Immunoglobulins (Ig) play an important role in the immune system both when expressed as antigen receptors on the cell surface of B cells and as antibodies secreted into extracellular fluids. The advent of high-throughput sequencing methods has enabled the investigation of human Ig repertoires at unprecedented depth. This has led to the discovery of many previously unreported germline Ig alleles. Moreover, it is becoming clear that convergent and stereotypic antibody responses are common where different individuals recognise defined antigenic epitopes with the use of the same Ig V genes. Thus, germline V gene variation is increasingly being linked to the differential capacity of generating an effective immune response, which might lead to varying disease susceptibility. Here, we review recent evidence of how germline variation in Ig genes impacts the Ig repertoire and its subsequent effects on the adaptive immune response in vaccination, infection, and autoimmunity.

INTRODUCTION

Adaptive immune responses carried out by B and T cells are central in the body’s fight against pathogens, and they are also implicated in autoimmune diseases. Both types of adaptive immune cells recognise antigenic determinants by specific receptors – the B-cell receptor (BCR) and the T-cell receptor (TCR). Whereas functionality of a BCR is dependent on complex formation with the CD79a/b co-receptor, in this review, we refer to BCR as the antigen-binding, cell-surface anchored immunoglobulin without this co-receptor. Similarly, we refer to the TCR as the antigen-binding receptor without the CD3 co-receptor. To be able to recognise the vast plethora of existing antigens, an immense diversity of both BCRs and TCRs is required. The collection of BCRs (or TCRs) in an individual is called the BCR (or TCR) repertoire. Immune repertoire diversity is mainly obtained by somatic gene rearrangements, so-called VDJ recombination, where one variable (V), one diversity (D) and one joining (J) gene segment combine in a supposedly stochastic fashion [1]. Note that D gene segments are only present in the heavy chain locus and are lacking in the kappa and lambda light chain loci. Although the majority of rearranged Ig sequences contain only one D gene segment, VDJ recombination has also been reported [2]. BCRs may further diversify by somatic hypermutation, which introduces point mutations in certain hotspots.

While the TCR only exists as a membrane-bound form, the antigen-binding component of BCR can be produced also as water-soluble Ig molecules – antibodies – that exert effector functions in extracellular fluids (serum, mucosa). Each B cell has its unique BCR and its specificity is determined by heavy and light chain pairing, V(D)J recombination, combinatorial and functional diversity and somatic hypermutation. The BCR of a naive B cell is in its germline configuration without any somatic hypermutation.

For a naive B cell to become activated, it needs to bind an antigen and the activated B cell can then give rise to effector B cells and memory B cells. The repertoire of a specific B-cell population can be studied separately by sorting the target population; for example, sorting naive B cells allows the exploration of a naive BCR repertoire.

In humans, the genes encoding BCRs and antibodies are located on Ig loci found on three chromosomes (chr). The heavy chain locus is located on chr14 (14q32.33) [3], the kappa chain locus is located on chr2 (2p11.2) and the lambda chain locus is located on chr22 (22q11.2) [4] (Fig. 1). Apart from the heavy, kappa and lambda loci, a number of Ig genes can also be found in clusters or alone on other human chromosomes (chr1, chr2, chr 8, chr 9, chr10, chr 15, chr16, chr18, chr21, chr22, chrY) [5, 6]. Such orphan Ig genes are generally perceived to be non-functional and have been described as pseudogenes or open-reading frames (ORFs) [6]. It is conceivable that genes outside the main Ig loci can also contribute to BCR formation. The findings of LAIR1 and LILRB1 gene sequences in rearranged BCRs in malaria-infected patients and also healthy individuals suggest that such a mechanism indeed is operative [7–10]. The LAIR1 and LILRB1 genes are located outside the IGH locus, yet templated insertions of LAIR1 and LILRB1 sequences were found to be introduced into rearranged IGH genes by various modalities. Both LAIR1 and LILRB1 encode receptors that bind to P. falciparum repetitive interspersed families of polypeptides (RIFINs)-variant surface antigens. These RIFINs antigens contribute to malaria pathogenesis by enterocyte aggregation. The non-canonicaly generated antibodies bind to P. falciparum-infected enterocytes, and the findings suggest that such non-conventional antibodies play a role in immune evasion. Importantly, these findings demonstrate that BCR repertoires are influenced directly by non-Ig genes.
Available information on the composition and sequence variation in the human Ig loci remains incomplete. This is due to a large number of highly similar and sometimes duplicated genes [11, 12], which render the correct assembly of the Ig loci from short sequences challenging. Advances in high-throughput sequencing methods and the development of specialised software tools have expanded our knowledge of Ig genes [13–17]. There is a high degree of genetic variation in the Ig loci. The numbers of allelic variants in databases for inferred variants from adaptive immune receptor repertoire sequence data [18, 19] and in the IMGT/GENE-DB database [20] have rapidly increased in the past few years. The genetic variation in Ig genes likely results from adaptation to pathogens. Consequently, functional effects of germline variation in Ig genes should be found when studying immune responses to pathogens and, therefore, may also be involved in autoimmune diseases. Germline, as used in the context of this review, refers to the unmutated configuration of a V gene, regardless of whether it is rearranged or not.

**CROSSTALK BETWEEN B CELLS AND T CELLS SHAPES ANTIBODY RESPONSES**

The shaping of the BCR repertoire after antigen encounter is orchestrated by T cells [21, 22]. An antigen bound by a BCR is endocytosed by a B cell and it gets processed into peptide fragments. These peptide fragments then bind to MHC class II molecules, which are transported to the cell surface for display to T cells. Upon TCR recognition of the peptide-MHC II complex, T cells become activated and provide help to B cells. Interaction of B cells with T cells typically takes place in organised lymphoid structures (e.g., spleen, lymph node). As a consequence of B-T-cell interaction, B cells may localise to germinal centres where BCR sequence diversification via somatic hypermutations as well as isotype switching takes place. The T cell - B cell crosstalk impacts both cell types; both T cells and B cells will undergo clonal expansion and some of the cells will persist as memory cells making up the memory BCR repertoire. A subset of post-germinal centre B cells will differentiate into effector cells (plasmablasts and plasma cells) that secrete soluble antibodies. In germinal centres, the B-cell clones with BCRs that have the highest affinity will be selected over those with lower BCR affinity in a process called antibody affinity maturation. It follows from this scheme that a BCR’s ability to bind antigen in the germline configuration is essential for the mounting of a B-cell response (Fig. 2). Therefore, the outcome of this response is highly influenced by the genetic variation affecting the sequence of the BCR itself and the number of circulating naive B cells with a BCR that can bind a given antigen [23].
LESSONS FROM STUDIES IN MICE

The notion that there is extensive genetic variation in Ig loci has been fuelled by comparative analysis of Ig genes in inbred mouse strains. Analysis of IgM- and IgG-associated VDJ rearrangements in C57BL/6 and BALB/c strains showed presence of 99 and 164 functional variable heavy (VH) genes in the genomes of the two strains, respectively [24]. Remarkably, only five VH sequences were common to both strains. Subsequent analysis of five additional mouse strains representing major subspecies revealed further large variation, and the majority of inferred germline sequences were unique to a single strain [25]. Based on this finding, it was suggested that the mouse IgH loci are complex mosaics of short haplotype blocks of disparate origins. The entire IgV region sequence of the C57BL/6 strain has been established [26]. However, it is clear from the comparative studies that the IgH locus of C57BL/6 is unable to serve as a map of genes for other strains, and further that the large genetic variation across mouse strains poses a challenge to establish suitable nomenclature [27]. Complete sequencing of the Ig gene loci of the different strains will be required to determine how much structural variation there is and how much of the strain difference can be ascribed to true allelic variation.

Genetic variation in Ig heavy loci can affect functional outcomes of immune responses. Early studies of immune responses to structurally simple haptens revealed that the hapten-specific antibodies were highly similar and utilised the same V gene combinations [28]. Structural studies revealed that the key residues in the Ig V genes that facilitated the hapten recognition in mice were germline-encoded [29]. More recent studies have indicated that genetic restriction in response also applies to complex antigens. In a study of the B-cell response to phycoerythrin, a fluorescent protein here used as an immunising antigen, it was observed that different inbred strains make qualitatively different responses [30]. The two strains of mice used in that study only differ in their Ig heavy chain loci, while the rest of their genes are identical. C57BL/6 and C.B-17 (Igh7) mice had larger numbers of phycoerythrin-specific B cells and generated smaller germinal centre responses and larger numbers of IgM memory cells and plasmablasts than BALB/c (Igh2) mice. BALB/c mice, on the contrary, elicited more affinity-matured switched Ig memory B cells. The properties of phycoerythrin-specific B cells in C57BL/6 mice correlated with the usage of a single VH that in the germline form of C57BL/6 but not in BALB/c mice afforded high-affinity phycoerythrin binding. These results suggest that some individuals may be genetically predisposed to generate non-canonical memory B-cell responses to certain antigens because of avid antigen binding via germline-encoded VH elements.

Genetic variation in murine Ig loci can also affect disease susceptibility as shown for autoimmune diseases [31]. In a study of collagen-induced arthritis, an experimental mouse model for human rheumatoid arthritis, it was found that a pathogenic antibody response to collagen II is genetically controlled by a strong influence of the Ig heavy chain locus. Antibodies targeting a major epitope of collagen II mainly utilised the same VH gene (mouse IGHV1-4). Comparing Ig V genes of the susceptible strain with that of the non-susceptible strain, the authors noted differences at two germline positions, which they considered to be allelic variants. Site-directed mutagenesis and X-ray crystallography revealed that these two residues were critical for recognition of the epitope thereby providing an explanation why the non-disease susceptible mouse strain would not produce antibodies against this major collagen epitope.

POLYMORPHIC VARIATION WITHIN THE HUMAN IG LOCI

Early studies

Before repertoire sequencing became widely available, early studies on Ig polymorphisms in disease association were performed by restriction fragment length polymorphism (RFLP) technique or detection of alleles by gene-specific polymerase-chain-reaction (PCR). One such study suggested that a polymorphism in IGHV1-69 was linked to susceptibility to rheumatoid arthritis in a specific population that lacked the HLA-DRB1 epitope [32], and several studies revealed a possible role of the IGHV locus in multiple sclerosis [33], particularly the IGHV2 gene family [34, 35]. Possible disease associations from older studies with less detailed locus characterisation (reviewed by Watson & Breden [36]) are difficult to interpret in the light of current knowledge of the Ig loci.

Studies of germline Ig variation by repertoire analysis

For now, high-throughput repertoire studies have been the most prevalent and accessible way to explore germline Ig variation. Such studies rely on rearranged Ig mRNA transcripts that are used to generate cDNA for the preparation of amplicon sequencing libraries. Repertoire studies have led to the inference of several novel Ig V alleles [13, 14, 37–41], however, new alleles must be validated from genomic DNA in order to be added to the IMGT/GENE-DB, which currently serves as the reference germline database. Since it is not always possible to obtain additional samples from the same individuals, the Inferred Allele Review sub-Committee (IARC) [42] has been set up to review previously unreported inferred Ig V alleles. The IARC and IMGT have been working closely to include inferred variants for which there is sufficient evidence from multiple studies in the IMGT/GENE-DB. IARC assesses novel allele candidates from inferred genotypes, which are deposited to the Open Germline Receptor Database (OGRDB) [18]. Currently, 16 inferred heavy chain and 2 light chain IgV alleles have been approved and added to the IMGT/GENE-DB.

Fig. 2 Different Ig V alleles might affect the immune response. The extent of antigen recognition by B cells has an impact on their activation, which in turn can influence the effectiveness of the immune response to a specific antigen.
The number of explored full-length Ig loci is still very limited, the fact that Ig loci from short reads are highly similar and frequent duplications make the assembly of Whole-genome studies are not suitable either, since the Ig genes are on the same chromosome, but it cannot determine their order, and such Ig repertoire data can only determine which genes are on the loci as well as D, J or C variation still remain underexplored compared to V genes.

Complete sequencing of human Ig haplotypes

Current knowledge of the non-coding Ig regions stems mostly from the studies that assembled full-length germline DNA sequences of the Ig loci using cosmid or BAC clones and Sanger sequencing (11, 46). There has been a lack of high-throughput methods suitable for exploring the Ig loci from gDNA. Although Ig haplotyping from repertoire data is possible, it has its limitations as repertoire studies that utilise rearranged Ig transcripts can only explore expressed genes. Furthermore, haplotype inference from such Ig repertoire data can only determine which genes are on the same chromosome, but it cannot determine their order, and neither can it tell us anything about the non-coding regions. Whole-genome studies are not suitable either, since the Ig genes are highly similar and frequent duplications make the assembly of Ig loci from short reads extremely difficult. Recently, a new method has been presented, which utilises long-read SMRT sequencing (Pacific Biosciences) to sequence the Ig loci (47). For analysis of such data, the authors have also developed a custom software, IGenotyper, which allows exploration of Ig haplotype configuration have affinity to certain antigens (Fig. 3) and can facilitate efficient immune response, the absence of certain genes or their duplication can affect the resulting Ig repertoire. Thus, haplotype variation can influence the immune response and could potentially have a profound impact on one’s ability to fight infections (48).

CONVERGENT ANTIBODY RESPONSES TO PATHOGENS INFLUENCED BY IG POLYMORPHISMS

As pathogens display a multitude of antigens, the antibody responses generated during infection are typically very diverse. Nevertheless, finding highly similar antibodies that are shared between different individuals in response to pathogens is not uncommon, and some examples of infections where such convergent responses have been observed are influenza (49), dengue (50), ebola (51) and SARS-CoV-2 (52, 53). Investigating responses to individual epitopes, for instance, when studying antibodies that have neutralising activity to viruses, it is common to find convergent and stereotyped antibody responses (54). This suggests that antibody fitness for a defined epitope governs the selection of BCRs engaged in the response and that the rules for selection are similar across individuals. Importantly, the phenomenon of convergent antibody responses provides the foundation for genetic effects to play out. Biases in the V gene usage as observed in antigen-specific/disease-specific BCR repertoires suggest that germline variation in these genes might have a functional effect (Fig. 4). It also suggests that to see genetic effects on recognition of antigen, studies will need to focus on recognition of single and defined epitopes (55). Examples of conserved germline-encoded residues that are crucial for recognition of defined epitopes are shown in Table 2. We expect many more such examples to be uncovered in the future. In the following, we will review selected antibody responses to particular antigens where such convergence has been observed and where Ig polymorphisms have been demonstrated to affect the antibody response.

Broadly neutralising antibodies against HIV

For vaccination against HIV, broadly neutralising antibodies targeting the HIV envelope are considered particularly useful. The VRC01-class of antibodies targeting the CD4-binding site on HIV-1 glycoprotein 120 [56] (gp120) are among the broadest neutralising antibodies [57, 58]. These antibodies share common

Table 1. Number of human Ig genes and alleles in IMGT/GENE-DB.

| Gene type | Functional genes a | Total (main locus) | Total (including orphans) |
|-----------|--------------------|--------------------|---------------------------|
| Genes     | Alleles b          | Genes              | Alleles                   | Genes | Alleles |
| IGHV      | 57                 | 311                | 159                       | 503   | 197     | 564    |
| IGKV      | 41                 | 72                 | 77                        | 115   | 109     | 149    |
| IGLV      | 34                 | 87                 | 80                        | 155   | 84      | 159    |
| IGHD      | 23                 | 30                 | 27                        | 34    | 37      | 44     |
| IGJH      | 6                  | 13                 | 9                         | 19    | –       | –      |
| IGKJ      | 5                  | 9                  | 5                         | 9     | –       | –      |
| IGLJ      | 5                  | 7                  | 7                         | 10    | –       | –      |
| IGHG      | 9                  | 84                 | 11                        | 91    | 12      | 93     |
| IGKC      | 1                  | 5                  | 1                         | 5     | –       | –      |
| IGLC      | 5                  | 17                 | 7                         | 21    | 10      | 25     |

These data are accurate as of 9 April 2021. Note that the IMGT/GENE-DB contains mostly mapped genomic Ig sequences but, as mentioned in this review, there is a lack of such data. The actual numbers of all possible Ig genes and alleles are likely higher than those presented in this table.

aGenes with at least 1 allele described as functional.

bSome alleles of the genes labelled as functional might not be functional after all.

[20] reference database following IARC assessment and recommendation.

Despite a large number of known germline V alleles (Table 1), more are still being discovered [37, 43]. This suggests that the amount of germline polymorphisms in Ig V genes is likely underestimated, especially since most of the known Ig V alleles are only from individuals of European ancestry [44]. Most studies have only investigated the coding region of V genes and left the non-coding regions unexplored. Although haplotypes and potential gene deletions can be inferred using specialised software [45], the non-coding regions of the loci as well as D, J or C variation still remain underexplored compared to V genes.
features including the highly conserved use of the IGHV1-2 gene paired with a light chain with a short complementarity determining region 3 of 5 amino acids [58, 59]. The light chain of VRC01 antibodies is frequently encoded by kappa genes, although some combinations with lambda have also been reported [58, 60]. Structural studies revealed that Trp50, Asn58, Arg71 and Trp100B in the IGHV1-2*02-encoded heavy chain form contact with conserved residues of the gp120 epitope via hydrogen bonds (Trp50-Asn280gp120; Asn58-Arg456gp120; Trp100B-Asn279gp120) and salt bridges (Arg71-Asp368gp120) [61, 62].

Due to differences between IGHV1-2 alleles, the ability to produce VRC01-class BCR depends on the alleles in an individual’s genotype [23, 60]. Allele IGHV1-2*05 was found incompatible with producing VRC01-class BCR, likely due to a difference in the amino-acid residue at position 50, where IGHV1-2*05 contains Arg50 instead of Trp50 [23, 60]. In a study by Lee and colleagues, the authors showed that although IGHV1-2*04 contains the critical residue Trp50, individuals heterozygous for IGHV1-2*04 had a lower amount of VRC01-precursor naive B cells [23]. The reason for this is unclear because the only difference between alleles *02 and *04 is located in the framework region 3 at position 66 (Arg in *02 and Trp in *04). Since IGHV1-2 alleles that decrease or hamper the chance of producing VRC01-class antibodies are relatively common in the population [23], not all individuals are equally able to produce potent neutralising antibodies to target the HIV-1 gp120 epitope.

Antibodies against Plasmodium falciparum circumsporozoite protein

For vaccination against malaria, elicitation of antibodies to the Plasmodium falciparum circumsporozoite protein (PfCSP) appears to provide protection in animal models [63]. These antibodies, which target a linear repeat region of the PfCSP antigen, show a
striking bias in usage of IGHV3-33/IGKV1-5 both when elicited in malaria-exposed individuals [64] or in malaria-vaccinated indi-
viduals [65–67]. A key residue for recognition of the linear PfCSP
repeat epitope, as also demonstrated by X-ray crystallography, is
the germline-encoded tryptophan (Trp) at position 52 in the
CDRH2 [65]. Genes that are highly similar to IGHV3-33, namely
IGHV3-30, IGHV3-30-3 and IGHV3-30-5; differ at position 52, where
their sequence encodes serine (Ser) instead of Trp. On mutating
high-affinity antibodies to encode Ser52 instead of Trp52, the
reactivity of such mutants to PfCSP was reduced. Interestingly, the
study of malaria-exposed individuals of Tanzania revealed a
putative novel variant of IGHV3-30 encoding Trp52 instead of
Ser52 [64].

Antibody response against the seasonal flu vaccine

Exploration of antibody repertoires in individuals vaccinated
against seasonal flu has revealed a stereotyped response against
the hemagglutinin stem of the influenza virus. Neutralising
antibodies against the influenza A hemagglutinin stem have a
biased usage of IGHV1-69 in their heavy chain [68, 69]. Structural
and site-directed mutagenesis studies have shown that a
hydrophobic residue at position 53 and phenylalanine (Phe) at
position 54 in IGHV1-69 are critical for binding [70]. Of note,
amin-o-acid position 54 in IGHV1-69 is polymorphic containing
either Phe or leucine (Leu). In a study of antibody repertoires in
individuals vaccinated with seasonal flu vaccine individuals, it was
observed that individuals who carried only Leu54 variants in their
genome did not make a proper response. Interestingly, there was
a difference in the frequency of these variants among ethnicities.
The Leu54 variants were more prevalent among Europeans and
nearly absent among Africans [71].

Antibody neutralisation of virulence factors of commensal
bacteria

Staphylococcus aureus (S.aureus) is commonly found on the skin
and in the upper respiratory tract of healthy individuals, but it can
also become an opportunistic pathogen. The bacteria require iron
to become pathogenic, and S.aureus obtains this critical factor by
using haem iron from the host [72]. This process involves an iron
surface determinant (Isd) system, particularly the IsdB surface
receptor, which interacts with the haem-binding part of human
haemoglobin [73]. By studying the antibody response against IsdB
in healthy individuals, it was observed that antibodies that bind
two domains of IsdB (NEAT1 and NEAT2) have a biased V gene
usage [74]. NEAT1 was mostly neutralised by antibodies utilising
the IGHV4-39 gene, and NEAT2 was neutralised by antibodies utilising
the IGHV1-69 gene. Interestingly, IGHV1-69 alleles that encoded Arg at position 50 instead of Gly or Ala led to
complete abolishment of the binding, thus losing the capacity for
neutralisation. Individuals with different germline variants of
IGHV1-69 could therefore have different susceptibility to infection
by S.aureus.

Fig. 4 Differences in germline Ig V alleles affect the immune response. A Ig V genes exist in many allelic forms, which differ in one or more
nucleotides. Non-synonymous polymorphisms that change the amino-acid sequence may also alter the antigen-binding capacity of the
resulting antibody. B For some V genes, the expression levels of alleles in heterozygous individuals are uneven, with one allele more
frequently used in the repertoire than the other allele [37, 38, 106]. C There exist differences in the distribution of alleles among different
population groups. As a result, vaccines relying on a specific Ig allele may not elicit equally efficient responses across ethnicities [23, 71].
**CONVERGENT ANTIBODIES IN COELIAC DISEASE**

Coeliac disease is caused by a harmful immune response to cereal gluten proteins. Unlike other autoimmune diseases with strong HLA associations, the antigens recognised by disease-specific T cells and B cells are known [75]. Coeliac patients have CD4+ T cells that recognise deamidated gluten peptides in the context of disease-associated HLA-DQ allotypes. In addition, these patients have B cells/plasma cells, which are specific to the autoantigen transglutaminase 2 and deamidated gluten peptides. Both types of antibodies have a biased usage of VH and VL pairs that is observed across individuals. Transglutaminase 2 specific plasma cells present in the coeliac disease gut lesion have a particularly prominent bias for the IGHV5-51:IGKV1-5 pair [76–78], but IGHV3-48:IGLV5-45 and IGHV4-34:IGKV1-39 are also frequently used [78]. The VH:VL usage of BCR/antibody depends on the epitope of transglutaminase 2 that is being recognised [79]. Similarly to transglutaminase specific cells, plasma cells with BCR specific to deamidated gluten peptides utilise stereotypical pairings of IGHV3-15:IGKV4-1; IGHV3-23:IGLV4-69 and IGHV3-74:IGKV4-1 [80].

An Arg residue at position 55 was found to be critical for binding of the stereotyped IGHV3-15:IGKV4-1 antibody to deamidated gluten peptides [82]. This residue was also critical for binding in IGHV3-43:IGKV4-1 antibodies [81]. Interestingly, one plasma cell-specific for deamidated gluten peptide isolated from the lesion of a coeliac disease patient carried the IGHV4-43:IGKV4-1 pair, and its heavy chain was encoded by the IGHV4-4*07 allele. This allele carries Arg at position 55 in contrast to the IGHV4-4*01 allele that carries Glu. Upon mutating the antibody to carry Glu at position 55, the antibody binding was lost [81]. This finding of one antibody from one patient does not say much about the role of allele IGHV4-4*01 in coeliac disease susceptibility, but it demonstrates that polymorphic variation in Ig genes influences the antibody response to deamidated gluten peptides. Further studies are required to understand how Ig polymorphisms shape the B-cell response to gluten in coeliac disease.

**CHANGES IN AFFINITY DUE TO FAB GLYCOSYLATION AT GERMLINE-ENCODED RESIDUES**

Ig molecules, like other proteins, can be subject to post-translational modifications, such as N-/O-linked glycosylation [83–86] or tyrosine sulfation [87]. Glycosylation is probably the most widely studied post-translational modification of antibodies, and it can occur both at the variable domain (antigen-binding fragment, Fab) or the constant region (crystallisable fragment, Fc). N-glycosylation usually occurs at specific motifs called sequons that are made of Asn-X-Ser/Thr, where asparagine (Asn) becomes glycosylated. The residues that are targeted for O-glycosylation are serine (Ser) or threonine (Thr) [88]. Modification of an amino-acid residue within the antigen-binding domain (Fab) of an antibody or a BCR can have an effect on its binding properties (Fig. 3D). Glycosylation status can be altered in response to various stimuli including infection, aging, smoking, etc [89].

The number of residues in the Fab domain that can be glycosylated depends on the germline gene used to make that Ig molecule since the variable region can be coded by many different genes. Glycosylation of the variable region is more frequent for antibodies encoded by the IGHV4 family and occurs less frequently in those encoded by IGHV1 or IGHV3 family [90]. The IGHV4-34*01 allele, in particular, contains a germline-encoded Asn-X-Ser/Thr sequon in its CDR2 region, which promotes N-linked glycosylation. Other genes with germline-encoded sequon include for example IGHV1-8 or IGHV5-10-1 [90]. Additional sequons within the variable region can be created via somatic hypermutation [86].

Glycosylation in the V region might affect the antigen-binding capacity of antibodies by either increasing [84, 85] or reducing their binding capacity [83]. In some cases, N-linked glycosylation of the variable domain can decrease the self-reactivity of an antibody [83]. A study conducted in mice has shown that an antibody produced by self-reactive B cells had reduced capacity to bind its self-antigen after undergoing a mutation of Ser to Asn at position 52; a change that can enable N-linked glycosylation [83]. When the Asn52 antibody was expressed in bacteria, where glycosylation does not occur, the self-reactivity of this antibody was restored, suggesting an impact of glycosylation on autoantigen binding. This study highlights the important role of post-translational modifications occurring in the Ig variable domain.

**CONSTANT REGION FUNCTION AND GERMLINE GENE VARIATION**

While the interaction of an antigen occurs at the Ig variable region, the constant region of an Ig molecule is responsible for interaction with other components of the immune system [91]. The Ig constant region determines the isotype and subclass of an Ig molecule, and it can be recognised and bound by various Fc receptors (FcR), through which antibodies can exert their effector functions [92, 93]. Binding of the constant region to FcR or the polymeric Ig receptor (pIgR) is important for transport across epithelial surfaces [94]. Antibody function may be further modified by post-translational modification. Glycosylation of a conserved Asn residue at position 279 in the constant region [95] by different sugar moieties can be used to alter the function of an antibody, as is often seen during infection [91].

Antibody effector functions can be also influenced by Ig allotypes [96], which are polymorphic variations identified in the constant regions of immunoglobulins. Allotypes are described

---

**Table 2.** Polymorphic residues in the Ig V domain that are critical for antibody-antigen binding.

| Gene | Amino-acid residue Katab (IMGT) numbering | Alleles | Region | Antigen | Reference |
|------|----------------------------------------|--------|--------|---------|-----------|
| IGHV1-2 | Trp50 (IMGT 55) | 02; 03; 04; 07 | FWR2/CDR2 | HIV-1 envelope glycoprotein 120 (CD4-binding region) | Scharf et al. [61, 62] |
| IGHV1-69 | Phe54 (IMGT 62) | 01; 03; 05; 06; 07; 12; 13; 14; 15; 17; 18; 19 | CDR2 | Influenza A virus hemagglutinin stem | Avnir et al. [70, 71] |
| IGHV1-69 | Gly50 (IMGT 55) | 01; 03; 05; 06; 10; 12; 13; 14; 16; 17; 19 | CDR2 | Staphylococcus aureus IsdB NEAT2 domain | Yeung et al. [74] |
| IGHV3-33 | Trp52 (IMGT 57) | 01; 02; 03; 04; 06; 07 | CDR2 | Plasmodium falciparum circumsporozoite protein | Imkeller et al. [65], Tan et al. [64], Murugan et al. [66], Oyen et al. [67] |
for IgG1, IgG2, IgG3, IgA2, and Ig kappa. Due to linkage of the human Ig heavy constant genes (ordered IGHG3, IGHG1, IGHA1, IGHG2, IGHG4, IGHE and IGHA2) Gm-Am haplotypes can be defined [97]. Differences in Ig allotypes seem to affect binding to Fc receptors and consequently, this can have a further effect on antibody-mediated immune responses [98, 99]. Allotypes may also play a role in the immune response to infections and in autoimmunity. Exploration of frequency of IgG1 allotypes in human cytomegalovirus (HCMV) infection revealed that indi-
uals with allotype G1m17 had higher levels of HCMV-specific IgG1 antibodies as well as higher amounts of total HCMV-specific Ig when compared to individuals homozygous for the G1m3 allotype [100, 101]. Similarly, subclass composition of IgG antibodies against bacterial antigens was also found to correlate with different allotypes [102]. In multiple sclerosis, G1m1 allotype was dominant among intrathecal antibody-producing B cells of G1m1/G1m3 heterozygous patients, and these cells also displayed preferential usage of the IGHV4 paired with IGKV1 [103]. The link of IgG constant region polymorphisms to stereotyped V gene usage could possibly reflect recognition of particular antigen by disease-related B cells in multiple sclerosis.

The constant regions of different Ig subclasses are encoded by different genes and alleles. It is important to note that the same allotype, characterised by a defined amino-acid residue at a specific position, can be encoded by different Ig constant alleles. Currently, there is a lack of studies describing the germline variation of the constant Ig genes. A relatively recent study from Brazil identified 28 novel IGHG alleles among 357 individuals from diverse population groups, suggesting there is a quite extensive germline variation in the constant that remains unexplored [104]. Some of the identified polymorphisms represented amino-acid changes, although the majority were synonymous mutations. Polymorphisms affecting residues important for glycosylation or FcR binding could have potential implications on the function of the immune system.

EFFECT OF NON-CODING POLYMORPHISMS ON THE IG REPERTOIRE

In addition to coding polymorphisms, the expression levels of different alleles in heterozygous individuals might also play a role in the ability to develop a sufficient immune response. Different expression levels of Ig V alleles in heterozygous individuals have been reported in multiple studies (Fig. 4B) [23, 37, 38], but the reason for such uneven expression remains unclear. It is likely that non-coding regions might play a role in the regulation of Ig expression, but this has been very little explored [105]. Although the non-coding regions have received slightly more attention in the past years [37, 106], polymorphisms in these regions remain poorly characterised.

Promoters

Each Ig V gene has its own promoter that regulates its transcription. Ig promoters are composed of conserved elements that include, but are not limited to, a TATA-box, an octamer/ decamer, and various other conserved motifs (Fig. 5) [107, 108]. Elements of the Ig promoter serve as binding sites for transcription factors that regulate the expression of various genes [109]. Promoters of the heavy, kappa, and lambda V genes are quite different from each other, and there are also differences among the different V gene families within the same Ig locus. An interesting feature of some Ig promoters is their directionality. The IGHV6-1 promoter, which has a TATA-box upstream and downstream of the octamer, can function in both directions [110]. Two putative TATA-boxes were also observed in the promoter of IGHV5-1 and IGHV3 genes by analysing sequences listed in the IMGT/GENE-DB [20, 37]. The functional significance of this phenomenon and its possible impact on antibody production is not clear.

5’UTRs and leader sequences

Germline polymorphisms in the 5’UTR and the leader sequences have been largely overlooked until recently [37]. Since high-throughput sequencing platforms have limitations in terms of the length of sequences, many library preparation protocols produce amplicons that only capture the V-REGION and the junction. Nevertheless, the 5’UTRs and leader sequences of Ig V genes also contain polymorphisms [37, 106]. The role of such upstream polymorphisms is unknown for now, but mutating the sequence downstream of the transcription initiation site (corresponding to the 5’UTR) was shown to decrease transcription efficiency [111]. The leader sequence is located downstream of the 5’UTR (Fig. 5) and is encoded by L-PART1 and L-PART2, which are spliced together in a functional Ig transcript. The leader sequence has an important role in in vitro antibody production since it is translated and serves as a peptide signal sequence before being cleaved off. Experimental studies aiming to optimise antibody production have shown that differences in the leader sequences had an influence on the efficiency of antibody expression in vitro [112, 113].

Intron

In the germline DNA, the leader sequences of an Ig V gene are separated by an intron, which is spliced out in the mature mRNA (Fig. 5). Since the intron is absent in Ig repertoire studies, germline variation in this region is unexplored. Introns might also play a role in shaping the Ig repertoire. Different splicing patterns of introns in kappa genes using cryptic splice acceptor sites were found to alter the mRNA levels of the same genes [114]. Recently, partial intron retention was observed in the Ig light chain transcripts from naive BCR repertoires, particularly in genes with low expression.
levels [106]. Intron retention in other genes often introduces a premature stop codon in the transcript, which makes the mRNA more susceptible to degradation. It is possible that alternative splicing could be a way to regulate the expression of light chain V genes.

**Recombination signal sequence**
The V recombination signal sequence (V-RS, also sometimes abbreviated as RSS) is crucial for the recombination of a V gene with a D or J gene, and producing a functional Ig transcript. Polymorphisms in this sequence can affect the recombination efficiency, as was observed in the case of IGKV2D-29 (previously known as A2 or VA2). Antibodies produced in response to vaccination by Haemophilus influenzae type B polysaccharide (Hib PS) were observed to preferentially utilise the IGKV2D-29 light chain gene [115]. It was demonstrated that a polymorphism at the V-RS sequence of this gene correlates with lower recombination frequency [116, 117]. Navajos have a 10-fold higher incidence of Haemophilus influenzae type B infection compared with control populations, and as this polymorphism is frequent among Navajos, it was hypothesised that this may be the underlying reason for higher disease susceptibility in this population [116, 117].

Despite the importance of the V-RS region, there has been very little research on V-RS polymorphisms in the past few years. Since the V-RS is spliced out and no longer present in the rearranged Ig mRNA, it is not possible to study this region from Ig repertoire data. Therefore, little is known about the variation in V-RS and the functional effects of polymorphisms.

**ENHANCERS AND THEIR ROLE IN REGULATING IG EXPRESSION**
The IgH loci contain three main enhancer clusters near the C genes (Fig. 1). These can be bound by important transcription factors such as NF-kB [118]. There is a lack of studies describing polymorphisms in these enhancers and their potential effect. Existing studies on Ig enhancer polymorphisms were done on a very small scale and there are conflicting conclusions about their potential role in diseases [119, 120]. However, mutations in Ig heavy enhancer have been previously implicated in B-cell lymphomas [121]. The 3′ regulatory region and the switch regions of human and murine Ig loci have been shown to contain hotspots for oestrogen receptor binding [122, 123]. A study in mice that investigated the role of Ig enhancers found differences between male and female BCR repertoire [123]. These in-vitro experiments showed that the addition of oestrogen to purified mouse splenic B cells increased the amount of Ig heavy chain transcripts. Additionally, the presence of oestrogen receptors bound to the switch regions was identified by chromatin immunoprecipitation. The authors suggested that such hormonal regulation could affect the expression of different Ig isotypes. These findings could also provide a potential explanation for the differences in Ig repertoires between males and females, and perhaps it could shed light on factors involved in the increased frequency of some autoimmune diseases among females. It is however unclear how frequent polymorphisms in this region of the loci are and what the functional effect of such polymorphisms would be.

**THE IMPACT OF BCR GENES ON SUSCEPTIBILITY FOR AUTOIMMUNE DISEASE: THE JURY IS STILL OUT**
Knowing that MHC genes with their central role in adaptive immunity by far are the chief determinant for autoimmunity, it is conceivable that other genes feeding into adaptive immune pathways are susceptibility genes for autoimmune diseases. For diseases with MHC associations, primary associations with MHC class I alleles are seen with seronegative diseases while primary associations with MHC class II alleles are seen in diseases with autoantibodies or where there is evidence for the involvement of B cells, such as multiple sclerosis [124]. As twin studies have demonstrated that a person’s BCR repertoire is affected by genes [125, 126], finding BCR genes among susceptibility genes for seropositive autoimmune diseases would be expected. In fact, very few studies have reported finding in keeping with this notion. A study of the Oceanic population reported that the IGHV4-61*02 allele is associated with a higher risk of rheumatic heart disease [127]. Further, a susceptibility locus for Kawasaki disease was identified among Ig genes [128], and in a later study, researchers identified a particular SNP within the IGHV3-66 gene as a risk factor for this disease [129]. However, the scarcity of results does not imply that BCR genes are not susceptibility genes. The coverage of genetic markers in the chips used for typing in genome-wide association analysis (GWAS) for the BCR (i.e., IGH, IGK and IGL) as well as TCR (TRA and TRB) loci is scarce and much less than that for MHC (HLA) (Table 3). Hence, these loci have yet to be scrutinised in extensive association analysis. This will only become possible when the genomic structures, knowledge of allelic variation, and knowledge of linkage disequilibrium in these loci have become more complete.

**CONCLUDING REMARKS AND FUTURE PERSPECTIVES**
Genomic variation in the Ig loci shapes a person’s Ig repertoire [48] and it can have a profound effect on the ability to develop a sufficient immune response to a specific antigen [130], as described above. Due to the large complexity of the Ig loci, the options to study these genomic regions have been somehow limited. High-throughput Ig repertoire studies have been central in characterising germline variants of Ig V genes (and recently also IGHD [131] and TRCβ [132] genes), however, such studies have their limitations. Since repertoire studies utilise mRNA (transcribed to cDNA) and not genomic DNA, they only provide information about the coding parts of the Ig loci. Yet, promoters, introns, V-RS and upstream sequences of Ig V genes also seem to play a role in regulating Ig expression. There is a need for studies that would characterise genetic variation in the non-coding regions of Ig loci and explore their functional implications. This could potentially help explain the uneven expression of Ig alleles in heterozygous individuals. The main issue in characterising non-coding genetic variation in Ig loci has been a lack of suitable high-throughput methods that would allow amplification and high-throughput sequencing of larger sections of the Ig loci. Recently, a novel method has been published, which utilises long-read high-throughput sequencing to explore germline Ig loci [47, 133]. This will hopefully enable researchers to study genomic Ig variation more effectively.

Another limitation of repertoire studies is their technical aspect, particularly the accuracy of sequencing methods and the software used for analysis. Although high-throughput methods are widely used for characterisation of polymorphisms in the genome, each method has its own biases and limitations that can affect the data

| Locus | SNPs | Size (kb) | SNP/kb |
|-------|------|-----------|--------|
| HLA   | 7125 | 3212      | 2.22   |
| TRA   | 62   | 930       | 0.07   |
| TRB   | 7    | 524       | 0.01   |
| IGK   | 4    | 1378      | 0.00   |
| IGL   | 15   | 997       | 0.02   |

**Table 3.** Locus coverage in GWAS.
interpretation [134–137]. Currently, there is a lack of benchmarking and proof-of-concept studies that would demonstrate the accuracy of polymorphism inference in a synthetic dataset containing defined sequences. Nevertheless, Ig repertoire studies remain a valuable tool that has played a key role in characterising genomic variation in Ig loci.

We expect that many more examples of functional impact of Ig polymorphisms will be described in the future, particularly in studies focusing on defined antigenic epitopes where there is usage of convergent VH/VL. In the relatively near future, we should also learn whether Ig polymorphisms have an impact on risk and development of autoimmune disease.

REFERENCES

1. Tonegawa S. Somatic generation of antibody diversity. Nature. 1983;302:575–81.
2. Safonova Y, Pevzner PA. VIDOJ recombination is an important and evolutionarily conserved mechanism for generating antibodies with unusually long CDRs. Genome Res. 2020;30:1547–58.
3. McBride OW, Batey J, Hollis GF, Swan DC, Siebenlist U, Leder P. Localization of human variable and constant region immunoglobulin heavy chain genes on subtelomeric band q32 of chromosome 14. Nucleic Acids Res. 1982;10:8155–70.
4. McBride OW, Heiter PA, Hollis GF, Swan D, Otey MC, Leder P. Chromosomal location of human kappa and lambda immunoglobulin light chain constant region genes. J Exp Med. 1982;155:1480–90.
5. Lütscher E, Zimmer FJ, Klopotock T, Grzeschik KH, Jaenichen R, Straubinger B, et al. Localization, analysis and evolution of transposed human immunoglobulin V kappa genes. Gene. 1989;88:215–23.
6. Lefranc M-P, Lefranc G. The Immunoglobulin FactsBook. Academic Press; 2001.
7. Tan J, Pieper K, Piccoli L, Abdí A, Perez MF, Geiger R, et al. A LAIR1 insertion generates broadly reactive antibodies against malaria variant antigens. Nature. 2016;529:105–9.
8. Pieper K, Tan J, Piccoli L, Foglierini M, Barbieri S, Chen Y, et al. Public antibodies to malaria antigens generated by two LAIR1 insertion modalities. Nature. 2017;547:589–601.
9. Koning MT, Vletter EM, Rademaker R, Vergroesen RD, Parren P, et al. Template insertions at VD and DJ junctions create unique B-cell receptors in the healthy B-cell repertoire. Eur J Immunol. 2020;50:2099–101.
10. Chen Y, Xu K, Piccoli L, Foglierini M, Tan J, Jin W, et al. Structural basis of malaria RIFIN binding by LILRB1-containing antibodies. Nature. 2021;592:639–43.
11. Matsuda F, Ishii K, Bourvagnet P, Kuma K-I, Hayashida H, Miyata T, et al. Templated insertions at VD and DJ junctions create unique B-cell receptors in the healthy B-cell repertoire. Eur J Immunol. 2020;50:2099–101.
12. Pallarés N, Lefebvre S, Contet V, Matsuda F, Lefranc M-P. The human immunoglobulin heavy chain repertoire is germline-focused and highly variable between inbred strains. Philos Trans R Soc Lond B Biol Sci. 2015;370:20140236. https://doi.org/10.1098/rstb.2014.0236.
13. Collins AM, Wang Y, Roskin KM, Marquis CP, Jackson KJL. The mouse antibody heavy chain repertoire is germline-focused and highly variable between inbred strains. Philos Trans R Soc Lond B Biol Sci. 2015;370:20140236. https://doi.org/10.1098/rstb.2014.0236.
14. Watson CT, Collins AM. A proposed new nomenclature for the immunoglobulin genes of Mus musculus. Front Immunol. 2019;10:2961.
15. Corcoran MM, Phad GE, Bernat NV, Stahl-Hennig C, Sumida N, Persson MAA, et al. Interpretation of high-throughput B-cell sequencing data reveals a high frequency of novel immunoglobulin V gene segment alleles. Proc Natl Acad Sci USA. 2015;112:E862–E870.
16. Gadala-Maria D, Gidoni M, Marquez S, Vander Heiden JA, Kos JT, Watson CT, et al. Identification of subject-specific immunoglobulin alleles from expressed repertoire sequencing data. Front Immunol. 2019. https://doi.org/10.3389/fimmu.2019.00129.
17. Corcoran MM, Phad GE, Bernat NV, Stahl-Hennig C, Sumida N, Persson MAA, et al. Production of individualized V gene databases reveals high levels of immunoglobulin genetic diversity. Nat Commun. 2016;7:13642.
18. Dandekar T, Eklund H, Porfiri M, Shlomchik M, Stehling-K曆ning A. A comprehensive database for human and mouse immunoglobulins. Nucleic Acids Res. 2004;32:D236–241.
19. Lanzavecchia A. Receptor-mediated antigen uptake and its effect on antigen presentation to class II-restricted T lymphocytes. Annu Rev Immunol. 2003. https://doi.org/10.1146/annurev.immunol.08.040190.004013.
variable, diversity, and joining genes and characterization of allelic and copy-number variation. Am J Hum Genet. 2013;92:530–46.

47. Rodriguez OL, Gibson WS, Parks T, Emery M, Powell J, Strahl M, et al. A novel framework for characterizing genomic haplotype diversity in the human immunoglobulin heavy chain locus. Front Immunol. 2020. https://doi.org/10.3389/fimmu.2020.02136.

48. Collins AM, Yaari G, Shepherd AJ, Lees W, Watson CT. Germline immunoglobulin genes: disease susceptibility genes hidden in plain sight? Curr Opin Syst Biol. 2020. https://doi.org/10.1016/j.coisb.2020.10.011.

49. Jackson KJL, Liu Y, Roskin KM, Glanville J, Hoh RA, See K, et al. Human responses to influenza vaccination show seroconversion signatures and convergent antibody rearrangements. Cell Host Microbe. 2014;16:105–14.

50. Paramekanti P, Liu Y, Roskin KM, Jackson KJL, Dixit VP, Lee J-Y, et al. Convergent antibody signatures in human dengue. Cell Host Microbe. 2013;13:691–700.

51. Davis CW, Jackson KJL, McElroy AK, Halfmann P, Huang J, Chennareddy C, et al. Longitudinal analysis of the human B cell response to Ebola virus infection. Cell. 2019;177:1566–1562.e17.

52. Robbiani DF, Gaebler C, Muecksfisch F, Lorenzi JCC, Wang Z, Cho A, et al. Convergent antibody responses to SARS-CoV-2 in convalescent individuals. Nature. 2020;584:437–42.

53. Nielsen SCA, Yang F, Jackson KJL, Hoh RA, Rölögten K, Jean GH, et al. Human B cell clonal expansion and convergent antibody responses to SARS-CoV-2. Cell Host Microbe. 2020;28:516–525.e5.

54. Dunand CJH, Wilson PC. Restricted, canonical, stereotyped and convergent immunoglobulin responses. Philos Trans R Soc Lond B Biol Sci. 2015;370:20140238.

55. Akbar R, Robert PA, Pavlovic M, Jeliazkov JR, Snapkov I, Slabodkin A, et al. A compact vocabulary of paratope-epitope interactions enables predictability of antibody-antigen binding. Cell Rep. 2021;34:108856.

56. Zhou T, Georgiev I, Wu X, Yang Z-Y, Dai K, Firai A, et al. Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. Science. 2010;329:811–7.

57. Li Y, O’Dell S, Walker LM, Wu X, Guenaga J, Feng Y, et al. Mechanism of neutralization by the broadly neutralizing HIV-1 monoclonal antibody VRC01. J Virol. 2011;85:8954–67.

58. Zhou T, Lynch RM, Chen L, Acharya P, Wu X, Doria-Rose NA, et al. Structural basis for the V1-V2 region of the HIV-1 gp120 envelope glycoprotein. Proc Natl Acad Sci USA. 2012;109:E2083–90.

59. Yacoob C, Pancera M, Vigdorovich V, Oliver BG, Glenn JA, Feng J, et al. Differences in allelic frequency and CDR3 region limit the engagement of HIV Env immunogens by putative VRC01 neutralizing antibody precursors. Cell Rep. 2016;17:1560–70.

60. Scharf L, West AP Jr, Dioksin R, Nussenzwerg MC, Bjorkman PJ. Structural basis for germline usage of a potential class of antibodies targeting the CD4 binding site of HIV-1 gp120. Proc Natl Acad Sci USA. 2003;100:E2083–90.

61. Yacoob C, Pancera M, Vgdorovich V, Oliver BG, Glenn JA, Feng J, et al. Differences in allelic frequency and CDR3 region limit the engagement of HIV Env immunogens by putative VRC01 neutralizing antibody precursors. Cell Rep. 2016;17:1560–70.

62. Scharf L, West AP Jr, Gao H, Lee T, Schedt JF, Nussenzwerg MC, et al. Structural basis for HIV-1 gp120 recognition by a germ-line version of a broadly neutralizing antibody. Proc Natl Acad Sci USA. 2013;110:6049–54.

63. Scharf L, West AP, Sievers SA, Chen C, Jiang S, Gao H, et al. Structural basis for germline antibody recognition of HIV-1 immunogens. Elife. 2016. https://doi.org/10.7554/eLife.08738.

64. Foquet L, Hermsen CC, van Gemert G-J, Van Braeckel E, Weening KE, Sauerwein WH, et al. Reversal of cryptic sites for canonical N-linked glycosylation of the immunoglobulin variable region. Proc Natl Acad Sci USA. 2000;97:12144–8.

65. Iversen R, Di Niro R, Stammes J, Lindeman I, Lundin KEA, et al. High abundance of plasma cells secreting trisaccharide 2-Specific IgA auto-antibodies with limited somatic hypermutation in celiac disease intestinal lesions. Cell. 2016;177:1566–70.

66. Iversen R, Ni R, Stammes J, Lindem I, Lundin KEA, Wilson PC. Solid-LM. Transglutaminase 2-specific autoantibodies in celiac disease target clustered, N-terminal epitopes not displayed on the surface of cells. J Immunol. 2013;190:5981–8.

67. Steinsbo O, Oqvi E, Chandrasekaran R, Huang M, Salgado-Ferrer M, Lundin KEA, et al. Restricted VH/VL usage and limited mutations in gluten-specific IgA of coeliac disease lesion plasma cells. Nat Commun. 2014;5:4041.

68. Lindeman I, Zhau C, Eggesbo LM, Miao Z, Polak J, Lundin KEA, et al. Longevity, clonal relationship, and transcriptional program of celiac disease-specific plasma cells. J Exp Med. 2021. https://doi.org/10.1084/jem.20200852.

69. Snir O, Chen X, Gidoni M, Fr MF du, Zhao Y, Steinsbo O, et al. Stereotyped antibody responses target posttranslationally modified gluten in celiac disease. JCI Insight. 2017. https://doi.org/10.1172/jci.insight.93961.

70. Saboury Z, Schofeld P, Horikawa K, Spierings E, Kipling D, Randall KL, et al. Redemptions of autoantibodies on anergic B cells by variable-region glycosylation and mutation away from self-reactivity. Proc Natl Acad Sci USA. 2014;111: E2567–75.

71. Wallick SC, Kabat EA, Morrison SL. Glycosylation of a VH residue of a monoclonal antibody against alpha (1-6) dextran increases its affinity for antigen. J Exp Med. 1998;188:1099–109.

72. Leibiger H, Wüstner D, Stigler RD, Marx U. Variable domain-linked oligosaccharides of a human monoclonal IgG: structure and in vivo function analyses reveal key features in Staphylococcus aureus IsdB-associated autoantibodies. Cell Immunol. 2018;333:65–72.

73. Bowden CFM, Chan ACK, Li EYW, Arrieta AL, Eltis LD, Murphy ME. Structure-function analyses reveal key features in Staphylococcus aureus IsdB-associated autoantibodies. Cell Immunol. 2018;333:65–72.
Huber R, Deisenhofer J, Colman PM, Matsushita M, Palm W. Crystallographic structure studies of an IgG molecule and an Fc fragment. Nature. 1976;264:415–20.

de Taeye SW, Bentlage AEH, Mebius MM, Meesters JI, Lissenberg-Thunnissen S, Falcik D, et al. FcγR Binding and ADCC Activity of Human IgG Allotypes. Front Immunol. 2020;11:740.

Lefranc M-P, Lefranc A, Human Gm, Km, and Am allotypes and their molecular characterization: a remarkable demonstration of polymorphism. In: Immunogenetics. Totowa, NJ: Humana Press; 2012, p. 635–80.

Aitherton A, Armour KL, Bell S, Minson AC, Clark MR. The herpes simplex virus type 1 Fc receptor discriminates between IgG1 allotypes. Eur J Immunol. 2000;30:2540–7.

Kotolich S, McKay PF, Chung AW, Kent SJ, Gilmour J, Shattock RJ. Immunoglobulin G1 allotype influences antibody subclass distribution in response to HIV gp140 vaccination. Front Immunol. 2017;8:1883.

Pandey JP, Kistner-Griffin E, Radwan FF, Kaur N, Nambodiri AM, Black L, et al. Immunoglobulin genes influence the magnitude of humoral immunity to cytomegalovirus glycoprotein B. J Infect Dis. 2014;210:1823–6.

Simon B, Weseslintner L, Gözzer I, Pollak K, Jaksch P, Klepetko W, et al. Subclass-specific antibody responses to human cytomegalovirus in lung transplant recipients and their association with constant heavy immunoglobulin G chain polymorphism and virus replication. J Heart Lung Transpl. 2016;35:370–7.

Carson RT, McDonald DF, Kehoe MA, Calvert JE. In Calonga-Solís V, Malheiros D, Beltrame MH, Vargas L de B, Dourado RM, et al. Lindeman I, Polak J, Qiao S-W, Holmøy T, Høglund RA, Vartdal F, et al. Stereocellular malignancies: Old protein learns new tricks. Biochim Biophys Acta. 2019;1861:792.

Falck D, et al. Fc ε Binding and ADCC Activity of Human IgG Allotypes. Front Immunol. 2020;11:740.

Bhardwaj V, Franceschetti M, Rao R, Pevzner PA, Safonova Y. Automated analysis figures were created with BioRender.com.

Vázquez-Arregui K, Tantin D. The Oct1 transcription factor and epithelial gene promoters from mice and humans: implications for transcriptional regulation and repertoire expression. Immunogenetics. 1998;47:183–95.

Vázquez-Arregui K, Tantin D. The Oct1 transcription factor and epithelial malignancies: Old protein learns new tricks. Biochim Biophys Acta. 196,2016;1859:792.

Sun Z, Kitchingman GR. Bidirectional transcription from the human immunoglobulin VH gene promoter. Nucleic Acids Res. 1994;22:861–8.

Pelletier MR, Hatada EN, Scholz G, Scheiderer C. Efficient transcription of an immunoglobulin κ gene requires an upstream fragment containing conserved sequence elements. Nature. 1984;310:71–74.

Bemark M, Liberg D, Leandersson T. Conserved sequence elements in κ promoters from mice and humans: implications for transcriptional regulation and repertoire expression. Immunogenetics. 1998;47:183–95.

Bemark M, Liberg D, Leandersson T. Conserved sequence elements in κ promoters from mice and humans: implications for transcriptional regulation and repertoire expression. Immunogenetics. 1998;47:183–95.

Fackler FG, Zachau HG. Correct transcription of an immunoglobulin κ gene requires an upstream fragment containing conserved sequence elements. Nature. 1984;310:71–74.

Vázquez-Arregui K, Tantin D. The Oct1 transcription factor and epithelial malignancies: Old protein learns new tricks. Biochim Biophys Acta. 196,2016;1859:792.

Sun Z, Kitchingman GR. Bidirectional transcription from the human immunoglobulin VH gene promoter. Nucleic Acids Res. 1994;22:861–8.

Pelletier MR, Hatada EN, Scholz G, Scheiderer C. Efficient transcription of an immunoglobulin κ gene requires a specific sequence elements overlapping with and downstream of the transcriptional start site. Nucleic Acids Res. 1997;25:3995–4003.

Haryadi R, Ho S, Kok YJ, Pu HK, Zheng L, Pereira NA, et al. Optimization of heavy chain and light chain signal peptides for high level expression of therapeutic antibodies in CHO cells. PLoS One. 2015;10:e0116878.

Gibson SJ, Bond NJ, Milne S, Lewis A, Sheriff A, Pettman G, et al. N-terminal or alternative splice variants and germline polymorphisms in human immunoglobulin light chain genes. bioRxiv. 2021;2021.02.05.429934.

Fackler FG, Zachau HG. Correct transcription of an immunoglobulin κ gene requires an upstream fragment containing conserved sequence elements. Nature. 1984;310:71–74.

198.3.

Smajlović I, Mikocziova et al. Human Gm, Km, and Am allotypes and their molecular characterization: a remarkable demonstration of polymorphism. In: Immunogenetics. Totowa, NJ: Humana Press; 2012, p. 635–80.

Bemark M, Liberg D, Leandersson T. Conserved sequence elements in κ promoters from mice and humans: implications for transcriptional regulation and repertoire expression. Immunogenetics. 1998;47:183–95.

Vázquez-Arregui K, Tantin D. The Oct1 transcription factor and epithelial malignancies: Old protein learns new tricks. Biochim Biophys Acta. 196,2016;1859:792.

Sun Z, Kitchingman GR. Bidirectional transcription from the human immunoglobulin VH gene promoter. Nucleic Acids Res. 1994;22:861–8.

Pelletier MR, Hatada EN, Scholz G, Scheiderer C. Efficient transcription of an immunoglobulin κ gene requires a specific sequence elements overlapping with and downstream of the transcriptional start site. Nucleic Acids Res. 1997;25:3995–4003.

Haryadi R, Ho S, Kok YJ, Pu HK, Zheng L, Pereira NA, et al. Optimization of heavy chain and light chain signal peptides for high level expression of therapeutic antibodies in CHO cells. PLoS One. 2015;10:e0116878.

Gibson SJ, Bond NJ, Milne S, Lewis A, Sheriff A, Pettman G, et al. N-terminal or alternative splice variants and germline polymorphisms in human immunoglobulin light chain genes. bioRxiv. 2021;2021.02.05.429934.

Fackler FG, Zachau HG. Correct transcription of an immunoglobulin κ gene requires an upstream fragment containing conserved sequence elements. Nature. 1984;310:71–74.

198.3.

Smajlović I, Mikocziova et al. Human Gm, Km, and Am allotypes and their molecular characterization: a remarkable demonstration of polymorphism. In: Immunogenetics. Totowa, NJ: Humana Press; 2012, p. 635–80.

Bemark M, Liberg D, Leandersson T. Conserved sequence elements in κ promoters from mice and humans: implications for transcriptional regulation and repertoire expression. Immunogenetics. 1998;47:183–95.

Vázquez-Arregui K, Tantin D. The Oct1 transcription factor and epithelial malignancies: Old protein learns new tricks. Biochim Biophys Acta. 196,2016;1859:792.

Sun Z, Kitchingman GR. Bidirectional transcription from the human immunoglobulin VH gene promoter. Nucleic Acids Res. 1994;22:861–8.

Pelletier MR, Hatada EN, Scholz G, Scheiderer C. Efficient transcription of an immunoglobulin κ gene requires a specific sequence elements overlapping with and downstream of the transcriptional start site. Nucleic Acids Res. 1997;25:3995–4003.

Haryadi R, Ho S, Kok YJ, Pu HK, Zheng L, Pereira NA, et al. Optimization of heavy chain and light chain signal peptides for high level expression of therapeutic antibodies in CHO cells. PLoS One. 2015;10:e0116878.

Gibson SJ, Bond NJ, Milne S, Lewis A, Sheriff A, Pettman G, et al. N-terminal or alternative splice variants and germline polymorphisms in human immunoglobulin light chain genes. bioRxiv. 2021;2021.02.05.429934.
COMPETING INTERESTS
VG declares advisory board positions in aiNET GmbH and Enpicom B.V. VG is a consultant for Roche/Genentech.

ADDITIONAL INFORMATION
Correspondence and requests for materials should be addressed to L.M.S.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.