Chapter 1

1. Introduction

1.1. Structure of the adaptive immune system

B-cells and T-cells are key to the immune response, and are crucial to the human body's ability to protect against infection and cancer by producing antibodies that can bind to pathogens and removing infected cells. Dysregulation of B- and T-cells can lead to life-threatening disorders, where understanding B- and T-cell population structures and dynamics are of considerable clinical importance (Tonegawa, 1983), particularly in response to infection (Foster, 2008), malignancies (Mao et al., 2007) and autoimmunity.

The adaptive immune response selectively expands B- and T- cell clones following antigen recognition by B- or T- cell receptors respectively. B-cell receptors (BCRs) mediate the humeral adaptive immune response by the binding antigens licensing B-cell clonal proliferation and antibody production. B-cells play a pivotal role in preventing and clearing infection as well as offering protection against antigen re-challenge. T-cells also play an important role in the adaptive immune response by an number of mechanisms, including co-stimulation of B-cells to differentiate and produce antibodies, stimulating clearance of antigen by other cells of the immune system, direct killing of infected cells, and regulation of immune responses. This section describes antibody structures and functions, B-cell development and the generation of B-cell BCR diversity.

1.1.1. Structure of antibodies

The main function of a B-cell is to produce and secrete immunoglobulin (Ig). Immunoglobulins are glycoproteins that bind antigens with high specificity to facilitate the clearance of antigen either by binding other parts of the immune system or by direct binding of antigen thus inhibiting antigen activity, known as neutralisation. The basic structural units of all immunoglobulins are very similar, consisting of two identical heavy chain (IgH) and two identical light (IgL) chain proteins, linked by disulphide bridges (**Figure 1.1**). The sites at the tip of the antigen-binding (Fab) regions are highly diversified and formed from the variable domains of the heavy (IgH) and light chains (IgL), both generated during B-cell development by

highly regulated gene rearrangements in the B-cell receptor gene loci, addressed in detail in Section 1.1.3 (Woof and Burton, 2004b, Lydyard et al., 2000, Tonegawa, 1983). The trunk of the heavy chain protein is known as the constant region, and is defined by the antibody isotype. Although the different isotypes of immunoglobulin have distinct biological activities, structures and distributions throughout the body, and trigger different effector mechanisms, all isotypes of immunoglobulin (IgA, IgD, IgE, IgG, and IgM) can be expressed as a membrane-associated form on the surface of the B-cell (B-cell receptor) or as a secreted form (antibody). The membrane-associated and secreted Ig forms differ only at the carboxy-terminus of the heavy chain, where a hydrophobic anchor sequence forms part of the membrane-associated Ig protein, and a hydrophilic sequence forms part of the secreted Ig sequence. Differential RNA splicing of the same RNA transcript, known as alternative splicing, generates these two Ig forms (Alt et al., 1980).

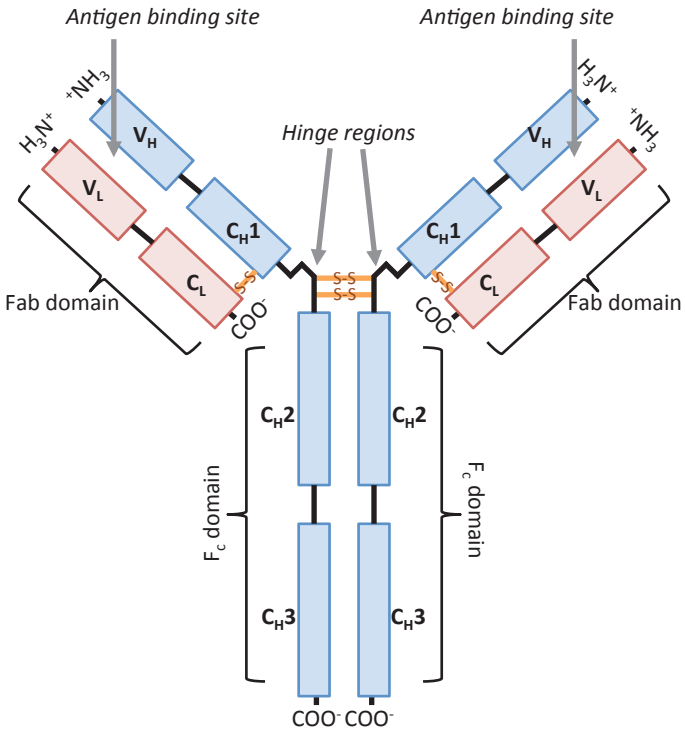

Figure 1.1. Representative structure of an antibody.

V_H indicates the heavy chain variable regions, generated from rearrangement of the Ig gene locus, and C_H1-3 are the constant regions of the antibody in the F_c domain. Likewise, V_L indicates the light chain variable region comprised of IgLV-J recombined genes. S-S indicated regions with disulphide bonds. Fab domains denote the antigen-binding regions. Adapted from Lydyard et al. (Lydyard et al., 2000).

1.1.2. Antibody isotypes

The structure of the heavy chain constant (C_H) gene defines the effector function of the immunoglobulin, and the paired IgH and IgL variable regions define the antigen specificity. The five isotypes of immunoglobulin (IgA, IgG, IgD, IgE, and IgM corresponding to α, γ, δ, ε and μ chains in the IgH gene locus) each form different effector functions. Despite the amino-acid differences between the isotypes, each C_H gene folds into similar structures consisting of β sheet linked together by inter-chain disulphide bonds. Each C_H gene is divided into domains defined as C_H1, C_H2 and C_H3 for IgA, IgD and IgG), and C_H1, C_H2, C_H3 and C_H4 for IgM and IgE. The C_H2-C_H3(-C_H4) domains comprise the region of the antibody, known as the F_c fragment, that mediates effector function by the binding of F_c receptors (FcRs) on effector cells or by activating other immune pathways such as complement activation. Each isotype differs in terms of size, complement fixation and receptor binding, such as to FcRs (Woof and Burton, 2004a). The immunoglobulin isotype also influences binding kinetics of the antibody by different binding efficiencies to the different FcRs. A summary of isotype properties is given in Table 1.1.

Table 1.1. Properties of immunoglobulin isotypes.

Adapted from (Schroeder and Cavacini, 2010) and (Burton and Woof, 1992).

Immunoglobulin isotype	Structure	Serum concentration (mg mL^{-1})	Half-life (days)	Placental transfer*	Complement activation*	Other functions
IgG1	Monomer	9	23	+++	+	Antiviral or secondary response
IgG2	Monomer	3	23	+	+	Neutralize toxins
IgG3	Monomer	1	7	+++	+++	Viral response
IgG4	Monomer	0.5	23	+++	-	Allergy
IgM	Pentamer/hexameric**	1.5	5	-	+++	Primary response
IgA1	Monomer or dimer	3	6	-	+	Direct neutralisation of toxins, viruses and bacteria
IgA2	Monomer or dimer	0.5	6	-	-	Direct neutralisation of toxins, viruses and bacteria
IgD	Monomer	0.04	3	-	-	Homeostasis
IgE	Monomer	0.0003	0.5	-	-	Allergy

*Major effector functions of each isotype are denoted by +++, lesser functions are denoted by +, and – denotes lack of corresponding function.
** (Davis and Shulman, 1989).

- ***IgM***

Monomeric IgM is expressed on the surface of naïve B-cells. After maturation and antigen stimulation, a pentameric form of IgM is secreted, where each unit is linked by disulphide bonds in the C_H4 region. This pentameric form of IgM is linked to joining chains, known as J-chains, by disulphide bonds, which helps mucosal surface secretion. IgM functions by binding antigen for destruction, known as opsonisation, and fixing complement. The monomeric form of IgM generally has low affinity for antigen as the B-cells that produce IgM are early in differentiation and the V(-D)-J regions have not undergone somatic hypermutation. However, the pentameric form may have a high total binding strength, known as avidity, due to the multimeric interactions, which is particularly enhanced if the antigen itself has multiple repeating units, thus is very efficient at opsonisation (Matsuda et al., 1998).

- ***IgD***

Low levels of circulating IgD are found in the serum, and the half-life of serum IgD is short. IgD antibodies have no known effector function, but IgD can react with specific bacterial proteins, such as *Moraxella catarrhalis* outer membrane

protein MID (Riesbeck and Nordstrom, 2006). Most B-cells expressing surface IgD also express surface IgM. It is thought that membrane-bound IgD contributes to the regulation of B-cell fate at particular developmental stages (Geisberger et al., 2006).

- ***IgG***

The predominant immunoglobulin isotype is IgG, which has the longest serum half-life. There are four subclasses of IgG (IgG1, IgG2, IgG3, and IgG4), numbered in order of their serum levels in the blood of healthy individuals. Each subclass differs in terms of their antibody flexibility, as shown by crystal structure analysis, and their affinities to different F_c receptors, and ability to fix complement. For example, the different subtypes of IgG also differ in terms of their disulfide bond structures, where multiple disulfide bond structures have been observed for IgG2 and IgG4 subtypes (Liu and May, 2012). These subclasses also differ in terms of their trans-placental transport and participation in secondary immune responses (summarised in Table 1.1). Response to protein antigens is generally facilitated by IgG1 and IgG3, whereas response to polysaccharide antigens is typically facilitated by IgG2 and IgG4. IgG antibodies directly neutralise toxins and viruses as well as activating other parts of the immune system, such as the classical pathway, which involves a cascade of immune protein production leading to antigen elimination (Cavacini et al., 2003, Scharf et al., 2001).

- ***IgA***

Although IgA antibodies have relatively high levels in the serum, they are predominantly observed on mucosal surfaces and in secretions, such as saliva and breast milk (Woof and Mestecky, 2005). Serum IgA generally exists as a monomer, but at the mucosal surfaces, secretary IgA is a dimer. The dimeric form associates with two other proteins, a J-chain and a secretary component chain, all linked by disulphide bonds. There are two subclasses of IgA (IgA1 and IgA2) that differ mainly in the hinge regions (indicated on **Figure 1.1**). The shorter hinge region in IgA2 decreases sensitivity to bacterial proteases, and predominates in many mucosal secretions such as the genital tract, whereas over 90% of serum IgA is of the IgA1 form. The main function of IgA is direct neutralisation of toxins, viruses and bacteria and the prevention of binding to mucosal surfaces. Intracellular IgA is thought to be

important in the prevention of bacterial and viral infections (Corthesy, 2007), where intracellularisation is thought to be mediated through polymeric immunoglobulin receptor (pIgR)-mediated endocytosis at the basolateral surfaces of epithelial cells followed by transcytosis (Mazanec et al., 1993, Lamm, 1997, Corthesy and Kraehenbuhl, 1999).

- ***IgE***

IgE is a very potent immunoglobulin even though it has the lowest serum levels and the shortest half-life. This immunoglobulin is made in response to parasitic worm infections, but also associated with hypersensitivity and allergic reactions. The high potency is, in part, due to the high affinity to the FcεRI receptor that is expressed on mast cells, eosinophils, basophils, and Langerhans cells (Corthesy, 2007).

1.1.3. Generation of antibody diversity

There are two different mechanisms for generating somatic diversity in B-cell receptor sequences: DNA rearrangement of the V, D and J germline segments and somatic mutation.

1.1.4. B-cell development

1.1.4.1. Immunoglobulin gene rearrangements

B-cells develop from hematopoietic stem cells and differentiate through several maturation stages in the bone marrow, after which they migrate through the peripheral blood to the secondary lymphoid organs (**Figure 1.2**). Survival and maturation of B-cells at each stage of development depends on signals transmitted through cell-surface ligands (Koopman et al., 1994, Mackay et al., 2005) and B-cell development requires the joint action of many cytokines and transcription factors that positively and negatively regulate gene expression (Milne and Paige, 2006, Hardy et al., 2007).

Early progenitor B cells (pro-B cells) are the first stage of differentiation, that rearrange DNA segments of the *Ig* loci to generate unique IgH sequence. The germline IgH chain gene locus encodes for multiple distinct copies of the variable (V), diversity (D) and joining (J) genes, which are separated by over 100 kbp from a much smaller number of DNA segments encoding the constant genes (**Figure 1.3** and **Figure 1.4**) (Lydyard et al., 2000). The total number of reportedly functional IgHV

genes in humans currently range from between 45 to 60 due to variable levels of gene loci heterozygosity (Boyd et al., 2010b), 27 IgHD genes and 6 IgHJ genes (Lefranc et al., 2009). This organisation is maintained in most somatic cell types, but in each individual mature B-cell a unique DNA rearrangement event has occurred. Functional immunoglobulin genes are generated by the process of site-specific recombination by recombination activating genes 1 (RAG1) and 2 (RAG2) through the deletion of intervening DNA (Schatz and Swanson, 2010), creating a IgH gene containing one V, one D and one J gene (VDJ) encoding the protein sequence for the antigen binding region of the IgH protein (**Figure 1.5**) (Lydyard et al., 2000, Latchman, 2005). The order of recombination events is highly regulated, where first the IgHD and IgHJ genes are brought together forming the pro-B-cell. Then the IgHV is recombined to form IgHV-D-J gene recombination of the pre-B-cell. The imprecise joining of the V, D and J gene segments leads to the introduction of random deletions and insertions of nucleotides during recombination events, resulting in sequence diversification at the junctional regions (Tonegawa, 1983). Further mechanisms that contribute to the generation of diversity include alternative IgHD reading frames, IgHD-IgHD fusions, and imprecise joining at the IgHD-J and IgHV-D junctions (Kalinina et al., 2011). These pre-B-cells are selected for functional heavy chain by IgV-D-J μ exon protein expression and IgH assembly by pairing and cell-surface expression with a surrogate light chain protein and Igα/Igβ (Vettermann and Schlissel, 2010, ten Boekel et al., 1998). This complex, along with the participation of the BCR signalling cascade, such as Syk, B-cell linker, and phosphoinositide 3-kinase, give the pre-B-cell signals for survival and proliferation (Fuentes-Panana et al., 2004).

Likewise, each IgL chain locus encodes for multiple distinct copies of variable (V) gene segments and joining (J) gene segments, in addition to a gene segment encoding the λ chain and the κ chain for the λ and the κ gene loci respectively, creating an immature B-cell (Woof and Burton, 2004a). When a cell has successfully rearranged a IgH gene, the B-cell begins to rearrange the κ light chain genes to bring together a κV and κJ. If this produces a functional κ light chain, the B-cell expresses and transcribes the heavy and κ light chains, else it then attempts to rearrange the κ light chain genes on the other chromosome. If the cell is unsuccessful at producing a functional κ light chain BCR from both chromosome κ light chain loci, then the B-cell attempts to rearrange the λ light chain genes to bring together a λV and λJ.

Likewise, if this produces a functional λ light chain, the B-cell expresses and transcribes the heavy and λ light chains.

Each mature, but antigen naïve, B-cell expresses a single BCR sequence, where the B-cell population has sufficient BCR diversity for initial recognition of all potential antigens in their environment (Dorner et al., 1998, Lydyard et al., 2000). After functional V-(D)-J recombination of IgH and IgL chain genes, naïve B-cells transcribe the IgH and IgL genes and are able to produce IgD and IgM immunoglobulin isotype by alternative splicing of the transcript to fuse the μ and δ exon to the IgHJ exon respectively (Geisberger et al., 2006). Later in development and in response to stimulation, B-cells can be signalled to produce other isotypes (IgA, IgG and IgE) (Schroeder and Cavacini, 2010) by alternative transcript splicing and class switch recombination (Section 1.1.6). These naïve B-cells typically migrate to the s such as spleen and lymph nodes where they may encounter antigens (Honjo et al., 2002). Each recombination creates a B-cell clone that can expand to form a lineage expressing a clonal B-cell receptor (a plasma-membrane anchored IgH/L complex) consisting of a heavy and light (either λ or κ) chains.

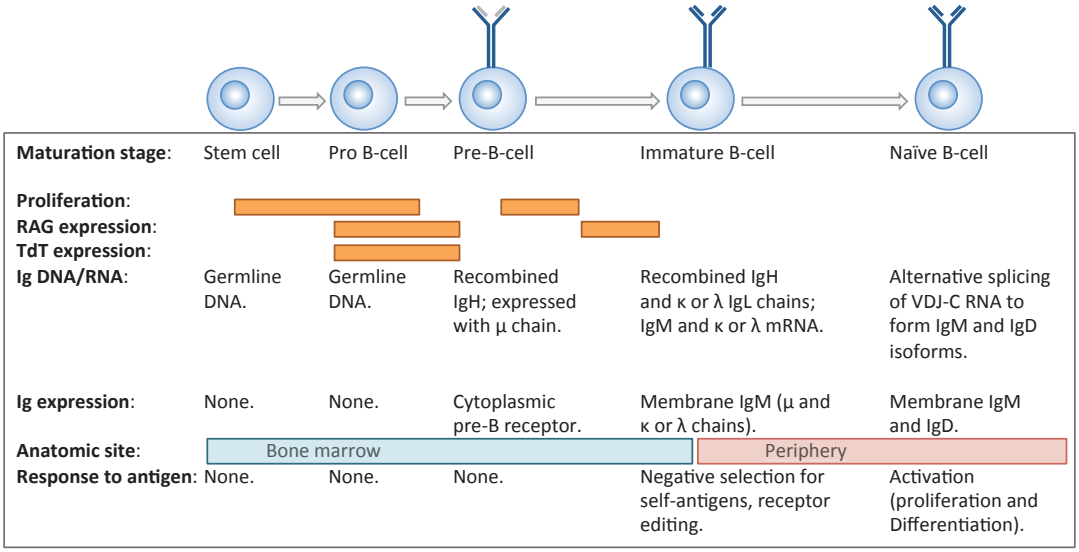

Figure 1.2. Stages of B-cell maturation.

Each stage is indicated by their Ig status, anatomical site and cell-surface marker expression, and the ability to respond to antigen. The haematopoietic stem cell develops into a naïve B-cell through a number of differentiation steps, where the IgH germline DNA is recombined in the pre-B-cell, leading to IgV-D-J μ exon protein expression and IgH assembly with a variant surrogate light chain protein. The light chain (λ or κ chain) then recombines to create an immature B-cell, with membrane assembly of IgM. Naïve B-cells are able to alternatively splice the constant region mRNA to express IgM and IgD. Increased proliferation rates, RAG and TdT expression is highly regulated and occur at specific points in differentiation, indicated by the orange bars, and the sites where each B-cell differentiation stage is indicated by the blue and red bars for the bone marrow and periphery respectively. B-cell Ig rearrangement status, Ig expression and ability to respond to antigen binding is indicated for each differentiation stage. Adapted from Edwards *et al*. (Edwards and Cambridge, 2006).

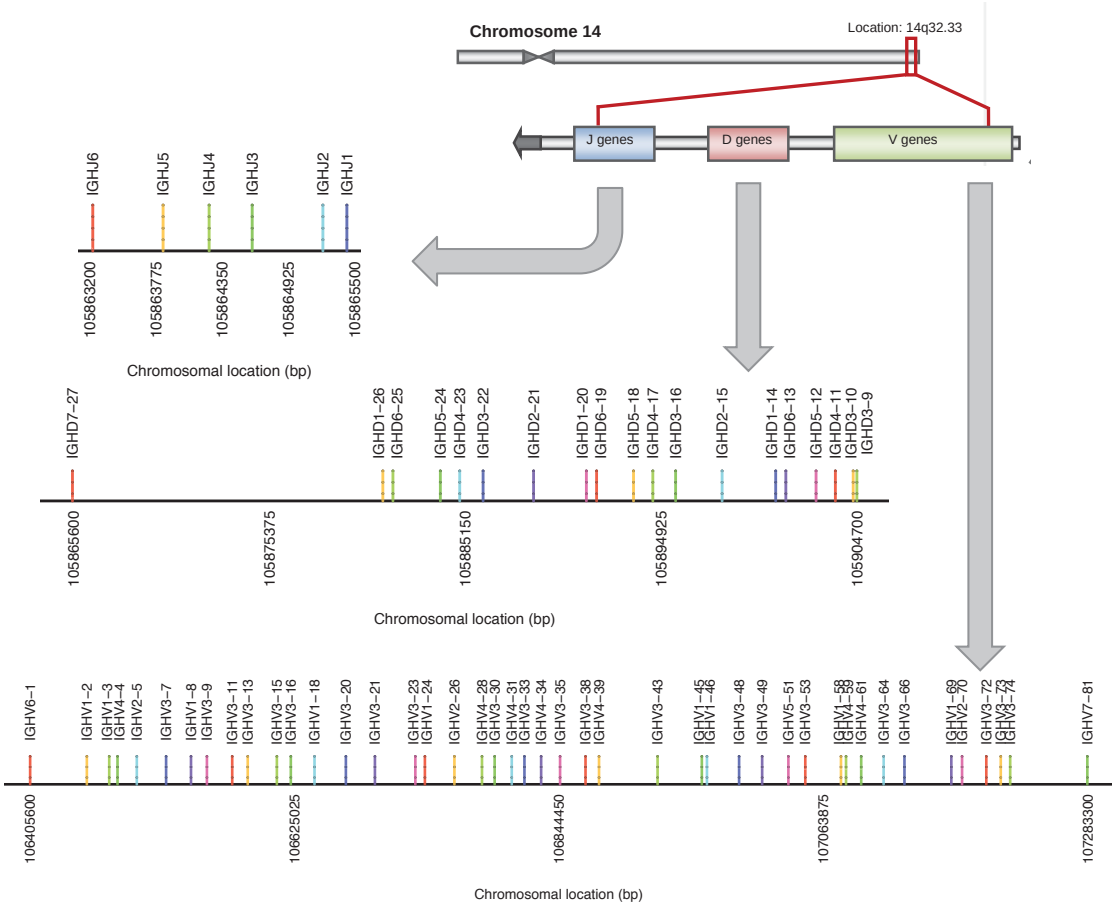

Figure 1.3. Arrangement of the human IgH gene locus on chromosome 14.

Schematic diagram of human chromosome 14, with the Ig locus marked in the red box. Within the locus are the multiple IgHJ, IgHD, and IgHV genes, where the chromosomal locations are marked to scale on the chromosomal sections at the top, middle and bottom respectively. Each coloured line along the chromosomal scale bar represents an IgH gene, with the name indicated below. Generated from gene coordinates from Ensembl.

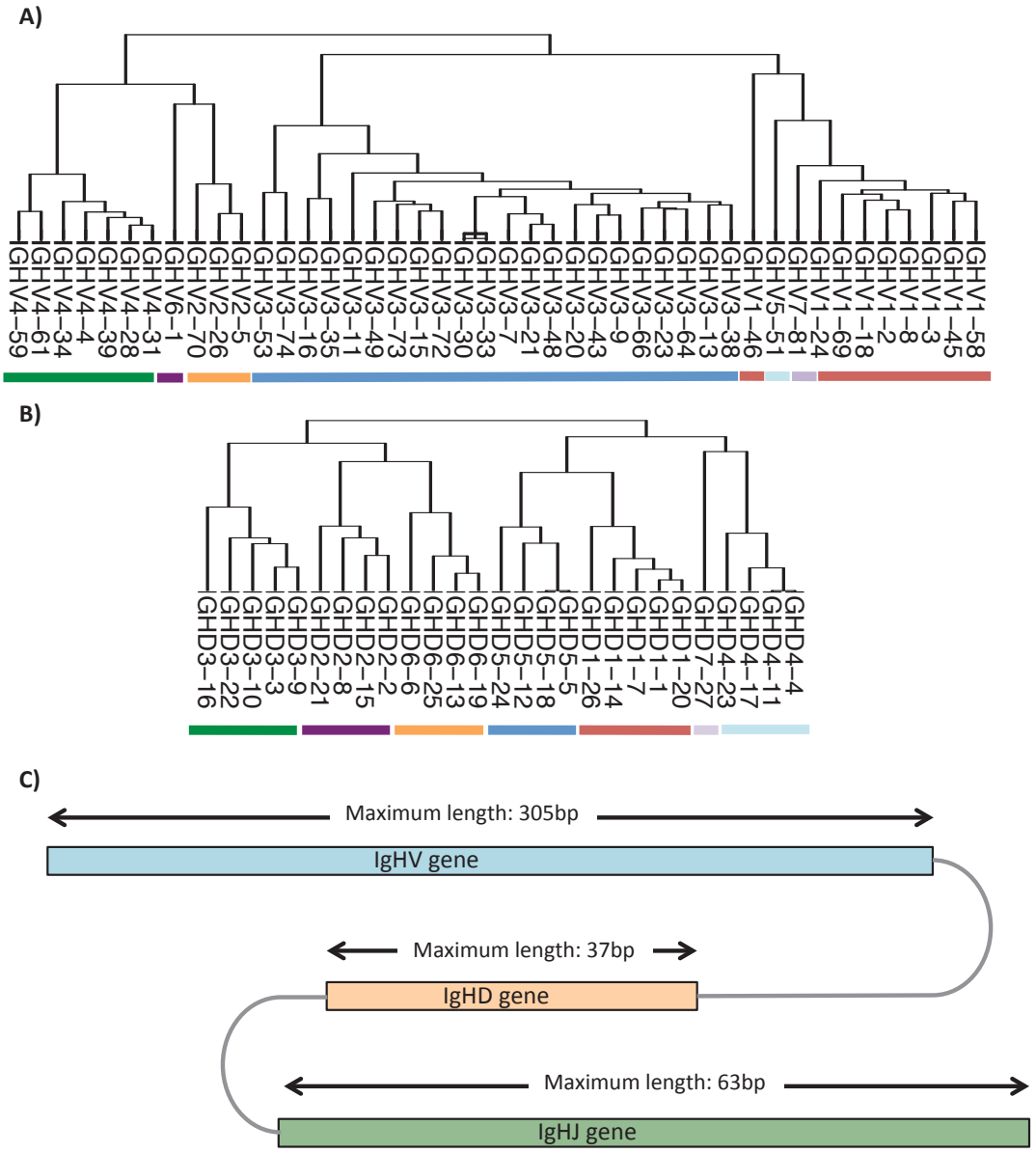

Figure 1.4. Phylogenetic sequence relationships between the human A) IgHV and B) IgHD genes.

Each IgHV or IgHD gene sequences was aligned using Mafft (Katoh and Standley, 2013) and a neighbour joining tree was fitted using BIONJ in R (Gascuel, 1997). The branch lengths represent the estimated evolutionary distance between gene sequences, and the tips are named by the corresponding gene names. Different coloured bars underneath each tree represent different IMGT gene groups (Lefranc et al., 2009) defined by (Giudicelli and Lefranc, 1999). As there are only 6 human IgHJ genes, phylogenetic sequence relationships of the IgHJ genes were not included. **C)** Schematic diagram of a functional BCR, with the maximum human IgHV, IgHD and IgHJ gene lengths indicated.

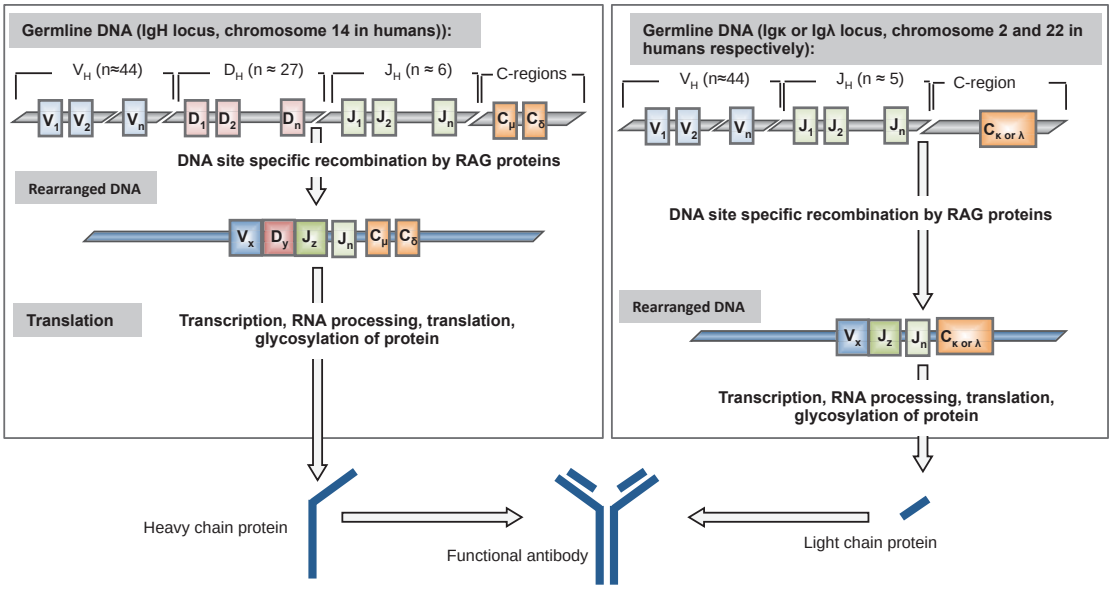

Figure 1.5. Stages of immunoglobulin gene rearrangement.

Gene rearrangements and assembly for the heavy and light to form a functional B-cell receptor. V_x, D_y, J_z denote the IgHV, IgHD, and IgHJ genes. C_μ and C_δ denote the μ and δ constant chains respectively, and $C_{\lambda \text{ or } \kappa}$ denotes the λ or κ chains constant chain. Adapted from Jackson et al (Jackson et al., 2013).

1.1.4.2. B-cell receptor editing and allelic exclusion

If a maturing naïve B-cell has high affinity for self-antigens or does not form a functional BCR, the cells are removed by induced programmed cell death in the bone marrow, known as negative selection. B-cells committed to cell death can be rescued by modifying the V-J light chain recombination so that the B-cell receptor no longer recognises self-antigens or creates a functional reading frame (Dorner et al., 1998). This occurs by the process of receptor editing where renewed IgHV-D-J rearrangement can result in expression of a functional or non-auto-reactive BCR, rescued further by expression of a different IgL chain. Receptor editing is under genetic control, where PLCγ2 is thought to play a role in regulating the recombination-activating (*rag*) genes, and therefore receptor editing (Verkoczy et al., 2007, Benschop et al., 1999, Derudder et al., 2009). Each mature B-cell typically expresses a single heavy chain and light chain allele. The expression of productive functional heavy and light chains suppresses subsequent immunoglobulin gene rearrangements as well as expression of other rearranged alleles, a process known as allelic exclusion (Kitamura and Rajewsky, 1992).

1.1.5. B-cell response to antigens

Naïve B-cells require multiple signals to become activated: the first signal is delivered through the binding of the IgM B-cell receptor to an antigen (a protein, peptide, carbohydrate or other substance that the immune system perceives as being foreign or harmful). IgM cross-linking on the cell surface causing localised IgM clustering. This assembly provides intracellular signalling to the B-cell through communication of the BCR complex via the Igα and Igβ complex. For T-cell dependent antigens (detailed in Section 1.1.5.1), the second signal is delivered through T-helper cell recognition of peptide fragments of antigen bound to MHC class II molecules on the B-cell surface, and the interaction between CD40 ligand (CD40L) on the T-cell surface and CD40 on the B-cell surface. For T-cell independent antigens (detailed in Section 1.1.5.2), the second signal is by interactions between the antigen itself and B-cell surface, or by non-T-cell accessory cells. The third signal is given by the binding of Toll-like receptors (TLRs), that are up-regulated in naïve B-cells upon BCR activation, as well as other co-receptors, such as the CD19:CD21:CD81 protein complex. An example of this is the T-cell independent protein LPS, which binds LPS-binding protein and CD14, that subsequently associates with the receptor TLR-4 on the B-cell, leading to increased B-cell activation. A fourth signal can be delivered through cytokines (LeBien and Tedder, 2008).

1.1.5.1. T-cell dependent B-cell responses

Antigens can be classified as either T-cell dependent (T_{dep}) or T-cell independent (T_{indep}), depending on whether T-cell stimulation is required. The differences between T-cell dependent or independent immune responses are based on antigen size, structure, and nature. The majority of antigens are T_{dep} antigens that cannot induce B-cell proliferation without T-cell help, (i.e. activation signals from T-helper cells that respond to the same antigen). T_{dep} antigen responses lead to the generation of high-affinity class-switched B-cell responses (i.e. antibodies with heavy chains classes of IgM to IgG, IgA or IgE, detailed in Section 1.1.6). However, naïve T-cells require co-stimulatory signals from professional antigen presenting cells (APCs), such as dendritic cells, B-cells and macrophages. For example, dendritic cells, on encounter with a pathogen or antigen, endocytose and display the processed antigen fragments or peptides on their cell surface complexed with MHC proteins.

These cells carry the peptides to local lymph nodes or organs, and undergo maturation in order to be able to active T-cells. However, activation of T-cells requires three protein signals from the APC. Firstly the MHC molecules with bound antigen or antigen fragment must be able to bind the T-cell receptor. Secondly, co-stimulatory proteins of the APC must be able to bind complementary receptors on the T-cell surface, and thirdly, the action of cell adhesion molecules of the T-cells and APCs to enable contact between the T-cell and APC for long enough for the T-cell to become active. However, if the T-cell does not receive all three signals, it may be triggered to apoptose or the activation suppressed, a process known as immunological tolerance. T-helper cells also act to regulate the immune response by cytokine secretion (Korn et al., 2009)

During a T_{dep} response, a small proportion of activated B-cells differentiate into short-lived low-affinity plasma cells within the B-cell regions of the secondary lymphoid organs. Recruitment of the remaining activated B-cells to the B-cell follicular regions of the secondary lymphoid tissues lead to formation of germinal centres (GCs). GCs are micro-anatomical structures that support antigen specific B-cell clonal expansion, positive selection based on antigen affinity and BCR diversification by somatic hypermutation (SHM) (McHeyzer-Williams and McHeyzer-Williams, 2005). SHM is a process that introduces point mutations and, occasionally, insertions and deletions into the variable regions of the heavy chain immunoglobulin, where some of the resulting populations are expanded through positive selection for higher affinity antigen binding (Gojobori and Nei, 1986). These lead to some B-cells improving their antigen specificity and affinity to the antigen, often by several orders of magnitude (Griffiths et al., 1984, Eisen and Siskind, 1964). These hypermutations occur only in B-cells expressing cell type-specific activation-induced cytosine deaminase (AID) and actively transcribed Ig genes. AID is thought to act on both IgH and IgL strands of DNA by deaminating cytosines to uracils. The resulting uracils therefore base-pair with adenines during the next round of B-cell genome and cell division, leading to C to T, or G to A conversions. The additional process of uracil excision by uracil glycosylases and error prone repair of replication of abasic sites leads to transition and transversion mutations at C/G bases (Batrak et al., 2011). The immunoglobulin genes in B-cells are diversified by hypermutation at a significantly higher rate compared to non-immunoglobulin genes in B-cells, and have been found to occur at a significantly higher level within the complementary

determining regions (CDRs) compared to the framework regions (FWRs) in the BCR (Lin et al., 1997). The estimated average rate of somatic hypermutation is 1.51% per nucleotide site, or 3.09% per amino acid (Gojobori and Nei, 1986), which is high enough for the mutation rate to play a significant role in generating antibody diversity (Baltimore, 1981, Tonegawa, 1983). However, if the B-cell does not receive required activation signals after SHM, it may be triggered to apoptose or the activation suppressed as part of immunological tolerance to select only B-cells with the optimal antigen binding properties and to reduce the risk of generation of auto-reactive B-cells. High-affinity B-cells in the GC are positively selected on the basis of antigen-binding affinity. These cells rapidly proliferate to expand the size of the antigen-reactive B-cell pool with more than $2x10^5$ cells in the dark zone and differentiate into either long-lived plasma cells or memory B-cells. Plasma cells typically migrate to the bone marrow and spleen, and secrete high-affinity antibodies for extended periods of time leading to clearance of antigen (Manz et al., 1997, Bernasconi et al., 2002).

1.1.5.2. T-cell independent B-cell responses

T_{indep} antigens can induce B-cell responses directly. Two main types of T_{indep} antigens exist: type I (T_{indep}-I) polyclonal B-cell stimulants, typically soluble antigen, and type II (T_{indep}-II) large non-protein polymeric molecules with repeated epitopes, typically cell-bound antigen. The T_{indep}-II antigens are able to cross-link multiple B-cell receptors on naïve B-cells leading to activation and stimulation of antibody production in the absence of T-cell help. However, TLR stimulation or complement activation with CD21 stimulation is typically additionally required for maximal stimulation. The development of long-term memory B-cells activated against T_{indep} antigens is limited (reviewed in Section 1.1.7.2) (Mond et al., 1995, Adderson, 2001), but T_{indep} antigens, such as polysaccharides, can be modified to produce T-cell dependent B-cell responses via conjugation to protein carriers, resulting in the initiation of longer-lived antibody memory responses (Kelly et al., 2006, Pollard et al., 2009).

1.1.6. Class switch recombination

B-cell activation and isotype switching from IgM to IgG, IgA or IgE through recombination and deletion process is initiated by the encounter of antigens. This process is achieved through deletional recombination via the introduction of DNA

double-stranded breaks in two participating switch regions, rejoining of the broken regions to each other, and deletion of the intervening sequences containing the various C_H genes, in a process known as class switch recombination (CSR) (Chaudhuri and Alt, 2004). The immunoglobulin heavy chain constant region locus consists of an ordered array of Ig C_H region genes each flanked at the 5' side by switch (S) regions. These S regions are composed of tandem repeat units. CSR occurs through the initiation of AID by looping and deletion of the genomic DNA. This generates an extra-chromosomal DNA recombination product, known as the switch circle (SC) (Muramatsu et al., 2000, Manis et al., 2002, Okazaki et al., 2002). As the Cμ gene is located in the most proximal to the IgHV-D-J exon, CSR between the Sμ and another S region at the 5' side brings another C_H gene adjacent to the IgHV-D-J exon (**Figure 1.6**). The specificity of CSR is determined by the chromatin accessibility of the target regions (Muramatsu et al., 2000). An alternative mechanism of CSR has been shown to occur through inter-chromosomal exchange between the target S regions in stimulated B-cells, which would give non-circular chromosomal products (Dougier et al., 2006).

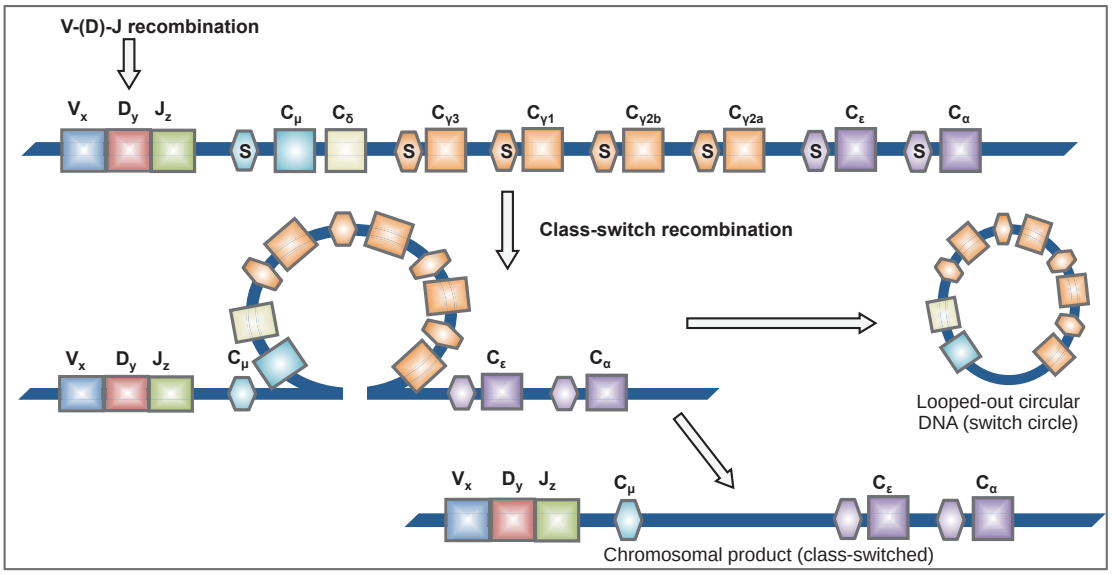

Figure 1.6. Mechanism of class-switch recombination.

The rectangles and hexagons represent exons and switch (S) regions respectively. V-(D)-J recombination occurs in the bone marrow, after which somatic hypermutation and class-switch recombination occurs in the peripheral lymph organs. Class-switch recombination involves the bringing together of heavy chain V-D-J exon with the target constant gene S region, and subsequent removal of intervening DNA. This results in a different chromosomal product and the looped out circular DNA. Adapted from Kinoshita *et al.* (Kinoshita and Honjo, 2001).

1.1.7. B-cell memory responses

The first exposure (the primary exposure) of a pathogen or antigen leads to the activation of naïve B- and T-cells. However, naïve B-cells require multiple signals to become activated (outlined in Section 1.1.5), but leads to the differentiation of antigen-specific antibody producing plasmablasts and memory B-cells, and differentiation of naïve T-cells to memory T-cells. These cells can persist for many years and maintained in a resting state in the absence of sustained antigen, thus immunological memory is established. Immunological memory allows the immune system to respond more rapidly to subsequent re-encounter to the same antigen. Resting memory B-cells are thought to have a low proliferation rate and the number of memory B-cells is highly regulated.

1.1.7.1. Generating T-cell dependent antigen immunological memory

Primary T_{dep} responses result in the interaction of antigen-stimulated B-cells with T-cells and other accessary cells (reviewed in Section 1.1.5.1), leading to the generation of short-lived plasma cells (PCs), GC-independent "early" memory B-cells and/or a GC reaction. The primary GC reactions persist following immunization for long periods (more than 8 months after initial antigen exposure) for certain types of antigen, as shown by monitoring memory B-cells over extended periods of time through the labelling of AID-expressing cells with yellow fluorescent protein (YFP) (Dogan et al., 2009). During this time in the GC, SHM and class-switch recombination can occur, resulting in the generation of high-affinity antigen-specific GC B-cells, that can differentiate into memory B-cells or long-lived PCs.

High-affinity antibody-producing long-lived PCs are thought to be integral to immunological memory, and reside in the in paracortical areas (immediately surrounding the medulla of the lymph nodes) to mature (Mohr et al., 2009). These cells migrate to the medullary regions, where CD93 expression is required for plasma cell survival in the bone marrow (Chevrier et al., 2009). Circulating antibodies from post-GC plasma cells therefore contribute to ongoing immune protection (Bernasconi et al., 2002). These plasma cells can also engage in antigen-specific immune regulation by negatively regulating the expression of BCL-6 and IL-21 in antigen-specific T_{FH} cells (Pelletier et al., 2010), and therefore modulating T-helper cell responses. It is thought that signalling through the BCR or MHC class II molecules in these post-GC plasma cells regulate the ongoing production of serum high-affinity

antibodies, but the mode of long-term antigen-presentation or regulation of post-GC plasma cells is not fully understood.

Humans and mice have been shown to generate memory B-cells expressing surface IgM (IgM^+ memory B-cells) as well as class-switched memory B-cells (expressing immunoglobulin isotypes other than IgM) (Weller et al., 2004, Tangye and Good, 2007). By tracking the murine memory B-cells during T_{dep} responses against sheep red blood cells (Dogan et al., 2009) and phycoerythrin (PE, a fluorescent T_{dep} antigen) (Pape et al., 2011), it was shown than IgM^+ memory B-cells persist longer than $IgG1^+$ memory B-cells. It has been shown in mice that, although the IgG memory B-cell population reduces in number over time, the number of IgM^+ memory B-cells remains constant after resolution of the primary response (Pape et al., 2011). In addition, IgM^+ memory B-cells have a slower turnover rate and typically contain lower levels of SHM than IgG^+ memory B-cells. Although IgM^+ memory B-cells are stimulated during subsequent antigen exposures, class-switched memory B-cells more rapidly differentiate into plasmablasts. IgM^+ memory B-cells can also be generated in T_{indep} responses in the presence of different adjuvants (molecular components that magnify or modulate response to antigen).

1.1.7.2. Generating T-cell independent antigen immunological memory

T_{indep}-II antigen can generate long-lived PCs that secrete IgM or IgG antibodies from secondary lymphoid organs (Hsu et al., 2006) and the bone marrow (Taillardet et al., 2009, Foote et al., 2012). However, long-lived PCs generated from T_{indep}-II antigen have been shown to secrete lower antibody levels compared to their T_{dep} counterparts (Taillardet et al., 2009). Although immunological memory can be employed against T_{indep}-II antigen, generated memory B-cells exhibit shorter longevity and different cell-surface phenotypes in T_{indep} responses compared to that of the T_{dep} response.

B-1 B-cells are a minor B-cell population and are able to contribute to the immune response against T_{indep}-II antigen. B1 B-cells express IgM and are thought to self-renew unlike “conventional” B2 B-cells. It has been shown in mice that B-1 B-cell populations expressing BCRs consisting of IgHV12 and IgHV11 in combination with IgHJ1, with no or low levels of SHM, is thought to be responsible for major natural antibody response against phosphatidylcholine (PtC), a ubiquitously expressed

membrane phospholipid found in both bacteria and mammalian cells (Mercolino et al., 1988, Yoshikawa et al., 2009, Arnold et al., 1994, Popi et al., 2009, Rowley et al., 2007). It is thought that anti-PtC antibodies comprise up to 15% of B-1 cells in the peritoneal cavity in most mouse strains (Arnold and Haughton, 1992), and the majority of B-1 B-cells are generated during foetal or neonatal development, and undergo self-renewal throughout life, thus considered a germline "memorised" B-cell response (Hardy and Hayakawa, 2001, Berland and Wortis, 2002). B1 B-cells can be divided into two subtypes, where B1-a B-cells make up the majority of the B-1 B-cells population and express CD5, and B1-b B-cells are the minor B-1 B-cell population that do not express CD5. It has been shown that, a population of IgM^+ mouse B1-b B-cells persists that confers protection against *Borrelia hermsii* infection on transfer to antigen naïve mice, suggesting that these are memory B-cells (Alugupalli et al., 2004). In addition, after immunisation with (4-hydroxy-3-nitrophenyl)-acetyl (NP)-Ficoll, a T_{indep}-II antigen, mouse IgG^+ and IgG^- B1-b B-cells were shown to persist for more than 4 months, where these cells rapidly divided on adoptively transfer into antigen naïve mice (Obukhanych and Nussenzweig, 2006). However, it is unclear whether the precursors to these anti-NP-Ficoll B-cells were of B1 B-cell or B2 B-cell origin, whether aborted GCs could have been generated (de Vinuesa et al., 2000), and the human counterpart to the B1 B-cell population.

1.1.7.3. Immunological memory recall

The positioning of memory B-cells in the antigen-draining sites of secondary lymphoid tissues, such as the tonsil mucosal epithelium and splenic marginal zone, and the enhanced expression of co-stimulatory molecules assists rapid presentation of antigen to specific T-cells, thus promoting strong secondary adaptive immune responses (summarised in **Figure 1.7**) (Liu et al., 1988, Tangye et al., 1998, Liu et al., 1995). Indeed, enhanced reactivity of memory B-cells over naïve B-cells is thought to be, in part, conveyed by the cytoplasmic domain of surface IgG, thus contribute to rapid secondary immune responses (Martin and Goodnow, 2002). Upon rechallenge with the same antigen, antigen-specific memory B-cells can return to the GCs, undergo further clonal expansion, and differentiate into effector cells, such as plasma cells that secrete high-affinity antibodies.

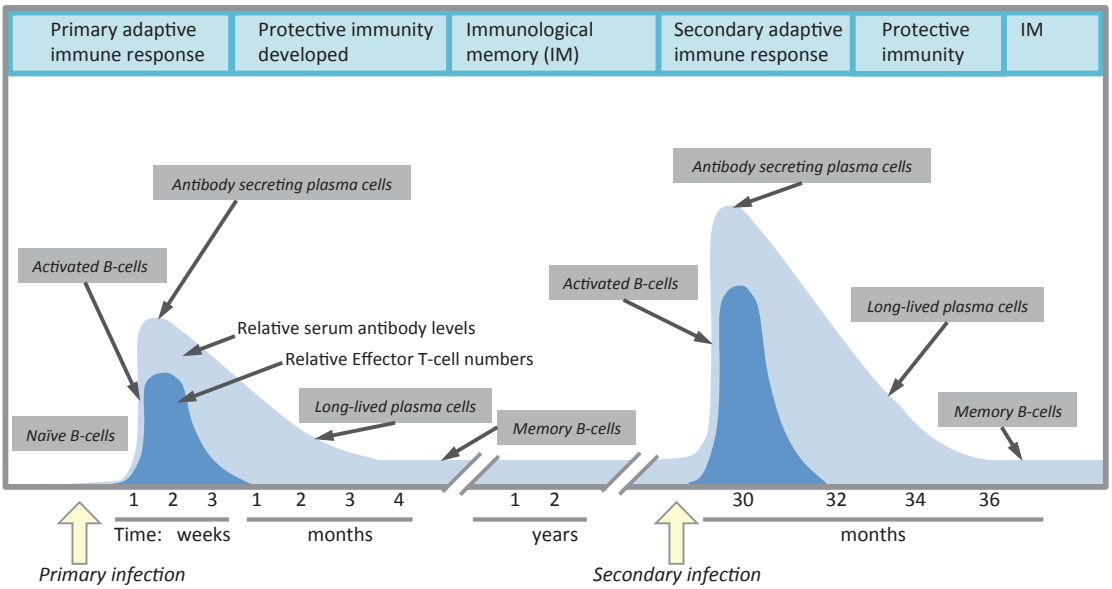

Figure 1.7. Features of primary and secondary response.

Primary exposure to antigen activates naïve B-cells with antigen specificity, resulting in clonal expansion of antigen-specific long-lived memory B-cells, plasma cells and memory T-cells. Plasma cells generate large amounts of serum antibody against the antigen, thus providing protective immunity. The majority of plasma cells in the primary response live for up to a few months, thus the serum antibody levels decline. Subsequent exposure to this antigen leads to the reactivation and proliferation of antigen-specific long-lived memory B- and T-cells and differentiation into effector cells. As memory B- and T-cells are sensitive to activation, and also may have already undergone class switch recombination and affinity maturation, secondary responses typically occur more rapidly and to a larger degree than primary exposure, and producing more effective antigen-specific antibodies. Adapted from McHeyzer-Williams *et al.* (McHeyzer-Williams et al., 2012).

1.2. Measuring B-cell population structure

Healthy humans have approximately $3x10^9$ B-cells in the peripheral blood, where the population of B-cells in an individual reflects both current B-cell responses and historical immune encounters from memory B-cell and plasma cell populations. As a B-cell clone expresses a unique BCR, the B-cell population structure can effectively be probed by analysing the repertoire of BCR sequences in a given B-cell sample, for example from peripheral blood or bone marrow sample. This section details the main advances in understanding B-cell population structures and dynamics in health and disease by B-cell BCR repertoire sequencing.

1.2.1. Low-throughput B-cell receptor analyses

Low-throughput analysis of the heavy and light chains in the 1990s has illuminated biological mechanisms involved in the generation of specific immune responses. The functional characterisation of antibodies was made possible by the cloning of immunoglobulin genes from single B-cells and the isolation of specific antibodies. An alternative route for expression of antibodies was made possible through B-cell immortalisation (Tiller et al., 2007, Corti et al., 2011, Corti and Lanzavecchia, 2013). These methods have led to the isolation of neutralising antibodies to a range of pathogens. A summary of six vaccine studies based on low-resolution B-cell repertoire characterisation is given in Table 1.2.

Table 1.2. Summary of vaccine studies based on low-resolution B-cell repertoire characterisation.

Adapted from (Galson et al., 2014)

Vaccine*	Cells used	Methodology	Key findings	References
Influenza				
TIV	IgG plasmablasts 7 days after vaccination	Single cell heavy and light chain PCR followed by Sanger sequencing	Study of 50 mAbs produced from 14 individuals against three different influenza strains, showing that influenza-specific antibody response is pauci-clonal, with extensive SHM-derived intraclonal diversification of the influenza-specific lineages.	(Wrammert et al., 2008)
TIV	IgA and IgM plasmablasts 7 days after vaccination	Single cell heavy and light chain PCR followed by 454 sequencing	384–768 sequences analysed from three individuals. eight mAbs from large clonal sequence families, and 12 mAbs from singleton sequences were cloned, where 75% of these mAbs bound and neutralized influenza. Most effective binding was from sequences from large clonal families, three of which bound more effectively to the HA from the previous influenza season than the vaccine strain.	(Tan et al., 2014)
Tetanus				
TT	Plasmablasts 6 days after three consecutive vaccinations, separated by at least 1.5 years	Single cell heavy and light chain linkage PCR, cloning into *Escherichia coli* and Sanger sequencing of TT-positive clones	The level of SHM in the BCR sequences were similar between individuals, and did not increase through the study, suggesting the limit had already been reached through previous routine vaccinations.	(Poulsen et al., 2011)
TT	TT-specific plasmablasts 7 days, and TT-specific memory B cells 9 days after vaccination	Single cell isotype-specific heavy and light chain PCR followed by Sanger sequencing	CDR3 length, IgHV-D-J gene usage, and distribution of SHMs were similar among TT-specific plasmablasts and memory cells.	(Frolich et al., 2010)
S. pneumoniae				
PS (23 valent)	IgG plasmablasts 7 days after vaccination	Single cell heavy and light chain PCR followed by Sanger sequencing	137 mAbs against 19 of the 23 vaccine serotypes from four individuals were cloned, and it was found that most antibodies were serotype-specific, but 12% cross reacted with two or more serotypes.	(Smith et al., 2013)
PS (23 valent)	PPS4 or PPS14 specific B cells 6 weeks after vaccination	Single cell culture followed by IgH PCR and Sanger sequencing of pooled, cultured cells	More than 1300 sequences from 40 individuals were analysed, showing significant differences in antibody repertoires between young and elderly individuals, where the latter had significantly more clonal with lower levels of SHM.	(Kolibab et al., 2005)
Hib				
PS or PS-DT or OC-CRM	Lymphocytes 7 days after vaccination	Fusion of lymphocytes to mouse myeloma cells followed by culture, heavy and light chain PCR and Sanger sequencing	15 cell lines that secreted antibody against Hib PS were sequenced from 10 individuals, where it was found that these mAbs had undergone SHM and demonstrated increased B-cell clonality after vaccination and bias towards use of the IgHV3 gene family.	(Adderson et al., 1993)
PS-DT	Lymphocytes 7 days after vaccination	Fusion of lymphocytes to mouse myeloma cells followed by culture, heavy and light chain PCR and Sanger sequencing	4 cell lines that secreted antibody against Hib PS from four individuals were sequenced, where all used IgHV3 genes, but 2 unique IgHD-IgHJ and IgHKV gene segments, indicating that the four cell lines were from two different lineages.	(Pinchuk et al., 1995)
PS or PS-DT	Lymphocytes 7 days after vaccination	Fusion of lymphocytes to mouse myeloma cells followed by culture, IgH PCR and Sanger sequencing	5 cell lines that secreted antibody against Hib PS from four individuals were sequenced, where all used IgHV3 genes, but unique IgHD-IgHJ gene combinations.	(Adderson et al., 1991)

*Abbreviations: DT = diphtheria toxoid; Hib = *Haemophilus influenzae* type b; OC = oligosaccharide; PPV23 = 23-valent pneumococcal polysaccharide vaccine; PS = polysaccharide; TIV = trivalent inactivated influenza vaccine.

1.2.2. High-throughput B-cell receptor analyses

As healthy peripheral blood contains 0.07-0.53 $x10^6$ B-cells per ml, high-throughput sequencing, currently able to produce $>10^7$ sequencing reads per run covering the variable BCR gene sequence, is well suited for sampling this BCR repertoire (Dimitrov, 2010, Benichou et al., 2012). To ensure that the maximum number of sequencing reads correspond to fully rearranged BCRs and reduce the number of non-specific sequencing reads, B-cell DNA samples require PCR amplification, and B-cell RNA samples require both cDNA synthesis and PCR amplification. The three main BCR amplification methods for sequencing BCR repertoires are PCR using IgH-specific multiplex primers (van Dongen et al., 2003), 5' rapid amplification of cDNA ends (5'RACE) (Freeman et al., 2009, Bertioli, 1997, Warren et al., 2011, Varadarajan et al., 2011) and RNA-capture using RNA bait/capture probes (Choi et al., 2009, Mamanova et al., 2010) (summarised in **Figure 1.8**).

Three sets of human IgH-specific multiplex PCR primers have been designed (van Dongen et al., 2003), and validated (van Krieken et al., 2007, Evans et al., 2007, Vargas et al., 2008, Lukowsky et al., 2006, Bruggemann et al., 2007), known as BIOMED-2 FR1, FR2 and FR3 primer sets. These primer sets use a single IgHJ specific primer that can potentially bind any IgHJ gene, and 6-7 IgHV primers that can potentially bind any of the reference IgHV genes. The annealing sites of the BIOMED-2 FR1, FR2 and FR3 IgHV primers are in the highly conserved FR1, FR2 and FR3 regions of the IgHV genes respectively (**Figure 1.9**). PCR amplification using the BIOMED-2 FR1 primer set gives the longest PCR product, therefore is the most popularly used primer set for biological studies (Campbell et al., 2008, Boyd et al., 2009, Sanchez et al., 2003, Maletzki et al., 2012, Boyd et al., 2010a, Lev et al., 2012, Jager et al., 2012, Krause et al., 2011, Bashford-Rogers et al., 2013). Similar primers have also been designed to amplify the B-cell light chains (van Dongen et al., 2003). This multiplex PCR method can be performed on either RNA or DNA and sensitive enough to amplify BCRs from even single cells.

RNA-capture is based around the methods used for human exome sequencing that uses RNA bait/capture probes and subsequent universal PCR amplification (Choi et al., 2009, Mamanova et al., 2010). Briefly, the cDNA is generated from RNA, typically using primers that allow for sample indexing and sequencer specific

adaptors. The resulting sampling is hybridised with 120bp biotinylated RNA-capture baits, designed to bind to any IgHV or IgH constant region (as well as, potentially, other regions, such as the light chains and T-cell receptors, **Figure 1.8**B). The hybridised sequences are specifically bound to magnetic streptavidin beads, after which the sequences are universally amplified and sequenced. This allows for enrichment, amplification and sequencing of TCRs (α, β, γ and δ chains) and BCRs (heavy and light chains) simultaneously. PCR and RNA-capture methods can use RNA or DNA, but have the potential for sequence-based differential annealing and biased capture.

5'RACE uses a single Ig-specific primers, either to the heavy or light chain J genes or constant regions, for first strand Ig cDNA synthesis and subsequent sequence-independent template switching primer for second strand cDNA synthesis (**Figure 1.8**C). This eliminates potential multiplex primer bias, but can have low efficiency, high non-specific amplification, and short fragment contamination from RNA degradation or incomplete cDNA synthesis and template switching (Freeman et al., 2009, Bertioli, 1997, Warren et al., 2011, Varadarajan et al., 2011). Also, as the RNA bait probes and multiplex PCR primers are generated from reference Ig and TCR gene databases, they may lack the same efficiency as 5'RACE for capturing certain human allelic variants of TCR or BCR segments that are not represented in the reference database.

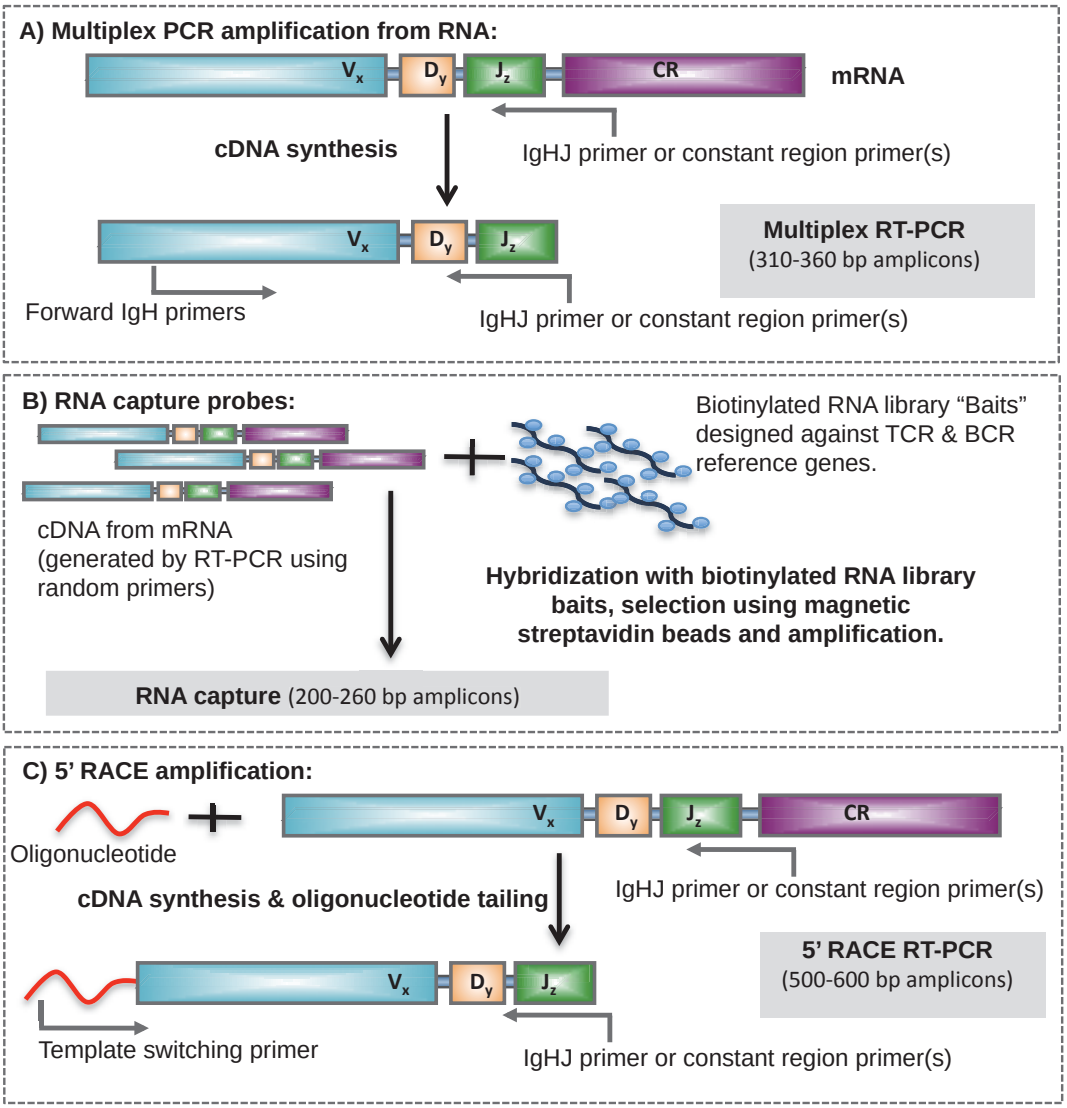

Figure 1.8. Different IgH RNA sequencing methods.

A schematic diagram of the different BCR amplification methods: RNA was extracted from peripheral blood samples and multiplex PCR, 5'RACE and RNA-capture. **A)** Multiplex RT-PCR of RNA uses degenerate primers located in conserved regions of the IgHV and IgHJ or constant region primer, where lengths of PCR products are based on the use of BIOMED-2 FR1 primer set. **B)** RNA-capture uses RNA bait probes and subsequent PCR amplification. **C)** 5' Rapid amplification of cDNA ends (5'RACE) of RNA uses a single IgHJ or constant region primer and a template switching primer.

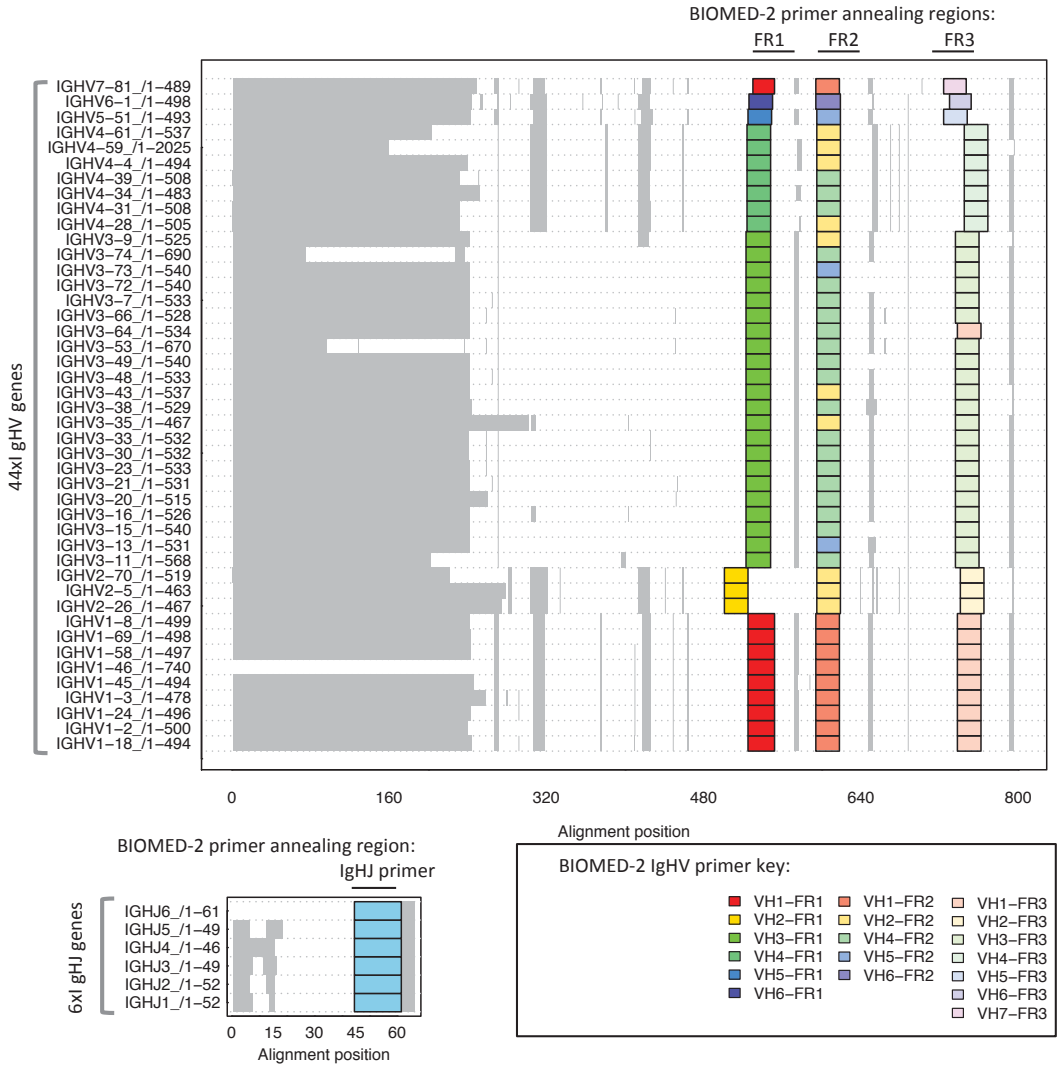

Figure 1.9. Alignment of human IgHV and J genes with BIOMED-2 primer annealing locations.

Each row of the figure represents the alignment of the corresponding IgHV or IgHJ gene sequence by CLustalW, with the gene name given on the left of the alignment box. Grey represents gaps in the alignment, and the coloured boxes represent the best annealing locations of the BIOMED-2 FR1, FR2 and FR3 primers. For the IgHV region, the colours of the primer annealing locations represent the primer that is most likely to anneal (with a maximum of 3 base-pair mismatches between the gene sequence and the primer sequence). For the IgHJ region, the light-blue boxes represent the primer annealing location of the BIOMED-2 IgHJ primer.

The limitation of sequencing of paired heavy and light chains from bulk cells in independent reactions is that information of the heavy and light pairings are lost. A solution to this is high-throughput single cell linkage PCR. This method can currently sequence more than 50,000 cells in a single experiment by depositing single cells in a high-density microwell plate and *in situ* lysis (DeKosky et al., 2013). mRNA is then captured in magnetic beads, on which RT-PCR is performed by emulsion VH:VL linkage PCR. These methods can be used to characterise antibodies of interest by generation of recombinant antibody by cloning the paired heavy and light chains into expression vectors, such as antibody variants of an isolated anti-HIV broadly neutralising antibody (Zhu et al., 2013a).

1.2.3. B-cell receptor repertoires

1.2.3.1. B-cell repertoires in model species

One of the first studies of the B-cell repertoire was performed on zebrafish (Weinstein et al., 2009). Zebrafish are ideal organisms for trialling high-throughput sequencing methods as they have recognisable adaptive immune system similar to humans that undergo IgHV-D-J recombination to generation functional BCRs with junctional diversity and the potential for somatic hypermutation. However, the zebrafish immune system contains only about 300,000 B-cells that produce antibodies, thus can be exhaustively sampled in a single sequencing run. Weinstein *et al.* found that not all possible IgHV-D-J combinations were used per zebrafish, but IgHV-D-J frequency distributions were highly correlated between individual zebrafish. A summary of studies of B-cell repertoires in model species is given in Table 1.3.

Table 1.3. Summary of studies of B-cell repertoires in model species.

Vaccine/ disease	Cells used	Methodology	Key findings	References
None	14 zebrafish	Multiplex PCR of IgH and 454 sequencing	Only between 50-86% of all possible IgHV-D-J combinations were used per zebrafish, where IgHV-D-J frequency distributions were highly correlated between individuals zebrafish. Zebrafish exhibited unique BCRs that were shared between different individuals.	(Weinstein et al., 2009)
None	3 healthy homozygous isogenic Teleost fish and 4 Viral Hemorrhagic Septicemia Virus (VHSV)-infected Teleost fish	Cloning of IgHs into pCR2.1 vector and 454 sequencing	IgM, IgD and IgT repertoires were distinct (where the IgT class is specific to fish and encoded by the τ constant region gene). Clonal expansions dominated by a small number of large public and private clones were observed in infected fish. Differences were seen between the mucosal IgT and IgM repertoires, indicating that both IgM^+ and IgT^+ splenic B-cells responded to systemic infection but to different degrees.	(Castro et al., 2013)
None	PB from two crossbred calves (Brown Swiss × Red Angus-Simmental) and two purebred Holstein calves	Multiplex PCR of IgH and 454 sequencing	The bimodal length distribution of unique CDR3 sequences, with common lengths of 5-6bp and 21-25bp amino acids. 19 extremely long CDR3 sequences (up to 62 amino acids in length) within IgG transcripts were observed, a phenomenon that is rarely observed in other species. In addition, there was a high number of cysteine residues in the CDR3 regions compared to human BCRs.	(Larsen and Smith, 2012)
Finegoldia magna	4 adult C57BL/6 mice received injections of the V_L-targeting superantigen, protein L (PpL), a product of the common commensal bacterial species, *Finegoldia magna*	5′ RACE of mouse light chains and 454 sequencing	Recurrent and consistent patterns of IgKV-J gene pairing was observed, where PpL exposure resulted in a significant shift in the repertoire in all exposed mice. Specifically, significant reductions in the frequencies of 14 specific IgKV gene combinations were observed. This suggests that microbial protein may modulate the composition of the B-cell repertoire, with potential implications for the relationship between the host-microbiome.	(Gronwall et al., 2012)
None	2 NOD-*scid-IL2Rγ*^{null} mice were engrafted with UCB $CD34^+$ hematopoietic progenitors and $CD19^+$ immature, naïve or total splenic B-cells were sampled at either 16 or 18 weeks of age	5′ RACE of mouse heavy and light chains and 454 sequencing	Naïve B-cell repertoires in humanised mice are indistinguishable from those in human PB B-cells, with high correlations between heavy and light V-(D)-J gene usage frequencies, and similar CDR3 length distributions and charged amino-acid content, and hydropathy. However, the CDR3 region was found to be highly diversified in these mice, and specific for each individual.	(Ippolito et al., 2012)

1.2.3.2. Diversity of the immune repertoire

The potential number of different IgHV-D-J combinations in humans is 7128 (44 IgHVs x 27 IgHD x 6 IgHJs), and the number of light chain combinations is 200 and 124 for the Igκ chains and Igλ chains respectively (Table 1.4). However, with the junctional diversity between the gene recombination regions comprising of non-template additions and deletions greatly increases the number of potential unique BCRs that an individual can produce to over approximately 10^{14}. Further diversification can occur during the process of somatic hypermutation. However, as each B-cell can only encode for a single BCR, the true number of unique BCRs in an individual is bounded by the number of B-cells present. With the healthy peripheral blood B-cell population contains approximately 80% naïve B-cells and 20% memory B-cells (Tangye and Good, 2007), where each naïve B-cell is antigen inexperienced so each naïve B-cell BCR is considered to be unique (i.e. not clonally expanded). Sequencing BCRs from only naïve B-cells therefore theoretically results in a diverse BCR population with all BCRs represented with equal probability. In fact, the number of unique BCRs in two healthy individuals was estimated to be $3x10^9$ -$3x10^{10}$ by high-throughput sequencing of the CDR3 regions (Arnaout et al., 2011, Glanville et al., 2009a). However, little is currently known about B-cell turnover, dynamics or the different tissue distributions or B-cell repertoires (Dimitrov, 2010).

Table 1.4. Number of potential human BCR gene segment combinations.

Ig chain	Number of gene Segments	Number of potential combinations
IgκV	40	
IgκJ	5	200 Igκ chains
IgλV	31	
IgλJ	4	124 Igλ chains
IgHV	44	
IgHD	27	
IgHJ	6	7,128 IgH chains
Total number of potential combinations:		$2.452x10^6$

High-throughput BCR sequencing of different B-cell subsets can distinguish between human transitional, naive repertoires, switched memory B-cell and IgM memory populations (Wu et al., 2010). Previous studies have qualitatively shown diverse IgH repertoires in healthy patients contrasting with clonal populations in

malignancies (Boyd et al., 2009, Campbell et al., 2008, Logan et al., 2011, Sanchez et al., 2003, Maletzki et al., 2012, Carulli et al., 2011, Bashford-Rogers et al., 2013), and that distinct subsets of B-cells, defined by difference cell surface markers, within the same individual have distinct repertoires (Wu et al., 2010). Other significant studies of healthy BCR repertoires by high-throughput sequencing are summarised in Table 1.5.

Table 1.5. Summary of studies of B-cell repertoires from healthy individuals.

Cells used	Methodology	Key findings	References
Using IgD, CD27 and CD10 to sort cells into transitional, naive, switched memory, and IgM memory populations	Multiplex PCR of IgH and 454 sequencing	Memory B-cell repertoires differ from transitional and naïve repertoires, where the IgM memory repertoire is distinct from that of class-switched memory B-cells.	(Wu et al., 2010)
654 healthy human donors	Multiplex PCR of heavy and light chains and 454 sequencing	Although length and sequence variability in the CDR3 regions, substantial contributions to somatic diversity were found in the CDR1 and CDR2 regions. Estimation of the total number of unique BCRs from this healthy donor was estimated to be at least 3.5×10^{10}.	(Glanville et al., 2009b)
10 different human tissues using RNA samples derived from a large number of individuals	Multiplex PCR of IgH and 454 sequencing	Unique B-cell gene repertoires were observed in mucosal tissues, such as stomach, intestine and lung that differed significantly from those found in lymphoid tissues or PB. Mucosal tissue BCRs were distinct from peripheral blood and lymphoid tissue repertoires in their mutation level, frequency of both insertions and deletions, and CDR3 region lengths.	(Briney et al., 2014)
Isolated naive, IgM memory and IgG memory B cells from 4 healthy individuals	Multiplex PCR of IgH and 454 sequencing	IgHV-D-D-J recombinations were present in approximately 1 in 800 circulating B-cells, where this frequency was significantly reduced in memory B-cell subsets. These recombinations were shown to occur with all IgHD genes, suggesting that all recombination signal sequences that flank the IgHD genes are able to undergo IgHV-D-D-J recombination.	(Briney et al., 2012)

Historically, it was thought that BCRs and TCRs would not typically be shared between individuals due to the potential number of unique BCRs and TCRs compared to the limited number of B- and T-cells. However, it has been shown in multiple studies that certain BCRs and TCRs are shared significantly between individuals, known as public BCRs and TCRs respectively (exemplified in **Figure 1.10**), and thought to be a result of germline encoded BCRs (i.e. gene combinations with no somatic hypermutations (Li et al., 2012, Agathangelidis et al., 2012, Darzentas and Stamatopoulos, 2013, Messmer et al., 2004, Rossi and Gaidano, 2010, Warren et al., 2013, Hoi and Ippolito, 2013).

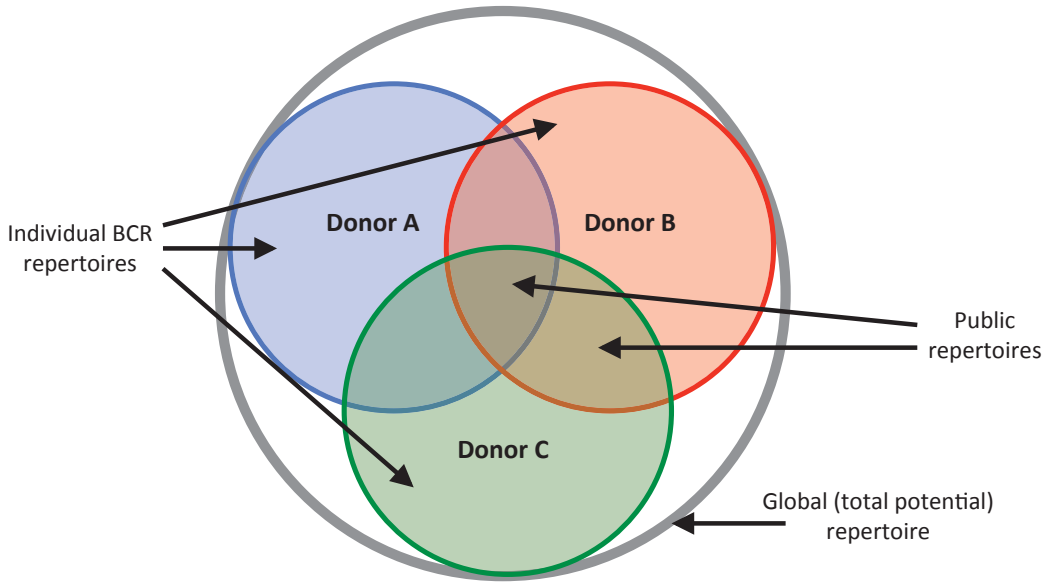

Figure 1.10. Schematic diagram of the different types of BCR repertoire.

Individual repertoires from each individual, where the BCRs shared between two or more donors are the public repertoires. The global repertoire is the set of all potential BCR sequences. Not drawn to scale, and adapted from (Finn and Crowe, 2013).

1.2.3.3. Immune repertoire variation with age

The adaptive immune system is not fully functional in human infants, and therefore infants can receive maternal antibodies (IgG) through their mother's milk. However, young infants are at increased risk of infectious diseases, such as influenza (Feeney et al., 2000). Although the Ig diversity is primarily thought to be a random process, evidence for deterministic, programmed repertoire development in foetal repertoires has been shown by overrepresentation of certain V segments in both mouse and humans (Perlmutter et al., 1985, Berman et al., 1991, Kalled and Brodeur, 1990). Preferential IgHV gene use in the adult B-cell repertoire is distinct from that of foetal, and young infant B-cell repertoires. For example, very young infant respiratory syncytial virus (RSV)-specific B-cells (<3 years of age, purified by immunoaffinity purification using RSV F protein) exhibited a biased repertoire with preferential by use of the IgHV1, IgHV3, and IgHV4 gene families, and less common use of the four immunodominant genes seen in the adult RSV F-specific B-cell response (IgHV3–23, IgHV3–30, IgHV3–33 and IgHV4–04) (Williams et al., 2009). The BCRs from children under three months of age possessed significantly fewer somatic mutations than those of adults, thus suggesting that younger children produce a different and potentially less optimised or weaker immune response than adults. The most frequently observed rearranged BCRs in healthy adult humans included IgHV4-59, IgHV4-61, IgHV3-23, and IgHV3-48 genes, where only 10 different IgHV genes account for more than half of all observed BCRs (Arnaout et al., 2011, Glanville et al., 2009b, Lloyd et al., 2009). Similarly, IgHD2-2, IgHD3-3, IgHD3-10, and IgHD3-22 were the highest observed IgHD genes and IgHJ4 and IgHJ6 the highest observed IgHJ genes in rearranged BCRs from healthy individuals, where this pattern of IgH gene recombination bias were shown to be consistent between multiple unrelated healthy adult individuals and in separate studies (Arnaout et al., 2011, Brezinschek et al., 1997).

The mechanisms so far implicating determinism versus stochasticity in the foetal repertoire are two-fold. Firstly, variation in recombination signal sequences which flank the V, D and J genes leads to favoured gene segments to be recognised by the recombinase (Feeney et al., 2000). Secondly, the observation that the expression of the terminal deoxyribonucleotidyl transferase enzyme is supressed in infants, therefore reducing the diversity generated through non-templated random nucleotide

insertions and deletions at IgHV-D and IgHD-J junctions) (Schroeder et al., 2001). Changes in B-cell repertoire structure have been associated with age in multiple studies, where increases in clonality and delays in immune response correlate with age and immunosenescence (summarised in Table 1.6).

Table 1.6. Summary of studies of immune repertoire variation with age.

Adapted from (Galson et al., 2014).

Vaccine*	Cells used	Methodology	Key findings	References
None	Zebrafish (ZF)	Multiplex PCR of IgH and 454 sequencing	At 2 weeks, ZF have highly preferential use of a small number of IgHV-D-J gene combinations, but this stereotypy decreases as the zebrafish mature by 1 month old. Evidence of complex diversification processes of antibody maturation observed due to clonal expansion during the affinity maturation process.	(Jiang et al., 2011)
None	14 healthy donors representing different gender and age groups	5' RACE of heavy and light chains and 454 sequencing	Donor B-cell repertoires separate into clusters of young adults and elderly donors (>50), thus suggests that clustering defines the onset of immune senescence at the age of fifty and beyond.	(Rubelt et al., 2012)
Influenza vaccine				
TIV or LAIV	Naïve B-cells, and plasmablasts on the day of vaccination and on days 7 or 8 and day 28 after vaccination from 17 participants from three age groups	IgH-specific multiplex PCR, and 454 sequencing	Higher clonality and SHM level was observed in the influenza-specific antibody repertoire in older individuals compared to younger individuals. In twins, the SHM level of the IgM repertoire was similar, but diverged for the IgG repertoire, indicating that the naïve repertoire is more influenced by individual genetics, but the memory repertoire is more influenced by environmental stimuli.	(Jiang et al., 2013)
Co-administered vaccines				
TIV and PPV23	PBMCs on the day of vaccination and on day 7 and day 28 after vaccination from 14 participants from two age groups	Semi-nested isotype and IgH-specific multiplex PCR, and 454 sequencing	At day 7 post-vaccination the repertoire changed, but returned to a baseline-like state after 28 days. Clonal expansion after vaccination is delayed in older individuals compared to young individuals, and age-related differences in IgA and IgM repertoire dynamics were observed.	(Wu et al., 2012)

*Abbreviations: DT=diphtheria toxoid; OC=oligosaccharide; PPV23=23-valent pneumococcal polysaccharide vaccine; PS=polysaccharide; TIV=trivalent inactivated influenza vaccine.

1.2.3.4. B-cell repertoire responses to vaccines and natural infections

Sequencing the B-cell immune response to infectious diseases and vaccines, such as human immunodeficiency virus and influenza, have been used to understand better the development of an antigen specific immune response and for identification of antigen-specific antibodies. Key studies and their findings are summarised in Table 1.7 and Table 1.8.

Table 1.7. Summary of studies of antigen-specific antibody repertoires.

Disease	Cells used	Methodology	Key findings	References
Human immunodeficiency virus (HIV)				
Natural HIV infection	Sera and PBMCs from a HIV-1 infected donor	5' RACE of IgH and 454 sequencing	Heavy and light-chain phylogenetic trees of identified anti-HIV 10E8 variants displayed similar architectures, where 10E8 variants reconstituted from matched and unmatched phylogenetic branches displayed significantly lower autoreactivity when matched.	(Zhu et al., 2013a)
Natural HIV infection	4 healthy controls, 4 subjects with SLE, and 4 HIV-1 infected therapy-naïve individuals and 4 HIV-1 infected individuals receiving combination antiretroviral therapy	Multiplex PCR of IgH and 454 sequencing	HIV or SLE subjects have increased clonality within their IgHM repertoire compared to healthy individuals. Antiretroviral therapy failed to reduce IgHM clonality in HIV-infected individuals, but IgHV-D-J gene combinations within the IgHM repertoire were found to be similar to known broadly neutralising HIV-1 antibodies.	(Yin et al., 2013)
Natural HIV infection	An acute HIV-1 infection donor was followed for more than 3 years starting from approximately 4 weeks after HIV-1 infection	Multiplex PCR of IgH and 454 sequencing	Study of the evolution and structure of a broadly neutralizing antibody from an African donor followed from the time of infection. This antibody, CH103, neutralized approximately 55% of HIV-1 isolates, which co-crystallised with the HIV-1 envelope protein gp120, thus revealing a new loop-based mechanism of CD4-binding-site recognition. This study showed virus and antibody co-evolution and maturation.	(Liao et al., 2013)
Plasmodium falciparum (Pf)				
Natural Pf infection	Classical memory B-cells and atypical memory B-cells from peripheral blood.	Single cell heavy and light chain PCR followed by Sanger sequencing	Natural Pf infection induces the development of memory B-cells from 67 healthy adults with neutralizing serum IgG activity against asexual blood stage parasites from a highly endemic area in Gabon that produce broadly neutralizing antibodies against blood stage Pf parasites, but only atypical memory B-cells, rather than classical memory B-cells, show signs of active antibody secretion	(Muellenbeck et al., 2013)

Table 1.8. Summary of studies of B-cell repertoires from vaccinations.

Vaccine*	Cells used	Methodology	Key findings	References
Influenza vaccine				
TIV	1 healthy volunteer vaccinated against seasonal influenza, hepatitis A, and hepatitis B, and 2 healthy volunteers vaccinated against seasonal flu vaccine, sampled at multiple time points before and after vaccinations	Multiplex PCR of IgH and 454 sequencing	IgHV and J gene usage differs between individuals, and is conserved within individuals over time. CDR3 clustering into clonal groups showed clonal expansion and contraction in response to the vaccine with different participants exhibiting different dynamics. A small number of highly mutated, persistent clones were found within all individuals, potentially corresponding to long-lived B-cell memory or indicative of chronic infection.	(Laserson et al., 2014)
TIV	Memory B-cells 14 days after vaccination from 1 participant	High-throughput single cell heavy and light chain linkage PCR and 2x250bp Illumina sequencing	The accuracy of heavy and light chain pairings identified using a high-throughput method was validated. Identification of 240 putatively influenza-specific heavy and light chain CDR3 pairings.	(DeKosky et al., 2013)
TIV or LAIV	PBMCs on the day of vaccination, and on day 7 and 28 after two vaccinations given a year apart from 28 individuals	IgH-specific RT and second strand synthesis, and PCR using barcoded priemrs. Custom 100x120bp Illumina sequencing.	Demonstrated different repertoire dynamics after TIV and LAIV vaccination. TIV induced a stronger response, with more abundant IgG lineages than LAIV. Shared antibody sequences on day 7 after two TIV vaccinations were found, where these lineages are present after the second vaccination potentially due to memory B-cell recall. Suggested that this method could be used to identify cross-reactive antibodies.	(Vollmers et al., 2013)
TIV	PBMCs on 18 time-points around two vaccinations given a year apart from 1 participant, and 10 time-points around one vaccination from 2 participants).	IgH-specific multiplex PCR, and 454 sequencing	IgHV and J gene usage differs between individuals, and is conserved within individuals over time. CDR3 clustering into clonal groups showed clonal expansion and contraction in response to the vaccine with different participants exhibiting different dynamics. A small number of highly mutated, persistent clones were found within all individuals, potentially corresponding to long-lived B-cell memory or indicative of chronic infection.	(Laserson et al., 2014)
Tetanus vaccine				
TT	Plasmablasts 7 days after vaccination (one participant)	High-throughput single cell heavy and light chain linkage PCR and 2x250bp Illumina sequencing	Identified 86 putatively TT-specific heavy and light chain pairings, 10 of which where cloned into HEK293K cells followed by competitive ELISA of the antibodies produced showed them to be TT-specific.	(DeKosky et al., 2013)
TT	Bulk plasmablasts, memory B-cells, and antigen-specific plasmablasts 7 days and 3 months after vaccination from 2 participants	Heavy and light chain-specific multiplex PCR, and 454 sequencing. High-throughput single cell of heavy and light chain linkage PCR from day	Analysed the serum antibody repertoire by using the IgH sequence database to interpret results from high-resolution liquid chromatography tandom mass spectrometry of the serum antibodies. This showed that ~5% of the plasmablast clonotypes identified by sequencing at day 7 could subsequently also be detected in the serological response 9 months after vaccination	(Lavinder et al., 2014)

*Abbreviations: LAIV=live attenuated influenza vaccine, TIV=trivalent influenza vaccine, TT=tetanus vaccine

1.2.3.5. *In vivo* B-cell evolutionary processes

The immune system is capable of continually learning and memorising immunological experiences. The study of B-cell dynamics by B-cell cell receptor sequencing has been useful in the understanding of affinity maturation and selection of resulting mutants. Initially, using small sequence datasets per clone, lineage analysis became a popular analytical tool to understand mutational processes (**Figure 1.11**) (Steiman-Shimony et al., 2006a, Dunn-Walters et al., 2004, Barak et al., 2008, Frumkin et al., 2005, Uduman et al., 2014, McIntyre et al., 2014, Messmer et al., 2005b, Steiman-Shimony et al., 2006b, Bankoti et al., 2014, Sok et al., 2013, Green et al., 2013, Bergqvist et al., 2013, von Budingen et al., 2012, Seifert and Kuppers, 2009, Spencer et al., 2009). Lineage trees describe the clonal relationships between related cells within a lineage, where the root of the tree is assumed to be the germline sequence. Both the internal nodes as well as the leaves can represent sequences, as intermediate sequences can be included in the sample. Furthermore, lineage trees are not necessarily binary as a single B-cell can produce a population of identical cells that can produce mutations (Barak et al., 2008).

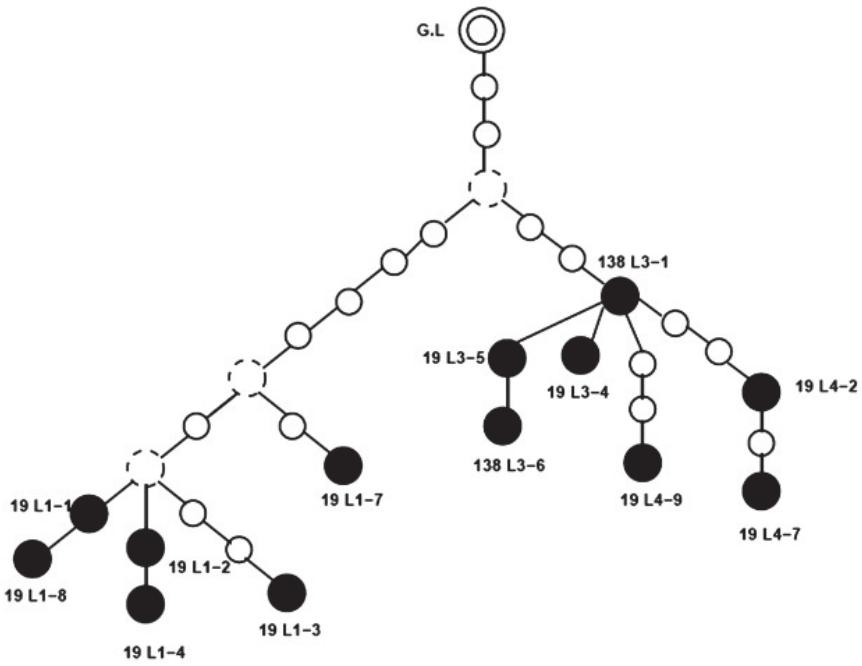

Figure 1.11. Lineage tree constructed by IgTree.

BCR sequences were from a patient with light chain amyloidosis (Manske et al., 2006). The trees are rooted with the nearest germline sequence (double circle). The filled circles represent the sampled sequences (named beside the nodes), dashed circles represent nodes that have more than one descendant, and solid white cycles represent nodes that have only a single descendant, inferred by IgTree©. Figure from (Barak et al., 2008).

Different phylogenetic methods are employed to analyse larger sequencing datasets, and using different hypotheses of evolution. Maximum likelihood methods estimate phylogenetic relationships by determining the theoretical likelihood of query sequences arising from a given ancestor by somatic hypermutation (Kepler et al., 2014, Kepler, 2013, Liao et al., 2013, Zhu et al., 2013b, Wu et al., 2011, Doria-Rose et al., 2014). Neighbour-joining methods use agglomerative clustering to generate trees representing sequencing relationships, and is typically faster for large datasets (Liao et al., 2013, Wu et al., 2011, Logan et al., 2011). Maximum parsimony methods assume that populations of cells, such as tumour cells or B-cell clones, develop in a parsimonious manner, such that the evolutionary process to create the population is minimal. Maximum parsimony assumes minimal number of explicit assumptions, thus useful when the true evolutionary process in B-cells is unknown or temporally variable (Campbell et al., 2008, Sutton et al., 2009, Rossi et al., 2012, Dagklis et al., 2012, Benichou et al., 2013). For example, Campbell *et al.* fitted unrooted parsimony models to generate phylogenetic trees for the malignant clones in 2 chronic lymphocytic leukaemia (CLL) samples to show the evolutionary relationships among the subclones and dominant clone of CLL cells (**Figure 1.12**) (Campbell et al., 2008). Although bootstrapping shows uncertainty in the ancestral relationships between individual subclones, there is strong support for 3 different classes of subclones: B-cells representing the intermediate stages between germline and the dominant clone, blind alleys representing divergent evolution away from the germline sequence, and ongoing evolution from the dominant clone. The persistence of B-cells from the intermediate stages suggest that initiating driver mutation(s) may have led to leukaemogenesis at the earliest branch-point of the tree.

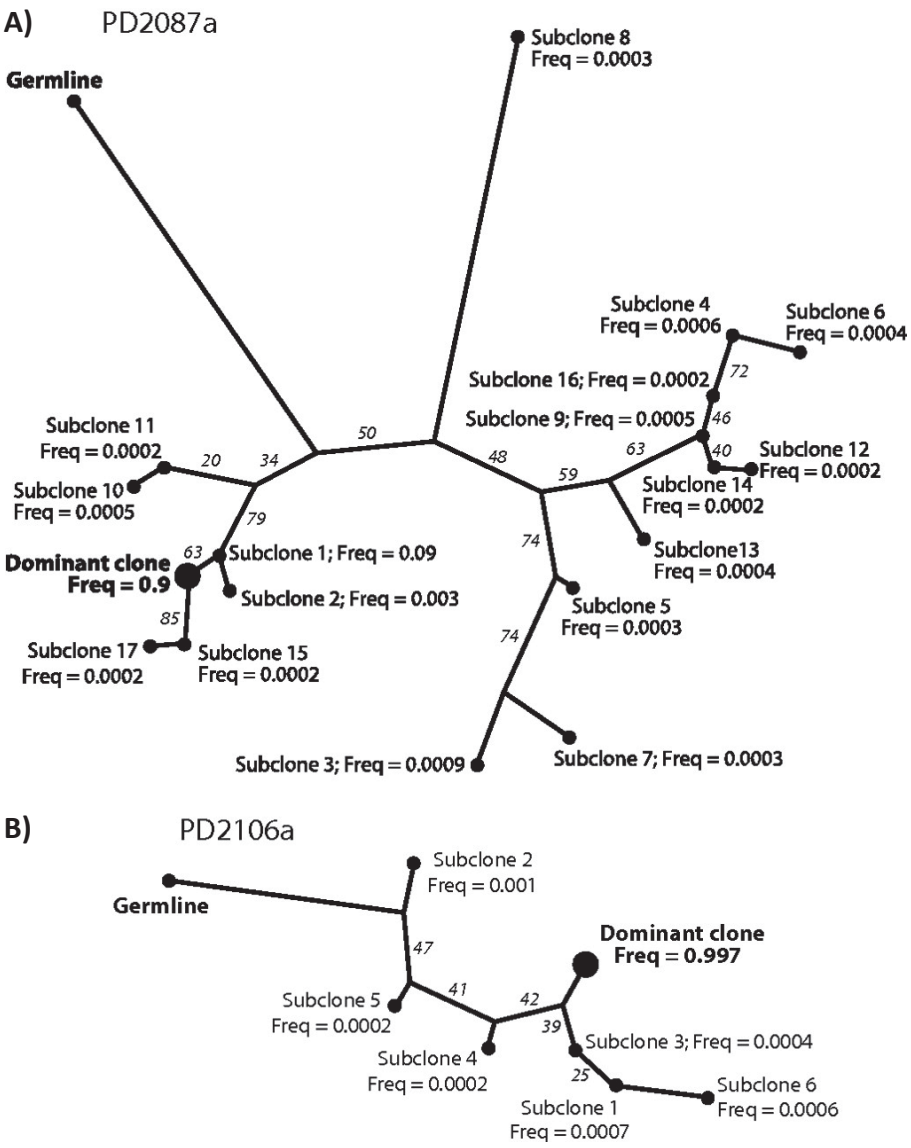

Figure 1.12. Maximum parsimony trees of B-clones.

The phylogenetic interrelationships between BCR clones for 2 chronic lymphocytic leukaemia patients, where the length of each branch proportional to the number of varying bases (evolutionary distance). The percentage support across 1000 bootstrap samples is given beside each intermediate branch. From (Campbell et al., 2008).

1.3. Chronic lymphocytic leukaemia (CLL)

1.3.1. Aetiology and epidemiology

Chronic lymphocytic leukaemia (CLL) is the most common form of leukaemia, representing 30% of all leukaemias. The incidence rate for CLL is 4.92 per 100,000 per year in Europe (Sant et al., 2010). The rates of CLL vary between populations; 35-40% of all leukaemia in Denmark is CLL, but only 3-5% in Chinese and Japanese populations (Redaelli et al., 2004). Incidence rates are higher for men than women, and increase with age, with two thirds of patients older than 60 years of age (Zenz et al., 2007). The clinical course of CLL is highly heterogeneous across individual patients (Morabito et al., 2011). Many CLL patients are asymptomatic, and remain treatment free for many decades, while an aggressive form of the disease can affect others. Patient conditions may deteriorate with the disease or may suffer from therapy related treatments (Morabito et al., 2011). Therefore, biological indicators of disease progression and prognosis are of great clinical importance. Identifying the risk factors associated with requirement early treatment or better prognosis estimation will decrease the treatments given to patients with the non-aggressive disease, with the majority of treatments carry significant toxicities.

1.3.2. Biology, pathogenesis and diagnosis of CLL

The diagnosis of CLL is made on two criteria. Firstly, if greater than 5×10^9 cells/L peripheral blood B-cells for at least 3 months, where clonality of circulating B-cells needs to be confirmed by flow-cytometry (Eichhorst et al., 2011, Hallek et al., 2008). CLL typically has preferential kappa or lambda immunoglobulin light chain usage at a ratio of greater than 3:1 or less that 1:0.3 respectively (Rozman and Montserrat, 1995, Cheson et al., 1996, Kilo and Dorfman, 1996). Secondly, leukaemia cells found in blood smears are small, mature B-cells with a narrow border of cytoplasm and dense nucleus with partially aggregated chromatin and lacking distinct nucleoli. CLL B-lymphocytes co-express CD19, CD5 and CD23, with weak or no expression of CD20, CD79b, FMC7 and surface immunoglobulin.

Monoclonal B-cell lymphocytosis (MBL) is thought to be a pre-clinical manifestation of CLL. The diagnostic criteria for MBL is exhibiting less than 5×10^9 cells/L peripheral blood B-cells for at least 3 months (Eichhorst et al., 2011, Hallek et al., 2008) along with either (a) kappa or lambda immunoglobulin light chain usage at

a ratio of greater than 3:1 or less that 1:0.3 respectively, (b) greater than 25% of B-cells expressing low-level or no surface immunoglobulin, or (c) a disease-specific immunophenotype, such as CD5+.

CLL manifests as an increasing collection of B-cells with related BCRs (malignant B-cell clone) that exhibit a wide range of phenotypic states, illustrated by the expression of different cell-surface proteins. Typical CLL is characterised by the accumulation of mature CD5+ B-cells in the blood, bone-marrow and secondary lymphoid organs (Chiorazzi et al., 2005). Unlike most tumour entities, only a small proportion of CLL cells proliferate, potentially acting as tumour stem cells (Messmer et al., 2005a), suggesting accumulation of CLL cells *in vivo* is not due to increased proliferation rates, but rather due to resistance to apoptosis (Chiorazzi et al., 2005). Evidence for CLL resistance to apoptosis includes both an anti-apoptotic expression profile, such as high expression of Bcl-2 protein (Inamdar and Bueso-Ramos, 2007, Mauro et al., 1999), and micro-environmental signals. Evidence for the latter is that CLL cells cultured without support *in vitro* rapidly undergo apoptosis, which can be prevented by co-culture with supporting stromal cells. Different types of stromal cells assist in survival of CLL cells *in vitro* and thought to be an integral part of the CLL microenvironment. These include monocyte-derived nurse-like cells (NLCs, a subset of large, round, adherent cells (CD14+ cells) that differentiate *in vitro* on co-culture with CLL or healthy B-cells) (Burger et al., 2000, Bhattacharya et al., 2011), CXCL12-expressing mesenchymal stromal cell (MSCs)(Burger et al., 2000, Eisele et al., 2009), or follicular dendritic cells (FDCs) (Pedersen et al., 2002). However, normal B-cells are not supported in this manner. It has recently been established that a signalling pathway for CLL B-cell survival and apoptotic resistance is activated by upregulation of protein kinase C (PKC)-βII expression on contact with stromal cells. Unregulated stromal PKC-βII in biopsies from patients with CLL, acute lymphoblastic leukaemia, and mantle cell lymphoma suggests that this pathway may commonly be activated in a variety of haematological malignancies (Lutzny et al., 2013).

The signs and symptoms of CLL gradually develop, therefore the onset of disease is difficult to identify. The disease is often discovered accidentally as a result of elevated lymphocyte counts during routine physician visits (Andritsos and Khoury, 2002). Asymptomatic CLL is seen in about 25% of patients, where the duration of the asymptomatic phase is highly variable (Inamdar and Bueso-Ramos, 2007). The early

signs of disease include persistent lymphocytosis, mild cervical, supraclavicular, and/or axillary nodes lymphadenopathy and splenomegaly. Thrombocytopenia and mild anaemia is seen in approximately 25% and 50% of patients respectively. Nodular and diffuse skin infiltrations, exfoliative dermatitis, erythroderma, and secondary skin infections are seen in about 5% of patients (Bonvalet et al., 1984, Cerroni et al., 1996). Disease progression can lead to organ infiltration, adenopathy with splenomegaly, hypersplenism, and peripheral cytopenias. These patients can present with weight loss, fever and night sweats. Advanced disease exhibits extensive bone marrow infiltration by neoplastic cells. Due to replacement of marrow by tumour cells, symptoms include severe anaemia, thrombocytopenia, and neutropenia (Inamdar and Bueso-Ramos, 2007).

CLL patients have increased frequency of abnormal immune manifestations, including immunodeficiency and autoimmunity despite the increased number of B-cells (Dearden, 2008). Approximately half of CLL patients have hypogammaglobulinemia (Hudson and Wilson, 1960). Bacterial infections are responsible for the majority of illnesses in patients with CLL, particularly infections of the respiratory tract, urinary tract, and skin, as well as viral infections. These infections contribute highly to patient morbidity and mortality. Many patients have poor responses to vaccination (Dearden, 2008, Shaw et al., 1960), where vaccine response is correlated with better CLL patient outcome and treatment-free survival (Dearden, 2008).

CLL is frequently associated with autoimmune conditions. Coombs' positive autoimmune haemolytic anaemia is seen in up to 25% of patients at some point during the course of the disease (Dameshek and Schwartz, 1959, Pisciotta and Hirschboeck, 1957). This condition involves the production of antibodies against red blood cells during or after developing CLL. Approximately 6% of patients develop red cell aplasia, and a subset of CLL patients develop auto-antibodies against platelets and neutrophils leading to thrombocytopenic purpura and neutropenia. Bence Jones paraproteinemia is seen in 65% of patients (Diehl and Ketchum, 1998).

1.3.3. Monoclonal B lymphocytosis as a possible pre-leukemic phase

Monoclonal B-cell lymphocytosis (MBL) is thought to be a pre-clinical manifestation of CLL, characterised by asymptomatic B-cell clonal expansions with surface phenotypes similar to that of CLL (Marti et al., 2007, Marti et al., 2005).

MBL has been detected in older adults in good health (Shim et al., 2014). The prevalence has been reported in a number of studies, ranging from <1% (Rachel et al., 2007, Shim et al., 2007) to 18%, depending on the detection methods and tested populations (Shim et al., 2010). MBL is more frequent in males, with prevalence significantly higher in individuals with relatives with CLL, and increases with age (Rawstron et al., 2002). However the incidence of MBL is approximately 100 times greater than that of CLL, and therefore cannot be taken to be a definitive sign of genuine neoplastic transformation (Ghia et al., 2000). Some CLL-like MBL clones can be present at much higher frequencies in the blood, with a 1-2% per year rate of progressing to symptomatic CLL (Rawstron et al., 2008, Shanafelt et al., 2009). The natural history of MBL is not well understood.

1.3.4. Disease staging in CLL

The Rai stage was first prognostic staging process to be developed for CLL, using a combination of lymphadenopathy (abnormal enlargement of lymph nodes), organimegaly (abnormal enlargement of organs), and cytopenias (anaemia and thrombocytopenia (platelet number reduction)) to determine five prognostic groups with median survivals given in Table 1.9 (Rai et al., 1975).

Table 1.9. Rai stage median survival.

Rai Stage	Risk level	Prognosis factors	Median survival
Stage 0	Low	Lymphocytosis	> 150 months
Stage 1	Intermediate	Lymphocytosis + Lymph node enlargement	101 months
Stage 2	Intermediate	Lymphocytosis + Spleen/liver enlargement	71 months
Stage 3	High	Lymphocytosis + anaemia	19 months
Stage 4	High	Lymphocytosis + thrombocytopenia	19 months

Adapted from Rai *et al.* (Rai et al., 1975).

The Binet staging system was also developed for CLL, which relied on the number of affected lymphal areas and cytopenias, summarized in Table 1.10 (Binet et al., 1977). The Rai and Binet staging systems provide prognosis for the patient as well as the appropriate time for patient therapy. However, there is significant heterogeneity of outcomes at the different stages, so new and more accurate prognostic markers in CLL are of great clinical interest.

Table 1.10. Binet stage median survival.

Binet Stage	Risk level	Prognosis factors	Median survival
Stage A	Low	Lymphocytosis + less than 3 enlarged lymphal areas	> 12 years
Stage B	Intermediate	Lymphocytosis + more than 3 enlarged lymphal areas	7 years
Stage C	Intermediate	Lymphocytosis + anaemia or thrombocytopenia	2 years

Adapted from Binet *et al.* (Binet et al., 1977).

1.3.5. Prognostic markers in CLL

Recurring genomic abnormalities with prognostic significance have been identified in genetic studies using interphase fluorescence *in situ* hybridisation (FISH) and chromosomal analysis in CLL (Oscier et al., 2002, Juliusson et al., 1990). Many reports have associated CLL prognosis with genomic aberrations, summarised in Table 1.11.

Table 1.11. Genomic markers in CLL associated with prognosis.

Marker type	Genomic/chromosomal markers	Relative prognosis	Reference
Deletions	Deletions in 11q, 17p	Poor	(Krober et al., 2006)
Deletions	Deletions in 13q	Good	(Krober et al., 2006)
Deletions	Deletion in 6q	Intermediate	(Cuneo et al., 2004)
Mutations	TP53, ATM (tumour suppressor genes)	Poor	(Zenz et al., 2010)
Mutations	IRF4, Bcl-2 polymorphism	Good	(Havelange et al., 2011)
Mutations	Bcl-6 mutation	Poor	(Sarsotti et al., 2004)
Mutations	MDM2 SNP	Poor	(Gryshchenko et al., 2008)
IgVH mutational status	IgVH mutated	Good	(Schroeder and Dighiero, 1994, Fais et al., 1998, Damle et al., 1999, Hamblin et al., 1999)
	IgVH unmutated	Poor	
Gene expression	ZAP-70 (correlates with mutational status)	Poor	(Krober et al., 2006)
	V3-21 gene usage	Poor	(Krober et al., 2006)
Micro RNAs	Micro RNA signature associated with prognosis	-	(Calin et al., 2005)
Telomere length	Longer telomere length (correlates with mutational status)	Good	(Grabowski et al., 2005)

There are four prognosis markers that are currently widely in clinical use:

1.3.5.1. IgHV mutational status

Studies of CLL in some patients have shown that CLL cells do not possess any somatic hypermutations in the complementary determining regions (CDRs) of the immunoglobulin genes, whereas other patients have highly mutated BCRs (Cai et al., 1992). It has been suggested that the two different mutational statuses of CLL patients may be derived from two different stages of B-cell ontology, with the unmutated CLL cases corresponding to pre-antigenic stimulation, and the mutated cases corresponding to post-antigenic stimulation (Hamblin et al., 1999, Damle et al., 1999). The examination of IgHV genes in CLL patients have shown that the two subsets of CLL that are prognostically significant, with studies suggesting an inferior survival and high likelihood of requiring early treatment in patients with unmutated IgHV. For example, Hamblin *et al.* found that the median survival for stage A patients with mutated CLL was 293 months (n=46) compared to 95 months for unmutated CLL (n=38) (p-value=0.0008) (Hamblin et al., 1999). Furthermore, leukemic cells from unmutated patients tend to produce polyreactive antibodies (Martin et al., 1992, Herve et al., 2005b), similar to natural autoantibodies (Caligaris-Cappio and Ghia, 2008, Mouquet and Nussenzweig, 2011).

Commercially available assays are used to determine mutational status of the CLL, which relies on capillary sequencing of reverse transcribed/PCR products of peripheral blood and bone marrow aspirate from each patient. The percentage BCR mutation is calculated by comparing the IgHV sequences to the germline sequence database (difference of > 2% from germline counterpart is classified as mutated BCR).

1.3.5.2. Interphase fluorescence in-situ hybridization

Interphase fluorescence in-situ hybridization **(**iFISH) has shown that trisomy 12, and deletions in 13q14 are correlated with IgHV mutational status in CLL (Hamblin et al., 1999, Damle et al., 1999). Trisomy 12, and chromosome 13 and 14 abnormalities are the most common genomic aberrations associated with CLL.

Juliusson *et al.* found that CLL patient with a normal karyotype (n=173) had a median overall survival of greater than 15 years, compared to 7.7 years for karyotypic abnormalities (n=218) (Juliusson et al., 1990). In addition, patients with single karyotypic abnormalities (n=113) had better prognosis than those with complex karyotypes (p-value<0.001). Within the subset of patients with single karyotypic abnormalities, patients with trisomy 12 (n=67) had poorer survival than patients with chromosome 13q (n = 51) (p-value=0.01), where the latter has the same survival as those with a normal karyotype. Patients with chromosome 14q had significantly poorer survival than those with abnormalities of chromosome 13q (p-value<0.05).

1.3.5.3. CD38 expression on CLL B-cells

CD38 is a single chain type II transmembrane glycoprotein, which is expressed in a discontinuous manner in normal B-cell development. CD38 can be detected in high levels in B-cell precursors, germinal centre cells and plasma cells, and lower expression is usually seen in the peripheral blood and tonsillar B-cells of health individuals. CD38 function is thought to include complex ectoenzymatic activity and signal transduction for regulation of cell proliferation and survival (Kumagai et al., 1995). However, CD38 expression has been seen in a proportion of CLL patients, and is correlated with survival outcomes and a increase probability of requiring treatment, including continuous chemotherapy or chemotherapy with two or more agents or regimens (Durig et al., 2002). Specifically a CD38-negative patient group required minimal or no treatment, remained treatment-free for a longer time period and had prolonged survival (p-value<0.05 between CD38-negative CLL subgroup (<20% of the CLL cells expressed membrane CD38, n = 77) and CD38-positive CLL subgroup (≥20% of the CLL cells expressed membrane CD38, n = 56)).

1.3.5.4. Zeta-associated protein-70 expression on CLL B-cells

Zeta-associated protein (ZAP-70) expression is involved in T-cell receptor signal transduction (Elder et al., 1994, Iwashima et al., 1994). Normal B-cells do not express ZAP-70, but has been shown to be overexpressed in IgHV unmutated CLL by microarray analyses, and serves as a surrogate marker for IgHV mutational status (Rosenwald et al., 2001, Klein et al., 2001, Wiestner et al., 2003). *In vitro*, ZAP-70 is

involved in the signal transduction cascade initiated by BCR stimulation in IgHV unmutated CLL (Chen et al., 2002, Chen et al., 2005, Crespo et al., 2003). Importantly, there is a statistically significant inferior clinical course for CLL patients with ZAP-70 expression (Krober et al., 2006, Schroers et al., 2005).

1.3.5.5. CLL proliferation centres

B-cell signalling pathways have been found to be associated with proliferative potential in neoplastic cells in CD38-positive, ZAP-70-positive and unmutated CLL patients, and it has been suggested that stimulation may occur in the "pseudo-follicular" proliferation centres (PCs) (Ferrarini and Chiorazzi, 2004). PCs are focal aggregates of variable sizes scattered in the lymph nodes, and their presence are observed in CLL (Ratech et al., 1988, Granziero et al., 2001, Schmid and Isaacson, 1994, Soma et al., 2006). Similar proliferation centres have also been seen in the inflamed tissues of patients with systematic autoimmune and inflammatory disorders, such as rheumatoid arthritis (Takemura et al., 2001) and multiple sclerosis (Corcione et al., 2005). It is thought that the clustering of pro-lymphocytes and CLL cells forms pseudo-follicular proliferation centres, where small lymphocytes accumulate and overflow into the peripheral blood. The PC microenvironment consists of pro-lymphocytes and CLL cells intermixed and surrounded by CD3+ T-cells (most of which are CD40 and CD40+), which are in close contact with the proliferating malignant B-cells (Ghia et al., 2002). Follicular dendritic cells have been observed in some PC (Ratech et al., 1988, Schmid and Isaacson, 1994), along with stromal cells and accessary cells interspersed with the small lymphocytes (Caligaris-Cappio, 2003, Burger and Kipps, 2006).

1.3.6. Current treatments for CLL

In CLL, the decision to treat is guided by clinical staging, symptoms, and disease activity. Patients in early stages of disease (Rai 0-II or Binet A) are generally only monitored but not treated unless associated CLL symptoms occur. It has only been shown that treatment is beneficial for patients at later stages (Rai III-IV or Binet B-C), but no statistical difference in outcome has been found by treating patients at earlier stages (Rai I-II or Binet A). Disease activity is typically monitored by

lymphocyte doubling time (the time it takes for the number of lymphocytes to double) of less than 6 months or by the rapid growth of lymph nodes, and is often an indication to commence treatment (Hallek and German, 2005). Although the aim for treatment is disease eradication, most patients who have a complete response typically have minimal residual disease. The term minimal residual disease (MRD) refers to low-level disease, often after incompletely effective chemotherapy (Paietta, 2002). The general consensus for MRD level is between 0.01% and 0.035% leukemic cells within a morphologically normal appearing bone marrow, as the detection of less than 0.01% of leukaemic cells by flow cytometry may not be reliably reached due to variability of technical expertise in different clinical laboratories (Campana, 2010, Coustan-Smith et al., 2000, Paietta, 2002). Detection of disease relapse after therapy is of great clinical importance, particularly to determine if further CLL therapy is required. Flow-cytometry and real-time quantitative polymerase chain reaction techniques are typically used for clinical monitoring of MRD (Moreton et al., 2005). The main therapies available to CLL patients are:

- *Alkylating agents ± prednisone (chlorambucil, cyclophosphamide)*

Chlorambucil and other alkylating agents can bind to cellular structures such as membranes, RNA, proteins and DNA. It is thought that DNA cross-linking is the most important mechanism for anti-tumour activity. Despite the relative benefit to some patients with the use of chlorambucil, drug resistance and relapse remains a problem. Furthermore, CLL cells are typically not highly proliferative, therefore raising questions about the anti-tumour activity of the drug (Begleiter et al., 1996). Prednisone is a corticosteroidal immunosuppressant drug which acts by repressing the activity of transcription factors such as activating protein-1 (AP-1) and nuclear factor-κB (NF-κB), thus inhibiting cytokine production, changes the expression of various oncogenes, induces cell-cycle arrest and apoptosis (Inaba and Pui, 2010).

- *Combination chemotherapy with alkylating agents (COP, R-CHOP)*

R-CHOP is composed of rituximab, cyclophosphamide, hydoxydaunorubicin, oncovin, prednisone respectively for the letter abbreviations, and COP is composed of cyclophosphamide, vincristine, prednisone. Hydoxydaunorubicin prevents DNA and RNA from replicating, oncovin inhibits during the M phase of the cell cycle and

prednisone is anti-inflammatory. The synthetic corticosteroid drug, prednisone, is an immunosuppressant used in the treatment of some inflammatory diseases, such as allergies, as well as cancer in higher doses. Prednisone is converted via hepatic metabolism to prednisolone, which irreversibly binds to the alpha and beta glucocorticoid receptors. The glucocorticoid receptor-prednisone complexes dimerise, and interact with nucleic DNA leading to gene transcription alterations. However, the long-term use of prednisone and other steroids has been associated with development of osteoporosis, where bone loss is observed in approximately 50% of patients taking 7.5mg prednisone for more than 3 months, and 25% of patients developed osteoporotic fractures, and further patients developed osteonecrosis (Van Staa et al., 2000). Prednisone and other steroids have been shown to promote apoptosis of osteoblasts and osteoclasts, as well as reducing the recruitment of osteoblasts and osteoclasts from progenitor cells (Weinstein et al., 1998). Therefore, the suitability of these combination therapies need to assessed on an individual basis, particularly in reference to the risk of developing osteo-related complications.

In a randomised study of 287 stage B CLL patients, treatment response was improved with CHOP (n=147) compared to chlorambucil plus prednisone (n=140) (p-value=0.007, chi-square test), but showed no difference in survival (p-value=0.33, score test). However, for stage C CLL patients, there were no significant differences treatment response and survival between CHOP (n=44) or CHOP plus methotrexate (n=46). Therefore, even though CHOP has been shown to improve therapy response, questions remain about its effectiveness at treating advanced CLL patients (Binet, 1994). These treatments are used in other B-cell malignancies, where 2-year and 5-year follow-up studies have shown that the outcome of elderly diffuse large B-cell lymphoma (DLBCL) patients on R-CHOP therapy regimens have shown significant increases the rate of complete response, decreases the rates of treatment failure and relapse, better event-free survival and overall survival compared to CHOP alone (Coiffier et al., 2002, Feugier et al., 2005).

- *Purine analogs*

Fludarabine, pentostatin, and cladribine are the three purine analogs currently used in CLL. Pentostatin inhibits adenosine deaminase by mimicking adenosine, thus reducing the cell's capability to process DNA (Sauter et al., 2008) and typically used in patients who have relapsed as well as those with acute graft-versus-host disease

(Bolanos-Meade et al., 2005). Cladribine works by a similar manner as pentostatin, with complete response and overall response rates similar to fludarabine (Robak, 2001), although 18-42% of patients experience fever side-effects after cladribine infusion (Van Den Neste et al., 1996, Saven et al., 1999). Cladribine is also used in the treatment of treatment of symptomatic hairy cell leukaemia, and is in clinical trials for use in the treatment of multiple sclerosis (Giovannoni et al., 2010). Fludarabine inhibits DNA synthesis by hindering ribonucleotide reductase and DNA polymerase. Fludarabine affects both resting and dividing cells, therefore works on both cancerous and healthy cells. Fludarabine monotherapy produces the best longer overall survival rates, but combination with Chlorambucil has shown some increased benefit (Rai et al., 2000, Johnson et al., 1996). Fludarabine has been combined with purine analogs (such as low-dose fludarabine with cyclophosphamide ± mitoxantrone), that have been shown to be effective in a subset of elderly CLL patients while with low infectious complications and negligible toxic side-effects (Marotta et al., 2000).

- *Monoclonal antibodies (campath-1H, rituximab)*

Campath-1H (alemtuzumab) is a humanised anti-CD52 antibody, an antigen on the surface of normal and malignant lymphocytes. This treatment has also been approved for the treatment of multiple sclerosis. However, the exact mechanism of campath-1H is not fully defined (Hu et al., 2009).

Rituximab is a CD20-specific monoclonal antibody that causes potent antibody-mediated B-cell cytotoxicity. Depletion of circulating B-cells from the pre-B-cell stage to the pre-plasma cell stage can lead to reduction and, in some cases, remission of CLL (Grillo-Lopez et al., 2002, Cragg et al., 2005). However, germinal centre B-cells have been found to be resistant to killing, potentially due to poor tissue penetration by rituximab (Grillo-Lopez et al., 2002). Rituximab is now being used in patients with autoimmune disease (Buch et al., 2011). Sometimes, a combination of chemotherapy and immunotherapy is used, such as fludarabine, cyclophosphamide, rituximab, fludarabine and campath-1H.

- *Transplantation (auto, allo, RIC)*

The use of auto-grafting in CLL is being an increasingly frequent treatment option. However, complications can be caused by fludarabine in stem-cell harvesting.

Hematopoietic stem cell transplantation (SCT) has been explored in clinical trials in younger patients with associated adverse disease risk factors. Although autologous SCT is not curative, it has a low treatment-associated mortality rate. Only a small number of patients are offered myeloablative allogeneic SCT due to high treatment-associated morbidity and mortality (Gribben, 2009).

1.3.7. B-cell receptors in CLL

B-cell malignancies have been found to typically express dominant clonal IgH receptors (Arber, 2000), and a variety of assays have been developed to assess B-cell clonality for diagnosis of B-cell cancers, such as in CLL and mantle cell lymphoma (MCL) (Campbell et al., 2008).

The suggestion that CLL B-cells are selected by antigenic pressure is reinforced by a number of studies showing highly restricted and biased IgHV gene usage in the B-cell repertoire of CLL patients compared to normal adult repertoire (Kipps et al., 1989, Herve et al., 2005a, Schroeder and Dighiero, 1994, Fais et al., 1998, Chiorazzi and Ferrarini, 2003, Stevenson and Caligaris-Cappio, 2004, Tobin et al., 2004a, Ghiotto et al., 2004, Messmer et al., 2004, Widhopf et al., 2004, Tobin et al., 2004b). Similar CLL BCRs are expressed between different CLL patients arising from common V-(D-)J gene usage in the heavy and light chains that share structural features such as CDR3 length, amino acid composition and joining regions, such as IgHV1-69 with IgHJ6 in the unmutated CLL, and IgHV4-34 in the mutated CLL. These stereotyped BCRs in CLL supports the hypothesis that BCR reactivity may play an important role in the CLL leukaemogenesis, potentially through activation by common antigen or auto-antigen (Tobin et al., 2004a, Ghiotto et al., 2004, Messmer et al., 2004, Widhopf et al., 2004, Tobin et al., 2004b).

CLL B-cells have been shown to express more than one IgHV allele in about 3.1% patients (Visco et al., 2013). This phenomenon can be explained either by the expression of two productive BCRs in a monoclonal CLL clone, or the presence of two distinct clonal expansions, known as bi-clonal CLL (Langerak et al., 2011). The prevalence of dual BCR expression in a single CLL clone has been reported in up to 5% of CLL cases, and thought to be due to incomplete allelic exclusion or secondary rearrangements of the IgH locus (Visco et al., 2013, Katayama et al., 2001, Rassenti and Kipps, 1997). Bi-clonal CLL is defined as the presence of two or more

phenotypically or morphologically distinct leukemic populations (Sanchez et al., 2003).

Multiple B-cell neoplasms are frequently encountered in patients, with associations of CLL with small lymphocytic lymphoma (SLL) follicular lymphoma (FL) (Boiocchi et al., 2012, Sanchez et al., 2006), and hairy cell leukaemia with CLL and SLL (Gine et al., 2002). Indolent B-cell lymphomas can develop into more aggressive disease, such as by Richter transformation or the transformation of FL to diffuse large B-cell lymphoma (DLBCL) (Boiocchi et al., 2012). Composite neoplasms can be clonally related, as suggested by related IgV gene rearrangements in cells from two lymphomas. Many of such cases have been shown to exhibit both shared and distinct somatic mutations, suggesting separate development of the lymphomas from a common premalignant precursor (Rosenquist et al., 2004b, Rosenquist et al., 2004a, Tinguely et al., 2003, van den Berg et al., 2002, Kuppers et al., 2001, Marafioti et al., 1999, Brauninger et al., 1999, Schmitz et al., 2005). Therefore, multiple B-cell neoplasms represent models to understand the transformation process in tumourigenesis and development of heterogeneous tumour populations from shared cancer precursors. Therefore, detection and monitoring of B-cell populations in lymphoid malignancies is of great clinical importance.

1.4. B-cell Acute lymphoblastic leukaemia

1.4.1. Aetiology and epidemiology

Acute lymphoblastic leukaemia (ALL) is the most common childhood leukaemia, where children account for two thirds of all ALL cases. Typically, children with ALL have a better prognosis than adult patients with ALL. Through the use of combinations of drug therapies, outlined in Section 1.4.4, between 80-90% of children are cured (Pui et al., 2008, Fielding, 2008), but the cure rate in adults is 30-40% (Pui et al., 2008). Relapse remains the leading cause of morbidity and mortality in children. The reasons for the difference in cure rates between children and adults is not fully understood, but thought to comprised of multiple factors including different therapeutic protocols between these groups and differences in biology between the disease groups.

1.4.2. Biology, pathogenesis and diagnosis of ALL

ALL is thought to develop from a single leukaemic progenitor cell with the capability of indefinite clonal expansion. Different subtypes of ALL are based on the stage of lymphoid differentiation at which leukaemogenesis occurred, either in the committed lymphoid B-cell (1-2%) or T-cell lineage (15-20%), or an early precursor B-cell (80%) or early precursor T-cell (~2%) (Reaman, 2002).

Diagnosis is confirmed by the presence of lymphoblasts on a bone marrow biopsy or aspirate, or peripheral blood smear, containing more than 20-25% of cells with the immunophenotype for ALL (Sabattini et al., 2010). Lymphoid lineage cells can be confirmed by immunophenotyping, which also distinguishes between B-cell and T-cell lineages as well as stage of differentiation (Huh and Ibrahim, 2000). Distinguishing acute myeloid leukemia (AML) from ALL is routinely achieved by staining leukemic cells for myeloperoxidase (MPO), where ALL is typically MPO-negative (Bennett et al., 1981, Bennett et al., 1976). Additional risk stratification and prognostic estimation of patients presenting with ALL include complete blood count, bone marrow and CNS involvement, cytogenetic studies, and tests for additional infections.

Common symptoms of ALL include fatigue (50%), fever (60%), pallor (skin paleness, 25%) and weight loss (26%). Bone pain caused by infiltration of blast cells

into the marrow cavity and periosteum occurs in 23% of patients (Dworzak and Panzer-Grumayer, 2003, Silverman and Sallan, 2003). The large burden of leukemic cells in patients with B- or T- cell ALL or B-cell precursor leukaemia commonly results in blood hyperkalemia (excess potassium), hyperuricemia (excess uric acid), and hyperphosphatemia (phosphate excess) with secondary hypocalcemia (low serum calcium). Therefore, intravenous hydration and sodium bicarbonate are often used to alkalise the urine, and hyperuricemia is treated with allopurinol, and hyperphosphatemia is treated with aluminum hydroxide or calcium carbonate (Pui et al., 1997). The peripheral blast-cell count can be reduced before chemotherapy by allopurinol, a purine synthesis inhibitor (Masson et al., 1996). Infiltration and involvement of the central nervous system (CNS) is found in <5% of children with ALL at presentation. The symptoms of CNS involvement include vomiting, headache, papilledema (swelling of the optic disc) and abducens nerve palsy (cranial nerve VI dysfunction) (Craig, 2003, Ma et al., 1997, Downing and Shannon, 2002). Fever is presented in at least half of ALL patients, either due to pyrogenic cytokines released from leukemic cells, including IL-1, IL-6 and tumor necrosis factor (Dinarello and Bunn, 1997) or from infection. These symptoms are often treated with broad-spectrum antibiotics until infection can be excluded (Hughes et al., 1987).

1.4.3. Prognostic markers in ALL

Individuals with Down's syndrome or ataxia telangiectasia have an increased risk of developing ALL (Hasle et al., 2000, Morrell et al., 1986). A number of prognostic factors in ALL have been determined, where some are co-associated with age (summarised in Table 1.12). Furthermore, additional genetic modifications have been found in relapsed ALL different to that seen in patients at presentation. Of significance are mutations in the histone acetyl transferase domain of cyclic adenosine monophosphate (cAMP) response element-binding protein found in approximately 20% of relapsed ALL cases, particularly in hyperdiploid ALL (seen in 60% of relapsed patients), thought to be a result of clonal selection during disease course rather than clonal evolution (Inthal et al., 2012).

Table 1.12. Genomic and cell-based prognostic factors in ALL.

Genomic aberration	Risk association*	Reference
Hyperdiploidy	Younger age and better prognosis	(Aguiar et al., 1996, Burmeister et al., 2010)
t(12;21) [ETV6/RUNX1] translocation	Younger age and better prognosis	(Aguiar et al., 1996, Burmeister et al., 2010)
t(9;22) [BCR/ABL1]	Older age and worse prognosis	(Secker-Walker et al., 1991)
Complex karyotype	Older age and worse prognosis	(Secker-Walker et al., 1991)
Hypodiploidy	Older age and worse prognosis	(Secker-Walker et al., 1991)
Janus kinase 1 and 2 mutations	Poor prognosis and associated with T-cell precursor ALL in adults	(Mullighan et al., 2009)
Ikaros family zinc finger protein 1 mutations	Worse prognosis	(Kuiper et al., 2010)
Cytokine receptor-like factor 2 translocations	Older age and worse prognosis	(Chen et al., 2012)
BCR-ABL1 translocations	Worse prognosis	(Roberts et al., 2012)
Intra-chromosomal amplifications of chromosome 21 (the gain of at least three copies of the RUNX1 region)	Worse prognosis	(Moorman et al., 2007)
Philadelphia chromosome	Worse prognosis	(Fielding et al., 2009)

Other risk factors	Risk association*	Reference
T-cell ALL	Worse prognosis	(Neumann et al., 2012)
B-cell ALL	Better prognosis	(Neumann et al., 2012)
Early T-cell precursor ALL ($CD3^{+}$, $CD5^{weak}$, $CD8^{-}$, $CD1a^{-}$ expression)	Worse prognosis	(Neumann et al., 2012)

* All studies based on comparing the outcomes of multiple patients with or without each corresponding risk factor, where statistically significant prognostic associations have p-values <0.05.

1.4.4. Current treatments for ALL

The ALL treatment regimen is typically determined by patient age and the Philadelphia chromosome status. Philadelphia chromosome is a chromosomal abnormality with the reciprocal translocation between chromosome 9 and 22 that has a statistically poor prognosis (Fielding et al., 2009). Children and young adults are treated with paediatric regimens, and Philadelphia chromosome-positive ALL patients receive tyrosine kinase inhibitor (TKI, such as such as imatinib) in addition to chemotherapy. Typically there are three treatment phases: (a) induction phase, (b) consolidation phase and (c) maintenance phase. The outline of these phases are discussed below (Larson et al., 1995, Kantarjian et al., 2004, Thomas et al., 2004, Rowe et al., 2005, Cortes et al., 1995).

Induction phase treatment

The aim of the induction phase is patient remission, defined by healthy blood cell counts, the absence of leukemic cells in the bone marrow and repopulation of the bone marrow with healthy cells. Combinations of chemotherapy drugs are used in this stage according to patient risk profile (such as defined in Table 1.12), but typically including vincristine (a mitotic inhibitor), dexamethasone or prednisone (as an anti-inflammatory and immunosuppressant drug), and doxorubicin, daunorubicin, or another anthracycline drug (DNA intercalating agent that inhibits DNA and RNA synthesis). Treatment of leukemic cells that have entered the CNS or to prevent leukaemic cells from entering CNS includes intrathecal chemotherapy often involving methotrexate (an antimetabolite). Radiation therapy may be used directly to the brain or spinal cord.

Consolidation phase treatment

For patients that achieve remission, a short course of chemotherapy is performed lasting a few months. Typically the same drugs are used as in the induction phase given in high doses. For high relapse risk patients (as defined in Table 1.12) and those with poor prognostic factors, allogeneic or autologous stem cell transplant can be given. CNS prophylaxis may be continued.

Maintenance phase treatment

A maintenance chemotherapy program of methotrexate and 6-mercaptopurine (an immunosuppressive drug) is given to patients after the consolidation phase. This phase usually lasts about 2 years. Additional drugs, such as imatinib, are given to Philadelphia chromosome-positive ALL patients, and CNS prophylaxis may be continued.

1.4.5. Monitoring minimal residual disease in ALL

MRD testing is routinely used in most paediatric ALL protocols and an increasing number of adult ALL trial protocols. The risk stratification and clinical significance of MRD depends intrinsically on the MRD assays used and time points tested (Bruggemann et al., 2010, Borowitz et al., 2003).

The clinical evaluation of treatment responses in ALL patients is achieved with a range of MRD assays. B-ALL and T-ALL cells have distinct clonal rearrangements in their B- or T-cell receptors respectively, and are often associated with the expression of gene fusions and leukaemia-associated immunophenotypes. Assays based on PCR or flow cytometry have the sensitivity to detect one ALL cell in at least 10^4-10^5 healthy cells from clinical samples (summarised in Table 1.13) (Campana, 2010). However, this may not be sensitive enough to detect MRD considering only a single or small number of B-cells is required for disease relapse and $>10^6$ B-cells are taken in a typical 10ml blood sample.

MRD is currently most frequently quantified using real-time quantitative PCR (qPCR) (van der Velden et al., 2007). Using immunoglobulin rearrangements in B-ALL patients or TCR rearrangements in T-ALL patients has proven to be sensitive and quantitative. However, in some patients, ongoing immunoglobulin or TCR rearrangements occur generating leukaemic subclones with distinct sequences, which can be undetected at diagnosis but become the dominant clone subsequently. Therefore, recommendations have been made to monitor two or more different rearrangement from diagnosis or to use additional MRD assays such as flow cytometry (van der Velden et al., 2007, Faham et al., 2012).

Genetic abnormalities are carried in most ALL cells. qPCR of gene fusions are most frequently used for MRD detection, such as BCR-ABL1, ETV6-RUNX1, MLL-AFF1, and TCF3-PBX1. These recurrent abnormalities are present and suitable for MRD monitoring in about 40% of ALL patients (Campana, 2010, Bruggemann et al., 2010). The benefits of qPCR are the rapidity of the procedure that does not require sequencing or patient-specific primer design, and the stable association of the gene fusion and the ALL clone. Although the use of mRNA for qPCR is typically more sensitive than DNA due to the higher copy number per cell, mRNA is prone to degradation leading to potential false negative results and the number of transcripts per cell for a fusion gene may be variable between patients, thus quantification is difficult (Gabert et al., 2003).

ALL cells express cell surface markers that resemble closely the origin of B- or T-lymphoid precursors. Healthy T-lymphoid precursors typically do not circulate but instead occupy the thymus. Therefore blood or bone marrow cells with cell surface markers resembling that of T-lymphoid precursors is sufficient to identify T-

ALL (Coustan-Smith et al., 2006). Detection of B-ALL cells typically relies on the aberrant expression of cell markers known as the leukaemia-associated immunophenotypes, which can be identified in more than 95% of B-ALL cases and typical for each ALL disease subtype (Campana, 2010).

An alternative MRD assay is next-generation sequencing of the BCR or TCR repertoires in B-ALL and T-ALL respectively. ALL cells typically arise from the leukaemic transformation of a single lymphoid precursor, therefore each B- or T-ALL cell has a unique B- or T-cell receptor rearrangement respectively that is a unique marker for the leukaemic clone in high-throughput sequencing data (Bruggemann et al., 2010). Markers for MRD in B-ALL can be determined at the time of diagnosis when the leukaemic B-cell load is greatest and the peripheral blood exhibits significant B-cell clonal expansion. B-cell BCR repertoire analysis has been used to identify IgHV-D-J gene usage in the leukaemic B-cell clone(s). Subsequent monitoring of the patient can be used to follow the leukemic B-cell population during and after therapy, enabling early detection of minimal residual disease (Brisco et al., 2009). This approach overcomes the requirement of patient-specific reagents while achieving a sensitivity of >1 in 10^6 cells (Logan et al., 2011, Faham et al., 2012, Gawad et al., 2012). Furthermore, this approach allows for the assessment of the total diversity of the B- or T-cell populations, thus able to follow the ALL sub-clonal evolution as well as the major malignant clone (Ladetto et al., 2013).

Table 1.13. The main clinical assays used to monitor MRD in acute lymphoblastic leukaemia.

Adapted from (Campana, 2010).

Moelcular target	Method	% of patients that can be monitored	Sensitivity for the detection of MRD	Strengths	Weaknesses
BCR and TCR gene rearrangements	qPCR	~90%	0.01%-0.001%	• High sensitivity • Consensus protocols established • Generally accurate quantification	• Laborious • Clonal evolution and secondary rearrangements may lead to false negatives • The requirement for more than one target reduces the applicability of the assay
Fusion transcripts	qPCR	~40%	0.01%-0.001%	• Rapid • Unambiguous association of fusion transcript with leukaemic/pre-leukaemic clone • Stable throughout therapy/ clonal evolution	• Quantification not certain • RNA instability may lead to false negatives
Leukaemic immunophenotype	Flow cytometry	~95%	0.01%	• Widely applicable • Rapid • Accurate quantification	• False positive/negatives without expertise in sample processing and data interpretation • Phenotype shifts during course of disease means multiple sets of molecular markers are required

1.5. Aims and hypotheses

Mapping of BCR and TCR repertoires promises to transform our understanding of adaptive immune dynamics, with applications ranging from identifying novel antibodies and determining evolutionary pathways for haematological malignancies to monitoring of minimal residual disease following chemotherapy (Weinstein et al., 2009, Woof and Burton, 2004b, Tonegawa, 1983). The aim of this thesis is to investigate B-cell diversity in health and disease as follows:

1. Investigating and developing robust methods for analysing high-throughput B-cell receptor sequencing repertoires.
2. Determining robust methods for deep-sequencing B-cell receptor populations.
3. Investigating B-cell repertoire dynamics following treatment of acute lymphoblastic leukaemia and investigating MRD and relapse.