



Immunoglobulin Light Chain Gene Rearrangements, Receptor Editing and the Development of a Self-Tolerant Antibody Repertoire

Andrew M. Collins^{1*} and Corey T. Watson²

¹ School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW, Australia, ² Department of Biochemistry and Molecular Genetics, University of Louisville School of Medicine, Louisville, KY, United States

OPEN ACCESS

Edited by:

Gur Yaari,
Bar-Ilan University, Israel

Reviewed by:

Gregory C. Ippolito,
University of Texas at Austin,
United States
Peter Daniel Burrows,
University of Alabama at Birmingham,
United States

*Correspondence:

Andrew M. Collins
a.collins@unsw.edu.au

Specialty section:

This article was submitted to
B Cell Biology,
a section of the journal
Frontiers in Immunology

Received: 15 June 2018

Accepted: 10 September 2018

Published: 08 October 2018

Citation:

Collins AM and Watson CT (2018)
Immunoglobulin Light Chain Gene
Rearrangements, Receptor Editing
and the Development of a
Self-Tolerant Antibody Repertoire.
Front. Immunol. 9:2249.
doi: 10.3389/fimmu.2018.02249

Discussion of the antibody repertoire usually emphasizes diversity, but a conspicuous feature of the light chain repertoire is its lack of diversity. The diversity of reported allelic variants of germline light chain genes is also limited, even in well-studied species. In this review, the implications of this lack of diversity are considered. We explore germline and rearranged light chain genes in a variety of species, with a particular focus on human and mouse genes. The importance of the number, organization and orientation of the genes for the control of repertoire development is discussed, and we consider how primary rearrangements and receptor editing together shape the expressed light chain repertoire. The resulting repertoire is dominated by just a handful of IGKV and IGLV genes. It has been hypothesized that an important function of the light chain is to guard against self-reactivity, and the role of secondary rearrangements in this process could explain the genomic organization of the light chain genes. It could also explain why the light chain repertoire is so limited. Heavy and light chain genes may have co-evolved to ensure that suitable light chain partners are usually available for each heavy chain that forms early in B cell development. We suggest that the co-evolved loci of the house mouse often became separated during the inbreeding of laboratory mice, resulting in new pairings of loci that are derived from different sub-species of the house mouse. A resulting vulnerability to self-reactivity could explain at least some mouse models of autoimmune disease.

Keywords: immunoglobulin light chain, receptor editing, self-tolerance, antibody repertoire, V(D)J rearrangement, models of autoimmune disease, sub-species of the house mouse

INTRODUCTION

The success of the humoral arm of the adaptive immune system depends upon a diversity of antibody specificities within an individual's population of circulating B cells. This diversity is made possible by the process of gene recombination that takes place during B cell development, creating functional antibody heavy and light chain V(D)J transcripts from relatively small sets of Variable (V), Diversity (D), and Joining (J) genes. The basic processes underlying V(D)J recombination are now well understood (1, 2) and recently, thanks to advances in sequencing technologies that allow millions of different V(D)J gene rearrangements to be explored in a single individual, much has been learnt about the nature of the expressed antibody repertoire (3–8). Most repertoire studies, however, have focused upon the heavy chain repertoire. The nature of the light chain repertoire is less clear.

The diversity of the antibody repertoire is a consequence of the permutations of heavy chain V, D, and J genes, and light chain V and J genes, that are possible given the size of these sets of genes within the genome, and of the permutations of heavy and light chain pairings. This component of the overall diversity is referred to as combinatorial diversity, and is a simple reflection of the number of available heavy chain genes and κ and λ light chain genes. Additional diversity is generated during the recombination processes by imprecise joining at the V(D)J junctions. This is referred to as junctional diversity, and is principally determined by the extent to which random nucleotides are inserted between joining genes (4, 6).

In this review, we highlight important consequences for repertoire development that result from the organization of light chain genes within the mammalian genome. In particular, this organization facilitates repeated rounds of light chain gene rearrangement through the process of receptor editing. This helps to ensure that virtually all developing B cells successfully generate productive light chain rearrangements.

A number of biases and constraints are discussed which lead to substantially less diversity in the light chain repertoire than is usually calculated, and this limited diversity appears to be present in a wide range of species. We conclude that diversity is not the *raison d'être* of the light chain repertoire. In light of substantial evidence for a special role for light chains in autoimmune reactivity, we propose that the co-evolution of heavy and light chain genes has resulted in a limited light chain repertoire that usually serves to avoid self-reactivity. This hypothesis is explored through an examination of the generation of light chain repertoires in inbred mouse strains that are widely used in models of autoimmune disease.

THE NUMBER AND ORGANIZATION OF LIGHT CHAIN GENES WITHIN THE MAMMALIAN GENOME

To properly understand how the heavy and light chain repertoires are generated, it is essential to have a detailed knowledge of the number of rearranging germline genes that give rise to these repertoires, and of their organization within the genome. The number of genes per species is highly variable as a result of dynamic evolutionary processes in these complex gene families. This can be appreciated by examining the phylogenetic relationships among genes within and between species, and is demonstrated in a phylogeny of functional human and mouse heavy and light chain variable genes (**Figure S1**). However, our understanding of the precise evolutionary histories of these genes across a larger range of species remains limited, largely due to a paucity of available genomic data.

The organization of the light chain genes is particularly complex, and quite different to that of the heavy chain genes. Heavy chain genes are found within a single gene locus (IGH), while light chain genes are generally found as two separate gene loci—the κ locus (IGK) and the λ locus (IGL). These two loci are found in virtually all mammalian species, while loci for

these and other light chain variants are found in bony and even cartilaginous fish (9, 10). Such a general distribution of light chain genes between separate loci is intriguing, and suggests that this genomic organization may carry evolutionary advantages.

Within the κ chain loci of humans, mice and most other species, genes are organized in a similar fashion to the genes of the heavy chain locus (11–15). That is, a cluster of IGKV genes are found 5' of a small number of IGKJ genes, with the IGKJ gene cluster located 5' of a single IGKC gene. The dog genome is unusual in that half the canine IGKV genes are located upstream, and half are located downstream of the IGKJ and IGKC genes (16).

The number of functional IGKV genes varies widely between species, and this number may have some relationship to species size (**Figure S1**). We have argued that small species may require more germline genes because of the small burst size of the germinal center reaction in those species (17). As the number of cells responding to antigen is limited in small species, there is less chance for important higher affinity antibodies to emerge from the germinal center through the process of somatic point mutation. Critical specificities must therefore be encoded in the germline.

Sequencing of the human κ locus has identified 44 functional IGKV genes and open reading frames, which are found in two clusters that arose through segmental duplication (18, 19). An additional three functional IGKV sequences may be present in some haplotypes (20). A comparable number of functional IGKV genes ($n = 54$) was recently characterized in genomic sequences from the rhesus macaque, a commonly used non-human primate model (21). In contrast, studies of the horse genome reference sequence identified only 19 apparently functional IGKV genes (22), while 111 and 135 potentially functional IGKV genes have been found in the guinea pig and rat genome reference sequences, respectively (23, 24). Among mammalian species studied to date, the microbat (*Myotis lucifugus*) is unique in that it lacks a κ locus (25). It has been suggested that this may be part of a general simplification of immune function in a species that has met the weight to muscle challenge that is necessary for flight capability (26). This hypothesis is indirectly supported by the fact that the κ locus is also absent from the genome of chickens (25) and zebra finches (27), and it may have been lost from the genomes of all bird species.

In line with numbers reported in the small rodent species mentioned above, the genome of C57BL/6 mice carries around 101 functional IGKV genes (28). This number may, however, not be accurate for other inbred mouse strains. We recently reported that the heavy chain loci of different inbred strains of mice are derived from different sub-species of the house mouse. As a result, the C57BL/6 strain carries 99 functional IGHV genes, but the BALB/c strain carries 163 functional IGHV genes (29). We have also noted that based on available whole-genome SNP data (30), almost all inbred strains carry κ chain loci that appear to be derived from the *Mus musculus domesticus* sub-species (17). NOD/ShiLtJ mice are unusual in that they carry a *M. m. castaneus*-derived κ locus (17). Interestingly, many of the distinct κ genes of this diabetes-prone strain are identical to κ

genes of the Systemic Lupus Erythematosus (SLE)-prone (NZB \times NZW)F1 and MRL strains (31). SNP analysis shows that parts of the IGKV loci of NZB and MRL mice are also derived from the *M. m. castaneus* sub-species (**Figure 1**). This confirms an observation from a very early study of BALB/c and NZB-derived myeloma proteins. It was reported that NZB and BALB/c mice share some κ light chain sequences, but that other κ genes differ markedly between the strains (32). SNP analysis shows that the chr6:67.5m–68.2m region of the κ locus of the NZB strain is of *M. m. castaneus* origin. The chr6:68.2m–68.5m region is of uncertain origin, and the remainder of the NZB κ locus is *M. m. domesticus*-derived. Genes of *M. m. castaneus* origin are also found in the MRL strain, in the region chr6:68.8m–70.7m (see **Figure 1**).

The λ locus of most species investigated to date includes a set of IGLV genes that are located 5' from a variable number of tandem cassettes, each made up of an IGLJ gene and an IGLC gene. The human locus includes as many as 38 functional IGLV genes and Open Reading Frames (18) and five functional J-C gene pairs (34). In the rhesus macaque, 47 IGLV genes are predicted to be functional based on genomic data (21), and in the pig, there are nine functional IGLV genes (35). In these species too, the IGKV genes are located 5' of functional J-C pairs, but this organization is not invariant. 144 IGLV genes of uncertain functionality have been identified in the horse, with 110 genes being located upstream and 34 being located downstream of the IGLJ/IGLC cluster (22, 36). The locus of the mouse is also differently organized.

The C57BL/6 genome includes just three IGLV genes (**Figure S1**), and there has been speculation that IGLV genes might have been lost during the inbreeding of laboratory strains. In fact, diversity is lacking in wild mice of all three *Mus musculus* subspecies (37). Two of the C57BL/6 IGLV genes are associated with one functional J-C gene pair, while the third IGLV gene is associated with a second J-C pair. Lambda rearrangement in the mouse takes place within each of the two VJC units, with little or no recombination between the units (38).

The genes of both the human and mouse λ IGLV loci are all in the same transcriptional orientation as the λ J-C gene clusters (18, 39). The V, D, and J genes of the heavy chain loci of mammalian species are also found in the same orientation as their associated constant region genes (40–42). The κ chain locus of these species, on the other hand, includes many IGKV genes that are found in the opposite orientation to their associated IGKJ and IGKC genes. In the human, the orientation of the distal κ gene cluster is opposite to that of the IGKJ genes and IGKC gene, while all but the two most 3' genes of the proximal gene cluster share their orientation with the IGKJ and IGKC genes (20). The κ locus of the mouse also includes IGKV genes in both the same and the opposite orientation to their respective IGKJ and IGKC genes (13). Such variable orientations of IGKV genes have also been reported in other species including the elephant (43), horse (22), pig (14), dog (16), and rhesus macaque (21).

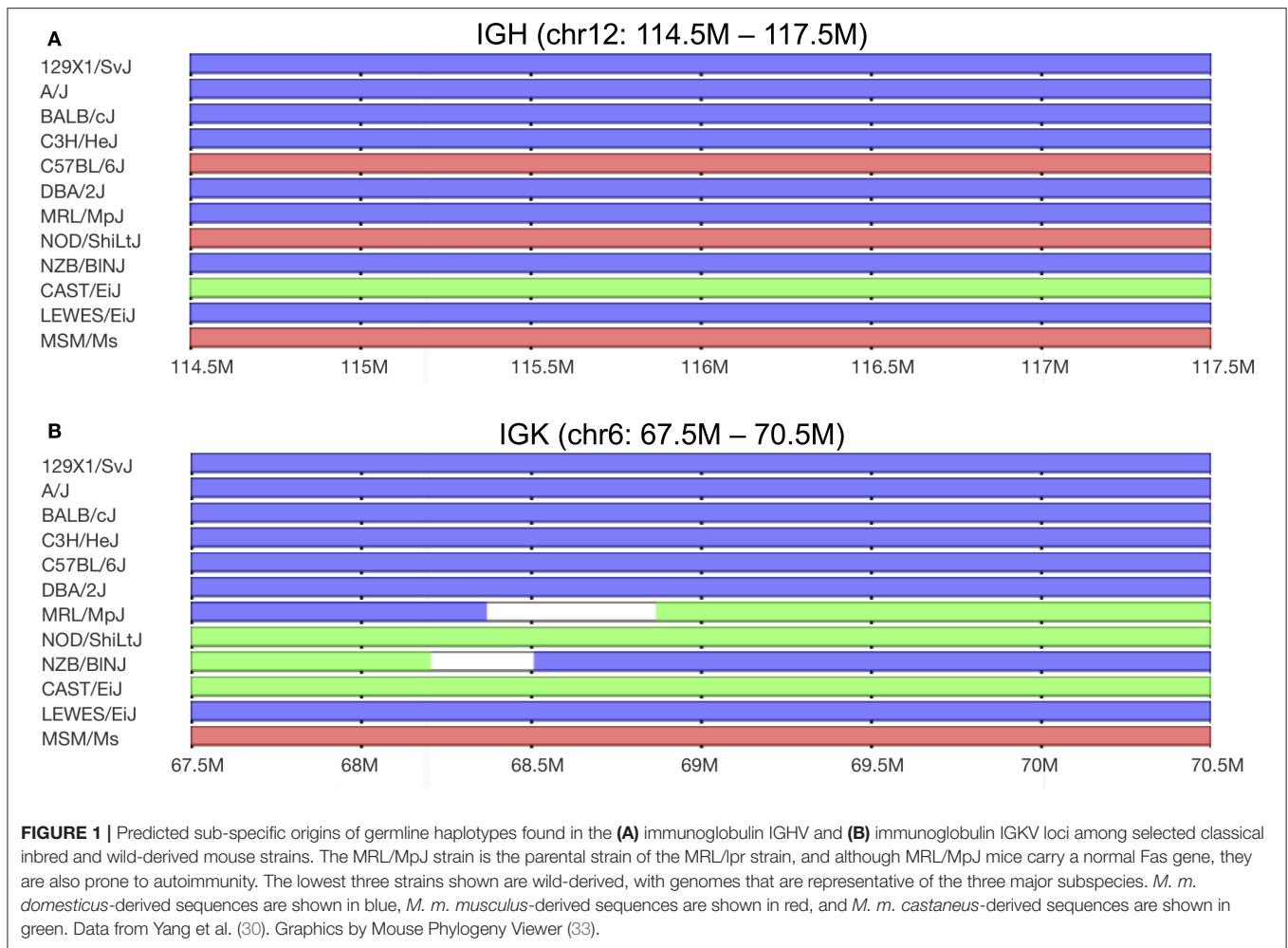
In the horse, which is a species with a λ -dominant repertoire, IGLV genes are found both upstream and downstream of the λ

J-C gene clusters. Many of these sequences are pseudogenes, but the few functional genes in the downstream cluster are found in the opposite orientation to that of the horse J-C genes, thereby allowing the genes to recombine (36).

The orientation of genes has other consequences for the generation of diversity. The opposite orientation of many IGKV genes within the murine and human κ loci means that primary rearrangements of such genes do not lead to deletion of the genes that are located between recombining IGKV and IGKJ genes (see **Figure 2**). This retention of genes becomes important if a rearrangement results in a non-productive chain or a self-reactive B cell receptor (BCR). In such situations, all other IGKV and IGKJ germline genes remain available for secondary rearrangements (see discussion below). In any cell that experiences such successive rounds of recombination, the order and orientation of the genes within the locus will be subject to complex changes, and this will have consequences for the repertoire that is generated by secondary rearrangements (**Figure 2**).

The frequencies with which different V, D, and J genes are utilized in gene rearrangements vary by many orders of magnitude. This appears to reflect, at least in part, the accessibility of genes, and their positions within the genome (28, 44, 45). Dramatic changes in the order of genes and in the distances between IGKV and IGKJ genes, arising from a primary rearrangement of genes, should therefore lead to changes to gene accessibility. This may mean that the utilization frequency of a gene can vary between primary, secondary and subsequent rearrangements. Complex changes could therefore compromise the tight regulation that otherwise appears to guide the generation of the antibody repertoire. In many species, this risk to the regulation of the repertoire may be mitigated by the action of Kappa Deleting Elements (KDE). The consecutive rearrangements that are possible within the κ locus can be terminated by KDE-mediated recombination, driving B cells to the expression of genes of the λ locus (46). This may also prevent or lead to the resolution of allelic inclusion, which can arise because of the orientation of IGKV genes within the locus (see **Figure 2**) (47).

Kappa Deleting Elements (KDE) are located downstream of IGKC genes, and they appear to be highly conserved within the mammalian genome (48). The mouse KDE is referred to as the Recombining Segment (RS), and it is distinct but very similar to the Recombination Signal Sequences (RSS) located adjacent to the 3' ends of the IGKV genes and the 5' ends of the IGKJ genes (49). KDEs of all species studied are made up of conserved heptamer and nonamer sequences separated by 23 base pair spacers (48). The KDEs function by allowing recombination between the KDEs and recombining elements that contain the palindromic heptamer CACAGTG. These are located within the IGKJ-IGKC intron and in the RSS at the 3' ends of the IGKV genes. Such recombination effectively terminates the involvement of the rearranging locus in the generation of diversity. This will drive recombination from the first to the second κ locus (i.e., on the alternate chromosome), or from the second κ locus to the λ gene-bearing chromosomes (see



Figures 2, 3). It is likely that despite the conservation of this element within the κ locus, the strength of action of the elements varies between species. The preponderance of κ chains in the expressed antibody repertoire of the mouse, for example, suggests that the mouse RS usually fails to drive rearrangement to the λ locus. Instead, each murine κ locus will likely be rearranged to exhaustion, and this will prevent the overexpression of the handful of λ genes that remain in the mouse genome. The activities of RS in different sub-species of the house mouse have not been explored.

Gene Rearrangement of the Light Chain Loci

The light chain repertoire is shaped by the order of gene rearrangement, and early studies in the mouse and human showed that rearrangement begins with the κ locus (34, 50). This may not be true for all species. It has recently been shown that the λ locus rearranges first in the fetal and neonatal pig (51, 52). Timing therefore requires further investigation, particularly in species with repertoires that are dominated by the λ light chain, for regulation of the expressed repertoire could be more difficult if the minor locus was to rearrange

first. If a species had just a handful of functional κ genes, and abundant functional λ genes, initial rearrangements of the κ locus would risk over-expression of the few available IGKV genes.

In the mouse and human, if an initial κ rearrangement is unproductive or self-reactive, additional rounds of secondary rearrangement can proceed, in a process known as receptor editing (53–55). Receptor editing is usually discussed as a pathway to resolution of auto-reactivity, either in developing B cells in which self-reactivity is generated by primary rearrangements (56), or in mature antigen-selected B cells where self-reactivity may result from somatic point mutations (57). Less attention has been paid to the more general role that receptor editing plays in shaping the formation of the repertoire.

The organization of genes within the light chain loci facilitates receptor editing, and this increases the likelihood that each B cell will form an in-frame light chain rearrangement (58). As long as unrearranged V genes remain 5' of a VJ rearrangement, and unrearranged J genes remain 3' of the rearrangement, receptor editing can continue (see **Figure 2**). In the mouse, the potential of κ chain receptor editing is maximized by a bias toward rearrangement of the 5' IGKJ1 gene (59), and this targeting

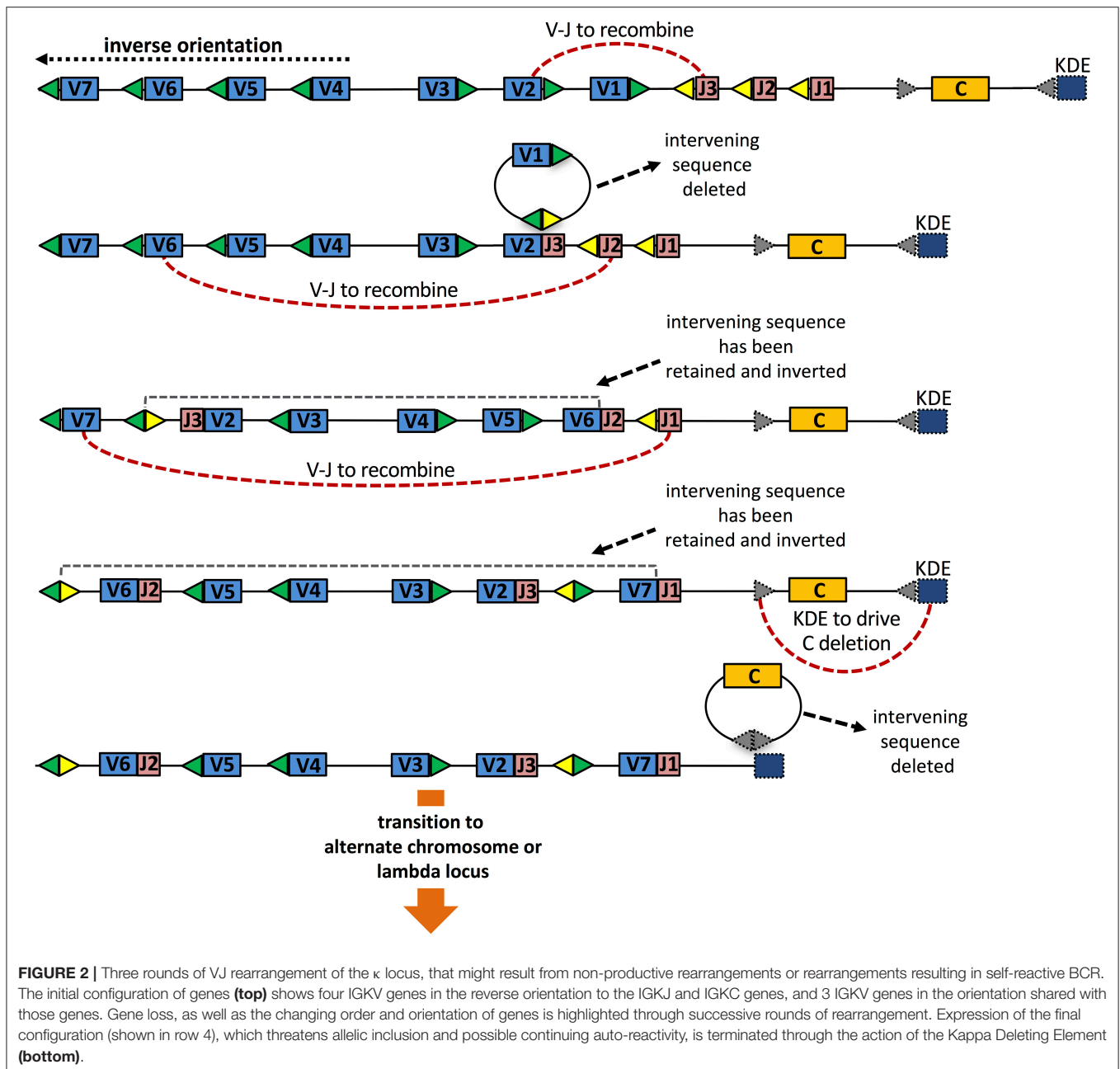


FIGURE 2 | Three rounds of VJ rearrangement of the κ locus, that might result from non-productive rearrangements or rearrangements resulting in self-reactive BCR. The initial configuration of genes (top) shows four IGKV genes in the reverse orientation to the IGKJ and IGKC genes, and 3 IGKV genes in the orientation shared with those genes. Gene loss, as well as the changing order and orientation of genes is highlighted through successive rounds of rearrangement. Expression of the final configuration (shown in row 4), which threatens allelic inclusion and possible continuing auto-reactivity, is terminated through the action of the Kappa Deleting Element (bottom).

results from the action of the proximal IGKJ germline transcript promoter (60).

A process of serial rearrangement of the κ chain locus may continue on one chromosome until all possibilities of recombination have been exhausted. Recombination will then proceed on the second κ chromosome (Figure 3). A failure to produce a productive, self-tolerant rearrangement on the second chromosome, after multiple rounds of rearrangement, will be followed by rearrangement of the λ loci.

The human λ locus also seems permissive of receptor editing (61), and the absence of deletional elements in the λ locus should maximize the potential of serial λ recombination in the

human. This should ensure that relatively few human B cells fail to make a suitable productive light chain rearrangement that is self-tolerant when expressed in association with the cell's heavy chain rearrangement. The possibility of repeated rounds of λ rearrangement could be particularly important for the avoidance of self-reactivity, for it has been suggested that λ -bearing human B cells are less prone to self-reactivity than κ -bearing B cells (61). The λ chains of these cells may also provide stability during an ongoing immune response, for it has been shown that the codon usage of λ genes reduces the likelihood of structural changes arising from accumulating somatic point mutations (62).

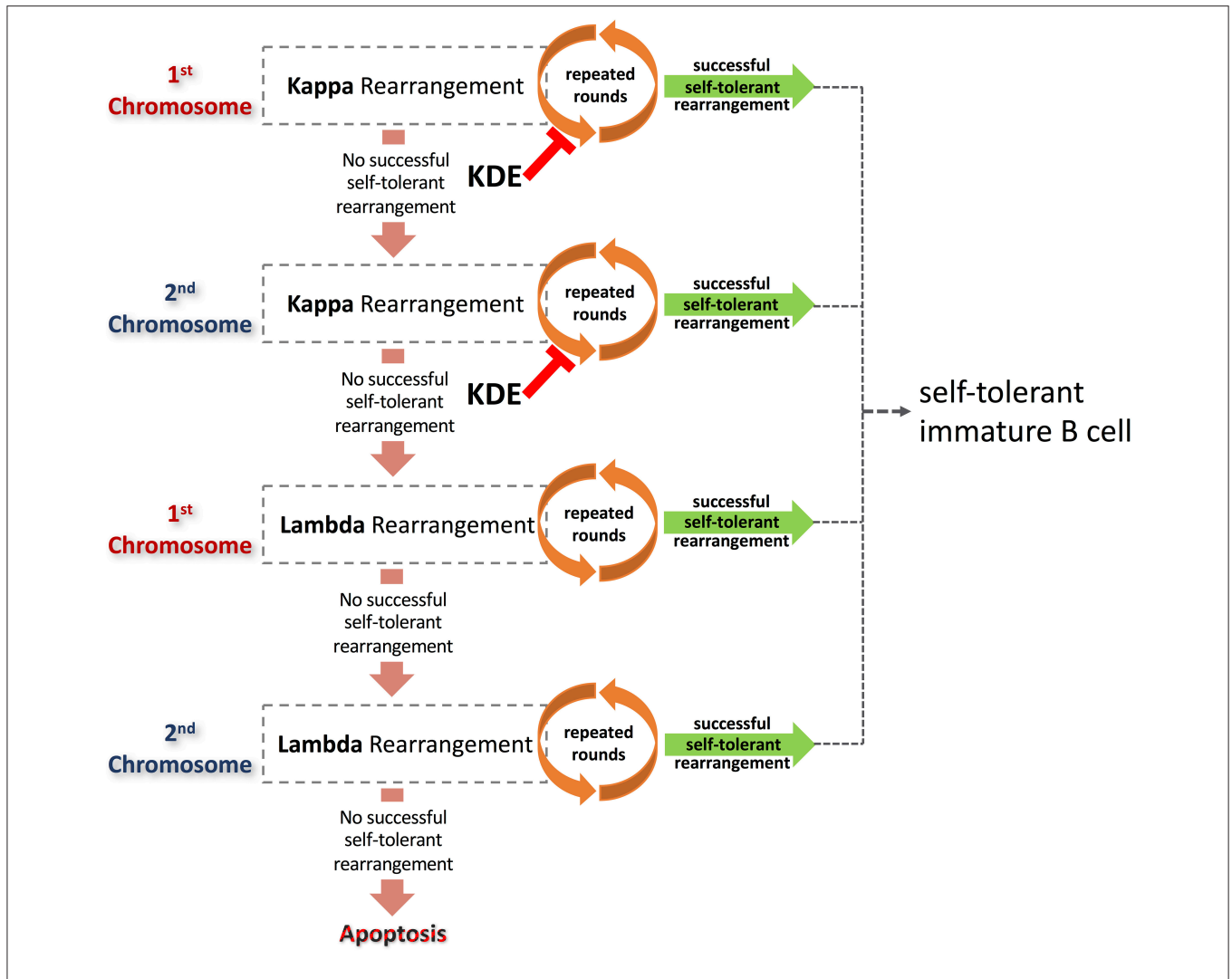


FIGURE 3 | Light chain rearrangements provide multiple pathways to the production of self-tolerant B cells. Alternative pathways are a consequence of rearrangements of the light chain loci that result in non-productive or auto-reactive BCR. Beginning with rearrangements of the κ locus on the first selected chromosome, a succession of light chains can be paired with an existing heavy chain that is already expressed on the surface of the pre-B cell. Each resulting BCR is assessed for affinity to self-antigen. Cells expressing autoreactive antibodies can be rescued via further rounds of receptor editing. If repeated rounds of rearrangements of a locus fail to generate a functional, self-tolerant antibody, and the possibilities of rearrangement are exhausted, the process may continue on other chromosomes. Rounds of κ rearrangement may be prematurely curtailed by the action of the kappa deleting element. B cells unable to generate self-tolerant antibodies despite multiple rounds of receptor editing will ultimately be deleted or rendered anergic. In contrast, B cells that generate successful light chain rearrangements that result in self-tolerant antibodies will go on to develop further into immature B cells.

Population Variation in the κ and λ Gene Loci

Combinatorial diversity is expanded by heterozygous gene loci, and such diversity appears to be of functional significance (63). It is therefore important that repertoire studies include a focus on alleles and gene heterozygosity. Although a few new allelic variants of human IGKV genes have recently been reported (19), the reported IGKV germline gene repertoire appears to be relatively complete (64). According to the IMGT reference directory, 26 IGKV genes have no known allelic variants, while 15 IGKV genes have only one reported variant and five have two known variants. The extent of allelic variation within the κ light

chain locus could be even less than is indicated in the IMGT reference directory, for some of the reported variants are likely to be artifacts arising from sequencing errors. This is certainly the case for many reported IGHV alleles that were identified in early sequencing studies (65).

The reported human IGLV germline gene repertoire may also be relatively complete, for only five new alleles have been reported since 1997. These sequences are more varied than genes and allelic variants of the IGH and κ loci (66), but like the IGKV repertoire, there appears to be relatively little allelic variation amongst the IGLV genes. Functional and ORF allelic variants have been reported for 24 IGLV genes, but not for 15 other

IGLV genes. No more than four alleles are identified in the IMGT reference directory for any IGLV gene (<http://www.imgt.org/vquest/refseqh.html>).

In contrast to the genes of the κ and λ loci, there is just a single functional IGHV gene (IGHV3-NL1) that lacks reported allelic variants in the IMGT reference directory or in the IgPdb database (<https://cgi.cse.unsw.edu.au/~ihimmune/IgPdb/>). So many common variants are known for some genes that heterozygosity in any individual is almost assured. For example there are 16 *IGHV1-69* gene sequences in the IMGT reference directory, and a further 13 alleles have been inferred from analysis of high throughput genomic and AIRR-seq data (67, 68). Although the larger number of IGHV allelic variants could reflect the greater attention that has been given to defining this set of germline genes, there is additional evidence that points to a lack of diversity in the light chain gene repertoire.

A lack of allelic variation in the human κ locus is supported by AIRR-seq studies of κ rearrangements. In a study of four individuals, involving the dominant three human IGKV gene families (IGKV1, IGKV2 and IGKV3), VJ rearrangements were seen involving between 20 and 25 genes (69). One individual was homozygous at all gene loci. In the three other individuals, heterozygosity was only seen at 1 or 2 of the IGKV loci (69). The contrast with the heavy chain locus is striking. A recent AIRR-seq study of 95 individuals explored heterozygosity at 50 heavy chain IGHV gene loci (70). Other than in three individuals from whom relatively few sequences were generated, study participants were shown to be heterozygous at between 20% and 40% of the loci. Six gene loci were heterozygous in over 50% of study participants. Only six genes that were relatively abundantly present in the datasets showed homozygosity in all individuals (70). Similar patterns of heterozygosity within IGHV coding segments have also been noted from targeted genomic sequencing data (67).

In addition to allelic variation, gene copy number variation is also enriched in the IGHV locus, relative to IGLV and IGKV. Greater than half of the known human functional/ORF IGHV genes have evidence of copy number variation (45, 67, 70–75), compared to only one and three IGLV and IGKV genes, respectively (76–78).

Additional albeit indirect evidence for an evolutionary drive to conserve rather than diversify the human κ locus comes from the similarity of the genes in the proximal and distal IGKV clusters. The large segmental duplication that gave rise to the human κ locus appears to have occurred since the divergence of the human lineage from the most recent shared ancestor with other great apes (11). There are 23 functional IGKV genes and ORFs in the proximal cluster, and 22 in the distal cluster. Eighteen paired sequences are found in both clusters, and no coding changes have evolved at eight of these paired gene loci. In addition, one sequence in each of two other sequence pairs are now non-expressible pseudogenes. Expressed variation is therefore concentrated in just 8 of the 18 sequences. Furthermore, comparisons of nucleotide variation across the entirety of the sequence comprising the large proximal and distal gene clusters reveal strong similarity. Diversity within the large

segmental duplications harboring these gene clusters appears to be much lower on average (>6 fold less) than that observed within segmental duplications found in the IGHV locus (19). We have speculated that this lack of diversity in IGKV may be the result of homogenizing effects of gene conversion events between the proximal and distal regions, as such events have been explicitly documented (19). We also reported that locus-wide IGHV diversity is \sim 3-fold higher than IGLV diversity; in fact, IGHV diversity appears to be generally higher than the genome average (19). Earlier analyses based on limited datasets have suggested that nucleotide and amino acid substitution patterns within V segments may differ between IGHV, IGKV, and IGLV loci (79); specifically, and consistent with decreased genomic diversity in κ locus haplotypes, Schwartz and Hershberg showed that, relative to κ chain V segments, heavy and λ chain genes exhibit greater amino acid diversity in both framework and complementarity determining regions (66). Together, these data suggest contrasting evolutionary histories that have resulted in different genetic features being associated with the human heavy and light chain loci.

The κ locus of the mouse seems to display the same lack of variation that is seen in the human locus. The IGKV locus was first mapped using YACs and BACs derived largely from C57BL/6 and C3H mice (12, 80, 81), and these sequences dominate the IMGT mouse IGKV database. An alternative assembly of the mouse κ locus was later produced based upon data from the C57BL/6, A/J, 129 and DBA/2 strains (82). Each of the IGKV genes previously reported by Zachau and colleagues were mapped to this new assembly, and they were all found to have >99% identity. Not a single allelic variant was reported from this study, although it is true that their approach means that some highly similar but previously unreported polymorphisms may have been overlooked (82).

Evidence of a lack of allelic variation amongst germline genes within mouse strains also comes from analysis of the IMGT database. Studies of light chain germline genes have included a sampling of a wide variety of inbred strains, and from wild-derived *M. m. musculus* and *M. m. castaneus* mice (83–85). Yet the IMGT database includes allelic variants for just 11 functional IGKV genes, and when analysis is confined to reports from studies of strains appearing to carry *M. m. domesticus*-derived genes, variants have only been seen for 6 IGKV genes. Confirmation that the apparent lack of variation is genuine, rather than reflecting insufficient investigation of mouse light chain genes, needs to be pursued through more comprehensive surveys of variation across wild mice from each of the sub-species.

The Diversity of the Expressed Light Chain Repertoire

It is generally held that a stupendous diversity is the defining characteristic of the antibody repertoires of all species. This was famously expressed by Peter Medawar as the miracle of immunology: “that a rabbit yet unborn will be able to make antibodies to an antigen not yet synthesized” (86). We have recently argued that the production of antibodies that target

molecules never before seen, and unlikely to be seen, could be a costly investment for many species (17). The immune repertoires of different species may have developed varying levels of diversity in response to the quite differing evolutionary pressures faced by each species. Some of the most significant pressures may result from basic aspects of the biology of species, including their differing reproductive strategies and longevity, and especially from their varying sizes. The antibody repertoires of small species are necessarily small, and there is therefore a greater need for regulatory processes to steer the development of their repertoires toward specificities that target key pathogens (17). This may explain why in comparison to the human antibody repertoire, the murine repertoire includes more heavy chain clonotypes that are shared by many individuals of the species (6, 17, 87).

Public clonotypes may be rare in the human heavy chain repertoire, but there is a surprising lack of diversity in the human light chain repertoire, and public clonotypes account for about 60% of the human κ (69) and λ (88) light chain repertoires. This is in part a consequence of very strong biases in light chain gene usage. Six IGKV sequences dominate reported human IGK rearrangements: IGKV3-20*01, IGKV3-15*01, IGKV3-11*01, IGKV1-5*01, IGKV2-30*01, and IGKV1-39*01/IGKV1D-39*01 (69). The IGKV3-20*01 gene alone is seen in over 30% of rearrangements in some individuals (69). On the other hand, some genes are utilized at very low frequencies. In fact, amongst 22,193 rearrangements analyzed from four individuals, we saw no sequences that utilized eight reportedly functional IGKV genes (69).

Similarly, while the mouse may have over 100 available IGKV genes, just seven genes are responsible for over 40% of rearrangements, and the utilization frequencies of some IGKV genes are as low as 0.001% (28).

Biased usage of λ IGLV genes is also seen. Three IGLV genes account for more than 50% of human rearrangements, and individual IGLV genes are used at frequencies that range from 0.02 to 27% (89). In the neonatal pig, biases are even more extreme, with three IGLV genes accounting for 70% of rearrangements (51). The utilization of the four functional human IGLJ genes are also affected by biases, with frequencies varying from just 5% for IGLJ1 to almost 55% for IGLJ7 (90, 91).

The lack of D genes in light chain rearrangements limits their diversity. Diversity is further limited by the fact that relatively few nucleotides are lost from κ and λ V and J gene ends by exonuclease removals and few N nucleotides are added to the junction of the joining genes. Public human κ clonotypes have on average just 0.4 added N nucleotides, while even private clonotypes have an average of only 2.5 N additions (69). Similarly, on average, public λ VJ junctions include a single N addition, and private junctions average around two additions (88). There is even less N addition in the mouse (92), and interestingly, this is also true in the humanized mouse (88). This severely limits junctional diversity of the complementarity determining region 3 (CDR3) of light chains in the mouse. Together with the lack of combinatorial diversity, this ensures that the light chain repertoire of the mouse and human are highly constrained. In an analysis of over 250,000 mouse κ chain VJ rearrangements from

59 mice, over 90% of the sequences encoded just 1000 amino acid sequences (28). A similar number of amino acid sequences dominate the human κ chain repertoire (69).

LIGHT CHAINS AND THE CONTROL OF SELF-REACTIVITY

The light chain repertoire is constrained, and there is an extensive body of research that suggests that an important factor that constrains the repertoire is the need for light chain rearrangements to minimize BCR self-reactivity. Human antibodies formed with κ chains may have a greater tendency toward self-reactivity (61), but through repeated rounds of κ rearrangement, and through similar rounds of λ rearrangement, much self-reactivity seems to be avoided. This may explain the recent observations that reduced light chain editing is associated with several autoimmune conditions in the human, including Systemic Lupus Erythematosus (SLE), type 1 diabetes (T1D), and myasthenia gravis (47, 93). This has also been observed in several mouse models of autoimmunity (47). It has also been shown that reduced KDE rearrangements can lead to dual κ and λ chain expression, through a failure to delete κ rearrangements in λ -switched cells, and this disturbance of light chain editing is associated with SLE (94).

The study of cells in which both κ and λ rearrangements are present has highlighted the fact that certain IGKV genes may be prone to self-reactivity. Biases in IGKV gene expression are seen when productive κ rearrangements are studied in λ -bearing B cells, and compared with κ rearrangements from κ -bearing cells (95). This comparison is possible because of the persistence of κ VJ rearrangements in cells that have switched to a λ light chain rearrangement as a consequence of the earlier generation of a self-reactive κ positive BCR. The biased gene expression therefore points to a tendency of some genes to mediate self-reactivity.

Some heavy chain IGHV genes are also associated with autoreactivity, and human *IGHV4-34* in particular has been implicated in anti-red blood cell autoimmune responses (96). It may be, however, that this association should be seen as resulting from a difficulty in finding a suitable light chain partner for *IGHV4-34*. The persistence of *IGHV4-34* in the human population, and its expression at relatively high frequency within the antibody repertoire, points to the value of *IGHV4-34* heavy chains when a self-tolerant light chain partner is found.

Evidence in support of a special role for light chains in the etiology of autoimmune diseases also comes from a consideration of mouse disease models. There are several types of mouse model of autoimmune disease (97). Autoimmunity can be induced by challenging animals with self-antigen in the presence of powerful adjuvants. An example is the Experimental Allergic Encephalitis (EAE) model that involves the challenge of SJL mice with spinal cord homogenate (97). Other models of autoimmune disease involve the spontaneous development of disease. This is the case with the NOD Type 1 Diabetes model and models of SLE using MRL/lpr mice and (NZB \times NZM)F1 mice (97). These spontaneous models may more closely approximate human disease than the antigen challenge models.

A third kind of model relies upon genetic manipulation of animals using gene knockout and transgenic techniques. These models have been particularly important for the study of self-reactive B and T cells, and how they are deleted or otherwise controlled. Some of these models involve the use of transgenic antigen and antibody pairs (eg HEL/anti-HEL) (98). Other models use transgenic immunoglobulin chains derived from autoreactive B cells arising in autoimmune-prone mice. For example, Andrews and colleagues recently published a study exploring receptor editing in mice that carry an IGKV4-IGKJ4 anti-DNA transgene (99). Although more self-reactive cells were seen when the transgene was expressed in autoimmune-prone MRL/lpr mice, self-reactive B cells were also generated when the transgene was expressed in C57BL/6 mice (99).

This IGKV4-IGKJ4 anti-DNA transgene sequence is derived from a monoclonal antibody that was first isolated from an MRL/lpr mouse in 1987 (100, 101). In describing this and other anti-DNA antibodies, the authors acknowledged their lack of knowledge of the germline genes in MRL/lpr mice, but concluded that the mAb antibody gene sequences were relatively unmutated, based upon a consensus sequence created from both the anti-DNA and other non-DNA-specific antibodies. The apparent presence of some somatic point mutations was, however, deemed to be highly significant. In fact studies describing these antibodies stand as the first evidence for the possibility that self-reactive B cells can arise from self-tolerant B cells by the accumulation of somatic point mutations within the germinal center reaction (100, 101).

Thirty years later, our understanding of MRL/lpr germline genes is still far from complete, but comparisons can now be made between the anti-DNA antibodies and the complete repertoire of C57BL/6 IGKV genes and other murine IGKV genes. This includes sequences that are likely to be NOD IGKV germline genes, many of which are identical to MPL-derived IGKV sequences in GenBank (31). The IGKV sequence in the transgene includes 18 nucleotide differences with respect to the nearest reported IGKV gene (*IGKV4-81*) in the IMGT reference directory. The sequence is, however, much more similar to a NOD sequence reported by Henry and colleagues, differing only within the CDR3 region of the sequence (31). We believe that the many differences with respect to C57BL/6 IGKV genes are a consequence of the separate evolutionary origins of the IGKV loci of the C57BL/6 and MRL/lpr mouse strains. Based upon the haplotype analysis depicted in **Figure 1**, the MRL/MpJ-derived transgene appears to be of *M. m. castaneus* origin. In the absence of further information about the MRL/lpr IGK locus, there can now be no certainty regarding the presence or absence of somatic point mutations in the anti-DNA sequences reported in 1987. Only when the germline IGKV genes of MRL mice have been properly documented will it be possible to say whether or not these anti-DNA antibodies arose through an accumulation of point mutations in previously self-tolerant cells.

We believe that the autoreactivity of the light chain product of the IGKV4-IGKJ4 transgene may be the result of its *M. m. castaneus* origin, and of its association with *M. m. domesticus* and *M. m. musculus*-derived heavy chains. We also believe that *M. m. castaneus* genes may also explain the spontaneous autoreactivity

that is seen in NOD and other inbred mice. The complete κ light chain locus of the NOD strain, and portions of the loci of the MRL/lpr and NZB strains, are derived from the *M. m. castaneus* sub-species of the house mouse (**Figure 1**). The heavy chain locus of the NOD mouse, on the other hand, comes from the *M. m. musculus* sub-species, while the MRL/lpr and NZB strains appear to carry IGH loci that are *M. m. domesticus*-derived.

The three major sub-species of the house mouse emerged from a common ancestor about 350,000 years ago (102), and it is reasonable to assume that their heavy and light chain genes co-evolved as the sub-species diverged. This co-evolution would be required to minimize self-reactivity, and to ensure that each heavy chain could successfully partner with light chains encoded by at least a subset of the IGKV genes. It appears, however, that the breeding histories of many laboratory mice have resulted in heavy and light chain gene sets that did not evolve together being found in their genomes. A few common laboratory strains, including BALB/c and 129 mice, carry matching *M. m. domesticus*-derived IGH and IGK loci, whereas others like the AEJ, C57BL/6, C57BL/10, and SJL strains carry a *M. m. domesticus*-derived IGH locus but a *M. m. musculus*-derived κ locus (**Figure 1**).

Not all inbred mice that have been reported to be prone to autoimmunity harbor obviously mismatched loci. For example, C57BLKS/J mice are diabetes-prone, but have heavy and light chain loci that appear to be derived from *M. m. domesticus* (103). DBA mice are used in a collagen-induced arthritis model of rheumatoid arthritis (104), and both their heavy and light chain loci also appear to be *M. m. domesticus*-derived. It is also true that not all strains that carry mismatched loci have been reported to be susceptible to autoimmunity. An example is the RF/J strain, which appears to have a *M. m. domesticus* IGH locus and a *M. m. castaneus* IGK locus. However, it is striking how many models of autoimmunity involve mismatched heavy and light chain gene loci. In addition to the NOD, MRL/lpr, and NZB models, SJL mice that are used in the EAE model of multiple sclerosis (97) appear to have a *M. m. musculus* IGH locus and a *M. m. domesticus* IGK locus. A model of autoreactivity to matrix collagen uses a C57BL/6-derived IGKV3 transgene in C57BL/6 hosts (105). In this strain, the IGH locus seems to be *M. m. musculus*-derived, while the IGK locus is *M. m. domesticus*-derived. Hybrid (129 \times C57BL/6) mice spontaneously develop an SLE-like condition (106), and these mice express *M. m. domesticus* heavy chains in association with both *M. m. domesticus* and *M. m. musculus*-derived κ chains. Finally, the BXSB mouse spontaneously develops lupus-like pathology (107). SNP analysis suggests that it has a *M. m. domesticus* IGH locus and a *M. m. musculus* κ locus (**Figure 1**).

For over 30 years, mouse models have provided profound insights into the nature of autoreactivity and self-tolerance. It may be though that an ignorance of the makeup of the immunoglobulin gene loci has kept hidden a key genetic contributor to autoimmunity. To determine if this may be the case, the repertoires of laboratory mice will need to be compared to repertoires generated in animals in which the heavy and light chain loci and all critical regulatory elements, as well as self-antigens, are all derived from the same sub-species of the house

mouse. The immunoglobulin genes of the different strains will also need to be properly characterized. It is possible that this may reveal that the apparently matched loci of some autoimmune-prone strains are derived from disparate sources. SNP analysis at present characterizes mouse strains with respect to the three major sub-species of the house mouse, but other minor sub-species may also have contributed genes to the laboratory mouse. We believe that such a focus on heavy and light chain pairings, in mouse models and in human studies, may help explain some of the mysteries that still surround autoimmune diseases.

REFERENCES

- Schatz DG. V(D)J recombination. *Immunol Rev.* (2004) 200:5–11. doi: 10.1111/j.0105-2896.2004.00173.x
- Schatz DG, Swanson PC. V(D)J recombination: mechanisms of initiation. *Annu Rev Genet.* (2011) 45:167–202. doi: 10.1146/annurev-genet-110410-132552
- Ellebedy AH, Jackson KJ, Kissick HT, Nakaya HI, Davis CW, Roskin KM, et al. Defining antigen-specific plasmablast and memory B cell subsets in human blood after viral infection or vaccination. *Nat Immunol.* (2016) 17:1226–34. doi: 10.1038/ni.3533
- Jackson KJ, Kidd MJ, Wang Y, Collins AM. The shape of the Lymphocyte receptor repertoire: lessons from the B cell receptor. *Front Immunol.* (2013) 4:263. doi: 10.3389/fimmu.2013.00263
- Georgiou G, Ippolito GC, Beausang J, Busse CE, Wardemann H, Quake SR. The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat Biotechnol.* (2014) 32:158–68. doi: 10.1038/nbt.2782
- Greiff V, Menzel U, Miho E, Weber C, Riedel R, Cook S, et al. Systems Analysis reveals high genetic and antigen-driven predetermination of antibody repertoires throughout b cell development. *Cell Rep.* (2017) 19:1467–78. doi: 10.1016/j.celrep.2017.04.054
- Kirik U, Persson H, Levander F, Greiff L, Ohlin M. Antibody heavy chain variable domains of different germline gene origins diversify through different paths. *Front Immunol.* (2017) 8:1433. doi: 10.3389/fimmu.2017.01433
- Rettig TA, Ward C, Bye BA, Pecaut MJ, Chapes SK. Characterization of the naive murine antibody repertoire using unamplified high-throughput sequencing. *PLoS ONE* (2018) 13:e0190982. doi: 10.1371/journal.pone.0190982
- Sun Y, Wei Z, Li N, Zhao Y. A comparative overview of immunoglobulin genes and the generation of their diversity in tetrapods. *Dev Comp Immunol.* (2013) 39:103–9. doi: 10.1016/j.dci.2012.02.008
- Mashoof S, Criscitiello MF. Fish immunoglobulins. *Biology* (2016) 5:E45. doi: 10.3390/biology5040045
- Ermert K, Mitlohner H, Schempp W, Zachau HG. The immunoglobulin kappa locus of primates. *Genomics* (1995) 25:623–9. doi: 10.1016/0888-7543(95)80003-5
- Kirschbaum T, Pourrajabi S, Zocher I, Schwendinger J, Heim V, Rosenthaler F, et al. The 3' part of the immunoglobulin kappa locus of the mouse. *Eur J Immunol.* (1998) 28:1458–66. doi: 10.1002/(SICI)1521-4141(199805)28:05<1458::AID-IMMU1458>3.0.CO;2-5
- Martinez-Jean C, Folch G, Lefranc MP. Nomenclature and overview of the mouse (*Mus musculus* and *Mus sp.*) immunoglobulin kappa (IGK) genes. *Exp Clin Immunogenet.* (2001) 18:255–79. doi: 10.1159/000049204
- Schwartz JC, Lefranc MP, Murtaugh MP. Evolution of the porcine (*Sus scrofa domestica*) immunoglobulin kappa locus through germline gene conversion. *Immunogenetics* (2012) 64:303–11. doi: 10.1007/s00251-011-0589-6
- Schwartz JC, Philp RL, Bickhart DM, Smith TPL, Hammond JA. The antibody loci of the domestic goat (*Capra hircus*). *Immunogenetics* (2018) 70:317–26. doi: 10.1007/s00251-017-1033-3
- Martin J, Ponstingl H, Lefranc MP, Archer J, Sargan D, Bradley A. Comprehensive annotation and evolutionary insights into the canine (*Canis lupus familiaris*) antigen receptor loci. *Immunogenetics* (2018) 70:223–36. doi: 10.1007/s00251-017-1028-0

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2018.02249/full#supplementary-material>

- Collins AM, Jackson KJL. On being the right size: antibody repertoire formation in the mouse and human. *Immunogenetics* (2018) 70:143–58. doi: 10.1007/s00251-017-1049-8
- Kawasaki K, Minoshima S, Nakato E, Shibuya K, Shintani A, Schmeits JL, et al. One-megabase sequence analysis of the human immunoglobulin lambda gene locus. *Genome Res.* (1997) 7:250–61. doi: 10.1101/gr.7.3.250
- Watson CT, Steinberg KM, Graves TA, Warren RL, Malig M, Schein J, et al. Sequencing of the human IG light chain loci from a hydattidiform mole BAC library reveals locus-specific signatures of genetic diversity. *Genes Immun.* (2015) 16:24–34. doi: 10.1038/gene.2014.56
- Lefranc MP. Nomenclature of the human immunoglobulin kappa (IGK) genes. *Exp Clin Immunogenet.* (2001) 18:161–74. doi: 10.1159/000049195
- Ramesh A, Darko S, Hua A, Overman G, Ransier A, Francica JR, et al. Structure and diversity of the *Rhesus macaque* immunoglobulin loci through multiple *de novo* genome assemblies. *Front Immunol.* (2017) 8:1407. doi: 10.3389/fimmu.2017.01407
- Sun Y, Wang C, Wang Y, Zhang T, Ren L, Hu X, et al. A comprehensive analysis of germline and expressed immunoglobulin repertoire in the horse. *Dev Comp Immunol.* (2010) 34:1009–20. doi: 10.1016/j.dci.2010.05.003
- Guo Y, Bao Y, Meng Q, Hu X, Meng Q, Ren L, et al. Immunoglobulin genomics in the guinea pig (*Cavia porcellus*). *PLoS ONE* (2012) 7:e39298. doi: 10.1371/journal.pone.0039298
- Dammers PM, Hendricks J, Terpstra P, Bos NA, Kroese FG. The organization of the immunoglobulin heavy- and light-chain loci in the rat. In: Kaushik AKaP, Y, editor. *Comparative Immunoglobulin Genetics*. Boca Raton, FL: CRC Press (2014). 107–30.
- Das S, Nikolaidis N, Klein J, Nei M. Evolutionary redefinition of immunoglobulin light chain isotypes in tetrapods using molecular markers. *Proc Natl Acad Sci USA.* (2008) 105:16647–52. doi: 10.1073/pnas.0808800105
- Butler JE, Wertz N, Baker ML. The immunoglobulin genes of bats. In: Kaushik AK, Pasman Y, editors. *Comparative Immunoglobulin Genetics*. Toronto, ON: Apple Academic Press (2014). 53–84.
- Das S, Mohamedy U, Hirano M, Nei M, Nikolaidis N. Analysis of the immunoglobulin light chain genes in zebra finch: evolutionary implications. *Mol Biol Evol.* (2010) 27:113–20. doi: 10.1093/molbev/msp212
- Aoki-Ota M, Torkamani A, Ota T, Schork N, Nemazee D. Skewed primary Igkappa repertoire and V-J joining in C57BL/6 mice: implications for recombination accessibility and receptor editing. *J Immunol.* (2012) 188:2305–15. doi: 10.4049/jimmunol.1103484
- Collins AM, Wang Y, Roskin KM, Marquis CP, Jackson KJ. 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. doi: 10.1098/rstb.2014.0236
- Yang H, Wang JR, Didion JP, Buus RJ, Bell TA, Welsh CE, et al. Subspecific origin and haplotype diversity in the laboratory mouse. *Nat Genet.* (2011) 43:648–55. doi: 10.1038/ng.847
- Henry RA, Kendall PL, Woodward EJ, Hulbert C, Thomas JW. Vkappa polymorphisms in NOD mice are spread throughout the entire immunoglobulin kappa locus and are shared by other autoimmune strains. *Immunogenetics* (2010) 62:507–20. doi: 10.1007/s00251-010-0457-9
- Loh E, Black B, Riblet R, Weigert M, Hood JM, Hood L. Myeloma proteins from NZB and BALB/c mice: structural and functional differences. *Proc Natl Acad Sci USA.* (1979) 76:1395–9. doi: 10.1073/pnas.76.3.1395

33. Wang JR, de Villena FP, McMillan L. Comparative analysis and visualization of multiple collinear genomes. *BMC Bioinf.* (2012) 13(Suppl. 3):S13. doi: 10.1186/1471-2105-13-S3-S13
34. Hieter PA, Korsmeyer SJ, Waldmann TA, Leder P. Human immunoglobulin kappa light-chain genes are deleted or rearranged in lambda-producing B cells. *Nature* (1981) 290:368–72. doi: 10.1038/290368a0
35. Schwartz JC, Lefranc MP, Murtaugh MP. Organization, complexity and allelic diversity of the porcine (*Sus scrofa domestica*) immunoglobulin lambda locus. *Immunogenetics* (2012) 64:399–407. doi: 10.1007/s00251-011-0594-9
36. Walther S, Rusitzka TV, Diesterbeck US, Czerny CP. Equine immunoglobulins and organization of immunoglobulin genes. *Dev Comp Immunol.* (2015) 53:303–19. doi: 10.1016/j.dci.2015.07.017
37. Scott CL, Potter M. Variation in V lambda genes in the genus *Mus*. *J Immunol.* (1984) 132:2638–43.
38. Sanchez P, Nadel B, Cazenave PA. V lambda-J lambda rearrangements are restricted within a V-J-C recombination unit in the mouse. *Eur J Immunol.* (1991) 21: 907–11. doi: 10.1002/eji.1830210408
39. Lefranc MP, LeFranc G. Immunoglobulin lambda (IGL) genes of human and mouse. In: Honjo T, Alt FW, Neuberger MS, editors. *Molecular Biology of B cells*. London: Elsevier Academic Press (2004). 37–59.
40. Matsuda F, Ishii K, Bourvagnet P, Kuma K, Hayashida H, Miyata T, et al. The complete nucleotide sequence of the human immunoglobulin heavy chain variable region locus. *J Exp Med.* (1998) 188:2151–62. doi: 10.1084/jem.188.11.2151
41. Riblet R. Immunoglobulin heavy chain genes in the mouse. In: Honjo T, Alt FW, Neuberger M, editors. *Molecular Biology of B cells*. London: Elsevier Academic Press (2004). p 19–26. doi: 10.1016/B978-012053641-2/50003-4
42. Ma L, Qin T, Chu D, Cheng X, Wang J, Wang X, et al. Internal duplications of DH, JH, and C region genes create an unusual IgH gene Locus in cattle. *J Immunol.* (2016) 196:4358–66. doi: 10.4049/jimmunol.1600158
43. Guo Y, Bao Y, Wang H, Hu X, Zhao Z, Li N, et al. A preliminary analysis of the immunoglobulin genes in the African elephant (*Loxodonta africana*). *PLoS ONE* (2011) 6:e16889. doi: 10.1371/journal.pone.0016889
44. Guo C, Gerasimova T, Hao H, Ivanova I, Chakraborty T, Selimyan R, et al. Two forms of loops generate the chromatin conformation of the immunoglobulin heavy-chain gene locus. *Cell* (2011) 147:332–43. doi: 10.1016/j.cell.2011.08.049
45. Kidd MJ, Jackson KJ, Boyd SD, Collins AM. DJ Pairing during VDJ recombination shows positional biases that vary among individuals with differing IGHD Locus immunogenotypes. *J Immunol.* (2016) 196:1158–64. doi: 10.4049/jimmunol.1501401
46. Langerak AW, Nadel B, De Torbal A, Wolvers-Tettero IL, van Gastel-Mol EJ, Verhaaf B, et al. Unraveling the consecutive recombination events in the human IGK locus. *J Immunol.* (2004) 173:3878–88. doi: 10.4049/jimmunol.173.6.3878
47. Panigrahi AK, Goodman NG, Eisenberg RA, Rickels MR, Naji A, Luning Prak ET. RS rearrangement frequency as a marker of receptor editing in lupus and type 1 diabetes. *J Exp Med.* (2008) 205:2985–94. doi: 10.1084/jem.20082053
48. Das S, Nikolaidis N, Nei M. Genomic organization and evolution of immunoglobulin kappa gene enhancers and kappa deleting element in mammals. *Mol Immunol.* (2009) 46:3171–7. doi: 10.1016/j.molimm.2009.05.180
49. Siminovitch KA, Moore MW, Durdik J, Selsing E. The human kappa deleting element and the mouse recombining segment share DNA sequence homology. *Nucleic Acids Res.* (1987) 15:2699–705. doi: 10.1093/nar/15.6.2699
50. Alt FW, Enea V, Bothwell AL, Baltimore D. Activity of multiple light chain genes in murine myeloma cells producing a single, functional light chain. *Cell* (1980) 21:1–12. doi: 10.1016/0092-8674(80)90109-9
51. Butler JE, Wertz N. The porcine antibody repertoire: variations on the textbook theme. *Front Immunol.* (2012) 3:153. doi: 10.3389/fimmu.2012.00153
52. Wertz N, Vazquez J, Wells K, Sun J, Butler JE. Antibody repertoire development in fetal and neonatal piglets. XII. Three IGLV genes comprise 70% of the pre-immune repertoire and there is little junctional diversity. *Mol Immunol.* (2013) 55:319–28. doi: 10.1016/j.molimm.2013.03.012
53. Prak EL, Weigert M. Light chain replacement: a new model for antibody gene rearrangement. *J Exp Med.* (1995) 182:541–8. doi: 10.1084/jem.182.2.541
54. Retter MW, Nemazee D. Receptor editing occurs frequently during normal B cell development. *J Exp Med.* (1998) 188:1231–8. doi: 10.1084/jem.188.7.1231
55. Halverson R, Torres RM, Pelanda R. Receptor editing is the main mechanism of B cell tolerance toward membrane antigens. *Nat Immunol.* (2004) 5:645–50. doi: 10.1038/ni1076
56. Lang J, Ota T, Kelly M, Strauch P, Freed BM, Torres RM, et al. Receptor editing and genetic variability in human autoreactive B cells. *J Exp Med.* (2016) 213:93–108. doi: 10.1084/jem.20151039
57. Detanico T, St Clair JB, Aviszus K, Kirchenbaum G, Guo W, Wysocki LJ. Somatic mutagenesis in autoimmunity. *Autoimmunity* (2013) 46:102–14. doi: 10.3109/08916934.2012.757597
58. Li S, Liu W, Li Y, Zhao S, Liu C, Hu M, et al. Contribution of secondary Igkappa rearrangement to primary immunoglobulin repertoire diversification. *Mol Immunol.* (2016) 78:193–206. doi: 10.1016/j.molimm.2016.09.006
59. Yamagami T, ten Boekel E, Andersson J, Rolink A, Melchers F. Frequencies of multiple IgL chain gene rearrangements in single normal or kappaL chain-deficient B lineage cells. *Immunity* (1999) 11:317–27. doi: 10.1016/S1074-7613(00)80107-7
60. Vettermann C, Timblin GA, Lim V, Lai EC, Schlissel MS. The proximal J kappa germline-transcript promoter facilitates receptor editing through control of ordered recombination. *PLoS ONE* (2015) 10:e0113824. doi: 10.1371/journal.pone.0113824
61. Wardemann H, Hammersen J, Nussenzweig MC. Human autoantibody silencing by immunoglobulin light chains. *J Exp Med.* (2004) 200:191–9. doi: 10.1084/jem.20040818
62. Hershberg U, Shlomchik MJ. Differences in potential for amino acid change after mutation reveals distinct strategies for kappa and lambda light-chain variation. *Proc Natl Acad Sci USA.* (2006) 103:15963–8. doi: 10.1073/pnas.0607581103
63. Avnir Y, Watson CT, Glanville J, Peterson EC, Tallarico AS, Bennett AS, et al. IGHV1-69 polymorphism modulates anti-influenza antibody repertoires, correlates with IGHV utilization shifts and varies by ethnicity. *Sci Rep.* (2016) 6:20842. doi: 10.1038/srep20842
64. Collins AM, Wang Y, Singh V, Yu P, Jackson KJ, Sewell WA. The reported germline repertoire of human immunoglobulin kappa chain genes is relatively complete and accurate. *Immunogenetics* (2008) 60:669–76. doi: 10.1007/s00251-008-0325-z
65. Wang Y, Jackson KJ, Sewell WA, Collins AM. Many human immunoglobulin heavy-chain IGHV gene polymorphisms have been reported in error. *Immunol Cell Biol.* (2008) 86:111–5. doi: 10.1038/sj.icb.7100144
66. Schwartz GW, Hershberg U. Conserved variation: identifying patterns of stability and variability in BCR and TCR V genes with different diversity and richness metrics. *Phys Biol.* (2013) 10:035005. doi: 10.1088/1478-3975/10/3/035005
67. Scheepers C, Shrestha RK, Lambson BE, Jackson KJ, Wright IA, Naicker D, et al. Ability to develop broadly neutralizing HIV-1 antibodies is not restricted by the germline Ig gene repertoire. *J Immunol.* (2015) 194:4371–8. doi: 10.4049/jimmunol.1500118
68. Sheng Z, Schramm CA, Kong R, Program NCS, Mullikin JC, Mascola JR, et al. Gene-specific substitution profiles describe the types and frequencies of amino acid changes during antibody somatic hypermutation. *Front Immunol.* (2017) 8:537. doi: 10.3389/fimmu.2017.00537
69. Jackson KJ, Wang Y, Gaeta BA, Pomat W, Siba P, Rimmer J, et al. Divergent human populations show extensive shared IGK rearrangements in peripheral blood B cells. *Immunogenetics* (2012) 64:3–14. doi: 10.1007/s00251-011-0559-z
70. Gidoni M, Snir O, Peres A, Polak P, Lindeman I, Clouser C, et al. Mosaic deletion patterns of the human antibody heavy chain gene locus. (2018) *BioRxiv. [Preprint]*. doi: 10.1101/314476
71. Boyd SD, Gaeta BA, Jackson KJ, Fire AZ, Marshall EL, Merker JD, et al. Individual variation in the germline Ig gene repertoire inferred from variable region gene rearrangements. *J Immunol.* (2010) 184:6986–92. doi: 10.4049/jimmunol.1000445

72. Chinge NO, Pramanik S, Hu G, Lin Y, Gao R, Shen L, et al. Determination of gene organization in the human IGHV region on single chromosomes. *Genes Immun.* (2005) 6:186–93. doi: 10.1038/sj.gene.6364176
73. Watson CT, Breden F. The immunoglobulin heavy chain locus: genetic variation, missing data, and implications for human disease. *Genes Immun.* (2012) 13:363–73. doi: 10.1038/gene.2012.12
74. Watson CT, Steinberg KM, Huddleston J, Warren RL, Malig M, Schein J, et al. Complete haplotype sequence of the human immunoglobulin heavy-chain variable, diversity, and joining genes and characterization of allelic and copy-number variation. *Am J Hum Genet.* (2013) 92:530–46. doi: 10.1016/j.ajhg.2013.03.004
75. Milner EC, Hufnagle WO, Glas AM, Suzuki I, Alexander C. Polymorphism and utilization of human VH Genes. *Ann N Y Acad Sci.* (1995) 764:50–61. doi: 10.1111/j.1749-6632.1995.tb55806.x
76. Juul L, Hougs L, Barington T. A new apparently functional IGVK gene (VkLa) present in some individuals only. *Immunogenetics* (1998) 48:40–6. doi: 10.1007/s002510050398
77. Lefranc MP, Pallares N, Fripiat JP. Allelic polymorphisms and RFLP in the human immunoglobulin lambda light chain locus. *Hum Genet.* (1999) 104:361–9. doi: 10.1007/s004390050969
78. Moraes Junta C, Passos GA. Genomic EcoRI polymorphism and cosmid sequencing reveal an insertion/deletion and a new IGLV5 allele in the human immunoglobulin lambda variable locus (22q11.2/IGLV). *Immunogenetics* (2003) 55:10–5. doi: 10.1007/s00251-003-0549-x
79. Romo-Gonzalez T, Vargas-Madrazo E. Substitution patterns in alleles of immunoglobulin V genes in humans and mice. *Mol Immunol.* (2006) 43:731–44. doi: 10.1016/j.molimm.2005.03.018
80. Kirschbaum T, Rosenthaler F, Bensch A, Holscher B, Lautner-Rieske A, Ohnrich M, et al. The central part of the mouse immunoglobulin kappa locus. *Eur J Immunol.* (1999) 29:2057–64. doi: 10.1002/(SICI)1521-4141(199907)29:07<2057::AID-IMMU2057>3.0.CO;2-P
81. Rosenthaler F, Kirschbaum T, Heim V, Kirschbaum V, Schable KF, Schwendinger J, et al. The 5' part of the mouse immunoglobulin kappa locus. *Eur J Immunol.* (1999) 29:2065–71. doi: 10.1002/1521-4141(2000012)30:123.0.CO;2-G
82. Brekke KM, Garrard WT. Assembly and analysis of the mouse immunoglobulin kappa gene sequence. *Immunogenetics* (2004) 56:490–505. doi: 10.1007/s00251-004-0659-0
83. Fitzsimmons SP, Clark KJ, Shapiro MA. Evolutionary relationships and polymorphisms among V kappa 10 genes from inbred and wild-derived inbred mice. *Immunogenetics* (2002) 54:9–19. doi: 10.1007/s00251-002-0438-8
84. Henderson TJ, Rudikoff S. Characterization of a V kappa family in *Mus musculus castaneus*: sequence analysis. *Immunogenetics* (1993) 37:426–36.
85. Kim SO, Sanz I, Williams C, Capra JD, Gottlieb PD. Polymorphism in V kappa 10 genes encoding L chains of antibodies bearing the Ars-A and A48 cross-reactive idiotypes. *Immunogenetics* (1991) 34:231–41. doi: 10.1007/BF00215258
86. Good RA. Contempo '81. *Immunology. JAMA* (1981) 245:2197–8. doi: 10.1001/jama.1981.03310460049021
87. Greiff V, Weber CR, Palme J, Bodenhofer U, Miho E, Menzel U, et al. Learning the high-dimensional immunogenomic features that predict public and private antibody repertoires. *J Immunol.* (2017) 199:2985–97. doi: 10.4049/jimmunol.1700594
88. Hoi KH, Ippolito GC. Intrinsic bias and public rearrangements in the human immunoglobulin V lambda light chain repertoire. *Genes Immun.* (2013) 14:271–6. doi: 10.1038/gene.2013.10
89. Ignatovich O, Tomlinson IM, Jones PT, Winter G. The creation of diversity in the human immunoglobulin V(lambda) repertoire. *J Mol Biol.* (1997) 268:69–77. doi: 10.1006/jmbi.1997.0956
90. Vasicek TJ, Leder P. Structure and expression of the human immunoglobulin lambda genes. *J Exp Med.* (1990) 172: 609–20. doi: 10.1084/jem.172.2.609
91. Farner NL, Dorner T, Lipsky PE. Molecular mechanisms and selection influence the generation of the human V lambda J lambda repertoire. *J Immunol.* (1999) 162:2137–45.
92. Longo NS, Rogosch T, Zemlin M, Zouali M, Lipsky PE. Mechanisms that shape human antibody repertoire development in mice transgenic for human Ig H and L chain Loci. *J Immunol.* (2017) 198:3963–77. doi: 10.4049/jimmunol.1700133
93. Vander Heiden JA, Stathopoulos P, Zhou JQ, Chen L, Gilbert TJ, Bolen CR, et al. Dysregulation of B cell repertoire formation in *Myasthenia Gravis* patients revealed through deep sequencing. *J Immunol.* (2017) 198:1460–73. doi: 10.4049/jimmunol.1601415
94. Fraser LD, Zhao Y, Lutalo PM, D'Cruz DP, Cason J, Silva JS, et al. Immunoglobulin light chain allelic inclusion in systemic lupus erythematosus. *Eur J Immunol.* (2015) 45:2409–19. doi: 10.1002/eji.201545599
95. Hehle V, Fraser LD, Tahir R, Kipling D, Wu YC, Lutalo PM, et al. Immunoglobulin kappa variable region gene selection during early human B cell development in health and systemic *Lupus erythematosus*. *Mol Immunol.* (2015) 65:215–23. doi: 10.1016/j.molimm.2015.01.017
96. Li Y, Spellerberg MB, Stevenson FK, Capra JD, Potter KN. The I binding specificity of human VH 4-34 (VH 4-21) encoded antibodies is determined by both VH framework region 1 and complementarity determining region 3. *J Mol Biol.* (1996) 256:577–89. doi: 10.1006/jmbi.1996.0110
97. Morel L. Mouse models of human autoimmune diseases: essential tools that require the proper controls. *PLoS Biol.* (2004) 2:E241. doi: 10.1371/journal.pbio.0020241
98. Burnett DL, Langley DB, Schofield P, Hermes JR, Chan TD, Jackson J, et al. Germinal center antibody mutation trajectories are determined by rapid self/foreign discrimination. *Science* (2018) 360:223–6. doi: 10.1126/science.aao3859
99. Andrews SF, Zhang Q, Lim S, Li L, Lee JH, Zheng NY, et al. Global analysis of B cell selection using an immunoglobulin light chain-mediated model of autoreactivity. *J Exp Med.* (2013) 210:125–42. doi: 10.1084/jem.20120525
100. Shlomchik MJ, Aucoin AH, Pisetsky DS, Weigert MG. Structure and function of anti-DNA autoantibodies derived from a single autoimmune mouse. *Proc Natl Acad Sci USA.* (1987) 84:9150–4. doi: 10.1073/pnas.84.24.9150
101. Shlomchik MJ, Marshak-Rothstein A, Wolfowicz CB, Rothstein TL, Weigert MG. The role of clonal selection and somatic mutation in autoimmunity. *Nature* (1987) 328:805–11. doi: 10.1038/328805a0
102. Geraldès A, Basset P, Smith KL, Nachman MW. Higher differentiation among subspecies of the house mouse (*Mus musculus*) in genomic regions with low recombination. *Mol Ecol.* (2011) 20:4722–36. doi: 10.1111/j.1365-294X.2011.05285.x
103. Mao HZ, Roussos ET, Peterfy M. Genetic analysis of the diabetes-prone C57BLKS/J mouse strain reveals genetic contribution from multiple strains. *Biochim Biophys Acta* (2006) 1762:440–6. doi: 10.1016/j.bbdis.2006.01.002
104. Joosten LA, Helsen MM, van de Loo FA, van den Berg WB. Anticytokine treatment of established type II collagen-induced arthritis in DBA/1 mice. A comparative study using anti-TNF α , anti-IL-1 α/β , and IL-1Ra. *Arthritis Rheum.* (1996) 39:797–809. doi: 10.1002/art.23363
105. Clark AG, Worni-Schudel IM, Korte FM, Foster MH. A murine Ig light chain transgene reveals IGKV3 gene contributions to anti-collagen types IV and II specificities. *Mol Immunol.* (2017) 91:49–56. doi: 10.1016/j.molimm.2017.08.015
106. Bygrave AE, Rose KL, Cortes-Hernandez J, Warren J, Rigby RJ, Cook HT, et al. Spontaneous autoimmunity in 129 and C57BL/6 mice—implications for autoimmunity described in gene-targeted mice. *PLoS Biol.* (2004) 2:E243. doi: 10.1371/journal.pbio.0020243
107. Bubier JA, Sproule TJ, Foreman O, Spolski R, Shaffer DJ, Morse HC, III, et al. A critical role for IL-21 receptor signaling in the pathogenesis of systemic lupus erythematosus in BXSb-Yaa mice. *Proc Natl Acad Sci USA.* (2009) 106:1518–23. doi: 10.1073/pnas.0807309106

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor declared a past co-authorship with one of the authors AC.

Copyright © 2018 Collins and Watson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.