Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

CellPress

Opinion

The Individual and Population Genetics of Antibody Immunity

Corey T. Watson,[1,*] Jacob Glanville,[2,*] and Wayne A. Marasco[3,4,*]

Antibodies (Abs) produced by immunoglobulin (IG) genes are the most diverse proteins expressed in humans. While part of this diversity is generated by recombination during B-cell development and mutations during affinity maturation, the germ-line IG loci are also diverse across human populations and ethnicities. Recently, proof-of-concept studies have demonstrated genotype–phenotype correlations between specific IG germ-line variants and the quality of Ab responses during vaccination and disease. However, the functional consequences of IG genetic variation in Ab function and immunological outcomes remain underexplored. In this opinion article, we outline interconnections between IG genomic diversity and Ab-expressed repertoires and structure. We further propose a strategy for integrating IG genotyping with functional Ab profiling data as a means to better predict and optimize humoral responses in genetically diverse human populations, with immediate implications for personalized medicine.

Trends

Genetic variation in human populations affects how individuals are able to mount functional antibody responses.

Different alleles can encode convergent binding motifs that result in successful Ab responses against specific infections and vaccinations.

Given the complexity of the IG loci and the diversity of the antibody repertoire, links between IG polymorphism and antibody repertoire variability have not been thoroughly explored.

We present a strategy to mine genotype–repertoire–disease associations.

The Molecular Basis for Antibody Diversity

Antibodies (Abs) have long been appreciated as key constituents of the adaptive immune response. Their function is to enable selective recognition and mediate immune responses to novel foreign antigens. This is accomplished through the somatic generation of vast repertoires of hundreds of millions of unique Ab receptors that can be selected, matured, and ultimately participate in the formation of long-term memory during B-cell development and activation. As a consequence of this diversity, even after nearly a century of research, the complexity of the Ab response within and between individuals is only beginning to be delineated at the molecular and genetic levels.

Hundreds of variable (V) and dozens of diversity (D) and joining (J) immunoglobulin (IG) germ-line gene segments across three primary loci in the human genome comprise the necessary building blocks of the expressed Ab heavy- and light-chain repertoires [1]. Whereas the heavy chain is encoded by genes at the IG heavy-chain locus (IGH), the light chain can be encoded by genes at either the IG kappa (IGK) or IG lambda (IGL) chain loci [1]. The naïve Ab repertoire is formed by assembling variants of these building blocks using a specialized V(D)J recombination process that somatically joins various V, D, and J segments (or V and J at IGK and IGL). The introduction and deletion of P and N nucleotides at V(D)J junctions and the pairing of different heavy and light chains dramatically increase diversity (Figure 1) [2]. Considering these processes alone, a given baseline or primary naïve repertoire can theoretically sample from 10^{15} different Abs [3]. The extraordinary diversity of the naïve repertoire ensures that it will likely contain a naïve Ab with at least weak initial binding against a vast array of antigens.

[1]Department of Biochemistry and Molecular Genetics, University of Louisville School of Medicine, Louisville, KY, USA
[2]Institute for Immunity, Transplantation and Infection, and Computational and Systems Immunology, Stanford University School of Medicine, Stanford, CA, USA
[3]Department of Cancer Immunology and Virology, Dana-Farber Cancer Institute, Boston, MA, USA
[4]Department of Medicine, Harvard Medical School, Boston, MA, USA

*Correspondence:
corey.watson@louisville.edu (C.T. Watson),
jake@distributedbio.com (J. Glanville),
wayne_marasco@dfci.harvard.edu (W.A. Marasco).

CrossMark

© 2017 Elsevier Ltd. All rights reserved.

CellPress

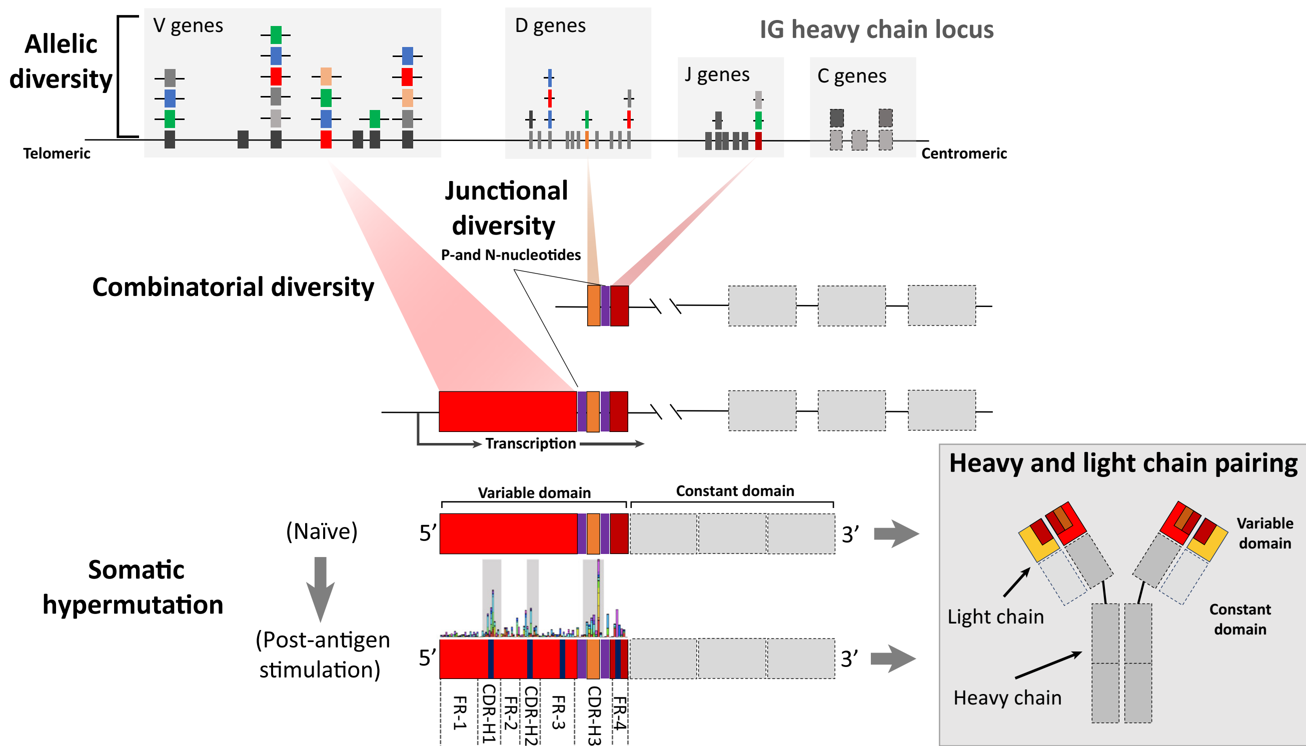

Figure 1. A Basic Overview of Key Elements That Contribute to the Diversity of Naïve and Memory Repertoires. A basic schematic of the germ-line IGH locus is shown (not to scale), consisting of clusters of tandemly arrayed IGH V, D, J, and constant (C) gene segments. For a subset of these segments, multiple alleles are shown, representing population-level 'allelic diversity' (see Table 1). During the initial formation of the naïve repertoire, single IGH V, D and J gene segment on one of two chromosomes in a given B cell are somatically recombined; at each of these steps, P and N nucleotides are added at the D–J and V–D junctions ('junctional diversity'), respectively. This process, known as V(D)J rearrangement, is the basis for 'combinatorial diversity'. The recombined V (red), D (orange), and J (maroon) segments will then be transcribed, and following splicing, will be paired with a C gene (gray). The somatic recombination process also occurs at one of two loci encoding the Ab light-chain gene segments [IGK and IGL; except it involves only V (yellow), J (maroon), and C (light gray) gene segments]. Two identical heavy chains and two identical light chains are ultimately paired through disulfide bonds to form a functional Ab; thus, additional diversity in the expressed Ab repertoire comes from 'heavy- and light-chain pairing'. Together, the V, D, and J segments depicted comprise the variable domain of the heavy chain of a functional antibody, and together with the variable domain of the light chain, encoded by V and J segments, are responsible for Ag binding. The C domains of both heavy and light chains provide structural and/or effector functions of the Ab. As shown here for the heavy chain, the variable domain is partitioned into four framework regions (FRs) and three complementarity-determining regions (CDRs). Following Ag stimulation, 'somatic hypermutations' introduce additional variation in the variable domain of the Ab (vertical purple bars), with the aim of improving binding affinity. Mutations that arise via SHM can occur across all FRs and CDRs, but these are most prevalent in CDRs, as illustrated by the hypothetical frequency histogram shown between the unmutated and mutated IG heavy-chain RNA. While the general molecular mechanisms outlined here have long been realized as the primary determinants of diversity within a given expressed Ab repertoire, there is a growing appreciation for the contribution of 'allelic diversity' as well, particularly as this pertains to repertoire differences observed between unrelated individuals. Ab, antibody; C, constant; D, diversity; IGH, immunoglobulin heavy-chain locus; IGK, immunoglobulin kappa; IGL, immunoglobulin lambda; J, joining; SHM, somatic hypermutation; V, variable.

Even so, this impressive baseline diversity can be subsequently augmented when a B cell encounters and is stimulated by an antigen to undergo somatic hypermutation (SHM; Figure 1), resulting in lineages of tens of thousands of clonally derived affinity maturation variants of the initial Ab. Specifically, SHM introduces somatic mutations throughout the variable portion of the Ab, including targeted hotspots residing within the antigen-contacting hypervariable complementarity-determining regions (CDRs). This process ultimately increases the affinity and specificity of the Ab for binding the target epitope, facilitating a highly focused antigen-specific response.

While the prevailing paradigm for investigating B-cell and Ab-mediated responses has placed emphasis on the importance of the unique molecular mechanisms cited earlier in the generation of key functional Abs, there is a growing appreciation for the fact that IG genes are highly variable at the germ-line level, exhibiting extreme allelic polymorphism and gene copy number

variation (CNV) between individuals and across populations [4–9]. Recent studies have begun to highlight that, in addition to diversity introduced during V(D)J recombination, heavy- and light-chain pairing, and SHM, IG germ-line variation (e.g., allelic variation; Figure 1) plays a vital part in determining the development of the naïve repertoire, with downstream impacts on signatures observed in the memory compartment, and the capacity of an individual to mount an Ab response to specific epitopes [10–16].

IG Loci Haplotype Diversity in the Human Population

Recent genomic sequencing indicates that IG loci, specifically IGH, may be among the most polymorphic in the human genome [17]. Across IGH, IGK, and IGL, there are currently >420 alleles cataloged in the ImMunoGeneTics information system database (IMGT) [18–21] that have been described from germ-line DNA in the human population, with an enrichment of nonsynonymous variants (Table 1). Although the validity of some alleles in IMGT has been called into question [22], the number of polymorphic alleles continues to grow [11,23,24], especially as IG gene sequencing is conducted in increasing numbers of non-Caucasian samples [7,9,25]. A recent study conducted in 28 indigenous South Africans identified 122 non-IMGT IGHV alleles [9]. In addition to IG allelic variation and single nucleotide polymorphisms (SNPs), CNVs, including large deletions, insertions, and duplications (~8–75 Kb in length), are also prevalent in IG regions (Table 1). Using IGH as an example, up to 29 of the 58 functional/open reading frame (ORF) IGHV genes may vary in genomic copy number [4,6,7,11,26–28]; CNVs of IGH D (diversity) and constant (C) region genes are also known [11,12,29]. Until recently, primarily due to technical difficulties associated with the complex genomic architecture of the IG loci, none of the known CNVs in IGHV had been sequenced at nucleotide resolution [7]; many likely remain undescribed at the genomic level.

The high prevalence of IG allelic and locus structural diversity translates into extreme levels of inter-individual haplotype variation [4–7]. For example, recent comparisons of the two available completed assemblies for the IGHV gene region (~1 Mb in length) revealed that two human chromosomes can vary by >100 Kb of sequence, with >2,800 SNPs, and CNVs of 10 IGHV functional/ORF genes [7,17]. In population sequencing experiments, extreme examples of heterozygosity have been noted, with evidence of some individuals carrying more than one allele at every IGHV coding gene [9]. Supporting earlier genetic mapping data [4,5], more recent analysis of inferred haplotypes from Ab repertoire data surveyed in nine individuals revealed that all 18 haplotypes characterized were unique [6]. Furthermore, at the population level, of the few SNPs and CNVs screened within IGH, allele and genotype frequencies have been shown to vary considerably between ethnic backgrounds [7–9,15], with evidence of selection [7]. Despite the evidence for elevated germ-line diversity, genomic resources for IG loci continue to lag behind other regions of the genome [26]. Because of this, the comprehensive and accurate genotyping of IG polymorphisms remains a significant challenge [26,30], and as a result, the full extent of IG polymorphism and the implications for human health are yet to be uncovered [26]. However, it is plausible that population-level diversity in the IG loci, particularly in IGH, will rival that of other complex immune gene families, such as the human leukocyte antigen (HLA) and killer cell IG-like receptor (KIR) genes. These genes are also characterized by extreme haplotype diversity, due to CNV and coding region variation [31,32]; HLA genes, for example, have thousands of known alleles [31]. In contrast to IG genes, HLA and KIR have been studied more extensively across human populations, and have demonstrated critical roles in disease [31,32].

Influence of IG Germ-Line Diversity in the Expressed Ab Repertoire and Ab Function

Our limited knowledge of IG population diversity has hindered our ability to comprehensively test for direct connections between IG germ-line polymorphisms, variation in the repertoire

CellPress

Table 1. Allelic, Copy Number, and Amino Acid Variation for IG Functional and Open Reading Frame Genes Cataloged in IMGT[a]

Family	Genes	Alleles	NS variants	S variants	CDR-H1 NS variants	CDR-H2 NS variants	Genes in CNV
IGHV1	11	40	19	13	2	3	6
IGHV2	4	23	26	9	3	1	1
IGHV3	27	109	82	57	9	17	12
IGHV4	10	78	92	71	11	8	8
IGHV5	2	9	4	4	0	0	1
IGHV6	1	2	0	1	0	0	0
IGHV7	2	6	4	0	0	0	1
Subtotal	58	267	227	155	25	29	29
IGKV1	20	35	33	17	4	1	1
IGKV2	11	18	14	4	1	1	0
IGKV3	8	18	24	9	2	1	0
IGKV4	1	1	NA	NA	NA	NA	0
IGKV5	1	1	NA	NA	NA	NA	0
IGKV6	3	5	2	0	0	0	0
IGKV7	0	0	NA	NA	NA	NA	0
Subtotal	44	78	73	30	7	3	1
IGLV1	7	12	4	2	0	2	1
IGLV2	6	20	13	8	2	3	0
IGLV3	11	18	14	5	3	3	0
IGLV4	3	6	2	1	0	0	0
IGLV5	5	10	3	2	0	0	1
IGLV6	1	2	2	0	0	0	0
IGLV7	2	3	1	0	0	0	0
IGLV8	1	3	1	1	0	0	1
IGLV9	1	3	0	2	0	0	0
IGLV10	1	3	4	1	1	0	0
IGLV11	1	2	1	1	0	0	0
Subtotal	39	82	45	23	6	8	3
Total	141	427	345	208	38	40	33

[a]Data accessed from IMGT February 2017. NS, nonsynonymous; S, synonymous.

generated after recombination, amino acid variation in the Ab produced, and ultimately Ab function. Advances in high-throughput sequencing technology now enable extensive characterization of the expressed Ab repertoire [33–35], creating opportunities for beginning to investigate the heritability of the Ab response at fine-scale resolution. Applications of these methods, collectively referred to as repertoire sequencing (‘IgSeq’ or ‘RepSeq’), have already led to a wealth of new discoveries in a range of contexts [33,36]. These include general observations that key features of the Ab repertoire show extensive variability between healthy individuals [10,11,13,14,37], and a limited overlap of B-cell receptor clones between individuals, even monozygotic (MZ) twins [10,13,14]. However, RepSeq studies have also revealed that these inter-individual differences are not necessarily random, but likely have a strong underlying genetic component, providing initial support for the importance of germ-line IG polymorphism

in determining the naïve and Ag-stimulated Ab repertoire. For example, several recent studies have revealed that V, D, and J gene usage in the naïve repertoire is much more highly correlated between MZ twins than between unrelated individuals [10,13,14], and that IG gene usage patterns are consistent across time points within a given individual [38]. A role for genetic factors can be seen for other repertoire features in twins as well, including the degree of SHM [13], and the distribution of CDR-H3 length and clone convergence [10,13,14]. Intriguingly, although existing data suggest that features in the memory compartment are more stochastic, likely reflective of random recruitment and transient proliferation, certain genes and repertoire features exhibit predictable patterns even in memory B cells [10,13,14,39].

Studies of repertoire heritability are consistent with a number of examples for which germ-line IG polymorphisms have been explicitly linked to features in the expressed Ab repertoire [12,15,40–42] (see Figure IA in Box 1 for hypothetical examples of IG genotype effects on the repertoire). Sasso *et al.* [40] reported the first direct connection to IG genotype, reporting

Box 1. Influence of IG Germ-Line Polymorphism on Ab Repertoire Variation and Functional Ab Structural Residues

Although the roles of IG germ-line variants have not been comprehensively studied, there is now convincing evidence that they can influence Ab repertoire variation and function in two main ways (i and ii). In addition, known functional variants exhibit allele frequency variation between human populations (iii):

(i) Gene copy number changes and coding/noncoding SNPs in IGHV genes have been shown to correlate with gene usage patterns in the naïve repertoire, the memory repertoire, patterns of SHM, class-switch frequency, and circulating Ab titers (Figure IA).

(ii) There are now many examples that provide evidence for functional effects of germ-line variants encoded in CDR-H1 and CDR-H2, many of which are polymorphic and vary between human populations. Based on known IGHV alleles in the IMGT database, residues within CDR-H1/H2 that have a higher probability of making Ag contact are also more likely to be associated with a polymorphic allele (Figure IB).

(iii) Several positions in IGHV genes that encode residues critical for antigen binding are polymorphic and exhibit different genotype frequencies between human populations and ethnicities (Figure IC).

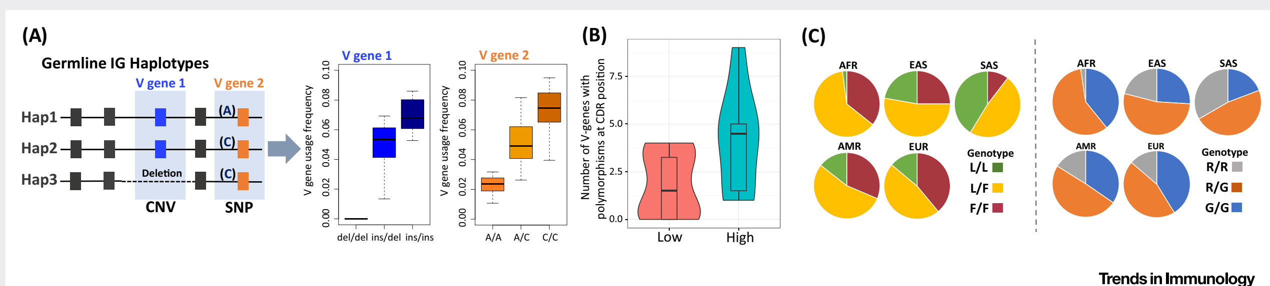

Figure I. Impacts of IG Germ-Line Polymorphism on Ab Repertoire/Structural Diversity. (A) Hypothetical examples of associations between IG gene region CNV (V gene 1 insertion/deletion) and SNP (noncoding regulatory variant, A/C) genotypes and V gene usage frequencies in the expressed Ab repertoire. (B) Violin plot showing nonsynonymous polymorphism rates in CDR positions with high (>0.6; 'high', blue) or low (<0.25; 'low', red) frequency of contact with antigen, as labeled on the *X* axis. The *Y* axis records, for each CDR-H1 and CDR-H2 position, the number of IMGT IGHV genes that have alleles with nonsynonymous polymorphisms at that position. The positional probability of antigen contact was calculated for each CDR position as the percentage of 150 crystal structures of antibody–antigen complexes from the protein database (PDB) where any atom of that residue is within 5 Å of any antigen atom. Allelic variation is enriched in antigen-contact sites, in that the number of IGHV genes with alleles containing nonsynonymous polymorphisms is greater for high contact probability positions. (C) Genotype frequency differences between five human ethnic groups [Africans (AFR); East Asians (EAS); South Asians (SAS); Central/South American (AMR); and Europeans (EUR)], published by the 1000 Genomes Project [80]*, at two SNPs in *IGHV1-69* that have been shown to encode functional residues critical for neutralizing Abs against the influenza HA stem (F54 and L54 amino acid-associated alleles; SNP rs55891010; left panel), and 'NEAT2' domain of *Staphylococcus aureus* (R50 and G50 alleles; SNP rs11845244; right panel). In the left panel, the F allele encodes the functional critical phenylalanine residue, and in the right panel, the primary glycine residue is encoded by the G allele. Interestingly, in both cases, the frequency of individuals lacking alleles encoding the critical residues varies among populations, with the L/L and R/R genotypes showing the lowest frequencies in Africans, and the highest frequencies in South Asians. rs55891010 and rs11845244 are in linkage disequilibrium, and thus R50 and L54 amino acids (and likewise, G50 and F54) tend to co-occur in alleles of *IGHV1-69*. This explains similarities in genotype frequency estimates between the two SNPs in each population. *Although these genotypes may contain error due to confounds of unrepresented CNV information, they can provide insight into potential population differences. Ab, antibody; CDR, complementarity-determining region; CNV, copy number variation; HA, hemagglutinin; IG, immunoglobulin; IMGT, ImMunoGeneTics information system database; SNP, single nucleotide polymorphism.

that CNV of *IGHV1-69* was tightly correlated with its relative usage in tonsillar B cells. Our own work has also demonstrated this relationship, but uncovered associations for *IGHV1-69* coding and potentially noncoding polymorphism as well as CNV [15]. Inferred deletions of IGHD genes have also been shown to associate with variation in D–J pairing frequencies, demonstrating that germ-line effects on the repertoire extend beyond V genes [12]. An interesting aspect of IGH CNVs is that, in addition to observed effects of these variants on the genes within the CNV event, they also can impact the usage of genes elsewhere in the locus [12,15]. For example, we recently observed apparent long-range effects of *IGHV1-69* CNV in the naïve and memory repertoire, in that individuals with fewer *IGHV1-69* germ-line copies and reduced usage showed consistently higher usage of IGHV genes over 200 Kb away [15]. The mechanisms underlying the observed effects of CNVs in human IG loci remain technically difficult to assess experimentally, but it has been speculated that these large changes in locus architecture (i.e., deletions and insertions) could alter regulatory systems related to V(D)J recombination [12,15], for example, by modifying the chromatin landscape, *cis*-regulatory elements and transcription factor binding, and/or the physical locations of the IG V, D, and J genes. All of these factors are known to be key determinants of IG gene accessibility and usage frequencies in mice [43,44].

A role for noncoding polymorphisms is also strongly supported by early work conducted in the human IGK region which directly showed that a variant associated with *Haemophilus influenzae* infection susceptibility in the recombination signal sequence (RSS) of *IGKV2-29* significantly decreased gene rearrangement frequency [42]. RSSs, which are critical for the recruitment of RAG1/2 proteins, have also been demonstrated to impact IGHV gene usage in mice [43,44]. Moreover, extensive work in the murine IG gene loci has uncovered important roles for other key *cis*-regulatory sequences and transcription factors as well [45,46]. Such analyses have not yet been comprehensively conducted in humans, and as a result, our knowledge of the IG regulatory elements involved in the formation of the expressed Ab repertoire is restricted to canonical RSS, promoter, enhancer elements, and class switch regions. However, even for these well-known noncoding regulatory regions, limited data on human population-level variation exist, and thus the broader consequences of polymorphism in these elements on Ab repertoire variability have not been explored.

Although direct links between repertoire variability and human IG CNVs and noncoding polymorphisms remain limited to the few examples discussed above, additional evidence from expressed Ab repertoire studies in unrelated individuals also highlights the clear potential for these variants to have pervasive impacts on Ab repertoire features, particularly gene usage in the naïve compartment. Most demonstrable is the fact that many of the genes with the most variability in naïve repertoire usage across individuals are also known to be in CNV, including examples of the complete absence of genes in the expressed Ab repertoires of some donors [6,10–12]. In addition, allele-specific usage in the naïve Ab repertoires of individuals heterozygous at a given IGHV gene has been demonstrated, also clearly suggesting a role for noncoding variation and CNV [11]. Moreover, although effects of germ-line IG polymorphism may be most evident on a per gene basis, it is worth noting that findings from MZ twins demonstrated that certain CDR-H3 features are highly heritable [13,14]. This indicates that even strong genetically determined biases on individual V, D, and J gene usage [and thus their nonrandom combination during V(D)J rearrangement] could also be directly linked to variation observed within CDR-H3. This is an important point given that CDR-H3 variation has classically been considered independent of the germ line [13,14].

In addition to effects of IG polymorphism on gene usage, functional CDR variants can also be directly encoded in the genome. For example, across the ~267 coding alleles cataloged in IMGT for functional and ORF IGHV genes, 60% of the 382 polymorphisms are nonsynonymous (Table 1), including sites located in CDR-H1 and CDR-H2 with predicted relevance to Ab

functional residue diversity (see Figure IB in Box 1). Although the CDR-H3 loop, formed at the V(D)J junction, is the most diverse region of an Ab and is a principal determinant of specificity [47,48], there is a growing appreciation for the importance of residues outside of CDR-H3 in antigen recognition and binding [15,49–51]. For example, recent analyses have shown that the median length of CDR-H2, which is solely encoded by germ-line V gene sequence, is substantially longer than that of CDR-H3, and typically forms the same number of interactions with antigen [52]. Specifically, analyses of antigen-binding region (ABRs; which roughly correspond to CDRs, but differ slightly in their boundaries) have shown that Abs contain a median of six, six, and four contact residues in the heavy-chain CDR-H3, H2, and H1 ABR regions, respectively. In addition, the overall percentage of energetically important Ag-binding residues within each ABR follows the same rank order, with ~31%, 23%, and 14% for H3, H2, and H1, respectively. Similar trends were noted for light-chain ABRs as well [52]. In addition, considering that many known nonsynonymous sites reside outside of CDRs (Table 1), it is worth highlighting the fact that there are also examples demonstrating indirect effects of framework region variants on Ag binding [53,54].

The Identification of Shared Ab Immune Response Signatures across Individuals

A critical question is whether the germ-line effects on the repertoire outlined above can also partially account for inter-individual variation of the Ab-mediated response in disease and clinical phenotypes. The initial observation from RepSeq studies that essentially no Ab clones were shared among individuals, including MZ twins, posed a challenge to comparative Ab repertoire analysis: how could correlates of protection be identified in the Ab repertoire if every individual was responding with different Abs? However, an answer began to emerge with the observation that in multiple settings, including viral and bacterial infection, different individuals have been shown to respond to a given antigen with Abs that share convergent amino acid signatures [13,49,54–58]. These convergent Abs are often encoded by common V genes or sets of V genes, and specific amino acid residues in their CDRs enable them to converge upon a common binding solution against a shared antigen. Critically, in some cases evaluated, convergent signatures include amino acid residues that are directly encoded in the germ line. The occurrence of such convergent Ab responses highlights the potential for tracking common immune responses across individuals, and understanding the role of genetic factors, even when each individual creates unique Abs. Importantly, the implications of this line of thinking could be broad, as IG gene biases have been observed in contexts other than infection, including autoimmunity and cancer [59,60]. Moreover, IG gene biases may also extend to usage patterns of D and J genes, light-chain genes, and heavy- and light-chain V gene pairing frequencies [56,61,62].

Structural Residues Critical for Ag Binding and Involved in Biased Gene Usage Are Encoded in the Germ Line and Exhibit Population Variability

There are now many instances for which functional contributions of biased IG genes have been traced back to specific germ-line-encoded residues, including sites that are polymorphic in the human population [15,16,50,53–55,63–65]. These examples illuminate a direct role of the IG germ line in disease-associated Ab responses. In the case of stem-directed broadly neutralizing Abs (BnAbs) against influenza hemagglutinin (HA), the most prevalent Abs use the heavy-chain gene *IGHV1-69* [66–70]. These *IGHV1-69* BnAbs recognize an overlapping epitope of group 1 influenza A viruses and only amino acids from IGHV make contact with HA. Importantly, of the 14 known alleles at *IGHV1-69*, only those encoding a critical phenylalanine at position 54 (F54) within CDR-H2 have a major role in shaping the BnAbs response [16,15,55,71]. Although *IGHV1-69* F54-encoding alleles are dominant, there is a growing list of additional HA-directed BnAbs that also show IG germ-line biases [51,56,72–74], including those also known to be polymorphic with respect to coding variants and CNVs.

Interestingly, there are additional instances of biased *IGHV1-69* allele usage in other disease contexts, with both overlapping and contrasting patterns to that observed for influenza. For example, F54 alleles are predominantly observed in *IGHV1-69*-expressing B cells associated with chronic lymphoid leukemia (CLL), whereas alleles encoding a leucine (L54) at this position are primarily used by non-neutralizing anti-gp41 Abs in HIV-1 [63,64]. Moreover, it has been shown that *IGHV1-69* F54 alleles, in comparison with L54 alleles, have lower usage in the memory B-cell pool [10,15]. This observation may be similar to trends noted for *IGHV4-34*, which is also significantly underrepresented in the memory compartment of healthy individuals [10], and presumes to reflect a selective pressure against autoreactive Abs [75,76].

Other polymorphic positions in the framework regions of *IGHV1-69*, in conjunction with CDR-H2 54, have also recently been shown to influence Ab binding of Middle East respiratory syndrome coronavirus (MERS-CoV) [53] and the *Staphylococcus aureus* NEAr iron transporter 2 (NEAT2) domain [54]. In the example of NEAT2, neutralizing Abs encoded by *IGHV1-69* alleles carrying an arginine (R) at position 50 in place of glycine (G) showed significantly reduced NEAT2 binding [54]. Interestingly, based on publicly available data, the frequencies of critical alleles within polymorphic positions of *IGHV1-69* vary across populations (see Figure IC in Box 1).

A Strategy for Defining Relationships between IG Polymorphisms, Expressed Ab Signatures, and Functional Outcomes

Considering the aforementioned evidence, we argue that the antigen-specific Ab repertoire is likely influenced by the host genotype. Although the genetic bases for repertoire and germ-line gene biases have not been comprehensively investigated, several recent studies provide a strategy for systematically integrating data on IG polymorphism and Ab responses at the population and molecular levels to provide unique insight into Ab signatures associated with disease.

We have begun to explore this idea in detail at the *IGHV1-69* locus in the context of influenza vaccination [15]. Providing strong proof-of-concept, by initially focusing on observed *IGHV1-69* allelic usage bias against a critical broadly neutralizing epitope, we genotyped the *IGHV1-69* F54/L54 allele and copy number frequencies in a cohort of 85 H5N1 vaccines, including 18 individuals with accompanying Ab repertoire data [15]. Drawing directly on aspects of repertoire heritability reviewed above, we found robust connections between these polymorphisms and repertoire gene usage in both the unmutated IgM (naïve) and IgG memory repertoires, with *IGHV1-69* germ-line gene usage increasing with the number of copies of F54 alleles. In addition to usage frequencies, *IGHV1-69* genotype also associated with *IGHV1-69* B-cell expansion, SHM, and Ig class switching. It is important to note that these genotype effects extended to levels of circulating anti-HA stem BnAbs postvaccination, with individuals carrying only germ-line-encoded CDR-H2 L54 alleles having lower *IGHV1-69* BnAbs. Furthermore, with direct repertoire sequencing, we were able to specifically demonstrate that only carriers of the *IGHV1-69* F54 alleles expressed convergent anti-BnAb signatures. These results are bolstered by similar observations recently made by two other groups that also carried out *IGHV1-69* F54/L54 allele genotyping in their cohorts [16,55]. Altogether, these data demonstrate that genetically determined baseline differences in the Ab repertoire can set the stage for disease-related responses.

A crucial aspect of this story (which is expected to emerge in other cases as well) is that the frequency of *IGHV1-69* F54 alleles and CNV varies considerably across populations [7,15]. Specifically, the number of individuals that would be predicted to lack the capacity to generate effective *IGHV1-69* BnAbs was much higher in some populations. However, we and others have shown that individuals lacking *IGHV1-69* F54 alleles likely utilize other germ-line genes in

CellPress

Key Figure

A New Paradigm for Integrating Genotypic Information into the Study of the Ab-Mediated Response in Disease and Clinical Phenotypes

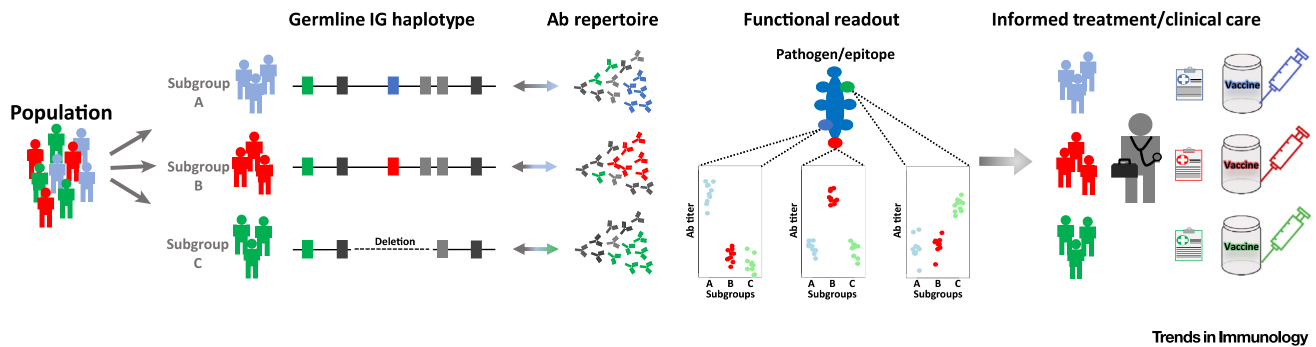

Figure 2. In the proposed paradigm, a population cohort is partitioned into subgroups based on functional genotypes/haplotypes that are directly associated with subgroup-specific signatures in the expressed repertoire and other relevant phenotypes (e.g., Ab titer; clinical outcome) associated with the Ab response to a given antigen/epitope. This partitioning can be used to inform tailored clinical care and treatment (e.g., vaccination regime). Ab, antibody.

place of *IGHV1-69* [51,55]. This finding in particular both highlights the complexity of the Ab response and demonstrates that the integration of genotyping information can help provide a more nuanced interpretation of the signatures discovered in the expressed repertoire. Moreover, it suggests that efforts should be made to study these complex responses in larger and more diverse cohorts, including individuals from presently understudied populations.

Building on findings in these initial studies [15,16,55], we propose a framework for integrating genotypic information into future studies of the Ab response in wellness and disease (Figure 2, Key Figure). The general strategy is as follows: (i) identify IG gene biases observed in a disease-related or epitope-specific response; (ii) characterize this response at the population level by performing comprehensive genotyping of coding, noncoding, and gene copy number variants at and around the locus of interest (and others if there is rationale); (iii) perform repertoire sequencing and analysis of the response in all relevant B-cell subsets to identify all Ab convergence groups with allele bias; and (iv) evaluate genotype–phenotype linkages of the functional Ab response and specific Ab convergence groups.

Concluding Remarks

We see a growing body of evidence to support the link between IG polymorphism and phenotype that may have important clinical applications (see Outstanding Questions). The most obvious of these correlations include potential effects of CNV and SNPs in non-translated and translated IG gene regions on expressed repertoire variability in naïve and memory B cell subsets. Some of these polymorphisms could be expected to more broadly impact variation in protective Ab responses [77] and quality of the memory B-cell pool. We anticipate that IG polymorphism will contribute to differences in expression of common (public) and unique (private) antibody signatures that are associated with protective responses in disease and in response to vaccination. We propose a model for the future in which cataloging these public signatures for biased gene use, V(D)J associations, SHMs, and heavy-light chain pairing in the context of IG germ-line variation should begin to provide us with information to advance our understanding of the immunogenetic potential of an individual's baseline naïve repertoire

(Figure 2), particularly when more complete data sets of biased Ab signatures to specific epitopes become available. Based on existing genetic data, it is probable that similar IG haplotypes will associate with overlapping signatures in baseline repertoire profiles, even if not to the degree of repertoire similarity observed in MZ twins. This IG polymorphism, as we and others have begun to show, may further influence the evolution of antigen-experienced B cells and plasma cells, where other genetic polymorphisms in the IG loci and environmental exposures come into play in continuing to shape affinity, epitope specificity, and fate. In addition, class-switched memory B-cell compartments will vary over time [37], and could be quantitated in the type and size of clonotypes with both public and private signatures against immunodominant epitopes.

Together, this knowledge should pave the way to using molecular and genetic signatures for mapping an individual's exposure history, current wellness state, and immune potential against future antigenic threats. For example, characterization of genotypes that specifically lead to common BnAb signatures in the repertoire should be useful for tailoring vaccines to responsive genotypes with the goal of achieving 100% 'universal vaccine' responsiveness at the population level (Figure 2). In addition, such information could lead to advances in the use of anti-idiotypic antibody and chimeric antigen receptor T-cell therapies that are directed against germ-line gene expressing B-cell clonotypes that are directly involved in autoimmune disease and hematologic malignancies [78,79]. We face tall hurdles to moving this paradigm forward, the greatest being the completion of a comprehensive catalogue of human IG haplotype variation [26]. However, with ever expanding advances in immunologic and genomic technologies, we believe that such integrative approaches are within our reach, and have the potential to transform our understanding of Ab-mediated immune responses in the clinical and research arenas.

Outstanding Questions

How large of an effect does IG polymorphism have on the development of the baseline naïve repertoire, and what types of genetic variation (CNV, coding variants, regulatory variants) matter most?

Do effects of IG genetic variants on the Ab repertoire correspond to known biases in disease and/or clinically relevant Ab responses?

What can population-level data on genetic and expressed Ab repertoire signatures tell us about an individual's exposure history, current wellness state, and immune potential against future antigenic threats?

Can we leverage integrated population-level data sets to inform clinical care, and more effective vaccine and therapeutic strategies?

Acknowledgments

This work was supported by the National Institute of Allergy & Infectious Disease of the US National Institutes of Health (NIH) under awards U01-AI074518, R56-AI109223, and R01-AI121285 to W.A.M.

References

1. Lefranc, M.-P. and Lefranc, G. (2001) *The Immunoglobulin Factsbook,* Academic Press
2. Tonegawa, S. (1983) Somatic generation of antibody diversity. *Nature* 302, 575–581
3. Schroeder, H.W. (2006) Similarity and divergence in the development and expression of the mouse and human antibody repertoires. *Dev. Comp. Immunol.* 30, 119–135
4. Chimge, N.-O. *et al.* (2005) Determination of gene organization in the human IGHV region on single chromosomes. *Genes Immun.* 6, 186–193
5. Li, H. *et al.* (2002) Genetic diversity of the human immunoglobulin heavy chain VH region. *Immunol. Rev.* 190, 53–68
6. Kidd, M.J. *et al.* (2012) The inference of phased haplotypes for the immunoglobulin H chain V region gene loci by analysis of VDJ gene rearrangements. *J. Immunol.* 188, 1333–1340
7. Watson, C.T. *et al.* (2013) 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.* 92, 530–546
8. Sasso, E.H. *et al.* (1995) Ethnic differences in polymorphism of an immunoglobulin VH3 gene. *J. Clin. Invest.* 96, 1591–1600
9. Scheepers, C. *et al.* (2015) Ability to develop broadly neutralizing HIV-1 antibodies is not restricted by the germline IG gene repertoire. *J. Immunol.* 194, 4371–4378
10. 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, 20066–20071
11. Boyd, S.D. *et al.* (2010) Individual variation in the germline Ig gene repertoire inferred from variable region gene rearrangements. *J. Immunol.* 184, 6986–6992
12. Kidd, M.J. *et al.* (2015) DJ pairing during VDJ recombination shows positional biases that vary among individuals with differing IGHD locus immunogenotypes. *J. Immunol.* 196, 1158–1164
13. Wang, C. *et al.* (2015) B-cell repertoire responses to varicella-zoster vaccination in human identical twins. *Proc. Natl. Acad. Sci. U. S. A.* 112, 500–505
14. Rubelt, F. *et al.* (2016) Individual heritable differences result in unique lymphocyte receptor repertoires of naïve and antigen-experienced cells. *Nat. Commun.* 6, 1–12
15. Avnir, Y. *et al.* (2016) IGHV1-69 polymorphism modulates anti-influenza antibody repertoires, correlates with IGHV utilization shifts and varies by ethnicity. *Sci. Rep.* 6, 20842
16. Wheatley, A.K. *et al.* (2015) H5N1 vaccine-elicited memory B cells are genetically constrained by the IGHV locus in the recognition of a neutralizing epitope in the hemagglutinin stem. *J. Immunol.* 195, 602–610
17. Watson, C.T. *et al.* (2014) Sequencing of the human IG light chain loci from a hydatidiform mole BAC library reveals locus-specific signatures of genetic diversity. *Genes Immun.* 16, 24–34
18. Pallarès, N. *et al.* (1999) The human immunoglobulin heavy variable genes. *Exp. Clin. Immunogenet.* 16, 36–60
19. Lefranc, M.-P. *et al.* (2014) IMGT®, the international ImMunoGeneTics information system® 25 years on. *Nucleic Acids Res.* 43, D413–D422
20. Pallarés, N. *et al.* (1998) The human immunoglobulin lambda variable (IGLV) genes and joining (IGLJ) segments. *Exp. Clin. Immunogenet.* 15, 8–18
21. Barbié, V. and Lefranc, M.P. (1998) The human immunoglobulin kappa variable (IGKV) genes and joining (IGKJ) segments. *Exp. Clin. Immunogenet.* 15, 171–183

22. Wang, Y. *et al.* (2008) Many human immunoglobulin heavy-chain IGHV gene polymorphisms have been reported in error. *Immunol. Cell Biol.* 86, 111–115
23. Gadala-Maria, D. *et al.* (2015) Automated analysis of high-throughput B-cell sequencing data reveals a high frequency of novel immunoglobulin V gene segment alleles. *Proc. Natl. Acad. Sci. U. S. A.* 112, E862–E870
24. Corcoran, M.M. *et al.* (2016) Production of individualized V gene databases reveals high levels of immunoglobulin genetic diversity. *Nat. Commun.* 7, 13642
25. Wang, Y. *et al.* (2011) Genomic screening by 454 pyrosequencing identifies a new human IGHV gene and sixteen other new IGHV allelic variants. *Immunogenetics* 63, 259–265
26. Watson, C.T. and Breden, F. (2012) The immunoglobulin heavy chain locus: genetic variation, missing data, and implications for human disease. *Genes Immun.* 13, 363–373
27. Milner, E.C. *et al.* (1995) Polymorphism and utilization of human VH genes. *Ann. N. Y. Acad. Sci.* 764, 50–61
28. Shin, E.K. *et al.* (1993) Polymorphism of the human immunoglobulin variable region segment V1-4.1. *Immunogenetics* 38, 304–306
29. Bottaro, A. *et al.* (1991) Pulsed-field electrophoresis screening for immunoglobulin heavy-chain constant-region (IGHC) multigene deletions and duplications. *Am. J. Hum. Genet.* 48, 745–756
30. Luo, S. *et al.* (2016) Estimating copy number and allelic variation at the immunoglobulin heavy chain locus using short reads. *PLoS Comput. Biol.* 12, 1–21
31. Trowsdale, J. and Knight, J.C. (2013) Major histocompatibility complex genomics and human disease. *Annu. Rev. Genomics Hum. Genet.* 14, 301–323
32. Parham, P. and Moffett, A. (2013) Variable NK cell receptors and their MHC class I ligands in immunity, reproduction and human evolution. *Nat. Rev. Immunol.* 13, 133–144
33. Georgiou, G. *et al.* (2014) The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat. Biotechnol.* 32, 158–168
34. Boyd, S.D. and Joshi, S.A. (2014) High-throughput DNA sequencing analysis of antibody repertoires. *Microbiol. Spectr.* 2, 1–13
35. Yaari, G. and Kleinstein, S.H. (2015) Practical guidelines for B-cell receptor repertoire sequencing analysis. *Genome Med.* 7, 121
36. Jackson, K.J.L. *et al.* (2013) The shape of the lymphocyte receptor repertoire: lessons from the B cell receptor. *Front. Immunol.* 4, 1–12
37. Galson, J.D. *et al.* (2015) In-depth assessment of within-individual and inter-individual variation in the B cell receptor repertoire. *Front. Immunol.* 6, 1–13
38. Laserson, U. *et al.* (2014) High-resolution antibody dynamics of vaccine-induced immune responses. *Proc. Natl. Acad. Sci. U. S. A.* 111, 4928–4933
39. Vollmers, C. *et al.* (2013) Genetic measurement of memory B-cell recall using antibody repertoire sequencing. *Proc. Natl. Acad. Sci. U. S. A.* 110, 13463–13468
40. Sasso, E.H. *et al.* (1996) Expression of the immunoglobulin VH gene 51p1 is proportional to its germline gene copy number. *J. Clin. Invest.* 97, 2074–2080
41. Sharon, E. *et al.* (2016) Genetic variation in MHC proteins is associated with T cell receptor expression biases. *Nat. Genet.* 48, 995–1002
42. Feeney, A.J. *et al.* (1996) A defective Vkappa A2 allele in Navajos which may play a role in increased susceptibility to *Haemophilus influenzae* type b disease. *J. Clin. Invest.* 97, 2277–2282
43. Feeney, A.J. (2009) Genetic and epigenetic control of V gene rearrangement frequency. *Adv. Exp. Med. Biol.* 650, 73–81
44. Choi, N.M. *et al.* (2013) Deep sequencing of the murine IgH repertoire reveals complex regulation of nonrandom V gene rearrangement frequencies. *J. Immunol.* 191, 2393–2402
45. Volpi, S.A. *et al.* (2012) Germline deletion of Igh 3′ regulatory region elements hs 5, 6, 7 (hs5-7) affects B cell-specific regulation, rearrangement, and insulation of the Igh locus. *J. Immunol.* 188, 2556–2566
46. Verma-Gaur, J. *et al.* (2012) Noncoding transcription within the Igh distal VH region at PAIR elements affects the 3D structure of the Igh locus in pro-B cells. *Proc. Natl. Acad. Sci. U. S. A.* 109, 17004–17009
47. Xu, J.L. and Davis, M.M. (2000) Diversity in the CDR3 region of V H is sufficient for most antibody specificities. *Immunity* 13, 37–45
48. Mahon, C.M. *et al.* (2013) Comprehensive interrogation of a minimalist synthetic CDR-H3 library and its ability to generate antibodies with therapeutic potential. *J. Mol. Biol.* 425, 1712–1730
49. Thomson, C.A. *et al.* (2008) Germline V-genes sculpt the binding site of a family of antibodies neutralizing human cytomegalovirus. *EMBO J.* 27, 2592–2602
50. Bryson, S. *et al.* (2016) Structures of preferred human IgV genes-based protective antibodies identify how conserved residues contact diverse antigens and assign source of specificity to CDR3 loop variation. *J. Immunol.* 196, 4723–4730
51. Fu, Y. *et al.* (2016) A broadly neutralizing anti-influenza antibody reveals ongoing capacity of haemagglutinin-specific memory B cells to evolve. *Nat. Commun.* 7, 12780
52. Kunik, V. and Ofran, Y. (2013) The indistinguishability of epitopes from protein surface is explained by the distinct binding preferences of each of the six antigen-binding loops. *Protein Eng. Des. Sel.* 26, 599–609
53. Ying, T. *et al.* (2015) Junctional and allele-specific residues are critical for MERS-CoV neutralization by an exceptionally potent germline-like antibody. *Nat. Commun.* 6, 8223
54. Yeung, Y.A. *et al.* (2016) Germline-encoded neutralization of a *Staphylococcus aureus* virulence factor by the human antibody repertoire. *Nat. Commun.* 7, 13376
55. Pappas, L. *et al.* (2014) Rapid development of broadly influenza neutralizing antibodies through redundant mutations. *Nature* 516, 418–422
56. Joyce, M.G. *et al.* (2016) Vaccine-induced antibodies that neutralize group 1 and group 2 influenza A viruses. *Cell* 166, 609–623
57. Parameswaran, P. *et al.* (2013) Article convergent antibody signatures in human dengue. *Cell Host Microbe* 13, 691–700
58. Strauli, N.B. and Hernandez, R.D. (2016) Statistical inference of a convergent antibody repertoire response to influenza vaccine. *Genome Med.* 8, 60
59. Johansen, J.N. *et al.* (2015) Intrathecal BCR transcriptome in multiple sclerosis versus other neuroinflammation: equally diverse and compartmentalized, but more mutated, biased and overlapping with the proteome. *Clin. Immunol.* 160, 211–225
60. Bomben, R. *et al.* (2010) Expression of mutated IGHV3-23 genes in chronic lymphocytic leukemia identifies a disease subset with peculiar clinical and biological features. *Clin. Cancer Res.* 16, 620–628
61. Forconi, F. *et al.* (2013) The IGHV1-69/IGHJ3 recombinations of unmutated CLL are distinct from those of normal B cells. *Blood* 119, 2106–2109
62. Zhu, D. *et al.* (2013) Biased immunoglobulin light chain use in the *Chlamydophila psittaci* negative ocular adnexal marginal zone lymphomas. *Am. J. Hematol* 88, 379–384
63. Hwang, K.K. *et al.* (2014) IGHV1-69 B cell chronic lymphocytic leukemia antibodies cross-react with HIV-1 and hepatitis C virus antigens as well as intestinal commensal bacteria. *PLoS One* 9, e90725
64. Williams, W.B. *et al.* (2015) HIV-1 vaccines. Diversion of HIV-1 vaccine-induced immunity by gp41-microbiota cross-reactive antibodies. *Science* 349, aab1253
65. 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, 336–338
66. Throsby, M. *et al.* (2008) Heterosubtypic neutralizing monoclonal antibodies cross-protective against H5N1 and H1N1 recovered from human IgM$^+$ memory B cells. *PLoS One* 3, e3942
67. 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, 181–193

68. Ekiert, D.C. *et al.* (2009) Antibody recognition of a highly conserved influenza virus epitope. *Science* 324, 246–251
69. Kashyap, A.K. *et al.* (2008) Combinatorial antibody libraries from survivors of the Turkish H5N1 avian influenza outbreak reveal virus neutralization strategies. *Proc. Natl. Acad. Sci. U. S. A.* 105, 5986–5991
70. 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, 850–856
71. Lingwood, D. *et al.* (2012) Structural and genetic basis for development of broadly neutralizing influenza antibodies. *Nature* 489, 566–570
72. Nakamura, G. *et al.* (2013) An *in vivo* human-plasmablast enrichment technique allows rapid identification of therapeutic influenza A antibodies. *Cell Host Microbe* 14, 93–103
73. Kallewaard, N.L. *et al.* (2016) Structure and function analysis of an antibody recognizing all influenza A subtypes. *Cell* 166, 596–608
74. Wu, Y. *et al.* (2015) A potent broad-spectrum protective human monoclonal antibody crosslinking two haemagglutinin monomers of influenza A virus. *Nat. Commun.* 6, 7708
75. Pugh-Bernard, A.E. (2001) Regulation of inherently autoreactive VH4-34 B cells in the maintenance of human B cell tolerance. *J. Clin. Invest.* 108, 1061–1070
76. Cappione, A.J. *et al.* (2004) Lupus IgG VH4.34 antibodies bind to a 220-kDa glycoform of CD45/B220 on the surface of human B lymphocytes. *J. Immunol.* 172, 4298–4307
77. Lee, J. *et al.* (2016) Molecular-level analysis of the serum antibody repertoire in young adults before and after seasonal influenza vaccination. *Nat. Med.* 22, 1456–1464
78. Fesnak, A.D. *et al.* (2016) Engineered T cells: the promise and challenges of cancer immunotherapy. *Nat. Rev. Cancer* 16, 566–581
79. Chang, D.K. *et al.* (2016) Humanized mouse G6 anti-idiotypic monoclonal antibody has therapeutic potential against IGHV1-69 germline gene-based B-CLL. *MAbs* 8, 787–798
80. Auton, A. *et al.* (2015) A global reference for human genetic variation. *Nature* 526, 68–74