reported an approach of encapsulating B cells within emulsion droplets, and then carrying out reverse transcription and V_H:V_L linkage PCR within these droplets (174). This is able to analyse >2 x 10^6 B cells per experiment, but requires custom-built equipment, and creates an amplicon that is currently too long to be fully sequenced using Illumina's technology. Another approach is to incorporate unique barcodes within the emulsion droplets so that after reverse transcription, each cDNA from a particular cell contains a unique barcode. V_H and V_L repertoires can then be amplified and sequenced, and paired based on having a shared barcode. Thus, although V_H:V_L pairing is not currently easy to determine, technological improvements make it likely that in the near future this will be commonplace.

1.5 Aims and Objectives of this thesis

This thesis aimed to develop robust laboratory and bioinformatic methods for high-throughput sequencing of the BCR repertoire in humans, and to then investigate the potential of these methods for describing vaccine-specific B cell responses. Such information would increase our understanding of the humoral immune response, and could be used to aid the development and testing of new vaccines. Specific aims were as follows:

1. To investigate BCR repertoire structure and diversity both within and between individ-

uals in the absence of any specific immune stimulus

2. To determine to what extent BCR repertoire sequencing can be used to investigate the B cell response to primary and secondary antigen exposure, using hepatitis B vaccination as a model system

3. To develop analytical models for identifying antigen-specific BCR sequences from the total BCR repertoire following antigen stimulation by combining data from both the total BCR repertoire, and the antigen-specific BCR repertoire from antigen-specific sorted cells

4. To investigate the use of BCR repertoire sequencing in understanding/elucidating immunological mechanisms of vaccine response in two case studies:

   (a) Tracking the B cell response to meningococcal polysaccharide and conjugate vaccines

   (b) Investigating the effect of adjuvant on influenza vaccine

2

General Methods

2.1 Laboratory Methods

2.1.1 List of buffers used

Phosphate buffered saline (PBS)

5 PBS tablets (Sigma) dissolved in 1000 ml H_2O and adjusted to pH 7.2-7.4 using 10M NaOH.

PBS-Tween20

1000 ml of 1xPBS to which 2.5 ml Tween20 (VWR International) had been added.

R0 medium

500 ml RPMI-1640 (Sigma-Aldrich) to which 5 ml of penicillin-streptromycin at 50 U/ml and 0.05 g/ml respectively (Sigma-Aldrich) and 5 ml of 2mM L-glutamine (Sigma-Aldrich) had been added.

R10 medium

450 ml R0 medium to which 50 ml heat-inactivated newborn bovine serum (Sigma-Adrich) had been added

Complete medium

500ml R10 medium to which 5 ml 10mM Non-essential amino acids (Invitrogen), 5 ml 100 mM Sodium pyruvate (Invitrogen) and 500 μl 50 mM 2-Mercaptoethanol (Invitrogen) had been

added.

Running Buffer

1000 ml 1xPBS to which 0.744 mg EDTA (ICH Biochemicals), and 5 ml heat-inactivated newborn bovine serum (Sigma-Adrich) had been added

2.1.2 Blood collection

Phlebotomy was performed on consenting volunteers in accordance with the Declaration of Helsinki, by trained doctors or nurses within the Oxford Vaccine Group. Blood was immediately transferred to Falcon tubes containing 15 μl heparin per 1 ml of whole blood to prevent clotting, and processed within 4 hours of collection.

2.1.3 PBMC separation

Heparinized blood was diluted 1:1 with R0 medium. Diluted blood was layered over Lymphoprep (Axis-Shield Diagnostics) for density gradient centrifugation at 1100 xg for 30 minutes. Peripheral blood mononuclear cells (PBMCs) in the buffy layer were removed, and washed in R0 medium, followed by centrifugation at 900 xg for 20 minutes. Supernatants were discarded, and the resulting cell pellet was resuspended in R10 medium.

2.1.4 Cell counting

Cell suspension was mixed 1:3 with Trypan Blue stain (VWR International) and 1xPBS. 10 μl of diluted cell suspension was transferred to a Neubauer Haemocytometer (VWR International), and the number of cells in 5 squares counted. Dead cells, and non-PBMCs were excluded from the count. Total cell count in the sample was determined by multiplying the cell count in 5 squares by 5 (to give the number of cells in all 25 squares), then by 3 (to account for diluting the cell suspension 1:3), then by 10^4 (to account for the volume of cell suspension in the chamber of the haemocytometer) and finally by the original volume of cells.

2.1.5 PBMC freezing/defrosting

Following counting, PBMCs were centrifuged at 500 xg for 5 minutes, and the supernatant completely removed. Cells were resuspended in ice-cold Recovery^{TM} Cell Freezing Medium (Invitrogen) at 10x10^6 cells/ml. Working rapidly, 1 ml aliquots of cell suspension were transferred to cryovials (Greiner), and placed in a chilled Mr Frosty freezing container (Fisher Scientific). Mr Frosty containers were left to freeze at -80 ^{o}C overnight, before removal of the cryovials and transfer to a LN_2 cryostore.

For defrosting, cryovials were placed in a 37 ^{o}C water bath. Defrosted cells were rapidly transferred to a centrifuge tube, and washed twice with ice-cold Complete Medium followed by centrifugation at 1000 xg for 20 minutes. Supernatants were discarded, and the cells resuspended in R10.

2.1.6 Cell culture

Cell culture was performed to stimulate memory B cells to secrete antibody, so that they can be detected by ELISpot (Section 2.1.7). PBMCs were cultured in 96 well round-bottomed culture plates. Each well contained 2x10^5 PBMCs, suspended in 200 μl R10, containing 1/5000 *Staphylococcus aureus* cowans strain Pansorbin cells (VWR International Ltd), 1/6000 Pokeweed mitogen (Sigma-Aldrich) and 1/40 CpG oligonucleotide (Invitrogen). Cells were incubated at 37 ^{o}C in 5% CO_2 for 6 days. After incubation, cells were resuspended, and washed once in R10, followed by centrifugation at 900 xg for 20 minutes, and then washed twice more in R10 medium, followed by centrifugation at 900 xg for 15 minutes. Supernatants were discarded, and the resulting cell pellet resuspended in R10.

2.1.7 Cell phenotyping using ELISpot

The enzyme-linked immunospot (ELISpot) assay was used to detect the presence of antibody secreting B cells. 96 well MultiScreen-IP ELISpot plates (Millipore) were coated with either

100 μl of antigen solution (test wells), or 1xPBS (control wells). Different concentrations of each antigen (TT; Statens Seruminstitute, Polyvalent Ig; Catlag, HBsAg; GlaxoSmithKline) were trialled to find the lowest concentration that still gave well-defined spots (Table 2.1). Coated plates were sealed and stored at 4 ^{o}C for a minimum of 24 hours, and a maximum of 1 month before use. On the day of use, coated plates were washed with 200 μl/well of 1xPBS three times. Plates were then blocked with R10, and incubated at 37 ^{o}C in 5% CO_2 for a minimum of 30 minutes before addition of cells.

Table 2.1: Ag dilutions for coating ELISpot plates

	TT	**Ig***	**HBsAg**
Stock (μg/ml)	1935	2500	1340
Dilution on plate (μg/ml)	5	10	2.5

*Goat anti-human polyvalent Ig

Cells were resuspended at 2x10^6 cells/ml in R10 medium. Cultured cells were used to detect the presence of memory cells, and uncultured PBMCs were used to detect the presence of PCs. 100 μl cell suspension was added to each well in the prepared plate, and incubated for 16-20 hours at 37 ^{o}C in 5% CO_2. The plate was then washed four times with 200 μl/well PBS-Tween20, and once with 200 μl/well 1xPBS. IgG alkaline phosphatase conjugate (Calbiochem) was diluted 1:5000 in R10, and then filtered before adding 100 μl to each well, and incubating for 4 hours at room temperature. The plate was then washed four times with 200 μl/well PBS-Tween20, and three times with 200 μl/well with H_2O. An alkaline phosphatase substrate kit (Bio-Rad #1706432) was used according to manufacturer's guidelines to create the alkaline phosphatase substrate, 50 μl of which was then added to each well in the plate. Spots were allowed to sufficiently develop, before the reaction was stopped by the addition of 200 μl/well of H_2O. The plate was then washed once more with 200 μl/well H_2O before being left to dry.

Plates were read and counted automatically using the AID ELISPOT reader, and Software Version 5.0 (Autoimmune Diagnostika). All plates were double-checked manually for erroneous

spots. Two to six replicate wells were conducted for each antigen depending on the number of cells available, and the mean spot count used. The number of spots related to background noise was then subtracted from each antigen well based on the number of spots in the PBS well.

2.1.8 Determining Anti-HBs antibody concentration

Blood serum was isolated by centrifuging whole blood at 3000 xg for 10 minutes, and removing the serum layer using a pipette. Serum was tested for the concentration of antibody specific for HepB surface antigen (HBsAg) IgG concentration (anti-HBs) using an automated enzyme-linked immunosorbant assay (ELISA) system (AxSYM HBsAg (Abbott Laboratories)) at the microbiology laboratory, John Radcliffe Hospital, Oxford.

2.1.9 Magnetic-activated cell sorting (MACS)

MACS was used to enrich CD19 positive B cells from total PBMCs. The manufacturer's guidelines for use of the CD19 microbead kit (Miltenyi Biotec) were followed. FcR block was used during the bead staining step to reduce non-specific binding. Separation was conducted on an autoMACS Pro separator using the 'Possel' setting, and automatic labelling. CD19 enriched B cells were counted, and either used directly for sequencing, or as input for FACS. B cells used for sequencing were split into aliquots of 500,000, centrifuged at 1000 xg for 2 minutes, and resuspended in 600 μl RLT lysis buffer (Qiagen). These were then snap-frozen prior to storage at -80 ^{o}C.

2.1.10 Fluorescence-activated cell sorting (FACS)

To prepare cells for flow cytometry and FACS, processed cells were pelleted at 1000 xg for 2 minutes, and resuspended in 250 μl running buffer per 10^7 cells. Cells were then stained for 30 minutes at 4 ^{o}C in the dark with the antibody cocktail (detailed separately in each section). All antibodies were titrated prior to use to determine the optimal concentration

to achieve maximal separation between background fluorescence, and the positive population. For competition assays, non-fluorescent proteins were also added at this stage. Stained cells were washed three times with 200 μl running buffer, followed by centrifugation at 1000 xg for 2 minutes at 4 ^{o}C. Cells were then resuspended in 50 μl cellFIX^{TM} (BD Biosciences), and incubated at 4 ^{o}C for a further 10 minutes. 2 ml running buffer per 10^7 cells was then added before proceeding to analysis.

Flow cytometry was conducted on either a 9 colour CyAn ADP analyser (Beckman Coulter), or a 4 colour FACScalibur (Beckman Coulter). Cell sorting was conducted on either a MoFlo cell sorter (Beckman Coulter) or FACSAria cell sorter. Machines were calibrated using BD^{TM} CompBeads. Fluorescence minus one (FMO) controls were used to set gating cutoffs. Summit 4.3 (Beckman Coultier) was used for data acquisition, and FlowJo 10.6 (Tree Star) was used for analysis. Sorted cells were immediately resuspended in 350 μl RLT lysis buffer, and snap-frozen prior to storage at -80 ^{o}C.

Validating HBsAg+ FACS

In order to isolate HBsAg-specific B cells by FACS, a custom stain of HBsAg conjugated to APC was created by Miltenyi Biotec. This should selectively bind to B cells expressing a BCR specific to HBsAg. The specificity of this stain was validated using blood samples from HepB vaccinated individuals (see Chapter 4 for study design) using three different methods. First, a competition assay was performed, where increasing amounts of free HBsAg were added during staining to determine the degree to which HBsAg-APC binding to B cells was inhibited. This experiment was conducted in the context of an antibody panel containing propidium iodine (eBiosciences), CD19-FiTC (HIB19) (eBioscience), CD27-PECy7 (O323) (eBioscience) and HBsAg-APC and run on a FACScalibur. Prior to FACS, CD19+ B cells were magnetically enriched. Binding inhibition was measured in the viable, CD19+, HBsAg+ gate, and indicated the specificity of HBsAg-APC staining to be at least 50% (Figure 2.1). While increased quantities of free

HBsAg may have inhibited binding further, this was not possible to test due to limitations in the amount of HBsAg available. There was no inhibition of binding when using a non-specific antigen (TT) in the competition assay.

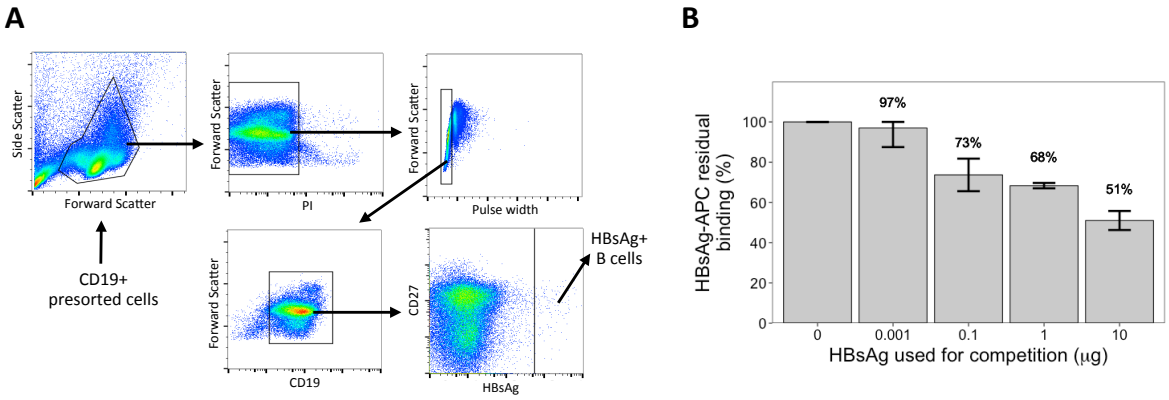

Figure 2.1: Validating HBsAg-APC by competitive binding - (**A**) FACS gating protocol for detecting of HBsAg+ B cells. (**B**) Results from the competition experiment, where unconjugated HBsAg was added to the antibody mix at the same time as the HBsAg-APC. Mean and ±SEM shown for three tests of each condition.

To further confirm specificity, the same antibody panel was used to determine the kinetics of viable, CD19+, CD27+, HBsAg+ B cells on days 0, 7, 14, 21 and 28 in four individuals following administration of a HepB booster vaccine, and compared to the kinetics of HBsAg-specific memory cells determined by ELISpot at the same times (Figure 2.2). HBsAg-specific B cells identified by flow cytometry show similar kinetics to those determined via ELISpot, giving evidence for good staining specificity. There is no increase in HBsAg-specific B cells determined by flow cytometry in the two participants who had a poor response determined by ELISpot, and an increase in the donors that had a good response determined by ELISpot. The only difference is in participant 2, where the HBsAg-specific B cells identified by flow cytometry peak at day 14 rather than 28.

Finally, for one sample HBsAg+ and HBsAg- cells were sorted using a MoFlo, cultured to promote antibody secretion, and phenotyped by ELISpot (Figure 2.3). The previously described cell culture method (Section 2.1.6) was modified for this to accommodate for lower

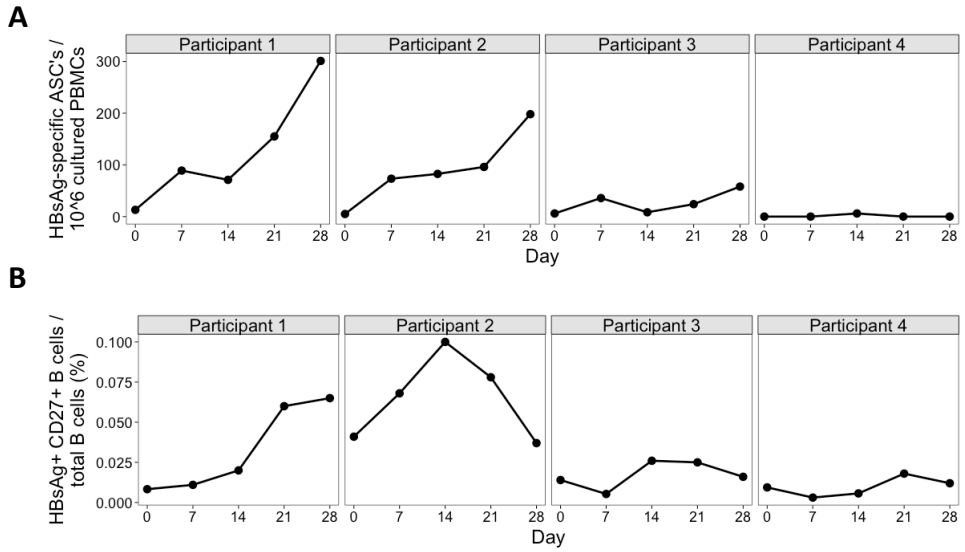

Figure 2.2: Validating HBsAg-APC by analysing kinetics following vaccination - (**A**) Number of HBsAg-specific ASC's detected following HepB booster vaccination in four individuals. (**B**) Percent of total B cells specific for HBsAg determined by flow cytometry.

cell numbers, and lack of accessory cells in the culture. V-bottomed culture plates were used to increase cell density, and the culture medium was supplemented with 60 U/ml IL-2 (BD biosciences), 40 ng/ml IL-10 (Calbiochem), 80 ng/ml anti-CD40 (BD biosciences) 80 ng/ml anti-CD27 (BD biosciences) and 200 ng/ml B cell activating factor (BioSupply). ELISpot was used to determined numbers of HBsAg-specific, TT-specific and total IgG secreting cells, and confirmed enrichment of HBsAg-specific cell in the HBsAg+ sorted fraction (Figure 2.3 B/C).

2.1.11 Sample preparation for sequencing

RNA extraction

Frozen cells were thawed, and homogenised by vortexing for one minute. Total RNA was extracted using the RNeasy Mini Kit (Qiagen) according to the manufacturer's guidelines. Fresh collection tubes were used at each step to reduce reagent carryover. RNA was eluted in 30 μl RNAse free water.

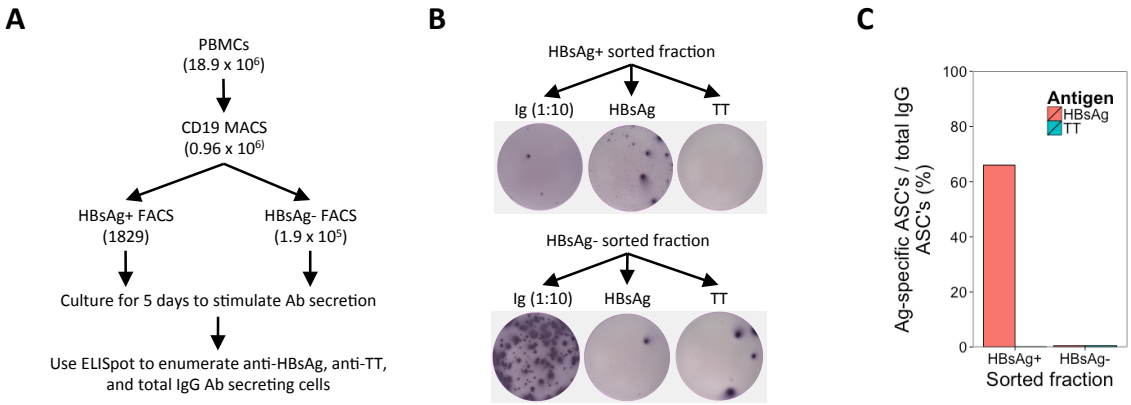

Figure 2.3: Validating HBsAg-APC by phenotyping sorted cells - (**A**) Experimental method for using ELISpot to phenotype the specificity of cells following FACS sorting into HBsAg+ and HBsAg- fractions. Numbers below each sorted fraction indicate the number of cells that were obtained. (**B**) Representative ELISpot wells from the two sorted fractions showing the number of total IgG, HBsAg-specific and TT-specific antibody secreting cells. (**C**) Percent of total IgG secreting cells that are specific for either HBsAg or TT in the two fractions.

Reverse transcription

Total cDNA synthesis was performed using the SuperScript III reverse transcriptase system (Invitrogen). Each 20 μl reaction contained 1 μl RNAsin (Promega), 1 μl 50 μM random hexamers (Applied Biosciences), 1 μl 10 mM dNTP mix (Qiagen), 4 μl 5x buffer, 2 μl DTT, 1 μl SuperScript III, and 10 μl template RNA. Reverse transcription was performed at 42 ^{o}C for 60 minutes, followed by 95 ^{o}C for 10 minutes using a DNA Engine PTC-200 thermocycler (MJ Research).

PCR

To amplify V_H genes from cDNA, a multiplex PCR reaction was used with a previously described primer set (122). This contained six forward primers (Table 2.2), which bind in the FR1 region of the seven V_H V gene segment families. The reverse primers bind in the constant region, and are isotype specific; there are separate primers for IgA, IgG and IgM. Primers that amplified a sequence of the β-Actin gene were used as positive controls. All primers were supplied as custom DNA oligos from Sigma-Aldrich, desalted, and diluted to 10 mM working stocks. Different polymerases (Taq (Qiagen), HotStarTaq (Qiagen), Phusion High-Fidelity (NE

BioLabs)) were trialled to find the most efficient for this application, with HotStarTaq, as part of the Multiplex PCR kit (Qiagen) being the most sensitive.

Table 2.2: V_H PCR primers. - obtained from Wu *et al.* (122)

Primer name	Sequence
IGHV1 Forward	CCTCAGTGAAGGTCTCCTGCAAGG
IGHV2 Forward	TCCTGCGCTGGTGAAACCCACACA
IGHV3 Forward	GGTCCCTGAGACTCTCCTGTGCA
IGHV4 Forward	TCGGAGACCCTGTCCCTCACCTGC
IGHV5 Forward	CAGTCTGGAGCAGAGGTGAAA
IGHV6 Forward	CCTGTGCCATCTCCGGGGACAGTG
CHA Reverse	GGCTCCTGGGGGAAGAAGCC
CHG Reverse	GAGTTCCACGACACCGTCAC
CHM Reverse	GGGGAATTCTCACAGGAGAC
β-actin Forward	CCAAGGCCAACCGCGAGAAGATGAC
β-actin Reverse	AGGGTACATGGTGGTGCCGCCAGAC

Each 50 μl PCR reaction contained 14 μl H_2O, 25 μl Multiplex PCR MasterMix, 1 μl each forward primer, 1 μl reverse primer, and 4 μl template cDNA. A no-template control was conducted in parallel to check for exogenous DNA contamination. When amplifying cDNA from 500,000 B cells, 30 PCR cycles was sufficient to generate enough DNA for sequencing, but when working with the smaller cell numbers obtained from FACS, 33 PCR cycles was required. After an initial denaturation at 94 ^{o}C for 15 minutes, 30 or 33 PCR cycles of 94^{o}C for 30 seconds, 58 ^{o}C for 90 seconds and 72 ^{o}C for 90 seconds were conducted. A final extension step was then carried out at 72 ^{o}C for 10 minutes.

Amplicon purification

PCR amplicons were purified prior to sequencing using the QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer's guidelines. DNA was eluted in 50 μl 10 mM Tris-Cl. For analysis of PCR performance, 5μl of purified DNA was mixed with 1.5 μl 5x loading buffer (Bioline), and loaded into a 2% agarose gel stained with ethidium bromide. For standardisation, 5 μl of HyperLadder IV (Bioline) was also loaded. Gels were run for 30 minutes at 120 Volts before being photographed with a G:BOX F3 transilluminator (Syngene). The supplied GeneTools software was used to analyse the gel. Band intensities were normalized based on

the DNA ladder, and the mass of loaded PCR product quantified. Prior to sequencing library preparation, purified PCR product was quantified using a Qubit flurometer (Invitrogen) with the high-sensitivity DNA kit, and normalised to 10 ng/μl.

2.1.12 Illumina sequencing

Library preparation and sequencing was carried out by the High Throughput Genomics Group, Wellcome Trust Centre for Human Genetics, Oxford, UK. Amplicons were end-repaired, A-tailed and adaptor ligated prior to size-selection and amplification. Samples were multiplexed depending on depth required, and sequenced using the 2x300 bp chemistry on the MiSeq (Illumina, San Diego, CA). Data alignment and initial quality control was conducted using the supplied MiSeq software.

2.2 Computational methods

A flexible computational pipeline was developed for the analysis of high-throughput BCR repertoire sequence data, as summarised in Figure 2.4 and below.

2.2.1 Raw sequence processing

Paired-end reads were joined to give a continuous sequence spanning the entire amplicon using fastq-join (ea-utils). Where there were fewer than 6 overlapping nucleotides, or more than 8% sequence difference in the overlapping region, the reads were removed. Quality filtering was then performed to remove any sequences containing unknown nucleotides, or with a Phred quality less than 30 over more than 15% of bases. Filtered data files were then converted to FASTA format, and submitted to IMGT/HighV-Quest for annotation (130). Sequences defined as unproductive (out of frame, or have a stop codon) by IMGT were removed.

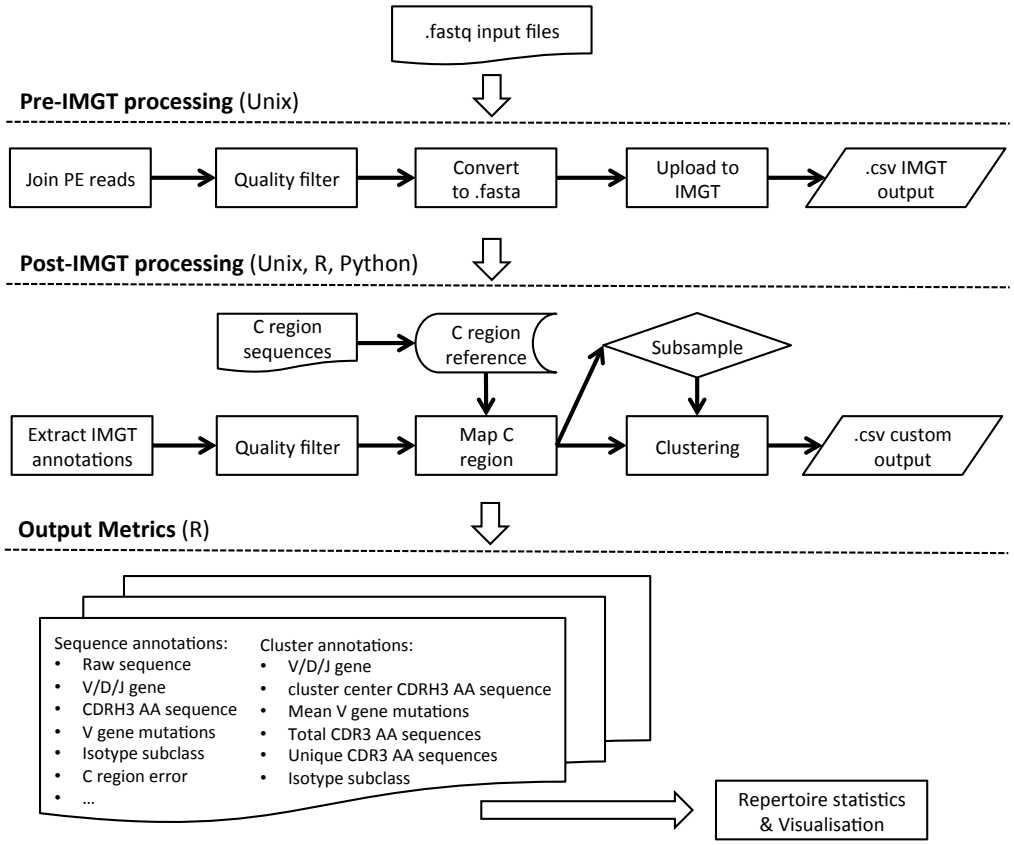

Figure 2.4: Sequence processing pipeline. - Raw sequence processing is split into two sections: pre-IMGT processing, and post-IMGT processing. Once raw sequence processing is complete, the sequence-level and cluster level output metrics can be determined.

2.2.2 Sequence-level annotation

IMGT output parsing

The IMGT output is extensive, comprising 11 excel spreadsheets, each with multiple annotations for each sequence. These data were selectively imported into the R software environment (175) for analysis. For each sequence, the annotations of functionality (productive/unproductive), V gene segment alignment, D gene segment alignment, J gene segment alignment, the number of nucleotide mismatches from germline in the V gene segment (mutation number), the amino acid sequence of the CDR3 region, and the nucleotide position within the sequence that marks the end of the variable region, and the start of the constant region were extracted from the IMGT output.

Assigning isotype subclass

The nucleotide position within the sequence that marks the end of the variable region was used to extract the constant region section of each sequence. This sequence was then mapped to a reference table of each constant region isotype subclass sequence using Stampy (176). The reference table was created using constant region sequence locations from the Vertebrate Genome Annotation database (177), and nucleotide sequences derived from the human_g1k_v37 reference genome. PCR and sequencing error rates were estimated from the number of nucleotide mismatches from the reference in the constant region, as these regions are not subject to SHM (178).

Finding nearest neighbours

For each read, CDR3 AA distance to its nearest neighbour was calculated by comparing it with all other reads from the same sample with the same CDR3 length. The read with the fewest AA mismatches in the CDR3 region is termed the nearest neighbour, and the distance is the number of AA differences in the CDR3.

2.2.3 Repertoire-level annotation

Mean CDR3 AA length, and number of V gene mutations were calculated for the entire repertoire of each sample. In addition, the relative proportion of the repertoire comprised by sequences using different V and J genes (and VJ combinations), and of different isotype subclasses was calculated for each sample.

2.2.4 Sequence clustering

Related reads were clustered together using parameters optimised to form groups of sequences that are either clonally related, or differ due to PCR and sequencing error. To be considered part of the same cluster, reads were required to have the same V and J gene segment annotation, the same length CDR3, and a similar CDR3 AA sequence (different thresholds were trialled,

between 75% and 96% identity). The same D gene annotation was not required for inclusion in the same cluster, as somatic hypermutation in this region makes accurate assignment of a gene segment difficult (179). We focused on the CDR3 region of the sequence, as it is highly variable, and has the dominant role in determining antigenic specificity of the sequence (180). Clusters were iteratively defined using an approach to identify cluster centres that gave the largest possible clusters. Briefly, clustering started with a set of unique sequences of the same length U(L) and an empty set of clusters C(L). The first cluster centre is defined as the sequence x∈U(L) that has the most neighbours |N(x)|; the set N(x) is then added to C(L), with cluster centre x. N(x) is then removed from U(L), and the process repeated until U(L) is empty and C(L) is full.

For each cluster, the 'cluster centre' was used to represent the CDR3 AA sequence. V, D, and J gene segment composition, and isotype subclass of the sequences within each cluster was also conserved. The average V gene mutation of sequences within each cluster was also determined. The number of total, and unique sequences within each cluster was also measured. Nearest neighbour distances were also calculated for cluster centre sequences in the same way they were calculated for the individual sequences.

2.2.5 Cluster-level annotation

Clusters were annotated for having potential binding specificity towards certain antigens based on comparison to either previously described sequence datasets, or datasets obtained as part of this thesis. For annotation of clusters based on similarity to previously described sequence datasets, CDR3 AA sequences known to be specific for certain antigens were obtained. The comparison of clusters to these known sequences used CDR3 AA sequence identity only, and was based on whether the known sequence would have fallen into that cluster during the clustering process. Note that a single previously described sequence could potentially be included in

multiple clusters, so when this is the case, all of the clusters are annotated as having potential specificity towards the antigen.

For annotation of clusters based on similarity to datasets obtained as part of this thesis, the antigen-enriched datasets were processed to obtain the CDR3 sequence, and V and J gene segment annotation of each sequence. The comparison of clusters to these antigen-enriched sequences then used CDR3 AA sequence identity, and V and J gene segment usage, and was also based on whether the antigen-enriched sequence would have fallen into that cluster during the clustering process.

Size estimates

To determine adequacy of sequencing depth, rarefaction analysis was conducted using the Vegan R package (181), with individual clusters representing species, and sampling done without replacement. Extrapolation of the rarefaction curves to give an estimation of species richness was conducted based on Chao's estimates (182), using the iNEXT package in R (183) with a q value of 0.

To estimate the effective cluster population size of the total repertoire, capture-recapture analysis was performed on biological replicate samples using the Chapman-Estimator formula (25). Clusters in total repertoire $= \frac{(S1+1)(S2+1)}{C+1} - 1$, where S1 is the total number of clusters in the first biological replicate, S2 is the total number of clusters in the second biological replicate, and C is the number of clusters present in both replicates.

Diversity measures

Repertoire diversity was calculated using three different single diversity metrics, which have previously been applied to repertoire studies. The Shannon index ($H' = -\sum_{i=1}^{S} p_i \ln p_i$) and Simpson index ($D_1 = 1-\sum_{i=1}^{S} {p_i}^2$) are derived from ecology and take into account species richness and abundance. For the purpose of studying the BCR repertoire, each cluster is considered

a distinct species. So, p_i is the proportion of the repertoire comprised by cluster i, and S is the total number of clusters. The Shannon index gives more weight to rare clusters while the Simpson index gives more weight to abundant clusters. In addition, a clonality index derived from cryptanalysis (the study of text-based cyphers) was used, which measures the probability that sequences selected from different PCR replicate samples belong to the same cluster (143). This clonality index was calculated as $\sum \frac{N_{ij} \times N_{ik}(j!=k)}{i,j,k} \div \sum \frac{T_j \times T_k(j!=k)}{j,k}$, where N_{ij} and N_{ik} are the total number of sequences within cluster i in the replicate libraries j and k, and T_j and T_k are the total number of sequences in the two replicate libraries

In addition to the use of single diversity metrics, diversity profiles were also determined using the framework developed by Greiff *et al.* (135). Here, diversity is calculated using the Renyi index, $H_\alpha = \frac{1}{1-\alpha} log \sum_{i=1}^{S} p_i^\alpha$, where as before, p_i is the proportion of the repertoire comprised by cluster i, and S is the total number of clusters. Changing the alpha value changes the weight given to abundant and rare clusters in the diversity calculation. Thus, using multiple values of alpha allows a diversity profile ($H_{\vec{\alpha}}$) to be created, where as alpha increases, more weight is given to the abundant clusters. Notably, the Shannon and Simpson indices can be related to the Renyi index, as $H'=H_{\alpha=1}$ and $D_1=H_{\alpha=2}$. Diversity profiles from different samples were compared based on Euclidean distance, and clustered using the complete linkage algorithm with the hclust function in the Stats R package (175).

Lineage trees

Lineage trees were constructed to show the clonal relationship between sequences within a cluster using the Alakazam R package (184). This constructs trees using the maximum parsimony model with the dnapars application of the Phylogeny Inference Package (PHYLIP). Trees are rooted to the inferred germline sequence of the cluster, and V and J segment DNA sequences only rather than the entire variable region used, as the germline sequence can not be reliably determined from the D segment. PHYLIP infers the sequences of common ancestors to

branches on the tree. When sequences in the dataset are identical to their inferred parent, they are moved up the tree to give internal nodes. Lineage trees were visualised using the iGraph R package (185).

2.2.6 Statistical analysis and graphing

All statistical analysis was conducted using the R software environment (175). Ggplot2 (186) was used for constructing graphs, and Circos (187) used for constructing circular plots. Principal component analysis was conducted using the prcomp R function in the Stats R package (175). T-tests or Mann-Whitney U tests were used where appropriate to compare groups. Correlations were calculated with Spearman or Pearson correlation coefficients where appropriate. Correlations were graded as low (r values between 0.2-0.39), moderate (0.4-0.59), strong (0.6-0.79), and very strong (0.8).

3

Assessment of the reproducibility of BCR repertoire sequencing, and variation in the repertoire over time and between individuals

3.1 Introduction

The vast diversity of the BCR repertoire has made it difficult to study, but NGS technology now makes it possible to capture a high-resolution snapshot of the circulating BCR repertoire in humans. BCR repertoire analysis has been used to increase understanding of the fundamental properties of B cells, including developmental processes (188), and responses to antigen (119, 124, 146, 149, 150, 152). In addition, a number of clinical applications are beginning to emerge, including identification of autoimmune irregularities (137), monitoring of B cell lymphoma minimal disease residue (23, 189), disease diagnostics (118) and the rapid identification of mAb sequences (155). These studies generally monitor global features of the BCR repertoire, such as diversity, mutation levels, isotype subclass usage and VDJ segment usage frequency as well as identifying specific B cell clones. Identifying B cells arising from the same clonal origin was initially conducted based on identifying common VDJ segment usage (23), but there is now a move towards incorporating CDR3 sequence identity into the definition (118, 188),

which forms at the junction of VDJ joining and is the most important region for determining antigen-binding properties (190).

B cell samples for repertoire studies in humans are usually obtained from peripheral blood, as this is an easy to sample compartment, but it is not possible to extract all peripheral blood from humans, and some B cells may be present in different compartments. This means that the entire repertoire cannot be sampled, so instead a representative sample is taken. It is therefore important to quantify exactly what proportion of the repertoire is being sampled, and how repeatable the sampling is. This is particularly pertinent for clinical applications where it is necessary to have a highly repeatable measure to detect the clones of interest. As BCR repertoire sequencing is a young technology, such repeatability studies have been done in mice (172), but not exhaustively in humans, who have a much larger repertoire (22). Many studies of the BCR repertoire also assess changes in various features of the repertoire over time in response to certain interventions (119, 146, 149, 150). However, there is relatively little known about how the much the repertoire naturally fluctuates over time in the absence of any specific intervention, making it difficult to discern natural fluctuations from intervention-induced fluctuations. Furthermore, as the total number of healthy individuals who have had their repertoire sequenced remains small, there is not a clear consensus of what can be construed as a 'normal' repertoire. It is uncertain to what extent similar B cell clones can be found in multiple individuals (the public repertoire) not undergoing a similar immune stimulus, and whether sharing is just due to chance, or due to historic expansion of similar B cells in different individuals from a common antigen and thus has some clinical significance (112, 118, 124, 149).

To shed light on these questions, repeat sampling from a single individual was carried out to assess the robustness of a BCR repertoire sequencing protocol, and also to determine how the repertoire changed over time in the absence of any intervention (Figure 3.1). Furthermore, sequencing of the BCR repertoire from 9 additional individuals was conducted to determine

how variable the repertoire was between them, and to interrogate properties of the public repertoire. In addition, as a database of BCR sequences with known specificity would be of benefit to interpreting BCR repertoire data, and uncovering the applications of BCR repertoire analysis (see Section 1.4.3), creation of such as database was initiated.

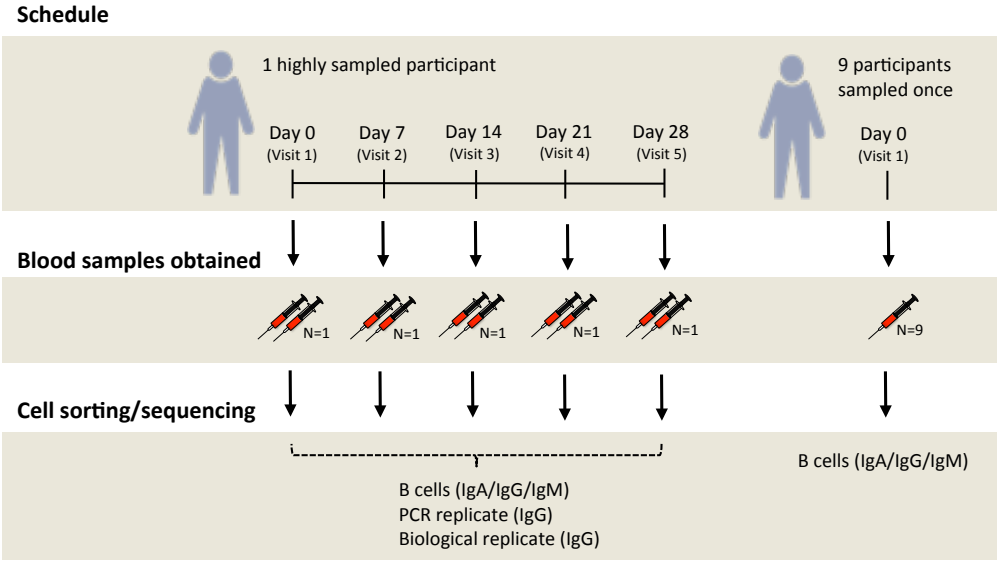

Figure 3.1: Study Design - For assessment of within-individual variation, one participant (participant AF01) had blood sampled at 5 timepoints one week apart (temporal replicates). For this participant, two aliquots of B cells were taken at each timepoint to give biological replicates. IgA, IgG and IgM isotype-specific PCRs were all conducted for one of the biological replicates, but only IgG for the second biological replicate. Additionally, for this participant, IgG PCR replicates were also conducted at each timepoint for one of the biological replicates. For assessment of inter-individual variation, 9 additional participants were sampled at a single timepoint, and had IgA, IgG and IgM PCRs conducted.

3.2 Methods

3.2.1 Study design

Healthy participants were recruited with informed consent in accordance with the Declaration of Helsinki, and under approval from the Northampton Research Ethics Committee (13/EM/0036). For the assessment of inter-individual variation, 10 participants had 50 ml of blood sampled at a single timepoint. These samples represented the baseline samples from the study discussed in Chapter 4. In addition, to assess within-individual variation, and repeatability of the protocol, there was a single highly sampled participant, who had 50 ml of

blood sampled on 5 consecutive weeks (day 0, 7, 14, 21 and 28) to give temporal replicates (Figure 3.1).

3.2.2 Sample processing

For each blood sample, total B cells were isolated by MACS, and an aliquot of 500,000 B cells used for BCR repertoire sequencing. An additional 500,000 B cell aliquot was also taken at each day from the highly sampled participant to give biological replicates. PCR was conducted independently for IgA, IgG and IgM transcripts for every sample, except the biological replicates, where only IgG reactions were performed. In addition, for one of the aliquots at each timepoint from the highly sampled participant, the IgG PCR was repeated to give PCR replicates. Sequencing was performed with 40-60 samples being multiplexed on each run.

3.2.3 Sequence processing

Sequences were processed using the pipeline described in Figure 2.4. To account for differences in the number of resulting sequences in different samples, all samples were randomly subsampled without replacement using the sample function in R to give 100,000 sequences per sample. To determine whether a cluster is present in more than one sample, data from all samples were clustered together. If two samples contribute at least one sequence to the same cluster, that cluster is defined as being present in both samples.

3.2.4 Generating a database of previously described antigen-specific sequences

A literature search was performed to identify previously described IgH BCR sequences with confirmed (by ELISA and/or surface plasmon resonance) specificity towards Influenza, TT and HBsAg. Where full-length sequences were provided, these were submitted to IMGT (130) to determine the V and J gene annotation, and CDR3 AA sequence. In some publications, only the CDR3 AA sequence, or the CDR3 AA sequence and V and J gene annotation was given,

not the full length sequence. As different definitions of CDR3 are used (see Section 1.4.4), this meant that in some cases, only truncated CDR3 AA sequences could be obtained, which are not comparable to the rest of the sequences. Truncated sequences had the conserved V and J region nucleotides manually added to either end so that they conformed to the IMGT definition of a CDR3 sequence.

3.3 Results

3.3.1 Determining the effect of PCR/sequencing error

Repertoire data were successfully obtained for all 52 samples (Table 9.1). The mean number of raw sequences per sample was 367,634 (216,878 - 1,516,275). Quality filtering removed on average 31% of raw sequences, leaving at least 100,000 sequences per sample for subsequent analysis. Sequences extended into the constant region on average 104 nt for IgA, 124 nt for IgG and 57 nt for IgM (Figure 3.2 A). This was sufficient to allow mapping to the constant region reference sequence to determine isotype subclass for all sequences in the dataset. The constant region is not subject to SHM (178), so assessing the number of mismatches from the reference in the constant region sequence can be used to give a rough estimate of PCR and sequencing error. Error rate estimates differed depending on the isotype of the sequence, and were 0.0021, 0.0079 and 0.0019 errors per nucleotide for IgA, IgG and IgM sequences respectively (Figure 3.2 B).

To determine the effect of this error on the resulting data, the nearest neighbour distribution of the CDR3 AA sequences was determined (Figure 3.2 C). This revealed a bimodal distribution: the first peak of sequences had a close neighbour 0-2 AAs away, and the second peak of sequences had a more distant neighbour 3-15 AAs away. The first peak was higher for IgA and IgG compared to IgM sequences, and the position of the second peak was shifted 1 AA towards the y-axis for the IgM sequences. It is therefore likely that the first peak contains sequences whose

nearest neighbour is either clonally related or differs due to error and the second peak likely contains sequences whose nearest neighbour arises from a distinct B cell clone. To correct for this error, we can therefore use a clustering approach to group together sequences which have their nearest neighbour in the first peak. Each cluster will then represent a group of sequences all derived from the same B cell clone, and will also group together sequences arising from error.

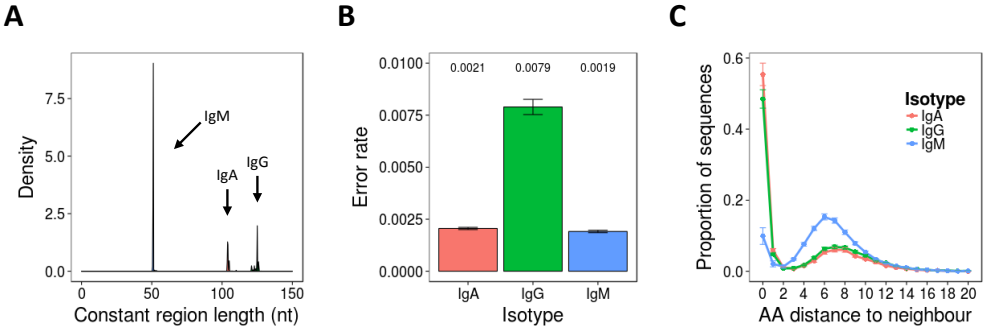

Figure 3.2: Error estimation. - (**A**) The length distribution of constant region sequences available for determining error of samples amplified using primers for the different isotypes. (**B**) Mean per nucleotide error rates of sequences obtained from each sample. For each sequence, error is calculated as number of mismatches in the constant region from germline divided by the length of the captured constant region sequence. (**C**) Nearest neighbour distributions of CDR3 AA sequences. The nearest neighbour of each sequence in each dataset is determined by comparing it to every other sequence of the same length in the dataset to find the closest match - that is its nearest neighbour. The distance is then the number of AA difference between the sequence and its nearest neighbour. For **B** and **C**, mean ±SEM values are plotted.

3.3.2 Using clustering to define clonally related sequences

Sequences were clustered together that had the same V and J gene segment annotation, and a similar CDR3 AA sequence. To find the threshold that best grouped together sequences with their nearest neighbour in the first peak, but not those with their nearest neighbour in the second peak, clustering was performed allowing different degrees of mismatch in the CDR3 AA sequence. The clustering thresholds ranged from allowing 1 AA mismatch per every 4 AA's (i.e, ≥75% similarity) to allowing 1 AA mismatch per every 26 AA's (≥96% similarity). Determining the nearest neighbour distribution of cluster centre CDR3 AA sequences following clustering then makes it possible to determine which thresholds are successful at grouping

together sequences with a neighbour in the 0-2AA peak. As the threshold stringency increases, a peak of sequences with a neighbour 2 AA's away appears, indicating that the most stringent threshold that can be used effectively is allowing 1 AA mismatch per every 12 AA's (Figure 3.3 A). When the threshold drops below 1 AA mismatch per 10 AA's, there is a sharp drop in the number of clusters formed (Figure 3.3 B), indicating that sequences from unrelated cells start to be grouped together. A threshold of 1 AA mismatch per 12 AA's was therefore chosen for the final analysis as the most stringent threshold that can effectively group together related sequences with a neighbour in the first peak, while limiting the risk of grouping unrelated sequences.

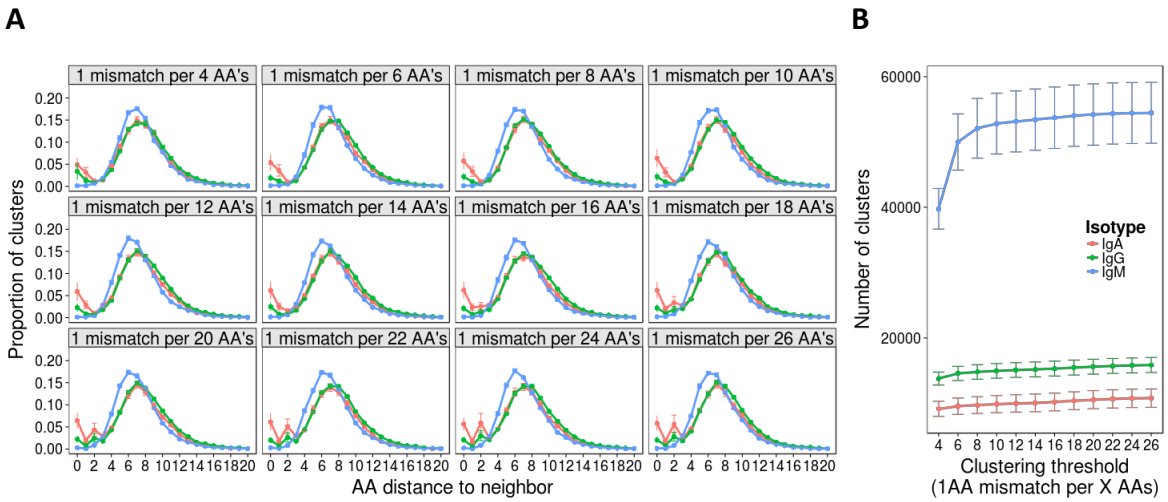

Figure 3.3: Trialling different clustering thresholds. - (**A**) Nearest neighbour distribution of cluster centre CDR3 AA sequences after clustering with different thresholds of mismatch allowed. The nearest neighbour of each sequence in each dataset is determined by comparing it to every other sequence of the same length in the dataset to find the closest match - that is its nearest neighbour. The distance is then the number of AA difference between the sequence and its nearest neighbour. (**B**) The number of clusters formed following clustering with different thresholds of mismatch allowed. Mean ±SEM values are plotted.

Following final clustering with the threshold of 1 AA mismatch per 12 AAs, the mean number of clusters differed for the different isotypes, and was 9,972 for IgA (1,658 - 22,940), 15,080 for IgG (1,599 - 25,000), and 53,150 for IgM (19,000 - 75,420) (Table 9.1). Following clustering, the amplitude of the first peak of the nearest neighbour distribution of cluster centre CDR3 AA sequences was greatly reduced for IgA and IgG sequences, and completely removed

for IgM sequences (Figure 3.3 A). Cluster sequences with a distance of 0 to their neighbour represent clusters with the same CDR3, but different V and/or J gene annotations. Although such sequences could be chimeras formed during the PCR reaction, they are not present in the IgM dataset, making this unlikely, and were thus retained for analysis. The size distribution of clusters was highly uneven, with a small number of abundant clusters, and a large number of rare clusters (Figure 3.4). Based on this, clusters were defined as abundant if they contained at least 10 sequences (i.e., comprised at least 0.01% of the total sequenced repertoire). Datasets contained mean 930 (518 - 1,537), 835 (195 - 1,983) and 642 (195 - 1,983) abundant clusters for IgA, IgG and IgM respectively. As each individual cluster can be considered to represent a distinct B cell clone, these abundant clusters will likely represent activated B cell clones - i.e, PCs which have high transcript levels, or proliferating cells where there will be many cells with the same BCR sequence.

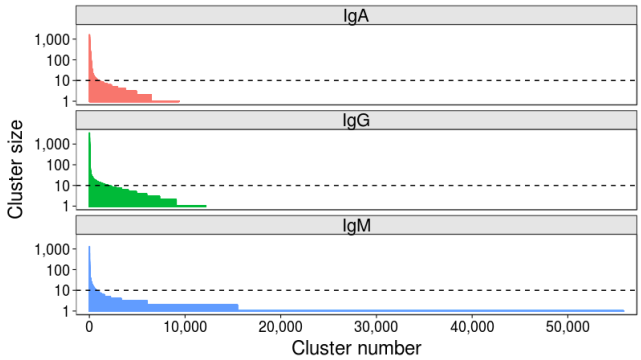

Figure 3.4: Cluster size distribution. - Clusters are ordered according to size, and the size of each cluster plotted. Representative data from one sample is shown (participant AF01, day 0). Horizontal dotted line intersects the y-axis at 10 sequences (0.01% of the total sequenced repertoire), and represents the cutoff between rare and abundant clusters.

3.3.3 Quantitative assessment of sequencing depth

Two methods were used to assess the adequacy of the sequencing depth used: rarefaction analysis, and comparison of the PCR replicates. Rarefaction is a technique used in ecology to estimate species richness; in the context of the BCR repertoire, each cluster is defined as

a unique species. Random samples of increasing size were taken from the total dataset to determine the number of clusters that were represented at increasing sequencing depths, and a curve drawn to show the number of clusters represented as a function of sequencing depth (Figure 3.5). As the curve plateaus, it indicates that sequencing depth is sufficient, and only the very rare clusters remain to be identified. The curves for IgA and IgG do show the beginning of a plateau, so it can be inferred that a sequencing depth of 100,000 is sufficient to capture most of the abundant clusters for these samples. For IgM samples, which had a much larger number of rare clusters (Figure 3.4), the curve does not begin to plateau, so it is unlikely that the full diversity of this population is being captured. If the curve does not plateau, it can be extrapolated to estimate where the plateau would occur, and give an estimate of the effective population size (the total number of clusters) of the population. Using the extrapolated curves to estimate total number of clusters for the different samples gives a mean value of 15,414 total clusters for IgA, 19,488 for IgG and 143,731 for IgM, indicating that our sequencing depth captures approximately 63%, 64% and 36% of all IgA, IgG and IgM clusters respectively. In order to capture 90% of clusters contained in a sample of 500,000 B cells, it would therefore be necessary to obtain approximately 280,000 sequences for IgA, 260,000 for IgG and 620,000 for IgM.

PCR replicates were only conducted for IgG samples, so comparison of these can only be used to assess sequencing depth for this isotype. Although simply re-sequencing the same library can also be used to assess sequencing depth, conducting the PCR again prior to sequencing is a more stringent measure that will also take into account differences in amplification efficiency of the template cDNA. Across the five samples where PCR replicates were available, the mean overlap of clusters present in both PCR replicates was 58%, but this increased to 95% when just considering the abundant clusters (Figure 3.6 A). In addition to identification of specific sequences, many repertoire studies also assess relative proportions of different VJ

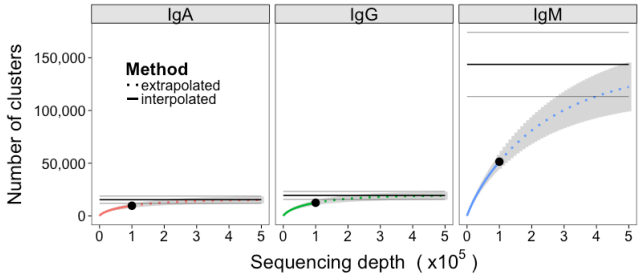

Figure 3.5: Extrapolated rarefaction curves for clusters of each isotype. - Rarefaction analysis (interpolated) was conducted by subsampling data without replacement at 1,000 sequence increments, and determining the number of clusters represented by these sequences. Cluster richness estimation of the sample (extrapolation) was based on Chao's estimator formula, and conducted up to a sequencing depth of 500,000 sequences. The curve shows the number of clusters identified as a function of sampling depth - as the curves start to plateau it indicates that increased sampling depth will yield few additional clusters, and sampling depth is sufficient. Results were calculated from samples from all 10 participants, and mean ±SEM (grey curve) plotted. The three horizontal grey lines show the mean ±SEM for the Chao estimate of total clusters in the sample.

combinations. To determine the reproducibility of this, the proportional representation of each VJ clone in the repertoire was determined for each sample, and correlated between the replicates. Strong correlations were seen, with Pearsons r $>$ 0.997 for all PCR replicate samples (Figure 3.6 B).

3.3.4 Quantitative assessment of sampling depth

To assess the adequacy of sampling 500,000 B cells for assessment of the total BCR repertoire, the biological replicates were used to estimate a lower bound for the total IgG BCR repertoire size using capture-recapture analysis. Across the five samples where biological replicates were available, the mean IgG effective repertoire size was estimated as 142,576 unique clusters; as a mean of 15,128 total IgG clusters were obtained from these samples, approximately 11% of the total repertoire was sampled at each timepoint (Table 3.1). Taking just the abundant clusters, the mean estimated BCR repertoire size of abundant clusters was 4,644, indicating that approximately 25% of the abundant repertoire was sampled at each timepoint. Estimates of repertoire size varied on the different days of sampling, and this variation in size estimates was more pronounced for the abundant repertoire (up to 8.9x size difference between days)

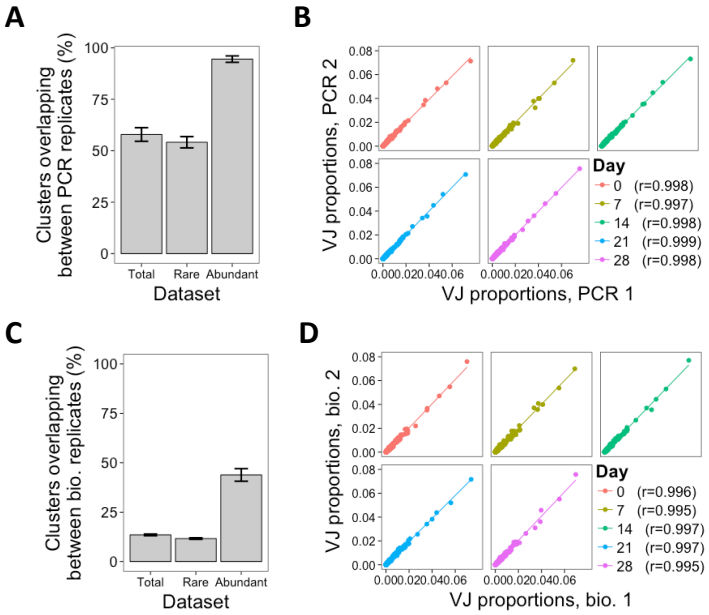

Figure 3.6: Overlap of biological and PCR replicates. - (**A**) For the PCR replicates, the percent of clusters present in both replicates was determined, where percent = (A∩B/sum(A,B))*100. This was determined for total clusters, rare clusters (<10 sequences), and abundant clusters (≥10 sequences) for all 5 samples, and mean ±SEM values plotted. (**B**) Correlation in proportion of the total repertoire comprised by each VJ gene combination in the two PCR replicates at each day. Each point represents the proportional representation of a particular VJ combination in the repertoire. R-values represent Pearsons correlation coefficients. (**C** & **D**) Same as **A** and **B**, but comparing the biological replicates.

compared to the total repertoire (up to 1.5x size difference between days).

Table 3.1: Repertoire size estimates - Number of total clusters, and abundant clusters sampled at each timepoint from participant AF01, size estimates of the total IgG and abundant IgG cluster repertoire based on capture-recapture analysis of biological replicates, and the percent of the total cluster repertoire, and abundant cluster repertoire that we captured.

	Total clusters			Abundant clusters		
Day	**No. sampled**	**Size estimate**	**% sampled**	**No. sampled**	**Size estimate**	**% sampled**
0	12,152	132,841	9.1	1,919	10,426	18.4
7	13,186	108,253	12.2	367	1,168	31.4
14	16,907	162,868	10.4	1,137	5,368	21.2
21	17,930	158,311	11.3	861	3,013	28.6
28	15,463	150,605	10.3	775	3,245	23.9
Mean	15,128	142,576	10.7	1,012	4,644	24.7

As with the PCR replicates, the overlap in clusters present in both of the biological replicates was also calculated (Figure 3.6 C-D). The mean overlap of total clusters present in both biological replicates was 14%, although this increased to 44% when considering the abundant clusters only. The reproducibility of VJ usage frequency remained strongly correlated between

the biological replicates, with Pearson's r > 0.995 for all samples.

3.3.5 Fluctuations in the repertoire over time

For a single participant, individual clusters were tracked across the samples collected at different times to see if they could be detected on multiple days. Although most clusters were present on just a single day, 5%, 14% and 7% were detected on more than 1 day for IgA, IgG and IgM respectively (Figure 3.7 A). In addition, there were a small number of clusters detected at all timepoints. Circos plots were constructed to show the relationship between the clusters present on different days, and show that it is primarily the abundant clusters that are present on more than one day (Figure 3.7 B).

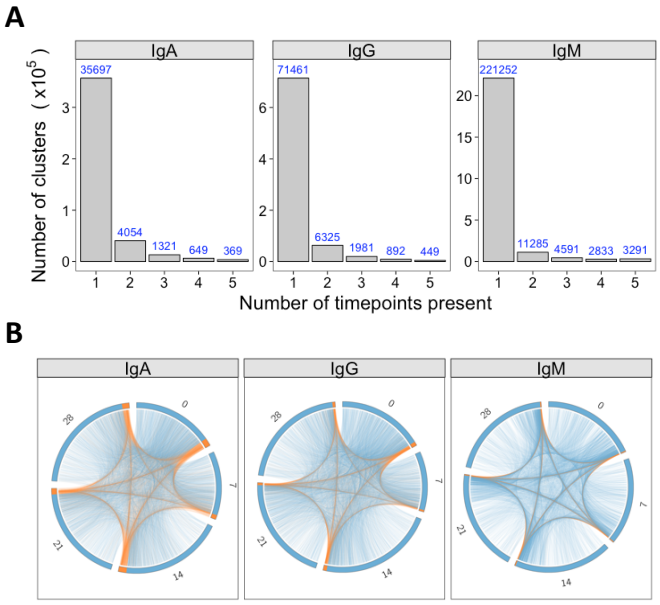

Figure 3.7: Persistence of clusters over time. - (**A**) For samples of each isotype from participant AF01, the total number of clusters across all timepoints was determined, and the number of these present at a different number of timepoints calculated. (**B**) Using the same data as **A**, circos plots were constructed to show the clusters present on different days. Each arc represents a different day, and clusters are ordered clockwise by abundance, with the abundant clusters coloured orange, and the rare clusters coloured blue. Lines join clusters present at more than one timepoint.

In addition to monitoring individual clusters, global repertoire metrics were also calculated to determine how they changed over time. Common global repertoire metrics that are used to give insight into B cell immunology include VJ segment usage, diversity, mutation, CDR3

length, and isotype subclass usage. A variety of diversity indices are interchangeably used to measure diversity of BCR repertoire datasets (143, 188); here, three different methods were used, but they all gave the same trend (Figure 3.8 A-C). There was a dramatic decrease in diversity at the day 7 timepoint, which was most pronounced in the IgG dataset. The use of diversity profiles has recently been reported as a more accurate method for determining immunological status from BCR data than the use of single diversity measures (135), and these results are shown in Figure 3.8 D. The difference between the diversity profiles of different samples can then be calculated, and used for hierarchical clustering of the samples based on diversity. This revealed that all day 7 samples clustered together regardless of isotype. All IgM samples also clustered together, but the distinction between IgA and IgG samples was not as clear. To determine whether the decrease in diversity at day 7 was due to a degree of immune activation, signatures of clonal expansion were investigated. Although there were actually fewer abundant clusters at day 7 than on the other days, the size of these abundant clusters tended to be greater (Figure 3.9), suggestive of clonal expansions occurring at this timepoint.

Although most clusters do not persist over time, and there is evidence of clonal expansion at day 7, VJ usage frequency remained highly correlated between all days, although was slightly reduced when comparing the day 7 sample to the other days (Figure 3.10 A). When calculating repertoire metrics such as VJ usage frequency, it is possible to correct for cluster size, by weighting clusters according to the number of sequences they contain. Performing such an adjustment can give more insight into features of the activated B cell clones. This reduces the correlation in VJ usage between the different days, and makes the day 7 difference more pronounced. When considering the metrics of mutation, CDR3 length, and isotype subclass usage, these are also fairly steady over time (Figure 3.10 B-E), but when adjusting for cluster size, there was an increase in mutation, an increase in CDR3 length, and a relative decrease

in IgG1, and increase in IgG3 usage at day 7. Despite these changes, the metrics were highly conserved between the PCR and biological replicates.

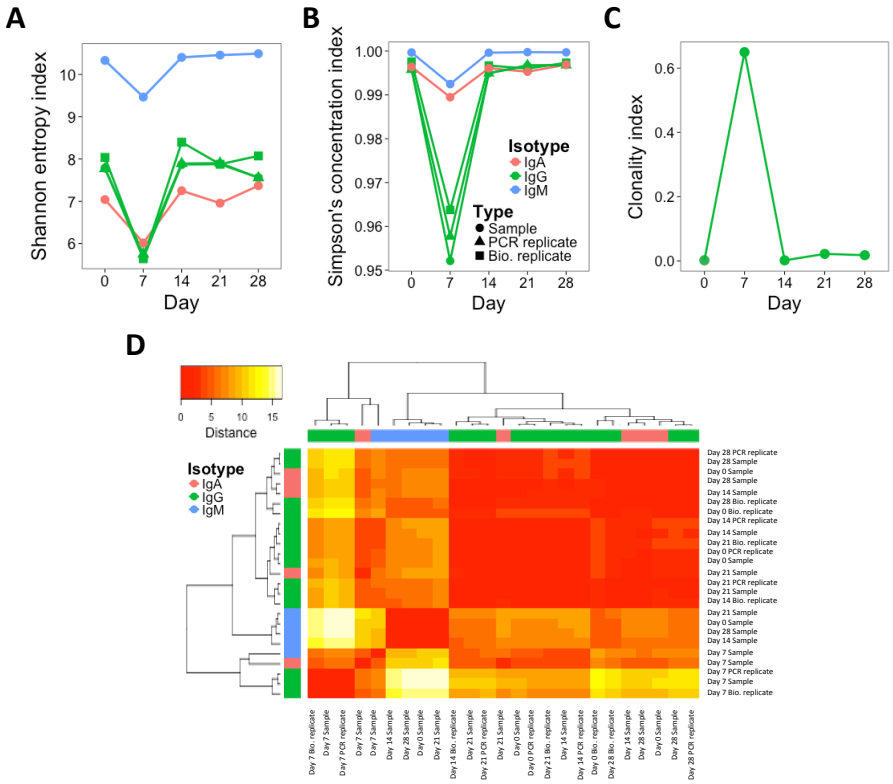

Figure 3.8: Comparison of diversity indices. - (**A**) Shannons entropy and (**B**) Simpsons concentration indices calculated from all samples from participant AF01. (**C**) Clonality index calculated from IgG sample from participant AF01 calculating this index requires biological replicates, so cannot be done for the other isotypes. (**D**) Euclidean distances between the diversity profiles of each sample were calculated, used for hierarchical clustering of the samples, and visualized as a heatmap. The colour of the heatmap indicates the similarity in the profile between two samples (red=high similarity, white=low similarity). The colour bars indicate the isotype of each sample.

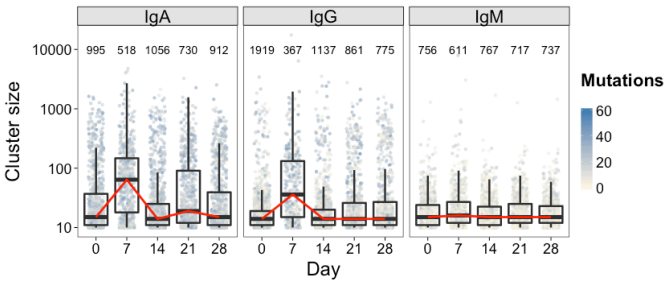

Figure 3.9: Size distribution of abundant clusters in participant AF01. - Every abundant cluster present at each day is plotted as a point (total number is above each day), and jittered to prevent over-plotting. Boxplots show locations of 25, 50 and 75th percentiles of cluster size, and whiskers extend to maximum values that lie within 1.5x the interquartile range above and below the quartiles. Red line connects median value at each timepoint.

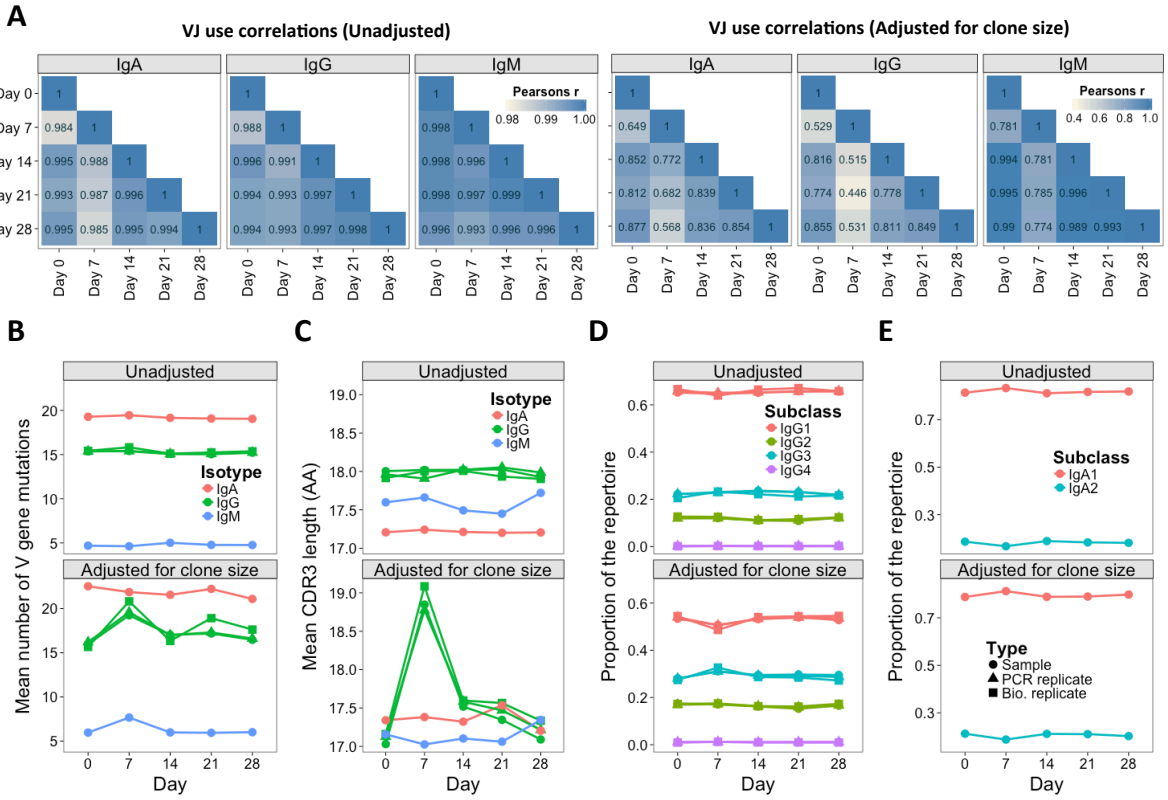

Figure 3.10: Persistence of repertoire metrics over time - (**A**) Correlation in relative usage frequency of each VJ gene segment combination from samples on different days from participant AF01. Stronger correlations are in darker blue. (**B**) Change in the mean number of V gene mutations, (**C**) mean CDR3 AA length, (**D**) IgG subclass use, and (**E**) IgA subclass use in the repertoire over time. For **A**-**E**, measurements were also calculated taking cluster (clone) size into account, thus giving more weight to the larger clusters.

3.3.6 Inter-individual variation in the repertoire

As well as the single highly sampled participant, 9 additional participants were sampled at a single timepoint to give insight into inter-individual variation in the repertoire. Although at the cluster sequence level, there was very little overlap between the repertoires of different participants (see Section 3.3.7), the repertoires of different participants were more comparable at the level of the general repertoire properties. Different V and J gene segments are not used in even proportions, and these biases in gene use are conserved between different participants (Figure 9.1). To more specifically assess any differences in VJ usage between participants, principle component analysis was used. This is a dimensional reduction technique that takes into account independent correlations in usage frequencies between variables (VJ combinations)

to give components that can explain the largest proportion of the total variability in the data. The first two principle components account for the largest proportion of variability, and in the case of VJ usage, account for approximately 40% of the variability. Plotting the samples according to the first two principle components shows that the samples from different participants cluster apart from each other compared to the samples on different days from participant AF01 which cluster closely together, indicating that VJ usage frequency is steady over time, and able to uniquely identify the different participants, despite apparent immune activation (Figure 3.11 A).

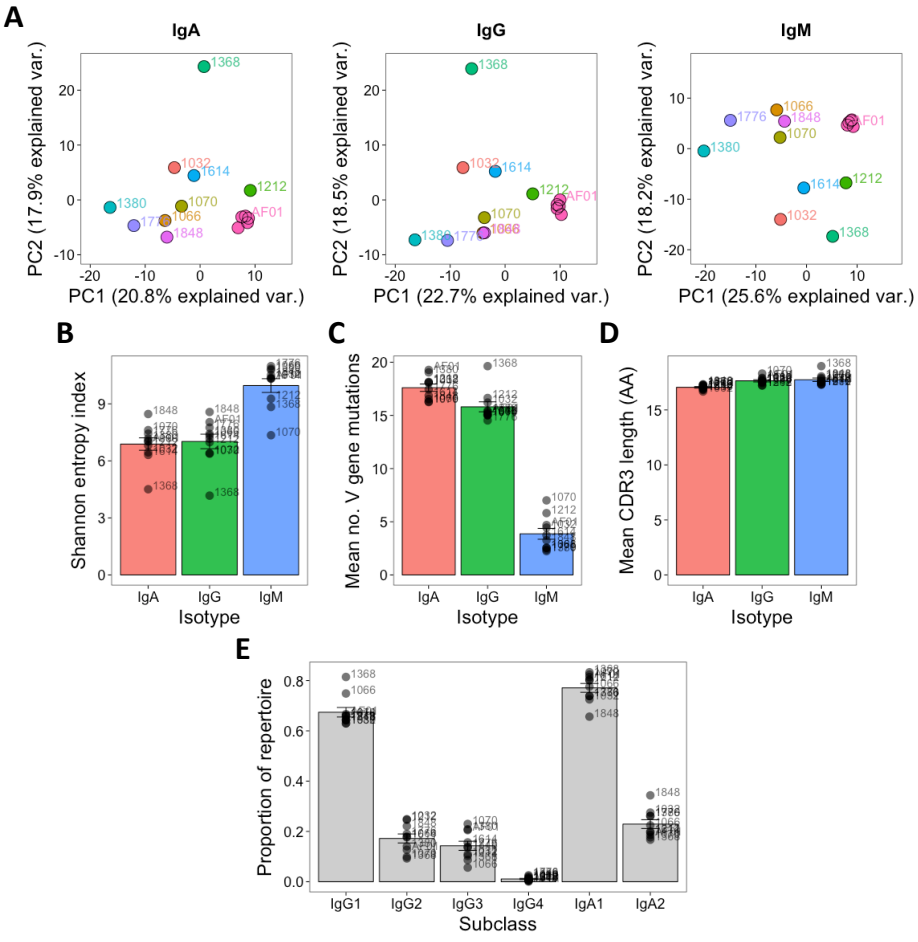

Figure 3.11: Inter-individual variation in global repertoire properties. - (**A**) Principal component analysis of VJ segment usage in each participant. Five samples from participant AF01 are included, each corresponding to a sample from a different day. Differences in repertoire diversity (calculated using the Shannon entropy index) (**B**), mean number of V gene mutations (**C**), mean CDR3 AA length (**D**), and proportion of the repertoire comprised by sequences of each IgG and IgA subclass (**E**) in each participant. For **B**, **C**, **D** and **E**, bars show mean values ±SEM.

The global repertoire properties of diversity (Figure 3.11 B), V gene mutation (Figure 3.11 C), CDR3 AA sequence length (Figure 3.11 D) and isotype subclass usage (Figure 3.11 E) were also determined for each participant. There was considerable variation in each of these properties between the participants. Participant 1368 appeared to consistently be on the extremes for nearly all measures, and notably for IgG had a highly mutated repertoire with low diversity, which is indicative of clonal expansions. The study was not powered to detect age-related differences in the repertoire, but there was an even mix of participants between the ages of 22 and 59, so each of these global properties was also correlated with age (Figure 9.2). Although not statistically significant, average CDR3 length increased with age, and there was also a change in IgG subclass distribution with age. IgG2 levels decreased with age, giving a relative increase in IgG1 and IgG3.

3.3.7 The public repertoire

The public repertoire is defined as the set of clusters within the repertoire that are common to multiple participants, whereas the private repertoire is the set of clusters that are unique to a particular participant. In an unstimulated setting, the public repertoire comprises a very minor part of the total repertoire, but is greater for IgM than IgG or IgA (1.4, 0.3, and 0.5 % of total clusters respectively). Considering the number of clusters present in different numbers of participants, there was a sharp reduction when considering clusters shared by increasing numbers of participants, and there was only a single cluster (IgM) that was present in all participants (Figure 3.12 A). Compared to the private repertoire, the public repertoire comprised larger clusters that had shorter CDR3s (Figure 3.12 B & D). For IgG, the public clusters were more mutated than the private clusters, but for IgM, the public clusters were actually less mutated than the private clusters, with no significant difference for IgA (Figure 3.12 C).

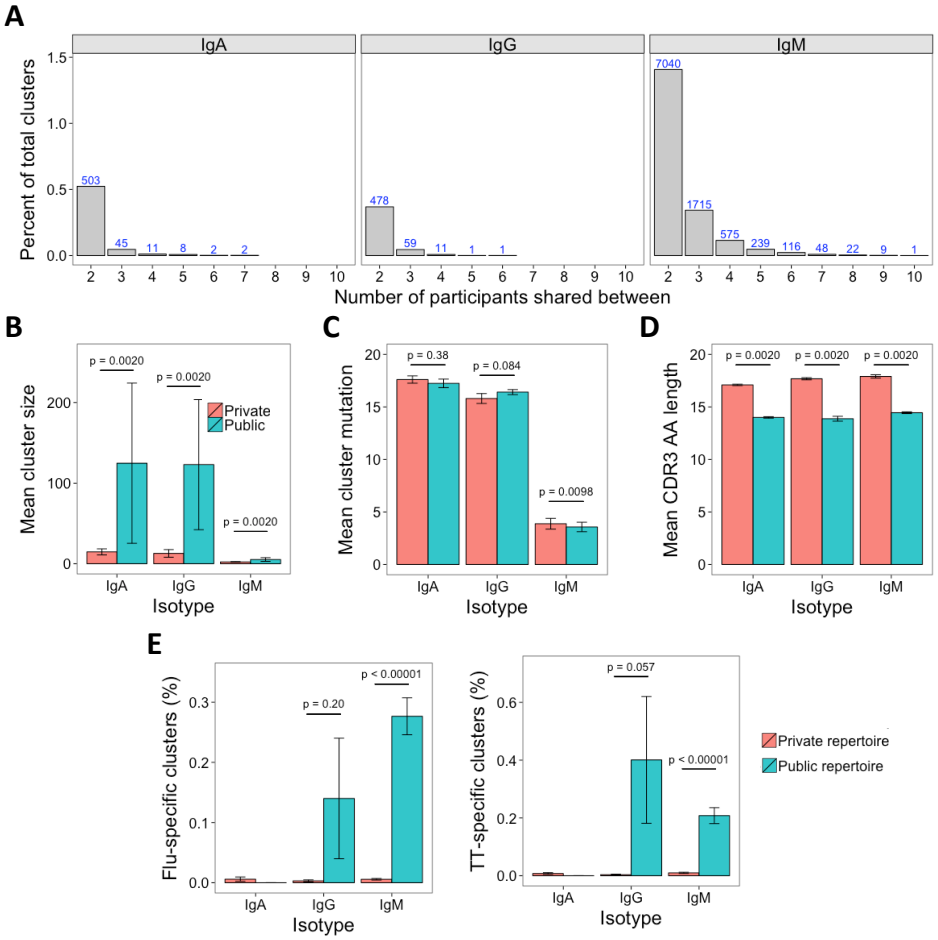

Figure 3.12: The public repertoire. - (**A**) The percent of total clusters that are present in different numbers of participants, where percent = (A∩B/sum(A,B))*100. The blue number above each bar shows the absolute number that is shared. (**B-D**) Mean cluster size, mutation, and CDR3 AA sequence length in the clusters that are unique to a participant (private repertoire) compared to those that are present in at least one other participant (public repertoire). (**E**) Percent of clusters in the private and public repertoire that are annotated as having specificity towards either TT or Influenza antigens. For **B-E**, mean values ±SEM are shown for the 10 participants. Comparisons performed using the paired Mann-Whitney U test.

3.3.8 Database of previously described antigen-specific sequences

A literature search identified 516 Influenza-specific sequences (from 14 publications), 449 TT-specific sequences (from 10 publications) and 114 HBsAg-specific sequences (from 12 publications), that were manually collated to give a comprehensive database of publicly available sequences with known antigenic-specificity (Table 9.2). Comparing these previously described sequences to the baseline dataset revealed 44 TT-specific, 23 influenza-specific and 13 HBsAg-specific sequences that mapped to clusters in the dataset. These clusters were therefore labelled

as potentially containing sequences with specificity towards TT, influenza and HBsAg respectively. These potential antigen-specific clusters were present in samples of at least one isotype from each participant, but only in small numbers. The mean number of potential TT-specific, influenza-specific and HBsAg-specific clusters was 1.5, 0.5 and 0.5 respectively for IgA, 1.3, 0.5 and 0.4 for IgG, and 7.2, 3.6 and 3.3 for IgM.

Despite the low numbers, where they were present, potential TT and influenza-specific clusters comprised a greater percentage of the public compared to the private repertoire (Figure 3.12 D). The identification of HBsAg-specific clusters in the public repertoire is discussed in more detail in Chapter 4.

3.4 Discussion

By collecting both repeat samples from a single participant as well as samples from multiple different participants, it has been possible to perform in-depth assessment of both the within-individual, and inter-individual variation in the BCR repertoire. There is high reproducibility in the methods used to sample the BCR repertoire, and although there is not exhaustive sampling of the entire repertoire at the individual B cell clone level, it is possible to calculate comparable global repertoire metrics, and routinely detect abundant clusters in repeat samples. In the absence of an immune stimulus given during the study, there was considerable variation in the BCR repertoire over time in a single individual, highlighting that it is a highly dynamic system that is constantly subject to immune stimulus and selective pressures. Nevertheless, there are certain features that remain steady, which enable unique identification of different individuals, and may be a contributing factor to genetic causes of variation in the immune response. Finally, it is shown that there is a small public repertoire, which has distinct features compared to the total repertoire, and which appears to be enriched for specificity towards commonly encountered antigens.

Although there is no standardized BCR repertoire sequencing protocol, different protocols give comparable results (191), so the data generated here are still informative for laboratories using alternate protocols. When conducting this study, it was necessary to use a simple and cost-effective protocol, so that it could be used on a large scale in the context of clinical vaccine trials. As input, 500,000 B cells were used - this number can normally be obtained from 2-8 ml of adult peripheral blood, so is a feasible quantity for routine sampling. To account for sequencing and PCR error in the dataset, a clustering-based approach was used that groups together related sequences. This approach is intended to group together sequences arising from the same cell that differ due to sequencing or PCR error, and also from clonally related cells. Each cluster can therefore be considered to represent a B cell clone, with the more abundant clusters representing proliferating clones, or clones with high transcript levels (i.e. PCs). Although using such an approach to define clones is commonplace (119, 192, 193), it should be considered a rough approximation, as B cells from distinct clonal origins could converge towards a similar sequence (150), and some clonally related B cells may diverge more significantly at the sequence level. Nevertheless, our use of the term clone will still represent a group of sequences that likely share similar antigenic specificity.

Obtaining both PCR and biological replicates allowed assessment of the adequacy of the sequencing depth (100,000 sequences per sample) to represent the true diversity of the 500,000 B cells in the sample, and the adequacy of the sampling depth to represent the true diversity of total B cells in a human. Sequencing depth and sampling depth are related in that there is no point collecting a large sample if sequencing depth is insufficient to capture the diversity of the sample. Under-sequencing results in a loss of information, whilst over-sequencing represents an unnecessary cost and increases the number of erroneous sequences in the dataset. Rarefaction analysis indicated that whilst the entire diversity of the 500,000 B cells is not captured, a sequencing depth of 100,000 sequences is sufficient to represent the most abundant clusters. So,

for most applications, this sequencing depth should be sufficient, but for rare cluster identification, a greater sequencing depth would be recommended, especially for the IgM population, which forms the greatest proportion of the total B cells. The estimates of total effective IgG repertoire size from the biological replicates were steady over time, and in line with previous studies (24, 25), indicating that up to 10% of the total peripheral blood IgG BCR repertoire was sampled. Considering just the abundant clusters, estimates varied by nearly a factor of 10 at the different times, indicating that the repertoire of these clusters is more likely to be affected by immune fluctuations.

At day 7, the evidence of clonal expansion indicates that there was a degree of subclinical immune activation in this participant. In addition, the abundant repertoire had the smallest estimated repertoire size at day 7, potentially due to a restriction towards more limited antigenic-specificity. Despite their reduction in number, the abundant clusters were large and mutated, so likely composed of rapidly proliferating B cells (194). This day 7 suspected immune stimulus highlights how dynamic the repertoire is over time, so for studies of the BCR repertoire following vaccination, it must therefore be considered that there could be natural stimulations of the repertoire that could affect the results. A larger study of more participants at multiple timepoints is necessary to determine exactly how common such immune stimulations are in normal healthy individuals, and how long they tend to last.

Despite the day 7 changes in the repertoire, the proportion of usage of different VJ gene combinations remained steady over time in a single individual, compared to the relatively large differences between individuals. As VJ usage therefore appears unique to an individual, and certain V genes are preferentially used in a protective response to certain antigens (101, 195), it could potentially be a cause of variation in disease and vaccine responses between individuals. Although isotype subclass usage, mutation, CDR3 length and diversity are more affected by immune stimulation, they also display considerable inter-individual variation. One explanation

for this inter-individual variation is the different ages of the individuals in the study; it has previously been observed that the repertoire in more elderly participants is less diverse and more mutated (119, 143). In this study, there was also a decrease in diversity, and increase in mutation with age, although this was not a statistically significant finding. Most striking, however, was the decrease in IgG2 levels with age, coinciding with an increase in IgG1 and IgG3. Such an observation is of potential importance, as the different IgG subclasses have different activities in different antigenic contexts (196), and change in their relative abundance could be a contributing factor to immunosenescence (142). For example, IgG2 is important for mounting immune responses against polysaccharide-encapsulated pathogens (such as *Neisseria meningitides* and *Streptococcus pneumonia*) (197). Such responses are reduced in older individuals (198), so this could be due to decreased IgG2 levels.

It appears that when different individuals are exposed to a common antigenic stimulus, there is a degree of similarity in the response (a public repertoire) at the BCR sequence level, and that this could be used to identify antigen-specific BCR sequences (118, 124, 149, 150). However, in this study, there is also a public repertoire in the absence of any common immune stimulation. The presence of such a public repertoire could have three possible causes: laboratory contamination of different samples, random overlap by chance, or historical common antigenic stimuli. Laboratory work was conducted under stringent conditions to minimize cross-sample contamination, and there are no clusters shared across all samples, making this an unlikely contributor to the public repertoire. If sharing was due to chance, it is expected that the public and private repertoires would have similar properties, but this is not the case. The public IgG repertoire comprises larger, more mutated clusters, with shorter CDR3s than the private repertoire; this is consistent with these clusters arising from more differentiated B cell subsets (124). In addition, considering the presumed antigenic specificity of the clusters, a greater proportion of the public repertoire comprised presumed TT or influenza-specific clus-

ters compared to the private repertoire. These are antigens to which all participants in the study are likely to have been exposed through either vaccination or infection, and thus provide support for using public repertoire analysis for identification of antigen-specific clusters following common antigen stimulation. However, it must be considered that this technique could then also enrich for sequences specific to antigens that are commonly encountered by the population. The public repertoire in the IgM dataset is approximately three times larger than that of the IgG or IgA datasets. This may be due to the presence of natural IgM antibodies, which target conserved microbial determinants and autoantigens (199); the most abundant of these is anti-gal, which constitutes approximately 1% of all human antibodies (200). This could explain why the public repertoire is less mutated than the private repertoire for IgM sequences, while the converse is true for IgG sequences. Unfortunately, there are currently no large sequence datasets of natural antibodies available to search in our dataset to confirm this hypothesis.

3.5 Conclusion

To summarise, this chapter presents a robust BCR repertoire sequencing method, and demonstrates that a sample of 500,000 B cells, and sequencing depth of 100,000 sequences per B cell isotype should be sufficient for most applications. The BCR repertoire varies substantially within a single individual over time, and between multiple different individuals. Assessing the complete range of such variation would benefit from further characterization in larger study cohorts. Nevertheless, there are certain conserved features of the repertoire within individuals that could be predictive of immune function. Finally, the public repertoire was investigated, and appears to be enriched for sequences with specificity towards antigens commonly encountered by the population.

4

Dynamics of the BCR repertoire following HepB booster vaccination

4.1 Introduction

In order to investigate whether BCR repertoire sequencing can give useful insight into B cell responses, and thus realise the utility of this technology for probing responses to vaccination and infection, it is important to have model systems in which to test it. The use of vaccines as a controlled model system to study B cell responses is an important approach in human immunology (201), as there can be highly controlled timing of antigen exposure and blood sampling. High-throughput BCR repertoire sequencing has now been applied to some vaccine studies (see Section 1.4.2; (25, 99, 119, 147, 150, 202)), and it does appear possible to detect vaccine-induced perturbations in the total repertoire that relate to the functional B cell response (97, 203). Of interest, despite BCR repertoire diversity there appears to be a degree of sequence convergence across individuals for a given antigen. Convergence has been seen seven days following vaccination with simple polysaccharide antigens (150), and more complex influenza antigens (203), as well as following dengue infection (118).

Despite this progress, it remains unclear how long perturbations in the BCR repertoire can be detected following vaccination. Most of the previous studies just focus on sampling blood around 7 days following vaccination, as this is when there is maximal egress of vaccine-specific

PCs into the peripheral blood, so thought to be when the greatest effects on the BCR repertoire will be seen (91, 95). Furthermore, the published studies focus on studying perturbations in the total repertoire, yet as detailed in Chapter 3, there may be confounding fluctuations in the total repertoire that are not related to vaccination. The ability to enrich for vaccine-specific BCR sequences from the total repertoire will be key in overcoming this limitation.

The HepB vaccine, is a safe and effective single-subunit vaccine, making it an ideal model antigen for studying a non-complex immune response. The vaccine is made from recombinant HBsAg, and alum adjuvant. In the vaccine formulation, the HBsAg self-assembles into spherical structures termed virus-like particles, which display the major immunogenic epitopes on the outside. In the United Kingdom, the HepB vaccine is not part of the routine vaccination schedule, and natural infection is rare, so it is possible to find individuals who have had no prior exposure to HBsAg. HepB vaccine is however recommended for those travelling to affected countries, so it is also possible to find previously immunised individuals. This vaccine therefore provides a unique opportunity to study both primary and secondary immune responses. A HepB vaccine study was therefore designed that consisted of two parts: one investigating booster vaccination, and one investigating primary vaccination. As booster vaccine is expected to give a more pronounced response, this was the first part of the study to be conducted in order to obtain detailed information about perturbations in the repertoire following vaccination. The primary vaccination study is considered in Chapter 5.

For the booster study described here, HepB vaccine was administered to nine participants, and blood sampled at multiple time points up to one month later (Figure 4.1). In addition to sequencing the total repertoire at each time point, cell sorting and sequencing of HBsAg-specific B cells and PCs was performed to generate a vaccine antigen-enriched sequence dataset. Results indicate that there are distinct time-limited signatures in the repertoire that appear to be related to the appearance of HBsAg-specific PCs in the blood at day 7, and HBsAg-specific

memory cells at later time points. Using this information, analytic approaches to enrich for HBsAg-specific sequences from the total repertoire data were investigated.

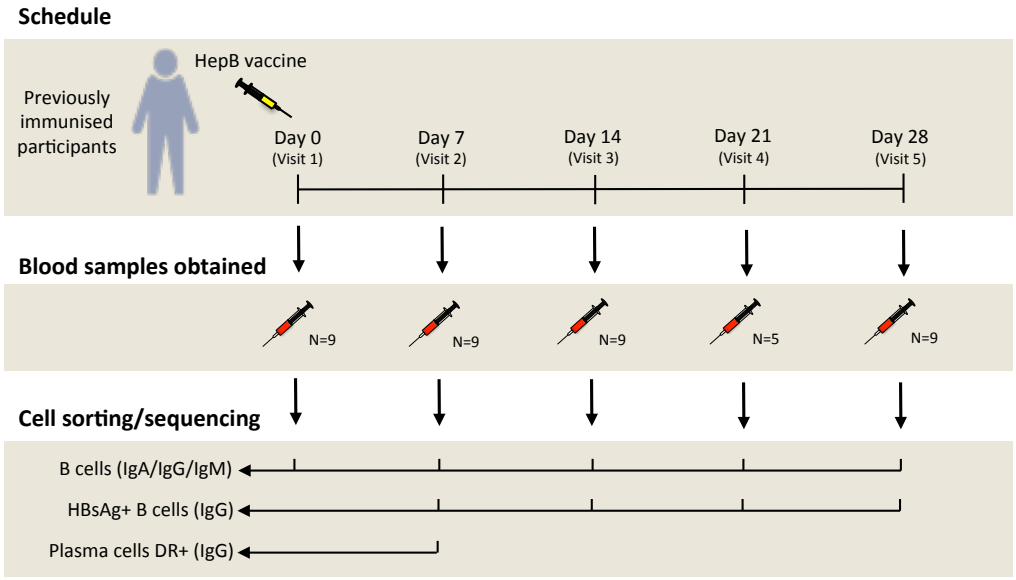

Figure 4.1: Study Design - Nine participants with previous history of HepB vaccination were enrolled to receive a HepB booster vaccine. Total B cells were sorted and sequenced (IgA, IgG and IgM transcripts) on the day of HepB booster vaccine, and on days 7, 14, 21 (samples missing from 4 participants), and 28 following vaccination. In addition, for 5 of the participants in this group, HBsAg+ B cells were isolated and sequenced on days 7, 14, 21 and 28 following vaccination, and PCs on day 7 following vaccination. Sequences from the HBsAg+ sorted cells, and sorted PCs were used to find sequence clusters in the total repertoire that appeared to have enriched specificity towards the vaccine.

4.2 Methods

4.2.1 Study design

Nine healthy subjects (aged 23-59), who had previously received a full primary course of HepB vaccination, were recruited with informed consent, under approval from the Northampton Research Ethics Committee (13/EM/0036). Participants were given a single intramuscular HepB booster vaccine containing 10 μg HBsAg, adsorbed on amorphous aluminium hydroxyphosphate sulfate (HBvaxPRO®, Sanofi Pasteur). A sample of 50 ml peripheral blood was taken immediately before vaccination as well as 7, 14, 21 and 28 days after vaccination (Figure 4.1). Blood was transferred to a heparinized tube for processing within 4 hours of collection.

4.2.2 Sample processing

For each blood sample, a 2 ml aliquot of blood was taken for serum separation, and determination of anti-HBs antibody concentration, and PBMCs extracted from the remainder. Half of the PBMCs were used for enumeration of HBsAg-specific PCs and memory cells by ELISpot, and the remainder used for BCR repertoire analysis. For BCR repertoire work, total CD19+ B cells were isolated by MACS, and an aliquot of 500,00 used for sequencing of the total BCR repertoire. The remainder were stained with Live/dead-Aqua (Life technologies), CD19-FiTC (HIB19) (eBioscience), CD20-APCH7 (2H7) (BD), CD27-PECy7 (O323) (eBioscience), CD38-PE (HIT2) (eBioscience), HLA-DR-PerCpCy5 (L243) (BioLegend) and HBsAg-APC (Miltenyi Biotec) for isolation of both HBsAg+ B cells (viable, CD19+, CD20+, HBsAg+) , and PCs (viable CD19+, CD20-, CD27+, CD38+, HLA-DR+) on visit 2, and isolation of HBsAg+ B cells on visits 2-5 (Figure 4.2). Cell sorting was performed on a MoFlo cell sorter by a specialist operator. PCR was conducted independently for IgA, IgG and IgM transcripts for total repertoire samples, and just for IgG transcripts for the HBsAg+ and PC samples. Total repertoire samples were sequenced with 40-60 samples multiplexed on each run, and the HBsAg+ and PC samples were sequenced with 80-120 multiplexed on each run.

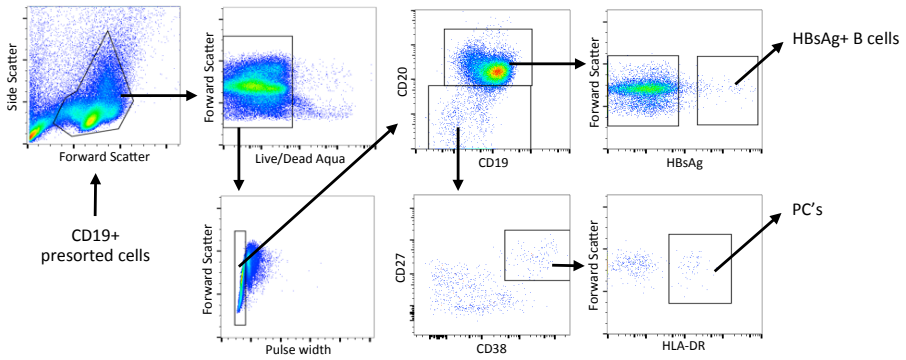

Figure 4.2: Gating strategy for isolation fo HBsAg+ B cells and PCs - HBsAg+ cells identified as viable, CD19+, CD20+, HBsAg+ and PCs identified as viable CD19+, CD20-, CD27+, CD38+, HLA-DR+.

4.2.3 Sequence processing

For total repertoire samples, sequences were processed using the pipeline described in Figure 2.4. To account for differences in the number of resulting sequences in different samples, all samples were randomly subsampled without replacement using the sample function in R to give 100,000 sequences per sample. Clustering was performed using the thresholds determined in Chapter 3 - to be included in the same cluster, sequences were required to have the same V and J gene annotation, the same length CDR3, and no more than 1 AA mismatch per 12 AA's in the CDR3. To determine whether a cluster is present in more than one sample, data from all samples were clustered together. If two samples contribute at least one sequence to the same cluster, that cluster is defined as being present in both samples.

For HBsAg+ and PC samples obtained from FACS sorting, sequences were processed using the same pipeline, except that sequence clustering was not performed. Instead, sequences were mapped to the clustered dataset from the total repertoire samples to see which clusters they would have been included in (i.e., same V and J gene annotation, and similar CDR3). These clusters in the total repertoire were then annotated as having similarity to HBsAg-specific B cells or PCs accordingly. As the specificity of the HBsAg+ cell sorting is not 100% (see Section 2.1.10), for this comparison, the HBsAg+ and PC sequences were only matched back to participants from whom those sequences were not obtained in order to reduce the effect of non-specific matching. The probability of any two individuals sharing a given randomly selected sequence is very low, so sequences are unlikely to match to the repertoire of a different individual unless they are specific to the common stimulus (HBsAg). This matching technique therefore provides an additional method for enriching for vaccine-specificity within the stained cells.

4.3 Results

4.3.1 Serological and cellular measures of vaccine response

All nine previously vaccinated participants had an increase in anti-HBs antibody concentration to greater than 100 mIU/ml by day 28 following vaccination (p = 0.0039). There was considerable inter-individual variation in both the pre-vaccination and post-vaccination antibody concentrations (Figure 4.3 A/C). HBsAg-specific IgG PC numbers in the peripheral blood, as determined by ELISpot, peaked 7 days following vaccination in all participants (mean 99, range 2-208 PCs/10^6 PBMCs), and were present at negligible numbers on all other days (Figure 4.3 B/D). HBsAg-specific IgG memory cell frequency, as determined by ELISpot, peaked at a later timepoint than the PCs; the greatest numbers were seen 28 days following vaccination (mean 72, range 0-163 memory B cells/10^6 cultured lymphocytes). Day 7 PC numbers were strongly correlated both with day 28 memory B cell numbers (Spearman's rho = 0.77, p = 0.0214; Figure 4.3 E), and the absolute increase in antibody concentration between day 0 and day 28 (Spearman's rho = 0.87, p = 0.0045; 4.3 F).

4.3.2 BCR repertoire sample QC

Sequence data were obtained for total IgA, IgG and IgM transcripts from all nine vaccinated participants on day 0, 7, 14 and 28, and from five participants also at day 21. On average 350,450 (203,151 - 1,023,663) raw reads were obtained per sample (Table 9.3). After processing, this dropped to 269,156 (109,499 - 1,195,321) per sample, of which 100,000 per sample were randomly subsampled for normalisation, and used for further analysis. Following clustering, there were on average 10,924 (1,677 - 22,878) clusters per sample for the IgA dataset, 14,560 (1,604 - 26,375) clusters per sample for the IgG dataset, and 55,612 (19,088 - 79,102) clusters per sample for the IgM dataset.

As demonstrated in Chapter 3, certain features of the repertoire remain steady over time

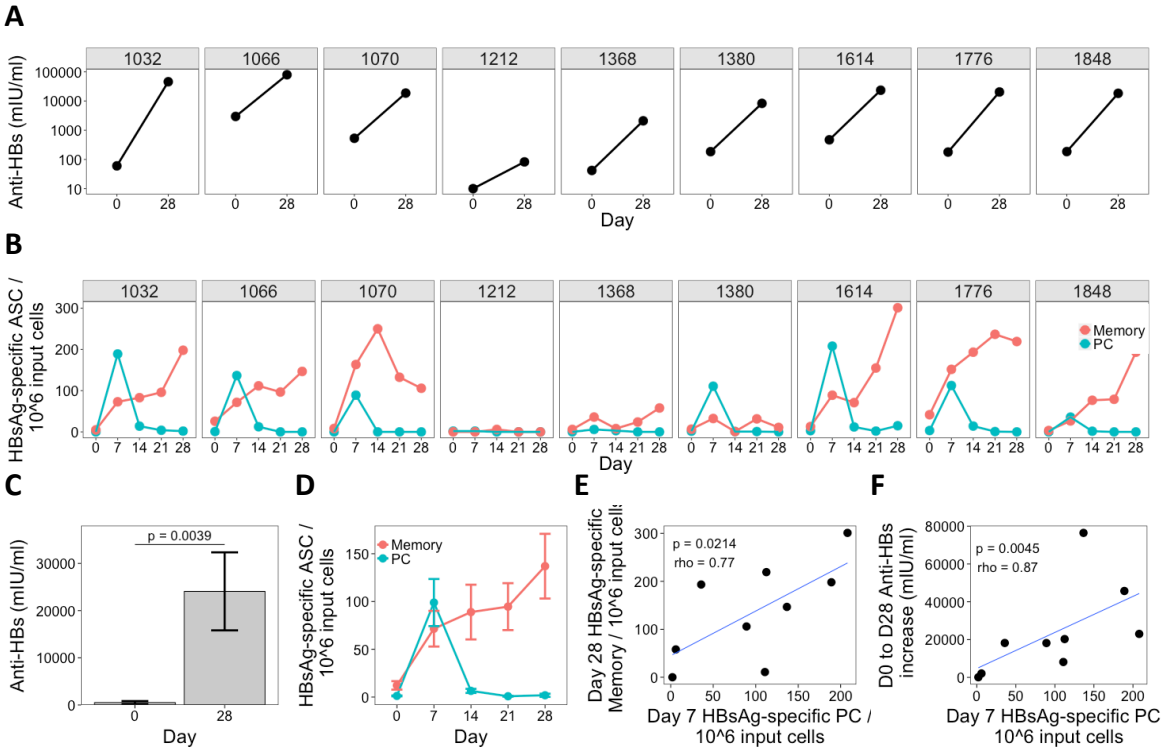

Figure 4.3: Clinical measures of vaccine response - (**A**) Anti-HBs antibody concentration of each participant at day 0 and day 28 following vaccination. (**B**) Kinetics of HBsAg-specific memory cells and PC numbers during the study period determined by ELISpot. Input cells are PBMCs for the PC detection assay, and cultured lymphocytes for the memory cell detection assay. (**C** & **D**) Same as A and B, but displaying mean values ±SEM of the nine participants. (**E**) Correlation between HBsAg-specific PC numbers at day 7, and HBsAg-specific memory cell numbers at day 28. (**F**) Correlation between HBsAg-specific PC numbers at day 7, and the absolute increase in antibody concentration between day 0 and 28. For **E** and **F**, rho represents Spearmans rank correlation coefficient.

even in the presence of an immune stimulus (V and J gene usage), whereas others appear to be associated with the degree of immune stimulus (repertoire diversity, V gene mutation, and average CDR3 sequence length). Analysis of these features via PCA can therefore be used as a form of quality control for the samples by looking for outliers. Outliers in V and J gene usage within a participant could indicate technical problems during sample preparation, whereas outliers in the general repertoire properties could indicate biological grounds for omitting samples from subsequent analysis. These data showed little change in V/J gene usage over time indicating that there were minimal technical biases (Figure 4.4 A). Analysing the general properties on the other hand did indicate some unusual features of the samples. As expected,

day 7 IgG samples formed a slightly distinct clustering, consistent with immune activation by the vaccine (Figure 4.4 B). The day 0 sample from participant 1368 appeared to be an outlier, most prominently in the IgG dataset, but to a lesser extent in the IgA and IgM datasets. In addition, all of the samples from participant 1070 were outliers in the IgM dataset.

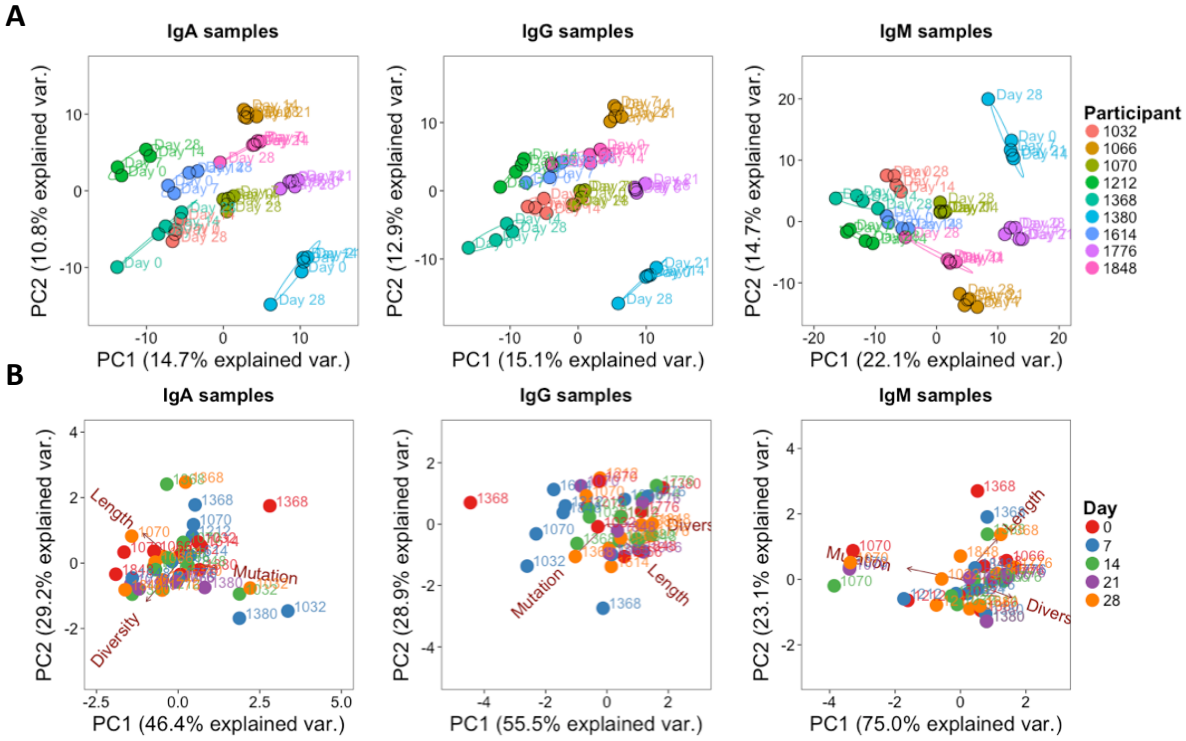

Figure 4.4: PCA of repertoire properties for QC - (**A**) For each sample, the proportion of the repertoire comprised by each VJ recombinant was calculated. Considering samples from each isotype separately, these data were used in a PCA to detect outliers. (**B**) For each sample, repertoire diversity, average V gene mutation, and average CDR3 AA length was determined. These measures were corrected for cluster size, to accentuate any differences between the samples. Considering samples from each isotype separately, these data were used in a PCA to detect samples with signs of B cell activation.

4.3.3 Repertoire expansions are seen 7 days post vaccination, and impact total repertoire properties

To further investigate the suspected immune activations identified by PCA, the kinetics of clonally expanded sequence clusters were determined. As shown in Chapter 3 (Figure 3.4), taking the 200 most frequent clusters can be used as a conservative methods to include all expanded clusters in a sample present at a frequency greater than 0.01%. Determining the

frequencies of these clusters at each timepoint makes it possible to track their dynamics during the course of the study (Figure 4.5 A & Figures 9.3 - 9.5). Although there was great inter-individual variability in the dynamics of these clusters, nearly all participants had a degree of expansion 7 days following vaccination, that tended to be most pronounced in the IgG samples. Of note, the repertoire of participant 1368 appeared to contain a large number of expanded clusters on the day of vaccination, and the repertoire from participant 1070 had a number of large expanded clusters that were present at all timepoints both in the IgG and IgM dataset. Both datasets also contained a single expanded cluster present at a frequency greater than 10% at all timepoints (although the identification of this cluster was different in the IgG compared to IgM dataset). Based on the finding of clonal expansions at day 0, and being an outlier by PCA, participant 1368 was excluded from downstream analysis due to suspected immune activation at day 0.

To quantify the clonal expansions, the top 200 most frequent clusters at each day for each participant were found, and the percent of the total repertoire that they account for calculated (Figure 4.5 B). Although the top 200 clusters comprise only a small percent of the total number of clusters (mean 1.7% for IgA, 1.3% for IgG and 0.4% for IgM), sequences contained in these clusters account for a large percentage of the total repertoire (mean 52% for IgA, 55% for IgG and 16% for IgM). The percentage of the repertoire comprised by these expanded clusters increased 7 days following vaccination ($p = 0.0124$) for the IgG dataset, and then decreased back to a baseline-like state by 14 days following vaccination. There was no significant change from baseline in the percent of the repertoire comprised by these expanded clusters at any day in the IgA or IgM datasets.

Comparing the 200 most frequent clusters to the remaining clusters for each sample showed them to be less diverse, more mutated, and having shorter CDR3 lengths than the remaining clusters (Figure 4.6). Considering the total repertoire, there were no significant changes in

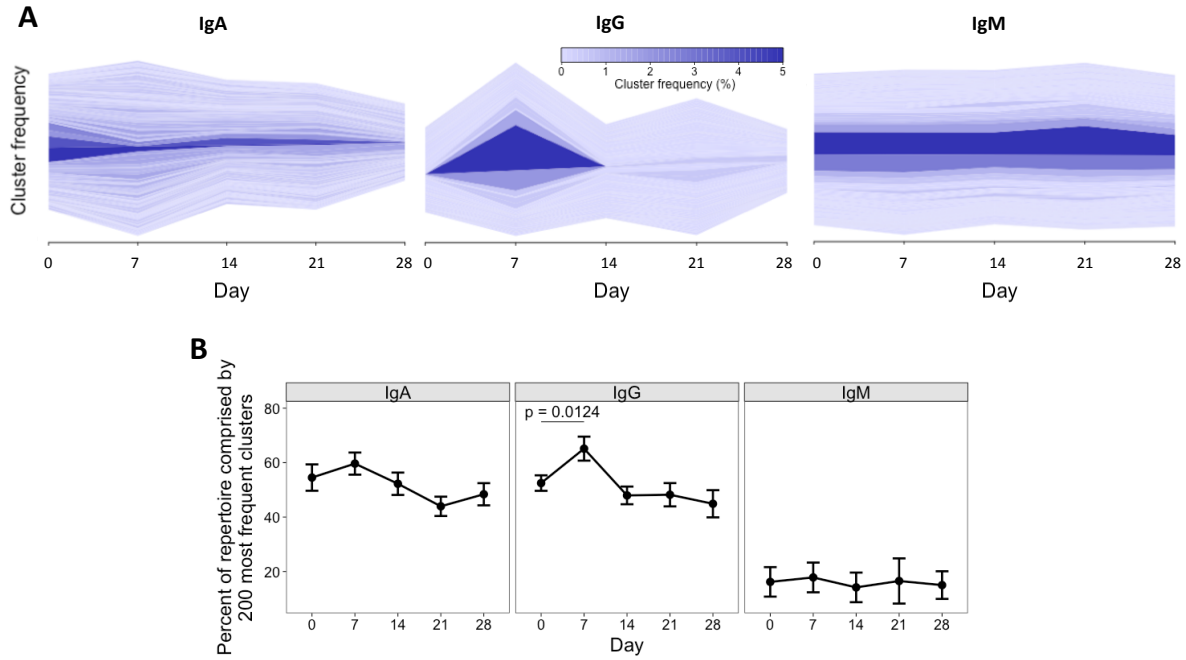

Figure 4.5: Day 7 repertoire expansions - (**A**) Representative cluster kinetics plots from participant 1848 for each isotype (see Figures 9.3-9.5 for plots from all participants). The 200 most frequent clusters at each day were found for this participant, and the frequencies of these clusters were plotted as a stacked bar chart, centered to the middle of the y-axis at each day. Clusters from each day were then joined using a horizontal stream to illustrate how the frequency of the clusters changes over time. The width and darkness of the stream represents the frequency of the cluster at that time. (**B**) Mean percent of the total repertoire comprised by the top 200 most frequent clusters at each day. Error bars indicate ±SEM from 8 participants. P value represents the result from a t-test.

these features over the course of the study for the IgM dataset. For the IgG dataset, the repertoire became less diverse ($p = 0.0320$), more mutated ($p = 0.0446$), and had a shorter average CDR3 AA length ($p = 0.0078$) at day 7 compared to day 0, consistent with increased B cell stimulation at day 7. For the IgM dataset, there was a shorter average CDR3 AA length ($p= 0.0234$) at day 7 compared to day 0, but mutation and diversity did not change. There were no differences between day 0 and any of the other timepoints for any isotype. Considering just the 200 most frequent clusters, these day 7 changes also occurred, but they were not as apparent when just considering the remaining less frequent clusters. It therefore appears to be the frequent clusters that drive the changes seen in the total repertoire at day 7, but even without enriching for these frequent clusters it is still possible to determine changes in the metrics of the total repertoire at day 7.

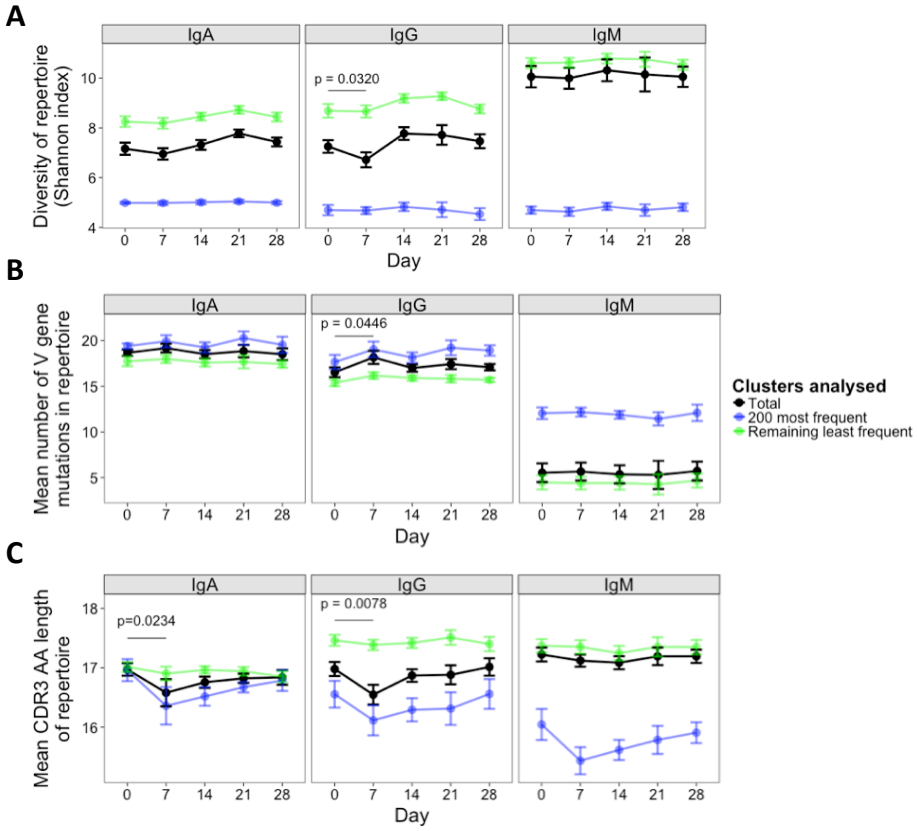

Figure 4.6: Expansion-induced changes in diversity, mutation, and CDR3 AA length - Changes in (**A**) mean repertoire diversity (measured using Shannons index), (**B**) mean number of V gene mutations from all sequences in the repertoire, and (**C**) mean CDR3 AA sequence length from all sequences in the repertoire. Mean ±SEM shown for 8 participants; black points are values from the total repertoire, green are from sequences contained within the 200 most frequent clusters from that sample, and blue are from the remaining clusters. P value represents the result from a t-test comparing day 0 and day 7 values from the total repertoire.

4.3.4 Repertoire convergence between participants peaks 14/21 days after vaccination

To investigate repertoire convergence following vaccination, every participant was compared to every other participant to see how many clusters were shared between each pair of two participants (Figure 4.7). This was relatively low at day 0 (mean of 26 (0.12%) for IgA, 20 (0.07%) for IgG and 601 (0.53%) for IgM). The number increased following vaccination, and peaked at day 21 (mean of 48 (0.16%) for IgA, 45 (0.11%) for IgG and 806 (0.61%) for IgM). For IgM, the increase in sharing from day 0 was only significant for day 14 ($p < 0.0001$), but for IgA and IgG the increase in sharing was significant for days 14 (IgA, $p = 0.0206$; IgG, $p =$

0.0014) and 21 (IgA, p = 0.0002; IgG, p <0.0001).

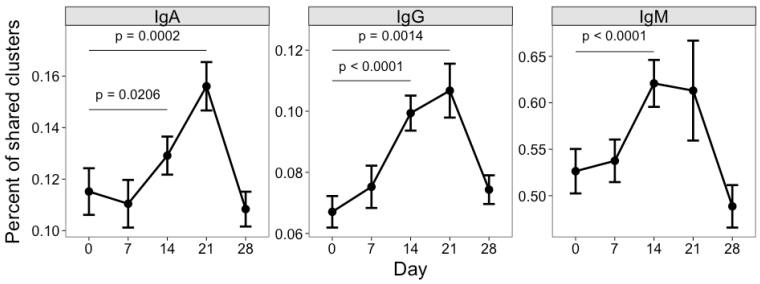

Figure 4.7: Repertoire convergence - At each day, the percent of clusters shared by each pair of two participants was determined (8 participants, giving 28 different pairings). Shown are the mean values ±SEM of the percent of clusters shared between each pair. Percent is calculated as (A∩B/sum(A,B))*100. P values represent the results from t-tests comparing day 0 to the other days.

4.3.5 Computational enrichment of vaccine-specific clusters from the total repertoire

Based on the observed repertoire signatures following vaccination, a simple computational model was developed to attempt to enrich for the vaccine-specific clusters from the total repertoire. For this analysis, only the IgG dataset was used, as the signals were stronger here, and class-switched IgG cells dominate the response to booster vaccination (90). Cluster frequency is a proxy for clonal expansion, and mutation is likely to be more common in clusters that have undergone clonal expansion. As there is an increase in cluster frequency and cluster mutation at day 7 seen in this study, these properties were included in the model. In addition, there is a clear signal of similar clusters seen after vaccination (albeit at a different timepoint), so sharing was also a requirement in the model. Although average CDR3 AA length also changes following vaccination, this was not included in the model, as stereotypic CDR3 lengths are not a universal property, but likely only specific for certain antigens or epitopes, so could result in some clusters being missed.

To be classified as a likely vaccine-specific cluster in our model, the cluster was required to be frequent, mutated, and shared by at least 2 participants. Conservative cutoff values of more

than 18 mutations, and a frequency greater than 0.01% were set based on the distributions of these variables (Figure 4.8 A). Increasingly more clusters were identified that fell into all three of these categories at the later timepoints, going from 65 at day 0 to 113 at day 28 (Figure 4.8 B). On each day, the overlap between each of these three categories was significantly more than expected by chance ($p < 0.0001$; fishers exact test). Ignoring timepoint, there were 516 clusters that fell into all three of these categories. Searching for these potentially vaccine-specific clusters in each sample shows that they are present at the lowest number on day 0, and increase after vaccination in all participants, to peak at day 7 (Figure 4.9 A). Additionally, the size of these clusters is greatest at day 7.

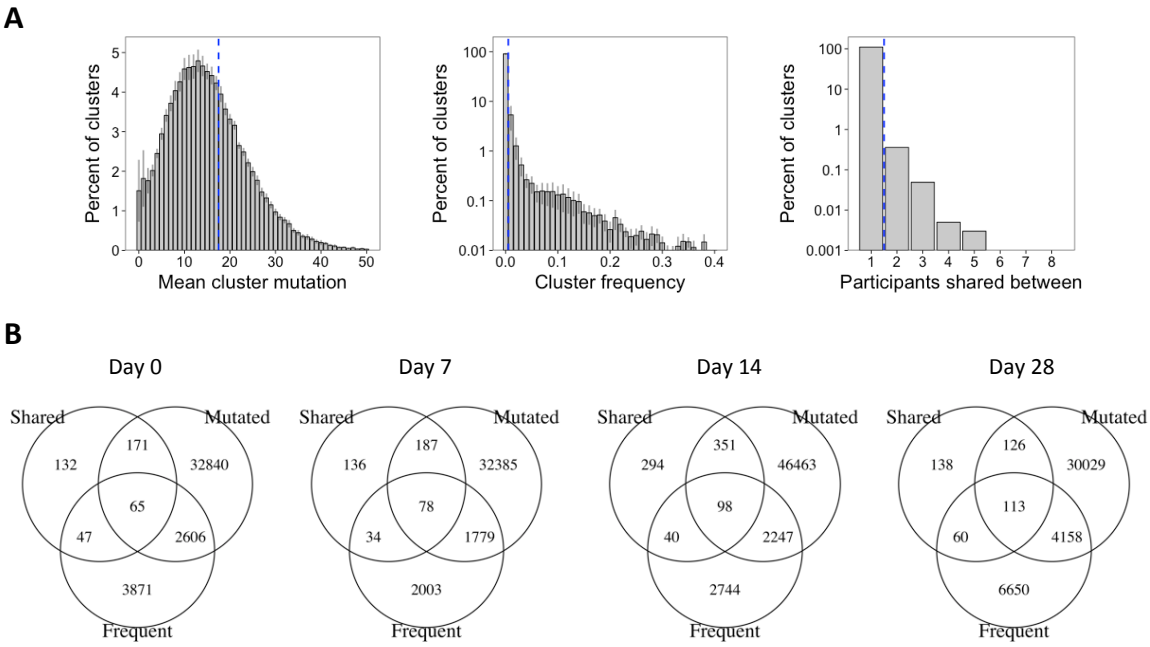

Figure 4.8: Computational enrichment of putative vaccine-specific clusters from the total IgG repertoire - (**A**) Distribution of IgG clusters in the total repertoire at day 0 according to mean number of mutations, frequency, and how many participants they are shared between. Mean values and ±SEM shown for 8 participants. Vertical dashed blue lines indicate where the cutoffs were chosen to be included in the model for vaccine-specific cluster enrichment. (**B**) Venn diagram showing the overlap of the shared (by more than two participants at any day), mutated (more than 18 V gene mutations on average at any day), and frequent (more than 0.01% of the total repertoire at any day) clusters from the total repertoire of IgG clusters from the 8 participants on each day. Day 21 is excluded, as samples from three participants were missing here.

4.3.6 FACS enrichment of vaccine-specific clusters from the total repertoire

FACS was used to isolate a total of 85,000 HBsAg+ B cells after vaccination, and 3,444 PCs at day 7 from five of the participants included in this study (Table 9.4). Using the sequence data from these samples to annotate the total IgG repertoire dataset showed that of the 482,849 total clusters, 1,056 were annotated as HBsAg+ and 167 were annotated as PC+. In total, there were 41 clusters that were annotated as both the HBsAg+ and PC+, which is significantly more than expected by chance (fishers exact test; $p < 0.0001$). As with the computationally enriched clusters, searching for these HBsAg+ and PC+ enriched clusters shows them to be present at the lowest number on day 0, and increasing after vaccination in all participants with a peak at day 7 (Figure 4.9 B/C). Despite considerable inter-individual variation, the clusters annotated as HBsAg+ or PC+ were more mutated, had greater frequency, and were more likely to be shared between participants than those not annotated (Figure 9.6), fitting with the use of these parameters in the computational enrichment model.

This kinetic was not present when searching for clusters annotated with specificity towards irrelevant antigens based on matching to the database of antigen-specific sequences (Figure 9.7). However, the numbers of these clusters identified was low at all timepoints due to limitations in the number of previously published sequences.

4.3.7 Computational and FACS enrichment identify the same clusters

The kinetics of the clusters identified as vaccine-specific both computationally and by FACS gives good evidence that these clusters are indeed enriched for vaccine-specificity. To further validate this, the vaccine-enriched clusters obtained computationally were compared to those obtained by FACS to determine whether the same clusters were identified using both methods (Figure 4.10). This was done separately for each metric in the model. By combining the metrics of sharing, mutation, and frequency, it was possible to obtain greater enrichment for vaccine-specific clusters than using either of these metrics on their own. Sharing and frequency had the

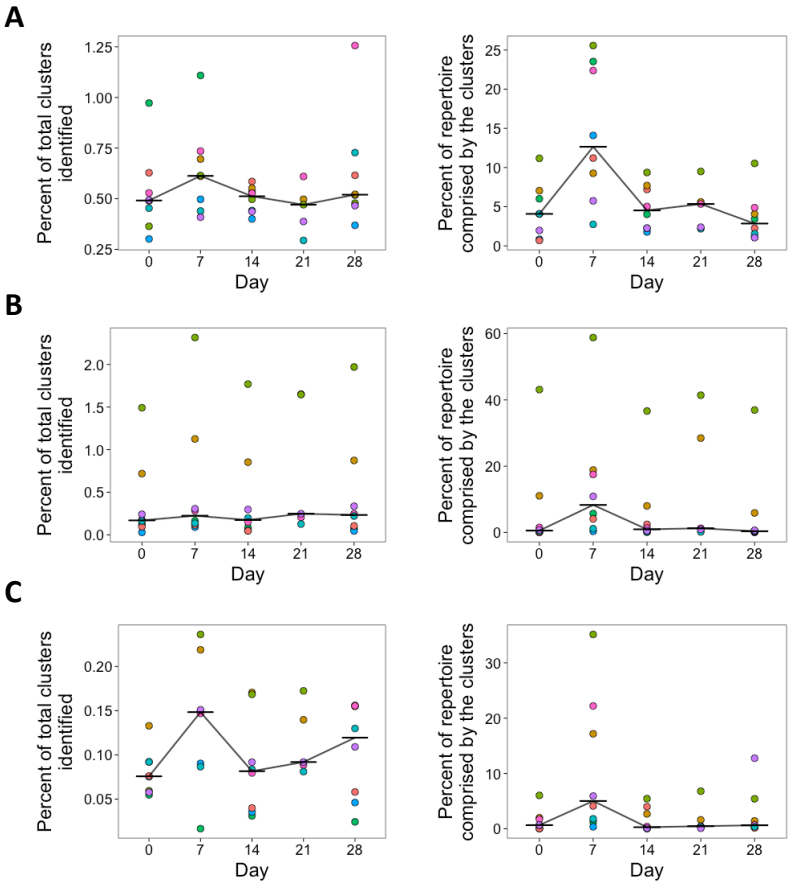

Figure 4.9: Kinetics of vaccine-enriched clusters - (**A**) Of the 516 total clusters computationally identified as putatively vaccine-specific, the number of these in each participant at each day was determined, and expressed as a percentage of the total number of clusters present in that participant at that day (left panel). In addition, the percent of the total repertoire comprised by the clusters identified (ie, corrected for cluster size) was determined (right panel). (**B**) Shows the same as **A**, but with the 1,056 clusters identified that matched to sequences from the HBsAg+ sorted cells. (**C**) Shows the same as **A** & **B**, but with the 167 clusters identified that matched to sequences from the PC+ sorted cells. For **B** and **C**, matching of sequences from sorted cells was only conducted for total sequence data sets from individuals from whom the sorted cells were not derived in order to improve specificity of the matching. For all plots, horizontal bars show the median value from the 8 participants. The colour of the point represents the participant.

greatest effect, while mutation added little sensitivity to the model. As well as identifying the fewest clusters with the computational model at day 0 (Figure 4.8 B), the percent of these also identified by FACS was also lowest on this day (9.2% HBsAg+ and 9.2% PC+). The greatest overlap was on day 7, where 19% of the 78 computationally-enriched clusters were annotated as HBsAg+ and 21% annotated as PC+. This overlap with the computationally-enriched clusters is 42x greater for HBsAg+ clusters, and 168x greater for PC+ clusters than considering the overlap with total clusters. Following day 7, the degree of overlap decreased at each subsequent

timepoint sampled, back to baseline-like levels by day 28

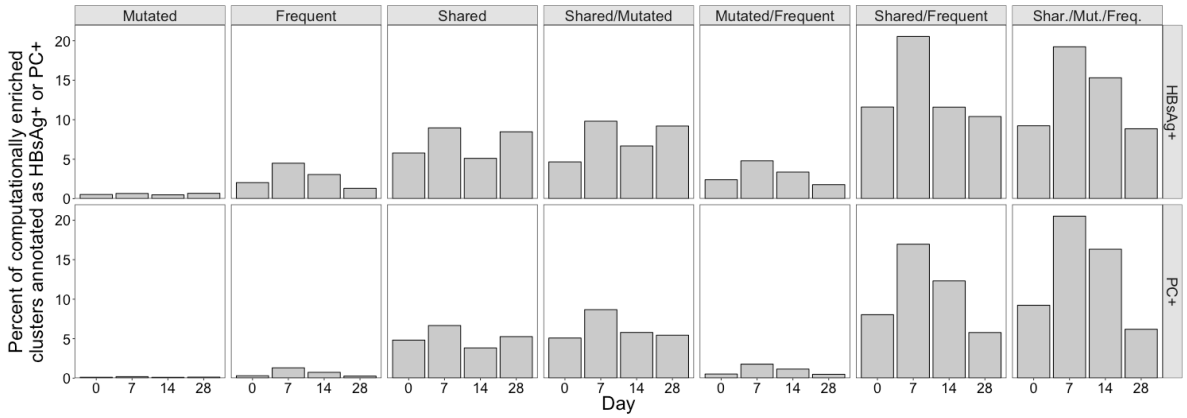

Figure 4.10: Comparison of computational and FACS enrichment - On each day, clusters that fitted each combination of the computational enrichment model parameters (shared, frequent and mutated) were determined. The percent of these clusters that were also annotated as HBsAg+ or PC+ based on the FACS data was then determined.

4.4 Discussion

Using HepB booster vaccination as a model system, in-depth sequence analysis of the BCR repertoire was conducted from circulating B cells to thoroughly characterize how the BCR repertoire responds to secondary antigen encounter. Although vaccine-specific B cells account for only a small proportion of the total peripheral B cells even at the peak of their appearance, it was still possible to detect perturbations in global repertoire properties 7 days following vaccination. These changes were characterized by a decrease in diversity, an increase in mutation, and a decrease in CDR3 sequence length, and can be related to the appearance of vaccine-induced PCs in the peripheral blood at this time. These changes were most prominent in the class-switched datasets, which is to be expected, as this is a secondary response so will be stimulating switched memory cells. Interestingly, despite these perturbations being most pronounced 7 days after vaccination, the greatest sequence similarity between the repertoires of different individuals was seen on days 14 and 21 after vaccination.

Although day 7 repertoire perturbations following vaccination have previously been ob-

served (147, 152, 203), repertoire similarity on days 14 and 21 has not been previously studied. The minimal repertoire overlap observed at day 0 is to be expected considering the enormous diversity of the repertoire, and the sampling depth of 100,000 reads (see Chapter 3). Given the diversity of the BCR repertoire, the finding of two identical CDR3 AA sequences in 2 different individuals is unlikely to occur by chance alone (151), so could be indicative of common selection pressures on the repertoire of different individuals. The increase in convergence seen following vaccination here is in line with the vaccine antigen either imposing such a selective pressure, causing convergent evolution of the repertoire in the different individuals to produce sequences specific for the antigen, or simply expanding pre-existing memory cells that have already been selected in this way. Such antigen-driven convergence has previously been seen following immunisation with highly repetitive polysaccharide antigens (150), and also in the context of influenza vaccination (203). Both of these studies investigated convergence at 7 days following vaccination, as this coincides with the peak of PCs in the peripheral blood in a secondary response, so is expected to give the greatest signal.

This study confirms that convergence is also seen in the context of a simple protein antigen (with alum adjuvant), but by sampling at more timepoints, shows that the increase in convergence at day 7 is relatively minor, and instead the most significant signals are seen on days 14 and 21. The kinetics of the increase in convergence is therefore more in line with the kinetics of vaccine-specific memory cells in the peripheral blood rather than PCs. One interpretation of this is that following vaccination there is rapid activation of circulating memory cells to terminally differentiate into low-affinity PCs by day 7, as demonstrated by the greatest repertoire expansions seen in our study at this time. However, it is likely that there is also formation of GCs, which mediate further proliferation and selection for antigen binding, and the production of higher-affinity memory cells at later timepoints (65, 204). The small increase in convergence at day 7 may therefore be a result of low-affinity PC sequences, while the increased convergence

later on may be a result of the higher-affinity memory cell sequences. Although there is not an increase in convergence at day 28 when there are the greatest number of circulating memory cells detectable by ELISpot, this may be because the circulating memory cells are all clonally related by this time, so sequences derived from these cells would all be incorporated into a single sequence cluster. Alternatively, this may be an artefact of the stimulation method used for memory cell detection by ELISpot in that the day 28 memory cells are more readily stimulated to secrete antibody than earlier memory cells. Nevertheless, these results suggest that careful consideration of timepoints should be taken when using BCR repertoire sequencing to find mAb sequences. Normally, day 7 PCs are isolated and sequenced, but the resulting mAbs can be of low affinity (104, 155); if later timepoints were instead chosen and the memory cell sequences identified, it may be possible to generate higher affinity mAbs.

To narrow down on the vaccine-specific repertoire from the total repertoire, two approaches were used in parallel: computational enrichment, and sorting and sequencing HBsAg+ (memory) and PCs followed by use of these sequences to annotate the total repertoire. Although neither of these methods is likely to be 100% specific, showing that the two methods gave similar results lends validity to both. That the HBsAg+ and PC+ clusters identified through the cell sorting data had increased convergence, mutation and frequency compared to the total repertoire, shows that cells specific to the vaccine were responsible for causing the observed repertoire signatures, and backs up their use in the computational model.

Comparing the computationally enriched clusters to those annotated as HBsAg+ or PC+ was ultimately used to validate each of the different parameters in the computational model, although it should be noted that there are a number of limitations to this. First, it was not possible to determine the specificity of the model, as the entire HBsAg+ and PC+ repertoire can not be sorted and sequenced, so there may be more clusters in the computationally-enriched population that are vaccine antigen-specific, but were not annotated as such based on the

HBsAg+ and PC+ data. Considering specificity of the sorted cells, HBsAg+ cell sorting was approximately 50% specific (see Section 2.1.10). However, by only matching the sequences back to the repertoire of individuals from whom the sequences were not obtained, this specificity should be increased, as the sequences are unlikely to match to the repertoire of a different individual unless they are specific to the common antigen. For the PC+ cell sorting, some of the PCs may not be specific to the vaccine, although HLA-DR+ was included in the FACS gating, which should select for a recently activated and thus vaccine-enriched population (205). Perhaps the major limitation is that it is not possible to clone and functionally confirm the specificity of the sequences identified, as just the V_H chain was sequenced, so there are no paired V_L chain sequences to express them with. Currently paired V_H:V_L chain sequencing is technically challenging to perform in high throughput, although since conducting this study, there have been significant advances in this area, offering promise for future work (174).

Despite the limitations, with sufficiently stringent parameters, it appears possible to computationally enrich for vaccine-specific sequences in populations that have encountered a common antigen, and thus achieve an enriched population even when antigen-specific cell sorting is not possible. It is interesting that it was possible to enrich for vaccine-specific clusters at baseline before the participants had received the common vaccine stimulus. This may be due to all participants having been previously immunised with HepB vaccine, making it possible to detect historic HBsAg-specific memory cells. Alternatively, this could be due to the detection of polyreactive sequences, or could represent the level of background noise in the system. The ability to *de novo* enrich for antigen-specific sequences (even if not 100% specifically), has important clinical implications. In the context of vaccination, identifying the responding sequences can be used to identify mAbs as previously mentioned. In addition, numbers of these sequences could potentially be used for novel correlates of vaccine-mediated immunity (150); although expanded clusters could not be correlated with PC numbers in this study, studies

with larger numbers of participants are required to further investigate this. In the context of disease, in many cases, there are no definitive diagnostic tests. If common sequences or repertoire signatures could be identified between individuals with certain diseases, these could be used as diagnostic candidates. Here, it is necessary that the signatures in the repertoire be persistent, and not limited to day 7 following antigen encounters.

Whilst it was possible to identify common signatures that could be used to identify vaccine-specific sequences, there was considerable inter-individual variation in the global repertoire properties, and response dynamics in different individuals, which can give tailored insight into individual responses. For example, there was one participant (1212) who did not have a peak of PCs appearing in the peripheral blood at day 7, but did have in increase of anti-HBs in the blood, so must have produced vaccine-specific PCs at some time during the study period. Sampling at day 7 may have missed the PC burst in the peripheral blood (95), or there may be limitations in the sensitivity of the ELISpot assay. At the sequence level, there were clear day 7 clonal expansions in this participant, and these expansions were actually greater than in other participants. It could be the case therefore that this participant just produced a small number of PCs (undetectable by ELISpot), but that these PCs produced large amounts of antibody (consistent with the large clonal expansions). In addition, in one participant (1070), there was a highly expanded cluster ($> 10\%$ of total repertoire) present at every timepoint in both the IgG and IgM dataset. Such expansions may be caused by chronic infection (143) such as from cytomegalovirus, and although this was not formally tested, it highlights the additional immunological insights that can be gained from this sequencing technology. Analysis of the enriched vaccine-specific sequence repertoire of this participant revealed one of the strongest day 7 signals, which is interesting in light of recent work suggesting some chronic infections may enhance the immune response to vaccination (206).

4.5 Conclusion

To summarise, this chapter presents the use of BCR repertoire sequencing for a thorough dissection of the post-vaccination B cell response. There are day 7 repertoire signatures which appear to correlate with the PCs appearing at this time. In addition, repertoire convergence increases up to 21 days following vaccination, which has a more similar kinetic to appearance of memory cells. Measuring such signatures can give a detailed insight of an individuals response to vaccination, and may be applicable to studies of vaccine immunogenicity and function. The knowledge of these signatures allowed the development of a simple computational model to enrich for vaccine-specific clusters from the total repertoire. Finding consensus between computational enrichment and FACS-based vaccine-specific sequence enrichment lends validity to both methods. Such ability to identify sequences of importance in the response is key for fully utilizing the potential of BCR sequencing to understand immune responses to disease and vaccination.

5

Dynamics of the BCR repertoire following sequential HepB primary vaccination

5.1 Introduction

It was shown in the previous chapter that the total BCR repertoire undergoes certain stereotypic changes following vaccination - there is an increase in mutation, and decrease in diversity of the repertoire seven days following vaccination, consistent with an increase in the number of mutated PCs released into the peripheral blood at this time. There also appears to be a small number of similar B cell clones produced in different individuals (the so called public repertoire) when they are administered the same antigen. By studying the total BCR repertoire, Laserson *et al.* also showed that there are rapid expansions and contractions of certain clones in response to vaccination, but these expansion dynamics were qualitatively different in different individuals, and were not related to vaccine type, or efficacy (147). Relating such changes in the global repertoire to vaccine response is challenging, as they may be confused with concurrent subclinical responses to irrelevant antigens. The effect of this is shown in the striking repertoire changes in the unimmunised participant in Chapter 3, and the suspected activation at baseline in one participant seen in Chapter 4. In the previous chapter, methods

were explored to enrich for the vaccine-specific repertoire both computationally, and by FACS sorting vaccine-specific cells, which should overcome some of these challenges.

The present study follows on from this, and focuses on primary immunisation in HepB-naive participants rather than booster vaccination in previously immunised participants. The combination of these studies should therefore give insight into differences in primary vs. secondary responses at the level of the BCR repertoire. It as also useful for further elucidating the nature of the presumed vaccine-specific sequences that were present at baseline in the booster study - i.e., whether these represent remnants from their primary vaccination, polyreactive sequences, or the level of background noise from incorrect annotation of vaccine-specific sequences. Furthermore, as primary HepB vaccination consists of three separate doses, this study will also be able to show how the response changes between the doses (Figure 5.1). Two different dose schedules were used to see if any effect of the interval between doses could be detected.

The present study focused on IgG transcripts from total B cells to gain insight into total repertoire dynamics, and show differences to the dynamics following booster vaccination. To overcome the difficulties of discerning the vaccine-specific from total repertoire, cell sorting was also performed to enrich for vaccine-specific cells. Sequences from these vaccine-specific cells were combined with those from the booster study to generate a large vaccine-enriched sequence database. Matching this database to the total repertoire, enabled confident identification of the sequence clusters within the total repertoire that were vaccine-specific. Selectively studying the vaccine-specific cluster dynamics reduced the noise in the system, and indicated a surprising role for previously generated memory B cells in the response to vaccination.

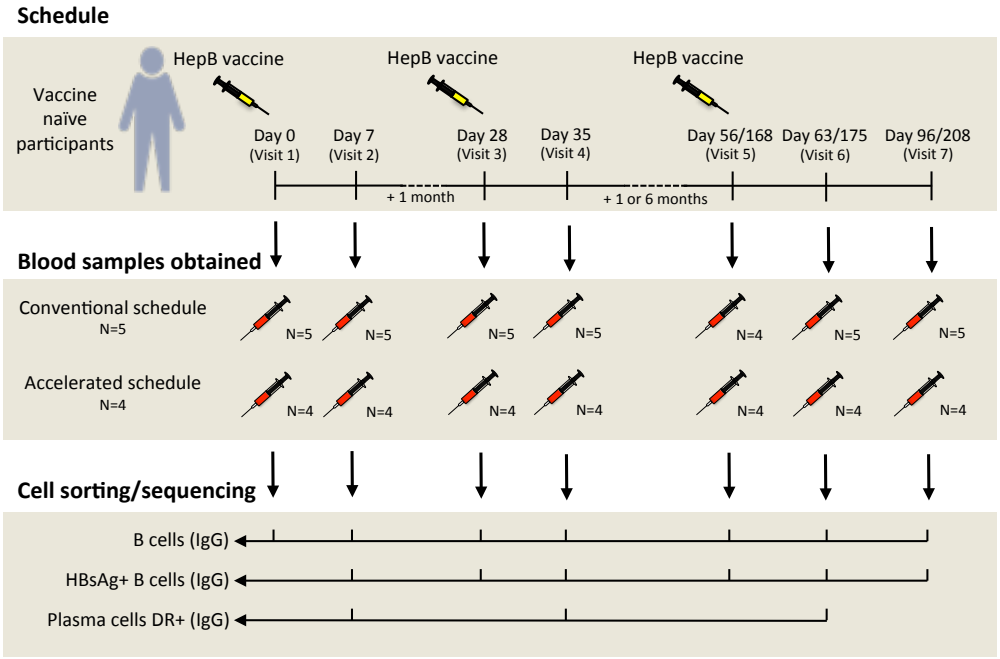

Figure 5.1: Study Design - Nine participants with no previous history of HepB vaccination were given a three dose primary course of HepB immunisation. Five participants were given the conventional schedule (month 0, 1 & 2), and four participants an accelerated schedule (month 0, 1 & 7). Blood was taken on the day of each vaccine, 7 days after each vaccine, and one month after the final vaccine. Total B cells were sorted and sequenced (IgG transcripts only) at each visit. HBsAg+ B cells were isolated and sequenced at each visit following the first vaccine, and PCs on the visit 7 days following each of the three vaccinations. Sequences from the HBsAg+ sorted cells, and sorted PCs both from this study, and the HepB booster study presented in Chapter 4 were used to find sequence clusters in the total repertoire that appeared to have enriched specificity towards the vaccine.

5.2 Methods

5.2.1 Study design

Nine healthy subjects (aged 20-38) with no prior history of HepB vaccination or infection were recruited with informed consent, under approval from the Northampton Research Ethics Committee (13/EM/0036). Participants were given a three-dose primary regime of monovalent HepB vaccine containing 10 μg HBsAg, adsorbed on amorphous aluminium hydroxyphosphate sulfate (HBvaxPRO, Sanofi Pasteur). Five participants were given a standard schedule (zero, one and seven months), and four participants were given an accelerated schedule (zero, one and two months). A sample of 50 ml peripheral blood was taken immediately before each vaccination as well as 7 days following each vaccination, and one month following the final

vaccine (Figure 5.1). Blood was transferred to a heparinized tube for processing within 4 hours of collection.

5.2.2 Sample processing

Sample processing followed the same protocol as described in Chapter 4, Section 4.2.3. Briefly, 2 ml of blood was used for determination of anti-HBs concentration, and PBMCs extracted from the remainder for enumeration of HBsAg-specific PCs and memory cells by ELISpot, and repertoire sequencing. At each visit, 500,000 B cells were sorted where available for sequencing the total repertoire. On visits 2, 4 and 6, both HBsAg+ and PCs were also sorted and sequenced, and on visits 3, 5 and 7, HBsAg+ cells only were sorted and sequenced (Figure 4.2). PCR was conducted for IgG transcripts only for all samples. Total repertoire samples were sequenced with 40-60 samples multiplexed on each run, and the HBsAg+ and PC samples were sequenced with 80-120 multiplexed on each run.

5.2.3 Sequence processing

For total repertoire samples, sequences were processed using the pipeline described in Figure 2.4. To account for differences in the number of resulting sequences in different samples, all samples were randomly subsampled without replacement using the sample function in R to give 75,000 sequences per sample. Clustering was performed using the thresholds determined in Chapter 3 - to be included in the same cluster, sequences were required to have the same V and J gene annotation, the same length CDR3, and no more than 1 AA mismatch per 12 AA's in the CDR3. To determine whether a cluster is present in more than one sample, data from all samples were clustered together. If two samples contribute at least one sequence to the same cluster, that cluster is defined as being present in both samples.

HBsAg+ and PC sequence data were combined with the data obtained in Chapter 4 to give four vaccine enriched datasets (HBsAg+ following booster, HBsAg+ following primary,

PC+ following booster, and PC+ following primary). Following quality control filtering and anntoation, sequences from these datasets were mapped to the total repertoire clustered dataset to see which clusters they would have been included in (i.e., same V and J gene annotation, and similar CDR3). These clusters in the total repertoire were then annotated according to which dataset the vaccine-enriched cluster came from. For the datasets obtained during this study, the HBsAg+ and PC sequences were only matched back to participants from whom those sequences were not obtained in order to reduce the effect of non-specific matching.

Lineage trees were generated for clusters with at least 50, or 25 sequences depending on the analysis, and subsampling performed if there were more than this number.

5.3 Results

5.3.1 Serological and cellular measures of vaccine response

All participants responded to the vaccination course, defined by having an anti-HBsAg antibody concentration greater than 100 mIU/ml at the end of the study. ELISpot was used to determine the number of HBsAg-specific IgG PCs and memory cells in the peripheral blood at each visit (Figure 5.2). The number of PCs detected 7 days following the first vaccine (visit 2) was negligible, and PCs were only produced in detectable numbers 7 days following the second and third vaccines (visits 4 and 6 respectively). Despite considerable inter-individual variation in the response, with some participants having a greater PC peak at visit 4, and some at visit 6 (Figure 5.2 B), the mean magnitude of the response was similar after each of these vaccines (visit 4: mean 13 PCs/10^6 PBMCs, range 0-58, visit 6: mean 14 PCs/10^6 PBMCs, range 0-44). The number of memory cells detected increased steadily throughout the course of the study, to reach the greatest number at the final visit.

Participants given the conventional schedule tended to have an increased B cell response after the third vaccine compared to those given the accelerated schedule. The mean PC num-

ber at visit 6 in the conventional schedule group was 19.7 PCs/10^6 PBMCs compared to 7.3 PCs/10^6 PBMCs in the accelerated schedule group (p = 0.0639; Mann-Whitney U test). The mean memory cell number at visit 7 in the conventional schedule group was 53.7 memory B cells/10^6 cultured lymphocytes compared to 9.6 memory B cells/10^6 cultured lymphocytes in the accelerated schedule group (p = 0.0651; Mann-Whitney U test).

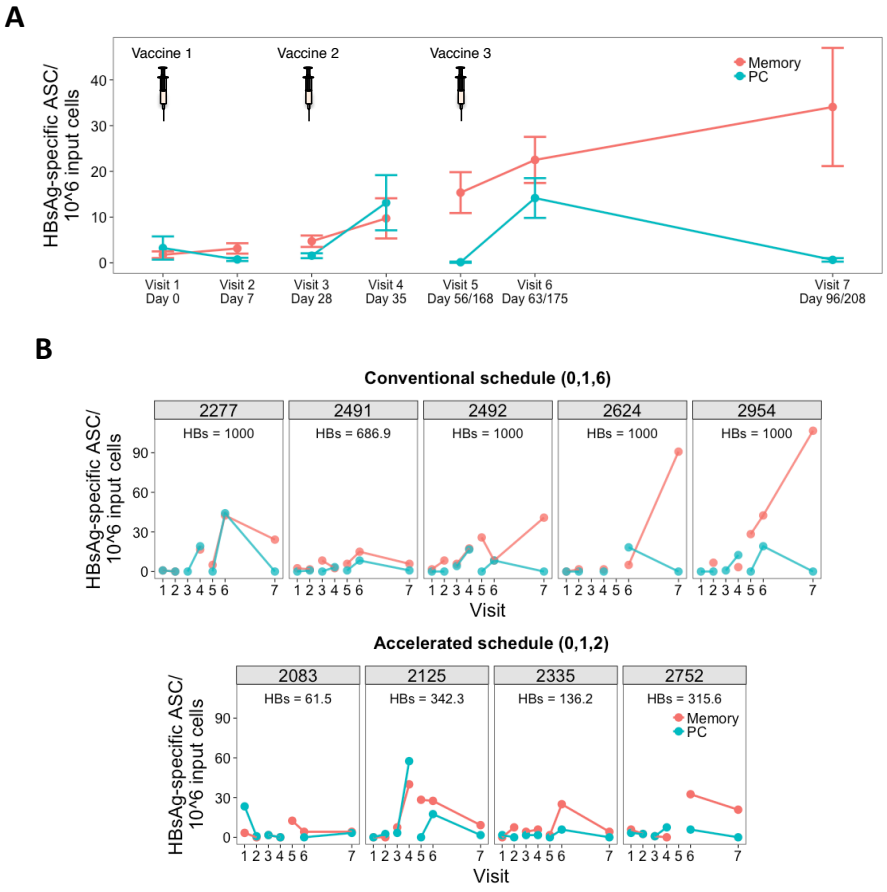

Figure 5.2: Clinical measures of vaccine response - (**A**) Kinetics of HBsAg-specific memory cells and PC numbers during the study period determined by ELISpot. Input cells are PBMCs for the PC detection assay, and cultured lymphocytes for the memory cell detection assay. Displayed are mean values ±SEM of the nine participants. (**B**) Same as **A**, but displaying the values for each individual participant. The number at the top of each plot shows the anti-HBsAg antibody concentration at the end of the study for that participant. Note that some points are missing due to either assay failure, or insufficient blood volume collected for the sample.

5.3.2 Dynamics of the total BCR repertoire

Total IgG repertoire data were successfully obtained from all samples except one, where a blood sample could not be obtained. On average, 308,100 (240,000 - 355,900) raw reads were

obtained for each sample, of which 111,200 (75,870 - 136,100) remained after all filtering steps (Table 9.5). Samples were normalized to give 75,000 sequences, which generated on average, 12,230 (4,163 - 19,240) clusters from each sample. PCA of V and J gene usage (Figure 9.8 A) as well as repertoire diversity, V gene mutation, and average CDR3 sequence length (Figure 9.8 B) was conducted to detect outliers. Samples from participant 2277 visit 7, and 2492 visit 4 appeared to be potential outliers based on repertoire diversity, mutation, and CDR3 length.

In the previous chapter, it was shown that there is an increase in cluster expansion, an increase in mutation, and a decrease in CDR3 AA sequence length seven days following administration of HepB booster vaccine. In addition, there was an increase in sequence convergence 14 and 21 days following vaccination. These same properties were analysed in the current dataset for comparison of primary and booster vaccination (Figure 5.3). There were still expansions of some clusters at each of the day 7 visits following each vaccine (Figure 5.3 A), however, this was only significant following the third vaccine, and there were also striking cluster expansions in two of the participants at baseline (Figure 9.9). Changes in mutation and CDR3 length were less pronounced in this study, and although the day 7 changes were in the same direction as after booster vaccine, they were not significant here (Figure 5.3 B/C). Changes in sequence convergence on the other hand were more pronounced in this study, than following the booster vaccine (Figure 5.3 D). The percent of convergent clusters was greater after each of the three vaccines than at baseline, and remained at higher levels than baseline even at the final visit ($p = 0.0001$; Mann-Whitney U test).

5.3.3 Enriching for the vaccine-specific repertoire

From each participant, on average 26,720 HBsAg+ B cells across visits 2-7, and 2,748 PCs on the day 7 visits were isolated and sequenced (Table 9.6). These data were combined with the sequence data from 85,000 HBsAg+ and 3,444 PCs obtained after the HepB booster vaccine study (Chapter 4) to create a large vaccine-enriched sequence database. Each of these four

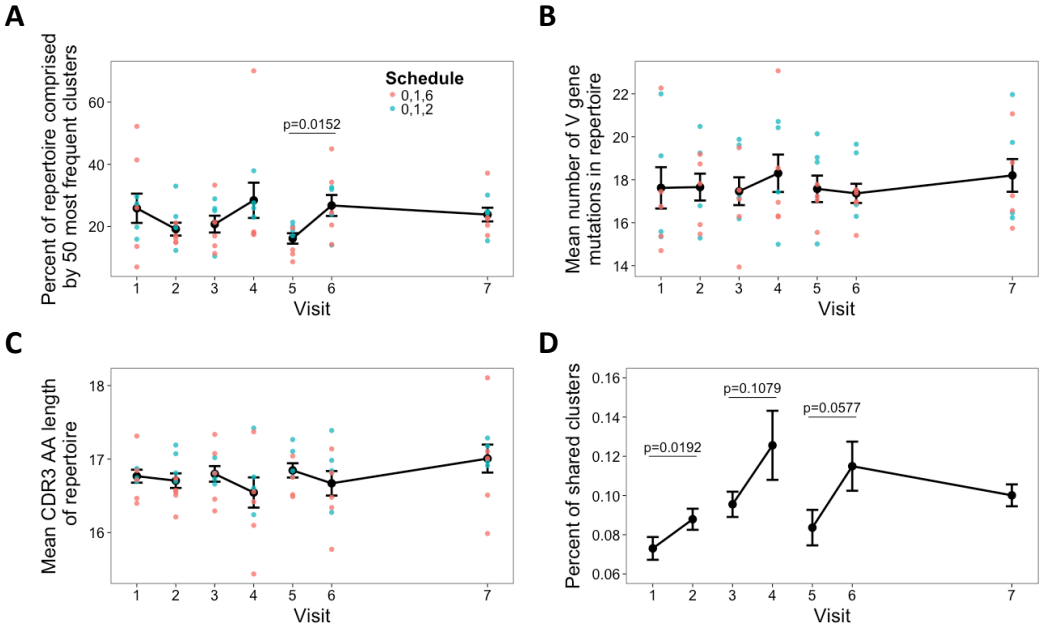

Figure 5.3: Changes in the total repertoire - (**A**) Mean percent of the total repertoire comprised by the 50 most frequent clusters at each day. (**B**) Mean number of V gene mutations from all sequences in the repertoire. (**C**) Mean CDR3 AA sequence length from all sequences in the repertoire. For **A-C**, error bars indicate ±SEM from 9 participants. Coloured points represent data from individual participants, and are split by vaccine schedule. (**D**) The percent of clusters shared by each pair of participants at each visit was determined (9 participants, giving 36 pairings). Shown are the mean values ±SEM of the percent of clusters shared between each pair. Percent is calculated as (A∩B/sum(A,B))*100. All p values were obtained from Mann-Whitney U tests.

vaccine-enriched sequence datasets (HBsAg+ after primary vaccine, HBsAg+ after booster vaccine, PC+ after primary vaccine and PC+ after booster vaccine) were used separately to annotate the total repertoire clusters. Irrespective of participant, of the 632,316 total clusters, 4,744 (0.71%) were annotated based on the HBsAg+ data obtained from this study, 2,445 (0.32%) based on the PC data from this study, 709 (0.11%) based on the HBsAg+ data from the booster study, and 419 (0.07%) based on the PC data from the booster study (Figure 5.4 A). The overlap in the clusters annotated by each of these datasets was significantly more than is expected by chance (Chi-square test; $p < 0.0001$ for each overlap).

As some of the sequences in these vaccine-enriched datasets may not actually be specific to the vaccine antigen, due either to non-specific staining for HBsAg+, or the inclusion of some non-specific PCs, a strict definition was used to find vaccine-specific clusters in the total

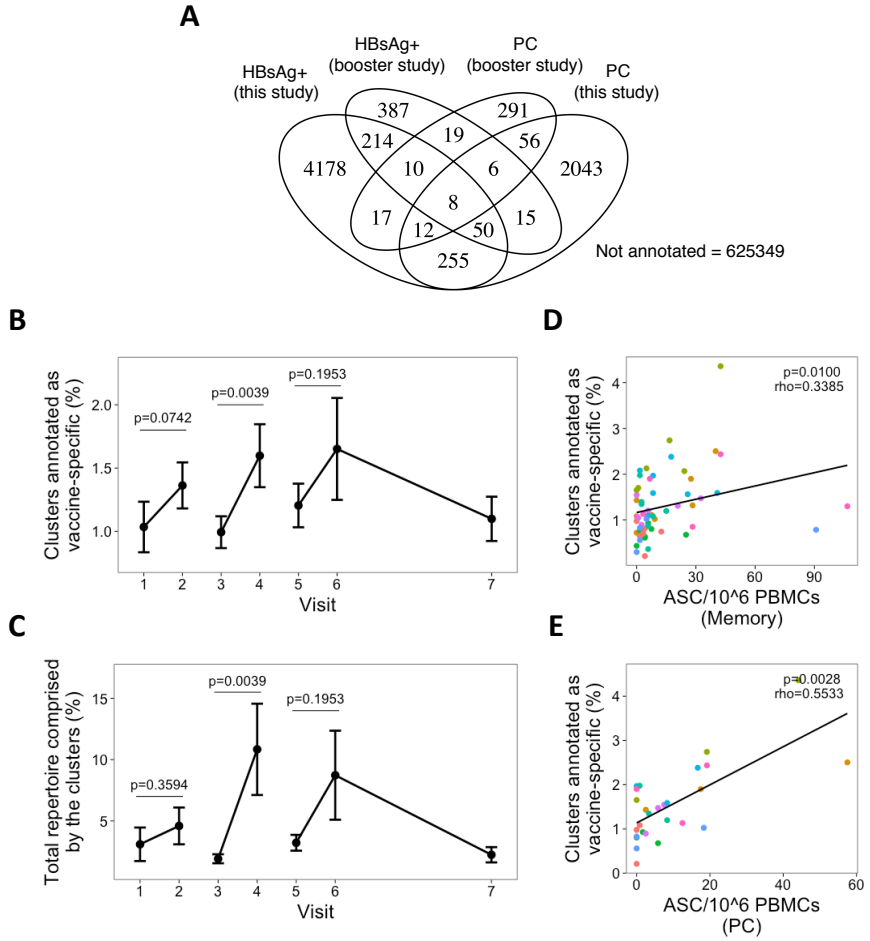

Figure 5.4: Enriching vaccine-specific clusters from the total repertoire - (**A**) Venn diagram illustrating the number of clusters annotated by the four vaccine-enriched datasets. (**B**) The percent of abundant (> 0.01% of total repertoire) clusters at each visit that were characterized as vaccine-specific based on being annotated by at least two of the vaccine-enriched datasets. (**C**) Same as **B**, but corrected for cluster size by considering the percent of the repertoire comprised by the vaccine-specific clusters. For **B** and **C**, mean values ±SEM are shown for all 9 participants, and p values were obtained from Mann-Whitney U tests. (**D**) Correlation (Spearman) between the percent of abundant clusters characterized as vaccine-specific, and PC numbers determined by ELISpot. Different colored points represent samples from the different participants. (**E**) Same as **D**, but correlated with memory cell numbers determined by ELISpot. For **D** and **E**, samples where no cells were detected by ELISpot were omitted.

repertoire dataset. To be considered vaccine-specific, a cluster had to be annotated by at least two of the vaccine-enriched datasets, and also be present at a frequency greater than 0.01% (ie, contain at least 8 sequences). For the annotation, the HBsAg+ and PC+ sequences were only matched back to total repertoire data from participants from whom the HBsAg+ and PC+ sequences were not obtained. On average 1.03% of the frequent clusters were annotated as vaccine-specific at visit 1, but this number increased to 1.36% by visit 2 (Figure 5.4 B). The

number increased further 7 days after each of the two subsequent vaccines to peak at 1.65% by visit 6. The clusters annotated as vaccine-specific tended to be large, so that when considering the percent of the repertoire comprised by these clusters, these post-vaccination changes were more pronounced (Figure 5.4 C). It is notable that a similar kinetic is seen, even when using a looser definition (annotation by just a single vaccine-enriched dataset) of what is considered a vaccine-specific clusters (Figure 9.10).

In total, 355 vaccine-enriched clusters were found, using the strict definition, across all participants regardless of which visit they were detected on. The percent of frequent clusters annotated as vaccine-specific in each participant at each visit correlated with the number of HBsAg-specific memory and PCs detected by ELISpot (Figure 5.4 D/E), giving validity to the technique for enriching vaccine-specific clusters.

5.3.4 The vaccine-specific repertoire has distinct features compared to the total repertoire

Next, the properties of the vaccine-specific clusters were compared to the total repertoire. For comparison, a random set of size matched clusters (Same size as each vaccine-specific cluster ±50 sequences) from the total repertoire were obtained that had no annotation for vaccine-specificity. A size-matched sample was taken to discount any differences between the datasets as being only due to cluster size rather than antigen specificity. For 13 of the vaccine-specific clusters, no size-matched counterpart could be found, so these were excluded, yielding 342 vaccine-specific and 342 size-matched clusters. In addition, a non size matched set of 342 random clusters that had no annotation for vaccine-specificity was also obtained. The size-matching approach was effective, giving an average cluster size of 249 for the vaccine-specific clusters and 241 for the size-matched cluster compared to just 11 for the random clusters (Figure 5.5 A). The vaccine-specific clusters, and size-matched clusters had similar levels of mutation, and this was greater than that of the random clusters ($p = 0.0061$; Figure 5.5 B).

The vaccine-specific clusters also had significantly shorter CDR3 regions than either the size-matched or random clusters ($p < 0.0001$; Figure 5.5 C). Looking within an individual at how many visits each cluster was present at showed that the vaccine-specific clusters were more likely to be present at multiple visits (Figure 5.5 D). Furthermore, the vaccine-specific clusters were also more likely to be present in multiple participants (Figure 5.5 E).

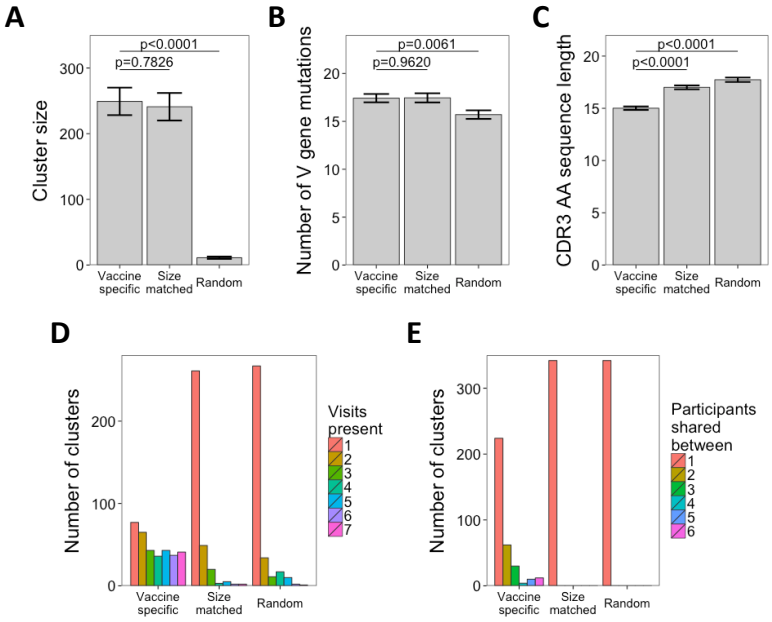

Figure 5.5: General properties of vaccine-specific clusters compared to the random and size-matched clusters - (**A-C**) Differences in size, number of V gene mutations, and CDR3 AA sequence length of the clusters belonging to the three datasets. Shown are the mean values ±SEM of the 312 clusters in each dataset. Comparisons were performed using a t-test. (**D**) The number of visits where at least a single sequence from each cluster is found was determined, and the number of clusters present at different numbers of visits in the different datasets plotted. (**E**) Same as **D**, but counting the number of participants where a similar cluster is found (same CDR3 cluster center sequence, and V/J gene usage). Shown are the mean values ±SEM. Comparisons were performed using a t-test.

In each participant, the proportion of clusters utilising different V gene segments was determined for the vaccine-specific, size-matched, and random clusters. While the size-matched, and random clusters had a similar distribution of V segment usage, the V segment usage of the vaccine-specific clusters differed to both of these (Figure 5.6). Most strikingly, usage of the V segment IGHV3-7 was the most commonly used in the vaccine-specific clusters, but was used to a much lesser extent in the size-matched and random clusters ($p < 0.0001$; t-test).

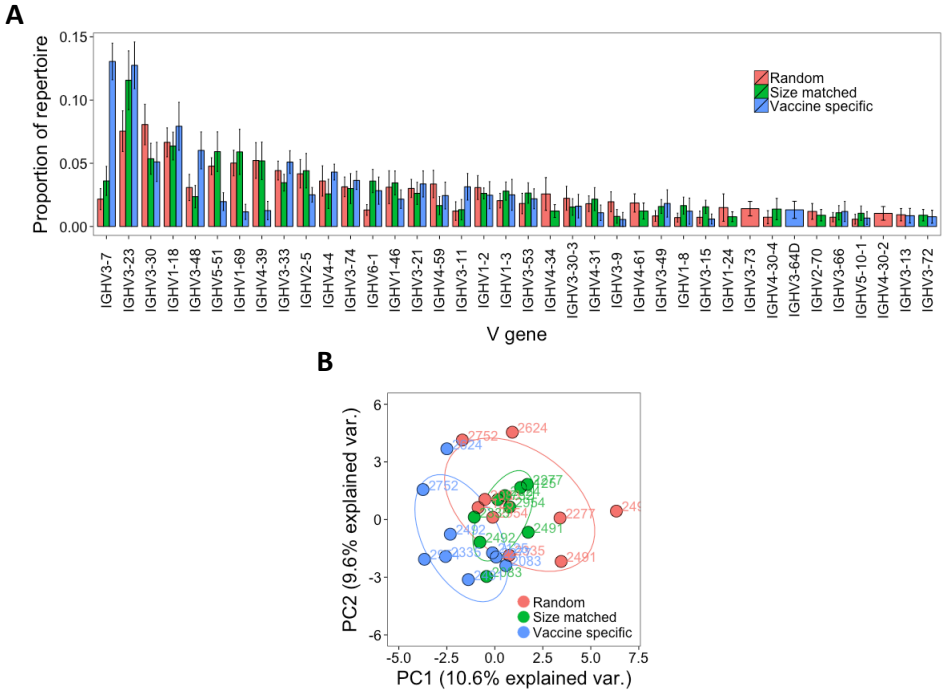

Figure 5.6: V gene usage of the vaccine-specific clusters compared to the random and size-matched clusters - (**A**) Histogram showing the proportion of clusters utilizing each of the different V genes. Error bars show ±SEM of the 9 participants. (**B**) Principle component analysis based on V gene usage for the vaccine-specific, random and size-matched clusters obtained from each of the 9 participants.

To investigate the structure of the clusters, lineage trees were constructed from each of the vaccine-specific, and size-matched clusters that contained at least 50 total sequences (Figure 5.7 A). Structure of the lineage trees can give insight into the degree of proliferation which the clonal B cells within the cluster are currently undergoing - a greater diversity of sequences within the cluster likely indicates greater proliferation. Furthermore, by estimating the most recent common ancestor to the clone, it is possible to determine how mutated this sequence is compared to the germline sequence (Figure 5.7 A; trunk length), and thus infer the maturation level of the initiating B cell for each clone (207). Vaccine-specific lineages both contained a greater diversity of sequences (Figure 5.7 B), and had a shorter trunk length (Figure 5.7 C) than the size-matched random lineages, indicating that they are undergoing greater proliferation, and are more closely related to germline.

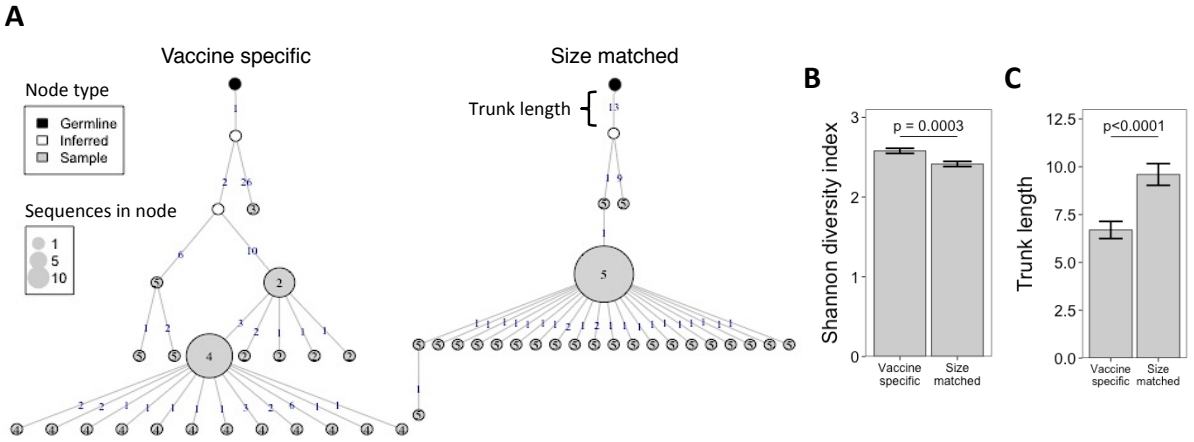

Figure 5.7: Lineage structure of the vaccine-specific and size-matched clusters - (**A**) Representative lineage trees of one vaccine-specific, and one size-matched cluster. Each node represents a unique sequence within the cluster, with the size of the node indicative of the number of duplicate sequences. The number within the node indicates the visit that the sequence is present at. Shading of the node represents whether the sequence is found in the cluster, an inferred common ancestor to sequences found in the cluster, or the germline sequence. Numbers on the edges of adjoining nodes show the number of mutations between the sequences. (**B**/**C**) Lineage trees were created for all clusters which contained at least 50 sequences in the dataset (n=232). Diversity was calculated for each cluster using the Shannon index, and trunk length is the number of mutations between the most recent common ancestor, and germline sequence. Shown are the mean values ±SEM. Comparisons were performed using a t-test.

5.3.5 Dynamics of the vaccine-specific repertoire, and evidence for vaccine-specific sequences prior to vaccination

Having identified a vaccine-specific repertoire with distinctive features, the kinetics of the vaccine-specific clusters were investigated in each participant over the course of the study (Figure 5.8). There were considerable expansions of vaccine-specific clusters 7 days after each vaccine dose in the majority of cases, and within an individual, many of the same clusters were recurrently expanded after each vaccine dose. While the number of these vaccine-specific clusters at each day 7 visit following vaccination was moderately correlated with the number of HBsAg-specific PCs detected by ELISpot in each participant (Figure 5.4 D), there were also striking expansions of the vaccine-specific clusters 7 days after the first vaccine, despite no PCs detected by ELISpot at this time. Furthermore, many of the clusters expanding after each vaccine were also present at low frequency at baseline, despite none of the participants having previously encountered the vaccine antigen. Although the percent of vaccine-specific clusters

after each vaccine, which were also present at baseline, was similar after each of the vaccine doses (Figure 5.9 A), correlating the percent of frequent clusters annotated as vaccine-specific with the ELISpot data revealed a strong correlation when considering only the vaccine-specific clusters not present at baseline, but no correlation with considering the vaccine-specific clusters that are present at baseline (Figure 5.9 B). This information suggests that the baseline clusters are inherently different to those first identified at later visits.

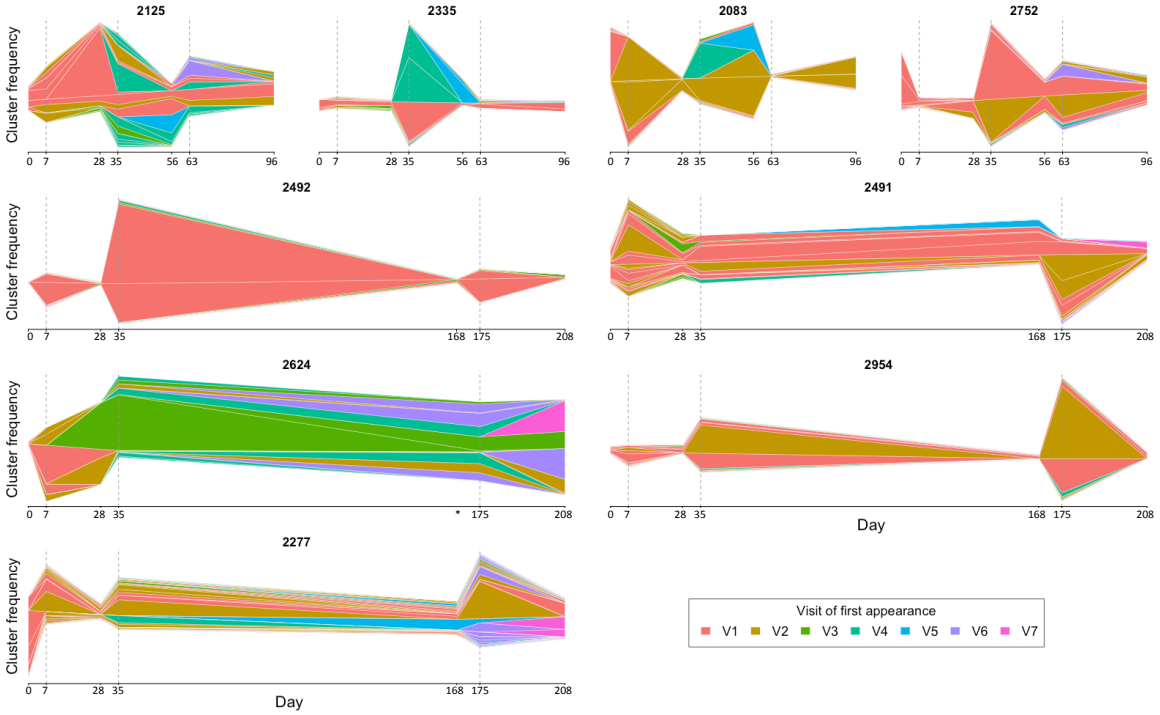

Figure 5.8: Kinetics of the vaccine-enriched clusters - The vaccine-enriched clusters were found in each participant, and at each day, the frequencies of these clusters plotted as a stacked bar chart, centered to the middle of the y axis. Clusters from each day are then joined using a horizontal stream to illustrate how the frequency of the clusters changes over time. The width of the stream represents the frequency of the cluster at that time, and the color of the stream represents the first visit at which sequences from the cluster can be found. The top four plots are from participants who were given the accelerated vaccine schedule, and the bottom five plots are from participants who were given the conventional vaccine schedule. Dotted vertical lines highlight the day 7 post vaccination visits. *The day 168 blood sample was missing from this participant, but the vaccine was still given on this day.

One potential cause of finding these vaccine-specific clusters at baseline is due to insufficient stringency in the definition of what comprises a vaccine-specific cluster, and that those present at baseline were erroneously identified, and are not actually specific to the vaccine.

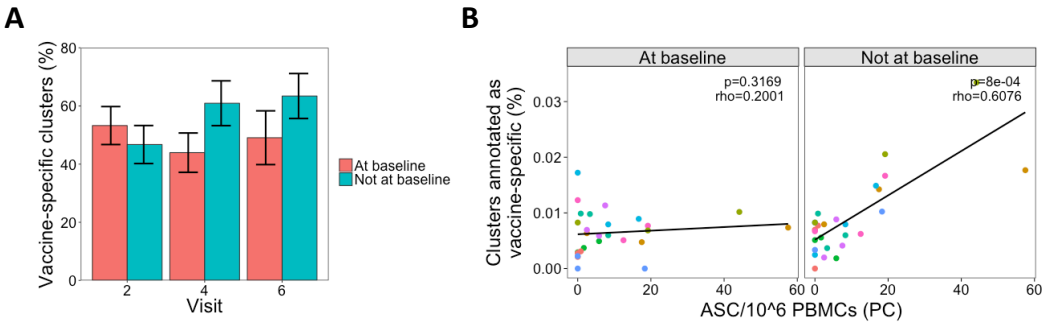

Figure 5.9: The number of vaccine-specific clusters after each vaccine that are also present at baseline - (**A**) In each of the nine participants, of the vaccine-specific clusters present at each day 7 post-vaccination timepoint (visit 2, 4, or 6), the percent of these that was also present at baseline was determined. Mean ±SEM shown. (**B**) Correlation (Spearman) between the percent of abundant clusters characterized as vaccine-specific, and PC numbers determined by ELISpot, split according to whether the vaccine-specific clusters are also present at baseline. Different coloured points represent samples from the different participants.

To investigate this, the previously characterised HBsAg-specific antibodies identified from the literature (Chapter 3, Table 9.2) were also compared to the dataset to see if any could be found at baseline. In total, there were 12 previously described sequences that mapped to clusters in the current dataset based on CDR3 AA sequence identity. Although there were not enough of these sequences present to construct detailed plots of their kinetics, they were found at baseline in six of the participants (Figure 9.11).

5.3.6 Investigating differences between the vaccine-specific clusters expanded after each vaccine

A hypothesis for why vaccine-specific clusters may be present at baseline is that they are derived from memory B cells stimulated previously by a different antigen, but that are also able to recognise HBsAg. These memory B cells would be expected to have a relatively low affinity for HBsAg, and might be less detectable by ELISpot. Higher affinity B cells detectable by ELISpot are then formed from activation of naive B cells following the second and third vaccines. Two different methods were used to investigate this hypothesis. First, to test whether the vaccine-specific clusters present at baseline were polyreactive, they were compared with sequences of known specificity to see if any of them resembled sequences with other known specificities.

Indeed, of the 137 vaccine-specific clusters present at baseline, 1 of these contained an identical CDR3 AA sequence to a previously described TT-specific sequence (80).

Next, the properties of the vaccine-specific clusters found 7 days following each of the three vaccinations was investigated in more detail. However, these properties are confounded by having two populations of clusters produced after each vaccine: those newly generated in response to the vaccine, and those that are also present at an earlier visit and just being re-stimulated (Figure 5.10 A). At each visit, the clusters were therefore split into groups according to the first visit at which they appeared. Considering the mutation level of these clusters, those first stimulated after the second and third vaccine doses were significantly less mutated than those first stimulated after the first vaccine dose (Figure 5.10 B). Those found after the first vaccine dose, and subsequently re-stimulated and found after the later vaccine doses retained a similar level of mutation throughout, whereas those found for the first time after the second vaccine dose, and subsequently re-stimulated after the third vaccine dose increased in mutation.

Where these clusters contained at least 25 sequences, lineage trees could be constructed from them. The trunk length of the lineages followed a similar pattern to mutation of the clusters. Lineages found after the first vaccine dose had longer trunk lengths than those first found after the second or third vaccine doses, indicating that the clusters found after the first vaccine dose are derived from more mature precursor cells than those found after the second and third doses (Figure 5.10 C). Considering the diversity of sequences within the lineages, this was similar for the clusters present first appearing after each vaccine dose (Figure 5.10 D). However, for the clusters that are first present after the first vaccine dose, and then re-stimulated, these became more diverse after the subsequent doses. Taken together, the data presented in Figure 5.10, gives evidence that the response to the first vaccine dose is dominated by activation of pre-mutated memory B cells, while less mutated naive cells are used more in the response to the subsequent doses. It also appears that these pre-mutated memory B cells

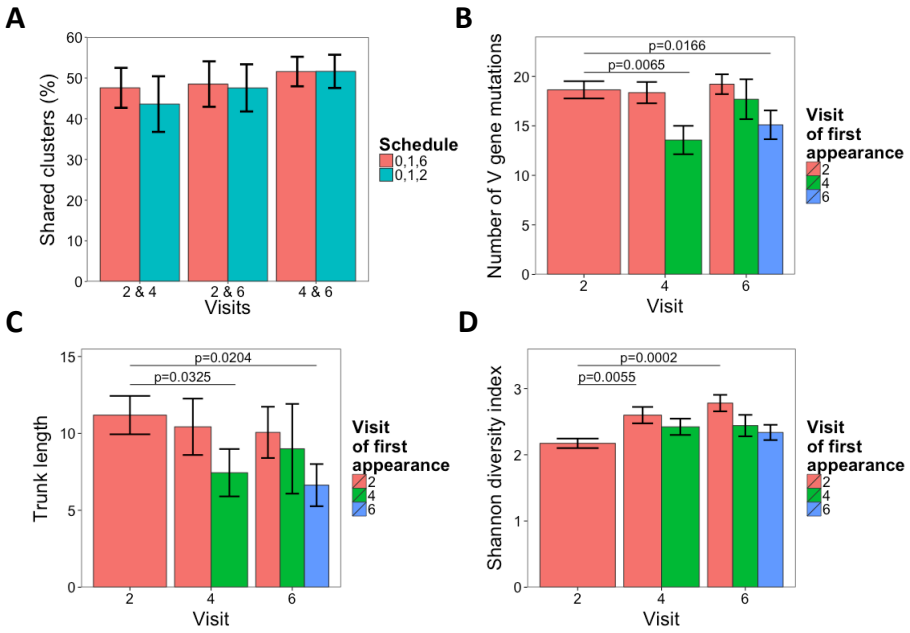

Figure 5.10: Properties of vaccine-specific clusters at each visit - (**A**) Percent of clusters that are shared between the different day 7 post-vaccination visits in each participant. At each of the day 7 post-vaccination visits, vaccine-specific clusters were labelled according to the first post-vaccination visit where they were identified. (**B**) Difference in the number of V gene mutations of the vaccine-specific clusters at each post-vaccination visit. Lineage trees were then constructed for each of these clusters where there were at least 25 sequences, and trunk length (**C**), and sequence diversity within the lineage (**D**) calculated. Shown are mean values ±SEM. Comparisons were performed using a Mann-Whitney U test.

require more diversification than the naive cells. Although at visit 6, the timing of the vaccine dose is different in the two vaccine groups, this did not have a significant effect on the properties of the vaccine-specific clusters at this time (Figure 9.12).

5.4 Discussion

In this chapter, a three dose HepB primary vaccination schedule was used as a model system to study the BCR repertoire response to a novel antigen. It was shown that changes in the total repertoire following vaccination involve a small minority of sequences, making them challenging to distinguish from background fluctuations. By focusing on the vaccine antigen-specific repertoire, clearer, and more detailed insight can be gained into the responding B cells than has previously been possible. It was thus possible to demonstrate that a large proportion of the responding sequence clusters are present at baseline in an IgG population, and have fea-

tures, which suggest that they are a mature population, previously generated in a response to a presumably unrelated antigen. The involvement of these clusters in response to HepB vaccine suggests that they also have a degree of cross-reactivity with HBsAg.

In the previous chapter, it was demonstrated in the context of a HepB booster vaccine, there are changes in features of the global B cell repertoire (increase in cluster expansion, increase in mutation, and decrease in CDR3 length). These changes were also observed to an extent in the current dataset, and most pronounced after the second and third vaccine doses. However, as before, overall changes in these properties were affected by background fluctuations in the non-specific repertoire of these individuals that obscured the vaccine-specific effects. Additional noise may also come from the use of two different vaccination schedules, although no significant differences were observed between the schedules. Furthermore, strong hallmarks of repertoire activation were observed in at least two of the participants in this study prior to any vaccination, highlighting the need to focus on the vaccine-specific repertoire rather than the total repertoire. Despite the background noise, the signal of repertoire convergence 7 days after vaccination was actually stronger in this dataset than after the booster vaccine. This may relate to the nature of a primary vs secondary response; convergence increases following the primary response, and then persists for long periods such that when given the booster vaccine, the increase in convergence is less pronounced. Nevertheless, this provides further evidence for the convergent repertoire being highly enriched for vaccine-specific sequences.

To circumvent the problems with background noise in the total repertoire, cell sorting was used to isolate and sequence vaccine-specific cells, and a strict procedure used to identify the vaccine-specific clusters within the total repertoire. However, despite referring to these clusters as vaccine-specific, it is not possible to be certain of their specificity, as they cannot be expressed and characterized due to the lack of a paired V_L chain. This is a limitation that should be removed in future studies, as techniques for high-throughput pairing of V_H and

V_L chains emerge (174). Comparison of the vaccine-specific clusters to random size-matched clusters does however reinforce the idea that they are specific to the vaccine, as they tend to be more recently activated from germline, and undergo increased diversification during the course of the study.

Studying the vaccine-specific repertoire yielded a number of interesting observations. It is striking that prior to vaccination all participants had clusters annotated as vaccine-specific despite never having previously encountered the vaccine antigen. Whilst this may be expected in the naive repertoire, only the class-switched IgG repertoire was sequenced here. This population represents previously activated cells, so is not expected to contain cells specific to the antigen. Although finding vaccine-specific sequences at baseline could represent an artifact from incorrect labeling of the vaccine-specific repertoire, sequences matching previously characterized HBsAg-specific antibodies from the literature were also found at baseline, indicating that this is not the case. It seems likely therefore, that the clusters present at baseline must represent B cells that have previously been activated in response to different antigens, and that are either polyreactive, or happen to have a degree of cross-reactivity with HBsAg. Such a finding is backed up by previous reports, which show that polyreactive B cells are a major constituent of the normal human B cell repertoire (208), and also specifically show that HBsAg is able to activate these (27, 209).

Tracking the dynamics of the vaccine-specific repertoire indicates a degree of response after each of the three vaccine doses. After the first vaccine dose, for most participants, the majority of the responding clusters are those that are also present at baseline, highlighting the large extent to which recruitment of these potentially cross-reactive B cells occurs. After subsequent vaccines, there is more recruitment of clusters that are not present at baseline, but some baseline clusters are still recruited. This is therefore indicative of recruitment of naive vaccine-specific cells occurring concomitantly to re-stimulation of the initial cross-reactive cells. It is notable

that despite detecting these responses in the sequence data after the first vaccine, there were no responses detected by ELISpot. It could be that repertoire sequencing is simply a more sensitive method for detecting the vaccine-specific effects (500,000 B cells used for repertoire sequencing vs. 200,000 PBMCs/∼20,000 B cells per well for ELISpot), or alternatively that, because most of the B cells activated after the first vaccine are suspected to be derived from a cross-reactive response, they are likely to have too low affinity for HBsAg for their antibody to be detected by ELISpot. Indeed, the finding that when selectively considering the vaccine-specific clusters present at baseline, their numbers do not correlate with the ELISpot data, but when selectively considering the vaccine-specific clusters not present at baseline, their numbers strongly correlate with the ELISpot data backs up this observation. Plasmablasts that have only recently differentiated, and may not yet be secreting large amounts of antibody could also preclude detection by ELISpot (210).

5.5 Conclusion

To summarise, these data have provided significant insight into B cell kinetics following repeated antigen stimulus in a naive population. Focusing on the vaccine-specific repertoire reduced the background noise, allowing these data to yield additional insights beyond those available from conventional techniques such as ELISpot. The finding of vaccine-enriched clusters following booster vaccine (Chapter 4) at baseline is backed up in this study, with further investigation indicating that they are derived from cross-reactive B cells. It will be interesting to investigate whether this also occurs in the response to other vaccines, and the degree to which cross-reactive activation affects the level of protection conferred by the vaccine.

6

Using BCR sequencing to track patterns of B cell activation following vaccination with meningococcal polysaccharide and protein-polysaccharide conjugates

6.1 Introduction

For most vaccines, protection is achieved via activation of B cells with vaccine antigen-specific receptors, which subsequently differentiate into PCs and produce functional antigen-specific antibody. Immunogenicity is conventionally assessed by measures of vaccine-specific antibody quantity and function, but this gives little insight into which B cell subsets were activated to generate the functional antibody response. It has been shown in Chapters 4 and 5 that NGS approaches to study BCR repertoires can be used to measure the diversity of B cell populations, and allow resolution of vaccine response at the level of individual B cell clones. To date, BCR repertoire studies of vaccine response have focused on total B cells, or PCs. However, diverse B cell subsets may be involved in a response, including naive, MZ and memory (IgM and IgG) B cells, depending on the type of antigen, previous exposure, and route of immunisation. Analysis of different B cell subsets has revealed differences in their sequence, and VDJ gene segment

composition (145, 211), and thus interrogation of the BCR repertoire on a subset-by-subset basis could potentially be used for fine delineation of which B cells are involved in vaccine responses.

Vaccines against polysaccharide-encapsulated pathogens (e.g. *N. meningitidis* and *S. pneumoniae*) are of great importance in controlling these predominantly childhood diseases. Polysaccharides themselves are TI antigens, have poor immunogenicity in infants (212), do not generate immunological memory (86), and can lead to hyporesponsiveness with repeated exposure (94). The conjugation of a carrier protein to a polysaccharide results in a TD antigen, which is immunogenic from early life, and primes for memory (87). Licensed quadrivalent ACWY *N. meningitidis* vaccines contain either plain purified capsular polysaccharides (polysaccharide vaccine), or the same polysaccharides conjugated to a carrier protein (conjugate vaccine). The difference in immunogenicity, and the different B cell subsets involved in the response to these related vaccines are still being elucidated. The different B cell subsets activated by conjugate and polysaccharide vaccines, have previously been investigated during a comparative study of pneumococcal conjugate and polysaccharide vaccines (93). The conjugate vaccine induced more circulating serotype-specific memory B cells than the polysaccharide vaccine (93). However, despite previous suggestion that polysaccharide antigens stimulate MZ B cells (213), there was no difference seen in the frequency of serotype-specific MZ B cells measured in peripheral blood after the two vaccines in this study, perhaps due to limitations in the sensitivity of the flow cytometry assay (93).

This study sought to determine the utility of BCR repertoire sequencing as a tool for investigating the different B cell subsets used in different vaccine responses, using meningococcal ACWY polysaccharide and conjugate vaccination as a model system. Participants were immunised with either a polysaccharide or conjugate vaccine, followed by a further immunisation with a conjugate vaccine 4 weeks later (Figure 6.1). Naive, MZ, IgM memory and IgG mem-

ory B cell subsets were isolated at baseline, in addition to PCs 7 days after each vaccination (Figure 6.1). This day 7 time point was chosen as previous work from our laboratory using the same vaccine regime has shown the presence of significant numbers of antigen-specific PCs 7 days after both vaccine doses (214), and induction of effective antibody responses in both groups (215). Hence, sorting PCs at this time point would capture a population enriched for vaccine antigen-specificity. Day 7 PCs were also further split based on HLA-DR expression, to distinguish recently activated PCs (HLA-DR+) likely to be enriched for vaccine specificity, from long-lived PCs (HLA-DR-), which are considered to be displaced from the bone marrow by arrival of the newly generated PCs and likely have a broader range of specificities (205). Characterising mutation levels, and IgG subclass usage of the PC sequences, as well as relating them to the sequences of the baseline B cell subsets, allowed determination of the B cell subsets involved in the response to the two vaccines, and to distinguish TI from TD responses.

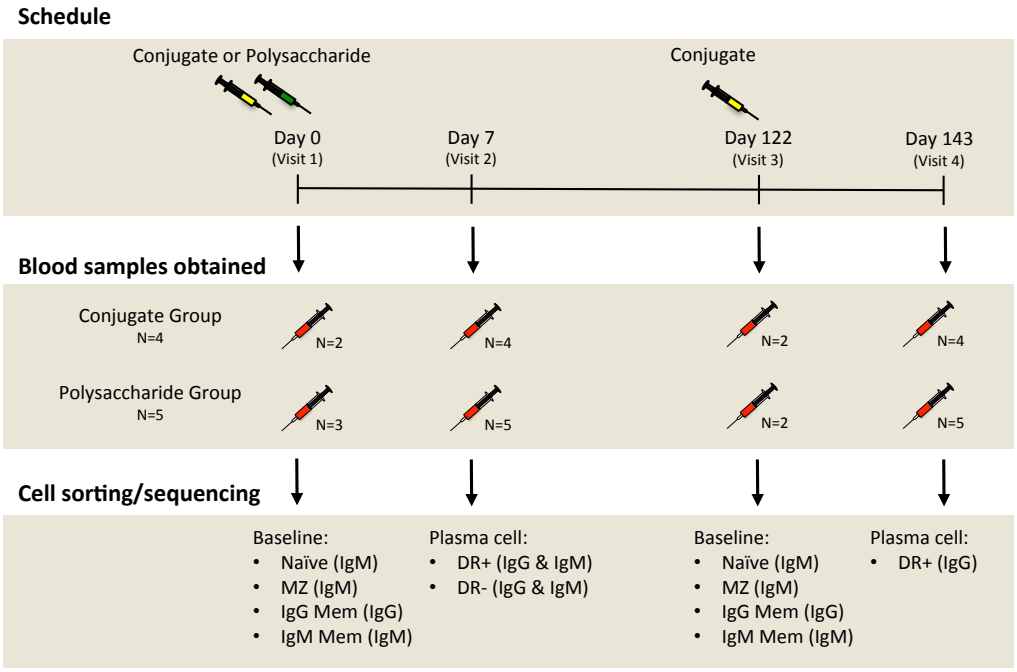

Figure 6.1: Study Design - Four participants were given a conjugate, and five participants given a polysaccharide MenACWY vaccine at day 0. All participants were given the conjugate vaccine at day 28. All participants had blood taken 7 days after each vaccine for PC repertoire analysis. Each participant also had blood taken on either the day of the first vaccine or the day of the second vaccine (but not both) for repertoire analysis of baseline B cell subsets. FACS was used to isolated the different B cell subsets at each visit.

6.2 Methods

6.2.1 Study design

Nine healthy participants were recruited with informed consent in accordance with the Declaration of Helsinki, and under approval from the Oxfordshire Research Ethics Committee (12/SC/0275). Participants were aged between 30 and 70 years, and had no prior meningococcal vaccine history. All participants were given two doses of meningococcal vaccine containing serogroup A, C, W and Y polysaccharides. Following a first dose of either ACWY polysaccharide vaccine (MenACWY-PS; ACWYVax; GSK Biologicals, Rixensart, Belgium) given to five individuals, or ACWY conjugate vaccine (MenACWY-CRM197; Menveo; Novartis Vaccines, Siena, Italy) given to four individuals (Figure 6.1), all individuals received a dose of MenACWY-CRM197 at day 28. All vaccines were administered as intramuscular injections. A sample of 50 ml peripheral blood was taken from participants on the day of the first vaccine (visit 1), 7 days after the first vaccine (visit 2), 28 days after the first vaccine (visit 3) and 7 days after the second vaccine (visit 4), and transferred to a heparinized tube for processing within 4 h of collection.

6.2.2 Sample processing

For each blood sample, total B cells were isolated by MACS. On visit 1, and visit 3, these B cells were stained with Live/dead-Aqua (Life technologies), CD19-FiTC (HIB19) (eBioscience), IgD-PE (IgD26) (Miltenyi Biotec), CD27-PECy7 (O323) (eBiosceince) and IgM-APCCy7 (MHM-88) (Biolegend), for isolation of naive (CD19+, CD27-, IgM+, IgD+), MZ (CD19+, CD27+, IgM+, IgD+), IgM memory (CD19+, CD27+, IgM+, IgD-) and IgG memory (CD19+, CD27+, IgM-, IgD-) B cell subsets (Figure 6.2). On visit 2 and visit 4, cells were stained with Live/dead-Aqua, CD19-FITC, CD38-PE (HIT2) (eBioscience), CD27-PECy7 and HLA-DR-APC (L243) (Biolegend) for isolation of recently activated PCs (CD19+, CD20-, CD27+, CD38+, HLA-

DR+), and long-lived PCs (CD19+, CD20-, CD27+, CD38+, HLA-DR-) (Figure 6.2). Cell sorting was performed on a MoFlo cell sorter by a specialist operator. From the naive, MZ and IgM memory cells, PCR was only conducted to amplify IgM transcripts, and for the IgG memory cells, PCR was only conducted to amplify the IgG transcripts. For the PC samples, both IgM and IgG transcripts were amplified in separate reactions. All samples were multiplexed and sequenced on a single run.

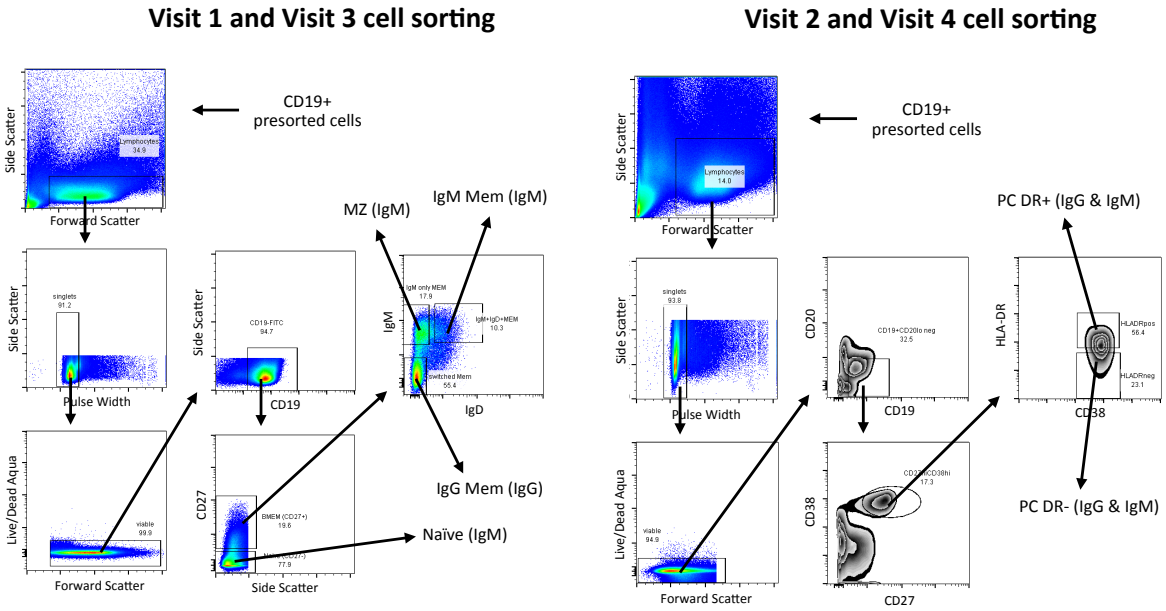

Figure 6.2: FACS gating strategy - On the vaccine administration visits (visit 1 and visit 3), naive, MZ, IgM memory and IgG memory B cell subsets were sorted based on their expression of CD19, CD27, IgM and IgD surface markers. IgM-specific BCR transcripts were PCR amplified from the naive, MZ and IgM memory subsets, and IgG-specific BCR transcripts from the IgG memory subset for sequencing. On the visits seven days post vaccination (visit 2 and visit 4), HLA-DR+ (PC DR+) and HLA-DR- (PC DR-) PCs were sorted based on their expression of CD19, CD20, CD27, CD38 and HLA-DR surface markers. Both IgM and IgG BCR transcripts were independently amplified from both PC subsets for sequencing.

6.2.3 Sequence processing

Sequences were processed using the pipeline described in Figure 2.4. Clustering was performed using the thresholds determined in Chapter 3 - to be included in the same cluster, sequences were required to have the same V and J gene annotation, the same length CDR3, and no more than 1 AA mismatch per 12 AA's in the CDR3. Shared clusters were defined as clusters present

in different individuals samples that shared the same CDR3 AA junction, and V and J gene segment usage.

6.2.4 Statistical analysis

The probability of PC clusters being related to clusters from each of the baseline subsets was calculated separately for each participant and each PC subset. PC clusters that were shared with any baseline subset were found, and the proportion of these clusters that mapped to each baseline subset was then determined (proportion shared PC shared with baseline subset = P_{SHARED}). To account for the different representation of baseline subsets, the proportion of each baseline subset of the total baseline clusters summed across all subsets was calculated (proportion baseline subset of total baseline = P_{BL}). The probability of PCs being related to a specific baseline subset was then determined as P_{SHARED}/P_{BL}. The resulting value was normalized so that the sum of the probabilities of being related to each baseline subset was equal to 1.

6.3 Results

6.3.1 Cell sorting

Nine participants were enrolled into the study; four were assigned to the conjugate group and five to the polysaccharide group (Figure 6.1). FACS was used to isolate baseline CD19+ B cell subsets bearing cell surface markers characteristic of naive (CD27-, IgM+, IgD+), circulating MZ (CD27+, IgM+, IgD+) (216), IgM (only) memory (CD27+, IgM+, IgD-) and switched IgG memory (CD27+, IgM-, IgD-) for five participants at visit 1 (day 0). The sorting failed for the other four participants (two from each vaccine group), so these subsets were sorted at visit 3 (day 28) instead, and treated as baseline. HLA-DR+ PCs (CD20-, CD27+, CD38+, HLA-DR+) and HLA-DR- PCs (CD20-, CD27+, CD38+, HLA-DR-) were successfully sorted at visit 2 (day 7), and HLA-DR+ PCs at visit 4 (day 35; day 7 after second vaccine) from all

participants. The mean number of sorted cells per sample was 75,900 (12,000 - 140,000) for naive cells, 6,753 (1,224 - 19,000) for MZ cells, 8,070 (1,888 - 19,000) for IgM memory cells, 27,145 (2,940 - 100,000) for IgG memory cells, 6,101 (110 - 50,000) for HLA-DR+ PCs and 4,311 (221 - 11,000) for HLA-DR- PCs (Table 9.7). For the visit 2 (day 7) HLA-DR+ and HLA-DR- PC populations, IgG and IgM transcripts were sequenced (except one IgG HLA-DR+ sample that failed). Only the IgG transcripts of the HLA-DR+ population were sequenced at visit 4 (day 35). For the baseline subsets, IgM transcripts were sequenced from all naive, MZ and IgM memory samples, and IgG transcripts sequences from all the IgG memory samples.

6.3.2 Sequence processing

Sequence data were obtained from 80 samples from 9 participants (Table 9.7). The mean number of raw sequences obtained for each sample was 133,190, of which 72% remained on average after initial filtering steps, giving 95,610 reads per sample for analysis. There was over-sequencing (more reads than input cells) for nearly all samples, which led to large numbers of erroneous sequences, especially for the samples with small numbers of cells, so that there were more unique sequences than cells in the sample (Figure 6.3). Following clustering, the number of sequences per sample was more representative of the number of cells for that sample (Figure 6.3), and on average 6% of the original number of raw reads. For some samples, there were still more sequences than cells, but the discrepancy was small, and could be due to inaccuracies in the number outputted by the cell sorter, a small number of erroneous sequences not being incorporated into the same cluster, exogenous DNA contamination, or the formation of chimeric sequences during the PCR reaction (Table 9.7). The mean number of clusters per sample was 41,100 (19,240 - 52,440) for naive cells, 7,873 (3,861 - 16,120) for MZ cells, 6,200 (3,096 - 9,673) for IgM memory cells, 3,652 (1,308 - 7,382) for IgG memory cells, 1,466 (436 - 2,786) for HLA-DR+ PCs, and 1,259 (601 - 2,277) for HLA-DR- PCs.

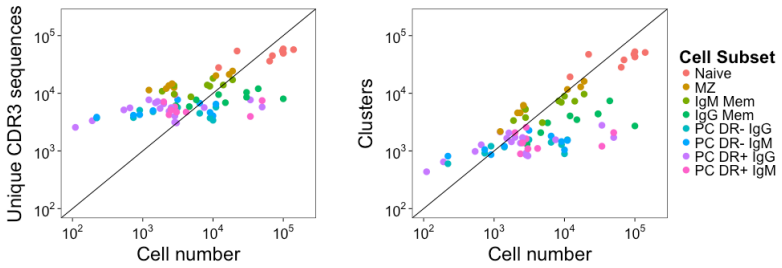

Figure 6.3: Using clustering to correct for over-sequencing. - Correlation between the number of cells for a particular sample, and the number of unique CDR3 AA sequences before clustering (left panel), and unique CDR3 AA cluster center sequences after clustering (right panel).

6.3.3 Post-vaccination PC repertoires have distinct properties compared with the baseline cell subset repertoires

Baseline B cell subsets could be distinguished from each other by a small number of global repertoire properties, which have been shown in previous chapters to be related to B cell activation: diversity (measured using the Shannon entropy index), V gene mutation from germline and CDR3 length (Figure 6.4 A-C). Naive B cells had the greatest diversity, few mutations and longest CDR3. Memory B cells had decreased diversity, more mutations, and shorter CDR3 length compared to naive cells, with IgM memory being less mutated and having longer CDR3 length than IgG memory cells. MZ cells had similar diversity to IgM memory cells, but had fewer mutations, and a shorter CDR3 length than IgM memory cells. PC subsets were the least diverse and most mutated populations. CDR3 length of PCs resembled that of memory cells with the same isotype. Within the PC populations, there were some differences between the IgG and IgM PCs (as with the memory cells, IgM were less mutated, and had shorter CDR3s than IgG), but no differences between the HLA-DR+ and HLA-DR- populations. Principal component analysis based on these variables showed distinct groupings of the different baseline subsets (Figure 6.4 D). Strikingly, PCs had much greater inter-individual variation in these global repertoire properties than other subsets. Whilst the PCs could still be clearly distinguished from the naive, MZ and IgM memory subsets, their global repertoire properties overlapped with the IgG memory cells. Within the PCs, populations could be distinguished

based on isotype, but not HLA-DR expression.

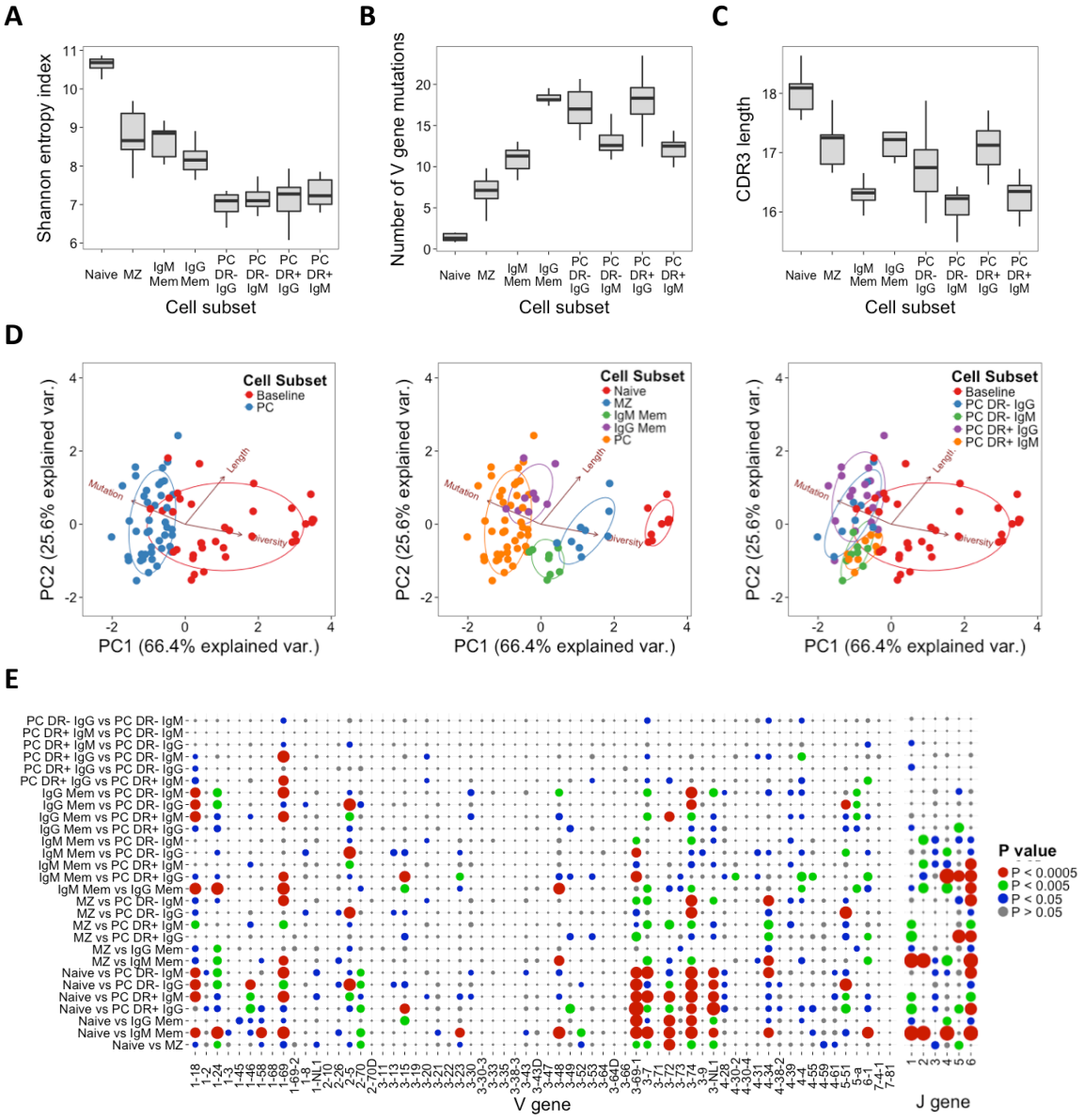

Figure 6.4: Repertoire differences in the different cell subsets. - Differences in (**A**) diversity, (**B**) number of V gene mutations, and (**C**), CDR3 AA length between the different cell subsets. For **A-C**, boxes show locations of 25, 50 and 75th percentiles. Whiskers show data with 1.5x the interquartile range. Samples from all participants and timepoints were included in the analysis. (**D**) Principle component analysis based on the diversity, V gene mutation, and CDR3 length variables. Different plots highlight grouping of different cell subsets. (**E**) Differences in V and J gene segment usage proportions between the different cell subsets. Comparisons performed using the pairwise Mann-Whitney U test. Size and colour of the points indicates the level of significance.

There were differences in V, and J gene segment usage between all subsets analysed, including numerous differences between PC and baseline subsets, and within the different baseline subsets (Figure 6.4 E). When comparing each PC population to other PC and baseline sub-

sets, the PCs were more similar to each other (whether IgG/IgM or HLA-DR+/-) than to any baseline subset.

6.3.4 PC repertoires show similarity between individuals at the sequence level

It has been shown in Chapter 5 and 4 that when administered the same antigen, there is a degree of repertoire convergence between different participants, so it would be expected that the cells responding to the vaccine (HLA-DR+ PCs) should have the greatest convergence. To determine if this was the case, shared clusters were searched for between the different participants within each cell subset. The PCs exhibited more convergence than the baseline cell subsets, with the naive cells having the least convergence (Figure 6.5 A). Convergence was generally greater in the IgM compared to IgG PCs. Comparing the two vaccine groups, there was more convergence between participants in the polysaccharide vaccine group compared to the conjugate vaccine group. After the first vaccine (visit 2), the percent of shared clusters was more than seven times greater in the polysaccharide compared to conjugate group for the HLA-DR+ IgM PCs and more than six times greater for the HLA-DR+ IgG PCs. After the second vaccine (visit 4), when both groups were given the same vaccine, convergence was equivalent for the two vaccine groups.

Comparing the HLA-DR- and HLA-DR+ PCs indicated that the HLA-DR- PCs tended to have more shared clusters than the HLA-DR+ PCs, despite HLA-DR- PCs being theorised to comprise a population of PCs not specific to the vaccine (205). To investigate this further, the database of previously described TT and influenza-specific sequences (Table 9.2) was mapped to the PC dataset to annotate clusters as potentially containing sequences with specificity towards these commonly encountered antigens. There were no clusters annotated as having specificity towards these antigens in the IgG PC dataset, but there were a small number in the IgM dataset (8 influenza-specific, and 7 TT-specific). The percent of these annotated clusters is

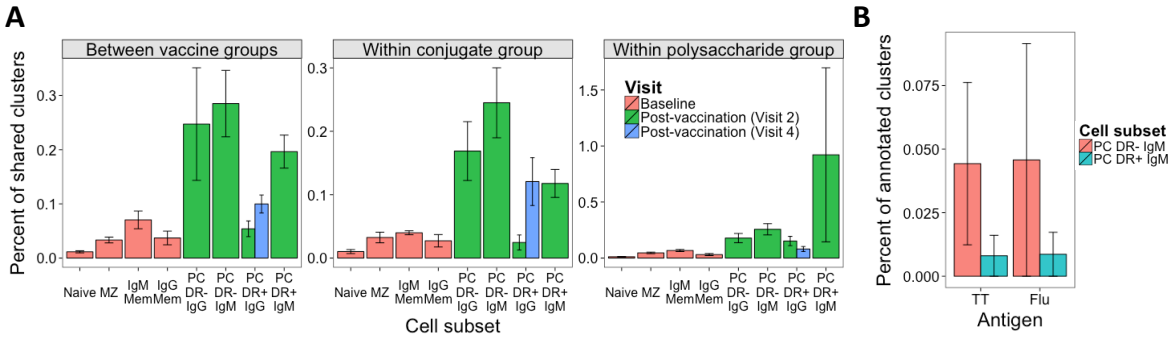

Figure 6.5: The convergent repertoire. - (**A**) For each cell subset, the percent of clusters shared by each pair of two participants was determined, both comparing the participants within each vaccine group (5 participants in the polysaccharide group, giving 10 different pairings, and 4 participants in the conjugate group, giving 6 different pairings), as well as between the two vaccine groups (20 different pairings). Shown are the mean values ±SEM of the percent of clusters shared between each pair. Percent is calculated as (A∩B/sum(A,B))*100. (**B**) Percent of clusters annotated as having similarity to previously described TT and influenza-specific sequences in the HLA-DR+ and HLA-DR- IgM PCs.

greater in the HLA-DR- compared to the HLA-DR+ PCs (Figure 6.5 B), although the numbers are too low to make valid statistical comparisons.

6.3.5 Analysis of PC activation distinguishes polysaccharide from conjugate responses

The PCs showed the greatest similarity between participants at the sequence level, but differed in terms of their general repertoire properties. To determine whether this was caused by the PCs being activated from different baseline B cell subsets, and subsequently converging at the sequence level, the relationship between the PCs and the baseline cell subsets was determined. Within each participant, clusters were searched for that were shared (based on having the same cluster center sequence, and V and J gene segment) between the HLA-DR+ PC subsets, and the baseline subsets. On average, 2.2% (30) of clusters of each PC subset were shared with one of the baseline subsets for each participant. Across all participants IgM PC sequences were predominantly shared with the MZ and IgM memory cells whereas the IgG PC sequences were predominantly shared with the IgG memory cells (Figure 6.6 A and Figure 9.13).

To investigate which baseline subsets were related to the vaccine-stimulated PCs, all PC

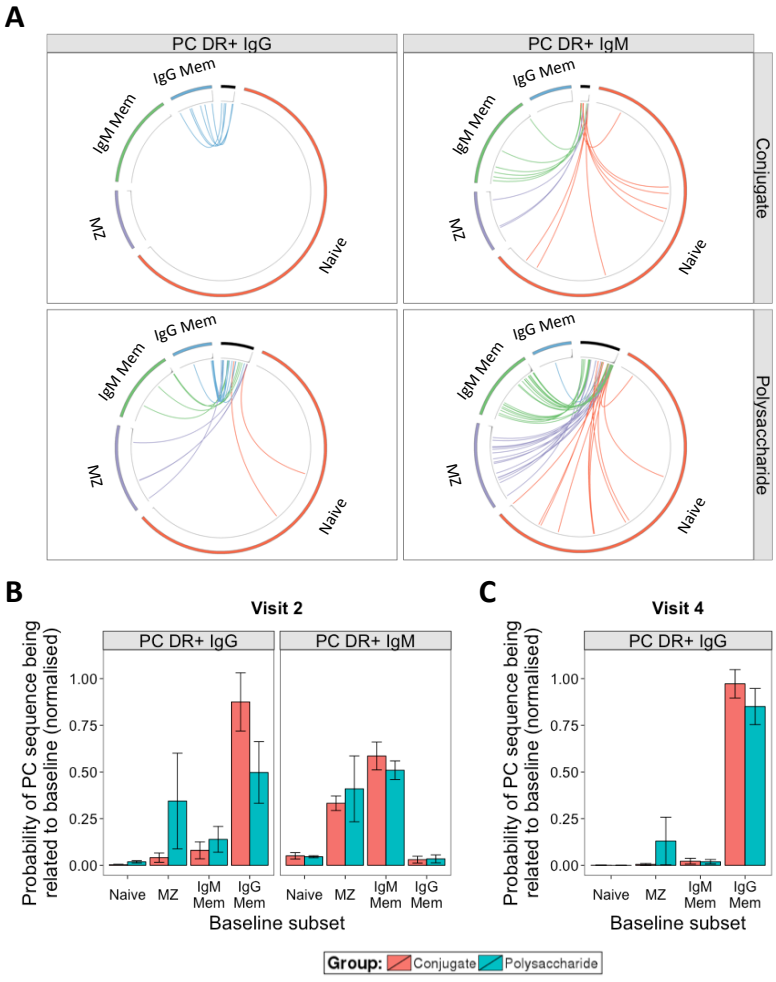

Figure 6.6: Relationship between baseline subsets and PCs - (**A**) Circos diagrams from one representative participant in the conjugate group, and one representative participant in the polysaccharide group showing the relationship between baseline subsets, and the HLA-DR+ IgM and HLA-DR+ IgG PCs after the first vaccination. The length of each section represents the number of clusters in that subset. The black section represents the PCs, and the coloured sections represent the baseline subsets. Clusters are ordered clockwise by size, which is represented by the grey histogram. Coloured lines join clusters that are present in the PCs, and any one of the baseline subsets. Clusters shared within the different baseline subsets are not shown. (**B**) Probability of shared PC clusters at visit 2 being shared with each baseline cell subset for the HLA-DR+ IgG and IgM PCs after the first vaccine (either conjugate or polysaccharide depending on the vaccine group) (**C**) Same as **B**, but with the HLA-DR+ IgG PCs at visit 4, after the second vaccine (conjugate for both vaccine groups). Error bars show ±SEM.

clusters which mapped to an identical cluster in one of the baseline subsets were found, and the probability of these clusters mapping to each baseline subset calculated (Figure 6.6 B-C). There was a low probability of IgG or IgM PC clusters being shared with the naive baseline subset at either visit, in either vaccine group. In the conjugate group, at Visit 2, the HLA-DR+ IgG PC sequences were more likely to be related to the IgG memory sequences than naive, MZ

or IgM memory sequences (Figure 6.6 B). In contrast, in the polysaccharide group, HLA-DR+ IgG PC sequences were equally likely to be related to MZ, IgM memory and IgG memory sequences. Therefore for the HLA-DR+ IgG PCs, the baseline subset that they were most likely to be shared with was related to the vaccine that was given; the potential confounding effect of having baseline subsets from Day 0 vs. Day 28 on this difference between the vaccine groups was investigated, but it was not significant (Figure 9.14). By comparison, for the HLA-DR+ IgM PC sequences, there was a similar probability of being similar to the MZ and IgM memory cells for both vaccine groups. Interestingly after the second dose of vaccine (conjugate for both groups), the probability of HLA-DR+ PCs being shared with each baseline subset was similar for both groups (Figure 6.6 C); the profile resembled that of the conjugate group after the first vaccination.

6.3.6 Vaccine response can further be distinguished based on IgG subclass usage and mutation

Differences in the response between the vaccine groups were then further characterised by comparing IgG subclass usage and V gene mutation from germline for the HLA-DR+ PC sequences generated after vaccination. Analysing the IgG memory cells as a control population showed that IgG subclasses were present in uneven frequencies, with IgG1 (58%) and IgG2 (34%) being the most common, and IgG3 (7%) and IgG4 (1%) being the least common. In the PCs produced after the first vaccination, in the conjugate group, IgG1 remained the most common subclass (58%), followed by IgG2 (33%), IgG4 (7%) and IgG3 (2%) (Figure 6.7 A). In the polysaccharide group on the other hand, the subclass distribution was different; IgG2 was the dominant subclass (58%), followed by IgG1 (37%), IgG3 (5%) and IgG4 (<1%). After the second vaccine, when both groups were given conjugate vaccine, the subclass usage of the PCs resembled that of the conjugate group in both vaccine groups.

Considering V gene mutation from germline, IgG PCs were more mutated than IgM PCs

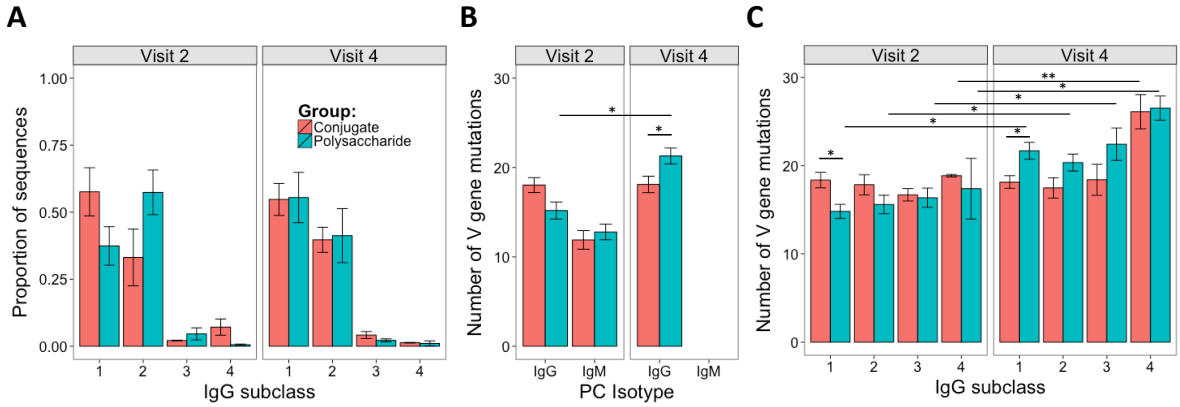

Figure 6.7: IgG subclass usage and mutation in the PCs produced after vaccination - (**A**) Proportion of sequences of each IgG subclass after the first and second vaccines for the HLA-DR+ PCs. (**B**) Mean number of mutated nucleotides in the V genes after the first and second vaccines for the IgG and IgM HLA-DR+ PCs. (**C**) Same as **B**, but split by IgG subclass. Comparisons were performed using a t-test. Error bars show ±SEM. *P<0.05, **P<0.005.

($p = 0.0024$). After the first vaccination, there was no significant difference in V gene mutation between the vaccine groups for either the total IgG or the IgM PCs (Figure 6.7 B). After the second vaccination (when both groups were given the conjugate vaccine), V gene mutation in the IgG PCs was greater in the group primed with the polysaccharide rather than conjugate vaccine. There was a significant increase in IgG PC V gene mutation in the polysaccharide group after the second vaccination (conjugate) compared to the first vaccine (polysaccharide). In the conjugate group, there was no difference in IgG PC V gene mutation after the second dose compared to the first dose of the two conjugate vaccines.

Calculating average V gene mutation separately for each IgG subclass revealed further differences between the vaccine groups. In the IgG memory cells, there are slight differences in the number of V gene mutations in the different subclasses going from IgG3 (17.4) to IgG1 (17.9), IgG2 (19.2) and IgG4 (20.5). Comparing the vaccine groups showed the only significant difference to be in IgG1 PC sequences. IgG1 sequences were significantly more mutated in the conjugate group after the first vaccine (18.4 in the conjugate group vs. 16.2 in the polysaccharide group), but significantly more mutated in the polysaccharide group after the second vaccine (18.1 in the conjugate group vs. 21.7 in the polysaccharide group) (Figure

6.7 C). Mutation for all four subclasses increased between the first vaccination and the second vaccination in the polysaccharide group. In the conjugate group, there was only a significant increase in mutation in the IgG4 sequences after the second vaccination.

6.4 Discussion

After vaccination with meningococcal conjugate or polysaccharide vaccines, antigen-specific PCs are detected in peripheral blood at high levels 7 days post-vaccination (91). Using BCR repertoire analysis of the response to two different meningococcal serogroup ACWY vaccines, this study suggests that insights into B cell activation can be obtained by relating BCR sequence data from post-vaccination PCs to a variety of baseline B cell subsets. This study has shown that a limited number of global BCR repertoire properties (mean CDR3 length, mutation and diversity) can be used to define different baseline subsets, and that these properties are well conserved between individuals. In contrast, antigen stimulated PCs display more inter-individual variation in these global repertoire properties. Despite this diversity in global repertoire properties, the PCs have a greater number of convergent clusters, when compared to the baseline subsets. The finding of PCs expressing the greatest difference between individuals in their global repertoire properties is consistent with them potentially being activated from a variety of different baseline B cell subsets, which will differ for each individual. The greater proportion of shared clusters in the PCs suggests that there is then some degree of convergence at the sequence level to the vaccine antigens stimulating their production. Furthermore these data suggest that immunisation with protein-polysaccharide conjugate vaccines stimulates the development of PCs from IgG memory cells, but plain polysaccharide vaccination results in significant numbers of B cells being derived from IgM memory and MZ B cells as well. Further dissection of the PC repertoire gives insight into differential IgG subclass usage, and V gene mutation induced by the two vaccines.

Differences have previously been observed in the BCR repertoire between naive, IgM memory and IgG memory cell subsets (122, 141) that are conserved across multiple individuals. This study extends these observations by including, MZ, and a variety of PC subsets in the same analysis. By measuring diversity, CDR3 length, and V gene mutation from germline (Figure 6.4), properties shown in Chapters 3 and 4 to be indicative of repertoire activation state, it is shown here that B cell subsets appear to have a characteristic sequential pattern of global repertoire properties in relation to their activation state (naive → MZ → IgM Memory → IgG Memory → PCs). The more activated (or mature) subsets have decreased diversity, shorter CDR3s, and more V gene mutations. In addition, repertoire convergence between individuals was greater in the more mature B cell subsets, in particular in the PCs, giving further evidence that when exposed to a common antigen, there is a degree of repertoire convergence between individuals, with convergent sequences likely to be antigen-specific. The presence of the circulating MZ cells being a distinct grouping from memory cells, yet located between the naive and IgM cells is interesting in light of recent work detailing the developmental pathway of MZ cells (1). It may be that MZ cells are a distinct cell subset arising from GC independent pathways, or alternatively that they are just in a transitional state between naive and IgM memory cells (37).

In the context of analysing vaccine response, only HLA-DR+ PCs were considered, as these are thought to be enriched for recently activated, and therefore vaccine-specific PCs (205, 217). Nevertheless, both HLA-DR+ and HLA-DR- PCs were sorted and sequenced to give further insight into differences between these populations, and further confirm this. Previous work has demonstrated that PCs show a gradient of HLA-DR expression from high levels in lymphoid tissue (tonsils) through to lower levels in blood and lowest expression in bone marrow (217). This is thought to correlate with maturation towards more long-lived PCs in the bone marrow. Studies of TT-specific PCs 6 days following TT immunisation suggested that the

vaccine antigen-specific population were HLA-DR+ and constituted a significant proportion of total HLA-DR+ PCs at that time point (205). HLA-DR- PCs at day 6 were not vaccine antigen-specific and were thought to have been displaced from their residence in the bone marrow by the vaccine generated PCs. Based on this evidence, it may be expected that HLA-DR- PCs would fall further along the continuum of activation than the HLA-DR+ PCs, but this was not the case, and the two populations were indistinguishable. In addition, there were limited differences observed in V and J gene usage between the HLA-DR+ and HLA-DR- populations, that were able to distinguish them from each other. Recently activated PCs have a limited stimulus (vaccine antigens), whereas previously generated PCs could potentially have been activated by a lifetime of different antigenic stimuli, so one may expect the HLA-DR+ PCs to show the greatest similarity between individuals. Intriguingly, in this study, there were generally more convergent clusters for the HLA-DR- than the HLA-DR+ population. One possible reason for this convergence in the HLA-DR- population is that there are certain common antigens that most individuals are exposed to during the course of their lifetime, and PCs specific for these common antigens dominate the HLA-DR- compartment. Despite having limited data on previously described antigen-specific sequences, a greater proportion of the HLA-DR- repertoire was comprised of TT and influenza-specific clusters than the HLA-DR+ repertoire, giving evidence for this hypothesis.

The dichotomy between PCs being the most similar population between individuals at the sequence level, but most divergent in terms of their general repertoire properties is consistent with the idea that a variety of precursor cells belonging to different baseline B cell subsets are activated to form the PCs of different individuals, but the process of affinity maturation then leads them to converge at the sequence level. The acquisition of sequence data from both baseline B cell subsets, as well as vaccine-activated PCs, allowed determination of which baseline B cell subsets were related to the PCs, by measuring the probability of clusters being

shared between them. Such cluster sharing can be used to indicate which baseline B cell subsets are stimulated to form PCs by the vaccine. Whilst the current study does not confirm the antigenic specificity of these PCs, a previous study using an identical vaccine has demonstrated that vaccine-specific PCs are detectable at high levels 7 days following either the first dose or second dose of the ACWY vaccine. Therefore, the HLA-DR+ PC populations sorted here are likely to be enriched for vaccine-specific PCs (214). Pre-existing immunity to *N. meningitidis* is common in adults, so the vaccinated individuals in this study are unlikely to mount naive responses (218). The observation that the lowest probability of sequence sharing was between the naive cells and the PCs in this study therefore supports our approach for determining B cell activation by showing that such cluster sharing is unlikely to occur by chance alone.

IgM PCs appeared to be most similar to the MZ and IgM memory cells, whereas the IgG PCs appeared to be most similar to the IgG memory cells, suggesting that PCs are mainly activated from memory cells as part of a recall response. Comparing the vaccine groups, the IgG PCs after polysaccharide vaccine are more likely to be related to MZ and IgM memory cells, and less likely to be related to IgG memory cells than after conjugate vaccine. This supports the hypothesis that the polysaccharide vaccine stimulates a more TI-like response, while the conjugate vaccine stimulates a more TD-like response. However, there was still a high probability of IgG PCs being related to IgG memory cells even after polysaccharide vaccination, indicating that categorising responses in this way is not clear-cut, and an antigen may activate both TD and TI pathways. Furthermore, as the response is expected to be dominated by memory recall, the effect of any differential activation will be reduced. While B cell memory recall has previously been assessed by testing antibody affinity (166), this method does not indicate which B cell subsets are activated in the recall response. The technique presented here for monitoring memory recall at the sequence level has clear application in vaccine development, where long-term protection requires such memory formation.

Knowing the exact sequences of the recalled cells not only allows relationships to precursor cell subsets to be explored, but also allows identification of IgG subclass and mutation levels, giving further insight into the vaccine response. Since the four IgG subclasses differ in their ability to activate complement, and bactericidal activity in different antigenic contexts (196), monitoring IgG subclass usage is useful. Normally, IgG2 is produced in response to polysaccharide antigens, IgG1 and IgG3 produced in response to protein antigens (219), while IgG4 serves a more regulatory function (220). The data here show that after polysaccharide vaccination, IgG2 was the predominant subclass produced, but after conjugate vaccination, IgG1 was the predominant subclass produced (even in individuals primed with the polysaccharide vaccine). As mutational levels are correlated with cellular proliferation in the GC (194), measuring changes in BCR sequence mutation may be a proxy for the magnitude of a high affinity PC response following vaccination. PC levels of IgG1 mutation were greater in the conjugate compared to the polysaccharide group after the first vaccine, which is consistent with increased GC activity and switching to IgG1 in the response to conjugate vaccination. The increase in frequency of mutations in all IgG subclasses between the first (polysaccharide) and second (conjugate) vaccines in the polysaccharide group further suggests that conjugate vaccination may initiate a more marked GC response than polysaccharide vaccination. However, an alternate explanation is simply that the conjugate vaccine stimulates B cells specific for the protein component of the vaccine, and these are more mutated than the B cells specific for the polysaccharide. In the conjugate group, only IgG4 sequences increased in mutation after the second vaccine, indicating that after repeat vaccination with such a highly stimulating TD antigen, IgG4 plays an active role in the response. The functional properties of IgG4 are poorly understood, but it is known to increase after prolonged antigen exposure, and may have a role in regulating the immune response (220), which would be interesting to explore further.

6.5 Conclusion

This study shows that BCR repertoire sequencing opens a novel window onto vaccine responses, not possible through conventional immunological measures. When combined with samples taken at different time points and from different B cell subsets, individual B cell clones can be tracked to determine activation and memory recall. Accumulating such data in vaccine development programmes can be used to show which B cell subsets (or even specific B cell sequences known to produce protective antibodies) are activated by the vaccine, so that the vaccine can then be altered to drive generation of optimal responses.

7

Using BCR sequencing to investigate the effect of AS03 adjuvant on the PC response to pH1N1 influenza vaccination

7.1 Introduction

Influenza virus causes seasonal outbreaks of clinical influenza, and has been responsible for four pandemics over the last 100 years (221). While seasonal outbreaks are associated with mutation of the HA protein on the viral surface to escape neutralization by antibodies generated in previous exposures, pandemics result from the introduction of completely new viruses into populations where there is little pre-existing immunity to that virus (222). The latest influenza pandemic arose in 2009, and was caused by a swine-origin A/California/7/2009 (H1N1) virus (pH1N1), and resulted in an estimated 300,000 deaths within the first 12 months (223). The pre-pandemic 2008/2009 seasonal trivalent influenza vaccines (TIV) did contain an H1N1 strain (A/Brisbane/59/2007), but this differed considerably at the structural level from the pandemic strain, with 24 AA differences at key antigenic sites (224), and thus offered only limited heterotypic protection (225, 226).

The capacity to rapidly develop and manufacture effective vaccines in large quantities is key

in combating influenza pandemics. Adjuvants can enhance vaccine immunogenicity, allowing a reduction in the quantity of antigen per dose and a consequent increase in the number of doses that can be manufactured in a given time-period. Many pH1N1 vaccines were therefore formulated with an oil-in-water adjuvant (AS03 or MF59), and these conferred greater immunogenicity than non-adjuvanted vaccines, even when using just a quarter of the antigen dose (227, 228). Despite the success of these adjuvants, the details of their mode of action in the context of influenza vaccine are still poorly understood.

AS03 and MF59 enhance innate immune responses by increasing antigen uptake and presentation in the local tissue. This in turn leads to increased CD4 T cell, and B cell responses (229, 230). For pandemic influenza vaccination, this suggests that the adjuvant could improve B cell responses by either increasing activation of naive B cells, or by increasing the activation and adaptation of pre-existing memory B cells generated through infection or immunization with seasonal influenza from earlier years to become specific towards the pandemic strain (231). In a previous study, GlaxoSmithKline investigated the effect of AS03 on the pH1N1 vaccine response, and also the effect of TIV priming on the subsequent pH1N1 response (228). This study indicated that prior TIV administration decreased both the humoral and T cell response to pH1N1 vaccine, but adjuvanting the pH1N1 vaccine helped to overcome this effect (228). Such a finding is potentially consistent with the adjuvant working by either stimulating more naive B cell activation, or by increasing adaptation of pre-existing memory B cells, but gives no mechanistic insight.

Understanding the mode of action of the adjuvant can be helped by studying the properties of the PCs produced in response to the vaccine. Khurana *et al.* used phage display libraries, and surface plasmon resonance to determine binding locations, and affinity of the antibodies produced in response to both adjuvanted and non-adjuvanted pandemic influenza vaccines (165, 232). They found that the antibodies produced in response to the adjuvanted vaccine

displayed a greater diversity of binding targets, had a shift away from targeting the conserved stem region of HA towards the more variable head region (Figure 7.1), and had a greater avidity than those produced in response to the non-adjuvanted vaccine (165, 232). These results suggested that the adjuvant mainly functioned by stimulating more of a naive vaccine response by activating B cells targeting different epitopes, and not through more extensive diversification of pre-existing memory cells.

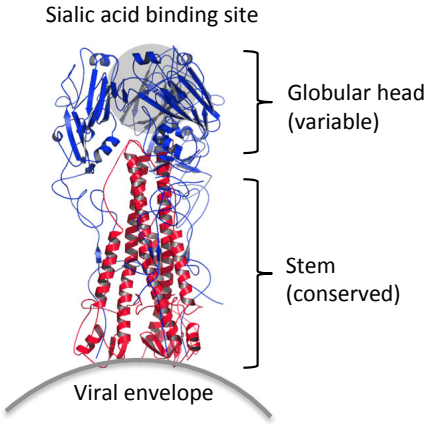

Figure 7.1: Structure of influenza HA - The HA antigen binds to host cell membranes containing sialic acid, and subsequently facilitates membrane fusion, and viral entry into the host cells. The structure of HA reflects these two functions, with a head region that contains the sialic acid binding sites, and a stem region containing the membrane fusion machinery. The head region is more variable between influenza strains as it is the main target of neutralizing antibodies, but effective neutralizing antibodies may also target the more conserved and difficult to access stem region.

As has been shown in the previous chapters, an increased understanding of the repertoire of PCs produced in response to vaccination could potentially be gained by sequencing their BCR repertoire. Knowing the exact nucleotide sequences allows determination of mutation numbers, which can be used to distinguish between the PCs activated from naive B cells versus pre-existing memory B cells. Other features of the repertoire, such as diversity, and isotype subclass usage can also be used to provide further insight into the immune mechanisms that generated the PCs. If sequential samples are taken following repeat vaccinations from a single participant, it is also possible to directly identify and characterize memory recall by looking for shared sequences between the two samples.

Here, samples from the previous GlaxoSmithKline study, which investigated the effect of TIV priming, and AS03 adjuvant on the pH1N1 response were used for high-throughput PC BCR sequencing (228). PC samples were sequenced 7 days following administration of the pH1N1 vaccine either with or without AS03 adjuvant (Figure 7.2). Day 7 was chosen as this represents the peak of the antigen-specific PC response to influenza vaccination, with previous studies demonstrating 33-80% of PCs isolated at this time to be specific to the vaccine (104, 233). In some participants, TIV was given four months prior to the pH1N1 vaccination, and in these participants, the PC repertoire was also sequenced 7 days following administration of this vaccine. Obtaining paired samples following both TIV and pH1N1 vaccination allowed direct investigation of memory recall between these two vaccines. The findings support the notion that the adjuvant functions through a combination of both increasing naive B cell activation, and also by increasing the adaptation of pre-existing memory B cells for greater specificity towards the pandemic strain.

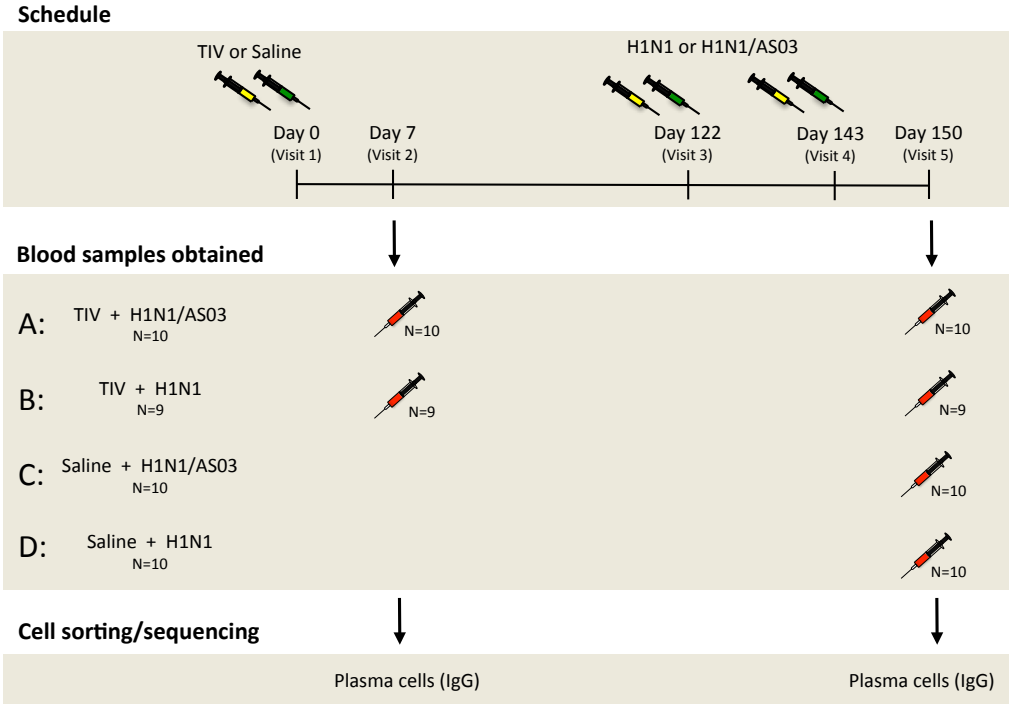

Figure 7.2: Study Design - Nineteen participants were given TIV at day 0, and 20 participants were given a saline placebo. On day 122 and 143 participants were given a pH1N1 vaccine either with or without adjuvant. FACS was used to isolated PCs for BCR repertoire analysis 7 days following the TIV vaccine, and 7 days following the second pH1N1 vaccine.

7.2 Methods

7.2.1 Study design

Of the 118 participants included in the original GlaxoSmithKline study (228), samples from a random subset of 39 participants were used for this study. Participants were recruited with informed consent across three study centers in the United States, under approval from local ethics committees, and in accordance with the Declaration of Helsinki. All participants were healthy, aged between 19-40 years, and had no prior immunisation history with the 2009 pH1N1 vaccine. Participants were randomized 1:1:1:1 to four study groups: A, B, C and D (Figure 7.2). For this study, there were samples from 10 participants from groups A, C and D, and 9 participants from group B. Group A and B were administered TIV, while groups C and D were administered saline at day 0. Peripheral blood samples were taken for PC sorting 7 days following TIV vaccination in groups A and B. Four months later, two doses of pH1N1 vaccine were given to all participants at a three week interval. Groups A and C received AS03-adjuvanted vaccine, while groups B and D received non-adjuvanted vaccine. Peripheral blood samples were taken for PC sorting and immunogenicity evaluations 7 days following the second pH1N1 vaccine.

7.2.2 Vaccines

GlaxoSmithKline, Quebec, Canada, manufactured the vaccines. The TIV was the 2009-2010 vaccine, containing 15 μg HA each of A/Brisbane/59/2007 (H1N1) IVR-148, A/ Uruguay/716/2007 (H3N2) NYMCX-175C and B/Brisbane/60/2008 (B). The pH1N1 vaccine was the 2009 pandemic vaccine containing HA of A/California/7/2009 (H1N1). The non-adjuvanted vaccine contained 15 μg HA. The adjuvanted vaccine contained 3.75 μg HA with the $AS03_A$ oil-in-water emulsion adjuvant.

7.2.3 Immunogenicity evaluations

The immunogenicity evaluations were carried out as part of the original study, and methods only briefly outlined here (228). Immunogenicity evaluations were obtained from the day 150 visit only (7 days following administration of the second pH1N1 vaccine). Haemagglutination inhibition (HAI) titers were measured as the highest serial dilution of serum that prevented hemagglutination. Seropositivity was defined as a titer $\geq$ 1:10.

Memory B cell ELISpot was used to determine frequencies of H1N1(California)-specific memory B cells. Cells were induced to differentiate into PCs by incubating in culture medium containing CpG DNA for 5 days prior to addition to an antigen-coated ELISpot plate. Memory B cell ELISpot was only conducted for a subset of the participants in the original study, which equated to 17 of the 39 participants included in this study.

Activated H1N1(California)-specific CD4 T cell numbers were assessed using intracellular cytokine staining, followed by flow cytometry. To be classed as activated, the T cells were required to express at least two of the following immune markers: interferon-gamma, interleukin-2, tumor necrosis factor-alpha and CD40 ligand.

7.2.4 Sample processing

Cells were defrosted, and stained with CD3-pacific blue (UCHT1), CD19-FiTC (HIB19), CD20-PECya7 (2H7), CD27-APC (M-T271) and CD38-PerCPCya5.5 (HIT2) (all BD Biosciences). CD3-, CD19+, CD20-, CD27+ and CD38+ PCs (Figure 7.2) were sorted using a FACSAria cell sorter. PCR was conducted to amplify IgG transcripts only, and all samples were multiplexed and sequenced on a single run.

7.2.5 Sequence processing

Sequences were processed using the pipeline described in Figure 2.4. Clustering was performed using the thresholds determined in Chapter 3 - to be included in the same cluster, sequences

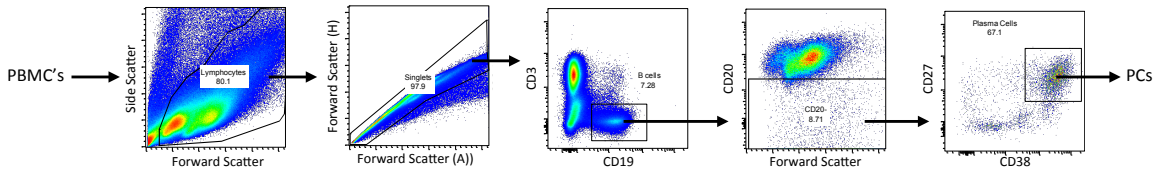

Figure 7.3: FACS gating strategy - PCs were isolated based on being CD3-, CD19+, CD20-CD27+ and CD38+ on day 7 following TIV vaccination (groups A and B only), and on day 7 following the second pH1N1 vaccine. IgG BCR transcripts were subsequently amplified and sequenced.

were required to have the same V and J gene annotation, the same length CDR3, and no more than 1 AA mismatch per 12 AA's in the CDR3. To find clusters shared within an individual after both their TIV and pH1N1 vaccines, sequences from the two samples were clustered together. If two samples contribute at least one sequence to the same cluster, that cluster is defined as being present in both samples.

Lineage trees were generated for clusters with at least 50 sequences, and subsampling performed if there were more than this number.

7.3 Results

7.3.1 Adjuvant improves the humoral and T cell response to pH1N1 vaccination

Of the 118 participants in the original study (228), 39 were used for PC repertoire sequencing in this study (Figure 7.2). In this study, the focus was on the response 7 days following the pH1N1 vaccine - more comprehensive immunogenicity analyses can be found in the original study (228). The humoral and T cell responses in the subset of 39 participants reflected those from the whole cohort. All participants were seropositive for 2009 H1N1(California) by 7 days following their first dose of pH1N1 vaccine. The HAI titer, the number of activated 2009 H1N1(California)-specific CD4 T cells, and the number of 2009 H1N1(California)-specific memory B cells following pH1N1 vaccination was greater in the group receiving the adjuvanted vaccine compared to non-adjuvanted vaccine (Figure 7.4 A-C and (228)). PC frequency, and

heterotypic seroconversion to the 2007 H1N1(Brisbane) viral strain in the TIV vaccine were similar between the adjuvanted and non-adjuvanted vaccine groups (Figure 7.4 D/E and (228)).

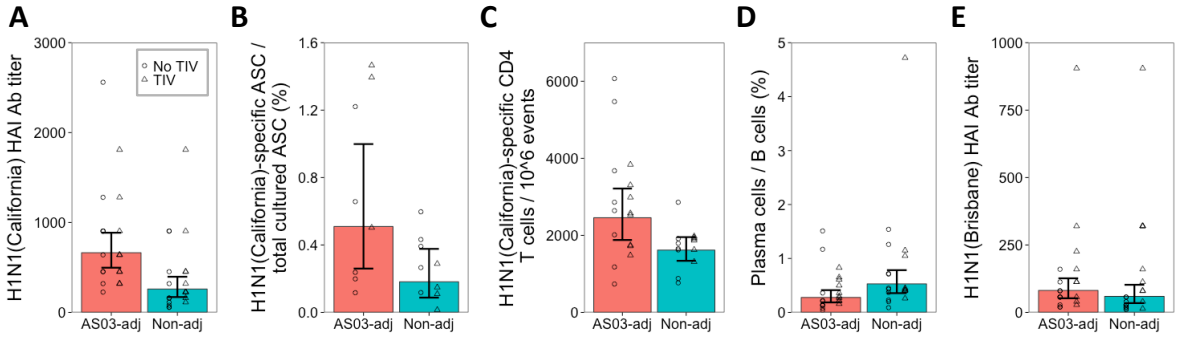

Figure 7.4: Serological and cellular measures of pH1N1 vaccine response taken 7 days following vaccination - (**A**) H1N1(California)-specific HAI antibody titer, (**B**) H1N1(California)-specific memory B cells determined by ELISpot, (**C**) H1N1(California)-specific activated CD4 T cells determined by intracellular cytokine staining and flow cytometry, (**D**) PCs determined by flow cytometry, and (**E**) H1N1(Brisbane)-specific HAI antibody titer. Measurements were carried out for all 39 participants, except for ELISpot, which was only conducted for a subset of 17 participants. Participants were grouped according to whether they received an AS03-adjuvanted or non-adjuvanted vaccine. Individual data points are shown, with the shape representative of whether the participant received the seasonal TIV prior to pH1N1 vaccination. Shown are the geometric mean values and 95% confidence interval.

Nineteen (10 in the adjuvanted group, and 9 in the non-adjuvanted group) of the participants had received the 2009 seasonal TIV vaccine four months previously. While the data in the original study indicated that prior TIV vaccination reduced the subsequent response to pandemic vaccine, this potentially confounding variable was not considered for the analyses presented here (228). The effect of TIV was small in relation to the effect of the adjuvant, and the reduced group sizes here make it unlikely that any differences would be detectable. Furthermore, the number of participants receiving TIV vs. not receiving TIV was similar in the non-adjuvanted and adjuvanted groups.

7.3.2 Sequencing the PC repertoire following pH1N1 and TIV vaccination

To further investigate the B cell responses, PCs were isolated at day 7 following both pH1N1 and TIV (where applicable) vaccination for BCR sequencing (Figure 7.2 & 7.3). On average, 3,433 (range: 464 - 10,651) PCs were isolated from each sample (Table 9.8). PC numbers were

similar in the adjuvanted and non-adjuvanted groups, but greater after TIV vaccination (mean numbers of 3,003, 2,754, and 4,577 respectively). A total of 19,101,557 raw sequencing reads were obtained, which reduced to 9,505,320 following sequence quality control and filtering steps. Following clustering, there were on average, 2,353 (range: 627 - 4,528) clusters generated for each sample. The number of clusters was closely related to the number of input cells, giving a 1:1.2 ratio of clusters to cells (Figure 9.15 A). As described previously, PCA of V and J gene usage (Figure 9.15 B) as well as repertoire diversity, V gene mutation, and average CDR3 sequence length (Figure 9.15 C) was conducted to detect outliers. No outliers were found in the current dataset using either of these measures, so all samples were retained for analysis.

7.3.3 PCs isolated following vaccination are enriched for vaccine specificity

While not all PCs produced 7 days following vaccination will be specific to the vaccinating antigen, previous studies with influenza vaccine have shown vaccine-specific PCs to be highly enriched at this time (104, 233). To verify that the PCs isolated in this study were enriched for vaccine-specificity, the database of previously described pH1N1-specific, and 2007-2009 seasonal influenza-specific BCR sequences collected in Chapter 3 were compared to the current dataset. In total, there were 182 pH1N1-specific, and 126 seasonal influenza-specific sequences in the database. These sequences were used to annotate 74 clusters in the current dataset as having seasonal influenza specificity, and 206 clusters as having pH1N1 specificity (based on sharing the same CDR3 AA region as sequences within the cluster). Clusters annotated as specific to seasonal influenza were most common after TIV vaccination, although some were still found after pH1N1 vaccination (Figure 7.5 A/B). The converse was true for clusters annotated as specific to pH1N1. The pH1N1 and seasonal influenza-specific sequences were also compared to the previously collected BCR sequence datasets from PCs collected following both Hepatitis B (Chapter 4 & 5) and Meningococcal ACWY (Chapter 6) vaccination, but no matches were found (Figure 7.5 A/B).

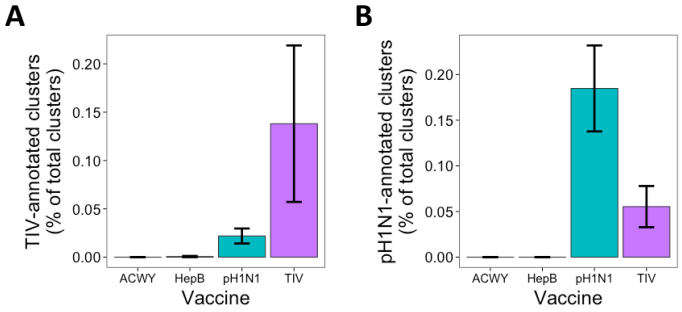

Figure 7.5: Specificity of PC sequence clusters following vaccination with different antigens - (**A**) Percent of total clusters annotated as having specificity to 2007-2009 seasonal influenza based on comparison to previous data. (**B**) Percent of total clusters annotated as having specificity to pH1N1 based on comparison to previous data. Bars show mean values ±SEM. N = 9 for Meningococcal ACWY vaccination, 14 for Hepatitis B vaccination, 39 for pH1N1 vaccination, and 19 for TIV vaccination.

7.3.4 Using mutation number to distinguish clusters derived from naive B cell activation vs. memory recall

The finding that some seasonal influenza-specific clusters are found following pH1N1 vaccination suggests that some B cells activated following TIV vaccination are subsequently recalled following pH1N1 vaccination. The number of V gene mutations of the seasonal influenza-specific and pH1N1-specific clusters was analyzed following pH1N1 vaccination, to try and distinguish between clusters closely related to germline (and are thus more likely to have been recently activated from naive B cells), compared to clusters which have diverged more from germline (and are thus more likely to have arising from memory recall). All of the seasonal influenza-specific clusters were highly mutated (minimum of 9 V gene mutations), whereas the pH1N1-specific clusters had a bimodal distribution with some highly mutated, and some with very little mutation (Figure 7.6 A). As the seasonal-influenza specific clusters must derive from memory recall, but the pH1N1 clusters can derive either from memory recall or naive activation, this information was used to place a mutation cutoff of 7 mutations (corresponding to approximately 2.5% of mutated bases) to distinguish unmutated clusters more likely to have arisen from naive B cells compared to mutated clusters more likely to have arisen from memory recalled B cells (although we cannot discount that some unmutated clusters may also

be derived from memory recall). While 2.5% mutation is more than is expected in naive IgM B cells (see Chapter 6), we are here interested in class-switched PCs recently derived from these naive B cells, so more mutation would be expected (72). A threshold of 2.5% mutation also ensures that sequencing errors are not overinterpereted as somatic hypermutation in the mutated clusters.

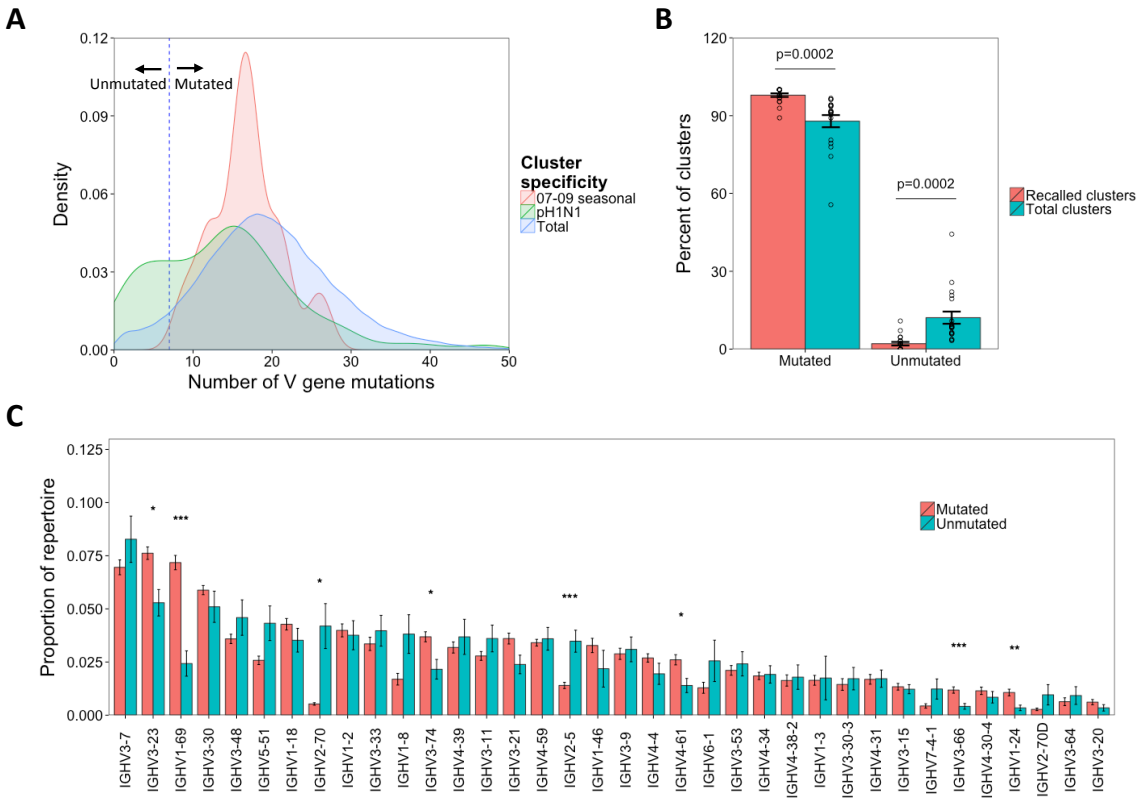

Figure 7.6: Distinguish clusters derived from naive B cell activation from clusters derived from memory recall following pH1N1 vaccination - (**A**) Density histogram showing the number of mutations of either total clusters following pH1N1 vaccination, or clusters annotated as having specificity to either pH1N1, or seasonal influenza based on comparison to previous data. Dotted vertical line separates clusters with less than 7 mean mutations (unmutated) from those with at least 8 mean mutations (mutated). (**B**) Clusters present following both TIV and pH1N1 vaccination were identified (recalled clusters), and the percent of these classed as either mutated or unmutated (based on the 7 mutation cutoff) was determined following pH1N1 vaccination. Percent of total clusters that were mutated or unmutated was then determined for comparison. P value shows the result from a paired two-sided t-test. (**C**) For each sample, the proportion of both mutated and unmutated clusters utilizing different V gene segments was determined. Bars show mean values ±SEM. Shown are the V genes with a proportion above 0.005 in at least one of the groups. Comparisons were performed using a two-sided paired t-test. * $p < 0.01$, ** $p < 0.001$, *** $p < 0.0001$.

As an independent measure to validate the 2.5% mutation threshold, specific recall between

the TIV to pH1N1 vaccines was determined by combined analysis of these two datasets. Co-clustering the data from the TIV and pH1N1 datasets for each participant revealed that on average 52 clusters were shared between their TIV and pH1N1 PC datasets, equating to 1% of the total number of clusters present in the two datasets. As the probability of the same CDR3 being produced during two independent recombination events during the lifetime of a single individual is practically zero, it is likely that these shared clusters represent memory recall of the same B cell lineage (25, 151). On average, 98% of these recalled clusters were classed as mutated following pH1N1 vaccination, compared to 88% of total clusters ($p = 0.0002$; Figure 7.6 B), which is what we would expect if mutated clusters represent those derived from memory recall.

7.3.5 Unmutated and mutated clusters have distinct V gene usage profiles

To further investigate the properties of the mutated and unmutated clusters following pH1N1 vaccination, for each sample, the proportion of the repertoire of these clusters comprised by different V genes was determined. V gene usage differed dramatically between the mutated and unmutated clusters, regardless of whether the dataset was split based on receipt of adjuvanted or non-adjuvanted vaccine, or previous receipt of TIV (Figure 7.6 C & 9.16). Usage of IGHV1-69, IGHV3-23, IGHV3-74, IGHV4-61, IGHV3-66 and IGHV1-24 was greater in the mutated clusters, while usage of IGHV2-70 and IGHV2-5 was greater in the unmutated clusters. No V gene usage differences between mutated and unmutated clusters were observed when carrying out the same analysis on a control dataset of PCs collected prior to immune stimulus as part of Chapters 4 and 5 (Figure 9.17).

While V gene usage alone cannot be used to determine what epitope a sequence will bind to, it is well documented that some V genes are preferentially used in the creation of BCRs with certain specificities, indicating that the mutated and unmutated clusters may target distinct HA epitopes. In the context of influenza vaccine, IGHV1-69 is associated with the production

of HA stem-binding (rather than head-binding) cross-reactive antibodies (234, 235). Usage of IGHV1-69 showed the most significant difference between the mutated and unmutated clusters; IGHV1-69 was three times more abundant in the mutated compared to unmutated clusters.

7.3.6 Adjuvant increases the proportion of unmutated clusters specific to the vaccine

Following on from the observation that the unmutated and mutated clusters appear to represent distinct populations which may represent PCs more recently activated from naive B cells versus those derived from memory recall, the effect of the adjuvant on the ratio of unmutated to mutated clusters was determined. The mean number of mutations was slightly lower in the participants receiving the adjuvanted compared to non-adjuvanted pH1N1 vaccine (18.5 vs. 19.8 respectively), and this was caused by a relative increase in the number of unmutated compared to mutated clusters in the adjuvant group (Figure 7.7 A). On average, in the adjuvant group, 10.6% of clusters were unmutated, compared to 5.8% in the non-adjuvanted group. Following TIV vaccine, where it is expected that the response is more dominated by memory recall than the response to pH1N1 vaccine, only 3.5% of clusters were unmutated (Figure 9.18).

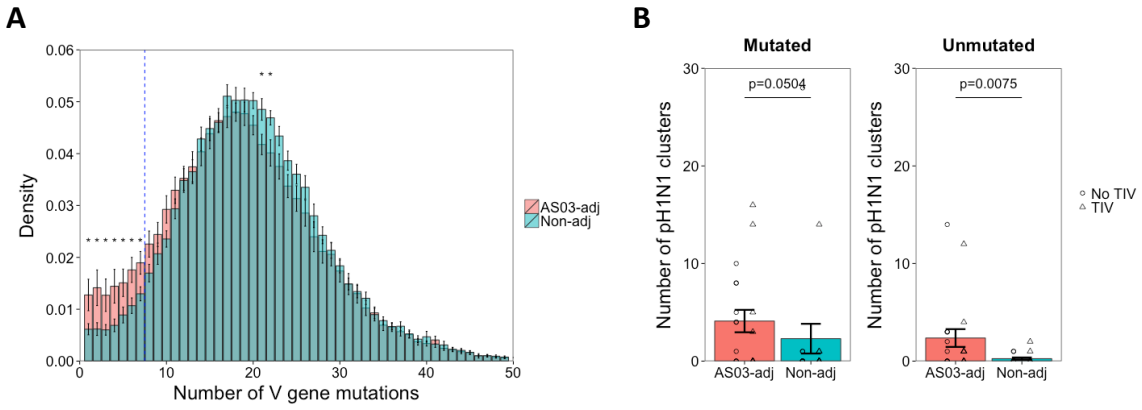

Figure 7.7: Effect of adjuvant on stimulation of naive vs. recalled cells - (**A**) Distribution of clusters with different mean numbers of V gene mutations, split by vaccine group. Dotted vertical line separates clusters defined as unmutated or mutated (threshold of 7). * $p < 0.05$; two-sided t-test. Bars show mean values ±SEM. (**B**) The number of mutated or unmutated clusters annotated as having specificity for pH1N1 based on comparison to previous data was determined. Bars show mean values ±SEM. Individual data points are shown, with the shape representative of whether the participant received the seasonal TIV prior to pH1N1 vaccination. P values show the result from a two-sided Mann-Whitney U test.

While the unmutated clusters represent a minority of the total number of clusters, they are more enriched for pH1N1-specificity (based on comparison to previously described sequences) than the mutated clusters. On average, 0.42% of unmutated clusters were annotated as pH1N1-specific, compared to 0.15% of the mutated clusters. Comparing the clusters annotated as pH1N1-specific between the vaccine groups showed that more pH1N1-specific clusters were present in the adjuvant compared to the non-adjuvanted group (Figure 7.7 B). While this difference remained regardless of whether unmutated or mutated clusters were considered, it was more significant for the unmutated ($p = 0.0075$; 9-fold increase) compared to the mutated ($p = 0.0504$; 2-fold increase) clusters. These data suggest that the increase in unmutated B cells stimulated by the adjuvanted vaccine also relates to an increase in vaccine-specific B cells.

7.3.7 Adjuvant increases adaptation of recalled cells

Next, the effect of the adjuvant on recalled clusters was investigated. While mutated clusters are likely to represent those derived from memory recall, their exact antigenic specificity is unknown. To study recalled clusters that are likely to be influenza-specific, those present in the TIV dataset, that then re-appear in the pH1N1 dataset were found for each participant. On average, there were 52 clusters for each participant found in both the TIV and pH1N1 datasets, equating to approximately 1% of the total number of clusters present in the two datasets. This number was similar regardless of whether the adjuvanted or non-adjuvanted pH1N1 vaccine was given (Figure 7.8 A). The change in the properties of these clusters from the TIV to the pH1N1 dataset was then determined. Following the pH1N1 vaccination, there was an increase in average cluster size and mutation than from following TIV vaccination, but there was no difference between the adjuvanted and non-adjuvanted vaccine groups (Figure 7.8 6B/C). Lineage trees were also generated from the clusters shared between the TIV and pH1N1 datasets. First, lineages were created just using the sequences from the TIV dataset, and then lineages were created using sequences from both the TIV and pH1N1 datasets. Adding in the

pH1N1 sequences led to an increase in diversity of the lineages. This increase in diversity of the lineages was greater in the adjuvanted compared to the non-adjuvanted group (Figure 7.8 D, p = 0.0211). Visualising the lineages showed that in the non-adjuvanted group, the TIV and pH1N1 sequences were similar, which is why there was only a small increase in diversity (Figure 7.8 E; left side lineages). In the adjuvanted group, the pH1N1 sequences diverged more from the TIV sequences, which is what lead to the greater increase in diversity of these lineages (Figure 7.8 E; right side lineages).

7.3.8 Adjuvant stimulates increased recall of cells with attributes of cross-reactivity

To further investigate properties of the recalled clusters between the TIV and pH1N1 vaccines, their V gene usage was determined. While biases in V gene use were broadly similar between the two vaccine groups, there were some differences between them (Figure 7.9). The greatest group differences were in IGHV1-18 (p = 0.0326), IGHV3-20 (p = 0.0432), and IGHV1-69 (p = 0.0582), which were used to a greater extent in the adjuvant compared to non-adjuvanted group. As IGHV1-69 is associated with HA stem binding cross-reactive antibodies, next the recalled clusters were compared to the database of 59 previously characterized cross-reactive HA stem binding monoclonal antibody sequences isolated from H1N1 vaccination studies (Table 9.2 and (166, 236)). Five of the recalled lineages contained sequences with the same CDR3 AA sequence, and V and J gene segment usage as those from the previous studies, and all of these lineages were present in participants receiving the adjuvanted vaccine only. From one of the published studies, the full V_H nucleotide sequence was available (236), so this could be compared to the sequences within the recalled lineages. Comparing the published sequence to its closest relative in each lineage indicated that there were only a small number of non-synonymous changes between them (12, 15 and 18 for the three lineages), so may not effect antigen-binding ability (Figure 9.19). Critically, the CDR2 Phe residue at position 54, and

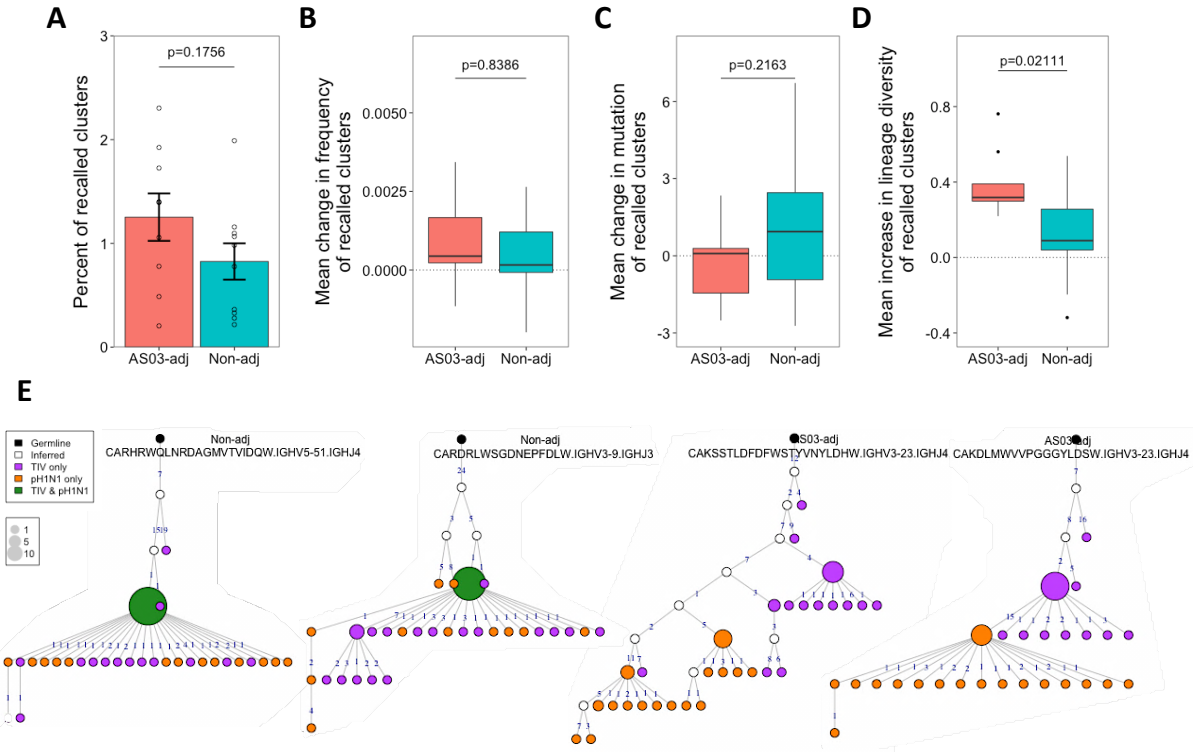

Figure 7.8: Measuring memory recall from TIV to pH1N1 vaccination - (**A**) For the 19 participants who were given a TIV vaccine four months prior to receiving the pH1N1 vaccine, the number of clusters that could be found in the PC sequence data following both vaccines was determined. Percent was then calculated as (A∩B/sum(A,B))*100. Bars show mean values ±SEM. For the clusters shared between the two vaccine datasets, the frequency of the cluster, the average mutation of the cluster, and the diversity of the cluster was calculated for each dataset. (**B**-**D**) Changes in cluster frequency, mutation, and lineage diversity from the TIV vaccine dataset to the pH1N1 vaccine dataset. Boxes show locations of the 25, 50, and 75th percentiles, and whiskers show data within 1.5x the interquartile range. For **A**-**D**, p values represent the result from a two-sided t-test. (**E**) Example lineages generated from the combined TIV and pH1N1 data using sequences from clusters present in both datasets. Two examples, chosen to best illustrate the differences, are shown for participants given the non-adjuvanted vaccine (left), and two examples are shown for participants given the adjuvanted vaccine (right). Each node in the lineage tree represents a unique sequence, and the size of the node represents the number of those sequences. The black node is the germline sequence, white nodes are inferred common ancestor sequences, and the coloured nodes are those found in the datasets (purple = TIV only, orange = pH1N1 only and green = TIV and pH1N1). Numbers on the edges of adjoining nodes show the number of mutations separating the sequences.

the CDR3 Tyr residue at position 98, which are required for optimal binding (234), remained unchanged.

7.3.9 Adjuvant stimulates subclass switching to IgG1 and IgG3

Subclass usage differed considerable between the vaccine groups (Figure 7.10 A). There was a higher proportion of IgG1 ($P < 0.0001$) and IgG3 ($P = 0.0074$) sequences, and a lower propor-

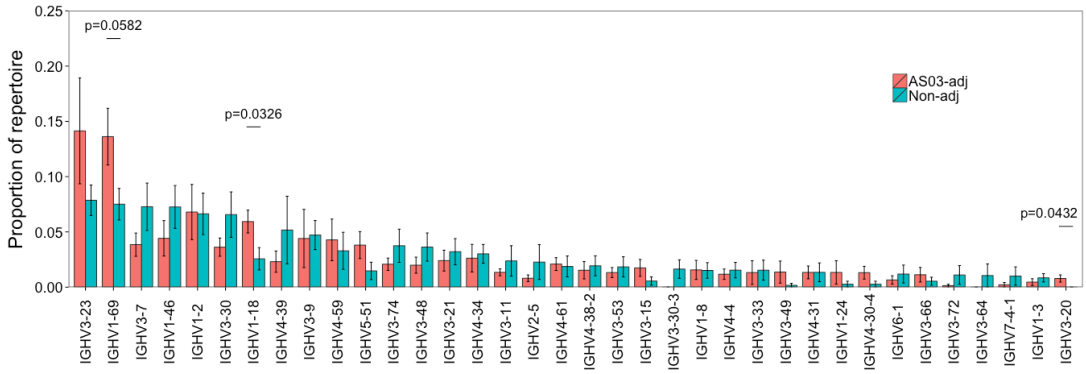

Figure 7.9: V gene usage of recalled clusters - For the clusters shared between the TIV and pH1N1 datasets, the proportion of these clusters utilizing different V gene segments was determined for each participant. Bars show mean values ±SEM. Shown are the V genes with a proportion above 0.005 in at least one of the vaccine groups. P values are shown for the three greatest differences between the adjuvanted and non-adjuvanted groups. Comparisons were performed using a two-sided t-test.

tion of IgG2 ($p < 0.0001$) and IgG4 ($p = 0.0721$) sequences in the adjuvanted compared with the non-adjuvanted vaccine groups. These differences remained the same when considering only the mutated clusters, but the difference between the vaccine groups in IgG3 usage disappeared when considering only the unmutated clusters (Figure 9.20).

The percent of sequences in the repertoire comprised of the different subclasses for each participant was then correlated with their vaccine response as measured by H1N1(California) HAI antibody titer (Figure 7.10 B). For both IgG1, and IgG3, the proportion of the repertoire comprised by these subclasses showed a moderate positive correlation with their vaccine response ($p = 0.0270$, $r = 0.354$ and $p = 0.0014$, $r = 0.493$ respectively). On the other hand, the proportion of the repertoire comprised by IgG2 sequences showed a negative correlation with vaccine response, while the proportion of the repertoire comprised by IgG4 sequences showed no correlation with vaccine response. The adjuvanted vaccine therefore promoted increased generation of IgG1 and IgG3, which correlated with an improved vaccine response.

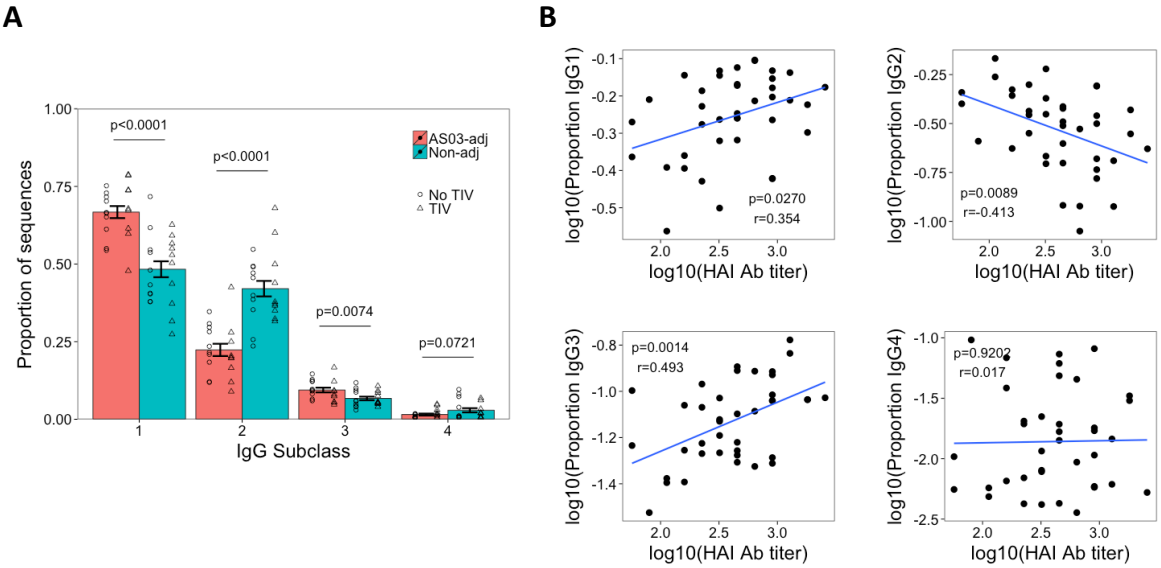

Figure 7.10: Analysis of IgG subclass usage - (**A**) For each participant, the proportion of sequences of each IgG subclass was determined. Bars show mean values ±SEM. P values represent the result from a two-sided t-test. Individual data points are shown, with the shape representative of whether the participant received the seasonal TIV prior to pH1N1 vaccination. (**B**) The proportion of the repertoire comprised by each of the four IgG subclasses was then correlated with the H1N1(California)-specific HAI antibody titer. R values show the Pearson product-moment correlation coefficient.

7.4 Discussion

In this study, high-throughput sequencing of the PC BCR repertoire was applied for detailed investigation of the effect of the AS03 adjuvant on pH1N1 influenza vaccination. It was found that the increased immunogenicity of the adjuvanted vaccine could potentially be explained by both an increase in the activation of naive B cells, as well as an increase in the adaptation of pre-existing memory B cells. There were a number of differences in the repertoire caused by the adjuvant, which could be used to distinguish the response to the two vaccines, and back up these findings.

A previous study of the PC response following seasonal influenza vaccine showed that the activated vaccine-specific PCs have high levels of mutation, indicating that they are likely derived from memory recall, and not activation of naive B cells (104). These PCs were still specific for the current vaccine antigen, indicating that the response to influenza tends to be dominated

by memory recall, and that the recalled cells can then fine-tune their specificity through further mutation (104). In the context of pandemic vaccine, where the difference between the vaccine strain and previously encountered strains is greater, this fine-tuning of specificity of the recalled cells may not be as efficient, thus leading to immune interference through original antigenic sin, and a reduction of vaccine immunogenicity (231). Indeed, previous studies have shown that prior receipt of TIV does reduce the subsequent response to pH1N1 vaccination (237, 238). In this study, it was observed that following adjuvanted vaccine there is an increase in the number of PCs with low mutation levels, indicative of recent activation from naive B cells. This suggests that the adjuvant potentially increases immunogenicity by overcoming this immune interference by instead activating more naive B cells specific to the new vaccine antigen. While it was not possible to functionally verify the antigenic-specificity of all the PCs that were isolated in this study, by comparing them to previously described pH1N1-specific sequences, it was possible to show that they were enriched for vaccine-specificity (Figure 7.5). In addition, previous studies have shown that 7 days following administration of pH1N1 vaccine, 33-80% of PCs isolated are specific to the vaccine (104, 233).

The distinct V gene usage profile seen when comparing the unmutated and mutated PCs may relate to distinct epitopes targeted by these two populations. There have been numerous studies showing that certain V genes are preferentially used in the response to different antigens or epitopes (112). In the context of influenza, IGHV1-69 has been well characterized as being important in the production of HA stem-binding cross-reactive antibodies (234). The reduction in IGHV1-69 among other V gene segments in the unmutated PC sequences may therefore relate to a decreased tendency of these sequences to bind the HA stem. This is a similar observation to that from Khurana *et al.*, who found that the MF59 adjuvant stimulated a switch away from stem binding towards head binding specificity (165, 232).

While the adjuvant did appear to increase the proportion of unmutated PCs likely derived

from naive B cells, these still formed a minority of the total PC response, so it is likely that the adjuvant functions through other mechanisms as well. When the recalled clusters between the TIV and pH1N1 vaccines were specifically studied, it was found the adjuvant stimulated greater diversification of the lineages. This is consistent with the adjuvant having a greater ability to fine-tune the specificity of the lineages through further rounds of affinity maturation (104). This increased ability of the adjuvant to drive re-diversification may be what enables the adjuvant to overcome the interference effects from prior receipt of TIV (237, 238). It was further found that some of the recalled clusters in the adjuvant group had properties consistent with them being derived from cross-reactive B cells that bound the HA stem. While the pH1N1 vaccine has previously been seen to induce cross-reactive stem binding B cells (166), this study indicates that cross-reactive B cell activation is enhanced if the vaccine is adjuvanted.

The AID enzyme, which drives somatic hypermutation and affinity maturation, and may therefore be responsible for lineage diversification, also drives isotype class switching (72). Specifically, switching to both IgG1 and IgG3 is related to TNF-α and IL-6 production, and this also correlates with the production of AID (239). It has previously been shown that IgG3 is the most important subclass used in the H1N1 response (239, 240). Focusing on IgG3 in the dataset in this study, showed IgG3 to be more prevalent following adjuvanted compared to non-adjuvanted vaccine, and that there was a positive correlation between the proportion of IgG3 sequences in the repertoire, and the HAI Ab titer. It may be that the Fc receptor of IgG3 is more effective at neutralizing influenza than other subclasses (10), or that the subclass switching is just a proxy for AID production, and it is the AID mediated lineage diversification which improves the neutralizing ability of the antibodies.

7.5 Conclusion

The data presented in this chapter support the use of the adjuvant for increasing the immunogenicity of influenza vaccines, particularly where the vaccine is highly different to previously encountered strains and thus requires increased naive B cell activation, and/or increased adaptation of pre-existing memory B cells. Furthermore, this study adds to the data from the previous chapters, demonstrating the utility of BCR sequencing for understanding vaccine responses, and the benefit of having previously described antigen-specific sequence data with which to annotate these datasets.

8

Summary

8.1 Re-iteration of aims

The specific aims of this thesis were as follows:

1. To investigate BCR repertoire structure and diversity both within and between individuals in the absence of any specific immune stimulus

2. To determine to what extent BCR repertoire sequencing can be used to investigate the B cell response to primary and secondary antigen exposure, using hepatitis B vaccination as a model system

3. To develop analytical models for identifying antigen-specific BCR sequences from the total BCR repertoire following antigen stimulation by combining data from both the total BCR repertoire, and the antigen-specific BCR repertoire from antigen-specific sorted cells

4. To investigate the use of BCR repertoire sequencing in understanding/elucidating immunological mechanisms of vaccine response in two case studies:

   (a) Tracking the B cell response to meningococcal polysaccharide and conjugate vaccines

   (b) Investigating the effect of adjuvant on influenza vaccine

8.2 Chapter summary

The broad aim of this thesis was to provide a comprehensive analysis of the BCR heavy chain repertoire in humans, and how this responds to vaccination. The first stage of this was to develop a robust protocol for both sequencing the BCR repertoire, and pipelines for data analysis. Optimisation of the laboratory protocol now allows reliable generation of BCR heavy chain amplicon libraries starting from whole blood in a single day, for approximately £10 per sample, with sequencing costs then coming to approximately £45 per sample. Although there is scope to further reduce costs, this represents a low enough cost such that this technology could potentially be used on a more routine basis in a clinical setting. The bioinformatic pipeline developed for initial sequence annotation and analysis takes approximately one week to complete for a batch of samples, using a standard desktop computer (3.4GHz Quad-Core Intel i7 processor with 32GB memory). While the exploratory nature of the downstream analyses presented here required extensive optimisation of the parameters used, and incorporation of novel analytical techniques, this process becomes faster with each new dataset analysed, and could ultimately be made more automated.

Using these methodological advances, Chapter 3 then explored the repeatability of the protocol, showing that common metrics used to characterise the repertoire are highly repeatable, and even at the clonal level, there is approximately 44% overlap of abundant clones in repeat IgG repertoire samples. Characterising the repertoire over time in a healthy participant in the absence of any known immune stimulus showed the striking degree to which some of these features (mutation, diversity and CDR3 length) could change over time, and that this was more dramatic than many of the changes seen in response to vaccination. Despite these changes, VDJ usage remained highly conserved over time, and appears to be unique to a participant. This information suggests that such features can be used for quality control of repertoire data to detect both biological and technical outliers, and indicates that when studying responses to

intervention, caution should be used so as not to over-interpret certain changes over time.

Chapters 4 and 5 then used HepB vaccine as a model to comprehensively study how the repertoire responds to both primary and booster vaccination. Combining total repertoire data with sequence data from vaccine-specific sorted cells allowed deconvolution of the vaccine-specific from background non-specific clusters in the total repertoire. Following the repertoire over multiple timepoints, showed that it was possible to track the dynamics of specific clones, and revealed certain stereotypic changes in the repertoire following vaccination. This information allowed determination of a model that could be used to enrich for vaccine-specific sequences in scenarios where it is not feasible to carry out antigen-specific cell sorting. A key feature of this is that of convergence: when multiple people are exposed to the same antigen, there is a degree of similarity in the B cell response. This study also showed that repertoire data could be related to ELISpot, indicating its potential to be used as a correlate of immunogenicity. Furthermore, the repertoire data gave additional information on the vaccine response not possible from ELISpot, and indicated a surprising dominance of cross-reactive sequences thought to be derived from memory recall even in the primary vaccine response.

Using the knowledge, and analytical tools developed during the previous chapters, BCR sequencing was then applied to the two specific case studies of meningococcal (Chapter 6) and influenza (Chapter 7) vaccination. Chapter 6 built on the ability to track specific B cell clones over time, by tracking the movement of specific B cell clones between different cell subsets. This information could then be used to determine which cell subsets were used in the response to a polysaccharide versus conjugate meningococcal vaccine. This study also showed the inherent differences there are between the different cell subsets, which can be related to their level of maturation.

Chapter 7 showed how repertoire data could be used to shed new light on the effect of an adjuvant on the pandemic influenza vaccine response. By splitting repertoire data into

clusters thought to have been recently generated from naive B cells in response to the vaccine, versus those thought to have been generated from recall of previously activated B cells, it was possible to show that the adjuvant functions via two mechanisms. First, the adjuvant stimulates increased activation of naive B cells, thus reducing the effect of immune interference with previous vaccine responses. Second, the adjuvant is able to increase the adaptability of the recalled cells to give improved specificity to the new vaccine antigen. This study was also able to show how analysis of V gene usage can give an indication of the epitopes likely to be targeted by certain clusters, and demonstrated that the clusters activated from naive B cells may be more likely to target the HA head while those activated from previously activated cells may be more likely to target the conserved HA stem.

8.3 Final discussion

The data presented here fulfil the aims of the thesis, and comprehensively describe the utility of BCR sequencing for studying the vaccine response. A large amount of this project has focused on the development of methods to analyse these data. There is currently little consensus in the field on the optimal ways to pre-process, cluster and analyse BCR data. While optimising the methods for this thesis, a large variety of the publicly available pre-processing tools were tested, and different clustering methods and thresholds were trialled before deciding on the final method. The final method chosen was based on a rational approach of exploring the data and setting thesholds that made biological sense as well as trying to minimise computing time and binning of potentially important sequences. It is worth noting however, that changing the pre-processing and clustering methods had remarkably little effect on the outcome of the subsequent analysis, and interpretation of the data (not presented due to space limitations). Analysis of BCR repertoire data therefore appears to be highly robust to different processing methods, which is important for considering the utility of this method as a clinical tool.

In the Introduction, the potential applications of BCR repertoire sequencing with relation to vaccination that were discussed were the generation of mAb sequences, tracking sequences with known specificity, investigating the breadth of the antibody response, understanding the immunological mechanisms of vaccination, and measuring vaccine immunogenicity. One of the key developments in this thesis which has allowed increased extraction of information from the vaccine datasets to help realise these applications, has been the deconvolution of the vaccine-specific from total repertoire. Three methods have been presented for achieving this deconvolution: comparison of the total repertoire to datasets derived from B cell subsets enriched for vaccine-specificity, computational enrichment based on cluster properties thought to be representative of recent activation to a common stimulus, and comparison to a previously described mAb sequence database. The mAb sequence database is perhaps the best resource for this, as it allows identification of sequences with a range of specificities, and can easily be used to annotate new datasets as they become available. It would be of great benefit to improve upon the database presented in Table 9.2 to make an online tool available for the whole community to use, where researchers could also upload new mAb sequences as they are described.

The utility of such a sequence database requires that there is a degree of convergence between individuals in the response to a particular antigen, which has been shown to be the case for at least the antigens described in this thesis. The database also needs to be sufficiently comprehensive to capture at least a representative sample of the potential sequences that can be involved in the response to a particular antigen. Currently it is hard to determine an estimate for the size of the total number of sequences that could be produced in response to a specific antigen. The largest database generated here is the HBsAg-specific sequence database, and rarefaction analysis gives no evidence of saturation. However, it may be the case that at the structural level the number of variants is smaller than at the sequence level due to different

sequences resulting in the formation of antibodies with similar tertiary structures. Tools for predicting antibody structure from sequence data alone are rapidly improving (241), and offer potential in the future for high-throughput analysis of BCR repertoire datasets at both the sequence and structural level (242).

Once the vaccine-specific repertoire has been identified, it has been demonstrated in this thesis how specific sequences can be tracked over time and between subsets, and compared between vaccine groups to increase understanding of the immunological mechanisms of vaccination. What is still uncertain is to what degree study of the vaccine-specific repertoire can be used to determine vaccine immunogenicity. While the number of vaccine-specific clusters can be related to ELISpot data, it has not been determined whether it is possible to use this information to predict how well the vaccine has worked in a particular individual. To investigate this in more detail, it will be necessary to have larger studies using vaccine antigens for which well validated indicators of protection exist, and then to assess how well features in the repertoire correlate to these indicators of protection.

The data obtained in this thesis has given insight into what constitutes a normal repertoire at baseline, as well as what happens to the repertoire in response to a controlled antigen stimulus. Having a clear definition of what can be considered normal for different repertoire metrics at baseline and in response to antigen stimulus is also essential for the interpretation of repertoire data from studies of the response to infection (118), and in the context of autoimmunity (137) and immune-deficiency (193). Increasing the number of participants used to generate the baseline metrics would be useful so that more accurate distributions can be determined. For investigating B cell defects, or B cell activation, the location of the sample metrics in relation to the distributions can then be measured to asses the probability of there being a clinically relevant deviation in the sample.

8.4 Future work

The data obtained for this thesis has given good insight into both the applications and limitations of BCR repertoire sequencing. The next step is to move towards testing more focused clinical applications of this technology. Although there are many potential avenues for future work, I will describe a few here which are direct outcomes of this thesis, and are currently underway.

As his been discussed though this thesis, PCR and sequencing error will introduce erroneous sequences into the datasets. While the clustering approach used can collapse erroneous sequences into a single cluster, so that individual clusters can be treated as proxies for error corrected B cells clones, it is not possible to correct for error when analysing the lineages of sequences within a cluster. While some methods do exist to try and account for this error (243), trialling these methods yielded little success, likely due to the inherent difficulties in discerning true error from SHM. Methods have now arisen for correcting error based on addition of UMI's, as was discussed in the Introduction (Figure 1.8), and adaptation and validation of such a method for use in our laboratory is under way.

One clinical application of BCR sequencing that has become apparent is that of disease diagnostics. As part of this thesis, good evidence for repertoire convergence in response to HepB and meningococcal vaccination has been obtained. Work conducted during this thesis, but not included in the thesis has shown convergence in the repertoire in multiple sclerosis patients, and in individuals infected with *Salmonella typhi*. Additionally, previous work from our lab has shown convergence in response to Hib polysaccharide vaccine (150), and from other labs has shown convergence in response to influenza vaccination (203) and dengue infection (118). Combined, this evidence suggests that there is a degree of persistent repertoire convergence in response to most antigens, so looking at convergent repertoires may be a way to find diagnostic markers of certain diseases. Future work is now planned to investigate convergence in

response to a range of bacterial (*Staphylococcus aureus*, *Streptococcus pneumoniae*, *Escherichia coli*), viral (Herpes simplex virus, respiratory syncytial virus, influenza) and mycobacterial (*Mycobacterium tuberculosis*) infections. Convergent signatures will be determined, and then validated against a test set of infected and healthy participants.

A limitation to much of the work described within this thesis is the inability to express and functionally characterise the BCR sequences of interest that were obtained. This limitation arises from the antibody protein consisting of both a V_H and a V_L chain that contribute to it's specificity, but the sequencing protocol only being able to capture V_H chain information. Linking V_H and V_L chains in a high-throughput manner is challenging, but approaches are starting to be developed to overcome this. Single-cell methods using high-density microwell plates and barcoded primers are becoming more common, but are limited to studying just thousands of cells at a time (244). To date, the most high-throughput approach is that developed by Dekosky *et al.*, using microfluidics to carry out in-droplet V_H:V_L linkage PCR which is able to analyse >2 x 10^6 B cells per experiment (174). However, both of these methods require customised equipment, so are challenging to transfer to other laboratories. Working with the Wellcome Trust Centre for Human Genetics single-cell group, we are currently trying to develop a microfluidic approach where all transcripts within an individual cell are given a unique barcode during reverse transcription. V_H and V_L repertoires can then be amplified and sequenced, and paired based on having a shared barcode.

The combination of the two aforementioned lines of work will result in a powerful method for the rapid identification of mAb sequences, which can subsequently be used for research or therapeutic purposes. Following antigen administration, PCs can be isolated and sequenced. The convergent repertoire can then be found based on the assumption that this will be enriched for antigen-specificity. To confirm the specificity, the sequences can then be cloned into a cell line for expression, and functionally characterised. It is planned that this method be applied

to an Ebola vaccine trial to find a diverse panel of anti-Ebola virus mAbs. Currently the most effective treatment for Ebola is ZMapp, which consists of three mAbs targeting distinct regions on the Ebola virus glycoprotein (245). Although the efficacy of this drug is not well documented, there are many cases where it has not worked, possibly due to the appearance of viral escape variants (246). The discovery of novel mAbs to create more advanced cocktails would therefore be of great clinical benefit.

9

Appendix

9.1 Appendix to Chapter 3

9.1.1 Figures

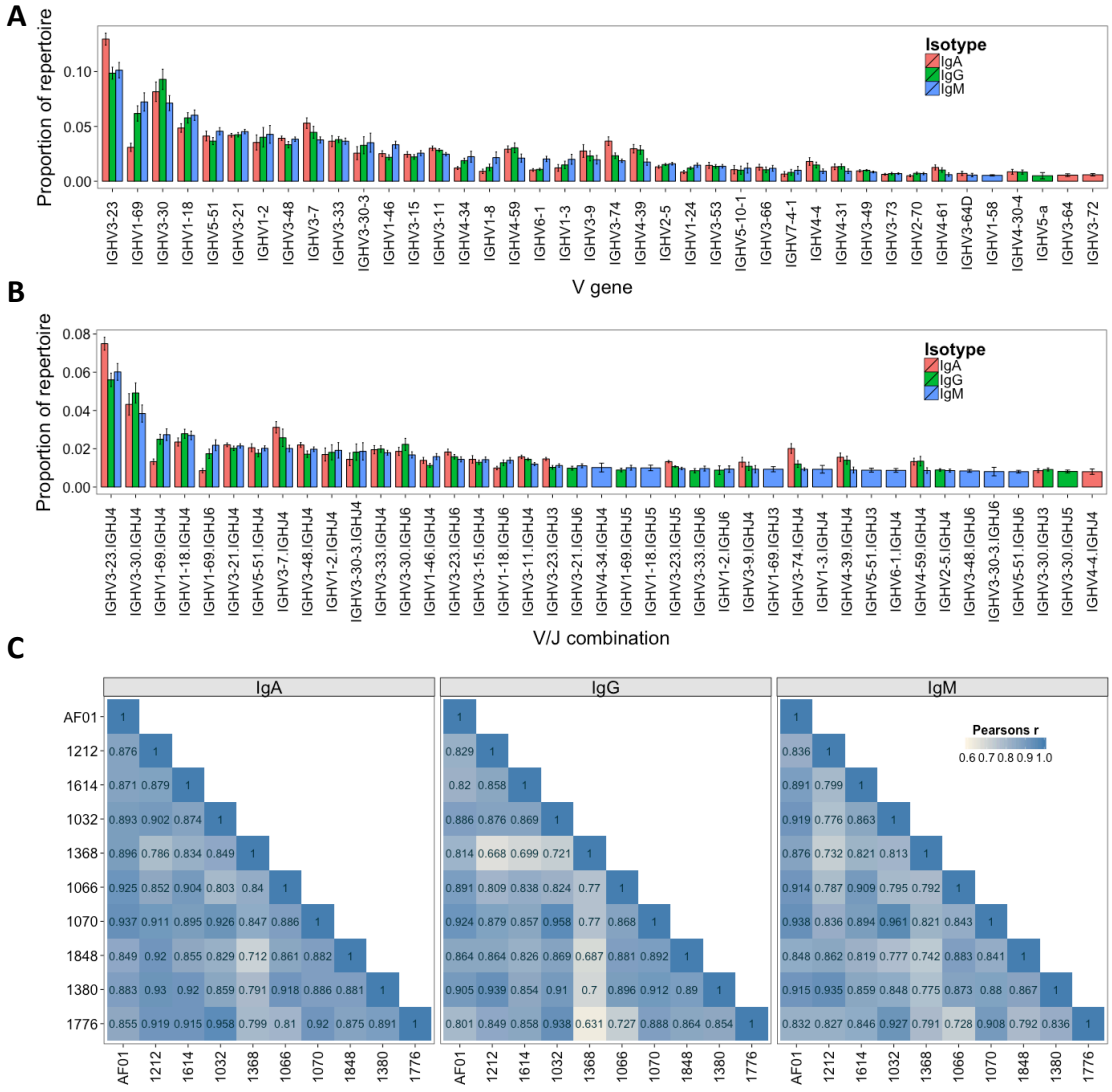

Figure 9.1: IgH V and VJ gene use - Usage frequency of different V genes (**A**) or V/J gene combination (**B**) in the repertoire. Of the 71 different V genes, and 379 different V/J combinations, only the 41 most frequent are shown. Ordering is by frequency in the IgM dataset. Bars show mean values from 10 participants, and error bars indicate ±SEM. (**C**) Correlation in relative usage proportion of each VJ gene combination from samples from different participants. Stronger correlations are in darker blue.

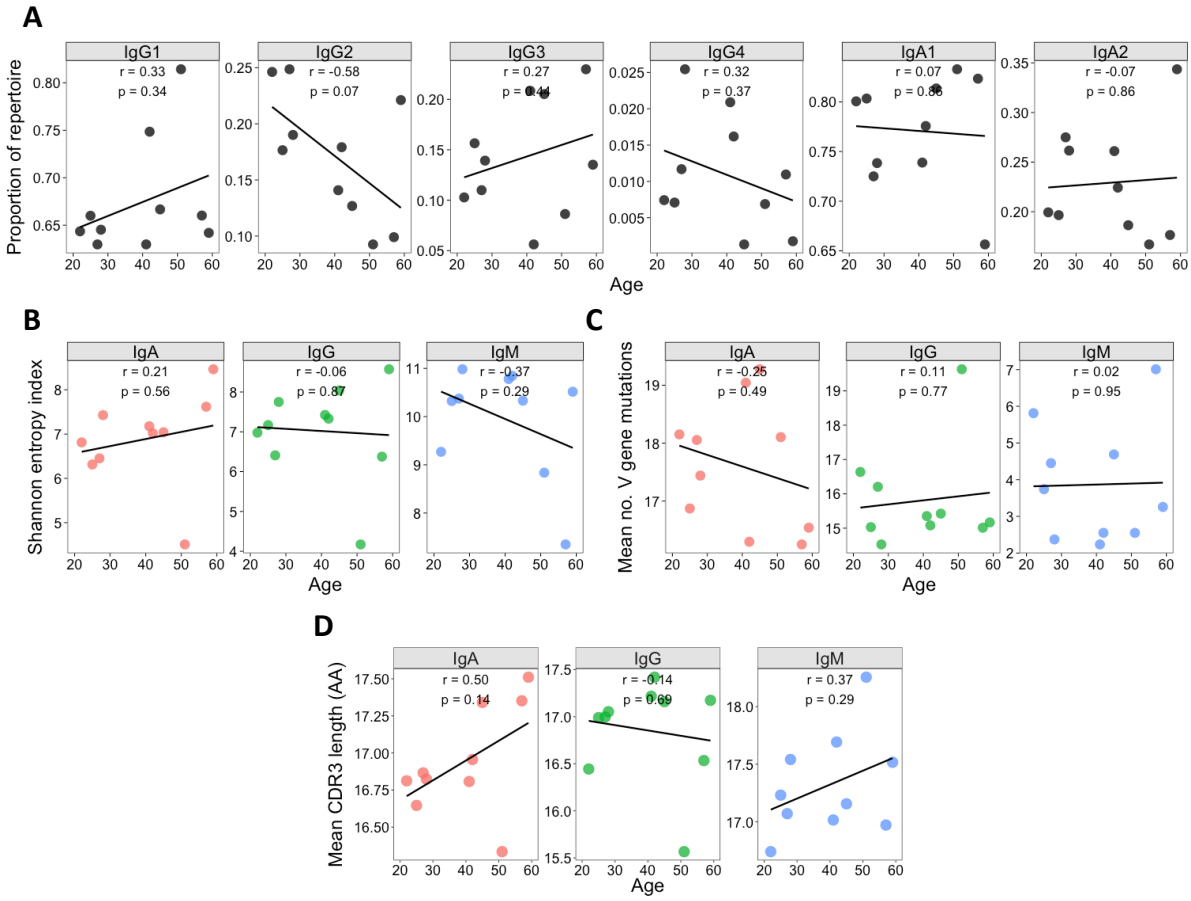

Figure 9.2: Correlating repertoire properties with age - Differences in the proportion of the repertoire comprised by sequences of each IgG and IgA subclass (**A**), repertoire diversity (calculated using the Shannon entropy index) (**B**), mean number of V gene mutations (**C**), and mean CDR3 AA length (**D**) in the different aged participants. r values calculated using Pearson's product-moment correlation.

9.1.2 Tables

Table 9.1: Summary of the samples used for sequencing, and sequence data obtained - IgG 2 samples represent PCR replicates, and IgG 3 samples represent biological replicates.

Participant	Day	Replicate	Cells	Isotype	Cell number	Raw seqs	Filtered seqs	Clusters
1032	0	Sample	Total B cells	IgA	500,000	281,579	100,000	4,946
1032	0	Sample	Total B cells	IgG	500,000	243,879	100,000	5,221
1032	0	Sample	Total B cells	IgM	500,000	331,231	100,000	47,051
1066	0	Sample	Total B cells	IgA	500,000	384,924	100,000	11,360
1066	0	Sample	Total B cells	IgG	500,000	336,302	100,000	16,560
1066	0	Sample	Total B cells	IgM	500,000	410,635	100,000	75,112
1070	0	Sample	Total B cells	IgA	500,000	367,806	100,000	11,684
1070	0	Sample	Total B cells	IgG	500,000	353,992	100,000	13,513
1070	0	Sample	Total B cells	IgM	500,000	533,009	100,000	24,623
1212	0	Sample	Total B cells	IgA	500,000	292,080	100,000	6,500
1212	0	Sample	Total B cells	IgG	500,000	280,407	100,000	7,278
1212	0	Sample	Total B cells	IgM	500,000	263,810	100,000	35,916
1368	0	Sample	Total B cells	IgA	500,000	324,406	100,000	1,658
1368	0	Sample	Total B cells	IgG	500,000	229,626	100,000	1,599
1368	0	Sample	Total B cells	IgM	500,000	327,464	100,000	18,998
1380	0	Sample	Total B cells	IgA	500,000	282,185	100,000	9,939
1380	0	Sample	Total B cells	IgG	500,000	376,162	100,000	14,116
1380	0	Sample	Total B cells	IgM	500,000	385,190	100,000	71,570
1614	0	Sample	Total B cells	IgA	500,000	265,633	100,000	5,825
1614	0	Sample	Total B cells	IgG	500,000	216,878	100,000	10,584
1614	0	Sample	Total B cells	IgM	500,000	295,468	100,000	45,118
1776	0	Sample	Total B cells	IgA	500,000	252,253	100,000	12,542
1776	0	Sample	Total B cells	IgG	500,000	309,604	100,000	19,061
1776	0	Sample	Total B cells	IgM	500,000	359,614	100,000	75,422
1848	0	Sample	Total B cells	IgA	500,000	387,084	100,000	22,942
1848	0	Sample	Total B cells	IgG	500,000	427,766	100,000	24,998
1848	0	Sample	Total B cells	IgM	500,000	473,752	100,000	64,903
AF01	0	Sample	Total B cells	IgA	500,000	369,765	100,000	9,356
AF01	0	Sample	Total B cells	IgG	200,000	366,431	100,000	12,152
AF01	0	Biological	Total B cells	IgG	500,000	1,516,275	100,000	17,390
AF01	0	PCR	Total B cells	IgG	200,000	339,893	100,000	12,447
AF01	0	Sample	Total B cells	IgM	500,000	381,663	100,000	55,821
AF01	7	Sample	Total B cells	IgA	500,000	416,194	100,000	7,274
AF01	7	Sample	Total B cells	IgG	200,000	269,345	100,000	13,186
AF01	7	Biological	Total B cells	IgG	500,000	324,996	100,000	11,812
AF01	7	PCR	Total B cells	IgG	200,000	394,522	100,000	13,350
AF01	7	Sample	Total B cells	IgM	500,000	300,980	100,000	51,019
AF01	14	Sample	Total B cells	IgA	500,000	293,845	100,000	11,814
AF01	14	Sample	Total B cells	IgG	400,000	406,863	100,000	16,907
AF01	14	Biological	Total B cells	IgG	500,000	350,925	100,000	23,262
AF01	14	PCR	Total B cells	IgG	400,000	380,327	100,000	17,679
AF01	14	Sample	Total B cells	IgM	500,000	271,986	100,000	58,103
AF01	21	Sample	Total B cells	IgA	500,000	310,648	100,000	11,259
AF01	21	Sample	Total B cells	IgG	400,000	480,514	100,000	17,930
AF01	21	Biological	Total B cells	IgG	500,000	296,139	100,000	21,374
AF01	21	PCR	Total B cells	IgG	400,000	333,238	100,000	18,200
AF01	21	Sample	Total B cells	IgM	500,000	381,581	100,000	59,674
AF01	28	Sample	Total B cells	IgA	500,000	258,421	100,000	12,506
AF01	28	Sample	Total B cells	IgG	280,000	439,999	100,000	15,463
AF01	28	Biological	Total B cells	IgG	500,000	464,562	100,000	22,029
AF01	28	PCR	Total B cells	IgG	280,000	344,963	100,000	15,784
AF01	28	Sample	Total B cells	IgM	500,000	430,162	100,000	60,765

Table 9.2: Previously described antigen-specific antibody sequences

V gene	D gene	J gene	CDRH3 AA Sequence	Antigen	Subtype	Ref.
3-33	2-21	3	CVREATDFVVKFDLW	HBsAg	NA	(247)
4-39	3-10	6	CTRLIKDYNGWALFQQYYFMDVW	HBsAg	NA	(247)
3-49	5-12	4	CTRAIPRGPTVWWLPGYYFDYW	HBsAg	NA	(247)
3-71	6-13	4	CTAGFDYW	HBsAg	NA	(247)
3-30	4-17	4	CIRDQTYGVHRFDSW	HBsAg	NA	(247)
3-30	5-18	4	CARTAFFNAYDFW	HBsAg	NA	(247)
1-46	4-11	6	CARSDIYYGNYNALDYW	HBsAg	NA	(247)
3-23	1-14	5	CARRNHMRTRW	HBsAg	NA	(247)
3-11	6-19	4	CARKLRNGRWPLVYW	HBsAg	NA	(247)
5-a	3-10	3	CARHVREKSMVQGVIIKDAFDIW	HBsAg	NA	(247)
3-21	1-26	5	CARGTVGYWFDPW	HBsAg	NA	(247)
4-4	3-9	4	CARGLSGFDYW	HBsAg	NA	(247)
1-46	3-22	5	CARGLDSSGYYYFWFDPW	HBsAg	NA	(247)
3-30	2-21	6	CARGGGVYYGINYAMDYW	HBsAg	NA	(247)
4-4	3-10	4	CARGGFGEFDYW	HBsAg	NA	(247)
3-33	6-13	4	CARERLIAAPAAFDLW	HBsAg	NA	(247)
3-33	3-3	4	CAREGLEWFPILDYW	HBsAg	NA	(247)
3-33	4-17	4	CAREELTLVTAFGYW	HBsAg	NA	(247)
3-33	4-23	4	CAREALLLWTIFDSW	HBsAg	NA	(247)
3-13		5	CARDLELW	HBsAg	NA	(247)
3-23	1-26	4	CAKGATRELLRLYSDYW	HBsAg	NA	(247)
3-20	3-3	6	CAKFARAWSGPQFTDYYYYMDVW	HBsAg	NA	(247)
3-23	6-13	4	CAKDRGRIAAAHFDYW	HBsAg	NA	(247)
3-30	3-10	4	CAKDQLYFGSQSPGHYW	HBsAg	NA	(247)
1-2	3-16	6	CAIMTTFLGEGYAMDYW	HBsAg	NA	(247)
1-3	6-19	4	CARDLLSNGWSLSYYFDFW	HBsAg	NA	(110)
3-30			CAKAARGGDGYGAYDDYDYVFQDLW	HBsAg	NA	(110)
4-34	1-26	6	CARGRRPTVVKYYYFYDGMDVW	HBsAg	NA	(110)
4-4	5-24	3	CARDGLRMDGGWIVTHGFDVW	HBsAg	NA	(110)
4-4	2-15	3	CARQCQGDGCYGSDAFDFW	HBsAg	NA	(110)
4-4	3-10	3	CARDGLRMDRGLIVTHGFDAW	HBsAg	NA	(110)
3-33	6-19	4	CARDTGYSSGWYLPDW	HBsAg	NA	(110)
3-33	1-26	4	CAREDRFAVGPGFDYW	HBsAg	NA	(110)
4-34	6-6	6	CARGRRPTLVKYYYFYHGMDVW	HBsAg	NA	(110)
4-39	3-3	3	CARQKVELRFLEWFRLHDSDVW	HBsAg	NA	(110)
1-3	1-26	4	CARGAKDSWRPEYYFDYW	HBsAg	NA	(110)
3-11	2-15	4	CARPGCWGGSCYPFTYW	HBsAg	NA	(110)
3-30	3-10	4	CAKSDILFGLSGGLDYW	HBsAg	NA	(110)
3-23	3-16	4	CAKDHGGLWLGEPPDYW	HBsAg	NA	(110)
3-11	2-15	4	CARPGCWGGSCYPFHYW	HBsAg	NA	(110)
4-39	3-3	5	CARQSYSDIWSGYGLHWFDPW	HBsAg	NA	(110)
3-30	3-9	6	CAKDSILLTSNPGVGIDVW	HBsAg	NA	(110)
7-4-1	2-2	6	CAREATNYAYAYGMDVW	HBsAg	NA	(110)
1-3	3-9	4	CAREGYDGSPFSPLDHW	HBsAg	NA	(110)
3-23		6	CAKDIASWYYYGMDVW	HBsAg	NA	(109)
3-23		6	CAKAIASWYYYGMDVW	HBsAg	NA	(109)
3-23		6	CAKDISSWYYYGMDVW	HBsAg	NA	(109)
3-21		3	CAKGIASFYFYGMGVW	HBsAg	NA	(109)
3-48		5	CVRDGRSCSGGFCHPYW	HBsAg	NA	(109)
3-48		5	CVRDGQSCSGGFCQPW	HBsAg	NA	(109)
3-48		5	CARDGRYCSGGICHPYW	HBsAg	NA	(109)
3-33			CAREVPHIYGPTFDLW	HBsAg	NA	(109)
3-21		6	CARRYSEGGSDRRTLYSYHHYMDVW	HBsAg	NA	(109)
3-21		6	CTRRYSEGGSDRRTLYSYHHYMDVW	HBsAg	NA	(109)
6-1		3	CARARIDLVRFYFYFYVDVW	HBsAg	NA	(109)
1-2		1	CAIEYGDTDCGGDCYSIW	HBsAg	NA	(111)
4-39		3	CARPLRLNYNSRSYYPGIGPFDMW	HBsAg	NA	(111)
4-39		2	CARPLTHDDFLTAYYPGGGYFDLW	HBsAg	NA	(111)
3-30		6	CARQYYDFWSGSSVGRNYDGMDVW	HBsAg	NA	(108)
3-33		2	CARERLIAAPAAFDLW	HBsAg	NA	(108)
1-69		6	CASGYYDFSGEGDILHYGLDVW	HBsAg	NA	(108)
3-33		6	CARGFYEAYMDVW	HBsAg	NA	(248)
1-66		3	CARSDIYYGNYNALDYW	HBsAg	NA	(249)
1-3		4	CARNYGYDESAYW	HBsAg	NA	(250)
3-21	3-10	6	CTRDGWLWGWDVRSNYYYNALDVW	HBsAg	NA	(251)
3-23	6-25	2	CAKDAILGSGHPWYFHVW	HBsAg	NA	(252)
			CTSRKSSSSDYW	HBsAg	NA	(253)
			CTRLSGRGVDYW	HBsAg	NA	(253)
			CTRKSSSSDYW	HBsAg	NA	(253)
			CTRTYSSSWYFDYW	HBsAg	NA	(253)
			CTRRGYYGSGSYYGDYW	HBsAg	NA	(253)
			CARGFHYW	HBsAg	NA	(253)
			CAKAVVDRARDGYNLGYW	HBsAg	NA	(253)
			CARGTYYGSGIGFDYW	HBsAg	NA	(253)
			CLINWGIRDW	HBsAg	NA	(253)
			CARHSEYYYDSSGYYLDYW	HBsAg	NA	(253)
			CARHLREAVADFPMDVW	HBsAg	NA	(253)
			CAKDIASWYYYGMDVW	HBsAg	NA	(253)
3-48		2	CARVSFASGTYWYFDLW	HBsAg	NA	(254)
4-1		1	CQQYFILPRTF	HBsAg	NA	(254)
3-30		6	CARGGGPTLISFYYYYYMDVW	HBsAg	NA	(254)
7-4-1		6	CARVVGIGDRRGYYYYYGMDW	HBsAg	NA	(254)
3-23		4	CAKDRGACSTNCDYW	HBsAg	NA	(254)
3-21		6	CTRRYSGGSDRRTYSYHHYMDVW	HBsAg	NA	(254)
3-21		6	CAKGIASYYGMGVW	HBsAg	NA	(254)
3-48		4	CARDGRYCSGGICHYW	HBsAg	NA	(254)
3-74		1	CVRDGSCSGGCWW	HBsAg	NA	(254)
3-21		4	CVRDGRSCSGGCHYW	HBsAg	NA	(254)
3-21		6	CARRYSGGSDRRTYSYHHYMDVW	HBsAg	NA	(254)
4-38-2		5	CASDSITIAGADSW	HBsAg	NA	(254)
3-23		6	CAKDIASWYYYGMDVW	HBsAg	NA	(254)
3-48		2	CARAGATVTSGYWYDW	HBsAg	NA	(254)

3-33		2	CARDDIAVASDW	HBsAg	NA	(254)
3-21		6	CARRYSGGSDRRTYSYHHYMDVW	HBsAg	NA	(254)
3-23		6	CAKHSATYDADVW	HBsAg	NA	(254)
6-1		3	CARARIDVRYYYVDVW	HBsAg	NA	(254)
3-48		2	CARAGATVTSGYWYDW	HBsAg	NA	(254)
3-48		2	CTRIGATTAGYWYDW	HBsAg	NA	(254)
3-33		4	CARIHVAVTDYW	HBsAg	NA	(254)
3-53		3	CASRAYYSGSSYVGYW	HBsAg	NA	(254)
3-33		4	CARDAIAVASDYW	HBsAg	NA	(254)
3-48		2	CARAGATVTSGYWYDW	HBsAg	NA	(254)
3-33		4	CARAITRGYW	HBsAg	NA	(254)
3-48		1	CVRDGSCSGGCWW	HBsAg	NA	(254)
3-33		2	CARVHIYGTDW	HBsAg	NA	(254)
4-38		5	CASDSITIAGADSW	HBsAg	NA	(254)
3-33		4	CARDAIAVASDYW	HBsAg	NA	(254)
4-38-2		5	CASDSITIAGADSW	HBsAg	NA	(254)
3-33		4	CARAITRGYW	HBsAg	NA	(254)
3-33		2	CARVHIYGTDW	HBsAg	NA	(254)
3-23		6	CAKAIASWYYYGMDVW	HBsAg	NA	(254)
3-53		3	CASRAYYSGSSYVGYW	HBsAg	NA	(254)
3-23		6	CAKDIASWYYYGMDVW	HBsAg	NA	(254)
3-9		6	CAKDAIAAAGNWDW	HBsAg	NA	(254)
1-69	5-12	6	CARGNGYSFAYGMDVW	Influenza	cross-reactive	(255)
1-69	3-10	6	CARRGFGDKYYYYGMDVW	Influenza	cross-reactive	(255)
1-69	4-11	5	CARAGTTLTRFNWFDPW	Influenza	cross-reactive	(255)
1-69	3-10	5	CARLGSGSYHNGPNWFDPW	Influenza	cross-reactive	(255)
1-69	1-26	4	CARGGRYYVDYFDYW	Influenza	cross-reactive	(255)
1-69	3-22	6	CARGQRYYYDRDGMDVW	Influenza	cross-reactive	(255)
1-69	1-26	1	CARGGGSYYGGEYFQHW	Influenza	cross-reactive	(255)
1-69	2-8	4	CARGKRPGYCSGGVCSSDYW	Influenza	cross-reactive	(255)
1-69	2-21	6	CASGLPSMVYGLDVW	Influenza	cross-reactive	(255)
3-7	6-13	6	CARTGSSWDTYYYYYAMDVW	Influenza	H1N1	(256)
3-7	6-13	6	CARSGSNWDTYYYYYGMDVW	Influenza	H1N1	(256)
3-7	4-17	6	CARQGSYADTYFYHYGMDVW	Influenza	H1N1	(256)
3-7	3-16	6	CARMGSYLDTYYYHYGMDVW	Influenza	H1N1	(256)
3-7	3-22	6	CAREGAPYDTYYYYYAMDVW	Influenza	H1N1	(256)
1-69	3-16	4	CARVGGEWGSGRYYLDHW	Influenza	H2N2/H3N2	(257)
3-33	5-12	6	CARQQDSGYSGPEVSYYSHYGMDVW	Influenza	H2N2/H3N2	(257)
4-39	3-10	4	CARLYGSGLDYW	Influenza	H2N2/H3N2	(257)
1-69	1-26	5	CARGISGSYGWFDPW	Influenza	H2N2/H3N2	(257)
4-4	2-21	1	CARAPGADAGPYSENIHHW	Influenza	H2N2/H3N2	(257)
1-69	3-10	4	CAGGSDDHAWGSFYW	Influenza	H1N1	(166)
4-31	6-19	4	CARGLEGITVGAYYFDFW	Influenza	H1N1	(166)
1-18	3-9	4	CARDRRDLLTGSLGDYW	Influenza	H1N1	(166)
1-69	5-18	1	CASPAYNSGFALLHW	Influenza	H1N1	(166)
1-69	3-22	4	CASPDLTMVFVPHTGPLDFW	Influenza	H1N1	(166)
4-59	5-12	6	CARDCSGFEDMDSFYYFMDVW	Influenza	H1N1	(166)
4-39	2-8	4	CARQLTGMVYAILLPSYFDFW	Influenza	H1N1	(166)
3-23	3-3	3	CAKDRILPYDTDAFDIW	Influenza	H1N1	(166)
3-23	6-6	4	CAKDRVVGRPWEYSLDFW	Influenza	H1N1	(166)
3-66	4-11	4	CASRHYNYDDDYG	Influenza	H1N1	(166)
3-7	3-10	5	CARAGSYGDYRPINNWFDPW	Influenza	H1N1	(166)
3-30	3-16	4	CARDPSNPPHWGNFDSW	Influenza	H1N1	(166)
3-23	4-17	4	CAKDLAVTPPAQGYLDRW	Influenza	H1N1	(166)
4-61	4-23	5	CARGIKGDYGGGANWFDPW	Influenza	H1N1	(166)
4-61	3-16	5	CARARFFGISNWFDPW	Influenza	H1N1	(166)
1-69	3-10	4	CARVGGALIRSSGSDYW	Influenza	H1N1	(166)
1-69	6-19	6	CARDDYMTVDRDYYYMDVW	Influenza	H1N1	(166)
3-7	5-24	6	CARVSREEWATVDDPHDYYYMDVW	Influenza	H1N1	(166)
3-7	5-24	6	CVRVSREEWATVDDPHDYYYMDVW	Influenza	H1N1	(166)
3-7	5-24	6	CARVSREEWATVDDPHDYYYMDVW	Influenza	H1N1	(166)
1-2	4-17	3	CARDFDYGDYRGSAFDIW	Influenza	H1N1	(166)
1-2	4-17	3	CARDIDTGDYRGADVLQMW	Influenza	H1N1	(166)
3-7	5-24	6	CARVSREEWATVDDPHDYYYMDVW	Influenza	H1N1	(166)
3-23	5-24	4	CAREEFTDTEMTITQGDFGYW	Influenza	H1N1	(166)
1-2	4-17	3	CARDIDSGDYRAADVFQIW	Influenza	H1N1	(166)
1-2	4-17	3	CARDIDSGDYRAADVFQIW	Influenza	H1N1	(166)
1-2	4-17	3	CARDIDSGDYRAADVFQIW	Influenza	H1N1	(166)
3-7	5-24	6	CARVSREEWATVDDPHDYYYMDVW	Influenza	H1N1	(166)
1-18	3-16	6	CAREGYDHLWGTYRFEAIDYYYTDVW	Influenza	H1N1	(166)
1-2	4-17	3	CARDIDFGDYRAADVFHIW	Influenza	H1N1	(166)
4-b	5-12	4	CARYIVSTINYFDDW	Influenza	H1N1	(166)
1-2	4-17	3	CARDFDYGDYRGSAFDIW	Influenza	H1N1	(166)
3-23	5-24	4	CAREEFTDTEMTINQGDFAYW	Influenza	H1N1	(166)
1-2	4-17	3	CARDFDYGDYRGSAFDIW	Influenza	H1N1	(166)
5-a	3-10	6	CTRDSFYDVDLSSFYMDVW	Influenza	H1N1	(166)
1-18	5-24	3	CARDRIDYVVYDAFDIW	Influenza	H1N1	(166)
1-18	4-17	3	CARRGDYGDYRGDAFDIW	Influenza	H1N1	(166)
3-21	5-24	3	CAKDRVRDGDNDWDSVDATYWGYGVFDTS	Influenza	H1N1	(166)
3-74	4-17	3	CVRDNDYGDYRGNAFDIW	Influenza	H1N1	(166)
3-21	5-24	3	CARDRVRDGDNYWDSVDATYWGYGAFDIC	Influenza	H1N1	(166)
4-59	3-3	6	CARAVSTLVSVDYYFYYIDVW	Influenza	H1N1	(166)
3-74	4-17	3	CARDHDYGDYRGNAYDIW	Influenza	H1N1	(166)
3-23	2-2	4	CAKDPRSSVPWVAYW	Influenza	H1N1	(166)
3-23	4-17	4	CANRMGLRPDYFDYW	Influenza	H1N1	(166)
3-23	6-13	4	CAKSPASSWYFDHW	Influenza	H1N1	(166)
4-39	4-23	5	CARHRVGTGPEVGDWFDPW	Influenza	H1N1	(166)
3-23	3-10	6	CRGWFGEGINGWDVW	Influenza	H1N1	(166)
3-30	2-2	6	CATLGGDIVLEPGTRSDYYYGLDVW	Influenza	H1N1	(166)
3-11	3-22	3	CARASAYYYDSSGRAAAFDIW	Influenza	H1N1	(166)
4-28			CATTIEVDITTEMGDCYFDSW	Influenza	H5N1	(258)
4-b			CARDGVLTYLDWLSKTHFDYW	Influenza	H5N1	(258)
1-69			CARAPYTYGTWWFDVW	Influenza	H5N1	(258)
1-18			CARDTTVTNEEINFYYGMDVW	Influenza	H1N1	(259)
1-18			CARDTEVTNEEINFYYGMDVW	Influenza	H1N1	(259)

1-18			CARDTEVTSEEINFYYGMDVW	Influenza	H1N1	(259)
1-18			CARDTTVTSEEINFYQGMDVW	Influenza	H1N1	(259)
1-46			CAREFGANGEDIYFYHGMDVW	Influenza	H1N1	(259)
2-5			CAHSIGGYDGEGIFYNHYGMDVW	Influenza	H1N1	(259)
3-7			CARDEWFGELGSSGMDVW	Influenza	H1N1	(259)
3-9			CAKDFAGEGHGSGSVDYW	Influenza	H1N1	(259)
3-21			CAKSATSYRDYLDRDFFYYALDVW	Influenza	H1N1	(259)
3-30			CARDHLNSEIVATITGFLDYW	Influenza	H1N1	(259)
3-30			CARDKLNSEMVATITGFLDYW	Influenza	H1N1	(259)
3-30			CARDKLNSEMVATITGFMDYW	Influenza	H1N1	(259)
3-30			CGRDKLNSDEVTTITGFLDYW	Influenza	H1N1	(259)
3-30			CARDNLNSELVATITGFLDHW	Influenza	H1N1	(259)
3-30			CGRDNLNSDEVATISGFLDYW	Influenza	H1N1	(259)
3-30			CARDYLNSEMVATITGFLDSW	Influenza	H1N1	(259)
3-30			CATEPSNTEDIRGIEGVFDYW	Influenza	H1N1	(259)
3-30			CARDAYSSGDTYYYGLDVW	Influenza	H1N1	(259)
3-43			CAKDRGTGEQIAVVTALIDYW	Influenza	H1N1	(259)
4-39			CARHGYGDYVGYFDYW	Influenza	H1N1	(259)
4-59			CARVLRWLGEEDADAFDIW	Influenza	H1N1	(259)
4-59			CARGFGMVGDTVDDLYNGMDVW	Influenza	H1N1	(259)
4-59			CARVQRPYGDYAAGAFDIW	Influenza	H1N1	(259)
4-59			CARVQRPYGDYITGAFDIW	Influenza	H1N1	(259)
5-51			CARRTWYYDGSGPDPSRDAFDIW	Influenza	H1N1	(259)
7-4-1			CARDLGNGEDIAVQPGTIGVDYW	Influenza	H1N1	(259)
7-4-1			CARDLGNGEDIAVQPGTTGVDYW	Influenza	H1N1	(259)
7-4-1			CARDLGNGEDIVVQPATIGVDYW	Influenza	H1N1	(259)
7-4-1			CASGTEVTTEEIYFYYGMDVW	Influenza	H1N1	(259)
7-4-1			CARGTEVTTEEINFYYGMDVW	Influenza	H1N1	(259)
1-3			CATAEKWLADYFYYFGMDVW	Influenza	H1N1	(259)
1-46			CARDREESLFAGAIYNYYYDMDVW	Influenza	H1N1	(259)
4-34			CARKGGAKLLYFDWLASAFDIW	Influenza	H1N1	(259)
1-69			CARGPNYYENFFDYW	Influenza	H1N1	(259)
1-69			CARGPNYYESYFDYW	Influenza	H1N1	(259)
1-69			CARGPNYYENYFDFW	Influenza	H1N1	(259)
1-69			CARGPNYYESYLDFW	Influenza	H1N1	(259)
1-69			CARGPNYFESYFDNW	Influenza	H1N1	(259)
1-69			CARGPNYYETYLDNW	Influenza	H1N1	(259)
1-69			CGRGPHYYESHLDYW	Influenza	H1N1	(259)
1-69			CAGGPHYYVSYFDSW	Influenza	H1N1	(259)
1-69			CARGNTYYSSYFDQW	Influenza	H1N1	(259)
1-69			CARGSTYYSSYFDQW	Influenza	H1N1	(259)
1-69			CATSGTYYVSYFDSW	Influenza	H1N1	(259)
1-69			CATSGTYYVSYLDSW	Influenza	H1N1	(259)
1-69			CATSGTYYVSFFDYW	Influenza	H1N1	(259)
1-69			CARSGSYYPDYFQYW	Influenza	H1N1	(259)
1-69			CARSPTYYPGALDMW	Influenza	H1N1	(259)
1-69			CARAPLIYNWYFDLW	Influenza	H1N1	(259)
1-69			CARAPLIYNWYYDLW	Influenza	H1N1	(259)
1-69			CAGHPTYHYGSAMDYW	Influenza	H1N1	(259)
1-69			CARHPTYYFGSAMEYW	Influenza	H1N1	(259)
1-69			CAGHPTYYYGSPMDYW	Influenza	H1N1	(259)
1-69			CAGHPMYHYGSAMDYW	Influenza	H1N1	(259)
1-69			CARHSGYHLIGYFDSW	Influenza	H1N1	(259)
1-69			CAKEEGYYYGSGPLDSW	Influenza	H1N1	(259)
1-69			CARNSGYHISGFYLDYW	Influenza	H1N1	(259)
1-69			CARSLGYHTQYNGMDVW	Influenza	H1N1	(259)
1-69			CASHPTYHFDKSGYRFDSW	Influenza	H1N1	(259)
1-69			CARSRGYSFGYGTDYFDYW	Influenza	H1N1	(259)
1-69			CARNYYGSGTYFNDAFDIW	Influenza	H1N1	(259)
1-69			CARYQSSDYYNSEYFQHW	Influenza	H1N1	(259)
4-31			CARGAPGVSGAMLDYYGMDVW	Influenza	H3N2	(260)
4-59			CARARPFYSDTSGHYFDYYGLDVW	Influenza	H3N2	(260)
4-59			CARARPFYSDSTGNYFDYYGLDVW	Influenza	H3N2	(260)
4-59			CARVGISITETGRVDWYFGVW	Influenza	H3N2	(260)
4-30-4			CARVAPTVRGVIADYYAMDVW	Influenza	H3N2	(260)
3-49			CTRTRGYGDYVDSYYYGIYVW	Influenza	H3N2	(260)
3-49			CTRTRGYGDYVDSYYYGIDVW	Influenza	H3N2	(260)
3-49			CTRTRGYGDYVDSYYYGIDAW	Influenza	H3N2	(260)
4-30-4			CVRLSPTSALDFYVMDVW	Influenza	H3N2	(260)
4-61			CARLRSYYETRGYSDYYAMDVW	Influenza	H3N2	(260)
4-31			CARLRPYSGYDFYGMDVW	Influenza	H3N2	(260)
4-31			CARLRPYSGYDFYAMDVW	Influenza	H3N2	(260)
4-31			CARLRPYRDYDFYAMDVW	Influenza	H3N2	(260)
4-31			CARLRPFIGYDFYGVDVW	Influenza	H3N2	(260)
4-31			CARLRPFIGYDFYGADVW	Influenza	H3N2	(260)
4-31			CARLRGFKILRGMTDEYGMDVW	Influenza	H3N2	(260)
4-30-4			CARLKDNVSSPGGNFYDYYAMDVW	Influenza	H3N2	(260)
4-30-4			CARLKDNVSSPGGGYYDYYAMDVW	Influenza	H3N2	(260)
4-30-4			CARLKDNVSSPGGGYYDHYAMDVW	Influenza	H3N2	(260)
4-30-4			CARLKDNLSSPGGGYHDYYAMDVW	Influenza	H3N2	(260)
4-31			CARLAGTAGDLGLDYYRVDVW	Influenza	H3N2	(260)
4-31			CARLAGTAGDLGLDYYRMDVW	Influenza	H3N2	(260)
4-31			CARLAGTAGDGGLDYYRMDVW	Influenza	H3N2	(260)
4-31			CARGRGFKYGSGMVDYYAMDVW	Influenza	H3N2	(260)
4-59			CASGARASYYDGDFDYW	Influenza	H3N2	(260)
1-2			CARERYHDSSGFLRDYYYYGLDVW	Influenza	H3N2	(260)
1-2			CARENYHDSSGNLRDYYYYGMDVW	Influenza	H3N2	(260)
1-2			CARENYHDSSGNFWDYYYYGMDVW	Influenza	H3N2	(260)
1-2			CARENYHDSSGFLRDYYYYGMDVW	Influenza	H3N2	(260)
4-34			CAREGYYGSGTYPVSYW	Influenza	H3N2	(260)
3-15			CTTDRNVAFEVW	Influenza	H3N2	(260)
3-15			CVTDRNVAFEVW	Influenza	H3N2	(260)
3-15			CVTDRNVAFEIW	Influenza	H3N2	(260)
3-15			CTTDRNVAFDIW	Influenza	H3N2	(260)
3-11	3	5	CVRNLGGRYSFGLYDRFDSW	Influenza	cross-reactive	(235)

1-2	3-9	5	CVRGEAVLQYFDWQINWFDTW	Influenza	cross-reactive	(235)
1-2	3-9	5	CVRGEAVLQHFDWQINWFDTW	Influenza	cross-reactive	(235)
1-8	3-16	5	CVREREHLTVSGMLSETVKLHRHFDPW	Influenza	cross-reactive	(235)
3-74	3-22	4	CVRDYLTSYYSSSGDDFDYW	Influenza	cross-reactive	(235)
3-33	3-10	4	CVRAPDHYGSGTVVEAFDYW	Influenza	cross-reactive	(235)
3-9	3-22	2	CVKDGGMGSSGYLYWHFDLW	Influenza	cross-reactive	(235)
3-15	3-10	6	CTTSMVRGVIRTADSSYYGMDVW	Influenza	cross-reactive	(235)
4-59	3	3	CTRDRTGINSLNALDVW	Influenza	cross-reactive	(235)
3-9	3-9	4	CTRDLYRGEAYYFDSW	Influenza	cross-reactive	(235)
3-9	2-8	5	CTKDRLLRAARTRGPPRGFDPW	Influenza	cross-reactive	(235)
4-31	2-15	6	CSRPGDSSLYYGMDVW	Influenza	cross-reactive	(235)
4-39	3-22	4	CSRHRDYSSGWGFDQW	Influenza	cross-reactive	(235)
3-49	2-15	4	CSGRYCSGSRCFDYW	Influenza	cross-reactive	(235)
1-69	3-10	4	CATSPTYYISHLDSW	Influenza	cross-reactive	(235)
1-69	3-22	4	CATSPTYYESSLDFW	Influenza	cross-reactive	(235)
1-69	3-10	4	CATSGTYYISHFDSW	Influenza	cross-reactive	(235)
1-69	3-9	4	CATSATYYISYFDSW	Influenza	cross-reactive	(235)
1-69	3-10	4	CATSATYYISHFDYW	Influenza	cross-reactive	(235)
3-30-3	3-3	6	CATGSGYYKRDYDYGMDVW	Influenza	cross-reactive	(235)
1-24	3-9	6	CATGANILTGPASRPYSYYGLDVW	Influenza	cross-reactive	(235)
4-31	1-26	4	CATGADGEYSGDGLLAGSDYW	Influenza	cross-reactive	(235)
1-69	3-3	6	CATDGRQIFGASNRSGMDVW	Influenza	cross-reactive	(235)
1-69	6-19	4	CASRYSSGWYYFDYW	Influenza	cross-reactive	(235)
1-69	2-15	1	CASPPAKTVVVSHAEYFQYW	Influenza	cross-reactive	(235)
1-69	1-26	6	CASGRDFYYYGMDVW	Influenza	cross-reactive	(235)
1-2	2-8	5	CARVTRYWSPFFGSW	Influenza	cross-reactive	(235)
4-61	3	4	CARVPRTGLDAVRYYFDYW	Influenza	cross-reactive	(235)
3-49	1-20	3	CARVGRNNWNARHAFDVW	Influenza	cross-reactive	(235)
1-69	5-24	4	CARTKSTIYPPYYFDYW	Influenza	cross-reactive	(235)
1-46	1-26	4	CARSWGIVGVTLHTSFDYW	Influenza	cross-reactive	(235)
1-69	3-16	4	CARSLASLTGDYW	Influenza	cross-reactive	(235)
4-4	3-10	4	CARSIGSGTYYNPRRVFDYW	Influenza	cross-reactive	(235)
1-69	3-16	6	CARRYGSTSGVNTGGVSGFYYYGMDVW	Influenza	cross-reactive	(235)
4-30-4	3-3	5	CARRPNIFGYNWFDPW	Influenza	cross-reactive	(235)
4-39	1-1	4	CARRIRLPHYFDYW	Influenza	cross-reactive	(235)
1-69	1-26	4	CARNREGSFDFW	Influenza	cross-reactive	(235)
4-30-4	3-16	3	CARLWGTGYDYVWGSFRYHAFDIW	Influenza	cross-reactive	(235)
4-39	6-13	4	CARLVNQMAAAGSLPGKRRFDSW	Influenza	cross-reactive	(235)
4-30-2	3-16	6	CARILSADYYYGMDVW	Influenza	cross-reactive	(235)
4-59	2	6	CARIITEAPGFRWGPKERYNGMDVW	Influenza	cross-reactive	(235)
4-34	5-12	5	CARHYGGYDWFDPW	Influenza	cross-reactive	(235)
1-69	5-12	4	CARHSGYHFQSYFDNW	Influenza	cross-reactive	(235)
5-51	3-3	4	CARHEGFYDRSGYTKIPDFW	Influenza	cross-reactive	(235)
1-69	6-25	4	CARGSGYHVTDYFDLW	Influenza	cross-reactive	(235)
1-69	3-3	4	CARGSGYHVRDYFDLW	Influenza	cross-reactive	(235)
1-69	1-26	4	CARGRNYYHTSLEYW	Influenza	cross-reactive	(235)
1-69	5-18	5	CARGRGYHFGDLVSW	Influenza	cross-reactive	(235)
1-46	3-22	4	CARGRFHYDSSGFYGLSDYW	Influenza	cross-reactive	(235)
3-33	6-19	2	CARGQQWHTPTSVNWYFDLW	Influenza	cross-reactive	(235)
6-1	3-10	2	CARGQHLWMTYWYFDVW	Influenza	cross-reactive	(235)
1-69	2-15	5	CARGQGYCISPNCPNWFDPW	Influenza	cross-reactive	(235)
1-69	2-15	5	CARGQGYCISASCPNYFDPW	Influenza	cross-reactive	(235)
1-69	2-15	5	CARGQGYCISANCPNWFDPW	Influenza	cross-reactive	(235)
1-69	3-10	4	CARGPHYYLNYFDYW	Influenza	cross-reactive	(235)
3-33	2	4	CARGPFGLLDYW	Influenza	cross-reactive	(235)
1-3	3-3	4	CARGNPNHDFWSSYYPGSFDCW	Influenza	cross-reactive	(235)
1-3	3-3	4	CARGNANSDFWSSYHPGTFDYW	Influenza	cross-reactive	(235)
1-69	1-26	4	CARGKNYYEDYFDTW	Influenza	cross-reactive	(235)
1-3	3-10	4	CARGHYYGSGSYFPPWAHW	Influenza	cross-reactive	(235)
1-69	1-7	4	CARGGNYYHTSLDYW	Influenza	cross-reactive	(235)
1-69	2	4	CARGGKYYHTGLDYW	Influenza	cross-reactive	(235)
1-46	2-8	4	CARGGEIVQFREHGDLDYW	Influenza	cross-reactive	(235)
1-2	3-9	5	CARGEAVLQYFDWQINWFDPW	Influenza	cross-reactive	(235)
1-2	3-9	5	CARGEAVFQYFDWQVNWFDTW	Influenza	cross-reactive	(235)
1-3	3-3	6	CARGAPTHDFWSAYYPYGMDVW	Influenza	cross-reactive	(235)
1-3	3-3	6	CARGAPTDDFWSAYYPNGMDVW	Influenza	cross-reactive	(235)
4-4	3-3	4	CARFNWYYDFLPGQSPDDFW	Influenza	cross-reactive	(235)
1-18	3-10	6	CARFGLGTTIVEVPYYYYGMDVW	Influenza	cross-reactive	(235)
1-46	3-22	5	CAREGTYYYDSSGLNWFDSW	Influenza	cross-reactive	(235)
1-18	3-3	5	CAREEGTYYDFWSANNWFDPW	Influenza	cross-reactive	(235)
4-59	1-14	4	CAREDGTVTSGGGHYFDLW	Influenza	cross-reactive	(235)
3-7	1-1	4	CAREDEWNAGADW	Influenza	cross-reactive	(235)
3-21	3-22	3	CAREANYQRFTTYEPFDVW	Influenza	cross-reactive	(235)
1-69	1-26	4	CAREAGYYSGSYYELW	Influenza	cross-reactive	(235)
1-69	3-3	4	CAREAAYYTGTYYEQW	Influenza	cross-reactive	(235)
1-3	2-15	4	CARDYEHCTGGSCLGYW	Influenza	cross-reactive	(235)
3-48	3-16	4	CARDSKGWRRVTRGSYDYW	Influenza	cross-reactive	(235)
4-59	3	3	CARDRTGINSLNALDVW	Influenza	cross-reactive	(235)
1-3	3-3	6	CARDRMELRFLEWLNYGMDVW	Influenza	cross-reactive	(235)
4-31	3-16	5	CARDRGYYDNIWGYNRPTQNWFDPW	Influenza	cross-reactive	(235)
3-30-3	1-1	4	CARDRATADTTFEFW	Influenza	cross-reactive	(235)
1-69	3-22	3	CARDPYFYDSSGYYTYSLDVW	Influenza	cross-reactive	(235)
3-30-3	2-15	4	CARDMRDCVTTRSCYRGPFDYW	Influenza	cross-reactive	(235)
4-59	3-10	4	CARDLTTLTMFQGVSYFDYW	Influenza	cross-reactive	(235)
4-30	2-15	6	CARDHLELGFCSGGSCYSGKLNYYYYGMW	Influenza	cross-reactive	(235)
3-7	1-7	4	CARDGRLGWITGTTWSW	Influenza	cross-reactive	(235)
1-69	5-24	4	CARDGFSKSWLDW	Influenza	cross-reactive	(235)
1-69	1-26	4	CARDAGYYSGSYYEQW	Influenza	cross-reactive	(235)
4-61	5-18	3	CARASPATPPSSRRYTGKVAPHVYDFW	Influenza	cross-reactive	(235)
4-61	5-12	3	CARASPAIPPSSRRYTGKVAPHVFDVW	Influenza	cross-reactive	(235)
4-61	3-16	6	CARAPDRRFTMNMFGGEFAPTGMDVW	Influenza	cross-reactive	(235)
4-61	3-16	6	CARAPDRRFTMNIFGGDFAPTGMDVW	Influenza	cross-reactive	(235)
4-61	3-16	6	CARAPDRRFSMTTFGGGSFAPTGLDVW	Influenza	cross-reactive	(235)
4-61	3-16	6	CARAPDRRFSMITFGGGSFAPTGMDVW	Influenza	cross-reactive	(235)
4-61	3-3	4	CARALRTGVEAVRYYFDCW	Influenza	cross-reactive	(235)

1-46	2-15	5	CARALGYCSGGTCSGASSVWFDTW	Influenza	cross-reactive	(235)
3-48	5-12	6	CARAITPGIRHALDVW	Influenza	cross-reactive	(235)
1-46	3-10	6	CARAGPFFFGSGTYYRSSYGLDVW	Influenza	cross-reactive	(235)
4-30-2	1-26	6	CARAGGGYSGRSSHYYSYGMDVW	Influenza	cross-reactive	(235)
1-69	3-10	4	CARAENYHGSGSYYRHW	Influenza	cross-reactive	(235)
3-30	1-26	3	CANDDVVGGSIRFGSSSRPPVGLKAFEIW	Influenza	cross-reactive	(235)
2-5	3-10	4	CALNVPPRPFDHW	Influenza	cross-reactive	(235)
1-46	4-11	4	CAKVQNDYSNYGPFDYW	Influenza	cross-reactive	(235)
3-30	2-2	4	CAKVASRYCITTSCYRSGRTVDYW	Influenza	cross-reactive	(235)
1-69	6-19	4	CAKRYSSGWYSFDYW	Influenza	cross-reactive	(235)
4-4	3-10	6	CAKLLVWFGTQIPNYYYAMDVW	Influenza	cross-reactive	(235)
1-69	3-3	4	CAKGSGYHVRDYFDYW	Influenza	cross-reactive	(235)
1-69	5-12	4	CAKGSGYHVRDHFDDW	Influenza	cross-reactive	(235)
3-30	3-9	4	CAKERPLRLLRYFDWLSGGANDYW	Influenza	cross-reactive	(235)
3-30	3-9	4	CAKERPLRLLRFFDWLSGGANDYW	Influenza	cross-reactive	(235)
3-30	3-9	6	CAKEGGARILRYFDWLAHDALDVW	Influenza	cross-reactive	(235)
3-30	3-9	3	CAKEEKARILRYFDWLSHAAFDIW	Influenza	cross-reactive	(235)
3-9	6-19	4	CAKDVRSAGLAVAGFESW	Influenza	cross-reactive	(235)
3-30	3-9	5	CAKDSQLRSLLYFEWLSQGYFDPW	Influenza	cross-reactive	(235)
3-30	3-9	4	CAKDSQLRSLLYFDWLSQGYFDHW	Influenza	cross-reactive	(235)
3-30	2-15	4	CAKDQTVVSVAAALFDYW	Influenza	cross-reactive	(235)
3-23	2-21	6	CAKDKWRCDGDCSSNYYGMDVW	Influenza	cross-reactive	(235)
1-69	3-22	6	CAKAQYYDSPRDAADFYFYAMDVW	Influenza	cross-reactive	(235)
3-23	1-1	4	CAKAGAGYYPRLERRASRHFYFDSW	Influenza	cross-reactive	(235)
3-9	5-24	4	CAKAFSGGDAYNWGLWDYFDYW	Influenza	cross-reactive	(235)
1-69	1-26	4	CAGSGTYYVSRFDYW	Influenza	cross-reactive	(235)
1-69	1-26	4	CAGSGTYYVSRFDSW	Influenza	cross-reactive	(235)
1-69	3-10	4	CAGSGTYFVSRFDYW	Influenza	cross-reactive	(235)
1-69	3-10	4	CAGSETYYVSRFDHW	Influenza	cross-reactive	(235)
1-69	5	4	CAGSETYFVSRFDYW	Influenza	cross-reactive	(235)
1-69	3-10	4	CAGSATYYVSRFDYW	Influenza	cross-reactive	(235)
1-69	3-10	4	CAGSATYYESRFDYW	Influenza	cross-reactive	(235)
1-69	3-16	4	CAGSATYYESRFDYW	Influenza	cross-reactive	(235)
1-69	1-26	4	CAESGTYFVSRFDSW	Influenza	cross-reactive	(235)
1-69	1-26	4	CAESGTYFVSRFDNW	Influenza	cross-reactive	(235)
3-30	3-9	4	CAKDSQLRSLLYFEWLSQGYFDPW	Influenza	cross-reactive	(235)
			CARDVHNYDFLTGYPLHLYGMDVW	Influenza	H5N1	(261)
			CARDFLSGPMEMPGGYYGLDVW	Influenza	H5N1	(261)
			CARSGDGYNYYFPLW	Influenza	H5N1	(261)
4-39	6-13	5	CAKHESDSSSWHTGWNWFDPW	Influenza	H1N1	(236)
4-39	3-22	5	CARHESDSSSWHTGWNWFDPW	Influenza	H1N1	(236)
1-69	3-16	4	CARGKKYYHDTLDYW	Influenza	H1N1	(236)
1-69	3-3	6	CARSITNLYYYYMDVW	Influenza	H1N1	(236)
1-69	1-14	6	CARGPKYYHSYMDVW	Influenza	H1N1	(236)
1-69	1-14	5	CARSGTTKTRYNWFDPW	Influenza	H1N1	(236)
1-69	3-22	4	CARHDSSGYHPLDYW	Influenza	H1N1	(236)
1-69	5-18	5	CARPNTYGYILPVYW	Influenza	H1N1	(236)
4-39	3-3	4	CARHHVTELRVLEWLPKSDYW	Influenza	H1N1	(236)
1-69	3-10	5	CARVCSFYGSGSYYNVFCYW	Influenza	H1N1	(236)
1-69	2-2	6	CAKHMGYQVRETMDVW	Influenza	cross-reactive	(262)
1-69	3-10	6	CARGPKYYSEYMDVW	Influenza	cross-reactive	(262)
1-69	3-22	1	CARSSGYYPAYLPHW	Influenza	cross-reactive	(262)
1-69	3-22	4	CARGSGYHISTPFDNW	Influenza	cross-reactive	(262)
1-69	5-12	4	CARGSGYTTRNYFDYW	Influenza	cross-reactive	(262)
1-69	3-22	4	CARGNYYYESSLDYW	Influenza	cross-reactive	(262)
1-69	NA	6	CARGPHYYSSYMDVW	Influenza	cross-reactive	(262)
1-69	NA	6	CARGPTYYYSYMDVW	Influenza	cross-reactive	(262)
1-69	3-10	5	CARLNYHDSGTYYNAPRGWFDPW	Influenza	cross-reactive	(262)
1-69	3-22	4	CARSSNYYDSVYDYW	Influenza	cross-reactive	(262)
1-69	3-10	4	CARIPHYNFGSGSYFDYW	Influenza	cross-reactive	(262)
1-69	3-10	6	CARDSDAYYYGSGGMDVW	Influenza	cross-reactive	(263)
1-69	1-14	3	CARTLSSYQPNNDAFAIW	Influenza	cross-reactive	(263)
1-69	3-22	4	CARSSGYHFRSHFDSW	Influenza	cross-reactive	(263)
1-69	3-22	4	CARGLYYYESSLDYW	Influenza	cross-reactive	(263)
1-69	1-14	3	CARTLSSYQPNNDAFAIW	Influenza	cross-reactive	(263)
1-69	2-8	4	CARSPSYICSGGTCVFDHW	Influenza	cross-reactive	(263)
1-69	2	3	CAREPGYYVGKNGFDVW	Influenza	cross-reactive	(263)
1-69	3-10	3	CARGPYYYGNSHLDFW	Influenza	cross-reactive	(264)
1-69	1-7	3	CATSAGGIVNYFLLFDIW	Influenza	cross-reactive	(264)
1-69	3-10	3	CATSAGGIVNYYLSFNIW	Influenza	cross-reactive	(264)
1-69	1-26	4	CASSSGSYYGDYFDYW	Influenza	cross-reactive	(264)
4-31		4	CARGMLMDTDMIAFDQW	Influenza	06-09 seasonal	(265)
4-31		4	CARGMLLDTDMIAFDHW	Influenza	06-09 seasonal	(265)
3-49		3	CSKGGEGDAFDIW	Influenza	06-09 seasonal	(265)
3-15		4	CATDRGDYYGSGTITGFLWVYW	Influenza	06-09 seasonal	(265)
1-69		6	CARSTYSGNFPRYYYYHGLDVW	Influenza	06-09 seasonal	(265)
3-66		6	CARDRYSDGLGYYYYYYYGMDVW	Influenza	06-09 seasonal	(265)
3-66		6	CARDRYFDSAGRYYYYYYGMDVW	Influenza	06-09 seasonal	(265)
1-69		3	CARYDYGGNSIFINPFDMW	Influenza	06-09 seasonal	(265)
4-39		3	CARVGGSAWKGGGADAFDFW	Influenza	06-09 seasonal	(265)
4-39		6	CARAREGYNYLYGMDVW	Influenza	06-09 seasonal	(265)
4-59		4	CASGSDNGYMERRAYYLDNW	Influenza	06-09 seasonal	(265)
4-39		6	CARQRAGWHYLYGMDVW	Influenza	06-09 seasonal	(265)
3-53		6	CARDRYYDASGYYYYYYGMDVW	Influenza	06-09 seasonal	(265)
1-02		4	CARSYDMRVVTVVSTGFDYW	Influenza	06-09 seasonal	(265)
3-66		6	CARDRYYDASGYYYYYYGMDVW	Influenza	06-09 seasonal	(265)
1-69		4	CAGTRGYYGDYEPFDYW	Influenza	06-09 seasonal	(265)
3-48		3	CARGKAGGSYLSDDALDIW	Influenza	06-09 seasonal	(265)
4-34		3	CATGRYSPNYFGSGSLAHAFDIW	Influenza	06-09 seasonal	(265)
3-21		4	CARDWGVKYADPEVFSDW	Influenza	06-09 seasonal	(265)
3-09		6	CAKDRIASSSLDYYYAYGMDVW	Influenza	06-09 seasonal	(265)
1-69		6	CASTKRGESGRGLDVW	Influenza	06-09 seasonal	(265)
3-09		6	CAKDVAASSSMDYYYYSGMDVW	Influenza	06-09 seasonal	(265)
3-09		6	CAKDRSSSSSMDDYYYSGMDVW	Influenza	06-09 seasonal	(265)
3-11		4	CVRCDSYSGYELDYW	Influenza	06-09 seasonal	(265)

3-15	3	CTNPTYGYYTEGGPADGYPFDLW	Influenza	06-09 seasonal	(265)
3-30	4	CAKEWDNSAYSLDYW	Influenza	06-09 seasonal	(265)
3-09	4	CAKDPTGPIDSVGDFYFDSW	Influenza	06-09 seasonal	(265)
3-15	3	CTNPTYGNYAEGGPAEGYPFDLW	Influenza	06-09 seasonal	(265)
3-20	3	CARGHEVSTSMDQHSDAFDVW	Influenza	06-09 seasonal	(265)
3-49	6	CARDQLELRGVIVRVGMDVW	Influenza	06-09 seasonal	(265)
3-15	4	CTTVSGYCGGGICAFFDNW	Influenza	06-09 seasonal	(265)
3-49	6	CARDQLDLREVIVRVGMDVW	Influenza	06-09 seasonal	(265)
4-34	5	CARGMSYYGSGSYYFFYNWFDW	Influenza	06-09 seasonal	(265)
3-15	4	CAKGYYYDTSGPYFFDSW	Influenza	06-09 seasonal	(265)
3-49	6	CARDQLDLREVIVRVGMDVW	Influenza	06-09 seasonal	(265)
3-48	4	CARDPFGYTDGHPFDSW	Influenza	06-09 seasonal	(265)
3-48	4	CARDPFGYTDGHLFDYW	Influenza	06-09 seasonal	(265)
4-39	6	CASGIGVLRYFDWFHAGMDVW	Influenza	06-09 seasonal	(265)
3-09	4	CAKDIAAGTHSSGYYNYW	Influenza	06-09 seasonal	(265)
1-02	3	CARVSVRYSDDAFDLW	Influenza	06-09 seasonal	(265)
3-23	4	CAKLQGYTSSWFGTMDYW	Influenza	06-09 seasonal	(265)
3-30	4	CAKVALGYFEPDRFPISYYFDSW	Influenza	06-09 seasonal	(265)
3-23	5	CAKSNEFVAVTGILDRW	Influenza	06-09 seasonal	(265)
1-69	3	CVLRGDEGAFHLW	Influenza	06-09 seasonal	(265)
3-09	4	CVKDIAAGTHSSGYYNYW	Influenza	06-09 seasonal	(265)
4-39	6	CARWAAGYYYFYGLDVW	Influenza	06-09 seasonal	(265)
3-30	6	CTKVATWELDHFYGLDVW	Influenza	06-09 seasonal	(265)
3-30-3	4	CARDESSGLIDYW	Influenza	06-09 seasonal	(265)
3-23	4	CAKLQGYTSSWYGTMDYW	Influenza	06-09 seasonal	(265)
4-39	6	CARRAGGYKYYYGLDVW	Influenza	06-09 seasonal	(265)
1-69	6	CARDPMPREVPNTYYSALDVW	Influenza	06-09 seasonal	(265)
3-15	6	CATSIAFLSGYYNYFYTGLDLW	Influenza	06-09 seasonal	(265)
3-30-3	4	CARAHQVAAAGTWTFDYW	Influenza	06-09 seasonal	(265)
1-18	6	CARHYLQGVVVVDPYSHAMDVW	Influenza	06-09 seasonal	(265)
3-48	6	CARALFVVVPVTGGMDVW	Influenza	06-09 seasonal	(265)
4-59	5	CARGKTIFGVVRNWFDSW	Influenza	06-09 seasonal	(265)
5-51	3	CARTTCTSSTCYSLEANDAFDIW	Influenza	06-09 seasonal	(265)
4-59	6	CARERHYYETMDRIDYGMDVW	Influenza	06-09 seasonal	(265)
4-59	6	CARERSYYETMDRIDYGMDVW	Influenza	06-09 seasonal	(265)
3-53	6	CARSGVTIFGVVRPYGMDVW	Influenza	06-09 seasonal	(265)
4-59	6	CARERSYYENMDKIDYGVDVW	Influenza	06-09 seasonal	(265)
4-30-4	4	CATSPSNYDNWNGFRLHFDSW	Influenza	06-09 seasonal	(265)
4-61	4	CARTYYYDSSDYSQVLDYFDSW	Influenza	06-09 seasonal	(265)
4-B	4	CARVLYDFYSGTFFDAW	Influenza	06-09 seasonal	(265)
4-30-4	6	CAGEQQHPQINYYALDVW	Influenza	06-09 seasonal	(265)
4-30-4	6	CAGEQHDLQINYYALDVW	Influenza	06-09 seasonal	(265)
4-30-4	6	CAGEQQYPQINYYSLDVW	Influenza	06-09 seasonal	(265)
1-02	4	CARDPSLDTSGYFDSW	Influenza	06-09 seasonal	(265)
1-69	6	CSRVWEWRRLNRMDYCGMDVW	Influenza	06-09 seasonal	(265)
3-23	3	CAKDALFYYGSGEEAFDVW	Influenza	06-09 seasonal	(265)
3-73	6	CSRPSYSSAWYLPNRMDVW	Influenza	06-09 seasonal	(265)
3-09	4	CAAAFYGDPLKADYW	Influenza	06-09 seasonal	(265)
3-30-3	6	CAKDGGGYFYYGMDVW	Influenza	06-09 seasonal	(265)
3-23	4	CAKDQVRVIPGTGYLDHW	Influenza	06-09 seasonal	(265)
3-30	6	CAKDPGWRTQLPLNPDYYYGMDVW	Influenza	06-09 seasonal	(265)
4-39	6	CAHSGPYTIMDNFSYYALDVW	Influenza	06-09 seasonal	(265)
4-39	6	CARRAGGYLYFYGMDVW	Influenza	06-09 seasonal	(265)
4-31	4	CARAPFNRRLVTLFGTYYFDTW	Influenza	06-09 seasonal	(265)
3-53	6	CAREATTVTTYGYYYYGMDVW	Influenza	06-09 seasonal	(265)
6-01	4	CARDRGYSGFGSFDYW	Influenza	06-09 seasonal	(265)
3-11	4	CARDRGEGTTTFDYW	Influenza	06-09 seasonal	(265)
4-39	6	CARRRDGYKYYYGMDVW	Influenza	06-09 seasonal	(265)
1-46	3	CARDLLELLVYTYDSAGSSDAFDIW	Influenza	06-09 seasonal	(265)
1-46	3	CARDLLELLVYTYDSAGSLDAFDIW	Influenza	06-09 seasonal	(265)
3-64	4	CVIRRGAATEYW	Influenza	06-09 seasonal	(265)
3-53	5	CVALQTYSYDTSGFTADGYYFDPW	Influenza	06-09 seasonal	(265)
3-23	4	CARRMIYGSGSFDSW	Influenza	06-09 seasonal	(265)
3-73	5	CIRPDDHANPRTW	Influenza	06-09 seasonal	(265)
3-48	3	CAGDPDTPFIRVFEMW	Influenza	06-09 seasonal	(265)
4-59	3	CARDRTGMTDINAFDIW	Influenza	06-09 seasonal	(265)
3-23	3	CAKLRSSGWYAPIDIW	Influenza	06-09 seasonal	(265)
3-30	5	CARDGNVLNGIDPW	Influenza	06-09 seasonal	(265)
4-B	4	CARRGSGSYFSFDYW	Influenza	06-09 seasonal	(265)
3-48	4	CAREEVVVAGTPFDSW	Influenza	06-09 seasonal	(265)
3-15	6	CATGGGSASGMYYYNGLDVW	Influenza	06-09 seasonal	(265)
3-48	4	CANMRAWFWYDPADDIGHW	Influenza	06-09 seasonal	(265)
4-39	6	CARDRGGEEYYYGMDVW	Influenza	06-09 seasonal	(265)
4-39	6	CAKINSGYEYYYGMDVW	Influenza	06-09 seasonal	(265)
4-59	5	CARSAQCTGDVCFGGPSWFDPW	Influenza	06-09 seasonal	(265)
3-48	4	CANMRAWFYYDPADDIGHW	Influenza	06-09 seasonal	(265)
3-48	4	CVREAVIVDGMPFEYW	Influenza	06-09 seasonal	(265)
4-59	5	CARSAHCTDDVCFGGPSWFDPW	Influenza	06-09 seasonal	(265)
1-18	3	CARGFDYGDYRGRAFDVW	Influenza	06-09 seasonal	(265)
4-39	6	CARVPDGYKYYYGMDVW	Influenza	06-09 seasonal	(265)
3-15	4	CTTASNPYNWNGPYFDYW	Influenza	06-09 seasonal	(265)
3-20	3	CAREMRDPYDAFDLW	Influenza	06-09 seasonal	(265)
1-69	4	CARRYHSAPGYYFDSW	Influenza	06-09 seasonal	(265)
1-18	3	CTRGFDYGDYRGRAFDVW	Influenza	06-09 seasonal	(265)
3-23	3	CAKNRGWQLGPDAFDIW	Influenza	06-09 seasonal	(265)
4-39	6	CARRCGGYQYYYGMDVW	Influenza	06-09 seasonal	(265)
3-48	4	CANMRAWFWYDPADDIGHW	Influenza	06-09 seasonal	(265)
4-59	4	CARFWREETYGSCFDYW	Influenza	06-09 seasonal	(265)
3-30	5	CVKDPGDPWSGWFDPW	Influenza	06-09 seasonal	(265)
4-30-2	6	CARVGDNGSSYGMDVW	Influenza	06-09 seasonal	(265)
3-48	4	CANMRAWFWYDPADDIGHW	Influenza	06-09 seasonal	(265)
4-39	6	CARQQEGYYYYYGMDVW	Influenza	06-09 seasonal	(265)
1-02	3	CARAVYYDLTSDFCVDCYDAFDVW	Influenza	06-09 seasonal	(265)
3-23	3	CVKNRGWQLGPDAFDSW	Influenza	06-09 seasonal	(265)
3-23	3	CARNRGWQLGPDAFAIW	Influenza	06-09 seasonal	(265)

3-48		4	CVREEVIVGGIPFDYW	Influenza	06-09 seasonal	(265)
3-23		3	CVKNRGWQLGPDAFDIW	Influenza	06-09 seasonal	(265)
3-23		3	CAKNRGWQLGPDGFDIW	Influenza	06-09 seasonal	(265)
4-B		5	CARQSIYTGNHFPKFDPW	Influenza	06-09 seasonal	(265)
4-59		6	CARTMDTSGWQPYYYGMDVW	Influenza	06-09 seasonal	(265)
3-30-3		4	CVRGQDSSHYFDAW	Influenza	06-09 seasonal	(265)
3-23		3	CAKNRGWQLGPDAFDIW	Influenza	06-09 seasonal	(265)
4-59		3	CARFHYYDNTGYYIDAFDIW	Influenza	06-09 seasonal	(265)
4-59		3	CARFHYYDSSGFFIDAFDIW	Influenza	06-09 seasonal	(265)
4-59		3	CARFHFYDSSGYYNDAFDIW	Influenza	06-09 seasonal	(265)
4-59		3	CARFHYYDSSGYFIDAFDIW	Influenza	06-09 seasonal	(265)
1-46		3	CARGGVGPTWGYDAFNMW	Influenza	06-09 seasonal	(265)
4-59		3	CARFHFYDSSGFLNDAFDVW	Influenza	06-09 seasonal	(265)
1-18		6	CARDLVQGAVSHYTGLDVW	Influenza	06-09 seasonal	(265)
3-48		4	CARVHVPYGDYIRTVLDYW	Influenza	06-09 seasonal	(265)
3-30		4	CARDYRHGLIDFW	Influenza	06-09 seasonal	(265)
3-74		4	CAREPYHPPQFDYW	Influenza	06-09 seasonal	(265)
4-39		6	CSRLQPYTNYGRSHYYFYALDVW	Influenza	06-09 seasonal	(265)
1-18		4	CARETLTYCSGGSCYARLSGNFDHW	Influenza	06-09 seasonal	(265)
7-81		6	CARRDGNYVWYFDVW	Influenza	06-09 seasonal	(265)
3-23		4	CAKGVGRIFMVRGFYFDYW	Influenza	06-09 seasonal	(265)
3-23		3	CAKDLPQYGSTYWRGSFDMW	Influenza	06-09 seasonal	(265)
3-30		6	CAKLPQRLDEGDYYFDMEVW	Influenza	06-09 seasonal	(265)
3-21		4	CARERGYSYGLDYW	Influenza	06-09 seasonal	(265)
3-23		5	CAKFPTIFGVVDSW	Influenza	06-09 seasonal	(265)
3-74		4	CARETLRGYSGYVSFEYW	Influenza	06-09 seasonal	(265)
4-39		5	CARRQYTNSPFDPW	Influenza	06-09 seasonal	(265)
3-30		6	CAKDRRRDSLPRYFDWLVVYNGMDVW	Influenza	06-09 seasonal	(265)
4-61		6	CARGARDFPRYVNYHMDVW	Influenza	06-09 seasonal	(265)
3-07		3	CARGGRGAYYDFWSAYIDIW	Influenza	06-09 seasonal	(265)
4-30-4		6	CAKTKTMSYFYTYMDVW	Influenza	06-09 seasonal	(265)
3-21		4	CARDRSSGWDPFDSW	Influenza	06-09 seasonal	(265)
1-03		6	CSRDSLIRSLEWSNPVRNFHFSYMDVW	Influenza	06-09 seasonal	(265)
3-30		6	CARVASMILVDIKYYFDYW	Influenza	06-09 seasonal	(265)
3-30		6	CAKVLYDILTGYRYGMDVW	Influenza	06-09 seasonal	(265)
3-09		4	CVKGRGGVYTGYGPRYLDYW	Influenza	06-09 seasonal	(265)
3-09		4	CAKAAKPGPARAPFDSW	Influenza	06-09 seasonal	(265)
4-61		4	CARTNYYDSSGYW	Influenza	06-09 seasonal	(265)
3-09		3	CTKPYLMRSTFDIW	Influenza	06-09 seasonal	(265)
3-15		4	CVTDFSGRRPYW	Influenza	06-09 seasonal	(265)
1-18		5	CARDRGSILGSDPWFDPW	Influenza	06-09 seasonal	(265)
3-30-3		6	CATRSVGATGYLSWGPKDYSYNDVDVW	Influenza	06-09 seasonal	(265)
3-23		4	CAKKSPNDYDHYW	Influenza	06-09 seasonal	(265)
1-69		6	CAKVRTPGPPTLPEMEGLFGDDYLADLCYGMDVW	Influenza	06-09 seasonal	(265)
4-34		4	CARGVTMLRGARPDVRQPKYDYW	Influenza	06-09 seasonal	(265)
3-09		3	CAKGENNWNDPDAFDIW	Influenza	06-09 seasonal	(265)
3-33		6	CARDGFSDYDTYYNGIDVW	Influenza	06-09 seasonal	(265)
4-61		4	CARDNVVSLGRFGEFALYIDYW	Influenza	06-09 seasonal	(265)
1-46		4	CATVPTGAGDYW	Influenza	06-09 seasonal	(265)
4-34		4	CARGYDYVWGSYRRRPYYFDYW	Influenza	06-09 seasonal	(265)
3-30		4	CAKDLVRYFEWVGADW	Influenza	06-09 seasonal	(265)
4-59		6	CARLGFGEVFSFKRYYYYGLDVW	Influenza	06-09 seasonal	(265)
4-61		4	CARGPSLVDFW	Influenza	06-09 seasonal	(265)
3-23		6	CVKDRGTMFRGAVYGMDVW	Influenza	06-09 seasonal	(265)
3-15		4	CTVDDSAYISQIDYW	Influenza	06-09 seasonal	(265)
4-59		4	CARVVTVAGLVNW	Influenza	06-09 seasonal	(265)
5-51		4	CVRGASSNLFDYW	Influenza	06-09 seasonal	(265)
1-02		4	CVREGPVRGLTPPDYW	Influenza	06-09 seasonal	(265)
3-23		6	CVKRGWFGELFGNYGMDVW	Influenza	06-09 seasonal	(265)
1-69		4	CSRDSNGGGAYDSW	Influenza	06-09 seasonal	(265)
3-30		4	CAKDLLPGWAAGGRYYFDNW	Influenza	06-09 seasonal	(265)
3-74		3	CAKSYKMNYSAFDIW	Influenza	06-09 seasonal	(265)
3-11		4	CARGSMRFSEWSPLGFW	Influenza	06-09 seasonal	(265)
3-9	1-7	6	CAKAPIIGPKYYFYMDVW	TT	NA	(148)
3-43	3-10	4	CGKSYDYIRENLDSW	TT	NA	(148)
1-3	2-2	4	CAKDRVRVVQAATTLDFW	TT	NA	(148)
2-70	2-2	4	CARGVVPAGIPFDFW	TT	NA	(148)
4-59	2-2	6	CARLHPTCASTRCPENYGMDVW	TT	NA	(148)
1-18	3-10	4	CARDYFHSGSQYFFDYW	TT	NA	(148)
4-30-4	3-10	4	CARARNYGFPHFFDFW	TT	NA	(148)
3-21	3-3	4	CARKGMGHYFDFW	TT	NA	(148)
4-4	2-15	4	CARGEDCVGGSCYSADW	TT	NA	(148)
1-18	3-22	5	CARKPRFYYDTSAWFEFW	TT	NA	(148)
			CARVLFQQLVLYAPFDIW	TT	NA	(266)
4-39		5	CAFTADNWFDPW	TT	NA	(106)
3-30		4	CAKTRGGTRDWYYFDYW	TT	NA	(106)
4-39		5	CALTYDNWFDPW	TT	NA	(106)
4-39		5	CANNRDNWFDPW	TT	NA	(106)
1-18		4	CARDPGNYAGHSHFDYW	TT	NA	(106)
1-18		4	CARDPRQTLGSSYFVYW	TT	NA	(106)
3-30		6	CAREVVRYSYGSGEYYYYGMDVW	TT	NA	(106)
5-51		3	CARGYCSGGSCFPADTFDVW	TT	NA	(106)
4-39		5	CARHADNWFDPW	TT	NA	(106)
4-39		5	CARQADNWFDPW	TT	NA	(106)
4-39		5	CARQTDNWFDPW	TT	NA	(106)
4-31		6	CARVFITTLGGVIDGHGMDVW	TT	NA	(106)
1-03		4	CARVPGMAAAGTEFDRW	TT	NA	(106)
3-30		4	CASTRSSTWYLDYW	TT	NA	(106)
5-51		3	CATVKYNSGWDGFDIW	TT	NA	(106)
3-74		2	CVRSTYAYGSGSDNWYLDFW	TT	NA	(106)
4-39		5	CVVTYDNWFDPW	TT	NA	(106)
			GKSYDYIRENLDS	TT	NA	(97)
			CARNLQGHYAMDVW	TT	NA	(97)
			CARLHPTCASTRCPENYGMDVW	TT	NA	(97)
			CARKPRFYYDTSAWFEFW	TT	NA	(97)

			CARKGMGHYFDFW	TT	NA	(97)
			CARGVVPAGIPFDFW	TT	NA	(97)
			CARGEPIRATVFGQPIPRGAWFDPW	TT	NA	(97)
			CARGEDCVGGSCYSADW	TT	NA	(97)
			CARDYFHSGSQYFFDYW	TT	NA	(97)
			CARDTVTPLGENLNYFAHW	TT	NA	(97)
			CARDSVTNLGENLNFFPYW	TT	NA	(97)
			CARARNYGFPHFFDFW	TT	NA	(97)
			CAKDRVRVVQAATTLDFW	TT	NA	(97)
			CAKAPIIGPKYYFYMDVW	TT	NA	(97)
			CAKAPIIGPKHYFYMDVW	TT	NA	(97)
			CARVLGGTRLYYALNVW	TT	NA	(267)
			CARDRGGTRHHYYMDVW	TT	NA	(267)
			CARDYGGTRHYYALDAW	TT	NA	(267)
			CARVVGGTRDYYALGFW	TT	NA	(267)
			CARVFGGTRLYYALNVW	TT	NA	(267)
			CARVLGGTRLYYALNVW	TT	NA	(267)
			CARYLGSTRGYYYMDVW	TT	NA	(267)
			CTRCRGGTRTYYYMDVW	TT	NA	(267)
			CTRCRGGTRTYYYMDIW	TT	NA	(267)
			CVSGSSLDYW	TT	NA	(267)
			CASGSTLDYW	TT	NA	(267)
			CVTGSSLDYW	TT	NA	(267)
			CVSGSSFDYW	TT	NA	(267)
			CVTGSSHDFW	TT	NA	(267)
			CARRPGWAATRAAGAFDIW	TT	NA	(267)
			CVTGSSLDYW	TT	NA	(267)
			CARDGVITLGGVIELRWYDPW	TT	NA	(267)
			CARAGSSWSLRPTTFDYW	TT	NA	(267)
			CARDFYSGTYRSFDYW	TT	NA	(267)
			CARARLLFCSGGRCDMDSW	TT	NA	(267)
			CARDRGRITLFGEVILRAGWFDSW	TT	NA	(267)
			CARDYYGSGSHYYFDYW	TT	NA	(267)
			CARVFGGTRLYYALNVW	TT	NA	(267)
			CARVLGGTRLYYALNVW	TT	NA	(267)
			CARVLGGTRLYYALNVW	TT	NA	(267)
			CARDYGGTRHYYALDAW	TT	NA	(267)
			CARLGIAARHYFYGVDVW	TT	NA	(267)
			CTRCRGGTRTYYMDVW	TT	NA	(267)
			CARYLGSTRGYYMDVW	TT	NA	(267)
			CATGVTMDYW	TT	NA	(267)
			CAKGLIFGVAAYYFDYW	TT	NA	(267)
			CAKGLIFGVPAYYFDSW	TT	NA	(267)
			CAKDLILGVPHYYFDSW	TT	NA	(267)
			CARHLDSYDVFTGYNLGGYMDVW	TT	NA	(267)
			CARHLDSDYDVFNGYNLGGYMDVW	TT	NA	(267)
			CVSAPRDTSTIAARFNRYFFDTW	TT	NA	(267)
			CVSAPRDTSTIAARFNRYFFDTW	TT	NA	(267)
			CASAPRDTSTIAARFNRYFFDFW	TT	NA	(267)
			CARDRGGTRHHYYMDVW	TT	NA	(267)
			CARARRTYSGYDSAFDYW	TT	NA	(267)
			CARRYDFWSGFLDYW	TT	NA	(267)
			CARVSGWGPRGGIYFDYW	TT	NA	(267)
			CARVSGWGPRGGIYFDYW	TT	NA	(267)
			CARRHYCSSTSCYDAFDIW	TT	NA	(267)
			CARIVGYNWKGEGNFDYW	TT	NA	(267)
			CARDVRRRFGEFLRPFDLW	TT	NA	(267)
			CARRIAIFSVVLRSGWFDPW	TT	NA	(267)
			CARLPKHYIAEAVTW	TT	NA	(267)
			CARSVVPATRAFDFW	TT	NA	(267)
			CARTVASLGTAFDYW	TT	NA	(267)
			CARIVGTHGFDYW	TT	NA	(267)
			CARIVGTHGFDYW	TT	NA	(267)
			CARTMGVVLPFDYW	TT	NA	(267)
			CARINGNVTIFGMILPRGWFDPW	TT	NA	(267)
			CWRINGNVTIFGMVLPRGWFDPW	TT	NA	(267)
			CARINGYVTVFGMILPRGWFDPW	TT	NA	(267)
			CARARLLFCSGGRCDMDSW	TT	NA	(267)
			CARDSAPLSRRGALGIW	TT	NA	(267)
			CARGSGTYSFLDNW	TT	NA	(267)
			CVKRRRQWLVNSSFDFW	TT	NA	(267)
			CAKGRKQWLVPDFDSW	TT	NA	(267)
			CAKYLSGGYAIDVW	TT	NA	(267)
			CVKYLWGGYAIDVW	TT	NA	(267)
3-21		4	CVTGSSLDYW	TT	NA	(80)
3-21		4	CVTGSSHDFW	TT	NA	(80)
1-69	6-6	4	CVTAPDDTGTILARHNRYYFDSW	TT	NA	(80)
3-21		4	CVSGSSLDYW	TT	NA	(80)
3-21		4	CVSGSSFDYW	TT	NA	(80)
3-21		4	CVSGGSLDYW	TT	NA	(80)
1-69	6-6	4	CVSAPRDTSTIAARFNRYFFDTW	TT	NA	(80)
3-48	2-21	3	CVRVLRWNNDAFHIW	TT	NA	(80)
3-74	2-8	6	CVRPRGYCADGLCYPALYFYYMDVW	TT	NA	(80)
4-59	3-16	4	CVRPPSWRYHYFDSW	TT	NA	(80)
4-59	3-3	2	CVRERVRVYGEIVHKYFDLW	TT	NA	(80)
3-48	3-9	4	CVRERQDGTMLTGLFNYFDHW	TT	NA	(80)
1-18	3-10	4	CVRDYNGSGKYYFEYW	TT	NA	(80)
1-69	6-6	4	CVRAPRGTSTIAARFNRYFFDSW	TT	NA	(80)
3-21	4-17	4	CVNGDYVV	TT	NA	(80)
3-21	1-26	1	CVNGDYVV	TT	NA	(80)
3-23	3-3	3	CVKYLWGGYAIDVW	TT	NA	(80)
3-23	3-3	6	CVKYLWGGYAIDVW	TT	NA	(80)
3-64	6-19	4	CVKRRRQWLVNSSFDFW	TT	NA	(80)
3-64		5	CVKDSGVTPSVW	TT	NA	(80)
3-64	3-22	3	CVKDPRRRTWFFHRSGFDIW	TT	NA	(80)

2-5	3-3	4	CVHLPPYRISLFGEDLYRPTYFDFW	TT	NA	(80)
1-24	3-3	4	CTTVGEALITLFGTVIRRARDFDSW	TT	NA	(80)
3-15	3-3	6	CTTVENFWSEFYGMDVW	TT	NA	(80)
3-15	7-27	5	CTTSAGALAGPGILW	TT	NA	(80)
3-15	4-23	4	CTTHDYW	TT	NA	(80)
3-15	3-10	4	CTTGTVVVPNDYW	TT	NA	(80)
3-15		3	CTTGGSLSMPMW	TT	NA	(80)
3-15	3-3	6	CTTDFSILRAYLLRFLEWPSYGMDVW	TT	NA	(80)
3-49		5	CTRLQYCSYVS	TT	NA	(80)
1-18	4-4	4	CTRDYSSPYYFEFW	TT	NA	(80)
3-49	3-10	4	CTRDREGSGSSKPPGPPHFDYW	TT	NA	(80)
3-48	1-26	3	CTRDLYTGSPVEAFDIW	TT	NA	(80)
3-21		6	CTRCRGGTRTYYYMDVW	TT	NA	(80)
3-21		6	CTRCRGGTRTYYYMDIW	TT	NA	(80)
1-3	3-22	6	CTRARNVHYTLPHYDMDVW	TT	NA	(80)
1-3	3-9	6	CTRARNAFYTLPHYDMDVW	TT	NA	(80)
4-39		5	CTNQMDNWFDPW	TT	NA	(80)
3-30	3-9	4	CTKDRYRGHVLTGNSFEHW	TT	NA	(80)
3-15	3-3	4	CTATYYDFWSGLSRGGYW	TT	NA	(80)
3-15	2-2	4	CTARYCTRTSCYGTPDYW	TT	NA	(80)
4-34	4-17	5	CSRVGPFDPW	TT	NA	(80)
3-21		5	CSRDEVNRPHYIW	TT	NA	(80)
3-49	2-2	6	CSPRGDCSSTNCYENFFHDLDVW	TT	NA	(80)
3-21		4	CLTGSSLDYW	TT	NA	(80)
1-69		6	CGRVLGGTRLYYALNVW	TT	NA	(80)
4-39	5-12	5	CGRQDATIRLNGPIDLW	TT	NA	(80)
3-53	1-7	4	CGNYDWNYGGVEFW	TT	NA	(80)
3-30	3-3	6	CGKDHVGYNSRSGYRPESYYYMDLW	TT	NA	(80)
3-30	3-3	6	CGKDHIGYNSRSGYRPESYYYMDLW	TT	NA	(80)
3-30	3-3	6	CGKDHIGYNSQSGYRPESYYYMDLW	TT	NA	(80)
1-69	6-6	4	CERAPPGTSTIAARFNRYFFDSW	TT	NA	(80)
1-8		5	CAVVAENWFDPW	TT	NA	(80)
1-2	2-2	4	CATWVSRESSFDYW	TT	NA	(80)
1-24	3-3	4	CATVGETRFTLFGTLMRRPRELRDW	TT	NA	(80)
3-21		4	CATGVTMDYW	TT	NA	(80)
3-21		4	CATGRTLDYW	TT	NA	(80)
1-24		4	CATGRGLEFDYW	TT	NA	(80)
3-21	4-23	4	CATGNTLDYW	TT	NA	(80)
3-15	1-26	1	CATGGDIGSYHSGGYFHHW	TT	NA	(80)
3-15		4	CATDGAGYSSNLW	TT	NA	(80)
1-69	6-6	4	CATAPRGTSTIAARFNRYFFDSW	TT	NA	(80)
1-69	2-2	4	CASGYCSSTSCYDYW	TT	NA	(80)
3-21		4	CASGVTHDYW	TT	NA	(80)
3-21		4	CASGRSLDYW	TT	NA	(80)
3-21		4	CASGNTLDYW	TT	NA	(80)
3-21		4	CASGNTHDYW	TT	NA	(80)
3-74	2-15	3	CASGDCNSGSCYFRDAFDIW	TT	NA	(80)
1-69	6-6	4	CASAPRGTSTIAARHNRYFFDYW	TT	NA	(80)
1-69	6-6	4	CASAPRGTSTIAARFNRYFFDFW	TT	NA	(80)
1-69	6-6	6	CASAPRGTSTIAARFNRYFFDFW	TT	NA	(80)
1-69	6-6	4	CASAPRDTSTIAARLNRYFFDFW	TT	NA	(80)
1-69	6-6	4	CASAPRDTSTIAARFNRYFFDFW	TT	NA	(80)
1-69	6-6	4	CASAPRDASTIAARFNRYFFDFW	TT	NA	(80)
3-21		6	CARYRGGTRGYYYMDVW	TT	NA	(80)
3-21		6	CARYLGSTRGYYYMDVW	TT	NA	(80)
3-7	5-24	4	CARWRWHQSEFDYW	TT	NA	(80)
5-51	3-9	4	CARVYYDW	TT	NA	(80)
1-69	1-26	6	CARVVGGTRSYYALGVW	TT	NA	(80)
1-69	1-26	6	CARVVGGTRSYYALGFW	TT	NA	(80)
1-69	1-26	6	CARVVGGTRPYYALGLW	TT	NA	(80)
1-3	4-17	1	CARVSRYGDYGPEYFQHW	TT	NA	(80)
4-59	6-25	4	CARVSGWGPRGGIYFDYW	TT	NA	(80)
4-59	6-25	4	CARVSGWGPRGGIYFDFW	TT	NA	(80)
4-59	3-3	2	CARVRVSLIGVVVHRYFDLW	TT	NA	(80)
4-59	3-3	5	CARVRINLYGALLLSWFDSW	TT	NA	(80)
2-26	3-3	4	CARVRGRRIPLFGGTIIRGARFDYW	TT	NA	(80)
1-69	6-13	4	CARVREGQQVLVFDSW	TT	NA	(80)
4-34	1-7	3	CARVPLFFSNWNLRAFDIW	TT	NA	(80)
2-26	3-3	5	CARVNGYVTVFGMILPRGWFDPW	TT	NA	(80)
4-59		6	CARVMKVGDYYFYYMDVW	TT	NA	(80)
1-69	6-19	6	CARVLGGTRVYYALNVW	TT	NA	(80)
1-69	1-26	6	CARVLGGTRLYYAQNVW	TT	NA	(80)
1-69	1-26	6	CARVLGGTRLYYAMNVW	TT	NA	(80)
1-69	2-2	6	CARVLGGTRLYYALNVW	TT	NA	(80)
1-69	2-2	6	CARVLGGTRLYYALHIW	TT	NA	(80)
4-4	2-15	3	CARVKGYCGGGRCHWGVSDIW	TT	NA	(80)
1-69	3-16	6	CARVGGGTRGYYYMAVW	TT	NA	(80)
1-69	3-16	6	CARVFGGTRLYYALNVW	TT	NA	(80)
4-61	7-27	4	CARVEWGSRRAFDSW	TT	NA	(80)
2-5	2-2	4	CARTVVPAGVGFDYW	TT	NA	(80)
2-5	3-10	4	CARTVRGVVPFDYW	TT	NA	(80)
2-5	2-2	4	CARTVLPATFAFDFW	TT	NA	(80)
2-5	3-3	4	CARTVGVVLPFDYW	TT	NA	(80)
2-5	3-10	4	CARTVGAFGRFPFDFW	TT	NA	(80)
2-5	6-13	4	CARTVASLGTAFDYW	TT	NA	(80)
6-1	1-26	6	CARTSGGVEWLSSIYYGMDVW	TT	NA	(80)
5-51	3-3	6	CARTQYDAWSGSVLEDIYYYYYMDVW	TT	NA	(80)
1-18	3-10	4	CARTNNFGSGNYDAWNYFFDYW	TT	NA	(80)
2-26	3-3	5	CARTNGNVTIFGMVLPRGWFDPW	TT	NA	(80)
2-5	3-3	4	CARTMGVVLPFDYW	TT	NA	(80)
2-5	2-2	4	CARSVVPATRSFDYW	TT	NA	(80)
2-5	2-2	4	CARSVVPATRAFDFW	TT	NA	(80)
2-5	1-1	4	CARSVFPVLPFDFW	TT	NA	(80)
3-7	5-24	4	CARSTHSSADYW	TT	NA	(80)
4-30	3-16	3	CARSSAYVYAFDIW	TT	NA	(80)

4-59	5-24	3	CARSREPYDMKAFDVW	TT	NA	(80)
2-5	3-10	4	CARSMRMVRGRIEQFDYW	TT	NA	(80)
4-61	5-5	4	CARSGPEYSYGYTFDYW	TT	NA	(80)
4-61	3-3	4	CARRYDFWSGFLDYW	TT	NA	(80)
3-11	7-27	2	CARRPNLWGSALWYFDLW	TT	NA	(80)
3-21	2-15	3	CARRPGWAATRAAGAFDIW	TT	NA	(80)
4-31	3-3	5	CARRIAIFSVVLRSGWFDPW	TT	NA	(80)
4-59	2-2	3	CARRHYCSSTSCYDAFDIW	TT	NA	(80)
4-59	2-2	3	CARRHYCRSTSCYDAFDTW	TT	NA	(80)
4-59	2-2	3	CARRHYCRSTSCYDAFDIW	TT	NA	(80)
1-69	3-22	5	CARRHGYSDQRNWFDSW	TT	NA	(80)
1-46	2-2	6	CARQYCSSASCYVSLRHSGQFYYFMDVW	TT	NA	(80)
1-46	2-2	6	CARQYCSSASCYASLRHSGQFYYFMDVW	TT	NA	(80)
4-39		5	CARQTDNWFDPW	TT	NA	(80)
4-39	6-19	4	CARQEPKQWLVDYYFDSW	TT	NA	(80)
5-51	3-10	4	CARQDSNTLYFFDLW	TT	NA	(80)
4-59	6-13	4	CARPPSWRYHYFDSW	TT	NA	(80)
4-39	1-14	3	CARPHVTWTSRTAFDIW	TT	NA	(80)
2-26		6	CARMRRIFINHYSSYMDVW	TT	NA	(80)
1-69	1-26	4	CARLYSGSYYPVEYW	TT	NA	(80)
4-39	6-6	4	CARLQSFYTSQNSVRRPFDSW	TT	NA	(80)
4-4	2-21	5	CARLPKHYIAEAVTW	TT	NA	(80)
4-34	5-12	6	CARLPIIGSGYDAVSLGNYGMDVW	TT	NA	(80)
5-51	6-6	5	CARLMRYSSPSGRLEGLVGRFDPW	TT	NA	(80)
1-69	1-26	6	CARLLGGTRHYYALNVW	TT	NA	(80)
3-74	7-27	3	CARLGLIW	TT	NA	(80)
1-46	2-21	6	CARLGIAARHYFYGVDVW	TT	NA	(80)
3-74	3-16	3	CARKLRVWGPFRLDAFDVW	TT	NA	(80)
2-5	7-27	5	CARKAPGWGPAFDWFGPW	TT	NA	(80)
4-39	1-1	4	CARIVGYNWKGEGNFDYW	TT	NA	(80)
2-5	1-26	4	CARIVGTHGFDYW	TT	NA	(80)
2-5	1-26	4	CARIVGTHGFDYC	TT	NA	(80)
2-26	6-13	6	CARIQVVSAADKEGLYYYNAMDVW	TT	NA	(80)
2-26	3-3	5	CARINGYVTVFGMILPRGWFDPW	TT	NA	(80)
2-26	3-3	5	CARINGNVTIFGMVPPRGWFDPW	TT	NA	(80)
2-26	3-3	5	CARINGNVTIFGMVLPRGWFDPW	TT	NA	(80)
2-26	3-3	5	CARINGNVTIFGMILPRGWFDPW	TT	NA	(80)
2-26	3-3	5	CARINGNVMVFGMVLPRGWFDPW	TT	NA	(80)
2-26	3-3	5	CARINGNVKIFGMVLPRGWFDPW	TT	NA	(80)
4-61	2-2	6	CARIKDVGHCSGGSCYSGAGWFDPW	TT	NA	(80)
5-51	2-15	4	CARHRGGWAKRGPFDYW	TT	NA	(80)
1-69	7-27	6	CARHRGGTLPYYYFDVW	TT	NA	(80)
5-51	3-9	6	CARHLDSYDVFNGYNLGGYMDVW	TT	NA	(80)
5-51	3-9	6	CARHLDSYDVFDGYNLGGYMDVW	TT	NA	(80)
5-51	3-9	6	CARHLDSHDVFTGYNLGGYMDLW	TT	NA	(80)
2-5	3-10	4	CARHIRGVTPFDYR	TT	NA	(80)
4-39	2-2	4	CARHGGYCSSASCYSGAFLTGYYFDYW	TT	NA	(80)
3-48	5-24	6	CARGVLGGYNNGLYYYYYLDVW	TT	NA	(80)
1-69		6	CARGVGTAGTGMDVW	TT	NA	(80)
1-2	1-26	4	CARGVFSRANYGFLYSFDSW	TT	NA	(80)
1-69	6-6	3	CARGVAGIASRRHAYDIW	TT	NA	(80)
3-33	3-10	4	CARGSGTYSFLDNW	TT	NA	(80)
4-34	2-2	6	CARGPCRSTSCPSYYYGMDVW	TT	NA	(80)
1-69	2-2	5	CARGKDCRSNNCYLSERSNWFDPW	TT	NA	(80)
1-69	2-2	5	CARGKDCRSNNCYLSERNNWFDPW	TT	NA	(80)
1-69	2-2	5	CARGKDCRANNCYLSERNNWFDPW	TT	NA	(80)
4-34	2-2	4	CARGHFKEEPTLPTMSRPTTPPYFDYW	TT	NA	(80)
3-30	6-19	4	CARGGVGVASKLSHW	TT	NA	(80)
1-8		3	CARGGSSAFDVW	TT	NA	(80)
4-59	5-5	4	CARGGRYSYGYASFFDYW	TT	NA	(80)
4-59	2-2	4	CARGGPSWTVIDHW	TT	NA	(80)
3-30	3-10	4	CARGGPGNVRVNLGRGVFDFW	TT	NA	(80)
4-59	3-10	6	CARGEPDLYYYGSGSYSPRRVGDYYMDVW	TT	NA	(80)
4-30	1-7	4	CARGDKWAFWFGEIDYW	TT	NA	(80)
3-49	2-21	6	CARGCPEIAVRAIYYYGMDVW	TT	NA	(80)
1-8		5	CARGAGAQGYDWFDPW	TT	NA	(80)
1-18		4	CAREYGDYKFDYW	TT	NA	(80)
4-59	3-3	2	CARERVRVYGEIIHKYFDLW	TT	NA	(80)
3-33	2-8	4	CARERGHDTNGQPDNW	TT	NA	(80)
4-31	2-8	4	CAREMVGHSAPADYW	TT	NA	(80)
4-59	6-19	4	CAREKESAGWNAHYFDYW	TT	NA	(80)
4-31	2-2	4	CAREGYCSSIYCSFDYW	TT	NA	(80)
4-59	6-6	5	CAREEFTSSSRWFDPW	TT	NA	(80)
1-18	3-10	4	CARDYYGSGSHYYFDYW	TT	NA	(80)
1-18	4-4	4	CARDYSSPYYFEHW	TT	NA	(80)
1-18	4-4	4	CARDYSSPYYFDYW	TT	NA	(80)
1-18	4-4	4	CARDYSSPYHFDYW	TT	NA	(80)
1-69	4-23	6	CARDYGGTRHYYALDAW	TT	NA	(80)
1-69	4-23	6	CARDYGGTRDYYALDAW	TT	NA	(80)
1-18	3-10	4	CARDYFGSGSVYYFDYW	TT	NA	(80)
1-18	3-10	4	CARDYFGSGSIYYFDYW	TT	NA	(80)
1-18	3-10	4	CARDYFGSGPIYYFDHW	TT	NA	(80)
1-18	3-10	4	CARDWNQGSGIYYISDW	TT	NA	(80)
4-61	3-10	3	CARDVRRRFGEFLRPFDLW	TT	NA	(80)
1-46	5-5	6	CARDVDLWLTADKGDYYGMDFW	TT	NA	(80)
1-69	6-19	6	CARDSRTRARGSGGWYRGNVYYYAMDVW	TT	NA	(80)
1-69	2-21	6	CARDSRTRARGSGGFYRGNVYYYAMDVW	TT	NA	(80)
3-33	6-25	3	CARDSAPLSRRGALGIW	TT	NA	(80)
1-3	1-14	4	CARDRWMGGYPRNFFDSW	TT	NA	(80)
1-69	7-27	6	CARDRLGTREYDWIFYGMEVW	TT	NA	(80)
1-18	3-3	5	CARDRGRITLFGEVILRAGWFDSW	TT	NA	(80)
1-69		6	CARDRGGTRHHYYMDVW	TT	NA	(80)
3-30	3-3	4	CARDRGAQITLFGAPLIRPSSFDSW	TT	NA	(80)
3-30	3-3	4	CARDRGAQITLFGAPLIRPSSFDSR	TT	NA	(80)
3-21	6-25	3	CARDRAQIWGYRRGGDALDVW	TT	NA	(80)

1-18	6-19	4	CARDPRTVAGMSYFLYW	TT	NA	(80)
1-18	6-19	4	CARDPRTVAGMSYFLHW	TT	NA	(80)
3-21	2-21	4	CARDPRARGLLFGNFDYW	TT	NA	(80)
1-46	2-2	4	CARDPKVWSASSGFDYW	TT	NA	(80)
4-31	2-15	4	CARDMGSRVAAPLDYW	TT	NA	(80)
3-30	3-10	5	CARDLYSGSGSNWATNRFDPW	TT	NA	(80)
3-74		4	CARDLGGIGSNW	TT	NA	(80)
3-48	3-3	4	CARDLFFGPLKGLFDSW	TT	NA	(80)
4-61	3-10	3	CARDIRRRFGEFFRPFDFW	TT	NA	(80)
3-30	2-15	5	CARDICSAGNCYPGNGLDPW	TT	NA	(80)
1-18	3-16	4	CARDHSSPYYFDYW	TT	NA	(80)
3-7	3-16	5	CARDGVITLGGVIELRWYDPW	TT	NA	(80)
1-18	4-23	4	CARDFYSGTYRSFDYW	TT	NA	(80)
1-18	4-23	4	CARDFYSGTYRSFDHW	TT	NA	(80)
1-18	4-23	4	CARDFYSGSYRSFDYW	TT	NA	(80)
1-18	1-26	4	CARDFYSGSYRSFDCW	TT	NA	(80)
3-30	2-15	3	CARDEDGVAGAFDIW	TT	NA	(80)
3-21	2-15	6	CARCRGGTRQYYYLDVW	TT	NA	(80)
2-5	2-2	4	CARAVVPARVPFDFW	TT	NA	(80)
1-46	3-10	5	CARATLTDWSGRFWFDPW	TT	NA	(80)
7-4-1	3-22	4	CARASYHYDSSGSSPQYYFDSW	TT	NA	(80)
3-7	6-13	4	CARASSSWSLRPTTFDYW	TT	NA	(80)
1-69	1-14	6	CARARYSYYDSSGYYRLDFW	TT	NA	(80)
4-61	5-12	4	CARARRTYSGYDSAFDYW	TT	NA	(80)
4-61	3-9	4	CARARRTFSGYYSAFDYW	TT	NA	(80)
4-61	5-12	4	CARARRTFSGYDSAFDYW	TT	NA	(80)
3-30	2-15	4	CARARLLFCSGGRCDMDSW	TT	NA	(80)
3-33	3-10	5	CARARGPYSGTYYATGWW	TT	NA	(80)
1-69	6-6	4	CARAPRGTSTISARFNRYFFESW	TT	NA	(80)
1-69	6-6	4	CARAPRGTSTIAARFNRYFFDSW	TT	NA	(80)
1-69	6-6	4	CARAPRDTSTIAARHNRYQFDSW	TT	NA	(80)
1-69	6-6	4	CARAPRDTSTIAARFNRYFFDSW	TT	NA	(80)
1-69	6-6	4	CARAPPGTSTIAARFNRYFFDSW	TT	NA	(80)
1-69	6-6	4	CARAPGGTSTIAARFNRYFFDSW	TT	NA	(80)
3-13	4-17	6	CARAKQGHTGLDVW	TT	NA	(80)
2-5	2-15	4	CARAGVPATRSFDFW	TT	NA	(80)
3-7	6-13	4	CARAGTSWSLRPTTFDYW	TT	NA	(80)
3-7	6-13	4	CARAGSSWSLRPTTFDYW	TT	NA	(80)
1-69	3-22	6	CARADHYDSSGSNFFFYGMDVW	TT	NA	(80)
4-34	4-17	5	CAQGVVGSW	TT	NA	(80)
1-69	6-19	4	CAMGDSSSKIEYW	TT	NA	(80)
1-69	5-24	4	CALRDGSNFVYFDFW	TT	NA	(80)
1-69	5-24	4	CALRDGNNLVYFDYW	TT	NA	(80)
1-69	5-24	4	CALRDGNNFVYFDYW	TT	NA	(80)
4-34	3-10	3	CALDHYYGSGSYYNLPAFDIW	TT	NA	(80)
3-23	6-19	3	CAKYRRQWLSNECFDIW	TT	NA	(80)
3-23	2-8	6	CAKYLSGGYAIDVW	TT	NA	(80)
3-23	6-6	4	CAKSPEPIPARLAPGHFDYW	TT	NA	(80)
3-23	3-3	4	CAKSLIFGVAAYYFDSW	TT	NA	(80)
1-69	1-26	4	CAKLPSGSYYDYFDSW	TT	NA	(80)
3-23	6-19	3	CAKHAMGGWYGFGAFDIW	TT	NA	(80)
3-23	6-19	4	CAKGRKQWLVPDFDSW	TT	NA	(80)
3-30	3-22	5	CAKGLSQALNYYGSGSPFL	TT	NA	(80)
3-30	2-2	6	CAKGLSQALNYYGSGSPFL	TT	NA	(80)
3-23	3-3	4	CAKGLIFGVPAYYFDSW	TT	NA	(80)
3-23	3-3	4	CAKGLIFGVAAYYFDSW	TT	NA	(80)
3-30	6-19	4	CAKGITRPGVAVRERFDHW	TT	NA	(80)
3-30	3-22	4	CAKEPAPFNYYDSSAYYGGGYYFDYW	TT	NA	(80)
3-30	3-16	3	CAKDRQLKDAFDIW	TT	NA	(80)
3-30		4	CAKDRITAADYW	TT	NA	(80)
3-23	4-17	4	CAKDLYGDYDLDYW	TT	NA	(80)
3-23	2-2	4	CAKDLVIGQCTTTKCPRFFDSW	TT	NA	(80)
3-30	3-10	5	CAKDLFSGSGSTWATNRLDPW	TT	NA	(80)
3-30	2-8	4	CAKDLESFYCTDGGRPFDYW	TT	NA	(80)
3-30	2-8	4	CAKDLESFYCTDGGCPFDYW	TT	NA	(80)
1-69	6-6	4	CAKAPRATSTIAARFNRYFFDAW	TT	NA	(80)
3-23	6-19	4	CAKAHKQWLAHYNFDYW	TT	NA	(80)
3-23	5-12	4	CAKAGIQWLRDYFYDSW	TT	NA	(80)
3-23	2-2	4	CAKADNIVLVPAALTRPVDYW	TT	NA	(80)
3-20	1-26	3	CAGAKVGVGGENAFDIW	TT	NA	(80)
3-23	3-3	4	CAEGLIFGVAAYYFDHW	TT	NA	(80)
3-33	7-27	4	CAAQSPFKGDHDYW	TT	NA	(80)
			CARTYGGKASIAARRDHYGMDVW	TT	NA	(253)
			CAKGGWNYYGSGSYYW	TT	NA	(253)
			CARAPGGRGVIIWAFDYW	TT	NA	(253)
			CAKDRTAAAGTPKRSWFDPW	TT	NA	(253)
			CAKQAGIAVAGSFDYW	TT	NA	(253)
1-9	3-2	3	CARKTVRATYPDW	TT	NA	(268)
3-6	5-5	3	CAREGVLPESW	TT	NA	(268)
1-35	2-10	1	CARRAYYGNSFSWPFDVW	TT	NA	(268)
1-39	2-10	1	CARRAYYGNSYWYFDVW	TT	NA	(268)
1-4	5-7	3	CARSASPLTWFAYW	TT	NA	(268)
5-2	2-1	3	CARGGYGNPFAYW	TT	NA	(268)
3-6	1-1	3	CAREGVIFGYW	TT	NA	(268)
			CARDHEDSLGGIWGYLEYW	TT	NA	(269)
			CARAPYDFWNGYYLDYW	TT	NA	(269)
			CTTGVTLDYW	TT	NA	(269)
			CAKASRQCVAEYYFDFDYW	TT	NA	(269)
			CAKAARQWLAEYYFDYW	TT	NA	(269)
			CARHGSQREITVFGTSDFFPYAMDIW	TT	NA	(269)
			CTTGITLDYW	TT	NA	(269)
			CARFYRYDYW	TT	NA	(270)
			CARLGYDGVALDYW	TT	NA	(270)
			CARWGYGDYSYAMDYW	TT	NA	(270)
			CARGFLLLYFDYW	TT	NA	(270)

CARHGDYGNYDWYFDVW	TT	NA	(270)
CATSPHYGSRYGYW	TT	NA	(270)
CARHYRYDYW	TT	NA	(270)
CSRDFDYGSYFDYW	TT	NA	(270)
CVRDFYYGNYEDYW	TT	NA	(270)
CARQDRYGFALDYW	TT	NA	(270)
CARQEAYGFALDYW	TT	NA	(270)
CARNFDYGNYFDYW	TT	NA	(270)
CARDFYYGNYFDYW	TT	NA	(270)
CARDFEYGSYFDYW	TT	NA	(270)
CARDFYFGGYFDYW	TT	NA	(270)
CARDRGAFDYW	TT	NA	(270)
CVRDGYRVMDYW	TT	NA	(270)
CARDFYYGSYFDYW	TT	NA	(270)
CAFNFYFDYW	TT	NA	(270)
CARSRYDGFYYAMDYW	TT	NA	(270)
CARGGFAYW	TT	NA	(270)
CAREGVLFGYW	TT	NA	(270)
CARQGYYGISPDYW	TT	NA	(270)
CARDFQYGNYIDYW	TT	NA	(270)
CARDFQYGNYFDYW	TT	NA	(270)
CVRTFHYGNYFDYW	TT	NA	(270)
CARGGKGFAYW	TT	NA	(270)
CARNFDYGNYEDYW	TT	NA	(270)
CARYFDYGNYFDYW	TT	NA	(270)
CARRFEYGNYDDYW	TT	NA	(270)
CARRFDYGNYDDYW	TT	NA	(270)
CTRRFDYGNYDDYW	TT	NA	(270)
CTRIFDYGYYFDYW	TT	NA	(270)
CARDGNWAFW	TT	NA	(270)
CARSGNYEFW	TT	NA	(270)
CVRGGNYAYW	TT	NA	(270)
CARDFYYGRYFDYW	TT	NA	(270)
CTRGGNYAYW	TT	NA	(270)
CSRGGNYAYW	TT	NA	(270)
CTRDGNWAYW	TT	NA	(270)
CARDFFYGRYFDYW	TT	NA	(270)

9.2 Appendix to Chapter 4

9.2.1 Figures

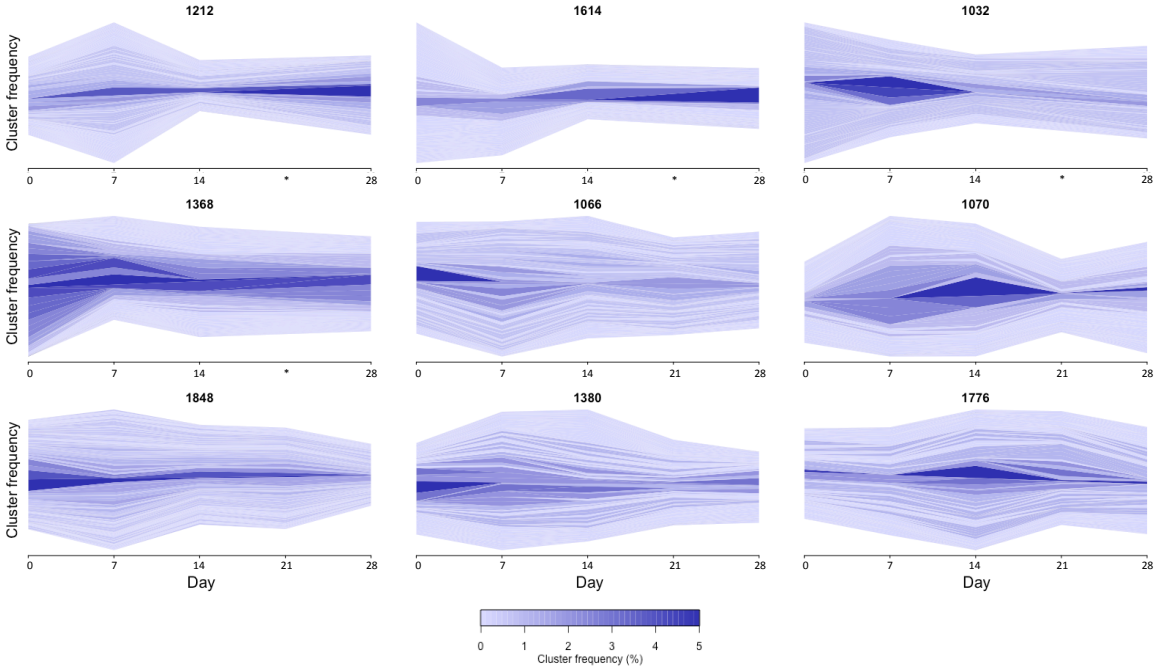

Figure 9.3: IgA cluster kinetics plot - For each participant, the 200 most frequent IgA clusters are found. At each day, the frequencies of these clusters are plotted as a stacked bar chart, centred to the middle of the y-axis. Clusters present on adjacent days are then joined using a horizontal stream to illustrate how the frequency of the clusters changes over time. The width and darkness of the stream represents the frequency of the cluster. * Sample not obtained at this day.

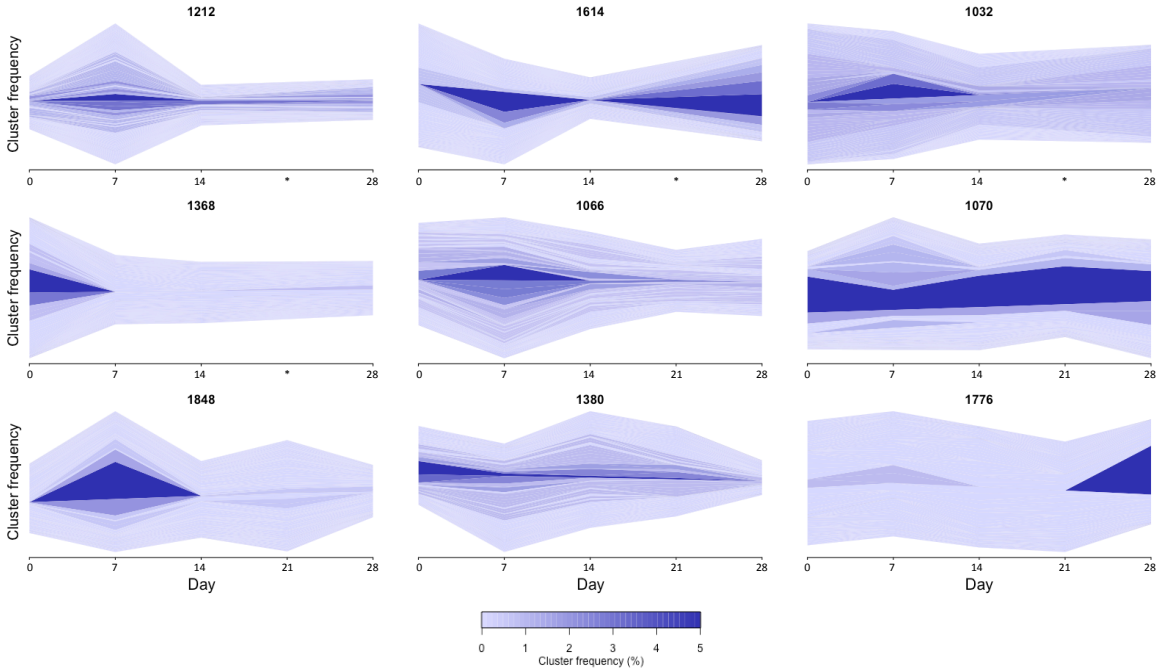

Figure 9.4: IgG cluster kinetics plot - Same as Figure 9.3, but for IgG samples

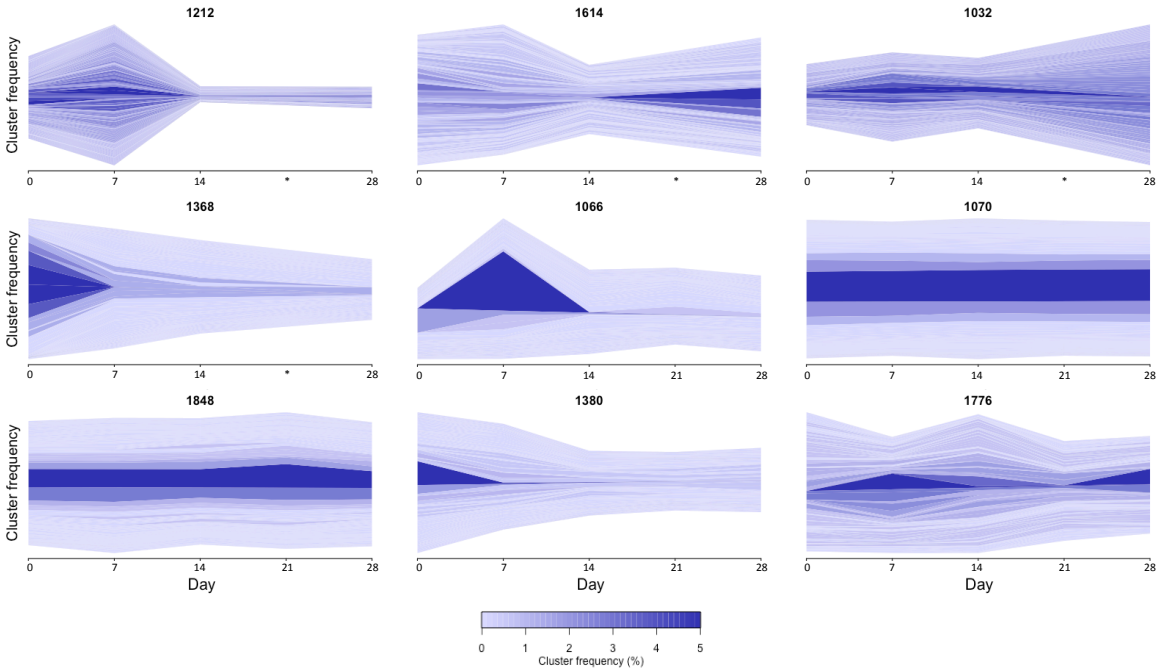

Figure 9.5: IgM cluster kinetics plot - Same as Figure 9.3, but for IgM samples

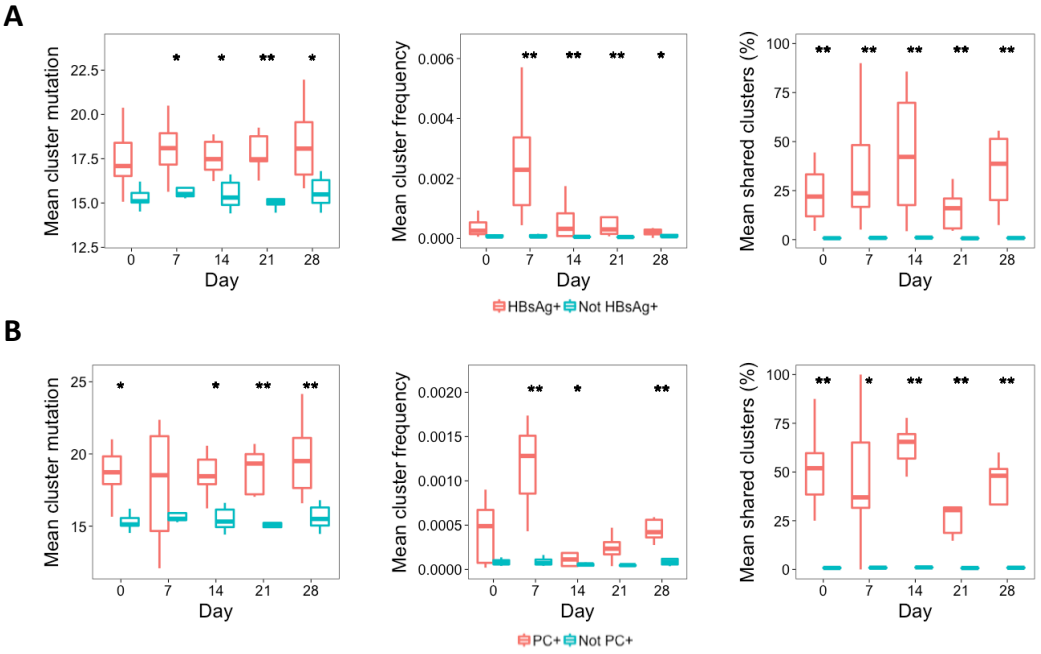

Figure 9.6: Properties of HBsAg+ and PC+ clusters - For each sample, clusters were annotated as HBsAg+ (**A**) or PC+ (**B**) based on whether sequences from the HBsAg+ or PC+ FACS enriched sequence datasets matched to them. The metrics of mean cluster mutation, mean cluster frequency, and mean percent of clusters shared by at least 2 participants were then calculated for each sample using just the clusters annotated as HBsAg+ or PC+, and the unannotated clusters. Boxes show locations of 25, 50, and 75th percentiles, and whiskers show data within 1.5x the interquartile range from 8 participants, except at visit 21 where there were only samples from 5 participants. * p <0.05, ** P <0.001 (Mann-Whitney U test).

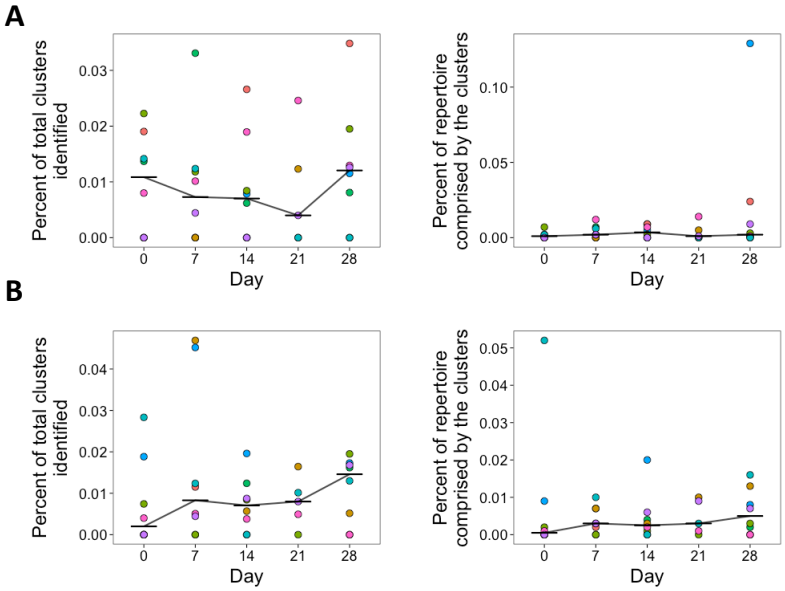

Figure 9.7: Kinetics of vaccine-irrelevant clusters - (**A**) Of the 30 total clusters annotated as TT-specific based on similarity to previously described sequences, the number of these in each participant at each day was determined, and expressed as a percentage of the total number of clusters present in that participant at that day (left panel). In addition, the percent of the total repertoire comprised by the clusters identified (ie, corrected for cluster size) was determined (right panel). (**B**) Shows the same as **A**, but with the 28 clusters identified as influenza-specific. For all plots, horizontal bars show the median value from the 8 participants. The colour of the point represents the participant.

9.2.2 Tables

Table 9.3: Summary of the total B cell samples used for sequencing, and sequence data obtained

Participant	Day	Cells	Isotype	Cell number	Raw seqs	Filtered seqs	Clusters
1032	0	Total B cells	IgA	500,000	281,579	100,000	4,983
1032	0	Total B cells	IgG	500,000	243,879	100,000	5,254
1032	0	Total B cells	IgM	500,000	331,231	100,000	47,192
1032	7	Total B cells	IgA	500,000	239,038	100,000	8,138
1032	7	Total B cells	IgG	500,000	288,204	100,000	8,679
1032	7	Total B cells	IgM	500,000	309,031	100,000	53,534
1032	14	Total B cells	IgA	500,000	258,896	100,000	10,248
1032	14	Total B cells	IgG	500,000	254,121	100,000	15,041
1032	14	Total B cells	IgM	500,000	302,460	100,000	62,369
1032	28	Total B cells	IgA	500,000	295,754	100,000	7,868
1032	28	Total B cells	IgG	500,000	265,032	100,000	8,613
1032	28	Total B cells	IgM	500,000	315,870	100,000	51,059
1066	0	Total B cells	IgA	500,000	384,924	100,000	11,360
1066	0	Total B cells	IgG	500,000	336,302	100,000	16,555
1066	0	Total B cells	IgM	500,000	410,635	100,000	75,014
1066	7	Total B cells	IgA	500,000	366,712	100,000	10,986
1066	7	Total B cells	IgG	500,000	455,339	100,000	12,793
1066	7	Total B cells	IgM	500,000	449,097	100,000	70,032
1066	14	Total B cells	IgA	500,000	417,423	100,000	11,122
1066	14	Total B cells	IgG	500,000	467,586	100,000	17,585
1066	14	Total B cells	IgM	500,000	521,114	100,000	70,223
1066	21	Total B cells	IgA	500,000	446,196	100,000	15,729
1066	21	Total B cells	IgG	500,000	482,029	100,000	24,339
1066	21	Total B cells	IgM	500,000	471,378	100,000	73,971
1066	28	Total B cells	IgA	500,000	383,578	100,000	12,424
1066	28	Total B cells	IgG	500,000	391,338	100,000	19,356
1066	28	Total B cells	IgM	500,000	468,921	100,000	71,217
1070	0	Total B cells	IgA	500,000	367,806	100,000	11,697
1070	0	Total B cells	IgG	500,000	353,992	100,000	13,473
1070	0	Total B cells	IgM	500,000	533,009	100,000	24,591
1070	7	Total B cells	IgA	500,000	392,759	100,000	8,630
1070	7	Total B cells	IgG	500,000	370,941	100,000	8,464
1070	7	Total B cells	IgM	500,000	463,595	100,000	25,396
1070	14	Total B cells	IgA	500,000	403,805	100,000	9,252
1070	14	Total B cells	IgG	500,000	408,608	100,000	11,870
1070	14	Total B cells	IgM	500,000	460,813	100,000	22,688
1070	21	Total B cells	IgA	500,000	412,018	100,000	12,721
1070	21	Total B cells	IgG	500,000	413,979	100,000	12,755
1070	21	Total B cells	IgM	500,000	471,874	100,000	25,600
1070	28	Total B cells	IgA	500,000	393,271	100,000	9,522
1070	28	Total B cells	IgG	500,000	359,743	100,000	10,254
1070	28	Total B cells	IgM	500,000	489,340	100,000	24,457
1212	0	Total B cells	IgA	500,000	292,080	100,000	6,526
1212	0	Total B cells	IgG	500,000	280,407	100,000	7,302
1212	0	Total B cells	IgM	500,000	263,810	100,000	35,981
1212	7	Total B cells	IgA	500,000	235,819	100,000	5,134
1212	7	Total B cells	IgG	500,000	236,620	100,000	6,045
1212	7	Total B cells	IgM	500,000	305,753	100,000	30,161
1212	14	Total B cells	IgA	500,000	230,734	100,000	11,206
1212	14	Total B cells	IgG	500,000	265,090	100,000	16,129
1212	14	Total B cells	IgM	500,000	342,949	100,000	58,609
1212	28	Total B cells	IgA	500,000	265,609	100,000	8,549
1212	28	Total B cells	IgG	500,000	239,019	100,000	12,359
1212	28	Total B cells	IgM	500,000	282,469	100,000	53,669
1368	0	Total B cells	IgA	500,000	324,406	100,000	1,677
1368	0	Total B cells	IgG	500,000	229,626	100,000	1,604
1368	0	Total B cells	IgM	500,000	327,464	100,000	19,088
1368	7	Total B cells	IgA	500,000	273,373	100,000	3,929
1368	7	Total B cells	IgG	500,000	263,905	100,000	4,337
1368	7	Total B cells	IgM	500,000	325,264	100,000	41,457
1368	14	Total B cells	IgA	500,000	232,884	100,000	5,321
1368	14	Total B cells	IgG	500,000	231,683	100,000	6,678
1368	14	Total B cells	IgM	500,000	249,160	100,000	53,032
1368	28	Total B cells	IgA	500,000	213,387	100,000	7,173
1368	28	Total B cells	IgG	500,000	203,151	100,000	8,224
1368	28	Total B cells	IgM	500,000	284,665	100,000	60,720
1380	0	Total B cells	IgA	500,000	282,185	100,000	9,958
1380	0	Total B cells	IgG	500,000	376,162	100,000	14,123
1380	0	Total B cells	IgM	500,000	385,190	100,000	71,583
1380	7	Total B cells	IgA	500,000	250,148	100,000	10,002
1380	7	Total B cells	IgG	500,000	378,682	100,000	16,157
1380	7	Total B cells	IgM	500,000	325,063	100,000	74,535
1380	14	Total B cells	IgA	500,000	234,427	100,000	22,513
1380	14	Total B cells	IgG	500,000	449,052	100,000	16,759
1380	14	Total B cells	IgM	500,000	382,603	100,000	75,301
1380	21	Total B cells	IgA	500,000	403,874	100,000	12,361
1380	21	Total B cells	IgG	500,000	489,708	100,000	19,728
1380	21	Total B cells	IgM	500,000	310,318	100,000	79,102
1380	28	Total B cells	IgA	500,000	223,590	100,000	5,645
1380	28	Total B cells	IgG	500,000	438,162	100,000	7,698
1380	28	Total B cells	IgM	500,000	265,220	100,000	46,375
1614	0	Total B cells	IgA	500,000	265,633	100,000	5,871
1614	0	Total B cells	IgG	500,000	216,878	100,000	10,622
1614	0	Total B cells	IgM	500,000	295,468	100,000	45,214
1614	7	Total B cells	IgA	500,000	269,052	100,000	6,138
1614	7	Total B cells	IgG	500,000	260,694	100,000	8,851

1614	7	Total B cells	IgM	500,000	331,460	100,000	44,247
1614	14	Total B cells	IgA	500,000	226,236	100,000	12,592
1614	14	Total B cells	IgG	500,000	272,195	100,000	25,498
1614	14	Total B cells	IgM	500,000	345,007	100,000	70,572
1614	28	Total B cells	IgA	500,000	262,190	100,000	10,757
1614	28	Total B cells	IgG	500,000	266,968	100,000	17,355
1614	28	Total B cells	IgM	500,000	378,363	100,000	66,281
1776	0	Total B cells	IgA	500,000	252,253	100,000	12,538
1776	0	Total B cells	IgG	500,000	309,604	100,000	19,060
1776	0	Total B cells	IgM	500,000	359,614	100,000	75,450
1776	7	Total B cells	IgA	500,000	347,406	100,000	14,463
1776	7	Total B cells	IgG	500,000	391,962	100,000	22,515
1776	7	Total B cells	IgM	500,000	359,678	100,000	77,343
1776	14	Total B cells	IgA	500,000	364,622	100,000	13,869
1776	14	Total B cells	IgG	500,000	402,395	100,000	22,891
1776	14	Total B cells	IgM	500,000	372,711	100,000	75,134
1776	21	Total B cells	IgA	500,000	302,786	100,000	15,741
1776	21	Total B cells	IgG	500,000	497,248	100,000	25,026
1776	21	Total B cells	IgM	500,000	313,332	100,000	77,758
1776	28	Total B cells	IgA	500,000	311,227	100,000	16,800
1776	28	Total B cells	IgG	500,000	302,880	100,000	23,840
1776	28	Total B cells	IgM	500,000	269,306	100,000	77,404
1848	0	Total B cells	IgA	500,000	387,084	100,000	22,878
1848	0	Total B cells	IgG	500,000	427,766	100,000	24,980
1848	0	Total B cells	IgM	500,000	473,752	100,000	64,859
1848	7	Total B cells	IgA	500,000	419,960	100,000	19,989
1848	7	Total B cells	IgG	500,000	412,862	100,000	19,727
1848	7	Total B cells	IgM	500,000	1,023,663	100,000	60,471
1848	14	Total B cells	IgA	500,000	403,322	100,000	10,339
1848	14	Total B cells	IgG	500,000	458,326	100,000	26,375
1848	14	Total B cells	IgM	500,000	329,791	100,000	64,802
1848	21	Total B cells	IgA	500,000	371,301	100,000	20,269
1848	21	Total B cells	IgG	500,000	439,611	100,000	20,333
1848	21	Total B cells	IgM	500,000	508,343	100,000	58,973
1848	28	Total B cells	IgA	500,000	415,186	100,000	10,957
1848	28	Total B cells	IgG	500,000	295,501	100,000	7,723
1848	28	Total B cells	IgM	500,000	464,751	100,000	24,627

Table 9.4: Summary of the HBsAg+ and PC samples used for sequencing, and sequence data obtained

Participant	Day	Cells	Isotype	Cell number	Raw seqs	Filtered seqs
1066	7	HBsAg+	IgG	1,760	327,019	2,749
1066	14	HBsAg+	IgG	1,359	326,266	4,499
1066	21	HBsAg+	IgG	979	82,093	2,105
1066	28	HBsAg+	IgG	1,204	63,774	1,874
1070	7	HBsAg+	IgG	6,187	342,243	5,067
1070	14	HBsAg+	IgG	2,068	271,512	5,470
1070	21	HBsAg+	IgG	3,344	315,878	6,512
1070	28	HBsAg+	IgG	1,227	106,439	2,683
1380	7	HBsAg+	IgG	6,583	323,395	5,338
1380	14	HBsAg+	IgG	954	62,840	6,850
1380	21	HBsAg+	IgG	4,464	297,629	6,443
1380	28	HBsAg+	IgG	5,208	35,536	1,069
1776	7	HBsAg+	IgG	8,420	269,573	6,023
1776	14	HBsAg+	IgG	15,000	231,163	7,617
1776	21	HBsAg+	IgG	6,008	242,903	6,221
1776	28	HBsAg+	IgG	9,266	304,594	8,856
1848	7	HBsAg+	IgG	1,598	77,476	2,038
1848	14	HBsAg+	IgG	6,169	348,983	1,774
1848	21	HBsAg+	IgG	1,328	62,811	1,983
1848	28	HBsAg+	IgG	1,874	21,648	527
1066	7	PC	IgG	679	118,301	5,456
1070	7	PC	IgG	479	96,005	2,737
1380	7	PC	IgG	1,548	118,290	6,525
1776	7	PC	IgG	333	108,157	4,289
1848	7	PC	IgG	405	113,655	2,821

9.3 Appendix to Chapter 5

9.3.1 Figures

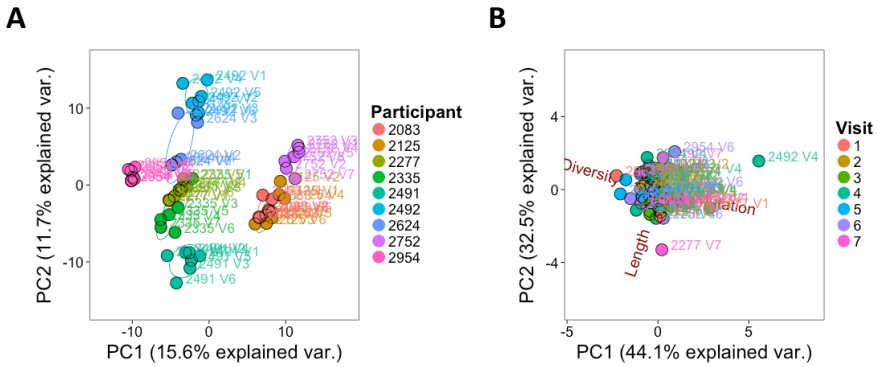

Figure 9.8: PCA of repertoire properties for QC - (**A**) For each sample, the proportion of the repertoire comprised by each VJ recombinant was calculated. These data were used in a PCA to detect outliers. (**B**) For each sample, repertoire diversity, average V gene mutation, and average CDR3 AA length was determined. These measures were corrected for cluster size, to accentuate any differences between the samples. These data were used in a PCA to detect samples with signs of B cell activation.

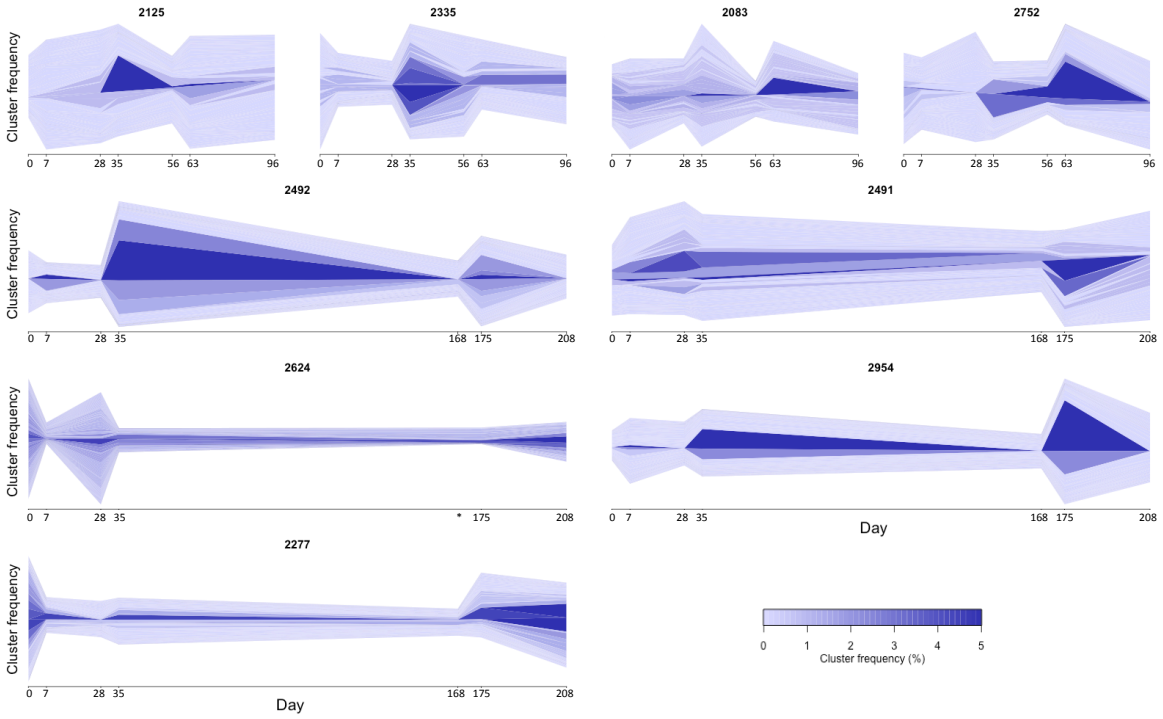

Figure 9.9: Cluster kinetics plot for the total repertoire - For each participant, the 200 most frequent clusters are found. At each day, the frequencies of these clusters are plotted as a stacked bar chart, centred to the middle of the y-axis. Clusters present on adjacent days are then joined using a horizontal stream to illustrate how the frequency of the clusters changes over time. The width and darkness of the stream represents the frequency of the cluster. The top four plots are from the participants administered the accelerated schedule, and the bottom five plots are from the participants administered the conventional schedule. * Sample not obtained at this day.

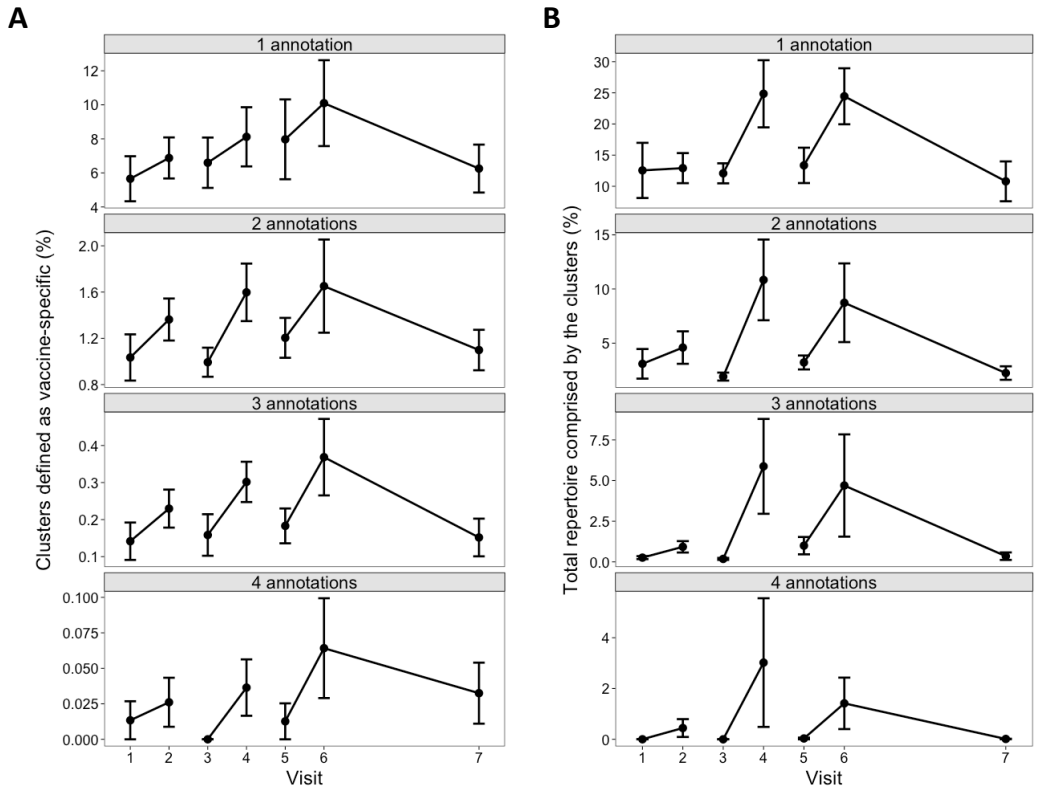

Figure 9.10: Thresholds for defining vaccine-specific clusters - Four thresholds were used for the definition of vaccine specific clusters. All clusters were annotated based on similarity to sequences in the HBsAg+ or PC+ datasets from the HepB booster vaccine or primary vaccine studies. The four thresholds are then based on the number of datasets that are able to independently annotate each cluster (ie, is the cluster annotated based on similarity to one, two three, or all four of these datasets). (**A**) The percent of abundant ($> 0.01\%$ of total repertoire) clusters at each visit that were characterized as vaccine-specific based on the four different thresholds. (**B**) Same as **A**, but corrected for cluster size by considering the percent of the repertoire comprised by the vaccine-specific clusters. Mean values $\pm$SEM are shown for all 9 participants.

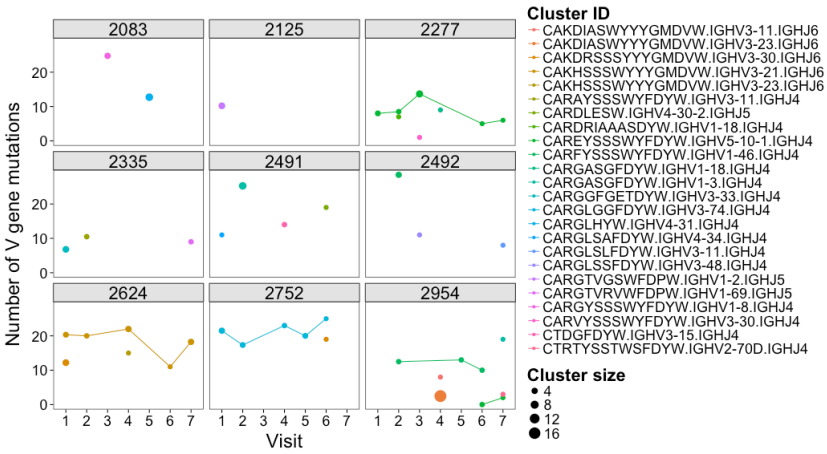

Figure 9.11: Identification of previously described HBsAg-specific sequences in the dataset - Previously described sequences were mapped to the dataset to see which clusters they would fall into based on CDR3 AA sequence identity. Shown are visits when these clusters are present in each participant. Size of the spot indicates the number of sequences in the cluster at that timepoint, and the position on the y-axis indicates the mean mutation of the sequences within the cluster at that timepoint.

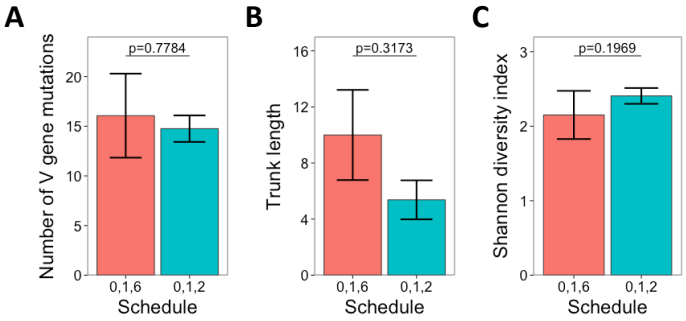

Figure 9.12: Differences in the properties of the visit 6 vaccine-specific clusters in the two vaccine groups - Differences between the two vaccine groups were investigated following the third vaccine only, which was given at a different time for the two groups. (**A**) The clusters unique to visit 6 were identified, and mean number of V gene mutations of these clusters determined. In addition, lineage trees were then constructed for each of these clusters where there were at least 25 sequences, and trunk length (**B**), and sequence diversity within the lineage (**C**) calculated. Shown are mean values ±SEM. Comparisons were performed using a Mann-Whitney U test.

9.3.2 Tables

Table 9.5: Summary of the total B cell samples used for sequencing, and sequence data obtained

Participant	Day	Cells	Isotype	Cell number	Raw seqs	Filtered seqs	Clusters
2083	0	Total B cells	IgG	500,000	285,665	75,000	7,862
2083	7	Total B cells	IgG	80,000	289,428	75,000	12,519
2083	28	Total B cells	IgG	500,000	316,854	75,000	14,671
2083	35	Total B cells	IgG	360,000	287,595	75,000	7,618
2083	56	Total B cells	IgG	500,000	290,731	75,000	18,180
2083	63	Total B cells	IgG	240,000	305,533	75,000	18,740
2083	96	Total B cells	IgG	440,000	241,392	75,000	18,780
2125	0	Total B cells	IgG	500,000	280,250	75,000	9,767
2125	7	Total B cells	IgG	500,000	248,724	75,000	6,605
2125	28	Total B cells	IgG	400,000	271,568	75,000	8,423
2125	35	Total B cells	IgG	480,000	282,535	75,000	10,894
2125	56	Total B cells	IgG	440,000	240,033	75,000	15,905
2125	63	Total B cells	IgG	480,000	273,632	75,000	14,375
2125	96	Total B cells	IgG	500,000	314,920	75,000	10,016
2277	0	Total B cells	IgG	500,000	342,026	75,000	7,268
2277	7	Total B cells	IgG	500,000	339,324	75,000	13,509
2277	28	Total B cells	IgG	500,000	327,746	75,000	11,888
2277	35	Total B cells	IgG	500,000	269,894	75,000	10,414
2277	168	Total B cells	IgG	500,000	336,158	75,000	13,788
2277	175	Total B cells	IgG	500,000	293,158	75,000	13,250
2277	208	Total B cells	IgG	500,000	315,752	75,000	10,329
2335	0	Total B cells	IgG	500,000	326,107	75,000	5,493
2335	7	Total B cells	IgG	500,000	265,573	75,000	7,622
2335	28	Total B cells	IgG	500,000	295,070	75,000	8,942
2335	35	Total B cells	IgG	500,000	292,074	75,000	8,576
2335	56	Total B cells	IgG	500,000	297,512	75,000	10,021
2335	63	Total B cells	IgG	500,000	278,008	75,000	12,387
2335	96	Total B cells	IgG	500,000	292,518	75,000	11,610
2491	0	Total B cells	IgG	500,000	326,473	75,000	16,126
2491	7	Total B cells	IgG	500,000	331,065	75,000	16,031
2491	28	Total B cells	IgG	500,000	332,481	75,000	14,045
2491	35	Total B cells	IgG	500,000	312,307	75,000	13,463
2491	168	Total B cells	IgG	500,000	352,448	75,000	17,648
2491	175	Total B cells	IgG	500,000	338,750	75,000	19,240
2491	208	Total B cells	IgG	500,000	355,851	75,000	8,768
2492	0	Total B cells	IgG	200,000	325,226	75,000	9,578
2492	7	Total B cells	IgG	280,000	342,157	75,000	13,850
2492	28	Total B cells	IgG	360,000	348,220	75,000	14,143
2492	35	Total B cells	IgG	500,000	337,566	75,000	4,163
2492	168	Total B cells	IgG	500,000	322,941	75,000	16,023
2492	175	Total B cells	IgG	500,000	332,830	75,000	8,798
2492	208	Total B cells	IgG	500,000	302,128	75,000	11,160
2624	0	Total B cells	IgG	120,000	277,691	75,000	4,275
2624	7	Total B cells	IgG	500,000	328,034	75,000	12,262
2624	28	Total B cells	IgG	120,000	301,307	75,000	4,231
2624	35	Total B cells	IgG	500,000	316,346	75,000	13,634
2624	175	Total B cells	IgG	500,000	249,755	75,000	15,662
2624	208	Total B cells	IgG	120,000	297,572	75,000	13,035
2752	0	Total B cells	IgG	500,000	329,479	75,000	8,323
2752	7	Total B cells	IgG	500,000	341,075	75,000	10,575
2752	28	Total B cells	IgG	500,000	307,994	75,000	9,778
2752	35	Total B cells	IgG	500,000	341,418	75,000	12,132
2752	56	Total B cells	IgG	500,000	332,007	75,000	16,018
2752	63	Total B cells	IgG	500,000	287,170	75,000	14,995
2752	96	Total B cells	IgG	500,000	345,096	75,000	11,489
2954	0	Total B cells	IgG	500,000	321,396	75,000	18,239
2954	7	Total B cells	IgG	500,000	317,080	75,000	13,241
2954	28	Total B cells	IgG	500,000	298,952	75,000	13,541
2954	35	Total B cells	IgG	500,000	314,153	75,000	14,459
2954	168	Total B cells	IgG	500,000	293,970	75,000	18,782
2954	175	Total B cells	IgG	500,000	324,087	75,000	14,837
2954	208	Total B cells	IgG	500,000	318,612	75,000	16,080

Table 9.6: Summary of the HBsAg+ and PC samples used for sequencing, and sequence data obtained

Participant	Day	Cells	Isotype	Cell number	Raw seqs	Filtered seqs
2083	7	HBsAg+	IgG	1,840	11,488	2,435
2083	28	HBsAg+	IgG	2,914	287,257	2,232
2083	35	HBsAg+	IgG	3,670	70,153	6,435
2083	56	HBsAg+	IgG	1,365	345,063	1,131
2083	63	HBsAg+	IgG	1,285	63,585	1,634
2083	96	HBsAg+	IgG	695	351,756	343
2125	7	HBsAg+	IgG	1,628	59,493	778
2125	28	HBsAg+	IgG	1,963	333,954	985
2125	35	HBsAg+	IgG	3,569	38,023	4,582
2125	56	HBsAg+	IgG	534	304,276	590
2125	63	HBsAg+	IgG	2,239	28,173	1,736
2125	96	HBsAg+	IgG	772	343,337	1,146
2277	7	HBsAg+	IgG	3,912	323,748	6,911
2277	28	HBsAg+	IgG	5,136	278,108	8,402
2277	35	HBsAg+	IgG	8,563	322,579	8,538
2277	168	HBsAg+	IgG	10,000	172,336	130
2277	175	HBsAg+	IgG	7,104	298,652	5,190
2277	208	HBsAg+	IgG	5,312	290,192	7,390
2335	7	HBsAg+	IgG	4,956	300,768	4,558
2335	28	HBsAg+	IgG	1,963	58,977	6,077
2335	35	HBsAg+	IgG	10,000	269,371	6,702
2335	56	HBsAg+	IgG	1,363	59,109	1,161
2335	63	HBsAg+	IgG	3,056	333,485	4,130
2335	96	HBsAg+	IgG	2,866	53,465	1,371
2491	7	HBsAg+	IgG	10,000	291,482	8,239
2491	28	HBsAg+	IgG	3,139	70,670	6,469
2491	35	HBsAg+	IgG	7,045	302,927	9,152
2491	168	HBsAg+	IgG	8,812	48,948	101
2491	175	HBsAg+	IgG	6,344	405,350	2,565
2491	208	HBsAg+	IgG	6,687	77,866	7,994
2492	7	HBsAg+	IgG	4,337	55,272	8,049
2492	28	HBsAg+	IgG	5,833	440,076	8,001
2492	35	HBsAg+	IgG	4,521	29,642	6,210
2492	168	HBsAg+	IgG	2,339	292,343	149
2492	175	HBsAg+	IgG	2,545	67,994	5,504
2492	208	HBsAg+	IgG	3,174	374,457	7,058
2624	7	HBsAg+	IgG	8,129	48,770	7,098
2624	28	HBsAg+	IgG	2,677	516,659	2,365
2624	35	HBsAg+	IgG	7,837	14,744	6,419
2624	175	HBsAg+	IgG	4,699	102,802	6,201
2624	208	HBsAg+	IgG	1,871	14,869	1,633
2752	7	HBsAg+	IgG	1,689	317,723	1,265
2752	28	HBsAg+	IgG	2,303	56,544	981
2752	35	HBsAg+	IgG	12,000	344,405	8,401
2752	56	HBsAg+	IgG	4,260	23,743	6,770
2752	63	HBsAg+	IgG	3,201	342,352	7,478
2752	96	HBsAg+	IgG	2,848	47,866	2,114
2954	7	HBsAg+	IgG	9,041	48,523	10,978
2954	28	HBsAg+	IgG	5,673	363,442	10,372
2954	35	HBsAg+	IgG	6,097	8,155	9,985
2954	168	HBsAg+	IgG	4,273	305,422	90
2954	175	HBsAg+	IgG	4,536	3,570	11,975
2954	208	HBsAg+	IgG	7,869	203,874	8,238
2083	7	PC	IgG	130	103,870	1,490
2083	35	PC	IgG	350	93,350	2,225
2083	63	PC	IgG	143	123,015	896
2125	7	PC	IgG	912	94,768	4,167
2125	35	PC	IgG	1,350	111,820	5,819
2125	63	PC	IgG	272	87,461	2,763
2277	7	PC	IgG	169	75,811	1,997
2277	35	PC	IgG	262	118,938	3,843
2277	175	PC	IgG	897	116,250	4,484
2335	7	PC	IgG	1,594	102,694	5,526
2335	35	PC	IgG	5,603	120,645	6,607
2335	63	PC	IgG	364	111,386	4,116
2491	7	PC	IgG	427	114,667	3,462
2491	35	PC	IgG	479	116,859	4,082
2491	175	PC	IgG	593	104,336	13,365
2492	7	PC	IgG	626	118,753	3,009
2492	35	PC	IgG	2,187	114,854	3,207
2492	175	PC	IgG	483	130,166	4,526
2624	7	PC	IgG	551	126,379	4,407
2624	35	PC	IgG	839	117,360	5,673
2624	175	PC	IgG	690	107,392	5,016
2752	7	PC	IgG	613	95,954	3,942
2752	35	PC	IgG	1,814	114,910	4,994
2752	63	PC	IgG	420	98,120	3,473
2954	7	PC	IgG	977	126,511	6,380
2954	35	PC	IgG	1,283	118,926	5,984
2954	175	PC	IgG	702	112,572	4,754

9.4 Appendix to Chapter 6

9.4.1 Figures

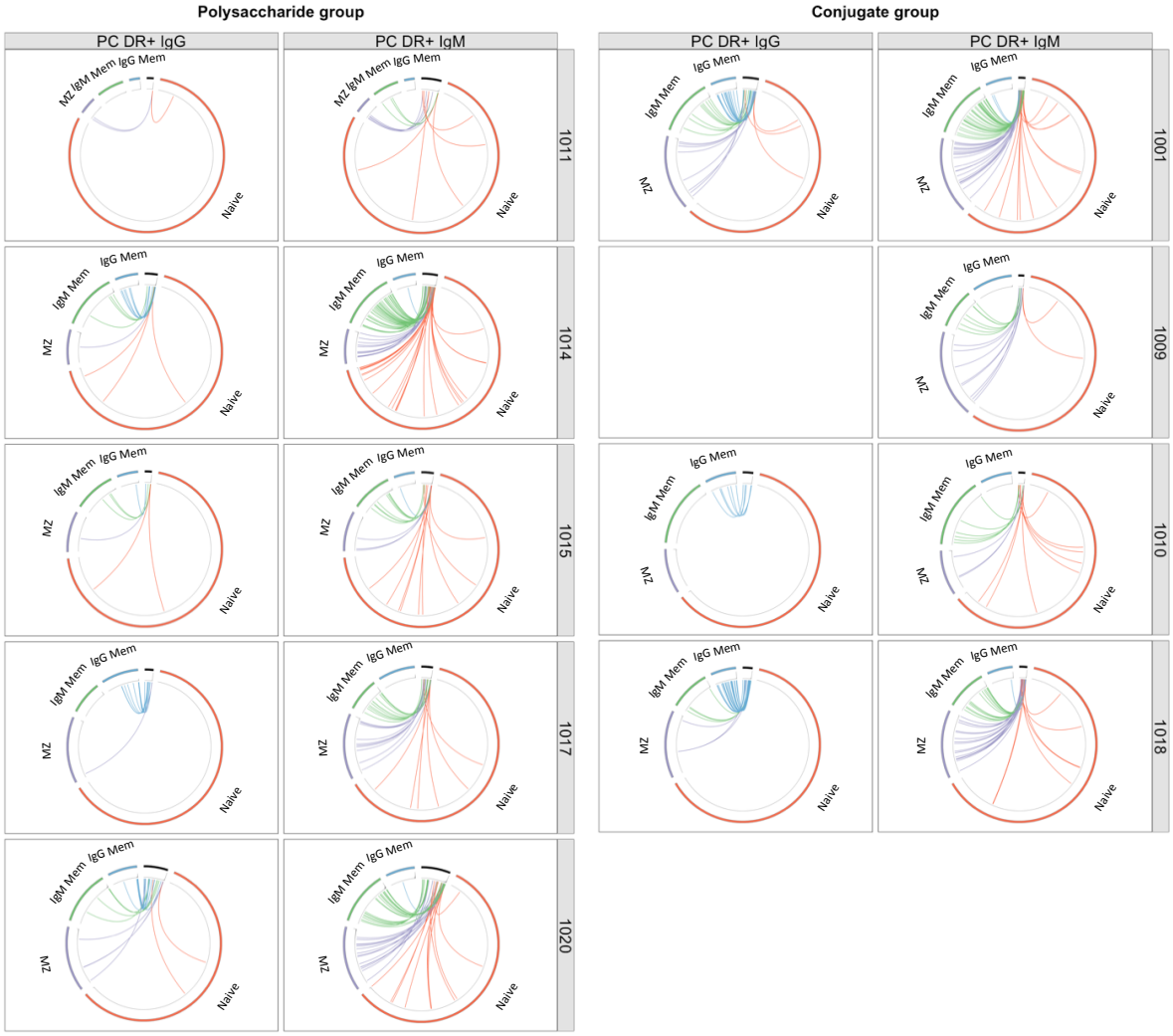

Figure 9.13: Relationship between baseline subsets and PC's - Circos diagrams from all participants, showing the relationship between baseline subsets, and the HLA-DR+ IgM and HLA-DR+ IgG PC's after the first vaccination. The length of each section represents the relative number of clusters in that subset. The black section represents the PC's, and the coloured sections represent the baseline subsets. Clusters are ordered clockwise by size, which is represented by the grey histogram. Coloured lines join clusters that are present in the PC's, and any one of the baseline subsets. Clusters shared within the different baseline subsets are not shown.

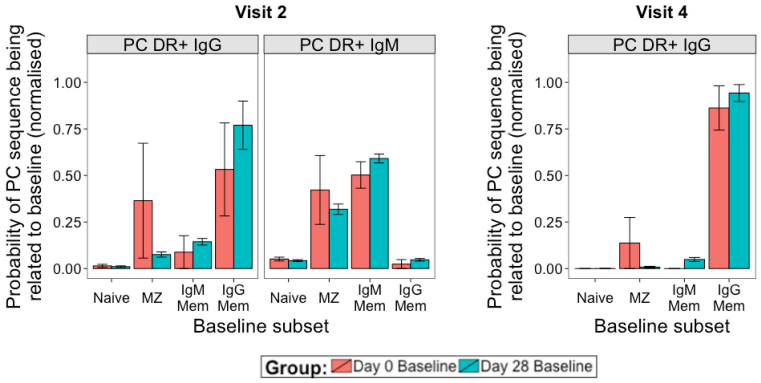

Figure 9.14: The effect of baseline visit on PC relationships - Probability of shared PC clusters being shared with each baseline cell subset for the HLA-DR+ IgG and IgM PC's at visit 2 after the first vaccine, and for HLA-DR+ IgG PC's at visit 4, after the second vaccine. Groups correspond to participants where baseline subsets were isolated at day 0 (N = 5) versus day 28 (N = 4). Error bars indicate ±SEM.

9.4.2 Tables

Table 9.7: Summary of the samples used for sequencing, and sequence data obtained.

Participant	Visit	Cells	Isotype	Cell number	Raw seqs	Filtered seqs	Clusters
1001	2	PC DR+	IgG	34,000	131,585	73,360	2,786
1001	2	PC DR-	IgG	1,400	120,966	73,869	1,167
1001	2	PC DR+	IgM	34,000	149,622	93,959	1,208
1001	2	PC DR-	IgM	1,400	140,271	113,928	1,211
1001	3	Naive	IgM	22,000	147,142	108,238	47,265
1001	3	MZ	IgM	17,000	137,404	108,214	12,887
1001	3	IgM Mem	IgM	19,000	143,281	110,264	9,673
1001	3	IgG Mem	IgG	30,000	118,486	74,469	4,406
1001	4	PC DR+	IgG	1,222	113,482	73,270	2,139
1009	2	PC DR-	IgG	3,134	130,565	65,223	1,566
1009	2	PC DR+	IgM	4,152	149,225	111,649	1,105
1009	2	PC DR-	IgM	3,134	142,544	112,549	2,277
1009	1	Naive	IgM	140,000	147,455	104,389	50,971
1009	1	MZ	IgM	19,000	140,810	110,489	16,124
1009	1	IgM Mem	IgM	14,000	150,452	118,839	7,281
1009	1	IgG Mem	IgG	44,000	116,925	66,685	7,382
1009	4	PC DR+	IgG	191	107,066	66,595	647
1010	2	PC DR+	IgG	2,369	109,273	65,857	1,371
1010	2	PC DR-	IgG	6,462	100,022	69,469	1,410
1010	2	PC DR+	IgM	2,369	133,201	104,043	891
1010	2	PC DR-	IgM	6,462	158,416	126,141	1,814
1010	1	Naive	IgM	70,000	143,386	101,837	37,903
1010	1	MZ	IgM	2,643	144,168	104,228	5,761
1010	1	IgM Mem	IgM	10,000	147,730	119,636	8,907
1010	1	IgG Mem	IgG	12,000	118,918	66,299	4,051
1010	4	PC DR+	IgG	110	116,052	76,475	436
1011	2	PC DR+	IgG	2,860	119,590	78,235	853
1011	2	PC DR-	IgG	10,000	106,711	79,434	893
1011	2	PC DR+	IgM	2,860	145,868	119,941	2,571
1011	2	PC DR-	IgM	10,000	106,686	81,546	1,046
1011	1	Naive	IgM	97,000	149,982	103,711	47,578
1011	1	MZ	IgM	1,224	147,823	118,588	2,175
1011	1	IgM Mem	IgM	1,888	139,335	106,526	3,365
1011	1	IgG Mem	IgG	2,940	122,018	79,345	1,308
1011	4	PC DR+	IgG	538	126,470	75,958	982
1014	2	PC DR+	IgG	1,984	119,126	88,336	1,724
1014	2	PC DR-	IgG	221	116,541	77,417	601
1014	2	PC DR+	IgM	1,984	142,845	109,075	2,076
1014	2	PC DR-	IgM	221	149,833	123,045	813
1014	3	Naive	IgM	100,000	158,575	110,940	42,659
1014	3	MZ	IgM	2,157	145,540	112,070	4,576
1014	3	IgM Mem	IgM	8,294	135,451	104,517	7,374
1014	3	IgG Mem	IgG	5,766	120,279	76,277	3,070
1014	4	PC DR+	IgG	7,491	100,654	67,928	1,562
1015	2	PC DR+	IgG	2,852	113,866	78,498	922
1015	2	PC DR-	IgG	9,081	121,421	73,844	1,285
1015	2	PC DR+	IgM	2,852	142,966	111,665	1,625
1015	2	PC DR-	IgM	9,081	154,623	115,372	1,457
1015	1	Naive	IgM	100,000	139,241	96,086	43,565
1015	1	MZ	IgM	2,787	119,250	93,460	5,220
1015	1	IgM Mem	IgM	2,783	143,934	113,632	5,343
1015	1	IgG Mem	IgG	100,000	114,550	73,642	2,714
1015	4	PC DR+	IgG	657	120,998	77,809	1,280
1017	2	PC DR+	IgG	3,025	105,639	70,821	821
1017	2	PC DR-	IgG	737	111,782	81,456	913
1017	2	PC DR+	IgM	3,025	139,101	112,616	1,091
1017	2	PC DR-	IgM	737	137,767	107,742	1,056
1017	1	Naive	IgM	64,000	134,352	106,141	28,237
1017	1	MZ	IgM	2,612	154,606	120,293	6,197
1017	1	IgM Mem	IgM	4,846	143,613	114,253	3,096
1017	1	IgG Mem	IgG	15,000	108,576	67,579	3,481
1017	4	PC DR+	IgG	1,765	120,886	71,007	1,443
1018	2	PC DR+	IgG	2,536	122,207	80,731	1,710
1018	2	PC DR-	IgG	905	127,309	74,675	1,209
1018	2	PC DR+	IgM	2,536	160,339	126,600	1,378
1018	2	PC DR-	IgM	905	145,834	117,478	862
1018	3	Naive	IgM	100,000	150,584	108,669	52,442
1018	3	MZ	IgM	11,000	164,406	120,101	11,660
1018	3	IgM Mem	IgM	9,015	154,232	118,549	6,973
1018	3	IgG Mem	IgG	30,000	111,747	75,530	4,378
1018	4	PC DR+	IgG	1,623	120,930	76,302	2,055
1020	2	PC DR+	IgG	50,000	113,195	63,911	1,712
1020	2	PC DR-	IgG	11,000	117,897	78,888	1,565
1020	2	PC DR+	IgM	50,000	157,507	124,651	2,078
1020	2	PC DR-	IgM	11,000	153,629	116,653	1,518
1020	3	Naive	IgM	12,000	150,115	111,326	19,245
1020	3	MZ	IgM	2,350	162,209	126,365	4,570
1020	3	IgM Mem	IgM	2,801	148,589	117,019	3,792
1020	3	IgG Mem	IgG	4,598	116,989	83,930	2,077
1020	4	PC DR+	IgG	1,514	117,743	86,175	1,662

9.5 Appendix to Chapter 7

9.5.1 Figures

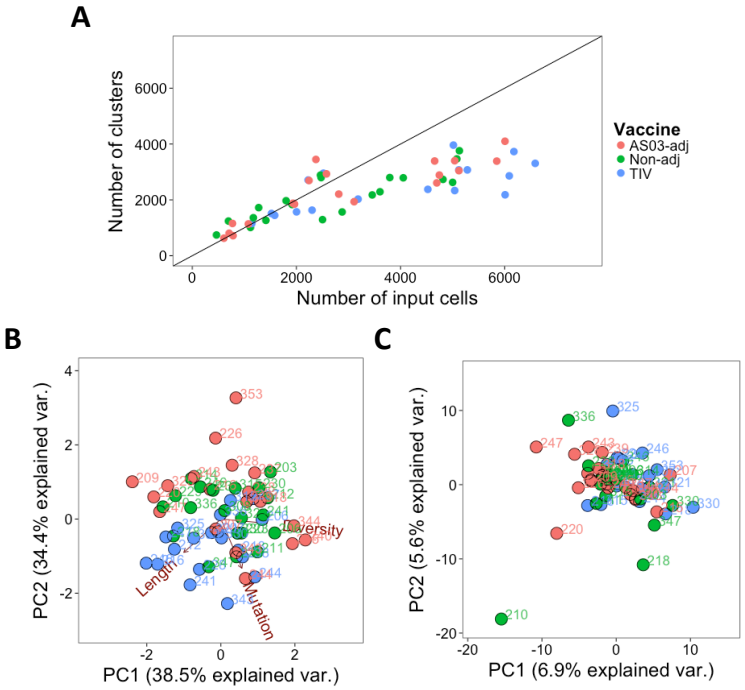

Figure 9.15: QC measures - (**A**) Relationship between the number of PCs isolated from each sample, and the number of clusters generated from the sequence data of that sample. Line represents x=y. (**B**) For each sample, the proportion of the repertoire comprised by each VJ recombinant was calculated. These data were used in a PCA to detect outliers. (**C**) For each sample, repertoire diversity, average V gene mutation, and average CDR3 AA length was determined. These data were used in a PCA to detect samples with signs of B cell activation.

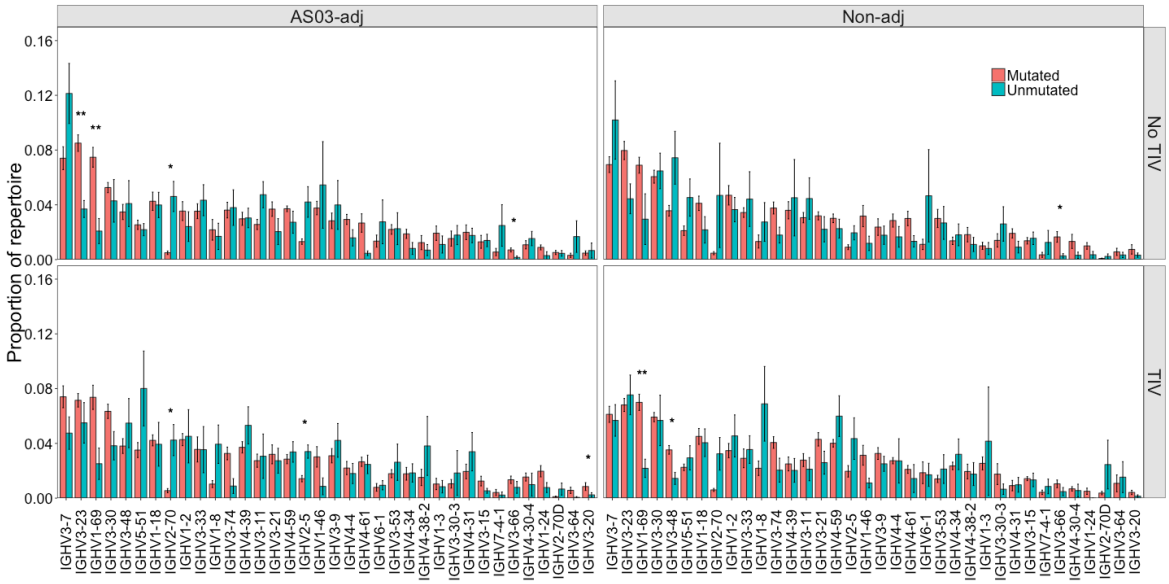

Figure 9.16: Comparing V gene usage of unmutated naive clusters to mutated clusters derived from memory recall split by vaccine group - For each sample, the proportion of both mutated and unmutated clusters utilizing different V gene segments was determined, and this was split by whether adjuvanted or non-adjuvanted vaccine was given, and whether there was prior TIV vaccination. Bars show mean values ±SEM. Shown are the V genes with a proportion above 0.005 in at least one of the groups. Comparisons were performed using a two-sided paired t-test. * $p < 0.01$, ** $p < 0.001$, *** $p < 0.0001$

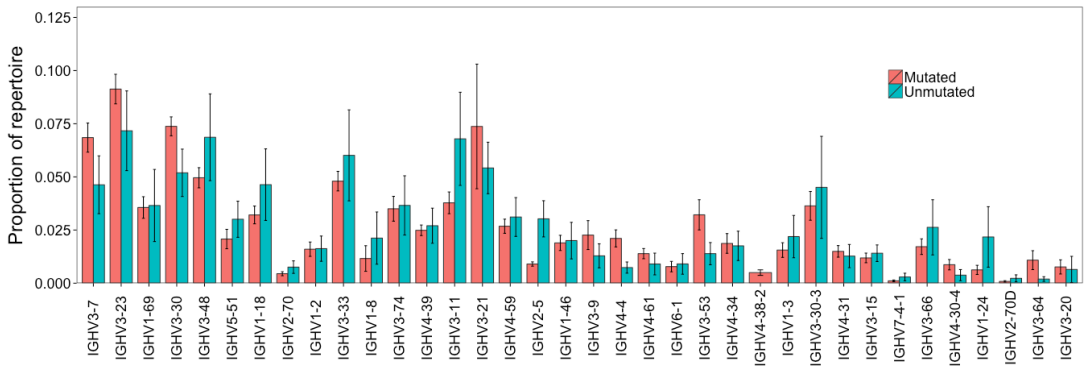

Figure 9.17: Comparing unmutated and mutated clusters derived from PCs in the absence of a specific immune stimulus. - Samples are derived from pre-vaccination samples from Chapters 4 and 5. For each sample, the proportion of both mutated and unmutated clusters utilizing different V gene segments was determined.

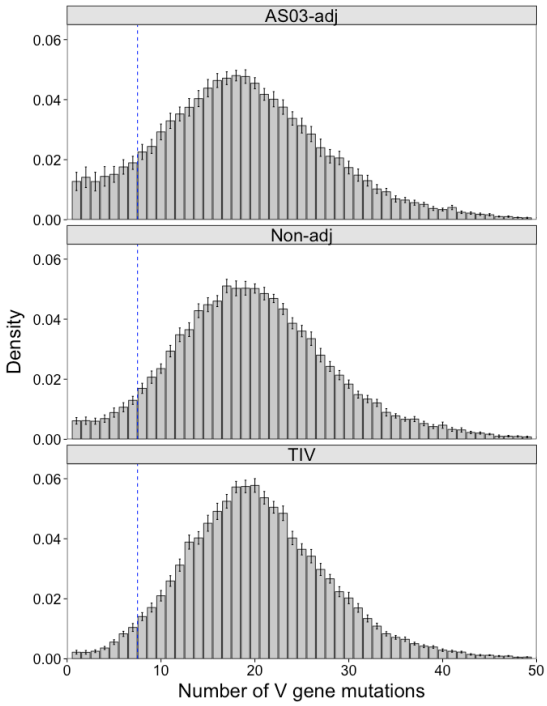

Figure 9.18: Distribution of clusters with different mean numbers of V gene mutations, split by vaccine group - Dotted vertical line separates clusters defined as unmutated or mutated (threshold of 7). Bars show mean values ±SEM.

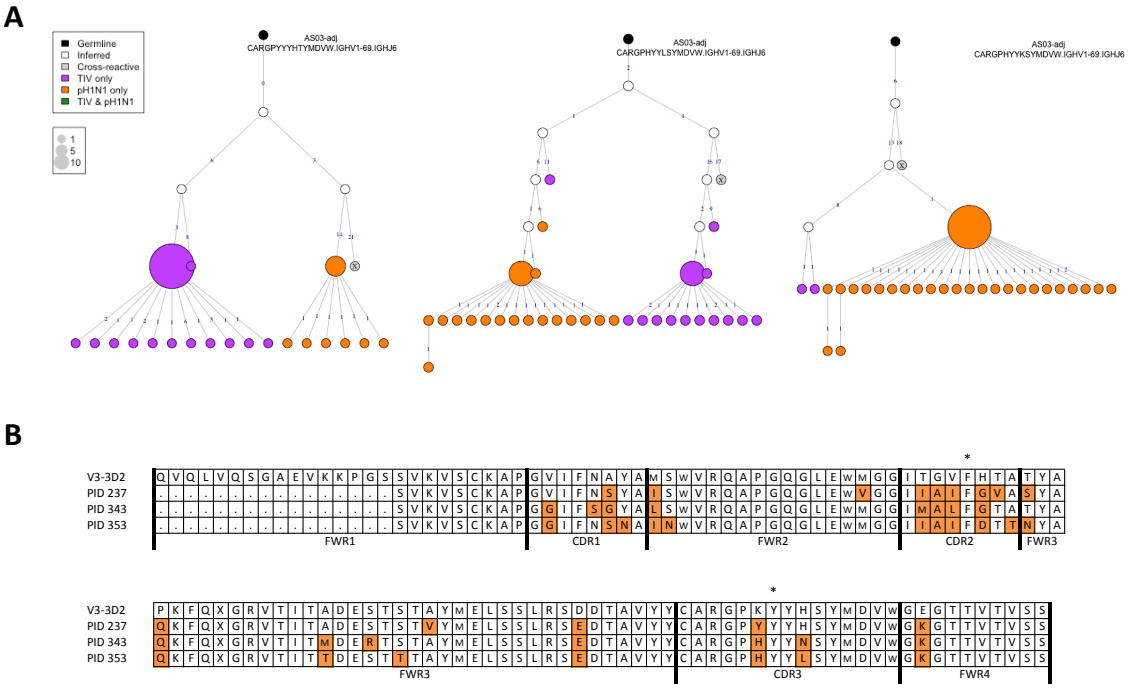

Figure 9.19: Similarity of the previously described cross-reactive sequence (V3-3DT) to sequences in our dataset - (**A**) Lineages generated from the combined TIV and pH1N1 data with similarity to V3-3DT. The V3-3DT sequence is included in the lineage generation. Each node in the lineage tree represents a unique sequence, and the size of the node represents the number of those sequences. The black node is the germline sequence, white nodes are inferred common ancestor sequences, the coloured nodes are those found in the datasets (purple = TIV only, orange = pH1N1), and the grey node with the 'X' is the V3-3DT sequence. Numbers on the edges of adjoining nodes show the number of mutations separating the sequences. (**B**) Alignment of the V3-3DT AA sequence to the closest match from the three lineages. Orange squares represent differences. * Residues with particular importance for antigen binding.

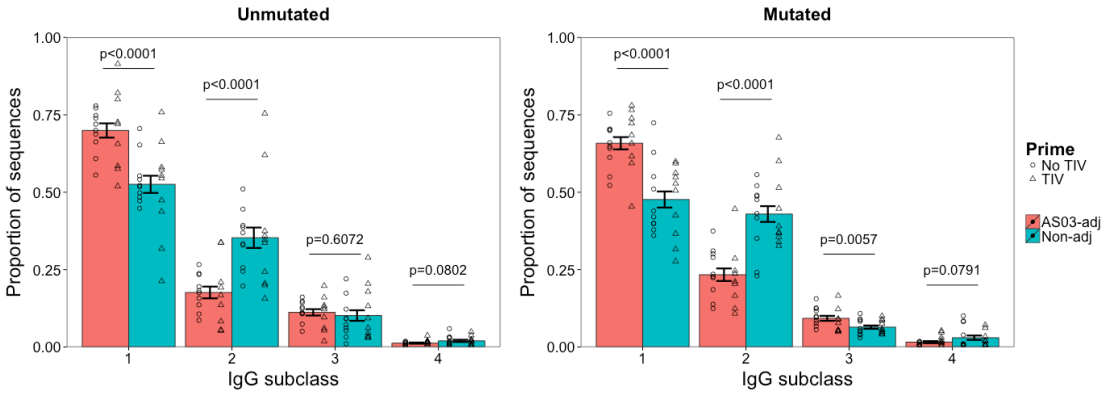

Figure 9.20: Analysis of IgG subclass usage split by mutated and unmutated sequences - For each participant, the clusters were split according to mutation level (less than 8 mutations = unmutated, and more than 8 = mutated), and the proportion of sequences of each IgG subclass was determined. Bars show mean values ±SEM. P values represent the result from a two-sided t-test. Individual data points are shown, with the shape representative of whether the participant received the seasonal TIV prior to pH1N1 vaccination.

9.5.2 Tables

Table 9.8: Number of cells, raw sequence reads, filtered sequence reads, and clusters obtained for each plasma cell sample.

Participant	Vaccine	Cell number	Raw Sequences	Filtered sequences	Clusters
207	AS03-adj	1,963	311,089	144,242	1,849
237	AS03-adj	4,745	381,734	193,031	2,888
238	AS03-adj	2,241	329,443	164,355	2,693
244	AS03-adj	1,933	315,791	155,593	1,879
247	AS03-adj	606	343,785	178,377	627
343	AS03-adj	3,108	310,611	150,935	1,935
353	AS03-adj	5,121	300,489	161,379	3,046
216	AS03-adj	2,813	286,853	139,327	2,207
325	AS03-adj	707	355,057	189,135	804
209	AS03-adj	782	318,189	160,831	718
215	AS03-adj	6,006	383,915	190,962	4,099
220	AS03-adj	771	326,679	161,972	1,156
226	AS03-adj	5,848	336,676	170,909	3,388
239	AS03-adj	4,696	227,580	112,361	2,608
240	AS03-adj	4,655	325,885	153,925	3,392
328	AS03-adj	2,577	290,798	136,686	2,931
344	AS03-adj	5,040	287,659	149,208	3,402
243	AS03-adj	1,076	374,145	150,293	1,138
318	AS03-adj	2,374	365,403	184,945	3,448
206	Non-adj	4,815	358,695	173,032	2,733
212	Non-adj	3,605	347,768	167,570	2,290
218	Non-adj	2,501	338,607	151,708	1,292
228	Non-adj	1,275	302,111	138,733	1,722
241	Non-adj	3,793	320,134	162,093	2,799
246	Non-adj	4,052	340,498	168,373	2,790
315	Non-adj	1,916	339,940	176,107	1,832
321	Non-adj	3,456	330,906	162,663	2,177
330	Non-adj	5,000	301,945	144,793	2,627
336	Non-adj	1,117	253,429	142,586	1,012
203	Non-adj	2,474	268,607	126,200	2,797
210	Non-adj	464	371,013	193,104	743
211	Non-adj	694	266,428	129,160	1,241
223	Non-adj	1,172	342,051	156,535	1,361
229	Non-adj	1,413	335,175	160,619	1,266
230	Non-adj	2,461	264,889	136,836	2,885
231	Non-adj	1,801	320,018	162,852	1,970
308	Non-adj	5,130	369,861	178,008	3,760
314	Non-adj	2,878	382,207	204,088	1,567
347	Non-adj	5,082	378,231	185,098	3,469
206	TIV	6,177	386,565	190,146	3,729
212	TIV	3,184	282,988	143,753	2,027
218	TIV	4,524	331,594	174,269	2,377
228	TIV	1,521	265,217	135,653	1,513
241	TIV	5,040	334,283	169,037	2,334
246	TIV	6,008	284,785	148,185	2,182
315	TIV	2,002	320,567	144,717	1,571
321	TIV	5,119	358,420	174,237	3,072
330	TIV	2,518	346,997	184,586	2,956
336	TIV	10,651	377,916	185,996	3,522
207	TIV	6,088	341,042	184,384	2,860
237	TIV	5,015	395,822	197,720	3,960
238	TIV	6,586	402,635	196,310	3,307
244	TIV	2,225	290,349	135,800	2,710
247	TIV	2,300	359,212	178,606	1,632
343	TIV	10,003	382,209	196,999	4,528
353	TIV	5,283	399,934	212,514	3,073
216	TIV	1,582	273,374	142,088	1,447
325	TIV	1,144	263,354	141,696	1,155

References

[1] Cerutti, A., Cols, M. and Puga, I. **Marginal zone B cells: virtues of innate-like antibody-producing lymphocytes.** *Nat. Rev. Immunol.*, **13**(2):118–32, 2013.

[2] Fried, A.J. and Bonilla, F.A. **Pathogenesis, diagnosis, and management of primary antibody deficiencies and infections**. *Clin. Microbiol. Rev.*, **22**(3):396–414, 2009.

[3] Corti, D. and Lanzavecchia, A. **Broadly neutralizing antiviral antibodies.** *Annu. Rev. Immunol.*, **31**:705–42, 2013.

[4] Roux, K.H. **Immunoglobulin structure and function as revealed by electron microscopy**. *Int. Arch. Allergy Immunol.*, **120**(2):85–99, 1999.

[5] Al-Lazikani, B., Lesk, A. and Chothia, C. **Standard conformations for the canonical structures of immunoglobulins**. *J. Mol. Biol.*, **273**(4):927–48, 1997.

[6] North, B., Lehmann, A. and Dunbrack Jr., R.L. **A new clustering of antibody CDR loop conformations**. *J. Mol. Biol.*, **406**(2):228–256, 2011.

[7] Murphy, K. *Janeway's Immunobiology*. Garland Science, 8 edition, 2011.

[8] Schroeder, H.W. and Cavacini, L. **Structure and Function of immunoglobulins**. *J. allergy Clin. Immunol.*, **125**(2 0 2):S41–52, 2013.

[9] Geisberger, R., Lamers, M. and Achatz, G. **The riddle of the dual expression of IgM and IgD.** *Immunology*, **118**(4):429–37, 2006.

[10] Irani, V., Guy, A.J., Andrew, D., Beeson, J.G., Ramsland, P.A. et al. **Molecular properties of human IgG subclasses and their implications for designing therapeutic monoclonal antibodies against infectious diseases**. *Mol. Immunol.*, **67**(2 Pt A):171–82, 2015.

[11] Woof, J.M. and Mestecky, J. **Mucosal immunoglobulins**. *Mucosal Immunol.*, **206**(1):64–82, 2005.

[12] Bouvet, J.P. and Dighiero, G. **From natural polyreactive autoantibodies to a la carte monoreactive antibodies to infectious agents: Is it a small world after all?** *Infect. Immun.*, **66**(1):1–4, 1998.

[13] Akira, S., Uematsu, S. and Takeuchi, O. **Pathogen Recognition and Innate Immunity**. *Cell*, **124**(4):783–801, 2006.

[14] Chen, X. and Jensen, P.E. **The role of B lymphocytes as antigen-presenting cells.** *Arch. Immunol. Ther. Exp. (Warsz).*, **56**(2):77–83, 2008.

[15] Coffman, R.L., Sher, A. and Seder, R.A. **Vaccine adjuvants: Putting innate immunity to work**. *Immunity*, **33**(4):492–503, 2010.

[16] Yuste, J., Sen, A., Truedsson, L., Jönsson, G., Tay, L.S. et al. **Impaired opsonization with C3b and phagocytosis of Streptococcus pneumoniae in sera from subjects with defects in the classical complement pathway**. *Infect. Immun.*, **76**(8):3761–3770, 2008.

[17] Haury, M., Sundblad, A., Grandien, A., Barreau, C., Coutinho, A. et al. **The repertoire of serum IgM in normal mice is largely independent of external antigenic contact**. *Eur. J. Immunol.*, **27**(6):1557–1563, 1997.

[18] Andersson, J., Sjöberg, O. and Möller, G. **Induction of immunoglobulin and antibody synthesis in vitro by lipopolysaccharides.** *Eur. J. Immunol.*, **2**(4):349–53, 1972.

[19] Wang, L., Radic, M.Z. and Galili, U. **Human anti-Gal heavy chain genes. Preferential use of VH3 and the presence of somatic mutations.** *J. Immunol.*, **155**(3):1276–1285, 1995.

[20] Hamadeh, R.M., Jarvis, G.A., Galili, U., Mandrell, R.E., Zhou, P. et al. **Human natural anti-Gal IgG regulates alternative complement pathway activation on bacterial surfaces.** *J. Clin. Invest.*, **89**(4):1223–35, 1992.

[21] Galili, U. **Anti-Gal: an abundant human natural antibody of multiple pathogeneses and clinical benefits.** *Immunology*, **140**(1):1–11, 2013.

[22] Glanville, J., Zhai, W., Berka, J., Telman, D., Huerta, G. et al. **Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire**. *Proc. Natl. Acad. Sci. U.S.A.*, **106**(48):20216–21, 2009.

[23] Boyd, S.D., Marshall, E.L., Merker, J.D., Maniar, J.M., Zhang, L.N. et al. **Measurement and Clinical Monitoring of Human Lymphocyte Clonality by Massively Parallel V-D-J Pyrosequencing**. *Sci. Transl. Med.*, **1**(12):12ra23, 2009.

[24] Arnaout, R., Lee, W., Cahill, P., Honan, T., Sparrow, T. et al. **High-resolution description of antibody heavy-chain repertoires in humans**. *PLoS One*, **6**(8):e22365, 2011.

[25] Vollmers, C., Sit, R., Weinstein, J.A., Dekker, C.L. and Quake, S.R. **Genetic measurement of memory B-cell recall using antibody repertoire sequencing**. *Proc. Natl. Acad. Sci. U.S.A.*, **110**(33):13463–8, 2013.

[26] Gallichotte, E.N., Widman, D.G., Yount, B.L., Wahala, W.M., Durbin, A. et al. **A new quaternary structure epitope on dengue virus serotype 2 is the target of durable type-specific neutralizing antibodies**. *MBio*, **6**(5):1–8, 2015.

[27] Xu, G.J., Kula, T., Xu, Q., Li, M.Z., Vernon, S.D. et al. **Comprehensive serological profiling of human populations using a synthetic human virome**. *Science (80-.).*, **348**(6239):aaa0698 1–9, 2015.

[28] Tarlinton, D. and Good-Jacobson, K. **Diversity among memory B cells: origin, consequences, and utility.** *Science*, **341**(6151):1205–11, 2013.

[29] Scheeren, F.A., Nagasawa, M., Weijer, K., Cupedo, T., Kirberg, J. et al. **T cell-independent development and induction of somatic hypermutation in human IgM+ IgD+ CD27+ B cells.** *J. Exp. Med.*, **205**(9):2033–2042, 2008.

[30] Covens, K., Verbinnen, B., Geukens, N., Meyts, I., Schuit, F. et al. **Characterization of proposed human B-1 cells reveals pre-plasmablast phenotype.** *Blood*, **121**(26):5176–83, 2013.

[31] Maecker, H.T., McCoy, J.P. and Nussenblatt, R. **Standardizing immunophenotyping for the Human Immunology Project.** *Nat. Rev. Immunol.*, **12**(3):191–200, 2012.

[32] Halliley, J.L., Tipton, C.M., Liesveld, J., Rosenberg, A.F., Darce, J. et al. **Long-Lived Plasma Cells Are Contained within the CD19CD38hiCD138+ Subset in Human Bone Marrow**. *Immunity*, **43**(1):132–145, 2015.

[33] Griffin, D.O., Holodick, N.E. and Rothstein, T.L. **Human B1 cells in umbilical cord and adult peripheral blood express the novel phenotype CD20+ CD27+ CD43+ CD70-.** *J. Exp. Med.*, **208**(1):67–80, 2011.

[34] Montecino-Rodriguez, E. and Dorshkind, K. **B-1 B Cell Development in the Fetus and Adult**. *Immunity*, **36**(1):13–23, 2012.

[35] Ho, F., Lortan, J.E., MacLennan, I.C. and Khan, M. **Distinct short-lived and long-lived antibody-producing cell populations.** *Eur. J. Immunol.*, **16**(10):1297–301, 1986.

[36] Martin, F., Oliver, A.M. and Kearney, J.F. **Marginal zone and B1 B cells unite in the early response against T-independent blood-borne particulate antigens**. *Immunity*, **14**(5):617–629, 2001.

[37] Berkowska, M., Driessen, G.J., Bikos, V., Grosserichter-wagener, C., Cerutti, A. et al. **Human memory B cells originate from three distinct germinal center-dependent and -independent maturation pathways**. *Blood*, **118**(8):2150–2159, 2011.

[38] Schatz, D.G. and Ji, Y. **Recombination centres and the orchestration of V(D)J recombination.** *Nat. Rev. Immunol.*, **11**(4):251–63, 2011.

[39] Tonegawa, S. **Somatic Generation of Antibody Diversity**. *Nature*, **302**(14):575–581, 1983.

[40] Oettinger, M.A., Schatz, D.G., Gorka, C. and Baltimore, D. **RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination.** *Science*, **248**(4962):1517–1523, 1990.

[41] Matsuda, F., Ishii, K., Bourvagnet, P., Kuma, K., Hayashida, H. et al. **The Complete Nucleotide Sequence of the Human Immunoglobulin Heavy Chain Variable Region Locus**. *J. Exp. Med.*, **188**(11):2151–2162, 1998.

[42] Mcbride, B.O.W., Hieter, P.A., Hollis, G.F., Swan, D., Otey, M.C. et al. **Chromosomal location of human kappa and lambda immunoglobulin light chain constant region genes**. *J. Exp. Med.*, **155**(5):1480–1490, 1982.

[43] Dunham, I., Shimizu, N., Roe, B.A., Chissoe, S., Hunt, A.R. et al. **The DNA sequence of human chromosome 22.** *Nature*, **402**(6761):489–495, 1999.

[44] Boyd, S.D., Gaëta, B.A., Jackson, K.J., Fire, A.Z., Marshall, E.L. et al. **Individual variation in the germline Ig gene repertoire inferred from variable region gene rearrangements**. *J. Immunol.*, **184**(12):6986–92, 2010.

[45] Fuentes-Pananá, E.M., Bannish, G. and Monroe, J.G. **Basal B-cell receptor signaling in B lymphocytes: Mechanisms of regulation and role in positive selection, differentiation, and peripheral survival**. *Immunol. Rev.*, **197**:26–40, 2004.

[46] Takeda, S., Sonoda, E. and Arakawa, H. **The kappa:lambda ratio of immature B cells.** *Immunol. Today*, **17**(4):200–1, 1996.

[47] Cornall, R.J., Goodnow, C.C. and Cyster, J.G. **The regulation of self-reactive B cells.** *Curr. Opin. Immunol.*, **7**(6):804–11, 1995.

[48] Roth, D.B., Menetski, J.P., Nakajima, P.B., Bosma, M.J. and Gellert, M. **V(D)J recombination: broken DNA molecules with covalently sealed (hairpin) coding ends in scid mouse thymocytes.** *Cell*, **70**(6):983–91, 1992.

[49] Nadel, B. and Feeney, A.J. **Nucleotide deletion and P addition in V(D)J recombination: a determinant role of the coding-end sequence.** *Mol. Cell. Biol.*, **17**(7):3768–3778, 1997.

[50] Jackson, K.J.L., Gaeta, B., Sewell, W. and Collins, A.M. **Exonuclease activity and P nucleotide addition in the generation of the expressed immunoglobulin repertoire.** *BMC Immunol.*, **5**:19, 2004.

[51] Melchers, F., ten Boekel, E., Seidl, T., Kong, X.C., Yamagami, T. et al. **Repertoire selection by pre-B-cell receptors and B-cell receptors, and genetic control of B-cell development from immature to mature B cells.** *Immunol. Rev.*, **175**:33–46, 2000.

[52] Casellas, R., Shih, T.A., Kleinewietfeld, M., Rakonjac, J., Nemazee, D. et al. **Contribution of receptor editing to the antibody repertoire.** *Science*, **291**(5508):1541–1544, 2001.

[53] Reth, M., Gehrmann, P., Petrac, E. and Wiese, P. **A novel VH to VHDJH joining mechanism in heavy-chain-negative (null) pre-B cells results in heavy-chain production.** *Nature*, **322**(6082):840–842, 1986.

[54] Covey, L.R., Ferrier, P. and Alt, F.W. **VH to VHDJH rearrangement is mediated by the internal VH heptamer.** *Int. Immunol.*, **2**(6):579–83, 1990.

[55] Zhang, Z., Zemlin, M., Wang, Y.H., Munfus, D., Huye, L.E. et al. **Contribution of VH gene replacement to the primary B cell repertoire**. *Immunity*, **19**(1):21–31, 2003.

[56] Lu, C.Y., Ni, Y.H., Chiang, B.L., Chen, P.J., Chang, M.H. et al. **Humoral and cellular immune responses to a hepatitis B vaccine booster 15-18 years after neonatal immunization.** *J. Infect. Dis.*, **197**(10):1419–26, 2008.

[57] Mond, J., Lees, A. and Snapper, C. **T Cell-Independent Antigens Type 2**. *Annu. Rev. Immunol.*, **13**:655–692, 1995.

[58] Vinuesa, C.G. and Chang, P.P. **Innate B cell helpers reveal novel types of antibody responses.** *Nat. Immunol.*, **14**(2):119–26, 2013.

[59] Obukhanych, T.V. and Nussenzweig, M.C. **T-independent type II immune responses generate memory B cells.** *J. Exp. Med.*, **203**(2):305–10, 2006.

[60] Alugupalli, K.R., Leong, J.M., Woodland, R.T., Muramatsu, M., Honjo, T. et al. **B1b lymphocytes confer T cell-independent long-lasting immunity.** *Immunity*, **21**(3):379–90, 2004.

[61] Amanna, I.J. and Slifka, M.K. **Mechanisms that determine plasma cell lifespan and the duration of humoral immunity.** *Immunol. Rev.*, **236**:125–38, 2010.

[62] Parker, D. **T Cell-Dependent B Cell Activation**. *Annu. Rev. Immunol.*, **11**:331–60, 1993.

[63] Ochsenbein, A.F., Pinschewer, D.D., Sierro, S., Horvath, E., Hengartner, H. et al. **Protective long-term antibody memory by antigen-driven and T help-dependent differentiation of long-lived memory B cells to short-lived plasma cells independent of secondary lymphoid organs.** *Proc. Natl. Acad. Sci. U.S.A.*, **97**(24):13263–8, 2000.

[64] Liu, Y.J., Zhang, J., Lane, P.J., Chan, E.Y. and MacLennan, I.C. **Sites of specific B cell activation in primary and secondary responses to T cell-dependent and T cell-independent antigens**. *Eur. J. Immunol.*, **21**(12):2951–62, 1991.

[65] Bachmann, M.F., Odermatt, B., Hengartner, H. and Zinkernagel, R.M. **Induction of long-lived germinal centers associated with persisting antigen after viral infection**. *J. Exp. Med.*, **183**(5):2259–69, 1996.

[66] Victora, G.D. and Nussenzweig, M.C. **Germinal centers.** *Annu. Rev. Immunol.*, **30**:429–57, 2012.

[67] Batrak, V., Blagodatski, A. and Buerstedde, J.M. **Understanding the immunoglobulin locus specificity of hypermutation.** *Methods Mol. Biol.*, **745**:311–26, 2011.

[68] Holder, M.J., Wang, H., Milner, A.E., Casamayor, M., Armitage, R. et al. **Suppression of apoptosis in normal and neoplastic human B lymphocytes by CD40 ligand is independent of Bc1-2 induction.** *Eur. J. Immunol.*, **23**(9):2368–71, 1993.

[69] Gojobori, T. and Nei, M. **Relative contributions of germline gene variation and somatic mutation to immunoglobulin diversity in the mouse.** *Mol. Biol. Evol.*, **3**(2):156–67, 1986.

[70] Zotos, D. and Tarlinton, D.M. **Determining germinal centre B cell fate**. *Trends Immunol.*, **33**(6):281–288, 2012.

[71] Phan, T.G., Paus, D., Chan, T.D., Turner, M.L., Nutt, S.L. et al. **High affinity germinal center B cells are actively selected into the plasma cell compartment.** *J. Exp. Med.*, **203**(11):2419–24, 2006.

[72] Stavnezer, J., Guikema, J.E.J. and Schrader, C.E. **Mechanism and regulation of class switch recombination.** *Annu. Rev. Immunol.*, **26**:261–92, 2008.

[73] Chaudhuri, J. and Alt, F.W. **Class-switch recombination: interplay of transcription, DNA deamination and DNA repair.** *Nat. Rev. Immunol.*, **4**(7):541–552, 2004.

[74] Ehreth, J. **The global value of vaccination**. *Vaccine*, **21**(7-8):596–600, 2003.

[75] Crotty, S. and Ahmed, R. **Immunological memory in humans**. *Semin. Immunol.*, **16**(3):197–203, 2004.

[76] Slifka, M.K., Antia, R., Whitmire, J.K. and Ahmed, R. **Humoral immunity due to long-lived plasma cells**. *Immunity*, **8**(3):363–372, 1998.

[77] Crotty, S., Felgner, P., Davies, H., Glidewell, J., Villarreal, L. et al. **Cutting Edge: Long-Term B Cell Memory in Humans after Smallpox Vaccination**. *J. Immunol.*, **171**(10):4969–4973, 2003.

[78] Ward, S.M., Phalora, P., Bradshaw, D., Leyendeckers, H. and Klenerman, P. **Direct ex vivo evaluation of long-lived protective antiviral memory B cell responses against hepatitis B virus.** *J. Infect. Dis.*, **198**(6):813–7, 2008.

[79] Pulendran, B. and Ahmed, R. **Immunological mechanisms of vaccination**. *Nat. Rev. Immunol.*, **12**(6):509–517, 2011.

[80] Poulsen, T.R., Jensen, A., Haurum, J.S. and Andersen, P.S. **Limits for antibody affinity maturation and repertoire diversification in hypervaccinated humans**. *J. Immunol.*, **187**(8):4229–35, 2011.

[81] Heath, P.T., Booy, R., Azzopardi, H.J., Slack, M.P., Bowen-Morris, J. et al. **Antibody concentration and clinical protection after Hib conjugate vaccination in the United Kingdom.** *JAMA*, **284**(18):2334–40, 2000.

[82] Bernasconi, N.L., Traggiai, E. and Lanzavecchia, A. **Maintenance of serological memory by polyclonal activation of human memory B cells**. *Science (80-.).*, **298**(5601):2199–202, 2002.

[83] Plotkin, S.A. **Vaccines: correlates of vaccine-induced immunity**. *Clin. Infect. Dis.*, **47**(3):401–9, 2008.

[84] McElhaney, J.E., Xie, D., Hager, W.D., Barry, M.B., Wang, Y. et al. **T cell responses are better correlates of vaccine protection in the elderly.** *J. Immunol.*, **176**(10):6333–9, 2006.

[85] Maslanka, S.E., Tappero, J.W., Plikaytis, B.D., Brumberg, R.S., Dykes, J.K. et al. **Age-dependent Neisseria meningitidis serogroup C class-specific antibody concentrations and bactericidal titers in sera from young children from Montana immunized with a licensed polysaccharide vaccine.** *Infect. Immun.*, **66**(6):2453–9, 1998.

[86] De Wals, P., De Serres, G. and Niyonsenga, T. **Effectiveness of a mass immunization campaign against serogroup C meningococcal disease in Quebec.** *JAMA*, **285**(2):177–81, 2001.

[87] Kelly, D.F., Snape, M.D., Cutterbuck, E.A., Green, S., Snowden, C. et al. **CRM197-conjugated serogroup C meningococcal capsular polysaccharide, but not the native polysaccharide, induces persistent antigen-specific memory B cells.** *Blood*, **108**(8):2642–7, 2006.

[88] O'Connor, D. and Pollard, A.J. **Characterizing vaccine responses using host genomic and transcriptomic analysis**. *Clin. Infect. Dis.*, **57**(6):860–9, 2013.

[89] Kelly, D.F., Snape, M.D., Perrett, K.P., Clutterbuck, E.A., Lewis, S. et al. **Plasma and memory B-cell kinetics in infants following a primary schedule of CRM 197-conjugated serogroup C meningococcal polysaccharide vaccine.** *Immunology*, **127**(1):134–43, 2009.

[90] Blanchard-Rohner, G., Pulickal, A.S., Jol-van der Zijde, C.M., Snape, M.D. and Pollard, A.J. **Appearance of peripheral blood plasma cells and memory B cells in a primary and secondary immune response in humans.** *Blood*, **114**(24):4998–5002, 2009.

[91] Mitchell, R., Kelly, D.F., Pollard, A.J. and Trück, J. **Polysaccharide-specific B cell responses to vaccination in humans**. *Hum. Vaccin. Immunother.*, **10**(6):1661–1668, 2014.

[92] Amanna, I.J., Carlson, N.E. and Slifka, M.K. **Duration of humoral immunity to common viral and vaccine antigens.** *N. Engl. J. Med.*, **357**(19):1903–1915, 2007.

[93] Clutterbuck, E.A., Lazarus, R., Yu, L.M., Bowman, J., Bateman, E.A.L. et al. **Pneumococcal conjugate and plain polysaccharide vaccines have divergent effects on antigen-specific B cells.** *J. Infect. Dis.*, **205**(9):1408–16, 2012.

[94] Jokhdar, H., Borrow, R., Sultan, A., Adi, M., Riley, C. et al. **Immunologic hyporesponsiveness to serogroup C but not serogroup A following repeated meningococcal A/C polysaccharide vaccination in Saudi Arabia.** *Clin. Diagn. Lab. Immunol.*, **11**(1):83–8, 2004.

[95] Henn, A.D., Wu, S., Qiu, X., Ruda, M., Stover, M. et al. **High-resolution temporal response patterns to influenza vaccine reveal a distinct human plasma cell gene signature.** *Sci. Rep.*, **3**(2327):1–12, 2013.

[96] Li, S., Rouphael, N., Duraisingham, S., Romero-Steiner, S., Presnell, S. et al. **Molecular signatures of antibody responses derived from a systems biology study of five human vaccines.** *Nat. Immunol.*, **15**(2):195–204, 2014.

[97] Lavinder, J.J., Wine, Y., Giesecke, C., Ippolito, G.C., Horton, A.P. et al. **Identification and characterization of the constituent human serum antibodies elicited by vaccination**. *Proc. Natl. Acad. Sci.*, **111**(6):2259–2264, 2014.

[98] Awdeh, Z.L., Williamson, A.R. and Askonas, B.A. **Isoelectric Focusing in Polyacrylamide Gel and its Application to Immunoglobulins**. *Nature*, **219**(5149):66–67, 1968.

[99] Ademokun, A., Wu, Y.C., Martin, V., Mitra, R., Sack, U. et al. **Vaccination-induced changes in human B-cell repertoire and pneumococcal IgM and IgA antibody at different ages**. *Aging Cell*, **10**(6):922–30, 2011.

[100] Seidman, J.G., Leder, A., Edgell, M.H., Polsky, F., Tilghman, S.M. et al. **Multiple related immunoglobulin variable-region genes identified by cloning and sequence analysis**. *Proc. Natl. Acad. Sci. U.S.A.*, **75**(8):3881–5, 1978.

[101] Adderson, E., Shackelford, P., Quinn, A., Wilson, P., Cunningham, M. et al. **Restricted immunoglobulin VH usage and VDJ combinations in the human response to Haemophilus influenzae type b capsular polysaccharide**. *J. Clin. Invest.*, **91**(6):2734–2743, 1993.

[102] Adderson, E.E., Shackelford, P.G., Quinn, A. and Carroll, W.L. **Restricted Ig H chain V gene usage in the human antibody response to Haemophilus influenzae type b capsular polysaccharide.** *J. Immunol.*, **147**(5):1667–74, 1991.

[103] Wilson, P.C. and Andrews, S.F. **Tools to therapeutically harness the human antibody response**. *Nat. Rev. Immunol.*, **12**(10):709–719, 2012.

[104] Wrammert, J., Smith, K., Miller, J., Langley, W.A., Kokko, K. et al. **Rapid cloning of high-affinity human monoclonal antibodies against influenza virus**. *Nature*, **453**(7195):667–71, 2008.

[105] Tan, Y.C., Blum, L.K., Kongpachith, S., Ju, C.H., Cai, X. et al. **High-throughput sequencing of natively paired antibody chains provides evidence for original antigenic sin shaping the antibody response to influenza vaccination**. *Clin. Immunol.*, **151**(1):55–65, 2014.

[106] Frölich, D., Giesecke, C. and Mei, H. **Secondary immunization generates clonally related antigen-specific plasma cells and memory B cells**. *J. Immunol.*, **185**(5):3103–3110, 2010.

[107] Pinchuk, G.V., Nottenburg, C. and Milner, E.C. **Predominant V-region gene configurations in the human antibody response to Haemophilus influenzae capsule polysaccharide.** *Scand. J. Immunol.*, **41**(4):324–30, 1995.

[108] Eren, R., Lubin, I., Terkieltaub, D., Ben-Moshe, O., Zauberman, a. et al. **Human monoclonal antibodies specific to hepatitis B virus generated in a human/mouse radiation chimera: The Trimera system**. *Immunology*, **93**(2):154–161, 1998.

[109] Jin, A., Ozawa, T., Tajiri, K., Obata, T., Kondo, S. et al. **A rapid and efficient single-cell manipulation method for screening antigen-specific antibody-secreting cells from human peripheral blood.** *Nat. Med.*, **15**(9):1088–1092, 2009.

[110] Sato, S., Beausoleil, S.a., Popova, L., Beaudet, J.G., Ramenani, R.K. et al. **Proteomics-directed cloning of circulating antiviral human monoclonal antibodies.** *Nat. Biotechnol.*, **30**(11):1039–43, 2012.

[111] Zebedee, S.L., Barbas, C.F., Hom, Y.L., Caothien, R.H., Graff, R. et al. **Human combinatorial antibody libraries to hepatitis B surface antigen.** *Proc. Natl. Acad. Sci. U. S. A.*, **89**(8):3175–3179, 1992.

[112] Galson, J.D., Pollard, A.J., Trück, J. and Kelly, D.F. **Studying the antibody repertoire after vaccination: practical applications**. *Trends Immunol.*, **35**(7):319–331, 2014.

[113] Lucas, A.H., Larrick, J.W. and Reason, D.C. **Variable region sequences of a protective human monoclonal antibody specific for the Haemophilus influenzae type b capsular polysaccharide.** *Infect. Immun.*, **62**(9):3873–80, 1994.

[114] Lucas, A.H. and Reason, D.C. **Polysaccharide vaccines as probes of antibody repertoires in man**. *Immunol. Rev.*, **171**(1):89–104, 1999.

[115] Shendure, J. and Ji, H. **Next-generation DNA sequencing**. *Nat. Biotechnol.*, **26**(10):1135–45, 2008.

[116] Bolotin, D.A., Mamedov, I.Z., Britanova, O.V., Zvyagin, I.V., Shagin, D. et al. **Next generation sequencing for TCR repertoire profiling: Platform-specific features and correction algorithms.** *Eur. J. Immunol.*, **42**(11):3073–3083, 2012.

[117] Loman, N.J., Misra, R.V., Dallman, T.J., Constantinidou, C., Gharbia, S.E. et al. **Performance comparison of benchtop high-throughput sequencing platforms.** *Nat. Biotechnol.*, **30**(5):434–9, 2012.

[118] Parameswaran, P., Liu, Y., Roskin, K.M., Jackson, K.K., Dixit, V.P. et al. **Convergent Antibody Signatures in Human Dengue**. *Cell Host Microbe*, **13**(6):691–700, 2013.

[119] Jiang, N., He, J., Weinstein, J.A., Penland, L., Sasaki, S. et al. **Lineage structure of the human antibody repertoire in response to influenza vaccination**. *Sci. Transl. Med.*, **5**(171):171ra19, 2013.

[120] Klein, U., Küppers, R. and Rajewsky, K. **Evidence for a large compartment of IgM-expressing memory B cells in humans.** *Blood*, **89**(4):1288–98, 1997.

[121] van Dongen, J.J.M., Langerak, A.W., Brüggemann, M., Evans, P.A.S., Hummel, M. et al. **Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936.** *Leukemia*, **17**(12):2257–317, 2003.

[122] Wu, Y.C., Kipling, D., Leong, H.S., Martin, V., Ademokun, A.A. et al. **High-throughput immunoglobulin repertoire analysis distinguishes between human IgM memory and switched memory B-cell populations**. *Blood*, **116**(7):1070–8, 2010.

[123] Vander Heiden, J.A., Yaari, G., Uduman, M., Stern, J.N.H., O'Connor, K.C. et al. **pRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires.** *Bioinformatics*, **30**(13):1930–2, 2014.

[124] Galson, J.D., Clutterbuck, E.A., Trück, J., Ramasamy, M.N., Münz, M. et al. **BCR repertoire sequencing: different patterns of B cell activation after two Meningococcal vaccines**. *Immunol. Cell Biol.*, **93**(10):885–95, 2015.

[125] Shugay, M., Britanova, O.V., Merzlyak, E.M., Turchaninova, M.a., Mamedov, I.Z. et al. **Towards error-free profiling of immune repertoires.** *Nat. Methods*, **11**:653–655, 2014.

[126] Gaëta, B.A., Malming, H.R., Jackson, K.J.L., Bain, M.E., Wilson, P. et al. **iHMMune-align: hidden Markov model-based alignment and identification of germline genes in rearranged immunoglobulin gene sequences.** *Bioinformatics*, **23**(13):1580–7, 2007.

[127] Ye, J., Ma, N., Madden, T.L. and Ostell, J.M. **IgBLAST: an immunoglobulin variable domain sequence analysis tool.** *Nucleic Acids Res.*, **41**(Web Server issue):W34–40, 2013.

[128] Volpe, J.M., Cowell, L.G. and Kepler, T.B. **SoDA: implementation of a 3D alignment algorithm for inference of antigen receptor recombinations.** *Bioinformatics*, **22**(4):438–44, 2006.

[129] Kuchenbecker, L., Nienen, M., Hecht, J., Avidan, U., Babel, N. et al. **IMSEQ - a fast and error aware approach to immunogenetic sequence analysis**. *Bioinforma. Advnaces*, **31**(18):2963–71, 2015.

[130] Brochet, X., Lefranc, M.P. and Giudicelli, V. **IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis.** *Nucleic Acids Res.*, **36**(Web Server issue):W503–8, 2008.

[131] Gadala-Maria, D., Yaari, G., Uduman, M. and Kleinstein, S.H. **Automated analysis of high-throughput B-cell sequencing data reveals a high frequency of novel immunoglobulin V gene segment alleles**. *Proc. Natl. Acad. Sci.*, **112**(8):E862–70, 2015.

[132] Yaari, G. and Kleinstein, S.H. **Practical guidelines for B-cell receptor repertoire sequencing analysis**. *Genome Med.*, **7**(1):121, 2015.

[133] Greiff, V., Miho, E., Menzel, U. and Reddy, S.T. **Bioinformatic and Statistical Analysis of Adaptive Immune Repertoires**. *Trends Immunol.*, **36**(11):738–749, 2015.

[134] Barak, M., Zuckerman, N.S., Edelman, H., Unger, R. and Mehr, R. **IgTree: creating Immunoglobulin variable region gene lineage trees.** *J. Immunol. Methods*, **338**(1-2):67–74, 2008.

[135] Greiff, V., Bhat, P., Cook, S.C., Menzel, U., Kang, W. et al. **A bioinformatic framework for immune repertoire diversity profiling enables detection of immunological status**. *Genome Med.*, **7**(1):3–5, 2015.

[136] Yaari, G., Uduman, M. and Kleinstein, S.H. **Quantifying selection in high-throughput Immunoglobulin sequencing data sets.** *Nucleic Acids Res.*, **40**(17):e134, 2012.

[137] Palanichamy, A., Apeltsin, L., Kuo, T.C., Sirota, M., Wang, S. et al. **Immunoglobulin class-switched B cells form an active immune axis between CNS and periphery in multiple sclerosis**. *Sci. Transl. Med.*, **6**(248):248ra106, 2014.

[138] Glanville, J., Kuo, T.C., von Büdingen, H.C., Guey, L., Berka, J. et al. **Naive antibody gene-segment frequencies are heritable and unaltered by chronic lymphocyte ablation**. *Proc. Natl. Acad. Sci. U.S.A.*, **108**(50):20066–71, 2011.

[139] Larimore, K., McCormick, M., Robins, H. and Greenberg, P. **Shaping of human germline IgH repertoires revealed by deep sequencing**. *J. Immunol.*, **189**(6):3221–30, 2012.

[140] Benichou, J., Glanville, J., Prak, E.T.L., Azran, R., Kuo, T.C. et al. **The restricted DH gene reading frame usage in the expressed human antibody repertoire is selected based upon its amino acid content.** *J. Immunol.*, **190**(11):5567–77, 2013.

[141] Briney, B.S., Willis, J.R., Hicar, M.D., Thomas, J.W. and Crowe, J.E. **Frequency and genetic characterization of V(DD)J recombinants in the human peripheral blood antibody repertoire.** *Immunology*, **137**(1):56–64, 2012.

[142] Boyd, S.D., Liu, Y., Wang, C., Martin, V. and Dunn-Walters, D.K. **Human lymphocyte repertoires in ageing**. *Curr. Opin. Immunol.*, **25**(4):511–5, 2013.

[143] Wang, C., Liu, Y., Xu, L.T., Jackson, K.J.L., Roskin, K.M. et al. **Effects of Aging, Cytomegalovirus Infection, and EBV Infection on Human B Cell Repertoires**. *J. Immunol.*, **192**(2):603–611, 2013.

[144] Markle, J.G. and Fish, E.N. **SeXX matters in immunity.** *Trends Immunol.*, **35**(3):97–104, 2013.

[145] Mroczek, E.S., Ippolito, G.C., Rogosch, T., Hoi, K.H., Hwangpo, T.A. et al. **Differences in the Composition of the Human Antibody Repertoire by B Cell Subsets in the Blood**. *Front. Immunol.*, **5**(March):1–14, 2014.

[146] Wu, Y.C.B., Kipling, D. and Dunn-Walters, D.K. **Age-Related Changes in Human Peripheral Blood IGH Repertoire Following Vaccination**. *Front. Immunol.*, **3**(July):1–12, 2012.

[147] Laserson, U., Vigneault, F., Gadala-Maria, D., Yaari, G., Uduman, M. et al. **High-resolution antibody dynamics of vaccine-induced immune responses.** *Proc. Natl. Acad. Sci. U. S. A.*, **111**(13):4928–33, 2014.

[148] Dekosky, B.J., Ippolito, G.C., Deschner, R.P., Lavinder, J.J., Wine, Y. et al. **High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire**. *Nat. Biotechnol.*, **31**(2):166–169, 2013.

[149] Jackson, K.J.L., Wang, Y. and Collins, A.M. **Human immunoglobulin classes and subclasses show variability in VDJ gene mutation levels.** *Immunol. Cell Biol.*, **92**(8):729–33, 2014.

[150] Trück, J., Ramasamy, M.N., Galson, J.D., Rance, R., Parkhill, J. et al. **Identification of Antigen-Specific B Cell Receptor Sequences Using Public Repertoire Analysis.** *J. Immunol.*, **194**(1):252–261, 2015.

[151] Saada, R., Weinberger, M., Shahaf, G. and Mehr, R. **Models for antigen receptor gene rearrangement: CDR3 length.** *Immunol. Cell Biol.*, **85**(4):323–332, 2007.

[152] Wang, Y., Jackson, K.J.L., Davies, J., Chen, Z., Gaeta, B.A. et al. **IgE-associated IGHV genes from venom and peanut allergic individuals lack mutational evidence of antigen selection.** *PLoS One*, **9**(2):e89730, 2014.

[153] Keller, M.A. and Stiehm, E.R. **Passive immunity in prevention and treatment of infectious diseases.** *Clin. Microbiol. Rev.*, **13**(4):602–14, 2000.

[154] Smith, K., Garman, L., Wrammert, J., Zheng, N.Y., Capra, J.D. et al. **Rapid generation of fully human monoclonal antibodies specific to a vaccinating antigen.** *Nat. Protoc.*, **4**(3):372–84, 2009.

[155] Reddy, S.T., Ge, X., Miklos, A.E., Hughes, R.A., Kang, S.H. et al. **Monoclonal antibodies isolated without screening by analyzing the variable-gene repertoire of plasma cells**. *Nat. Biotechnol.*, **28**(9):965–9, 2010.

[156] Smith, K., Muther, J.J., Duke, A.L., McKee, E., Zheng, N.Y. et al. **Fully human monoclonal antibodies from antibody secreting cells after vaccination with Pneumovax®23 are serotype specific and facilitate opsonophagocytosis**. *Immunobiology*, **218**(5):745–54, 2013.

[157] Büdingen, H.V. and Kuo, T. **B cell exchange across the blood-brain barrier in multiple sclerosis**. *J. Clin. Invest.*, **122**(12):4533–4543, 2012.

[158] Xiao, X., Chen, W., Feng, Y., Zhu, Z., Prabakaran, P. et al. **Germline-like predecessors of broadly neutralizing antibodies lack measurable binding to HIV-1 envelope glycoproteins: implications for evasion of immune responses and design of vaccine immunogens**. *Biochem. Biophys. Res. Commun.*, **390**(3):404–9, 2009.

[159] Kwong, P.D., Mascola, J.R. and Nabel, G.J. **Broadly neutralizing antibodies and the search for an HIV-1 vaccine: the end of the beginning.** *Nat. Rev. Immunol.*, **13**(9):693–701, 2013.

[160] Liao, H.X., Chen, X., Munshaw, S., Zhang, R., Marshall, D.J. et al. **Initial antibodies binding to HIV-1 gp41 in acutely infected subjects are polyreactive and highly mutated.** *J. Exp. Med.*, **208**(11):2237–49, 2011.

[161] Liao, H.X., Lynch, R., Zhou, T., Gao, F., Alam, S.M. et al. **Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus.** *Nature*, **496**(7446):469–76, 2013.

[162] Dimitrov, D.S. **Therapeutic antibodies, vaccines and antibodyomes**. *MAbs*, **2**(3):347–56, 2010.

[163] Sepúlveda, N., Daniel, C. and Carneiro, J. **Estimation of T-cell repertoire diversity and clonal size distribution by Poisson abundance models**. *J. Immunol. Methods*, **353**(1-2):124–137, 2010.

[164] Wiley, S.R., Raman, V.S., Desbien, A., Bailor, H.R., Bhardwaj, R. et al. **Targeting TLRs expands the antibody repertoire in response to a malaria vaccine**. *Sci. Transl. Med.*, **3**(93):93ra69, 2011.

[165] Khurana, S., Verma, N., Yewdell, J.W., Hilbert, A.K., Castellino, F. et al. **MF59 adjuvant enhances diversity and affinity of antibody-mediated immune response to pandemic influenza vaccines.** *Sci. Transl. Med.*, **3**(85):85ra48, 2011.

[166] Li, G.M., Chiu, C., Wrammert, J., McCausland, M., Andrews, S.F. et al. **Pandemic H1N1 influenza vaccine induces a recall response in humans that favors broadly cross-reactive memory B cells.** *Proc. Natl. Acad. Sci. U. S. A.*, **109**(23):9047–52, 2012.

[167] Jilg, W., Schmidt, M. and Deinhardt, F. **Vaccination against hepatitis B: comparison of three different vaccination schedules**. *J. Infect. Dis.*, **160**(5):766–9, 1989.

[168] Regules, J.A., Cicatelli, S.B., Bennett, J.W., Paolino, K.M., Twomey, P.S. et al. **Fractional Third and Fourth Dose of RTS , S / AS01 Malaria Candidate Vaccine : A Phase 2a Controlled Human Malaria Parasite Infection and Immunogenicity Study**. *J. Infect. Dis.*, **214**(5):762–71, 2016.

[169] Benichou, J., Ben-Hamo, R., Louzoun, Y. and Efroni, S. **Rep-Seq: uncovering the immunological repertoire through next-generation sequencing.** *Immunology*, **135**(3):183–91, 2012.

[170] Baum, P.D., Venturi, V. and Price, D.a. **Wrestling with the repertoire: The promise and perils of next generation sequencing for antigen receptors.** *Eur. J. Immunol.*, **42**(11):2834–9, 2012.

[171] Kidd, B.a., Peters, L.a., Schadt, E.E. and Dudley, J.T. **Unifying immunology with informatics and multiscale biology**. *Nat. Immunol.*, **15**(2):118–127, 2014.

[172] Greiff, V., Menzel, U., Haessler, U., Cook, S.C., Friedensohn, S. et al. **Quantitative assessment of the robustness of next-generation sequencing of antibody variable gene repertoires from immunized mice**. *BMC Immunol.*, **15**(40):1–14, 2014.

[173] Zhu, J., Ofek, G., Yang, Y., Zhang, B., Louder, M.K. et al. **Mining the antibodyome for HIV-1-neutralizing antibodies with next-generation sequencing and phylogenetic pairing of heavy/light chains**. *Proc. Natl. Acad. Sci. U.S.A.*, **110**(16):6470–5, 2013.

[174] DeKosky, B.J., Kojima, T., Rodin, A., Charab, W., Ippolito, G.C. et al. **In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire**. *Nat. Med.*, **21**(1):86–91, 2015.

[175] Team, R.D.C. **R: A language and environment for statistical computing**. *R Found. Stat. Comput. Vienna, Austria*, 2008.

[176] Lunter, G. and Goodson, M. **Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads**. *Genome Res.*, **21**(6):936–9, 2011.

[177] Wilming, L.G., Gilbert, J.G.R., Howe, K., Trevanion, S., Hubbard, T. et al. **The vertebrate genome annotation (Vega) database.** *Nucleic Acids Res.*, **36**(Database issue):D753–60, 2008.

[178] Longerich, S., Tanaka, A., Bozek, G., Nicolae, D. and Storb, U. **The very 5' end and the constant region of Ig genes are spared from somatic mutation because AID does not access these regions.** *J. Exp. Med.*, **202**(10):1443–1454, 2005.

[179] Jackson, K.J.L., Boyd, S., Gaëta, B.a. and Collins, A.M. **Benchmarking the performance of human antibody gene alignment utilities using a 454 sequence dataset**. *Bioinformatics*, **26**(24):3129–30, 2010.

[180] Xu, J.L. and Davis, M.M. **Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities.** *Immunity*, **13**(1):37–45, 2000.

[181] Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R. et al. **vegan: Community Ecology Package**. *R Packag. version 2.2.1*, 2015.

[182] Chao, A., Gotelli, N.J., Hsieh, T.C., Sander, E.L., Ma, K.H. et al. **Rarefaction and extrapolation with Hill numbers: A framework for sampling and estimation in species diversity studies**. *Ecol. Monogr.*, **84**(1):45–67, 2014.

[183] Hsieh, T.C., Ma, K.H. and Chao, A. **iNEXT online: interpolation and extrapolation**. *R Packag. version 1.0*, 2013.

[184] Vander Heiden, J.A. and Gupta, N. **alakazam: Immunoglobulin Clonal Lineage and Diversity Analysis**. *R Packag. version 0.2.0*, 2015.

[185] Csárdi, G. and Nepusz, T. **The igraph software package for complex network research**. *InterJournal Complex Syst.*, **1695**, 2006.

[186] Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. Springer; 1st ed. 2009. Corr. 3rd printing 2010 edition, 2009.

[187] Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R. et al. **Circos: an information aesthetic for comparative genomics**. *Genome Res.*, **19**(9):1639–45, 2009.

[188] Rechavi, E., Lev, A., Lee, Y.N., Simon, A.J., Yinon, Y. et al. **Timely and spatially regulated maturation of B and T cell repertoire during human fetal development**. *Sci. Transl. Med.*, **7**(276):1–11, 2015.

[189] Logan, A.C., Zhang, B., Narasimhan, B., Carlton, V., Zheng, J. et al. **Minimal residual disease quantification using consensus primers and high-throughput IGH sequencing predicts post-transplant relapse in chronic lymphocytic leukemia.** *Leukemia*, **27**(8):1659–65, 2013.

[190] Liberman, G., Benichou, J., Tsaban, L., Glanville, J. and Louzoun, Y. **Multi Step Selection in Ig H Chains is Initially Focused on CDR3 and Then on Other CDR Regions**. *Front. Immunol.*, **4**(September):1–10, 2013.

[191] Bashford-Rogers, R.J., Palser, A.L., Idris, S.F., Carter, L., Epstein, M. et al. **Capturing needles in haystacks: a comparison of B-cell receptor sequencing methods**. *BMC Immunol.*, **15**(1):29, 2014.

[192] Lindner, C., Thomsen, I., Wahl, B., Ugur, M., Sethi, M.K. et al. **Diversification of memory B cells drives the continuous adaptation of secretory antibodies to gut microbiota**. *Nat. Immunol.*, **16**(8), 2015.

[193] Roskin, K.M., Simchoni, N., Liu, Y., Lee, J.y., Seo, K. et al. **IgH sequences in common variable immune deficiency reveal altered B cell development and selection**. *Sci. Transl. Med.*, **7**(302), 2015.

[194] Gitlin, A.D., Shulman, Z. and Nussenzweig, M.C. **Clonal selection in the germinal centre by regulated proliferation and hypermutation.** *Nature*, **509**(7502):637–40, 2014.

[195] Racanelli, V., Brunetti, C., De Re, V., Caggiari, L., De Zorzi, M. et al. **Antibody V(h) repertoire differences between resolving and chronically evolving hepatitis C virus infections**. *PLoS One*, **6**(9):e25606, 2011.

[196] Michaelsen, T.E., Garred, P. and Aase, A. **Human IgG subclass pattern of inducing complement-mediated cytolysis depends on antigen concentration and to a lesser extent on epitope patchiness, antibody affinity and complement concentration**. *Eur. J. Immunol.*, **21**(1):11–6, 1991.

[197] Barrett, D.J. and Ayoub, E.M. **IgG2 subclass restriction of antibody to pneumococcal polysaccharides.** *Clin. Exp. Immunol.*, **63**(1):127–34, 1986.

[198] Lottenbach, K.R., Mink, C.M., Barenkamp, S.J., Anderson, E.L., Homan, S.M. et al. **Age-associated differences in immunoglobulin G1 (IgG1) and IgG2 subclass antibodies to pneumococcal polysaccharides following vaccination.** *Infect. Immun.*, **67**(9):4935–8, 1999.

[199] Grönwall, C., Vas, J. and Silverman, G.J. **Protective roles of natural IgM antibodies**. *Front. Immunol.*, **3**(66):1–10, 2012.

[200] Galili, U., Anaraki, F., Thall, A., Hill-Black, C. and Radic, M. **One percent of human circulating B lymphocytes are capable of producing the natural anti-Gal antibody**. *Blood*, **82**(8):2485–93, 1993.

[201] Lambert, P.H., Liu, M. and Siegrist, C.A. **Can successful vaccines teach us how to induce efficient protective immune responses?** *Nat. Med.*, **11**(4 Suppl):S54–62, 2005.

[202] Wang, C., Liu, Y., Cavanagh, M.M., Le Saux, S., Qi, Q. et al. **B-cell repertoire responses to varicella-zoster vaccination in human identical twins**. *Proc. Natl. Acad. Sci. U. S. A.*, **112**(2):500–5, 2014.

[203] Jackson, K.J., Liu, Y., Roskin, K.M., Glanville, J., Hoh, R.A. et al. **Human Responses to Influenza Vaccination Show Seroconversion Signatures and Convergent Antibody Rearrangements**. *Cell Host Microbe*, **16**(1):105–14, 2014.

[204] Mcheyzer-williams, L.J., Milpied, P.J., Okitsu, S.L. and Mcheyzer-williams, M.G. **Class-switched memory B cells remodel BCRs within secondary germinal centers**. *Nat. Immunol.*, **16**(3):1–12, 2015.

[205] Odendahl, M., Mei, H., Hoyer, B.F., Jacobi, A.M., Hansen, A. et al. **Generation of migratory antigen-specific plasma blasts and mobilization of resident plasma cells in a secondary immune response**. *Blood*, **105**(4):1614–21, 2005.

[206] Furman, D., Jojic, V., Sharma, S., Shen-orr, S.S., Angel, C.J.L. et al. **Cytomegalovirus infection enhances the immune response to influenza**. *Sci. Transl. Med.*, **7**(281), 2015.

[207] Tsioris, K., Gupta, N.T., Ogunniyi, A.O., Zimnisky, R.M., Qian, F. et al. **Neutralizing antibodies against West Nile virus identified directly from human B cells by single-cell analysis and next generation sequencing**. *Integr. Biol.*, **7**(12):1587–97, 2015.

[208] Chen, Z.J., Wheeler, C.J., Shi, W., Wu, A.J., Yarboro, C.H. et al. **Polyreactive antigen-binding B cells are the predominant cell type in the newborn B cell repertoire.** *Eur. J. Immunol.*, **28**(3):989–94, 1998.

[209] Matter, M., Mumprecht, S., Pinschewer, D.D., Pavelic, V., Yagita, H. et al. **Virus-induced polyclonal B cell activation improves protective CTL memory via retained CD27 expression on memory CTL**. *Eur J Immunol*, **35**(11):3229–3239, 2005.

[210] Shi, W., Liao, Y., Willis, S.N., Taubenheim, N., Inouye, M. et al. **Transcriptional profiling of mouse B cell terminal differentiation defines a signature for antibody-secreting plasma cells**. *Nat. Immunol.*, **16**(6):663–673, 2015.

[211] Kaplinsky, J., Li, A., Sun, A., Coffre, M., Koralov, S.B. et al. **Antibody repertoire deep sequencing reveals antigen-independent selection in maturing B cells**. *Proc. Natl. Acad. Sci.*, **111**(25):2622–9, 2014.

[212] Gold, R., Lepow, M.L., Goldschneider, I., Draper, T.L. and Gotschlich, E.C. **Clinical evaluation of group A and group C meningococcal polysaccharide vaccines in infants.** *J. Clin. Invest.*, **56**(6):1536–47, 1975.

[213] Weill, J.C., Weller, S. and Reynaud, C.A. **Human marginal zone B cells.** *Annu. Rev. Immunol.*, **27**:267–85, 2009.

[214] Ramasamy, M. **B cell responses to conjugate and polysaccharide meningococcal vaccines**. *PhD Thesis. Univ. Oxford, UK*, 2012.

[215] Ramasamy, M.N., Clutterbuck, E.A., Bowman, J., Omar, O., Amber, J. et al. **United Kingdom Randomized Clinical Trial To Evaluate the Immunogenicity of Quadrivalent Meningococcal Conjugate and Polysaccharide Vaccines**. *Clin. vaccine Immunol.*, 2014.

[216] Weller, S., Braun, M.C., Tan, B.K., Rosenwald, A., Cordier, C. et al. **Human blood IgM "memory" B cells are circulating splenic marginal zone B cells harboring a prediversified immunoglobulin repertoire.** *Blood*, **104**(12):3647–54, 2004.

[217] Medina, F., Segundo, C., Jiménez-Gómez, G., González-García, I., Campos-Caro, A. et al. **Higher maturity and connective tissue association distinguish resident from recently generated human tonsil plasma cells.** *J. Leukoc. Biol.*, **82**:1430–1436, 2007.

[218] Christensen, H., May, M., Bowen, L., Hickman, M. and Trotter, C.L. **Meningococcal carriage by age: a systematic review and meta-analysis.** *Lancet Infect. Dis.*, **10**(12):853–61, 2010.

[219] Hjelholt, A., Christiansen, G., Sørensen, U.S. and Birkelund, S. **IgG subclass profiles in normal human sera of antibodies specific to five kinds of microbial antigens.** *Pathog. Dis.*, **67**(3):206–13, 2013.

[220] Aalberse, R.C., Stapel, S.O., Schuurman, J. and Rispens, T. **Immunoglobulin G4: an odd antibody.** *Clin. Exp. Allergy*, **39**(4):469–77, 2009.

[221] Glezen, W.P. **Emerging Infections: Pandemic Influenza**. *Epidemiol. Rev.*, **18**(1):64–76, 1996.

[222] Bouvier, N.M. and Palese, P. **The biology of influenza viruses**. *Vaccine*, **26**(4):49–53, 2008.

[223] Dawood, F.S., Iuliano, A.D., Reed, C., Meltzer, M.I., Shay, D.K. et al. **Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: A modelling study**. *Lancet Infect. Dis.*, **12**(9):687–695, 2012.

[224] Igarashi, M., Ito, K., Yoshida, R., Tomabechi, D., Kida, H. et al. **Predicting the antigenic structure of the pandemic (H1N1) 2009 influenza virus hemagglutinin**. *PLoS One*, **5**(1):1–7, 2010.

[225] Lee, V.J., Tay, J.K., Chen, M.I.C., Phoon, M.C., Xie, M.L. et al. **Inactivated trivalent seasonal influenza vaccine induces limited cross-reactive neutralizing antibody responses against 2009 pandemic and 1934 PR8 H1N1 strains**. *Vaccine*, **28**(42):6852–6857, 2010.

[226] Johns, M.C., Eick, A.A., Blazes, D.L., Lee, S.E., Perdue, C.L. et al. **Seasonal influenza vaccine and protection against pandemic (H1N1) 2009-associated illness among US military personnel**. *PLoS One*, **5**(5):1–7, 2010.

[227] Roman, F., Clément, F., Dewé, W., Walravens, K., Maes, C. et al. **Effect on cellular and humoral immune responses of the AS03 adjuvant system in an A/H1N1/2009 influenza virus vaccine administered to adults during two randomized controlled trials.** *Clin. Vaccine Immunol.*, **18**(5):835–43, 2011.

[228] Roy-Ghanta, S., Van der Most, R., Li, P. and Vaughn, D.W. **Responses to A(H1N1)pdm09 Influenza Vaccines in Participants Previously Vaccinated With Seasonal Influenza Vaccine: A Randomized, Observer-Blind, Controlled Study.** *J. Infect. Dis.*, **210**(9):1419–30, 2014.

[229] Garçon, N., Chomez, P. and Mechelen, M.V. **GlaxoSmithKline adjuvant systems in vaccines : concepts , achievements and perspectives**. *Expert Rev. Vaccines*, **6**(5):723–739, 2007.

[230] O'Hagan, D.T., Ott, G.S., De Gregorio, E. and Seubert, A. **The mechanism of action of MF59 - An innately attractive adjuvant formulation**. *Vaccine*, **30**(29):4341–4348, 2012.

[231] van der Most, R. and Roman, F. **Seeking Help: B Cells Adapting to Flu Variability**. *Sci. Transl. Med.*, **6**(246):4–9, 2014.

[232] Khurana, S., Chearwae, W., Castellino, F., Manischewitz, J., King, L.R. et al. **Vaccines with MF59 adjuvant expand the antibody repertoire to target protective sites of pandemic avian H5N1 influenza virus.** *Sci. Transl. Med.*, **2**(15):15ra5, 2010.

[233] Wrammert, J., Koutsonanos, D., Li, G.M., Edupuganti, S., Sui, J. et al. **Broadly cross-reactive antibodies dominate the human B cell response against 2009 pandemic H1N1 influenza virus infection.** *J. Exp. Med.*, **208**(1):181–93, 2011.

[234] Avnir, Y., Tallarico, A.S., Zhu, Q., Bennett, A.S., Connelly, G. et al. **Molecular signatures of hemagglutinin stem-directed heterosubtypic human neutralizing antibodies against influenza A viruses.** *PLoS Pathog.*, **10**(5):e1004103, 2014.

[235] Pappas, L., Foglierini, M., Piccoli, L., Kallewaard, N.L., Turrini, F. et al. **Rapid development of broadly influenza neutralizing antibodies through redundant mutations**. *Nature*, **516**(7531):418–22, 2014.

[236] Thomson, C.A., Wang, Y., Jackson, L.M., Olson, M., Wang, W. et al. **Pandemic H1N1 Influenza Infection and Vaccination in Humans Induces Cross-Protective Antibodies that Target the Hemagglutinin Stem**. *Front. Immunol.*, **3**(May):87, 2012.

[237] Uno, S., Kimachi, K., Kei, J., Miyazaki, K., Oohama, A. et al. **Effect of prior vaccination with a seasonal trivalent influenza vaccine on the antibody response to the influenza pandemic H1N1 2009 vaccine: A randomized controlled trial**. *Microbiol. Immunol.*, **55**(11):783–789, 2011.

[238] Choi, Y.S., Baek, Y.H., Kang, W., Nam, S.J., Lee, J. et al. **Reduced antibody responses to the pandemic (H1N1) 2009 vaccine after recent seasonal influenza vaccination**. *Clin. Vaccine Immunol.*, **18**(9):1519–1523, 2011.

[239] Frasca, D., Diaz, A., Romero, M., Mendez, N.V., Landin, A.M. et al. **Effects of age on H1N1-specific serum IgG1 and IgG3 levels evaluated during the 2011-2012 influenza vaccine season**. *Immun Ageing*, **10**(1):14, 2013.

[240] Sakai, E., Yamamoto, T., Yamamoto, K., Mizoguchi, Y., Kaneno, H. et al. **IgG3 deficiency and severity of 2009 pandemic H1N1 influenza**. *Pediatr. Int.*, **54**(6):758–761, 2012.

[241] Choi, Y. and Deane, C.M. **FREAD revisited: Accurate loop structure prediction using a database search algorithm**. *Proteins Struct. Funct. Bioinforma.*, **78**(6):1431–1440, 2009.

[242] DeKosky, B.J., Lungu, O.I., Park, D., Johnson, E.L., Charab, W. et al. **Large-scale sequence and structural comparisons of human naive and antigen-experienced antibody repertoires**. *Proc. Natl. Acad. Sci.*, page 201525510, 2016.

[243] Safonova, Y., Bonissone, S., Kurpilyansky, E., Starostina, E., Lapidus, A. et al. **Ig Repertoire Constructor: A novel algorithm for antibody repertoire construction and immunoproteogenomics analysis**. *Bioinformatics*, **31**(12):i53–i61, 2015.

[244] Busse, C.E., Czogiel, I., Braun, P., Arndt, P.F. and Wardemann, H. **Single-cell based high-throughput sequencing of full-length immunoglobulin heavy and light chain genes**. *Eur. J. Immunol.*, **44**(2):597–603, 2014.

[245] Qiu, X., Wong, G., Audet, J., Bello, A., Fernando, L. et al. **Reversion of advanced Ebola virus disease in nonhuman primates with ZMapp**. *Nature*, **514**(7520):47–53, 2014.

[246] Kugelman, J., Kugelman-Tonos, J., Ladner, J., Pettit, J., Keeton, C. et al. **Emergence of Ebola Virus Escape Variants in Infected Nonhuman Primates Treated with the MB-003 Antibody Cocktail**. *Cell Rep.*, **12**(12):2111–20, 2015.

[247] Giudicelli, V., Duroux, P., Ginestoux, C., Folch, G., Jabado-Michaloud, J. et al. **IMGT/LIGM-DB, the IMGT comprehensive database of immunoglobulin and T cell receptor nucleotide sequences.** *Nucleic Acids Res.*, **34**(Database issue):D781–D784, 2006.

[248] Ryu, C.J., Chung, H.K. and Hong, H.J. **Cloning and sequence analysis of cDNAs encoding the heavy and light chain variable regions of a human monoclonal antibody with specificity for hepatitis B surface antigen.** *Biochim. Biophys. Acta*, **1380**(2):151–155, 1998.

[249] Tiwari, A., Khanna, N., Acharya, S.K. and Sinha, S. **Humanization of high affinity anti-HBs antibody by using human consensus sequence and modification of selected minimal positional template and packing residues**. *Vaccine*, **27**(17):2356–2366, 2009.

[250] Kim, K.S., Kim, H.J., Han, B.W., Myung, P.K. and Hong, H.J. **Construction of a humanized antibody to hepatitis B surface antigen by specificity-determining residues (SDR)-grafting and de-immunization**. *Biochem. Biophys. Res. Commun.*, **396**(2):231–237, 2010.

[251] Kim, S.H. and Park, S.Y. **Selection and characterization of human antibodies against hepatitis B virus surface antigen (HBsAg) by phage-display.** *Hybrid. Hybridomics*, **21**(5):385–392, 2002.

[252] Cerino, A., Bremer, C.M., Glebe, D. and Mondelli, M.U. **A Human Monoclonal Antibody against Hepatitis B Surface Antigen with Potent Neutralizing Activity**. *PLoS One*, **10**(4):e0125704, 2015.

[253] Becker, P.D., Legrand, N., van Geelen, C.M.M., Noerder, M., Huntington, N.D. et al. **Generation of human antigen-specific monoclonal IgM antibodies using vaccinated "human immune system" mice**. *PLoS One*, **5**(10):1–10, 2010.

[254] Tajiri, K., Ozawa, T., Jin, A., Tokimitsu, Y., Minemura, M. et al. **Analysis of the epitope and neutralizing capacity of human monoclonal antibodies induced by hepatitis B vaccine**. *Antiviral Res.*, **87**(1):40–49, 2010.

[255] Corti, D., Jr, A.L.S., Pinna, D., Silacci, C., Fernandez-rodriguez, B.M. et al. **Heterosubtypic neutralizing antibodies are produced by individuals immunized with a seasonal influenza vaccine**. *J. Clin. Invest.*, **120**(5):1663–1673, 2010.

[256] Krause, J.C., Tsibane, T., Tumpey, T.M., Huffman, C.J., Briney, B.S. et al. **Epitope-specific human influenza antibody repertoires diversify by B cell intraclonal sequence divergence and interclonal convergence**. *J. Immunol.*, **187**(7):3704–11, 2011.

[257] Krause, J.C., Tsibane, T., Tumpey, T.M., Huffman, C.J., Albrecht, R. et al. **Human monoclonal antibodies to pandemic 1957 H2N2 and pandemic 1968 H3N2 influenza viruses**. *J. Virol.*, **86**(11):6334–40, 2012.

[258] Meng, W., Pan, W., Zhang, A.J.X., Li, Z., Wei, G. et al. **Rapid Generation of Human-Like Neutralizing Monoclonal Antibodies in Urgent Preparedness for Influenza Pandemics and Virulent Infectious Diseases**. *PLoS One*, **8**(6):1–14, 2013.

[259] Ohshima, N., Kubota-Koketsu, R., Iba, Y., Okuno, Y. and Kurosawa, Y. **Two types of antibodies are induced by vaccination with A/California/2009 pdm virus: binding near the sialic acid-binding pocket and neutralizing both H1N1 and H5N1 viruses**. *PLoS One*, **9**(2):e87305, 2014.

[260] Okada, J., Ohshima, N., Kubota-Koketsu, R., Ota, S., Takase, W. et al. **Monoclonal antibodies in man that neutralized H3N2 influenza viruses were classified into three groups with distinct strain specificity: 1968-1973, 1977-1993 and 1997-2003**. *Virology*, **397**(2):322–330, 2010.

[261] Simmons, C.P., Bernasconi, N.L., Suguitan, A.L., Mills, K., Ward, J.M. et al. **Prophylactic and therapeutic efficacy of human monoclonal antibodies against H5N1 influenza.** *PLoS Med.*, **4**(5):e178, 2007.

[262] van den Brink, E., De Kruif, C. and Throsby, M. **Human binding molecules capable of neutralizing influenza virus H5N1 and uses thereof**, 2014.

[263] Marasco, W., Sui, J. and Liddington, R. **Antibodies against influenza virus and methods of use thereof**. *Pat. US20110038935 A1*, 2011.

[264] Ahmed, R., Wrammert, J. and Wilson, P. **Antibodies directed against influenza**. *Pat. WO2012096994 A3*, 2013.

[265] Kaur, K., Zheng, N.Y., Smith, K., Huang, M., Li, L. et al. **High Affinity Antibodies against Influenza Characterize the Plasmablast Response in SLE Patients After Vaccination**. *PLoS One*, **10**(5):e0125618, 2015.

[266] Faber, C., Shan, L., Fan, Z., Guddat, L.W., Furebring, C. et al. **Three-dimensional structure of a human Fab with high affinity for tetanus toxoid.** *Immunotechnology*, **3**(4):253–270, 1998.

[267] Poulsen, T.R., Meijer, P.J., Jensen, A., Nielsen, L.S. and Andersen, P.S. **Kinetic, affinity, and diversity limits of human polyclonal antibody responses against tetanus toxoid.** *J. Immunol.*, **179**(6):3841–3850, 2007.

[268] Yousefi, M., Khosravi-eghbal, R., Mahmoudi, A.R., Jeddi-tehrani, M. and Rabbani, H. **Comparative in vitro and in vivo assessment of toxin neutralization by anti-tetanus toxin monoclonal antibodies**. *Hum. Vaccin. Immunother.*, **10**(2):344–351, 2014.

[269] Persson, M.a., Caothien, R.H. and Burton, D.R. **Generation of diverse high-affinity human monoclonal antibodies by repertoire cloning.** *Proc. Natl. Acad. Sci. U. S. A.*, **88**(6):2432–2436, 1991.

[270] Sorouri, M., Fitzsimmons, S.P., Aydanian, A.G., Bennett, S. and Shapiro, M.a. **Diversity of the Antibody Response to Tetanus Toxoid: Comparison of Hybridoma Library to Phage Display Library**. *PLoS One*, **9**(9):e106699, 2014.