Studying the Antibody Repertoire after Vaccination: Practical Applications

Jacob D. Galson[1], Andrew J. Pollard[1], Johannes Trück[1] & Dominic F. Kelly[1]

[1]Oxford Vaccine Group, Department of Paediatrics, University of Oxford and the NIHR Oxford Biomedical Research Centre, Oxford, UK

Abstract

Nearly all licensed vaccines have been developed to confer protection against infectious diseases by stimulating the production of antibodies by B cells, but the nature of a successful antibody response has been difficult to capture. Recent advances in next-generation sequencing technology have allowed high-resolution characterization of the antibody repertoire, and of the changes that occur following vaccination. These approaches have yielded important insights into the B cell response, and have raised the possibility of using specific antibody sequences as measures of vaccine immunogenicity. Here, we review recent findings based on antibody repertoire sequencing, and discuss potential applications of these new technologies and of the analyses of the increasing volume of antibody sequence data in the context of vaccine development.

Corresponding authors: Galson, J.D. (jacob.galson@paediatrics.ox.ac.uk)

Pollard, A.J. (andrew.pollard@paediatrics.ox.ac.uk)

Keywords: antibody repertoire; B cell receptor repertoire; VDJ; vaccination

Introduction

Since the development of the first vaccine by Edward Jenner in 1796, vaccination has become one of the most important public health interventions worldwide. Where effective vaccines are used, they have greatly reduced the burden of infectious disease, preventing millions of deaths each year [1]. Most licensed vaccines confer protection against disease by stimulating B cells to produce pathogen-specific antibodies. The number of circulating B cell clones with distinct antibody heavy chain rearrangements is estimated to be at least 10^6 [2–4], although there is the theoretical potential for every one of a human adult's 10^{11} B cells to produce a unique antibody variant [5]; such diversity is a key element in the recognition of any potential pathogen. The sum of all circulating antibodies is known as the antibody repertoire.

Antibodies are large molecules, consisting of paired heavy and light polypeptide chains. These form a variable (V) antigen-binding region (known as V_H and V_L for the heavy and light chains respectively), as well as a constant (C) region [6]. Variability is concentrated in the three complementarity-determining regions (CDR1-3), which form a variety of tertiary structures to bind different antigens [7]. Initial $V_{H/L}$ diversity is generated during B cell development by somatic recombination of multiple variable (V), diversity (D; V_H only) and joining (J) gene segments, to form functional V_H and V_L regions [8]. Diversity is increased at the junctions between segments by the addition of palindromic (P) and nontemplated (N) nucleotides, and exonuclease activity leading to potential nucleotide deletion. During response to an antigen, further $V_{H/L}$ diversification occurs through rounds of somatic hypermutation, followed by selection of B cells for improved antigen binding in the germinal center (GC) [9].

An individual's antibody repertoire may be determined by a variety of factors including their genotype and chromatin structure [10,11], antigen exposure history [12] and age [12–17]. The diversity of the antibody repertoire makes it difficult to study, but improvements in next-generation sequencing (NGS) over the last decade now allow parallel sequencing of millions of antibody sequences (Box 1), making in-depth studies of the repertoire possible. There is an increasing body of data characterizing changes in the antibody repertoire following vaccination (influenza [4,13,16,18–21], tetanus [18,22,23], *Haemophilus influenzae* type b [24,25] and *Streptococcus pneumoniae* [13,26]), and natural infection (including influenza

[27–32], rotavirus [33–36], HIV [37–44], hepatitis C [45], cytomegalovirus (CMV) [12], Epstein-Barr virus (EBV) [12], *Staphylococcus aureus* [46] and dengue [47–51] among others). The majority of the vaccine studies are based on low-resolution methods for characterizing the repertoire, but there are six recent studies, that use NGS for in-depth characterization (Table 1).

Despite the increasing number of studies of the antibody repertoire, there is little mention of the application of these data, with most publications instead focusing on methods development [18,52]. Here, we summarize recent advances in repertoire sequencing, and the insight this has given into the antibody repertoire. We discuss both the use of vaccines as a tool with which to investigate the antibody repertoire, and also the use of antibody repertoire sequencing as a tool for vaccinology that can potentially be applied to aid the development and assessment of new vaccines and vaccine schedules.

Methods for studying the Antibody Repertoire

Low-resolution methods

The earliest studies of the antibody repertoire used isoelectric focusing of antibodies on polyacrylamide gels [53], resolving the antibodies into patterns of discrete bands based on their isoelectric pH-values. CDR3 size spectratyping, a polymerase chain reaction (PCR) and electrophoresis-based method for determining CDR3 diversity based on nucleotide length distribution, can also give some insight into the antibody repertoire after vaccination [14]. The advent of Sanger sequencing allowed the exact nucleotide sequences coding for specific antibodies to be determined, albeit in small numbers [54]. Lymphocytes can be isolated after vaccination, followed by production of immortalized cell lines. Rearranged immunoglobulin DNA can then be amplified from these cell lines, and sequenced [24,55]. Additionally, cloning amplified DNA into expression vectors allows functional characterization of sequences for antigen-specificity [56]. Fluorescence-activated single-cell sorting (FACS) has allowed more precise definition of the B cell subsets being studied. Sufficient DNA for sequencing or cloning can be generated from single cells by clonal expansion using culture [33], or single-cell reverse transcription PCR [20,57–59].

Such methods have been used to investigate small numbers of monoclonal antibodies (mAbs) generated against a variety of antigens, and although limited by cell numbers, we now know some antigen-specific sequences used in response to vaccines against influenza, tetanus, *Haemophilus influenzae* type b (Hib) and some serotypes of *Streptococcus pneumoniae* (Table 1), as well as natural infection (reviewed in [56]). These studies have shown that the B cell response to an antigen with a simple biochemical structure, such as Hib polysaccharide, appears markedly oligoclonal, with similar clones (predominantly V_H3-23 [24,55,60,61], with a conserved 'GYGMD' CDR3 amino acid motif) dominating the repertoire in different individuals (although it should be noted that these studies were limited to a total of 19 individuals, and 25 sequences between them) [62]. Repertoire diversity is also restricted for protein antigens (e.g. tetanus toxoid, influenza haemagglutinin (HA)) [20,22,23], but the expanded clones differ more between individuals (again limited to at most 14 individuals, and <500 sequences per individual in any study) [20,22]. Even in a single individual, there appears to be little similarity in the response after repeated tetanus toxoid vaccination, with only one third of the clones sequenced being shared between vaccination events [22].

NGS methods

The advent of NGS (Box 1) enabled the sequencing of antibody genes from millions of cells simultaneously, giving greater insight into the entire antibody repertoire. An in-depth antibody repertoire study was initially conducted in zebrafish; this is a useful model system due to its small number of antibody-producing B cells (approximately 300,000), which is 5 orders of magnitude lower than in humans [63]. This study gave the first insights into the entire antibody repertoire of an organism, showing that zebrafish use 50-86% of all possible VDJ combinations, and the VDJ frequency distribution is similar between individual zebrafish. It is not possible to sequence every B cell in humans, so representative samples (generally derived from peripheral blood) are taken. The first human studies sought to discover frequencies of VDJ segment usage, the similarity in VDJ segment usage, and the CDR3 sequence between individuals, and to estimate the size of the antibody repertoire [2,3,5,11,64–67]. Size-estimates are difficult as samples taken are a small representation of the entire repertoire, but sequencing independent replicate libraries in a capture-recapture analysis indicates a minimum bound of 10^6 unique V_H rearrangements [2]. This estimate neglects size

differences in the repertoire of different B cell subsets, so there are likely to be a greater number of unique naïve B cells, but fewer unique memory cells than this [4]. Comparing the repertoire across individuals showed that VDJ segments were used in unequal frequencies in the repertoire, but that their pattern of usage was similar between individuals, indicating inherent biases in the VDJ recombination process, and preferential use of core genes [2,3,64]. Despite similar VDJ usage though, there was limited overlap in the CDR3 repertoires between the two individuals studied by Arnaout *et al*. [3].

The antibody repertoire after vaccination has been examined using NGS for four different vaccines (Table 1). These studies have been able to broadly demonstrate minor changes in VDJ segment usage, and the size and diversity of the different B cell lineages after vaccination [13,16]. By comparing NGS data to sequences from known antigen-specific cells, it appears that the repertoire obtained by NGS includes antibody sequences that are targeted against vaccine antigens [16,18]. A recent study used V_H sequence databases generated after vaccination to allow characterization of the serum antibody repertoire using high-resolution liquid chromatography tandem mass spectrometry [68]. This technique has great potential for demonstrating the relationship between the V_H sequence repertoire, and the repertoire of serum antibodies, but is currently limited by the requirement for individual sequence profiles on which to map the mass spectrometry data, so proteins represented in the serum, but not the sequence data, cannot be identified. Two studies assessed the difference between the trivalent-inactivated (TIV), and the live-attenuated influenza vaccine (LAIV), on the post-vaccination repertoire, and showed that TIV causes increased production of abundant IgG lineages compared to LAIV [4,16]. To determine whether age is a confounding factor in repertoire sequencing, two studies investigated differences in antibody repertoire with age, which are discussed later [13,16]. There remain several important challenges to making the most effective use of NGS data for studying vaccine responses. These are primarily related to the methodology used for sequencing/analysis (Box 2) and also our ability to reliably identify the small number of antigen-specific sequences from large datasets of the total repertoire (Figure 1).

Identification of the antigen-specific repertoire, and the use of vaccination as a model system

The ability to distinguish between the total antibody repertoire, and antigen-specific antibody repertoire is key for any attempt to reliably utilize NGS data for elucidation of B cell responses to vaccination (Figure 1). Whereas NGS methods are well suited to studying perturbations in the total sampled repertoire, to date the low-resolution methods have provided the most insight into the antigen-specific repertoire after vaccination (Table 1 and [56]). To assess the full diversity of the antigen-specific repertoire, it is necessary to have methods for identification of antigen-specific sequences from total NGS repertoire data.

One way to identify such sequences is to search for those that have a degree of identity to previously described sequences, a technique that Zhu *et al*. used for de-novo identification of new VRC01 antibodies (from a new donor) against HIV [69]. Cross-donor phylogenetic analysis using known VRC01 sequences, and sequences from the new donor, was used to identify novel VRC01 V_H sequences based on evolutionary similarity. V_L sequences were identified based on the presence of a five AA motif present in the previously described sequences, and when reconstituted with the V_H genes were able to neutralize HIV-1 with varying potency.

An *a priori* method for the identification of antigen-specific sequences from total repertoire data is by analysis of the antibody sequences that are shared between multiple individuals following recent exposure (through infection, immunization or autoimmunity) to a common antigen – this is termed the 'convergent repertoire'. A study of the antibody repertoire in the cerebrospinal fluid, and peripheral blood of six multiple sclerosis (MS) patients, revealed some minor biases in V_H gene usage in the cerebrospinal fluid repertoire of the MS patients, which were not seen in the control patients [70]. However, it should be noted that only two V_H genes were found at significantly greater levels in the MS patients, and the level of significance was low (p=0.04/0.01). A study comparing the antibody repertoire in 60 dengue patients during acute disease, and at convalescence, indicated the presence of convergent CDR3 signatures (predominantly 'ARLDYYYYYGMDL') between individuals which were enriched during acute disease compared to at convalescence or in healthy controls [47]. However, no work was subsequently done to confirm the

specificity of these sequences for dengue antigens. The extent to which an individual's B cell repertoire for an antigen is shared with other individuals remains uncertain and is a key question to answer in order to determine the utility of 'convergent repertoire' analysis. Given the degree of antibody diversity high-throughput NGS methods will be invaluable in clarifying this important area.

In the analysis the of convergent repertoire, sequence similarity may not only be in the form of exact sequence matches, but also specific shared motifs within the sequences, such as CDR3 length, V_H usage, and amino acid usage in key locations, which could require sophisticated algorithms for identification [47,62]. Vaccine studies are an ideal tool for investigating the degree to which convergent sequences are likely to be antigen-specific. The ability to control the timing of antigen administration and sampling in vaccine studies increases our ability to sample the repertoire at timepoints when there is maximum egress of antigen-specific cells into the sampled compartment. Generally, an antigen-specific plasma cell burst is seen on days 4-10 in the peripheral blood following vaccination (with different kinetics for a primary vs secondary response [71–73]). If sampling is performed both before, and 4-10 days post-vaccination, filtering for sequences which increase, and persist at multiple timepoints after vaccination, in addition to analysis of convergence, would facilitate identification of antigen-specific sequences. The increase in sequence abundance will be even more pronounced if mRNA is used as the template (see Box 2, point 1), due to the high level of transcriptional activity in plasma cells [74]. In addition, FACS can be used to enrich for plasma cells or antigen-specific cells, so that a greater percentage of the sampled repertoire is antigen-specific (Figure 1). An additional consideration is that different B cell subsets have different kinetics after vaccination [23]. The majority of NGS studies have focused on sampling around day 7, where plasma cells are most likely to peak, yet with sufficient depth of sequencing it should also be possible to detect convergent repertoires related to antigen-specific memory cell populations present at later time points as well [23].

Vaccine studies which include both sequencing of the total repertoire, and the repertoire of sorted antigen-specific cells (or antigen-enriched plasma cells), will facilitate analysis of the extent to which antigen-specific sequences are shared (convergent or 'public' repertoire), compared to those antigen-specific sequences, which may be specific for a particular individual ('private' repertoire). Furthermore, functional

characterization of the putatively antigen-specific sequences will demonstrate whether the most functional antibodies are predominantly present in the private, or shared part of the antigen-specific repertoire.

Sequencing the antibody repertoire as a tool for vaccinology

Antibody repertoire sequencing technology has potential applications in multiple areas of vaccine development, including measurement of vaccine response, predicting vaccine safety, and guiding the development of more effective vaccines by increasing our understanding of B cell and GC immunology.

Generating mAbs

Perhaps the most well defined use of antibody sequencing after vaccination is in the generation of mAbs. mAbs can be used for passive immunization against certain diseases [75], and have important research applications, including verification of methods for identification of antigen-specific sequences from NGS repertoire data. After antigen administration, plasma cells (or antigen-enriched cells) can be isolated from peripheral blood by FACS and sequenced to identify the sequences of the antibodies that are produced in response to the antigen. Generally, the most abundant sequences are then cloned into an expression vector for functional characterization. The requirement of both V_H and V_L sequences to be obtained for mAb generation resulted in initial studies using a single-cell sorting approach followed by V_H and V_L sequencing of individual cells [20,59]. As methods are developed for linking the V_H:V_L repertoires (Box 2), it will be possible to identify mAbs in a more high-throughput manner (as demonstrated by Reddy *et al*. [76]*)*. Despite the exact method used, such studies have demonstrated 6-76% [22,59,76,77] of cell lines created produce antigen-specific antibodies, resulting in a significant decrease in the time taken to generate effective mAbs compared to the conventional methods of screening large combinatorial libraries.

Tracking known sequences

There are numerous antibody sequences in the literature that have a known specificity (Table 1 and [56]). Incorporating repertoire sequencing into studies of vaccines, autoimmunity and infectious diseases will aid in the creation of increasingly comprehensive databases containing antibody sequences with defined specificity

and function (Figure 2). In theory, if these databases contained sequences known to be specific for a certain vaccine antigen, we could search for these previously described sequences in the post-immunisation repertoire during new studies of the vaccine antigen, and infer the degree to which a protective immune response has been generated. Current data from a limited range of antigens suggest that a vaccine response is likely to be complex, and in-depth studies of the antibody repertoire will be required to fully understand the relationship between NGS data and immunogenicity, as discussed later. Although there are currently few known sequences for auto-reactive antibodies, the identification of known auto-reactive sequences in the post-vaccination repertoire, would be of interest in terms of signaling the potential for certain vaccines to increase the risk of autoimmune disease. However, the large number of 'self' epitopes, and the potential diversity of auto-reactive sequences that could target these epitopes make it unlikely that all potential auto-reactive sequences could ever be identified. A recent study of the antibody repertoire in MS patients identified expansion of some sequences potentially related to the disease [70] (although this was not confirmed with functional assays), and as such sequences are discovered, it would be beneficial to routinely test for them in vaccine trials as a precautionary measure.

Considering a specific example of the utility of tracking known sequences for vaccine development, in the case of HIV, neutralizing antibodies have been discovered against the envelope glycoprotein (Env) in HIV-infected individuals, but stimulating formation of these antibodies by vaccination is difficult, as their germline precursors have poor affinity for Env [78,79]. In some HIV infected individuals, it appears that neutralizing antibodies may be formed from mutated, cross-reactive precursors that were initially generated during previous non-HIV infections, and then further mutate to become specific for Env [80]. It has been observed in one study, that the unmutated precursor of an effective anti-HIV antibody was able to bind to the founder HIV virus. Sequencing the viral Env and antibody genes at multiple timepoints after virus transmission indicated that antibody maturation was preceded by viral evolution, suggesting that such co-evolution is a potential method for driving formation of anti-Env neutralizing antibodies [81]. It seems for HIV, that an effective vaccine requires prior identification of the correct germline precursor antibody sequence. The germline precursor could then be stimulated using an alternate immunogen to Env, and

successive immunogens then administered to guide B cell maturation to form the desired final antibody [82]. The success of this work is being driven by the ability to identify low-abundance antibody sequences in high-depth antibody sequencing experiments [79].

Investigating breadth of the antibody response

As well as identifying specific sequences, NGS data can give insight into the breadth of the antibody repertoire (a measure of the diversity, in terms of the number and abundance of different sequence clones [83]) used in an immune response. For some vaccines, the antigenic site targeted is highly variable (HIV-1 Env/*Plasmodium vivax* Duffy-binding protein), or undergoes seasonal change (influenza HA); a broader response can potentially confer protection against more antigenic variants. Antibody repertoire sequencing has been instrumental in demonstrating that use of a Toll-like receptor (TLR) agonist in an oil-in-water adjuvanted malaria vaccine formulation is able to broaden the antibody response in mice. Use of the TLR agonist in the vaccine led to improved antigen neutralization, and efficacy against more varied malarial strains [84]. Similar observations have been made using oil-in-water adjuvanted influenza vaccines, which induced broadly cross-reactive responses against the HA head, providing some protection against drifted strains of the virus [85].

While it seems reasonable to suggest that the breadth of the antibody repertoire is a measure of ability to neutralize diverse antigenic strains, it is also possible that single antibodies can have broadly neutralizing activity by targeting conserved regions on otherwise variable antigens. This appears to be the case with influenza HA, where broad neutralization of multiple variants can arise from a clonal antibody response targeting the conserved stem region or a diverse response targeting the variable head region [86]. Stem-binding antibodies are not found after seasonal influenza vaccination, but of the 28 HA-binding mAbs generated from eight subjects after 2009 pandemic H1N1 influenza vaccination, Li *et al*. found that three of these were able to bind the stem region and elicit cross-reactive protection [86]. The authors hypothesize that as the influenza response is predominately driven by memory cells, and the pandemic virus has a highly divergent HA head compared to seasonal variants, only the cross-reactive memory cells would be stimulated. This is a contested issue though, as a study by Wrammert *et al*. in which over 50 influenza-specific mAbs

were generated from plasma cell sequences after influenza vaccination showed that they exhibited the highest affinity for the current circulating strain of influenza [20]. The conflicting nature of these studies indicate that the high-throughput nature of NGS data would be better able to give insight into the true breadth of the response after seasonal influenza vaccination. Vollmers *et al*. have developed a novel method of barcoding individual RNA molecules, and building consensus reads to generate accurate antibody sequence data, which they used to demonstrate the appearance of some identical sequences after two different annual influenza vaccines in the same individuals (Table 1) [4]. Using conventional immunological assays in parallel with antibody repertoire sequencing would also be useful to improve our understanding of how repertoire data relates to functional measures of cross-reactivity.

Understanding Immunological Mechanisms of Vaccination

The majority of vaccines and immunization schedules were initially developed empirically, rather than through a detailed understanding of the immunological mechanisms of vaccination. Data generated from antibody repertoire sequencing both after vaccination, and in populations without vaccine challenge, is starting to give great insight into the underlying immunology of vaccination. For vaccines such as hepatitis B, where different dose schedules are available, the schedule used affects vaccine immunogenicity [87]. Using repertoire sequencing to improve understanding of the GC reaction and antibody maturation pathways could aid the design of optimal dose schedules. Booster doses of a vaccine are used to increase the number and affinity of antigen-specific antibodies. Using surface plasmon resonance to analyze the antibody repertoire after repeated doses of tetanus toxoid (TT) vaccination in two individuals, the limits of antibody affinity maturation were determined to be reached after just two doses for this antigen [22]. Sequencing plasma cells produced after these vaccinations indicated that the limit of somatic hypermutation was also reached after two doses. A third dose of TT did not increase either the affinity of the anti-TT antibodies, or the diversity of the repertoire [22]. Repeat doses of vaccine given close together may increase the magnitude of the antibody response by re-stimulating GC's, or by stimulating new GC's to form, resulting in the production of new plasma cells. Sorting and sequencing antigen-specific B cells present in the peripheral blood after vaccination may allow capture of recent GC emigrants. Observing the relative numbers of these

cells, and their mutational load, could give insight into proliferation and affinity maturation within the GC after repeat vaccination. It would be instructive to understand the interaction between dose interval and antibody maturation; the repertoire may become more clonal after repeat doses due to selection in the GC. Dose schedules could then be altered accordingly to optimize responses, and prevent unnecessary vaccinations.

It is also important to consider that the total B cell repertoire is actually composed of a combination of repertoires from different B cell subsets [12,88–90], and that these different B cell subsets could be responding differently to vaccination [91]. To determine whether IgM memory cells were early emigrants from the T-dependent GC response (and so related to IgG memory cells), or were instead formed after stimulation with T-independent antigens, the repertoire in IgM memory and IgG memory cells was compared in a population with no vaccine challenge. Wu *et al*. and Briney *et al*. both observed differences in VDJ segment composition in the IgM and IgG memory repertoires [88,89]. Wu *et al*. also found the IgM repertoire to contain fewer negatively charged amino acids, and greater levels of tyrosine, as well as lower hydrophobicity and aliphatic index, supporting the notion that IgG and IgM memory cells could comprise populations with distinct origins, which have been formed through responses to different antigenic stimuli (although it cannot be ruled out that IgG and IgM memory cells have the same origins, but are subject to distinct regulatory mechanisms) [88]. However, by defining IgM memory by having a V gene mutation frequency of greater than 1%, rather than by sorting CD27+ cells, Wang *et al*. found little difference between VDJ segment composition in IgM, IgD, IgG or IgA sequences [12]. This topic could clearly benefit from further study, and a clarification on the optimal definitions of how to sort cell subsets either physically before sequencing, or computationally from the sequence data. Nevertheless, it is clearly the case that cell subsets and isotype should be taken into account in studies of the antibody repertoire after vaccination.

Measuring vaccine immunogenicity

Developing new vaccines, and predicting efficacy of available vaccines in new populations requires measuring correlates of immunity. Correlates are generally measures of the immune response, such as antibody concentration, or function (e.g. serum bactericidal assay for group C *N. meningitides* or HA

inhibition assay for influenza), which relate to vaccine-induced protection against infection. For many vaccines, such correlates are often unavailable, or hard to measure and standardize between laboratories [92]. Sequencing the antibody repertoire after vaccination can provide a detailed dissection of the vaccine-induced B cells underlying the antibody response, and a potential application of this is in deriving novel correlates of protection. Although repertoire data can be used to identify known antigen-specific sequences, and the breadth of the response, it remains to be demonstrated whether it can be used as a valid standalone measure of immunogenicity. A critical first step is to distinguish the vaccine-specific repertoire from the total repertoire, and as discussed, this is increasingly feasible. The vaccine-specific repertoire (in terms of number and abundance of sequences) would then need to be correlated with a known indicator of vaccine response, or with protection from infection. Current NGS depth and cost (Box 1) makes it an increasingly feasible proposition to carry out repertoire sequencing in conjunction with routine clinical tests on a scale compatible with that of clinical vaccine trials necessary for investigating this application. Recent methods allowing high-throughput identification of serum antibodies will also help to show to which antigen-specific sequences ultimately give rise to long-lived plasma cells and mediate long-term protection [52,68]. This technique has so far been used to investigate the serum antibody repertoire in two individuals after TT vaccination, and showed that <5% of antigen-specific plasma cell sequences identified 7 days post-vaccination, are found to encode antibodies that are also present in the serum response 9 months post-vaccination at sufficient levels to be detected [68]. Other aspects that must be considered for a good correlate of immunity are that the signal detected is not transient, and that it is not confounded by other individual factors such as genotype, antigen exposure history, and age. Signal persistence in antibody repertoire studies has not been well investigated. It is clear that perturbations can be seen on days 7, 8 and 14 after vaccination (Table 1), and some sequences are found to persist in individuals for at least a year, and are identifiably re-stimulated upon further vaccination [4]. It is less clear to what degree such perturbations in the total repertoire reflect perturbations in the antigen-specific repertoire, but as methods for identifying the antigen-specific repertoire improve (Figure 1), this should become clearer.

The potential confounding effects of genotype, antigen exposure history, and age have been investigated to different degrees. By studying the antibody repertoire in two pairs of monozygotic twins, Glanville *at al*. investigated the relative influence of genetic and environmental factors in structuring the naïve repertoire [11]. Although limited to just four individuals, they found that certain V_H and D_H segments were used at significantly different frequencies between the different twins compared to within twin pairs, indicating that individual genetic differences should be taken into account when studying changes in VDJ segment usage after vaccination. There was limited overlap in the CDR3 repertoires both between and within twins, which may be due to limitations in sequencing depth (so there may be more overlap if more sequences were obtained), but is indicative of the random nature of VDJ junctional diversification during B cell development and somatic hypermutation during antigen-response, which drive diversification of the antibody repertoire. An additional complication is that the genotype at other loci as well as the immunoglobulin locus may also affect the antibody repertoire, as has been shown by the impact of an individuals human leukocyte antigen locus genotype on their T cell receptor repertoire [93]. It would be of interest to conduct a similar study to investigate the relationship between features of the antibody repertoire (such as VDJ composition), and genotype at other immunologically important loci.

As well as varying between individuals, vaccine response varies with age [94], but the reasons for this are not fully known. Jiang *et al*. studied the antibody repertoire before and after influenza vaccination in children (8-17 years), young adults (18-30 years) and the elderly (70-100 years) [14]. As there were only four individuals in the elderly age group, statistical comparisons were not performed, but this study did show that in two of the elderly individuals, the repertoire was highly clonal and had a greater mutational load compared to the younger individuals. In another study, the isotype-specific antibody repertoire was studied after simultaneous administration with pneumococcal and influenza vaccination in six younger (19-45 years) and six older (70-89 years) individuals [13]. There was a slight increase in IgG mutation in the older age group, but the most striking differences were in the IgA and IgM repertoires, which displayed slower clonal expansion as well as less mutation, and longer CDR3 regions in the elderly. Even with no vaccine stimulation, there appear to be age-related differences in the naïve repertoire, with elderly individuals having

different VDJ recombination frequencies [95], longer CDR3 regions, greater persistence of large clones, and increased mutational loads [12] compared to younger individuals. Interestingly, sex does not appear to effect the repertoire, despite the presence of other known sex-specific effects on the immune system [96]. Hence, it is clear that there are age-related and isotype-dependent effects on the repertoire, but it is necessary to conduct larger vaccine trials in individuals of different ages (including infants, which have been previously omitted), so that statistical comparisons can be made to confirm these findings.

Considering antigen exposure history, the potential effect of previous influenza vaccination on subsequent influenza vaccinations has been discussed, indicating the importance of documenting vaccine and disease history where possible. In addition to previous infections impacting how the repertoire responds to vaccination, the effect of chronic infection needs to be considered. Wang *et al*. documented chronic CMV and EBV infection status in a repertoire study of 27 individuals over two years [12]. CMV infection resulted in increased V_H mutation in the IgG and IgM repertoire, while EBV infection resulted in an increased number of persistent clonal groups in the repertoire Neither infection resulted in altered VDJ gene segment usage in the repertoire, but segment usage does appear to be different between healthy donors, patients with chronic hepatitis C infection, and in patients who have recovered from hepatitis C infection [45]. As the number of antigens an individual has been exposed to, and the rate of chronic viral infection increase with age, the effect of antigen exposure history, and age on the antibody repertoire will be tightly linked.

So, although antibody repertoire sequencing has promise as a novel tool to measure vaccine immunogenicity, more work is first required in order to improve and validate methods for identifying the antigen-specific repertoire (Figure 1). It is also essential that repertoire sequencing studies are conducted on larger sample sets so that more robust statistical comparisons can be made to investigate the effect of age, antigen exposure history and genotype, as well as to allow correlations to be drawn between repertoire data, and known correlates of immunity.

Concluding remarks

Understanding of the antibody repertoire has increased enormously over the last decade, due mainly to improved sequencing technology. There is an increasingly comprehensive view of the structure of the baseline antibody repertoire, and how this differs between individuals and B cell subsets. An appreciation of how vaccine antigens dynamically restructure the antibody repertoire, and how certain adjuvants affect this process is also emerging.

Laboratory and bioinformatic techniques available for studying the antibody repertoire are constantly improving. Major advances in the last year now allow high-throughput identification of paired V_H:V_L sequences [18], quantification of PCR-related bias and mutation [4], and identification of putatively antigen-specific sequences through repertoire convergence [47]. As groups working on vaccine development increasingly use repertoire sequencing, it is important that these different methods are synthesized, and a 'gold standard' sequencing protocol agreed so that data generated from different laboratories can be reliably compared. To make full use of available data, the creation of an open-access database for the accumulation of repertoire data after vaccination, natural infection and in autoimmune disease needs to be considered (Figure 2). Researchers can then concentrate on how to reliably interpret and make practical use of the resulting data, in the face of the potential confounding factors of genotype, age and antigen exposure history.

Analysis of the post-vaccination antibody repertoire is providing a fascinating view of the B cell response to vaccines. There is a compelling case that the practical application of these data will have a profound affect on the development of new vaccines. The use of vaccines as a model system for studying the repertoire also prefigures the use of repertoire sequencing technology in the more complex immune responses apparent during autoimmunity and disease.

Acknowledgements

J.D.G. receives funding from the BBSRC and GlaxoSmithKline in the form of an iCASE studentship. J.T. is a James Martin Fellow and was supported by an ESPID Fellowship Award. A.J.P. is a Jenner Investigator

and James Martin Senior Fellow. A.J.P. & D.F.K are supported by the NIHR Oxford Biomedical Research Centre.

Competing interests

A.J.P. acts as an investigator on behalf of Oxford University for clinical trials of vaccines for which the University receives grants from Pfizer, Novartis, GlaxoSmithKline and Sanofi Pasteur MSD, but does not receive any personal payments from them. J.T. and D.F.K. have received support from pharmaceutical companies to attend scientific meetings.

References

1 Ehreth, J. (2003) The global value of vaccination. *Vaccine* 21, 596–600

2 Boyd, S.D. *et al*. (2009) Measurement and Clinical Monitoring of Human Lymphocyte Clonality by Massively Parallel V-D-J Pyrosequencing. *Sci. Transl. Med*. 1, 12ra23–12ra23

3 Arnaout, R. *et al*. (2011) High-resolution description of antibody heavy-chain repertoires in humans. *PLoS One* 6, e22365

4 Vollmers, C. *et al*. (2013) Genetic measurement of memory B-cell recall using antibody repertoire sequencing. *Proc. Natl. Acad. Sci. U.S.A*. 110, 13463–8

5 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. U.S.A*. 106, 20216–21

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

7 North, B. *et al*. (2011) A new clustering of antibody CDR loop conformations. *J. Mol. Biol*. 406, 228–256

8 Tonegawa, S. (1983) Somatic Generation of Antibody Diversity. *Nature* 302, 575–581

9 Victora, G.D. and Nussenzweig, M.C. (2012) Germinal centers. *Annu. Rev. Immunol*. 30, 429–57

10 Shih, H. and Krangel, M.S. (2013) Chromatin Architecture, CCCTC-Binding Factor, and V(D)J Recombination: Managing Long-Distance Relationships at Antigen Receptor Loci. *J. Immunol*. 190, 4915–4921

11 Glanville, J. *et al*. (2011) Naive antibody gene-segment frequencies are heritable and unaltered by chronic lymphocyte ablation. *Proc. Natl. Acad. Sci. U.S.A*. 108, 20066–71

12 Wang, C. *et al*. (2013) Effects of Aging, Cytomegalovirus Infection, and EBV Infection on Human B Cell Repertoires. *J. Immunol*. DOI: 10.4049/jimmunol.1301384

13 Wu, Y.-C.B. *et al*. (2012) Age-Related Changes in Human Peripheral Blood IGH Repertoire Following Vaccination. *Front. Immunol.* 3, 193

14 Ademokun, A. *et al*. (2011) Vaccination-induced changes in human B-cell repertoire and pneumococcal IgM and IgA antibody at different ages. *Aging Cell* 10, 922–30

15 Dunn-Walters, D.K. and Ademokun, A. a (2010) B cell repertoire and ageing. *Curr. Opin. Immunol.* 22, 514–20

16 Jiang, N. *et al*. (2013) Lineage structure of the human antibody repertoire in response to influenza vaccination. *Sci. Transl. Med.* 5, 171ra19

17 Boyd, S.D. *et al*. (2013) Human lymphocyte repertoires in ageing. *Curr. Opin. Immunol.* 25, 511–5

18 Dekosky, B.J. *et al*. (2013) High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nat. Biotechnol.* 31, 166–169

19 Krause, J.C. *et al*. (2011) Epitope-specific human influenza antibody repertoires diversify by B cell intraclonal sequence divergence and interclonal convergence. *J. Immunol.* 187, 3704–11

20 Wrammert, J. *et al*. (2008) Rapid cloning of high-affinity human monoclonal antibodies against influenza virus. *Nature* 453, 667–71

21 Tan, Y.-C. *et al*. (2014) High-throughput sequencing of natively paired antibody chains provides evidence for original antigenic sin shaping the antibody response to influenza vaccination. *Clin. Immunol.* 151, 55–65

22 Poulsen, T.R. *et al*. (2011) Limits for antibody affinity maturation and repertoire diversification in hypervaccinated humans. *J. Immunol.* 187, 4229–35

23 Frölich, D. *et al*. (2010) Secondary immunization generates clonally related antigen-specific plasma cells and memory B cells. *J. Immunol.* 185, 3103–3110

24 Adderson, E. *et al*. (1993) Restricted immunoglobulin VH usage and VDJ combinations in the human response to Haemophilus influenzae type b capsular polysaccharide. *J. Clin. Invest.* 91, 2734–2743

25 Liu, L. and Lucas, A.H. (2003) IGH V3-23*01 and its allele V3-23*03 differ in their capacity to form the canonical human antibody combining site specific for the capsular polysaccharide of Haemophilus influenzae type b. *Immunogenetics* 55, 336–8

26 Rohatgi, S. *et al*. (2009) Molecular dissection of antibody responses against pneumococcal surface protein A: evidence for diverse DH-less heavy chain gene usage and avidity maturation. *J. Immunol.* 182, 5570–85

27 Wrammert, J. *et al*. (2011) Broadly cross-reactive antibodies dominate the human B cell response against 2009 pandemic H1N1 influenza virus infection. *J. Exp. Med.* 208, 181–93

28 Yu, X. *et al*. (2008) Neutralizing antibodies derived from the B cells of 1918 influenza pandemic survivors. *Nature* 455, 532–6

29 Krause, J.C. *et al*. (2012) Human monoclonal antibodies to pandemic 1957 H2N2 and pandemic 1968 H3N2 influenza viruses. *J. Virol.* 86, 6334–40

30 Simmons, C.P. *et al*. (2007) Prophylactic and therapeutic efficacy of human monoclonal antibodies against H5N1 influenza. *PLoS Med.* 4, e178

31 Whittle, J.R.R. *et al*. (2011) Broadly neutralizing human antibody that recognizes the receptor-binding pocket of influenza virus hemagglutinin. *Proc. Natl. Acad. Sci. U. S. A.* 108, 14216–21

32 Corti, D. *et al*. (2011) A neutralizing antibody selected from plasma cells that binds to group 1 and group 2 influenza A hemagglutinins. *Science* 333, 850–6

33 Weitkamp, J.-H. *et al*. (2003) Infant and adult human B cell responses to rotavirus share common immunodominant variable gene repertoires. *J. Immunol.* 171, 4680–8

34 Weitkamp, J.-H. *et al*. (2003) Generation of recombinant human monoclonal antibodies to rotavirus from single antigen-specific B cells selected with fluorescent virus-like particles. *J. Immunol. Methods* 275, 223–237

35 Weitkamp, J.-H. *et al*. (2005) Natural evolution of a human virus-specific antibody gene repertoire by somatic hypermutation requires both hotspot-directed and randomly-directed processes. *Hum. Immunol.* 66, 666–76

36 Tian, C. *et al*. (2008) Immunodominance of the VH1-46 antibody gene segment in the primary repertoire of human rotavirus-specific B cells is reduced in the memory compartment through somatic mutation of nondominant clones. *J. Immunol.* 180, 3279–88

37 Wu, X. *et al*. (2011) Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. *Science (80-.).* 333, 1593–602

38 Zhu, J. *et al*. (2012) Somatic Populations of PGT135-137 HIV-1-Neutralizing Antibodies Identified by 454 Pyrosequencing and Bioinformatics. *Front. Microbiol.* 3, 315

39 Scheid, J.F. *et al*. (2009) Broad diversity of neutralizing antibodies isolated from memory B cells in HIV-infected individuals. *Nature* 458, 636–40

40 Zhu, J. *et al*. (2013) Mining the antibodyome for HIV-1 – neutralizing antibodies with next-generation sequencing and phylogenetic pairing of heavy / light chains. DOI: 10.1073/pnas.1219320110

41 Huang, C. *et al*. (2004) Structural basis of tyrosine sulfation and VH-gene usage in antibodies that recognize the HIV type 1 coreceptor-binding site on gp120. *Proc. Natl. Acad. Sci. U. S. A.* 101, 2706–11

42 Bonsignori, M. *et al*. (2011) Analysis of a clonal lineage of HIV-1 envelope V2/V3 conformational epitope-specific broadly neutralizing antibodies and their inferred unmutated common ancestors. *J. Virol.* 85, 9998–10009

43 Corti, D. *et al*. (2010) Analysis of memory B cell responses and isolation of novel monoclonal antibodies with neutralizing breadth from HIV-1-infected individuals. *PLoS One* 5, e8805

44 Scheid, J.F. *et al*. (2011) Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science* 333, 1633–7

45 Racanelli, V. *et al*. (2011) Antibody V(h) repertoire differences between resolving and chronically evolving hepatitis C virus infections. *PLoS One* 6, e25606

46 Lu, D.R. *et al*. (2014) Identification of functional anti-Staphylococcus aureus antibodies by sequencing patient plasmablast antibody repertoires. *Clin. Immunol.* DOI: 10.1016/j.clim.2014.02.010

47 Parameswaran, P. *et al*. (2013) Convergent Antibody Signatures in Human Dengue. *Cell Host Microbe* 13, 691–700

48 Beltramello, M. *et al*. (2010) The human immune response to Dengue virus is dominated by highly cross-reactive antibodies endowed with neutralizing and enhancing activity. *Cell Host Microbe* 8, 271–83

49 Smith, S. a *et al*. (2012) Persistence of circulating memory B cell clones with potential for dengue virus disease enhancement for decades following infection. *J. Virol*. 86, 2665–75

50 De Alwis, R. *et al*. (2011) In-depth analysis of the antibody response of individuals exposed to primary dengue virus infection. *PLoS Negl. Trop. Dis*. 5, e1188

51 Dejnirattisai, W. *et al*. (2010) Cross-reacting antibodies enhance dengue virus infection in humans. *Science* 328, 745–8

52 Wine, Y. *et al*. (2013) Molecular deconvolution of the monoclonal antibodies that comprise the polyclonal serum response. *Proc. Natl. Acad. Sci. U.S.A*. 110, 2993–8

53 Awdeh, Z.L. *et al*. (1968) Isoelectric Focusing in Polyacrylamide Gel and its Application to Immunoglobulins. *Nature* 219, 66–67

54 Seidman, J.G. *et al*. (1978) Multiple related immunoglobulin variable-region genes identified by cloning and sequence analysis. *Proc. Natl. Acad. Sci. U.S.A*. 75, 3881–5

55 Adderson, E.E. *et al*. (1991) Restricted Ig H chain V gene usage in the human antibody response to Haemophilus influenzae type b capsular polysaccharide. *J. Immunol*. 147, 1667–74

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

57 Wang, X. and Stollar, B.D. (2000) Human immunoglobulin variable region gene analysis by single cell RT-PCR. *J. Immunol. Methods* 244, 217–25

58 Tiller, T. *et al*. (2008) Efficient generation of monoclonal antibodies from single human B cells by single cell RT-PCR and expression vector cloning. *J. Immunol. Methods*. 329, 112–124

59 Smith, K. *et al*. (2009) Rapid generation of fully human monoclonal antibodies specific to a vaccinating antigen. *Nat. Protoc*. 4, 372–84

60 Lucas, A.H. *et al*. (1994) Variable region sequences of a protective human monoclonal antibody specific for the Haemophilus influenzae type b capsular polysaccharide. *Infect. Immun*. 62, 3873–80

61 Pinchuk, G. V *et al*. (1995) Predominant V-region gene configurations in the human antibody response to Haemophilus influenzae capsule polysaccharide. *Scand. J. Immunol*. 41, 324–30

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

63 Weinstein, J.A. *et al*. (2009) High-throughput sequencing of the zebrafish antibody repertoire. *Science (80-.).* 324, 807–10

64 Boyd, S.D. *et al*. (2010) Individual variation in the germline Ig gene repertoire inferred from variable region gene rearrangements. *J. Immunol*. 184, 6986–92

65 Larimore, K. and McCormick, M. (2012) Shaping of human germline IgH repertoires revealed by deep sequencing. *J. …* DOI: 10.4049/jimmunol.1201303

66 Benichou, J. *et al*. (2013) The restricted DH gene reading frame usage in the expressed human antibody repertoire is selected based upon its amino acid content. *J. Immunol*. 190, 5567–77

67 Briney, B.S. *et al*. (2012) Frequency and genetic characterization of V(DD)J recombinants in the human peripheral blood antibody repertoire. *Immunology* 137, 56–64

68 Lavinder, J.J. *et al*. (2014) Identification and characterization of the constituent human serum antibodies elicited by vaccination. *Proc. Natl. Acad. Sci*. DOI: 10.1073/pnas.1317793111

69 Zhu, J. *et al*. (2013) De novo identification of VRC01 class HIV-1-neutralizing antibodies by next-generation sequencing of B-cell transcripts. *Proc. Natl. Acad. Sci. U.S.A*. DOI: 10.1073/pnas.1306262110

70 Büdingen, H. Von and Kuo, T. (2012) B cell exchange across the blood-brain barrier in multiple sclerosis. *J. Clin. Invest*. 122, 4533–4543

71 Blanchard-Rohner, G. *et al*. (2009) Appearance of peripheral blood plasma cells and memory B cells in a primary and secondary immune response in humans. *Blood* 114, 4998–5002

72 Henn, A.D. *et al*. (2013) High-resolution temporal response patterns to influenza vaccine reveal a distinct human plasma cell gene signature. *Sci. Rep*. 3, 2327

73 Mitchell, R. *et al*. (2014) Polysaccharide-specific B cell responses to vaccination in humans. *Hum. Vaccin. Immunother*. 10, 0-1

74 Klein, U. *et al*. (1997) Evidence for a large compartment of IgM-expressing memory B cells in humans. *Blood* 89, 1288–98

75 Keller, M. a and Stiehm, E.R. (2000) Passive immunity in prevention and treatment of infectious diseases. *Clin. Microbiol. Rev*. 13, 602–14

76 Reddy, S.T. *et al*. (2010) Monoclonal antibodies isolated without screening by analyzing the variable-gene repertoire of plasma cells. *Nat. Biotechnol*. 28, 965–9

77 Smith, K. *et al*. (2013) Fully human monoclonal antibodies from antibody secreting cells after vaccination with Pneumovax®23 are serotype specific and facilitate opsonophagocytosis. *Immunobiology* 218, 745–54

78 Xiao, X. *et al*. (2009) Germline-like predecessors of broadly neutralizing antibodies lack measurable binding to HIV-1 envelope glycoproteins: implications for evasion of immune responses and design of vaccine immunogens. *Biochem. Biophys. Res. Commun*. 390, 404–9

79 Kwong, P.D. *et al*. (2013) Broadly neutralizing antibodies and the search for an HIV-1 vaccine: the end of the beginning. *Nat. Rev. Immunol*. 13, 693–701

80 Liao, H.-X. *et al*. (2011) Initial antibodies binding to HIV-1 gp41 in acutely infected subjects are polyreactive and highly mutated. *J. Exp. Med*. 208, 2237–49

81 Liao, H.-X. *et al*. (2013) Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature* 496, 469–76

82 Dimitrov, D.S. (2010) Therapeutic antibodies, vaccines and antibodyomes. *MAbs* 2, 347–56

83 Sepúlveda, N. *et al*. (2010) Estimation of T-cell repertoire diversity and clonal size distribution by Poisson abundance models. *J. Immunol. Methods* 353, 124–137

84 Wiley, S.R. *et al*. (2011) Targeting TLRs expands the antibody repertoire in response to a malaria vaccine. *Sci. Transl. Med*. 3, 93ra69

85 Khurana, S. *et al*. (2011) MF59 adjuvant enhances diversity and affinity of antibody-mediated immune response to pandemic influenza vaccines. *Sci. Transl. Med*. 3, 85ra48

86 Li, G.-M. *et al*. (2012) Pandemic H1N1 influenza vaccine induces a recall response in humans that favors broadly cross-reactive memory B cells. *Proc. Natl. Acad. Sci. U.S.A*. 109, 9047–52

87 Jilg, W. *et al*. (1989) Vaccination against hepatitis B: comparison of three different vaccination schedules. *J. Infect. Dis*. 160, 766–9

88 Wu, Y.-C. *et al*. (2010) High-throughput immunoglobulin repertoire analysis distinguishes between human IgM memory and switched memory B-cell populations. *Blood* 116, 1070–8

89 Briney, B.S. *et al*. (2012) High-throughput antibody sequencing reveals genetic evidence of global regulation of the naïve and memory repertoires that extends across individuals. *Genes Immun*. 13, 469–73

90 Mroczek, E.S. *et al*. (2014) Differences in the Composition of the Human Antibody Repertoire by B Cell Subsets in the Blood. *Front. Immunol*. 5, 1–14

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

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

93 Warren, R. *et al*. (2011) Exhaustive T-cell repertoire sequencing of human peripheral blood samples reveals signatures of antigen selection and a directly measured repertoire size of at least of at least 1 million clonotypes. *Genome Res*. 21, 790–797

94 Perrett, K.P. *et al*. (2010) Antibody persistence after serogroup C meningococcal conjugate immunization of United Kingdom primary-school children in 1999-2000 and response to a booster: a phase 4 clinical trial. *Clin. Infect. Dis*. 50, 1601–10

95 Rubelt, F. *et al*. (2012) Onset of immune senescence defined by unbiased pyrosequencing of human immunoglobulin mRNA repertoires. *PLoS One* 7, e49774

96 Markle, J.G. and Fish, E.N. (2013) SeXX matters in immunity. *Trends Immunol*. DOI: 10.1016/j.it.2013.10.006

97 Kolibab, K. *et al*. (2005) Immune Response to Pneumococcal Polysaccharides 4 and 14 in Elderly and Young Adults : Analysis of the Variable Heavy Chain Repertoire. *Infect. Immun*. 73, 7465–76

98 Laserson, U. *et al*. (2014) High-resolution antibody dynamics of vaccine-induced immune responses. *Proc. Natl. Acad. Sci. U. S. A*. DOI: 10.1073/pnas.1323862111

99 Shendure, J. and Ji, H. (2008) Next-generation DNA sequencing. *Nat. Biotechnol*. 26, 1135–45

100 Gao, F. *et al*. (2013) Characterizing immunoglobulin repertoire from whole blood by a personal genome sequencer. *PLoS One* 8, e75294

101 Bolotin, D. a *et al*. (2012) Next generation sequencing for TCR repertoire profiling: Platform-specific features and correction algorithms. *Eur. J. Immunol*. DOI: 10.1002/eji.201242517

102 Loman, N.J. *et al*. (2012) Performance comparison of benchtop high-throughput sequencing platforms. *Nat. Biotechnol*. 30, 434–9

103 Benichou, J. *et al*. (2012) Rep-Seq: uncovering the immunological repertoire through next-generation sequencing. *Immunology* 135, 183–91

104 Baum, P.D. *et al*. (2012) Wrestling with the repertoire: The promise and perils of next generation sequencing for antigen receptors. *Eur. J. Immunol*. 42, 2834–9

105 Matsuda, F. *et al*. (1998) The Complete Nucleotide Sequence of the Human Immunoglobulin Heavy Chain Variable Region Locus. *J. Exp. Med*. 188, 2151–2162

106 Zhu, J. *et al*. (2013) Mining the antibodyome for HIV-1-neutralizing antibodies with next-generation sequencing and phylogenetic pairing of heavy/light chains. *Proc. Natl. Acad. Sci. U.S.A*. 110, 6470–5

107 Kidd, B. a *et al*. (2014) Unifying immunology with informatics and multiscale biology. *Nat. Immunol*. 15, 118–127

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

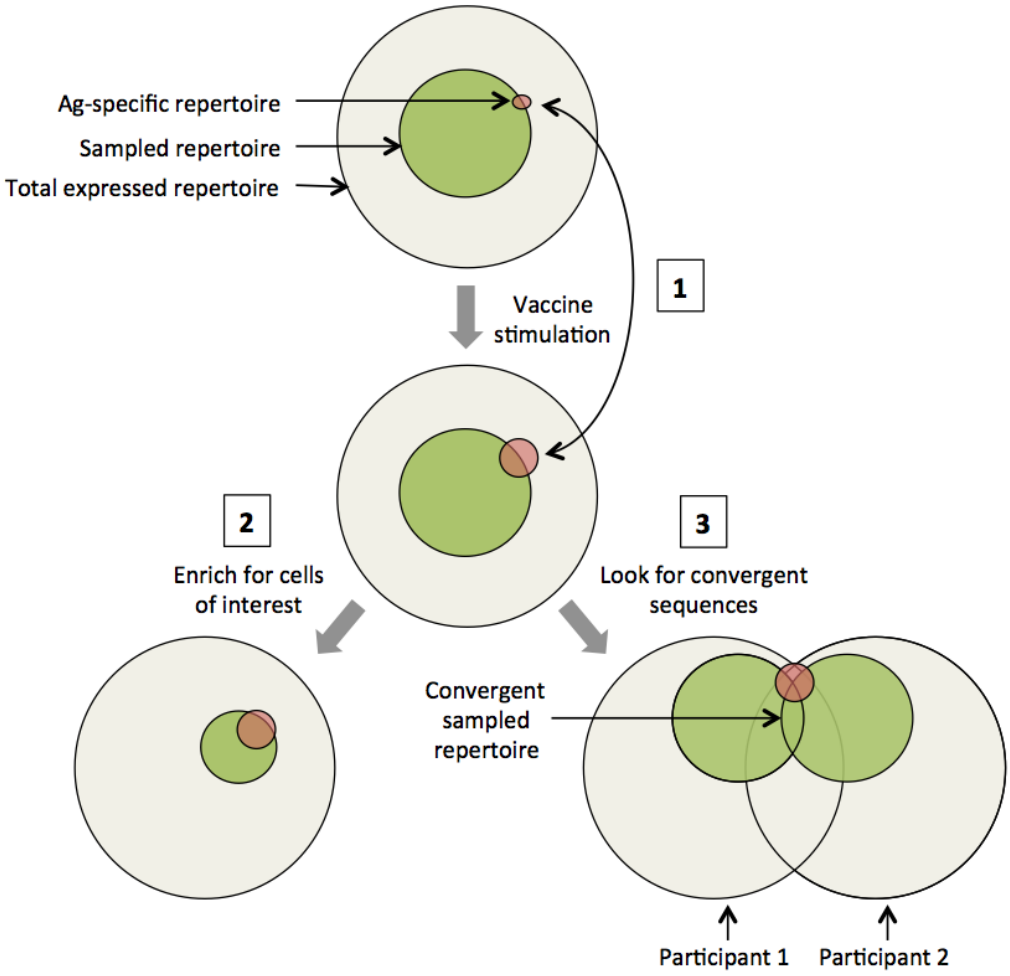
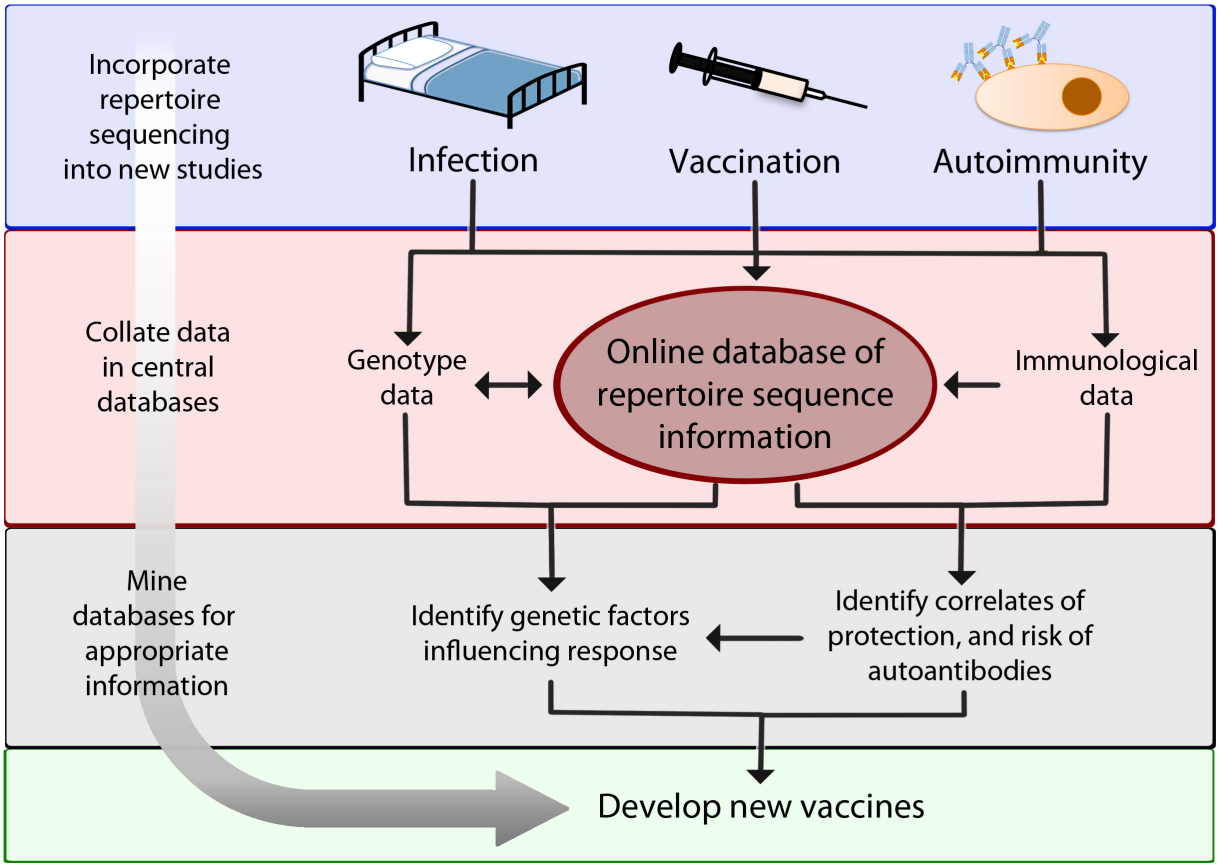


Figure 2. Proposed flow of information from repertoire sequencing projects to aid in the generation of data with practical uses. For such data to be suitable for meta-analysis, it is also important that a standardized sequencing protocol is used.



Box 1. Sequencing platforms used to study the antibody repertoire

The use of Sanger sequencing for characterization of single antibody genes from small numbers of cells is highly robust, as there is a high signal-to-noise ratio, low error rates, and a long read length. For characterizing whole repertoires of antibody genes from large numbers of cells, Sanger sequencing is too labor intensive, so NGS is used instead. NGS platforms allow multiple different DNA fragments to be sequenced simultaneously but generally have higher error rates, and shorter read lengths compared to Sanger sequencing. There are a number of different NGS platforms currently on the market, each utilizing different template amplification strategies, sequencing chemistries, and detection methods [99]. This leads to different types and degrees of sequence error (indels and substitutions) in the output data from each platform as well as different limitations such as read length, depth of sequencing (number of sequences that can be sequenced simultaneously), and cost.

High-throughput antibody repertoire sequencing was initially conducted using Roche's 454 sequencing technology, as this was the only platform with a sufficient read length (~500bp) to cover the entire V_H region [3,64]. Recent advances in the Illumina chemistry now permit 300 bp paired-end sequencing (http://www.illumina.com/systems/sequencing.ilmn, accessed April 2014) on their MiSeq platform, so more laboratories are now moving towards this technology instead [4,18]. The large output from Illumina sequencing means that 96 samples can be multiplexed in a single run, while still giving more than 100,000 reads for each sample, thus offering significant cost benefits over 454 sequencing. Recently, the Ion Torrent has also been used to sequence the antibody repertoire, where the emphasis was on speed of data acquisition rather than depth of sequencing [100]. By using whole blood rather than sorted B cells to characterize the repertoire, the authors were able to generate sequence data from a sample in a single day using the Ion Torrent. Illumina, 454 and Ion Torrent, were recently compared by Bolotin *et al*. for their application in profiling the T cell receptor repertoire [101]. Data from this study was used to estimate the error rate of each of these platforms, as shown below. Costs are derived from [102].

Platform	Mechanism	Error rate/type	Cost/Mb	Read Length	Depth

454	Pyrosequencing	1.4% of reads homopolymer associated indels	~$31	500 bp	10^5 - 10^6 reads
Illumina	Dye terminator sequencing	3.2% of reads random substitutions	~$0.5	2 x 300 bp	> 10^7 reads
Ion Torrent (314 chip)	Semiconductor sequencing	1.2% of reads homopolymer associated indels	~$22.5	400 bp	10^5 - 10^6 reads

NGS technology is rapidly advancing, with the major vendors frequently releasing updates to their systems, which reduce error and increase read length and depth. Costs are also constantly decreasing, making these technologies more accessible for use by laboratories with different specializations. As the technology develops, allowing deeper sequencing of the antibody repertoire, it is likely that more of the potential applications of this technology will be realized.

Table 1. Summary of publications where the antibody repertoire has been investigated after vaccination.

Vaccine	Cells used	Methodology	Key findings	Ref.
Publications where the antibody repertoire has been investigated using a low-resolution approach				
Influenza				
TIV	IgG plasmablasts 7 days after vaccination	Single-cell V_H and V_L PCR followed by Sanger sequencing	• Produced 50 mAbs from 14 individuals, against three different influenza strains. • Showed that the influenza-specific antibody response is pauci-clonal, with extensive intraclonal diversification of the influenza-specific lineages due to somatic hypermutation.	[20]
TIV	IgA- IgM- plasmablasts 7 days after vaccination	Single-cell V_H and V_L PCR (including unique barcodes for each cell), followed by 454 sequencing	• Analyzed 384-768 sequences from each of three individuals. • Cloned 8 mAbs from large clonal sequence families, and 12 mAbs from singleton sequences. 75% of these mAbs bound and neutralized influenza, but those from large clonal families were most effective. Three of these mAbs bound more effectively to HA from previous influenza seasons than the HA used in the vaccine	[21]
Tetanus				
TT	Plasmablasts 6 days after three consecutive vaccinations, separated by at least 1.5 years	Single-cell V_H:V_L linkage PCR. DNA cloning into *E. coli* and Sanger sequencing of TT-positive clones	• The number of somatic hypermutations was similar between individuals, and did not increase through the study, indicating the limit had already been reached through previous routine vaccinations.	[22]
TT	TT-specific plasmablasts 7 days, and TT-specific memory B cells 9 days after vaccination	Single-cell isotype-specific V_H and V_L PCR followed by Sanger sequencing	• VDJ segment usage, CDR3 length and distribution of somatic hypermutations were similar among TT-specific plasmablasts and memory cells.	[23]
Streptococcus pneumonia				
PS (23 valent)	IgG plasmablasts 7 days after vaccination	Single-cell V_H and V_L PCR followed by Sanger sequencing	• Cloned 137 mAbs against 19 of the 23 vaccine serotypes from 4 individuals. Most antibodies were serotype-specific, but 12% cross-reacted with 2 or more serotypes.	[77]
PS (23 valent)	PPS4 or PPS14 specific B cells 6 weeks after vaccination	Single-cell culture followed by V_H PCR and Sanger sequencing of pooled, cultured cells	• Analyzed over 1300 sequences from 40 individuals. Showed significant differences in the antibody repertoire in young and elderly participants. Elderly participants had a more clonal repertoire with less somatic hypermutations.	[97]
Haemophilus influenza B				
PS or PS-DT or OC-CRM	Lymphocytes 7 days after vaccination	Fusion of lymphocytes to mouse myeloma cells followed by culture, V_H and V_L PCR and Sanger sequencing	• Sequenced 15 cell lines from 10 individuals, secreting antibody against Hib PS. Showed that these mAbs had undergone somatic hypermutation and demonstrated clonality (all usingV_H3) of the antibody repertoire after vaccination.	[24]
PS-DT	Lymphocytes 7 days after vaccination	Fusion of lymphocytes to mouse myeloma cells followed by culture, V_H and V_L PCR and Sanger sequencing	• Sequenced 4 cell lines from 4 individuals, secreting antibody against Hib PS. All used V_H3, but were composed of two different configurations of D and J segment, and V_k genes, indicating that the 4 cell lines were from 2 different lineages.	[61]
PS or PS-DT	Lymphocytes 7 days after vaccination	Fusion of lymphocytes to mouse myeloma cells followed by culture, V_H PCR and Sanger sequencing	• Sequenced 5 cell lines from 4 individuals, secreting antibody against Hib PS. Demonstrated that they all used V_H3, but different D and J segments.	[55]
Publications where the antibody repertoire has been investigated using a high-resolution approach and next-generation sequencing				
Influenza				
TIV	Memory B cells 14 days after vaccination (one participant)	High-throughput single-cell V_H:V_L linkage PCR and 2x250 bp Illumina sequencing	• Validated accuracy of V_H:V_L pairings identified using their high-throughput method. • Identified 240 putatively influenza-specific CDR-H3:CDR-L3 pairings.	[18]
TIV or LAIV	PBMCs on the day of vaccination, and on day 7 and 28 after two vaccinations given a year apart (28	V_H-specific reverse transcription and 2nd strand synthesis, with the incorporation of random nucleotide tags, followed by PCR.	• Showed different repertoire dynamics after TIV and LAIV vaccination. TIV induced a stronger response, with more abundant IgG lineages than LAIV. • Found shared antibody sequences on day 7 after two TIV vaccinations. Hypothesized that these lineages are present after the second vaccination due to memory B cell recall. Suggested that this method could be used to identify	[4]

	participants)	Custom 100 x 120 bp Illumina sequencing protocol, and clustering of sequences based on tags to account for PCR bias and error	cross-specific antibodies.	
TIV or LAIV	Naïve B cells, and plasmablasts, on the day of vaccination and on days 7 or 8 and day 28 after vaccination (17 participants, three age groups)	V_H-specific multiplex PCR, and Roche 454 sequencing	• The influenza-specific antibody repertoire in older individuals was more clonal, and had a greater mutational load than the repertoire in younger individuals. • In twins, the mutational load of the IgM repertoire was similar, but diverged for the IgG repertoire, indicating that the naïve repertoire is more influenced by individual genetics, but the memory repertoire is more influenced by environmental stimuli.	[16]
TIV	PBMC's on 18 timepoints around two vaccinations given a year apart (one participant), or 10 timepoints around one vaccination (two participants).	V_H-specific multiplex PCR, and Roche 454 sequencing	• V and J segment usage differs between individuals, and is conserved within individuals over time. • Clustering sequences into clonal groups based on CDR3 identity showed clonal expansion and contraction in response to the vaccine with different participants exhibiting different dynamics. • There are a small number of highly mutated, persistent clones found within all individuals, potentially corresponding to long-lived B cell memory or indicative of chronic infection.	[98]
Tetanus				
TT	Plasmablasts 7 days after vaccination (one participant)	High-throughput single-cell V_H:V_L linkage PCR and 2x250 bp Illumina sequencing	• Identified 86 putatively TT-specific CDR-H3:CDR-L3 pairings. Cloning ten of these into HEK293K cells followed by competitive ELISA of the antibodies produced showed them to be TT-specific.	[18]
TT	Bulk plasmablasts, memory B cells, and antigen-specific plasmablasts 7 days and 3 months after vaccination (two participants)	V_H and V_L-specific multiplex PCR, and Roche 454 sequencing. Half the day 7 plasmablasts were also used for high-throughput single-cell V_H:V_L linkage PCR. Also conducted proteomic analysis of TT-specific serum antibodies	• Analyzed the serum antibody repertoire by using the V_H sequence database to interpret results from high-resolution liquid chromatography tandom mass spectrometry of the serum antibodies. • 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	[68]
Co-administered				
TIV & PPV23	PBMCs on the day of vaccination and on day 7 and 28 after vaccination (14 participants, two age groups)	Semi-nested isotype and V_H-specific multiplex PCR, and Roche 454 sequencing	• The repertoire changed at day 7 post-vaccination, but returns to a baseline-like state after 28 days. • Comparing the repertoire in young and elderly individuals after vaccination indicated clonal expansion is delayed in older individuals. Showed age-related differences in IgA and IgM repertoire dynamics.	[13]

Abbreviations: TT, tetanus toxoid; PS, polysaccharide; OC, oligosaccharide; DT, diphtheria toxoid; TIV, trivalent inactivated influenza vaccine; LAIV, live attenuated influenza vaccine; PPV23, 23-valent pneumococcal polysaccharide vaccine.