Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Immunology

Deep sequencing and human antibody repertoire analysis

Scott D Boyd[1] and James E Crowe Jr.[2]

In the past decade, high-throughput DNA sequencing (HTS) methods and improved approaches for isolating antigen-specific B cells and their antibody genes have been applied in many areas of human immunology. This work has greatly increased our understanding of human antibody repertoires and the specific clones responsible for protective immunity or immune-mediated pathogenesis. Although the principles underlying selection of individual B cell clones in the intact immune system are still under investigation, the combination of more powerful genetic tracking of antibody lineage development and functional testing of the encoded proteins promises to transform therapeutic antibody discovery and optimization. Here, we highlight recent advances in this fast-moving field.

Addresses
[1] Department of Pathology, Stanford University, Stanford, CA 94305, United States
[2] Vanderbilt Vaccine Center, Vanderbilt University Medical Center, Nashville, TN 37232-0417, United States

Corresponding authors: Boyd, Scott D (sboyd1@stanford.edu) and Crowe, James E (james.crowe@vanderbilt.edu)

Current Opinion in Immunology 2016, **40**:103–109

This review comes from a themed issue on **Special section: New concepts in antibody therapeutics**

Edited by **Janine Schuurman** and **Paul WHI Parren**

For a complete overview see the Issue and the Editorial

Available online 8th April 2016

http://dx.doi.org/10.1016/j.coi.2016.03.008

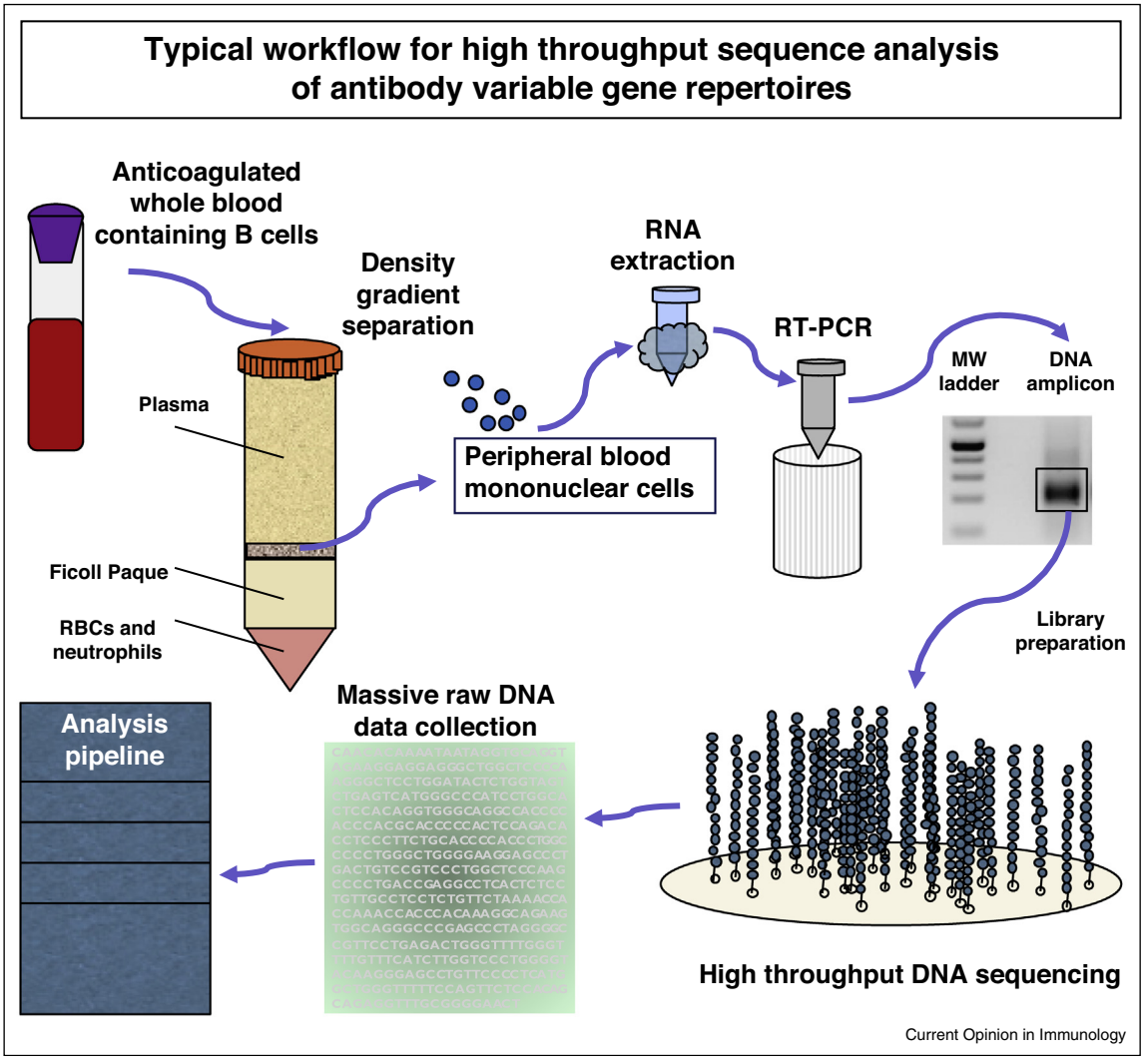
0952-7915/Published by Elsevier Ltd.

Sequence analysis techniques for antibody variable genes

Conventional technologies for cloning and sequencing antibody variable genes have been available since 1989 [1]. Usually, messenger or total RNA from some type of human tissue containing B cells (usually blood or bone marrow) is extracted, reverse-transcribed and then amplified by PCR. Most investigators study RNA, since the objective is to study transcribed antibody genes encoding expressed proteins. However, there are some advantages to using genomic DNA encoding recombined antibody gene segments as the template for analysis. Use of genomic DNA template eliminates the effect of transcript copy number on the composition of the resulting amplicons populations; the antibody mRNA copy number in plasma cells is extraordinarily high, while that in resting memory B cells appears to be on the order of hundreds. Multiplex panels of oligonucleotide primers designed to amplify most or all antibody variable genes have improved over time, as additional large-scale sequencing efforts have progressed. PCR can introduce amplification biases during multi-cycle amplification. This bias has been addressed by some investigators by using a molecular barcoding strategy at the time of the reverse transcription step, to identify individual transcripts in resulting sequence repertoires. This strategy does not, however, reveal the number of cells that produced the transcripts. Alternatively, some investigators have resorted to a 5′ RACE (rapid amplification of cDNA ends) procedure that reduces amplification bias, but tends to result in less efficient capture of individual transcripts and incomplete representation of the diversity within a sample. It is possible, but complex and time-consuming, to develop DNA standards that facilitate optimization of PCR protocols for semi-linear amplification. The capability for amplification of large numbers of diverse antibody genes into amplicons facilitated the early development of phage display antibody libraries, which formed the cornerstone of human antibody discovery efforts in the early 1990s. The sequence of individual clones derived from such libraries conventionally was determined by Sanger sequence analysis of individually cloned DNAs.

The human genome project spurred the development of HTS technologies and protocols have been developed for sequence analysis of immune repertoire gene amplicons (Figure 1) on most of the instrument platforms used for genomic studies [2]. The 454 Life Sciences technology (now owned by Roche), used a large-scale parallel pyrosequencing system, which is essentially detection of pyrophosphate release on nucleotide incorporation during synthesis. Compared to competing technologies, the technique allowed reads with good length for amplicon sequencing, but also had a relatively high cost, lower throughput, and increased frequency of indels, which can also occur in natural antibody sequences. Illumina has marketed sequencing using reversible-terminator technology, with instruments that allow millions (MiSeq) or billions (HiSeq) of amplicon sequences to be acquired in single experiments. The technique uses paired end sequencing of both ends of a fragment to generate alignable sequence data. Using a (2×250) or (2×300) base pair sequence technique, typically one can stitch the two reverse-orientation reads into a single contig that contains

Figure 1



Typical workflow for high throughput sequence analysis of antibody variable gene repertoires. Usually, cells are separated from the starting tissue (often peripheral blood, but alternatively another tissue source containing B cells). RNA is extracted and reverse transcribed (or alternatively, genomic DNA can be used). Multiplex PCR for all antibody variable genes is performed. The antibody gene amplicon is isolated, libraries prepared, then subjected to high throughput sequencing. The resulting sequence sets must be processed and then analyzed.

the entire coding sequence of the variable portion of the heavy or the light chain (which are usually in the 300–500 base pair range of length, depending on the primer sets used for amplification). Sequence analysis in HiSeq experiments using a (2 × 150) base pair approach can achieve extraordinary depth, but in this case the length of the sequences typically allows analysis only of the V–D–J or V–J junction, which encodes the CDR3 region of highest variability. Such partial sequences cannot be cloned and expressed in their natural configuration for validation experiments, however. Pacific Biosciences has developed a sequencing technology, based on a zero-mode waveguide (essentially an optical waveguide that guides light energy into a very small volume compared to the wavelength of the light used). This approach has lower single read fidelity but achieves high fidelity in amplicon sequencing, because of the high number of repetitive reads possible. The limitation in this approach in the past has been restricted depth of sequencing.

The HTS approaches described above collect heavy or light chain repertoires separately, losing information on the pairing of heavy and light chains in naturally-occurring antibodies. Heavy chain pairing with diverse light chains is known to be quite promiscuous, based on extensive data from phage display library experiments. Therefore, it is challenging to select with any confidence pairs of heavy and light chain genes from such separate

repertoire databases for recombinant expression of immunoglobulins for validation experiments. One approach is to use statistical methods, basically pairing the most common heavy chains with the most common light chains in sequence sets, and measuring binding of the resulting panels of antibodies to an antigen. Another is to isolate the naturally-paired heavy and light chains from one member of a B cell clone, using single-cell techniques and then pair one of those chains with somatic variants of the other chain identified by unpaired repertoire sequencing from the same individual, to create antibodies that are highly related to the natural antibody. Linking of the heavy and light chain genes from single cells is preferred for repertoire studies, but is challenging at large scale. It is possible to physically distribute a relatively large number of single cells as single entities into individual wells for RT-PCR amplification using a flow cytometer or a microfluidics device [3,4], but ultimately there are moderate physical and expense limits on this type of single-cell approach. Newer approaches use microfluidic devices to introduce single cells into individual water-in-oil emulsion droplets containing RT-PCR reagents that amplify and physically link the heavy and light chain amplicons [5•] or achieve the identical RNA barcoding of heavy and light chain genes from the same vesicle [6•]. The longer, physically-linked [heavy + light chain] amplicons are too long for currently available [2 × 300] paired-end Illumina sequencing, so in that approach, one is required to use PacBio sequencing [7]. Alternatively, the sequencing submission is divided into three different amplicons generated by PCR using linker and heavy or light chain primers or the full length amplicon.

Analysis of large-scale antibody gene repertoire sets is challenging at the present time. Typically, small sets of sequences have been analyzed using the publicly available international ImMunoGeneTics information system (IMGT; http://www.imgt.org) website. The IMGT group also has developed a High VQuest search capability that allows submission of moderate-sized HTS sequence sets [8]. Currently, HTS experiments are beginning to far exceed this scale, however. Therefore, many groups have been developing, in parallel, proprietary or open-source tools for massive-scale analysis that can be deployed either on individual servers, or on supercomputing clusters [9,10]. Furthermore, determination of the optimal bioinformatics techniques for assigning variable, diversity and joining genes, identifying somatic mutations, determining the limits and nature of the V–D, D–J and V–J junctions, and other features of antibody sequences is still an area of active study. Benchmarking experiments reveal that the currently available techniques agree in assignments in only about 90% of the cases. We can expect a lot of development in this area in the next several years.

Human antibody repertoires

HTS analysis of human antibody gene repertoires has clarified features of the formation and selection of antibodies in the B cells of healthy subjects exposed to various vaccines or infections, as well as highlighting perturbations in antibody repertoires associated with conditions such as aging, immunodeficiency, autoimmunity, and allergy. The initial description of somatic gene segment rearrangement that gives rise to diverse antibody repertoires in vertebrates was published four decades ago [11], but it is only in recent years that measurement methods for DNA sequences have become able to capture the data from the hundreds of thousands to millions of B cells in a typical clinical sample of human blood. Analysis of the sequences of antibody gene rearrangements provides several distinct kinds of data, each of which contribute to understanding the clonal history and function of the B cell expressing the antibody. The underlying germline V, D and J gene segments of antibody heavy chains, and the V and J segments of light chains, as well as the imprecise junctions between segments that encode the complementarity-determining region 3 (CDR3) that is the most diverse peptide loop in each antibody chain, can be inferred from collections of antibody gene rearrangements from an individual. Information about potential antibody function is obtained by identifying the heavy chain constant region isotype used by each B cell (for humans, IgM, IgD, IgG subtypes 1 through 4, IgA subtypes 1 and 2, and IgE), as these each confer distinct functional properties to the antibody by mediating binding to constant region receptors expressed by other cells in the immune system, or by activating the complement system of soluble anti-microbial factors. Evidence of prior antigen-driven stimulation of the B cell expressing the antibody can be detected in the form of somatic mutations in the DNA of the V(D)J rearrangement; these are usually introduced by the activation-induced cytidine deaminase enzyme (AID) and accompanying DNA repair pathways that are activated in B cells in the germinal centers of secondary lymphoid tissues. Here, we will summarize recent insights from human antibody repertoire sequencing, focusing on: (a) the diversity of germline gene segments and alleles in different population groups, (b) formation of the V(D)J junctions making up CDR3, and the effect on CDR3 of subsequent selection of B cells, and (c) the size, somatic mutation states and isotype expression relationships between B cells that are clonally related in an individual, that is, the daughter cells that arise from cell division of an initial naïve B cell. These features of antibody gene repertoires, prior to obtaining direct experimental evidence of antigen-specific binding of individual B cells, provide information about the antibody repertoires from which specific clones are selected, and can identify clones of interest that may not be detected in a given experimental design.

Germline antibody gene segment repertoires

Mapping of the extreme allelic diversity of key molecules in the adaptive immune system, such as the human leukocyte antigen (HLA) class I and II proteins that present peptides for detection by the T cell receptors,

has been driven by the medical necessity of typing organs and tissues for transplantation. It has been less clear to what extent allelic diversity in the gene segments of antibody heavy and light chains may contribute to B cell function and antibody diversity. Initial studies using HTS of antibody gene rearrangements reinforced and extended the findings from earlier literature reports that there is extensive allelic variation and even copy number variation of antibody gene segments in the germline genome [12]. For example, deletions in the heavy chain D segment locus that remove several contiguous segments are frequent in North American populations, such that some individuals completely lack common D segments in their repertoires [13]. Another pair of studies used the unique genetic resource of an effectively haploid cell line derived from a human hydatidiform mole to be able to cleanly resequence the human immunoglobulin heavy chain and light chain loci, and identified additional gene segments and copy number variation [14,15]. A very recent study of 28 HIV-infected South African women from the Centre for the AIDS Programme of Research in South Africa (CAPRISA) highlights the limited coverage of human germline immunoglobulin gene segment genetic diversity in current reference sequence databases, particularly when non-European populations are studied; these investigators identified 85 entirely novel immunoglobulin heavy chain variable region gene (IGHV) alleles, and 38 alleles that were only present in rearranged antibody gene sequences, rather than being reported as germline gene segments in public databases [16•]. The ability of the HIV-infected individuals in this study to elicit broadly-neutralizing antibodies against HIV was not clearly associated with their germline antibody heavy chain locus genotypes. The more general question of whether allelic variation in human Ig repertoires ever contributes to protective immunity will require further investigation of a variety of infectious diseases, with cohorts of genotyped subjects.

V(D)J rearrangement, CDR-H3 formation, and selection

HTS analysis of rearranged immunoglobulin genes in human subjects indicates that both the germline genome and the environment play important roles in shaping the antibody receptor populations that an individual uses to recognize antigens. Studies in identical twins show that members of a twin pair have highly similar frequencies of V, D and J gene segment usage in their peripheral blood B cells, particularly in B cells expressing unmutated antibodies, that likely have not previously encountered antigen [17]. These similarities in twins may be related to the strengths of the recombination signal sequences (RSS) that direct segment recombination by the recombination-activating gene (RAG) protein complex, but could also be affected by chromatin accessibility differences. Twins, compared to unrelated individuals, also showed more similar lengths of CDR-H3 regions formed by VDJ gene segment rearrangement combined with exonuclease chew-back of segment ends and addition of non-templated bases at the V–D and D–J junctions, suggesting that the activity of the various components in the rearrangement machinery are influenced by the germline genome [17].

The effects of cellular selection on the antibody gene repertoires seen at different stages of B cell development and differentiation are readily seen from HTS data sets. Amplification of rearranged antibody genes from genomic DNA templates permits analysis of unproductive rearrangements, which have not been subjected to selection on the basis of the encoded proteins, and can therefore give a measure of the features of the pre-selected repertoire of antibody gene rearrangements. Such out-of-frame sequences in healthy subjects usually have longer CDR-H3 regions than the productive, but unmutated antibody gene rearrangements seen in naïve B cells in the peripheral blood. This length difference indicates that there is a selection against B cells with long CDR-H3 sequences after precursor B-cell development in the bone marrow, consistent with earlier data from Sanger sequencing analysis of antibody genes in human early precursor B cells, and increased self-reactivity of the antibodies encoded by such B cells [18,19]. A second step of selection against antibodies with long CDR-H3 regions appears to occur during the process of somatic mutation and selection that gives rise to affinity maturation, since the CDR-H3s in mutated antibody genes from memory B cells are even shorter than those found in naïve B cells [19,20].

Several analysis approaches can be applied to antibody HTS repertoire data from peripheral blood samples of human subjects who have recently been vaccinated or received other antigenic stimuli, to highlight B cell clones that are likely to be antigen-specific. If replicate samples from pre-vaccination and post-vaccination time points are available, either in the form of RNA isolated from subsamples of the blood at each time point, or by taking aliquots of genomic DNA template from each time point (since each B cell only contains one productively rearranged heavy chain gene copy in its genome), then analysis of sequencing libraries made from each replicate sample can be used to identify clonally-expanded B cells, because the antibody gene rearrangements from such clones will appear in multiple replicate libraries [21–23]. In contrast, B-cell populations from baseline or pre-vaccination samples usually contain few clones that are expanded enough to be detected in this manner. Another signal that can provide evidence for antigen specificity of a B-cell clone after vaccination is sequence similarity or 'convergence' of antibody gene rearrangements to those seen in other individuals who have received the same antigen exposure. Early examples of convergent antibody sequences were reported for *H. influenzae* and pneumococcal vaccinations [24,25], but several recent studies have demonstrated that the presence of specific convergent antibodies in humans

exposed to a variety of infectious agents or vaccines are the rule, rather than the exception. For example, subjects with primary or secondary dengue virus infection demonstrate many convergent antibodies [22] and sequence similarities are detected between highly-selected antibodies with neutralizing breadth against HIV [26]. Subjects vaccinated against influenza virus show convergent antibody rearrangements [21], as do individuals vaccinated with tetanus toxoid and group C meningococcal polysaccharides [27•]. Other recent approaches to identify antibody gene rearrangements of potential biological interest from HTS data have made use of anatomic information and carefully chosen samples, such as samples from the central nervous system of patients with multiple sclerosis, that may be enriched for B cells involved in the autoimmune disease [28].

Isotype and mutational relationships within B cell clones

The antibody gene rearrangements expressed by clonally-related B cells derived from a single original naïve B cell co-expressing IgM and IgD can be recognized based on the shared V(D)J sequences of their heavy and light chains, even if some members of the clone have undergone extensive somatic mutation and switching of their heavy chain constant regions to other isotypes. HTS data sets derived from millions of individual B cells in individual subjects now make it possible to probe the relationships between members of B cell clones in new depth. One recent use of this methodology has been to study antibody rearrangements expressed as IgE, the antibody isotype that confers binding to high-affinity Fc receptors on mast cells and basophils, enabling these effector cells to become activated in an allergen-specific manner in allergic disease. HTS of allergic patient and healthy control indicated that IgE antibody repertoires do have somatic mutations, indicating that members of the B cell clone, either before or after switching to IgE, have activated their mutational mechanisms, and may have undergone affinity maturation to increase their binding affinity to antigen. Comparison of IgE isotype antibodies to members of the same clone expressing other isotypes indicated that IgG1 antibodies shared the largest number of somatic mutation positions with the IgE clone members, and provided some evidence that switching to IgE may occur from a prior IgG-expressing precursor within the clone [29]. A recent study of clonal relationships between memory B cells in healthy human peripheral blood identified extensive clonal overlap between IgM-expressing memory B cells and those expressing IgG, indicating that a large number of memory B cell clones specific for particular pathogens will contain members expressing a variety of isotypes [30].

Somatic mutations provide a measure of the amount of antigen stimulation a B cell clone has received. Antibody genes in the B cells of umbilical cord blood samples of newborn infants show a predominance of unmutated sequences, with some limited evidence of somatic mutation in rare cells [31]. The initial years of life bring infectious disease exposures and other antigenic encounters that result in the development of differentiated memory B cells and plasma cells that have accumulated somatic mutations. The rate at which somatic mutations are accumulated in human life, and the factors that affect this, are being elucidated with HTS analysis of B cell repertoires in humans of various ages. A study of identical twins found that twin-pair members, in the sixth decade of life had more similar levels of somatic mutation in their antibody repertoires than unrelated individuals, and with different levels identified for different twin pairs, suggesting that the rate of accumulation of somatic mutations may be under genetic control that differs between individuals [17]. A caveat to this result, however, is that no measures of the similarity of the environments and infectious disease exposures for different individuals were available, so additional research on separated twins compared to those who have lived in similar settings may be required to separate these variables. Is there an optimal level of somatic mutation in human antibodies needed to achieve high-affinity binding without deleterious mutations that affect antibody stability? Analysis of mutation levels in the isotype-switched mutated B cells in healthy young adults (20–35 years old) compared to healthy older adults (65–91 years old) clearly indicates that mutation levels show only modest increases over many decades of adult life, with most cells having IGHV gene mutation levels of 5–10% [20]. Potential explanations for this observation include decreased ability to carry out somatic mutation with age, or selection against B cells expressing more highly mutated antibodies as a result of decreased antibody protein stability or increased autoreactivity. Despite this general boundary of mutation levels seen in memory B cells in healthy subjects, it is clear that pathogenic circumstances, such as HIV infection, chronic viral antigenic stimulation and a dysregulated immune system, can give rise to much higher levels of somatic mutation, including mutation in the framework regions, in individual B cell clones, as seen in some broadly-neutralizing antibody lineages [32].

Summary

HTS methods and improved antigen-specific B cell isolation and single-cell antibody cloning approaches have added tremendous analytical power to recent research in human immune responses. Arguably, no other cellular system in the body retains such an informative genetic record of the history of individual cell clones, as is the case for the somatic mutations that enable improved antigen binding by members of a B cell clone. Most of the studies reviewed here have aimed to uncover the basic features of human B cell repertoires involved in effective responses to pathogens and vaccination. However, the ability to monitor the selection of individual members of hundreds to thousands of B cell clones as they improve their

binding to antigen, either *in vivo*, or in artificial selection systems, holds great promise for more efficient and more thorough development of antibodies for a wide range of therapeutic purposes. Increasing integration of methods for selecting individual antigen-specific B cells, and monitoring the evolution of many additional members of those clones under antigen-driven selection, should yield both new basic understanding of humoral immunity, and practical results for the benefit of patients.

Acknowledgements

The authors acknowledge the following grant support: NIH grant U19 AI117905 (JEC), NIH contract HHSN2720100007C (JEC), and NIH grants U19 AI09525801, UM1AI100663, and U01AI104342 (SDB).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Larrick JW, Danielsson L, Brenner CA, Abrahamson M, Fry KE, Borrebaeck CA: **Rapid cloning of rearranged immunoglobulin genes from human hybridoma cells using mixed primers and the polymerase chain reaction**. *Biochem Biophys Res Commun* 1989, **160**:1250-1256.

2. Boyd SD: **Diagnostic applications of high-throughput DNA sequencing**. *Annu Rev Pathol* 2013, **8**:381-410.

3. DeKosky BJ, Ippolito GC, Deschner RP, Lavinder JJ, Wine Y, Rawlings BM, Varadarajan N, Giesecke C, Dorner T, Andrews SF *et al.*: **High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire**. *Nat Biotechnol* 2013, **31**:166-169.

4. Love JC, Ronan JL, Grotenbreg GM, van der Veen AG, Ploegh HL: **A microengraving method for rapid selection of single cells producing antigen-specific antibodies**. *Nat Biotechnol* 2006, **24**:703-707.

5. • McDaniel JR, DeKosky BJ, Tanno H, Ellington AD, Georgiou G: **Ultra-high-throughput sequencing of the immune receptor repertoire from millions of lymphocytes**. *Nat Protoc* 2016, **11**:429-442.

This paper gives detailed protocols for building hardware and performing experiments that physically link naturally paired heavy and light chain antibody variable gene amplicons for large-scale repertoire studies.

6. • Briggs AW, Vigneault, F. Single cell bar-coding for antibody discovery. Abvitro, Inc., assignee. Patent WO/2014/144495. 18 September 2014. This patent describes methods for immune repertoire sequencing using single cell barcoding, enabling next generation sequencing data sets that retain the natural heavy and light chain pairing.

7. DeKosky BJ, Kojima T, Rodin A, Charab W, Ippolito GC, Ellington AD, Georgiou G: **In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire**. *Nat Med* 2015, **21**:86-91.

8. Alamyar E, Duroux P, Lefranc MP, Giudicelli V: **IMGT((R)) tools for the nucleotide analysis of immunoglobulin (IG) and T cell receptor (TR) V–(D)–J repertoires, polymorphisms, and IG mutations: IMGT/V-QUEST and IMGT/HighV-QUEST for NGS**. *Methods Mol Biol* 2012, **882**:569-604.

9. Gupta NT, Vander Heiden JA, Uduman M, Gadala-Maria D, Yaari G, Kleinstein SH: **Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data**. *Bioinformatics* 2015, **31**:3356-3358.

10. Vander Heiden JA, Yaari G, Uduman M, Stern JN, O'Connor KC, Hafler DA, Vigneault F, Kleinstein SH: **pRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires**. *Bioinformatics* 2014, **30**:1930-1932.

11. Hozumi N, Tonegawa S: **Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions**. *Proc Natl Acad Sci U S A* 1976, **73**:3628-3632.

12. Boyd SD, Gaeta BA, Jackson KJ, Fire AZ, Marshall EL, Merker JD, Maniar JM, Zhang LN, Sahaf B, Jones CD *et al.*: **Individual variation in the germline Ig gene repertoire inferred from variable region gene rearrangements**. *J Immunol* 2010, **184**:6986-6992.

13. Kidd MJ, Chen Z, Wang Y, Jackson KJ, Zhang L, Boyd SD, Fire AZ, Tanaka MM, Gaeta BA, Collins AM: **The inference of phased haplotypes for the immunoglobulin H chain V region gene loci by analysis of VDJ gene rearrangements**. *J Immunol* 2012, **188**:1333-1340.

14. Watson CT, Steinberg KM, Huddleston J, Warren RL, Malig M, Schein J, Willsey AJ, Joy JB, Scott JK, Graves TA *et al.*: **Complete haplotype sequence of the human immunoglobulin heavy-chain variable, diversity, and joining genes and characterization of allelic and copy-number variation**. *Am J Hum Genet* 2013, **92**:530-546.

15. Watson CT, Steinberg KM, Graves TA, Warren RL, Malig M, Schein J, Wilson RK, Holt RA, Eichler EE, Breden F: **Sequencing of the human IG light chain loci from a hydatidiform mole BAC library reveals locus-specific signatures of genetic diversity**. *Genes Immun* 2015, **16**:24-34.

16. • Scheepers C, Shrestha RK, Lambson BE, Jackson KJ, Wright IA, Naicker D, Goosen M, Berrie L, Ismail A, Garrett N *et al.*: **Ability to develop broadly neutralizing HIV-1 antibodies is not restricted by the germline Ig gene repertoire**. *J Immunol* 2015, **194**:4371-4378.

This paper provides evidence of the tremendous diversity of germline immunoglobulin gene segment alleles in human populations that are incompletely represented in current reference sequence archives.

17. Wang C, Liu Y, Cavanagh MM, Le Saux S, Qi Q, Roskin KM, Looney TJ, Lee JY, Dixit V, Dekker CL *et al.*: **B-cell repertoire responses to varicella-zoster vaccination in human identical twins**. *Proc Natl Acad Sci U S A* 2015, **112**:500-505.

18. Wardemann H, Yurasov S, Schaefer A, Young JW, Meffre E, Nussenzweig MC: **Predominant autoantibody production by early human B cell precursors**. *Science* 2003, **301**:1374-1377.

19. Roskin KM, Simchoni N, Liu Y, Lee JY, Seo K, Hoh RA, Pham T, Park JH, Furman D, Dekker CL *et al.*: **IgH sequences in common variable immune deficiency reveal altered B cell development and selection**. *Sci Transl Med* 2015, **7**:302ra135.

20. Wang C, Liu Y, Xu LT, Jackson KJ, Roskin KM, Pham TD, Laserson J, Marshall EL, Seo K, Lee JY *et al.*: **Effects of aging, cytomegalovirus infection, and EBV infection on human B cell repertoires**. *J Immunol* 2014, **192**:603-611.

21. Jackson KJ, Liu Y, Roskin KM, Glanville J, Hoh RA, Seo K, Marshall EL, Gurley TC, Moody MA, Haynes BF *et al.*: **Human responses to influenza vaccination show seroconversion signatures and convergent antibody rearrangements**. *Cell Host Microbe* 2014, **16**:105-114.

22. Parameswaran P, Liu Y, Roskin KM, Jackson KK, Dixit VP, Lee JY, Artiles KL, Zompi S, Vargas MJ, Simen BB *et al.*: **Convergent antibody signatures in human dengue**. *Cell Host Microbe* 2013, **13**:691-700.

23. Boyd SD, Marshall EL, Merker JD, Maniar JM, Zhang LN, Sahaf B, Jones CD, Simen BB, Hanczaruk B, Nguyen KD *et al.*: **Measurement and clinical monitoring of human lymphocyte clonality by massively parallel VDJ pyrosequencing**. *Sci Transl Med* 2009, **1**:12ra23.

24. Lucas AH, McLean GR, Reason DC, O'Connor AP, Felton MC, Moulton KD: **Molecular ontogeny of the human antibody repertoire to the Haemophilus influenzae type B polysaccharide: expression of canonical variable regions and their variants in vaccinated infants**. *Clin Immunol* 2003, **108**:119-127.

25. Zhou J, Lottenbach KR, Barenkamp SJ, Lucas AH, Reason DC: **Recurrent variable region gene usage and somatic mutation in the human antibody response to the capsular polysaccharide of *Streptococcus pneumoniae* type 23F**. *Infect Immun* 2002, **70**:4083-4091.

26. Wu X, Zhou T, Zhu J, Zhang B, Georgiev I, Wang C, Chen X, Longo NS, Louder M, McKee K *et al.*: **Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing**. *Science* 2011, **333**:1593-1602.

27. • Truck J, Ramasamy MN, Galson JD, Rance R, Parkhill J, Lunter G, Pollard AJ, Kelly DF: **Identification of antigen-specific B cell receptor sequences using public repertoire analysis**. *J Immunol* 2015, **194**:252-261.

These authors identify convergent, or public, antibody heavy chain sequences selected by the same infectious agents in different individuals, for several bacterial pathogens. With sufficient depth of antibody repertoire analysis in enough individuals, it seems likely that convergent antibody sequences will be identified for most pathogens, providing a host sequence signature of the infectious agent.

28. Stern JN, Yaari G, Vander Heiden JA, Church G, Donahue WF, Hintzen RQ, Huttner AJ, Laman JD, Nagra RM, Nylander A *et al.*: **B cells populating the multiple sclerosis brain mature in the draining cervical lymph nodes**. *Sci Transl Med* 2014, **6**:248ra107.

29. Looney TJ, Lee JY, Roskin KM, Hoh RA, King J, Glanville J, Liu Y, Pham TD, Dekker CL, Davis MM *et al.*: **Human B-cell isotype switching origins of IgE**. *J Allergy Clin Immunol* 2016, **137** 579–586.e7.

30. Budeus B, Schweigle de Reynoso S, Przekopowitz M, Hoffmann D, Seifert M, Kuppers R: **Complexity of the human memory B-cell compartment is determined by the versatility of clonal diversification in germinal centers**. *Proc Natl Acad Sci U S A* 2015, **112**:E5281-E5289.

31. Rechavi E, Lev A, Lee YN, Simon AJ, Yinon Y, Lipitz S, Amariglio N, Weisz B, Notarangelo LD, Somech R: **Timely and spatially regulated maturation of B and T cell repertoire during human fetal development**. *Sci Transl Med* 2015, **7**:276ra225.

32. Klein F, Diskin R, Scheid JF, Gaebler C, Mouquet H, Georgiev IS, Pancera M, Zhou T, Incesu RB, Fu BZ *et al.*: **Somatic mutations of the immunoglobulin framework are generally required for broad and potent HIV-1 neutralization**. *Cell* 2013, **153**:126-138.