Chapter 1
Background

1.1 Antibodies and Antibody Repertoire Development

The last 30 years of the biotechnology revolution have to a great extent been fueled by the discovery and application of monoclonal antibodies for research and therapeutic purposes [1, 2]. Antibodies are Y-shaped molecules produced by the immune system of many vertebrates to recognize and neutralize foreign pathogens or toxins with very high specificity. The highly specific nature of antibody-antigen interactions provides a means of molecular targeting that is extremely useful both as scientific reagents and as clinical therapeutics. An antibody molecule is comprised of two major portions: the constant region (or Fc region), which does not vary among distinct antibody molecules with different specificities, and the variable region, which comprises a unique sequence for each antibody and is the region responsible for antigen recognition (Fig. 1.1). Binding to molecular targets occurs at the exposed ends of the variable region. Different antibody variable regions confer distinct antibody binding specificities conferred by their unique antibody amino acid sequences, which are in turn derived from somatic recombination of the immunoglobulin genes and subsequent clonal selection that occurs during B-cell development.

The antibody variable region can be further subdivided into framework regions and complementarity-determining regions (Fig. 1.1). The relatively conserved framework regions (FRs) consist of antiparallel β strands which form a β-sandwich structure called the immunoglobulin fold [3]. Within the antibody variable region are distinct areas of increased variability which are termed the complementarity-determining regions (CDRs), or hypervariable loops. The CDRs contain much higher variation across different antibodies than the more conserved FRs. The precise area of the variable region where antibody binding occurs is called the paratope, whereas the binding region on an antibody's molecular target is termed the epitope. CDRs often comprise the antibody paratope.

B. DeKosky, *Decoding the Antibody Repertoire*,
Springer Theses, DOI 10.1007/978-3-319-58518-5_1

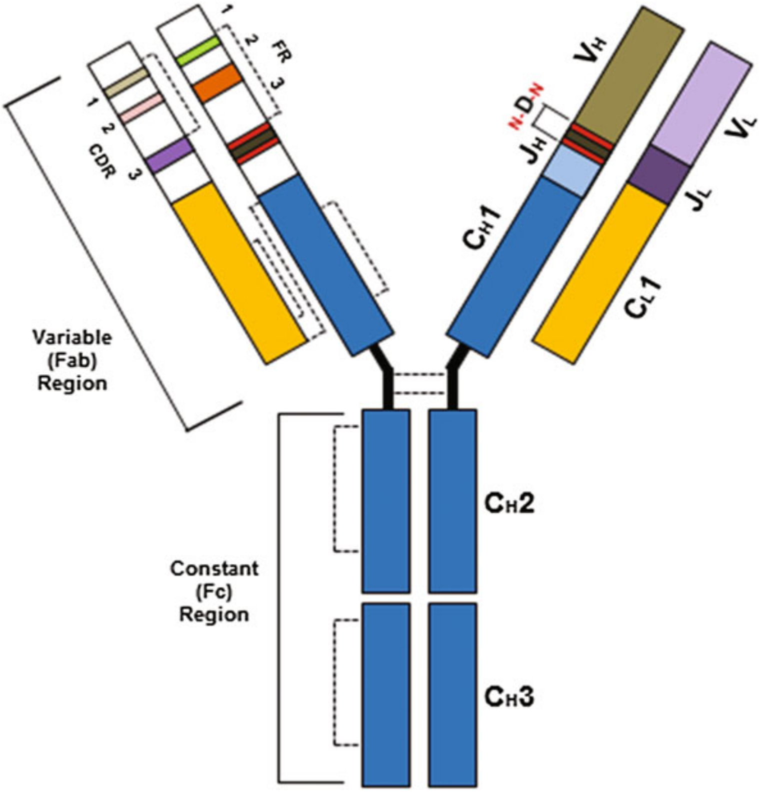

Fig. 1.1 An overview of antibody structure. Antibodies are Y-shaped molecules which consist of a variable region that imparts binding specificity and a constant region that determines the functionality of the antibody molecule. The variable region is subdivided into framework regions (FRs) which are somewhat conserved across antibodies, as well as complimentarity-determining regions (CDRs) which are highly variable across different antibodies; dashed lines indicate the location of disulfide bonds. The variable and constant regions are comprised of both heavy chain and light chain genes. The heavy chain variable region is comprised of the recombined V_H-D_H-J_H genes, whereas the light chain variable region is composed of recombined V_L-J_L genes

In addition to specificity conferred by the antibody variable region, antibody functionality is further determined by the characteristics of the antibody constant region (Fig. 1.1), termed the antibody class or isotype. Five major classes of antibody constant regions exist (IgD, IgM, IgG, IgA, and IgE), and several of these contain subclasses (e.g. IgG1, IgG2, IgA1, IgA2, etc.) Each antibody class and sub-class possesses a unique functional profile for conferring protection against pathogens and toxins. For example, IgM is expressed early in B-cell development, is secreted as a pentamer, and excels at activating immune complement (or the killing of pathogenic cells via activation of innate immune response mechanisms), whereas human IgG is expressed later in B-cell development, is secreted as a

Table 1.1 Overview of antibody constant regions, adapted from [4]

Isotype	Secreted form	Molecular weight (kDa)	Serum conc. (mean adult mg/mL)	Serum half-life (days)
IgD	Monomer	188	0.03	3
IgM	Pentamer	970	1.5	10
IgG1	Monomer	146	9	21
IgG2	Monomer	146	3	20
IgG3	Monomer	165	1	7
IgG4	Monomer	146	0.5	21
IgA1	Dimer	160	3.0	6
IgA2	Dimer	160	0.5	6
IgE	Monomer	188	5.E-05	2

monomer, and is a less potent activator of complement. An overview of antibody isotypes is presented in Table 1.1.

Membrane-bound B-cell receptors (BCR), the form of antibodies that is expressed on the surface of B-cells, are assembled via somatic recombination of antibody V-, (D-,) and J-genes within developing B-cells in the bone marrow (Figs. 1.1 and 1.2) [2, 4]. Antibody variable region gene recombination is mediated by the recombination activating genes (RAG-1 and RAG-2). The RAG enzymes recognize nucleotide patterns in the genome known as recombination signal

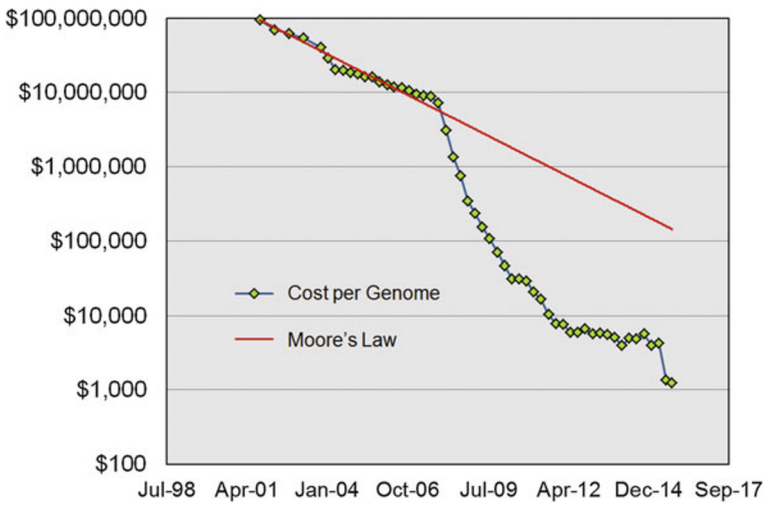

Fig. 1.2 High-throughput DNA sequencing costs have dropped dramatically in recent years beginning with the introduction of the first next-generation sequencing platforms in 2008 [32]. Log-scale reductions in DNA sequencing costs have directly enabled high-throughput analysis of individual antibody repertoires. New high-throughput DNA sequencing methods are fundamental drivers of recent advances in antibody repertoire sequencing and antibody discovery

sequences (RSS) [4]. Expression of RAG genes is highest during the pro-B and pre-B stages when V-(D)-J recombination occurs. RAG recombination can also generate junctional diversity at single-stranded DNA break sites via p-nucleotide addition and junctional trimming. Another major enzyme in B-cell development is terminal deoxynucelotidyl transferase (TdT), which is a specialized polymerase expressed predominantly during heavy chain rearrangement at the pro-B stage, and some expression can also occur in pre-B cells. TdT adds non-templated nucleotides to D-J and V-J junctions, thereby dramatically increasing the junctional diversity formed by recombination. The higher expression of TdT during heavy chain formation leads to variation in the length of heavy chain CDR3 loops (which comprise the V_H-D_H-J_H genes) because TdT adds a random number of nucleotides in each junction (typically, 0–30 nt per junction). Lower expression of TdT in pre-B cells and the lack of a D-gene causes light chain CDR3 loop lengths to be shorter and more narrowly distributed than heavy chain CDR3 loops [4].

The B-cell receptor heavy chain variable region is comprised of three genes: a V_H gene, D_H gene, and a J_H gene, and the site of heavy chain recombination is fully contained by the CDR-H3 (Fig. 1.1). In contrast, the light chain comprises a V_L gene and a J_L gene, and the V_L-J_L junction is contained within the CDR-L3 (Fig. 1.1). After a stem cell differentiates into a pro-B cell, the pro-B cell rearranges first the heavy chain D_H-J_H junction, then recombines the D_H-J_H product with a V_H gene to form the V_H-D_H-J_H region, with all these recombination events mediated by RAG expression. Recombination is stopped when the newly formed heavy chain and a paired surrogate light chain are both expressed on the cell surface. (The surrogate light chain covers the newly-generated heavy chain variable region to inhibit aberrant B-cell receptor binding.) This important step ensures that a productive, in-frame heavy chain junction without stop codons has been generated; RAG and TdT have no innate mechanism to maintain amino acid reading frames and thus only one in every three V-(D-)J recombinations will yield a productive sequence. After surface expression of the heavy chain, TdT expression decreases sharply and the B-cell is now denoted as a pre-B cell. At this stage, pre-B cells with newly formed heavy chains undergo pre-B expansion that consists of up to eight cell divisions that create multiple B-cell clones with the same heavy chain [5], which has been observed directly in mice and inferred in humans. Next, light chain V_L-J_L recombination occurs in expanded pre-B cells in a similar manner as heavy chains recombined using the RAG enzyme, but with very low levels of TdT expression that leads to generally restricted light chain CDR-L3 lengths [4]. As in pro-B cells, the final developmental checkpoint after the pre-B stage consists of B-cell receptor surface display, this time with the somatically generated heavy and light chains paired together. RAG expression and light chain recombination ceases, and the cell is now termed an immature B-cell.

Next, immature B-cells undergo an immune tolerance checkpoint to ensure that autoreactive antibodies do not advance in the antibody maturation pathway, as the process of random gene recombination without functional selection can generate self-reactive antibodies. Tolerance checkpoints are thought to occur via clonal deletion of B-cells encoding BCRs that bind to self-proteins, or alternatively, receptor editing can rescue self-reactive B-cells at this stage [6]. A known hallmark of potential autoreactivity is the expression of a very long heavy chain CDR-H3 loop, and BCRs with long CDR-H3 s are preferentially deleted from the repertoire during the pre-B tolerance checkpoint [7]. After passing tolerance checkpoints, B-cells express both IgD and IgM constant regions as B-cell receptors on the cell surface and are then considered mature, or naïve, B-cells. The newly generated mature B-cells are ready for positive selection in germinal center reactions of the spleen and lymph nodes over the course of an immune response.

1.2 Adaptive Immune Responses Lead to B-Cell Activation and Antibody Secretion

Major reactions of the adaptive immune response occur in the the spleen and tissue-draining lymph nodes (collectively called the secondary lymphoid organs) [4]. There, specialized antigen-presenting cells (APCs) named dendritic cells and follicular dendritic cells present captured foreign proteins and present antigens for B- and T-cell surveillance. In the process of dendritic cell antigen presentation, peptides from digested foreign proteins are presented on the cell surface in the major histocompatibility complex (MHC), a specialized cell surface protein for display of peptides to T cells. T cells survey MHC using their somatically generated T-cell receptors, or TCR. TCR are formed in a highly analogous process to BCR, with the notable difference that T cells do not normally express AID and therefore TCR do not exhibit somatic hypermutation or class-switch recombination [4, 8]. If a T cell's TCR binds tightly to a peptide presented on MHC by APCs during the course of infection, the T cell will be given signals to expand and proliferate. Known activating signals include a combination of the secreted cytokines like IL-6, IL-12, and TGF-β, and especially the B7 co-receptors which bind to CD28 (activating) and CTLA-4 (inhibiting) receptors on the T-cell surface. If a naïve T cell recognizes its peptide antigen in the absence of professional antigen-presenting cell co-receptor signals, the T cell will undergo functional inactivation (also known as anergy) or clonal deletion. The requirement for dedicated APCs to mediate T-cell activation is a major immune control mechanism that minimizes autoimmunogenic T-cell proliferation and helps to ensure that T-cell activation occurs only as a result of the response to true infections.

Effector T cells comprise two major classes:

1. $CD8^+$ cytotoxic T cells: bind MHC class I proteins expressed on all cells in the body and and survey cytosol (endoplasmic reticulum-derived) proteins.
2. $CD4^+$ helper T cells: bind MHC class II proteins that are expressed on dedicated antigen-presenting cells (e.g. macrophages and B-cells) and survey vesicular (extracellular-derived) proteins.

$CD8^+$ effector T cells (or cytotoxic T lymphocytes, CTLs) are cytotoxic cells that clonally expand and survey cells in the body for signs of infection. When CD8 T cells identify a cell expressing its targeted foreign peptide-MHC (e.g. a viral peptide from a virally infected cell) the CTL kills the target cell or induces the target cell to undergo programmed cell death and thereby help eliminate the infection. As MHC class I is expressed by all cells in the body and constantly presents peptides onto the cell surface derived from inside the cell, CTLs circulate throughout the body and can survey ongoing protein expression in most cells and tissues, serving as a major adaptive immune effector mechanism against intracellular pathogens. Alternatively, $CD4^+$ T cells are activated by APCs to become helper T cells, and their major role occurs in the secondary lymphoid organs. $CD4^+$ T cells perform a critical role in enabling the antibody (or extraceullar) immune response. The five major CD4 effector T subclasses (T_{FH}, T_H1, T_H2, T_H17, and T_{reg}) are summarized in Table 1.2.

Naïve B-cells are activated by both antigen-presenting cells and $CD4^+$ helper T cells (specifically, T_{FH} cells, see Table 1.2). B-cell selection occurs in germinal center (GC) reactions of the spleen and lymph node [4, 9]. Within germinal centers, B-cells that encode BCR will bind to antigen presented on follicular dendritic cells or antigen presented through other pathways. Following receptor endocytosis of the BCR in a bound complex to antigen, the B-cells digest the captured and display the resulting peptides via MHC class II on the B-cell surface. $CD4^+$ follicular helper T (T_{FH}) cells that have been activated by other APCs will in turn induce those antigen-specific B-cells to activate via TCR-peptide-MHC contacts, along with other with interactions between the B-cell co-receptor CD40 and the T cell-derived CD40 ligand (CD40L, or CD154). $CD4^+$ helper T-cells also secrete cytokines that stimulate B-cell proliferation and differentiation such as IL-4, IL-5, and IL-6.

When naïve B-cells receive the proper signals from antigen-specific T cells they also are activated and begin to undergo somatic hypermutation mediated by the enzyme activation-induced cytidine deaminase (AID) that is linked to successive rounds of cell division in the dark zone of the GC reaction [10]. Further rounds of positive selection in the light zone successively enhance antibody affinity to the

Table 1.2 Overview of $CD4^+$ T cell effector subsets. Each subset is induced to differentiate from naïve T cells via a unique third signal provided by antigen presenting cells (Signal 1 comprises the TCR-peptide-MHC interaction, and Signal 2 consists mainly of CD28-B7 interactions) [4]

CD4 effector subset	APC 3rd signal	Subset function
T_{FH}	IL-6	B-cell activation
T_H1	IL-12 + IFN-γ	Intracellular bacterial response
T_H2	IL-4	Parasitic response (eosinophils, mast cells, IgE B-cell activation)
T_H17	TGF-β + IL-6	Extracellular bacteria/fungi, induce epithelial/stromal cells to secrete cytokines for neutrophil recruitment
T_{reg}	TGF-β	Suppress T-cell activity, prevent autoimmunity

antigen of interest [11]. Activated B-cells can undergo class-switch recombination, also mediated by AID, which alters the antibody isotype class via genomic deletion of IgM/IgD constant regions that switches the isotype to IgG, IgA, or IgE. Positive selection in the GC light zone induces B-cells to asymetrically proliferate into antigen-secreting plasmablasts, antigen-secreting plasma cells (which home to bone marrow for long-term persistence), and memory B-cells, which comprise the effector cells of the antibody response [12, 13].

Importantly, other pathways can lead to B-cell activation beyond germinal center reactions. B-cell maturation has been reported at the border of the splenic T cell zone and red pulp [14], and for some antigens (termed T-independent antigens) B-cell activation occurs in the absence of T-cell help. T-independent antigens often have a component that triggers a receptor of the innate immune system on the B-cell surface, or alternatively can extensively cross-link IgM BCR molecules on the B-cell surface [4]. While some antigens are able to induce robust antibody responses in the absence of T-cell help, the vast majority of proteins will not activate the B-cell response in the absence of CD4$^+$ T-cell assistance.

The three B effector cell subsets (plasmablasts, plasma cells, and memory B-cells) each perform a distinct role in fighting disease. Plasmablasts are short-term mediators of serological immunity; they secrete high levels of antibody and circulate in peripheral blood for a relatively limited amount of time (days) before undergoing apoptosis, resulting in a rapid increase in serum antibody concentrations that dissipates in accordance with antibody half-life in serum (approximately two weeks) [15, 16]. Similar to plasmablasts, plasma cells (PCs) also secrete antibody but are long-lived and home to bone marrow where they can persist for many years [17–23] and continuously secrete high levels of immunoglobulin (estimated at 10,000–20,000 molecules/cell-sec) [2]. Long-term serological memory is mediated by plasma cell populations residing in the bone marrow. Like plasma cells, memory B-cells are also long-lived but do not secrete antibody. Instead, memory B-cells express immunoglobulin as a B-cell receptor on their cell surface and circulate in peripheral blood (and pass through secondary lymphoid organs) until encountering its cognate antigen again. Upon re-encounter with their specific antigen, memory B-cells rapidly differentiate into plasmablasts and/or plasma cells to enable long-term immune memory. Importantly, the kinetics of memory B-cell activation (or secondary responses) are much faster than the initial (or primary) responses to a particular antigen. A comparison of various B-cell subset characteristics with relevance to antibody repertoire sequence analysis is provided in Table 1.3.

Table 1.3 Selected characteristics of major B-cell subsets relevant to antibody repertoire sequence analysis. B-cells are negatively selected in central and peripheral tolerance, and positively selected for antibody affinity to antigen in GC reactions

Repertoire subset	Selection mechanisms	Isotypes	SHM?
Mature (naïve) B-cell	Negative	IgM/IgD	No
Memory B-cell	Positive and Negative	IgM, IgG, IgA, IgE	Often
Plasmablast/plasma cell	Positive and Negative	IgM, IgG, IgA, IgE	Often

Germinal center B-cell activation results in multiple variants of high-affinity antibodies specific to a target antigen. The resulting effector B-cell clones can persist for a very long time (e.g. nearly 90 years in the case of memory B-cells [24]), and the entire collection of antibodies encoded by an individual resulting from a lifetime of immune responses comprises that individual's antibody repertoire. Given the large number of unique B-cell clones in humans (likely exceeding 2×10^6 unique antibodies in peripheral blood alone [25, 26]), a thorough and comprehensive analysis of the antibody repertoire requires high-throughput means of sequence data collection and analysis. Beginning in 2008, the rapidly decreasing costs of gene sequencing for the first time permitted economic repertoire-scale, high-resolution DNA sequence analysis of B-cell populations (Fig. 1.2). These new experimental techniques, collectively known as high-throughput sequencing or next-generation DNA sequencing, are fundamentally transforming the major analytical methods for B- and T-cells as well as our understanding of adaptive immune responses.

1.3 High Throughput Antibody Sequencing

Tremendous advances in scale and decreases in costs of next-generation DNA sequencing technologies have dramatically accelerated the pace of biological research over the last eight years (Fig. 1.2). High-throughput DNA sequencing has been a transformative method for studying adaptive immunity by permitting repertoire-scale analysis of the vast number of unique BCRs and TCRs in the adaptive immune system [2, 27]. Exact measurement of total human B-cell repertoire size is difficult to determine due to tissue sampling limitations (peripheral blood, bone marrow, and secondary lymphoid organs), combined with high-throughput sequencing error (typically ~0.5% of sequenced bases [28, 29]) that contributes noise to secondary data analysis. However, lower-bound estimates of repertoire size are approximately 2 × 10 [6] unique B-cell clones (expressing distinct BCRs) in peripheral blood alone [25, 26]. Upper bounds on repertoire size are difficult to estimate but are several orders of magnitude higher, with theoretical B-cell receptor diversity exceeding 10^{13} and individual limitations at around 10^{11} B-cells in the human body [2, 4, 30, 31].

Standard immune repertoire high-throughput sequencing protocols begin with collection of 10^3–10^7 lymphocytes, followed by bulk cell lysis and recovery of cellular mRNA. Next, mRNA is reverse transcribed and a PCR multiplex primer set which targets all known V-genes is used for PCR amplification of antibody or TCR genes. In the case of antibody analysis, the 5' PCR primers usually target V genes, while the 3' primers target either the J genes or constant regions (e.g. IgG, IgA, etc.) to perform sequence analysis of the entire antibody variable region with minimal coverage of the constant region (Fig. 1.1) [2]. (Some protocols omit V-gene-specific primers and incorporate RACE PCR to reduce multiplex PCR amplification bias [33]). For antibody analysis, the heavy chain, kappa light chain, and lambda light chains are each amplified in separate PCR reactions. Finally, high-throughput sequencing and bioinformatic analyses are performed to quantitatively determine the composition of the input immune repertoire encoded by the cells originally isolated from experimental samples [2, 29, 34, 35]. High-throughput repertoire sequencing has been applied in a variety of applications ranging from characterization of the repertoire in healthy and disease states [36–39], to analysis of antibody-pathogen interactions [40–42], and for rapid antibody discovery [41, 43].

Despite the tremendous recent advances, all currently available techniques for antibody repertoire analysis have one severe limitation: high-throughput antibody sequencing is unable to resolve the pairing between antibody heavy and light chains. Using the high-throughput sequencing methods described above, B-cell populations are lysed in bulk to collect mRNA for downstream sequence analysis. Recombined heavy and light V-(D-)J junctions are located on separate chromosomes and expressed as distinct mRNA strands, and the bulk B-cell lysis required for high-throughput sequencing confounds the pairing between heavy and light mRNAs expressed by individual B-cells. Next-generation sequencing techniques can sequence only one mRNA strand at a time, which further complicates efforts to preserve heavy and light chain pairing information [2, 40, 44, 45]. Without the ability to sequence paired heavy and light chain sequences at high throughput with single-cell resolution, the full antibody clonotype (both heavy and light chains) cannot be resolved on a repertoire scale, nor can the resulting antibody sequences be expressed to test for function, nor can the antibody proteins be modeled computationally.

1.4 Next Generation Antibody Sequencing Data Analysis

The rapid growth of antibody sequencing data has also fueled rapid growth in our capabilities to analyze and interpret antibody sequence datasets. In particular, the errors introduced by Next Generation sequencing technologies greatly complicate efforts to analyze B-cell sequence data because B-cells are known undergo somatic hypermutation, creating two B-cell clones that differ by only a single nucleotide substitution in the variable region. Given that the errors (or noise) introduced by NextGen sequencing are greater than the single-nucleotide changes that can be generated via SHM, it can be very difficult to confidently assess whether two closely related antibody sequences derive from sequence error or from two true somatic variants of the same clonal lineage.

NextGen sequencing technologies exhibit an average error rate of approximately 0.5% [28, 46, 47], which consist of a mix of single base pair substitution errors and insertion/deletion errors. Different NextGen platforms exhibit distinct error rate patterns. For example, the widely-used Roche 454 sequencing platform, which is based on real-time observation of base incorporation and with pauses in between the addition of A/T/C/G bases, has difficulty determining the length of homopolymer stretches of DNA because homopolymer runs incorporate in such rapid succession that it is difficult for the camera and software to determine the exact number of bases incorporated. Thus, 454 (along with most other real time sequencing platforms such as those sold by Pacific Biosciences) has a high rate of homopolymer insertions and deletions (indels). Indels are often trivial to correct manually, but can prove especially problematic for high-throughput analyses because they introduce frameshift errors to the amino acid translations. 454 also maintains a consistent error rate across the course of a sequence read. In contrast, the Illumina sequencing platform relies on reversible terminator chemistry that permits the observation of each base incorporation event, similar to a reversible Sanger sequencing technology. For Illumina data the probability of insertion and deletion events is very low compared to 454 data, whereas the single base substitution error rate is comparably higher. In addition, due to the accumulation of imperfections such as incomplete removal of the reversible terminator across cycles, the read quality of Illumina sequences degrades as the sequence read progresses. For example in the MiSeq platform, the first ~150 bases of the read are of comparably high quality, whereas quality rapidly degrades in the final 100 bases of a 250 bp-length read. Thus, Illumina sequencing technologies require error correction that takes into account not only the overall average error rates of Illumina sequences, but also the characteristics of error introduction over the course of different sections in the sequence read.

High-confidence methods do exist for deconvoluting sequence error using molecular barcodes incorporated into reverse transcription primers [48–50]. These methods utilize a unique barcode region that is incorporated during first-strand cDNA synthesis and that is maintained throughout the following rounds of PCR. During sequence data analysis, the barcodes resulting from each RT event can be pooled, counted, and a consensus sequence, or average of all the different reads, can be generated for each barcoded RT event with a sufficient number of observations. This approach can identify the original 1st-strand cDNA sequences with very high accuracy, given sufficient coverage of the individual barcodes (minimum of 3 reads per barcode for establishing consensus.) However, in practice these methods have a limited throughput because the high number of RT events and skewing of repertoire distributions via PCR results in a low number of sequences that pass the minimum observation threshold for establishing a consensus [48–50]. Additionally, these techniques cannot deconvolute errors introduced by reverse transcription, which is an important source of sequence errors in immune repertoire sequence data.

Each antibody sequencing platform reports the estimated confidence of a particular base being correct or incorrect in the form of a quality score, similar to the Phred scores reported in DNA sequencing data via capillary electrophoresis. An essential first step in any NextGen data analysis pipeline is to filter sequence data by quality scores, which removes many of the most error-prone sequence reads in a NextGen dataset. However, it is critical to remember that sequence quality filtering is inadequate for sequence error mitigation on its own for several reasons. First, the quality score is a probabilistic measurement, which measures only the likelihood that a given base is right or wrong as calculated based on raw data collected in the sequencing platform, often using signal:noise ratios or other metrics for each sequenced base. Filtering out the bases with the highest probability of containing errors is not the same as removing all errors from the data. Second, quality scores report only the likelihood of an error introduced by the *sequencing* process. Another major source of error introduction occurs as a result of mismatched base incorporation during reverse transcription, which cannot be identified via quality score sequence analysis as the error was introduced prior to DNA sequencing.

A very powerful tool for identifying errors is comparative frequency analysis. NextGen sequencing identifies millions of DNA sequence reads, and because the per-base error rates are low and (to a first approximation) random, the probability of a correct base sequenced is high. Thus, by investigating a number of highly similar cDNA sequences, it becomes clear that the bases observed most frequently are the highest confidence for corresponding to the true initial RNA sequence. Minimum observation cutoffs are very helpful in this regard—if a sequence has been observed only once, then it is much more likely that that particular sequence contains a sequence error than a sequence observed many times. A related technique is the establishment of consensus sequences. In this process, multiple sequence reads are aligned and "averaged" such that a final consensus sequence is constructed from a minimum of 3 NextGen sequence reads. Both minimum observation cutoffs and consensus sequence generation are extremely helpful tools to minimize the effects of sequence error in NextGen antibody sequencing datasets.

Another critical tool for Nextgen antibody sequence analysis is clustering, or the grouping of similar antibody variants by sequence similarity. Several clustering techniques and platforms have been reported [51–53]. Clustering groups highly similar sequences based on defined cutoffs and other user-defined parameters, and because sequence errors are introduced at low-levels, clustering can accurately identify antibody clones, clonotypes, and consensus sequences. However, one caveat to clustering is that it removes most low-level somatic hypermutation of a given antibody lineage, which is an important feature of any clustering analysis. Because antibody sequences derive from common genes and are highly similar, it is recommended in almost all cases to perform clustering of a high-variation subset of the full sequence (e.g., the CDR3) rather than clustering complete antibody sequences. This concentrated clustering analysis uses the highest variation region of the antibody sequence to avoid collapsing genetically similar clones with a distinct origin (i.e., different CDR3 regions) together into the same cluster.

Many different iterations of the above analysis and error correction techniques can lead to robust antibody sequence data analysis. Different techniques may be required for distinct experimental sample preparation protocols, sequencing platforms, and intended applications of the experiment. An important step in the evaluation of any bioinformatic pipeline is to evaluate performance using spike-in controls of a known sequence. These spike-in controls can use immortalized B-cell lines that express a known antibody sequence, for example the human clones IM-9 and ARH-77, and the mouse clones MOPC-315 and MOPC-21, and sometimes spike-in RNA or cDNA of a known sequence. It is important to verify that a single spike-in sequence can be collapsed to a single antibody sequence by any experimental and bioinformatic pipeline as a test of that pipeline's ablity to identify *bona fide* antibody sequences.

For our work here and verified using spike-in controls from immortalized B-cell lines, we have found that a pipeline of quality filtering, CDR3 extraction and analysis (as the CDR3 has the highest per base variation across antibody clones), compilation by CDR3 sequence and V(D)J gene assignments, removal of single-read paired CDR-H3:CDR-L3 variants, and CDR-H3 clustering leads to a robust repertoire where most somatic variants are removed for antibody clones and where immortalized cell spike-in controls collapsed to a single sequence variant. While no methods can achieve perfection given the limitations of high NextGen error rate profiles and the unique characteristics of B-cells to undergo somatic hypermutation, the methods reported here present a highly useful compromise that allows for the annotation of antibody clusters (or lineages/clonotypes) contained in the data and provides a workable platform for robust antibody repertoire analyses.

1.5 Monoclonal Antibody Discovery Technologies

The utility of serum antibodies for treating disease was first established in the late 19th century through the work of Emil von Behring, Kitasato Shibasaburo, and Emile Roux in developing serum therapies to diphtheria toxins. These early methods were based on polyclonal antibodies, or a mixture of all the different antibody specificities contained in human or animal sera. Research continued with polyclonal antibody mixtures until the 1970's when Georges Köhler and César Milstein published a method to generate hybridomas, or a B-cell fused with an immortalized myeloma cell that allowed the resulting cell hybrid (called a hybridoma) to secrete antibody continuously in culture [1]. The discovery of hybridomas ushered in a new era of biotechnology as monoclonal antibodies (mAbs) against a wide variety of antigens could be isolated from mice following challenge with the antigen of interest.

In the hybridoma process, B-cells and myeloma cells are fused as described above, then cells are divided into individual wells by limiting dilution and cultured as they produce and secrete antibody. Culture supernatant from each well containing secreted antibody is screened for binding to antigen via enzyme-linked immunosorbent assay (ELISA), and cells from any positive-binding wells can be retrieved, further expanded, cloned, and sequenced, while the monoclonal antibody itself can be purified from hybridoma culture supernatant. Hybridoma methods continue to have tremendous impact. The basic techniques were outlined around 40 years ago and are still relevant today, but hybridoma techniques have improved such that an antibody can be developed toward a particular target much more rapidly and with high reliability. In particular, recent key methods include enhancing the efficiency of hybridoma generation with human cells [54] and humanization of mouse antibodies or the development of human transgenic [55–58] or humanized [59–61] mouse models to reduce immunogenicity of the resulting monoclonal antibody therapeutics in human patients. Several protocols have permitted faster and more economical hybridoma screening to accelerate the discovery of human or humanized monoclonal antibodies [62–64]. In particular, transgenic mice have recently been used for isolating human monoclonal antibodies against human proteins (the response to self-antigen is limited in humans), and resulting mAbs can be used to agonize or block human surface receptors or target expressed oncogenes for cancer therapeutics. Despite tremendous advances in tried-and-true hybridoma technologies, mAb discovery using hybridomas remains time-consuming and expensive due to the single-cell limiting dilution needed and the time required hybridomas to expand from a single cell to a cell population (several weeks). The large number of culture supernatant screens required also make hybridoma mAb discovery a resource-intensive experimental toolkit.

An important alternative technology to hybridoma mAb discovery is antibody isolation via in vitro screening of combinatorial libraries. The most widely used combinatorial library discovery platform technology uses scFv display on M13 bacterial phage, where antibody variable regions are PCR amplified from B-cell populations of interest and expressed for display on the surface of bacteriophage. Then, the phage can be selected for binding to the surface of antigen-coated plates or tubes, and bound phage are eluted from tubes and amplified via re-infection of bacteria (along with mutations acquired in each round). Finally after several rounds of phage panning a high-affinity monoclonal antibody can be isolated [65–71]. Phage panning has several key advantages including lower cost and the capability to affinity mature antibodies in vitro, however phage panning library construction requires combinatorial shuffling of heavy and light chain pairs during library generation, leading to a non-natural (synthetic) antibody library. Another major limitation to phage panning is that the resulting antibodies have not been screened by central or peripheral tolerance checkpoints in the immune system. Thus while it is a highly effective method for isolation of research and diagnostic antibodies, phage panning's inability to isolate native heavy and light chain pairings and the accrual of mutations throughout phage panning pose risks for immunogenicity and off-target binding in humans, and therefor phage panning has more limited applications for therapeutic antibody discovery [72].

A more recent method for monoclonal antibody discovery has applied high-throughput sequencing of B-cell receptors to identify antibodies of interest [40, 43, 73]. Next Generation sequencing is revolutionizing our ability to analyze and interpret antibody repertoires because it is rapid and efficient, and it also provides information on the entire repertoire of antibodies elicited in the individual. High-throughput sequencing of the cellular repertoire can also be used to construct a database for proteomic analysis of human serum antibodies via mass spectrometry [33, 74–77]. These techniques quantify the serum antibodies generated in response to vaccination and disease and have proven useful for antibody discovery by linking antibody function (i.e. binding to a particular antigen) to the antibody sequence [74, 75, 78]. High-throughput and proteomic techniques for antibody discovery will become more widely used in the coming years as DNA sequencing and protein mass spectrometry costs continue to decrease.

Unfortunately the heavy and light chain pairing information is irreversibly lost during conventional high-throughput antibody sequencing [40, 44, 45], and this inability to deconvolute paired heavy and light sequences using NextGen sequencing has severely complicated efforts to rapidly discover new antibodies and applications of high-throughput sequencing for antibody discovery. Some progress has been made toward more rapid antibody discovery using NextGen sequencing. Important early work demonstrated that frequency-based pairing of highly enriched cell populations can lead to productive antibodies [43]. Additionally, phylogenetic algorithms can be effective for inferring the heavy:light pairing of highly mutated antibody lineages in some cases [42]. New antibody variants identified by heavy chain-only antibody sequencing can also provide important insights regarding antibody lineage development [40, 41, 73, 79, 80]. In addition, such sequences can be used in combination with single-cell RT-PCR data to generate new antibody variants and test antibody performance in vitro [45].

High-throughput approaches could be optimal for antibody discovery and immune repertoire analysis if a new technology were available to gather single-cell heavy and light chain pairing information at high-throughput. The state of current (low-throughput) sequence-based alternatives to high-throughput sequencing, collectively known as single-cell RT-PCR, and applications of single-cell sequencing to antibody discovery are discussed in the following section.

1.6 Single-Cell Sequencing Techniques

As mentioned above, existing immune repertoire high throughput sequenc-ing technologies yield data on only one of the two chains of immune receptors and cannot provide information about the identity of immune receptor pairs encoded by individual B or T lymphocytes [40, 44, 45]. Due to this major limitation, lower-throughput single-cell techniques must be used when paired heavy and light chain information is required. Several experimental techniques have been employed for detection or sequencing of genomic DNA or cDNA from single cells; however these techniques are limited by low efficiency or low cell throughput (<200–500 cells) and further, they require fabrication and operation of complicated microfluidic devices [81–85]. As a result of these limitations, sequence analysis of VH: VL pairs is currently performed by microtiter-well sorting of individual B-cells followed by single-cell RT-PCR (scRT-PCR) and Sanger sequencing [7, 45, 86–88]. Once the sequence of a B-cell has been isolated it can then be cloned into bacteria and tested for antigen binding to a protein of interest [45, 86, 87, 89], or alternatively each B-cell can be induced to secrete antibody in vitro prior to screening single-cell culture supernatant by microneutralization [79]. A significant time savings can be achieved via linkage of heavy and light chains in the RT-PCR, thereby reducing cloning steps by a factor of two [86, 90, 91].

Single-cell sequencing in the small volumes required for manageable high-throughput experiments ($\sim 10^2$ pL per reaction) is severely limited by inhibition of the RT-PCR reaction by cell lysate, which poses a lower bound on microwell or droplet volume at around 5 nL/cell for one-pot cell encapsulation and RT-PCR [83]. Incomplete cell lysis and RNA degradation during thermal cell lysis can further reduce yield of linked cDNA products using one-pot reactions. Furthermore, the cell lysis and mRNA recovery steps are non-trivial to perform at high-throughput and with single-cell fidelity. These experimental complications have made high-throughput sequencing of multiple mRNA transcripts from single cells a critical unsolved problem. Potential solutions for sequencing of multiple mRNA strands derived from single cells would have important applications for in-depth analysis of antibody and TCR repertoires as well as provide a tremendous boost to currently available antibody discovery workflows [2, 40, 44, 45].

1.7 Synopsis

This dissertation directly addresses the aforementioned limitations of currently available high-throughput methods in resolving heavy and light chain pairings via the development and application of new high-throughput, single-cell sequencing technologies. In Chapter Three (*High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire*) [92], we report a new method for sequencing B-cell heavy and light chains at single-cell resolution by capturing cells in micropatterned PDMS wells and lysing cells individually, with the capacity to analyze up to 5×10^4 B-cells in a single experiment. After validating our technique, we analyzed human antibody responses in healthy and vaccinated donors and applied our high-throughput paired heavy and light chain sequencing platform for human antibody discovery.

Next, Chapter Four (*In-depth determination and analysis of the human paired heavy and light chain antibody repertoire*) [93] relates a modification of the single-cell sequencing methods described in Chapter Three that provided greatly enhanced cell throughput. We constructed and validated a new flow-focusing nozzle system for single B-cell isolation and analysis with the capacity to emulsify up to 3×10^6 B-cells per hour. After demonstrating that our technique was >97% accurate by analyzing technical replicates of expanded B-cell populations, we characterized several previously unreported features of human antibody repertoires including a quantitative analysis of promiscuous and public light chain junctions (i.e. light chains expressed by multiple B-cell clones both within and across human donors) and the high-throughput detection and sequence characterization of allelically included human B-cells.

The fifth chapter of this report (*Paired VH:VL analysis of naïve B-cell repertoires and comparison to antigen-experienced B-cell repertoires in healthy human donors*) applied the new high-throughput techniques developed in Chapters Three and Four toward a comprehensive, high-resolution analysis of naïve and antigen-experienced B-cells in the same individuals. Comprising the first high-throughput analysis of heavy and light chains to compare multiple B-cell compartments, our analysis of gene usage and antibody biochemical composition generated new insights regarding the antibody development, maturation, and selection processes across several human donors.

References

1. Kohler G, Milstein C (1975) Continuous cultures of fused cells secreting antibody of predefined specificity. Nat 256:495–497
2. Georgiou G et al (2014) The promise and challenge of high-throughput sequencing of the antibody repertoire. Nat Biotechnol 32:158–168
3. Berg JM et al (2002) Biochemistry. Freeman, WH
4. Murphy K, Travers P, Walport M, Janeway C (2012) Janeway's immunobiology. Garland Science, US
5. Hess J et al (2001) Induction of pre-B cell proliferation after de novo synthesis of the pre-B cell receptor. Proc Natl Acad Sci 98:1745–1750
6. Lee J, Monson NL, Lipsky PE (2000) The VλJλ repertoire in human fetal spleen: evidence for positive selection and extensive receptor editing. J Immunol 165:6322–6333
7. Wardemann H et al (2003) Predominant autoantibody production by early human B-cell precursors. Sci 301:1374–1377
8. Qi Q et al (2014) Diversity and clonal selection in the human T-cell repertoire. Proc Natl Acad Sci 111:13139–13144
9. Kuppers R, Zhao M, Hansmann ML, Rajewsky K (1993) Tracing B-cell development in human germinal centers by molecular analysis of single cells picked from histological sections. EMBO J 12:4955–4967
10. Fernández D et al (2013) The proto-oncogene c-myc regulates antibody secretion and ig class switch recombination. J Immunol 190:6135–6144
11. Gitlin AD, Shulman Z, Nussenzweig MC (2014) Clonal selection in the germinal centre by regulated proliferation and hypermutation. Nat 509:637–640
12. Barnett BE et al (2012) Asymmetric B-cell division in the germinal center reaction. Sci 335:342–344
13. Klein U, Dalla-Favera R (2008) Germinal centres: role in B-cell physiology and malignancy. Nat Rev Immunol 8:22–33
14. William J, Euler C, Christensen S, Shlomchik MJ (2002) Evolution of autoantibody responses via somatic hypermutation outside of germinal centers. Sci 297:2066–2070
15. Hinton PR et al (2006) An engineered human IgG1 antibody with longer serum half-life. J Immunol 176:346–356
16. Kyu SY et al (2009) Frequencies of human influenza-specific antibody secreting cells or plasmablasts post vaccination from fresh and frozen peripheral blood mononuclear cells. J Immunol Methods 340:42–47
17. Manz RA, Löhning M, Cassese G, Thiel A, Radbruch A (1998) Survival of long-lived plasma cells is independent of antigen. Int Immunol 10:1703–1711
18. O'Connor BP, Cascalho M, Noelle RJ (2002) Short-lived and long-lived bone marrow plasma cells are derived from a novel precursor population. J Exp Med 195:737–745
19. Amanna IJ, Carlson NE, Slifka MK (2007) Duration of humoral immunity to common viral and vaccine antigens. N Engl J Med 357:1903–1915
20. Amanna IJ, Slifka MK (2010) Mechanisms that determine plasma cell lifespan and the duration of humoral immunity. Immunol Rev 236:125–138
21. Dorner T, et al (2011) Long-lived autoreactive plasma cells drive persistent autoimmune inflammation. Nat Rev Rheumatol 7:170

22. Mahevas M, Michel M, Weill J.-C, Reynaud C-A (2013) Long-lived plasma cells in autoimmunity: lessons from B-Cell depleting therapy. Front Immunol 4
23. Halliley JL et al (2015) Long-lived plasma cells are contained within the CD19− CD38 hi CD138+ subset in human bone marrow. Immun 43:132–145
24. Yu X et al (2008) Neutralizing antibodies derived from the B-cells of 1918 influenza pandemic survivors. Nat 455:532–536
25. Boyd SD, et al (2009) Measurement and clinical monitoring of human lymphocyte clonality by massively parallel V-D-J pyrosequencing. Sci Transl Med 1:12ra23
26. Arnaout R et al (2011) High-resolution description of antibody heavy-chain repertoires in humans. PLoS ONE 6:e22365
27. Warren EH, Matsen FA, Chou J (2013) High-throughput sequencing of B- and T-lymphocyte antigen receptors in hematology. Blood 122:19–22
28. Loman NJ et al (2012) Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotech 30:434–439
29. Bashford-Rogers RJ et al (2014) Capturing needles in haystacks: a comparison of B-cell receptor sequencing methods. BMC Immunol 15:29
30. Schroeder HW Jr (2006) Similarity and divergence in the development and expression of the mouse and human antibody repertoires. Dev Comp Immunol 30:119–135
31. Apostoaei AI, Trabalka JR (2012) Review, synthesis, and application of information on the human lymphatic system to radiation dosimetry for chronic lymphocytic leukemia. SENES Oak Ridge, Inc., Tennessee
32. Wetterstrand KA. DNA sequencing costs: data from the NHGRI Genome Sequencing Program (GSP). Available at: www.genome.gov/sequencingcostsdata. Accessed 27 Feb 2017
33. Wine Y et al (2013) Molecular deconvolution of the monoclonal antibodies that comprise the polyclonal serum response. Proc Natl Acad Sci 110:2993–2998
34. Menzel U et al (2014) Comprehensive evaluation and optimization of amplicon library preparation methods for high-throughput antibody sequencing. PLoS ONE 9:e96727
35. Greiff V et al (2014) Quantitative assessment of the robustness of next-generation sequencing of antibody variable gene repertoires from immunized mice. BMC Immunol 15:40
36. Glanville J et al (2009) Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. Proc Natl Acad Sci USA 106:20216–20221
37. Wu Y-CB, Kipling D, Dunn-Walters DK (2012) Age-related changes in human peripheral blood IGH repertoire following vaccination. Front Immunol 3
38. Schoettler N, Ni D, Weigert M (2012) B-cell receptor light chain repertoires show signs of selection with differences between groups of healthy individuals and SLE patients. Mol Immunol 51:273–282
39. Hoi KH, Ippolito GC (2013) Intrinsic bias and public rearrangements in the human immunoglobulin V[lambda] light chain repertoire. Genes Immun 14:271–276
40. Wu X et al (2011) Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. Sci 333:1593–1602
41. Zhu J et al (2013a) De novo identification of VRC01 class HIV-1–neutralizing antibodies by next-generation sequencing of B-cell transcripts. Proc Natl Acad Sci 110:E4088–E4097
42. Zhu J et al (2013b) Mining the antibodyome for HIV-1-neutralizing antibodies with next-generation sequencing and phylogenetic pairing of heavy/light chains. Proc Natl Acad Sci 110:6470–6475
43. Reddy ST et al (2010) Monoclonal antibodies isolated without screening by analyzing the variable-gene repertoire of plasma cells. Nat Biotechnol 28:965–969
44. Fischer N (2011) Sequencing antibody repertoires: the next generation. MAbs 3:17–20
45. Wilson PC, Andrews SF (2012) Tools to therapeutically harness the human antibody response. Nat Rev Immunol 12:709–719
46. Prabakaran P, Streaker E, Chen W, Dimitrov DS (2011) 454 antibody sequencing-error characterization and correction. BMC Res. Notes 4:404

47. Schirmer M et al (2015) Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. Nucleic Acids Res gku1341. doi:10.1093/nar/gku1341
48. Vollmers C, Sit RV, Weinstein JA, Dekker CL, Quake SR (2013) Genetic measurement of memory B-cell recall using antibody repertoire sequencing. Proc Natl Acad Sci 110:13463–13468
49. Shugay M et al (2014) Towards error-free profiling of immune repertoires. Nat Methods 11:653–655
50. Khan TA et al (2016) Accurate and predictive antibody repertoire profiling by molecular amplification fingerprinting. Sci Adv 2:e1501371
51. Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinform 26:2460–2461
52. Fu L, Niu B, Zhu Z, Wu S, Li W (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinform 28:3150–3152
53. Li W, Fu L, Niu B, Wu S, Wooley J (2012) Ultrafast clustering algorithms for metagenomic sequence analysis. Brief Bioinform 13:656–668
54. Yu X, McGraw PA, House FS, Crowe JE Jr (2008) An optimized electrofusion-based protocol for generating virus-specific human monoclonal antibodies. J Immunol Methods 336:142–151
55. Lonberg N et al (1994) Antigen-specific human antibodies from mice comprising four distinct genetic modifications. Nat 368:856–859
56. Mendez MJ et al (1997) Functional transplant of megabase human immunoglobulin loci recapitulates human antibody response in mice. Nat Genet 15:146–156
57. Lonberg N (2005) Human antibodies from transgenic animals. Nat Biotechnol 23:1117–1125
58. Murphy AJ et al (2014) Mice with megabase humanization of their immunoglobulin genes generate antibodies as efficiently as normal mice. Proc Natl Acad Sci 111:5153–5158
59. McCune JM et al (1988) The SCID-hu mouse: murine model for the analysis of human hematolymphoid differentiation and function. Sci 241:1632–1639
60. Hiramatsu H et al (2003) Complete reconstitution of human lymphocytes from cord blood CD34+ cells using the NOD/SCID/γcnull mice model. Blood 102:873–880
61. Ippolito GC et al (2012) Antibody repertoires in humanized NOD-scid-IL2R gamma(null) mice and human B-cells reveals human-like diversification and tolerance checkpoints in the mouse. PLoS ONE 7:e35497
62. Rieger M, Cervino C, Sauceda JC, Niessner R, Knopp D (2009) Efficient hybridoma screening technique using capture antibody based microarrays. Anal Chem 81:2373–2377
63. Ogunniyi A, Story C, Papa E, Guillen E, Love J (2009) Screening individual hybridomas by microengraving to discover monoclonal antibodies. Nat Protoc 4:767–782
64. Debs BE, Utharala R, Balyasnikova IV, Griffiths AD, Merten CA (2012) Functional single-cell hybridoma screening using droplet-based microfluidics. Proc Natl Acad Sci 109:11570–11575
65. Smith G (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. Sci 228:1315–1317
66. McCafferty J, Griffiths AD, Winter G, Chiswell DJ (1990) Phage antibodies: filamentous phage displaying antibody variable domains. Nat 348:552–554
67. Marks JD et al (1992) Bypassing immunization–building high-affinity human antibodies by chain shuffling. Bio-Technol 10:779–783
68. Griffiths AD et al (1994) Isolation of high-affinity human antibodies directly from large synthetic repertoires. EMBO J 13:3245–3260
69. Sblattero D, Bradbury A (2000) Exploiting recombination in single bacteria to make large phage antibody libraries. Nat Biotechnol 18:75–80
70. Mazor Y, Van Blarcom T, Carroll S, Georgiou G (2010) Selection of full-length IgGs by tandem display on filamentous phage particles and Escherichia coli fluorescence-activated cell sorting screening. FEBS J 277:2291–2303
71. D'Angelo S et al (2014) From deep sequencing to actual clones. Protein Eng Des Sel 27:301–307

72. Chan CEZ, Lim, APC, MacAry, PA, Hanson BJ (2014) The role of phage display in therapeutic antibody discovery. Int Immunol dxu082. doi:10.1093/intimm/dxu082
73. Zhou T et al (2013) Multidonor analysis reveals structural elements, genetic determinants, and maturation pathway for HIV-1 neutralization by VRC01-Class antibodies. Immun 39:245–258
74. Lavinder JJ et al (2014) Identification and characterization of the constituent human serum antibodies elicited by vaccination. Proc Natl Acad Sci 111:2259–2264
75. Boutz DR et al (2014) Proteomic identification of monoclonal antibodies from serum. Anal Chem 86:4758–4766
76. Sato S et al (2012) Proteomics-directed cloning of circulating antiviral human monoclonal antibodies. Nat Biotech 30:1039–1043
77. Cheung WC et al (2012) A proteomics approach for the identification and cloning of monoclonal antibodies from serum. Nat Biotech 30:447–452
78. Lee J et al (2016) Quantitative, molecular-level analysis of the serum antibody repertoire in young adults before and after seasonal influenza vaccination. Nat Med, Accepted
79. Doria-Rose NA et al (2014) Developmental pathway for potent V1V2-directed HIV-neutralizing antibodies. Nature 509:55–62
80. Zhu J, et al (2012) Somatic populations of PGT135-137 HIV-1-neutralizing antibodies identified by 454 pyrosequencing and bioinformatics. Front Microbiol 3
81. Marcus JS, Anderson WF, Quake SR (2006) Microfluidic single-cell mRNA isolation and analysis. Anal Chem 78:3084–3089
82. Toriello NM et al (2008) Integrated microfluidic bioprocessor for single-cell gene expression analysis. Proc Natl Acad Sci USA 105:20173–20178
83. White AK et al (2011) High-throughput microfluidic single-cell RT-qPCR. Proc Natl Acad Sci USA 108:13999–14004
84. Furutani S, Nagai H, Takamura Y, Aoyama Y, Kubo I (2012) Detection of expressed gene in isolated single cells in microchambers by a novel hot cell-direct RT-PCR method. Anal 137:2951–2957
85. Turchaninova MA et al (2013) Pairing of T-cell receptor chains via emulsion PCR. Eur J Immunol 43:2507–2515
86. Meijer P et al (2006) Isolation of human antibody repertoires with preservation of the natural heavy and light chain pairing. J Mol Biol 358:764–772
87. Smith K et al (2009) Rapid generation of fully human monoclonal antibodies specific to a vaccinating antigen. Nat Protoc 4:372–384
88. Frölich D et al (2010) Secondary immunization generates clonally related antigen-specific plasma cells and memory B-cells. J Immunol 185:3103–3110
89. Smith K et al (2013) Fully human monoclonal antibodies from antibody secreting cells after vaccination with Pneumovax®23 are serotype specific and facilitate opsonophagocytosis. Immunobiol 218:745–754
90. Poulsen TR, Meijer P-J, Jensen A, Nielsen LS, Andersen PS (2007) Kinetic, affinity, and diversity limits of human polyclonal antibody responses against tetanus toxoid. J Immunol 179:3841–3850
91. Meijer P-J, Nielsen LS, Lantto J, Jensen A (2009) Human antibody repertoires. In: Dimitrov AS (ed), Therapeutic antibodies: Methods and protocols. New York, USA: Humana Press, 525:261–277
92. DeKosky BJ et al (2013) High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. Nat Biotech 31:166–169
93. DeKosky BJ et al (2015) In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire. Nat Med 21:86–91