Individual Variation in the Germline Ig Gene Repertoire Inferred from Variable Region Gene Rearrangements


*J Immunol* published online 21 May 2010
http://www.jimmunol.org/content/early/2010/05/21/jimmunol.1000445

**Supplementary Material**
http://www.jimmunol.org/content/suppl/2010/05/19/jimmunol.1000445.SDC1

**Subscription**
Information about subscribing to *The Journal of Immunology* is online at:
http://jimmunol.org/subscription

**Permissions**
Submit copyright permission requests at:
http://www.aai.org/About/Publications/JI/copyright.html

**Email Alerts**
Receive free email-alerts when new articles cite this article. Sign up at:
http://jimmunol.org/alerts

*The Journal of Immunology* is published twice each month by
The American Association of Immunologists, Inc.,
1451 Rockville Pike, Suite 650, Rockville, MD 20852
All rights reserved.
Print ISSN: 0022-1767 Online ISSN: 1550-6606.
Individual Variation in the Germline Ig Gene Repertoire Inferred from Variable Region Gene Rearrangements

Scott D. Boyd,* Bruno A. Gaëta,† Katherine J. Jackson,‡ Andrew Z. Fire,*§ Kari C. Nadeau,** Michael Egholm, # David B. Miklos, †† James L. Zehnder,*†† and Andrew M. Collins‡

Individual variation in the Ig germline gene repertoire leads to individual differences in the combinatorial diversity of the Ab repertoire, but the study of such variation has been problematic. The application of high-throughput DNA sequencing to the study of rearranged Ig genes now makes this possible. The sequencing of thousands of VDJ rearrangements from an individual, either from genomic DNA or expressed mRNA, should allow their germline IGHV, IGHD, and IGHJ repertoires to be inferred. In addition, where previously mere glimpses of diversity could be gained from sequencing studies, new large data sets should allow the rearrangement frequency of different genes and alleles to be seen with clarity. We analyzed the DNA of 108,210 human IgH chain rearrangements from 12 individuals and determined their individual IGH genotypes. The number of reportedly functional IGHV genes and allelic variants ranged from 45 to 60, principally because of variable levels of gene heterozygosity, and included 14 previously unreported IGHV polymorphisms. New polymorphisms of the IGHD3-16 and IGHJ6 genes were also seen. At heterozygous loci, remarkably different rearrangement frequencies were seen for the various IGHV alleles, and these frequencies were consistent between individuals. The specific alleles that make up an individual's Ig genotype may therefore be critical in shaping the combinatorial repertoire. The extent of genotypic variation between individuals is highlighted by an individual with aplastic anemia who appears to lack six contiguous IGHD genes on both chromosomes. These deletions significantly alter the potential expressed IGH repertoire, and possibly immune function, in this individual. The Journal of Immunology, 2010, 184: 000–000.

The mammalian immune system has the ability to respond to almost any Ag to which it is exposed because of the incredible diversity of lymphocyte receptor molecules. The diversity of Ab molecules is made possible by multiple sets of highly similar genes, which recombine to form functional VDJ gene rearrangements encoding the IgH chain and VJ gene rearrangements encoding the IgL chain (1). This combinatorial diversity is expanded still further by junctional diversity arising from exonuclease trimming of the recombining gene ends and from the essentially random addition of nucleotides by the enzyme TdT (2).

Although the complex IgHV gene locus was carefully mapped in the 1990s (3), the difficulties associated with the sequencing of these highly similar genes and pseudogenes has meant that the locus of a single individual has only been fully explored once (4). Even the initial versions of the human genome (5–7) were based on the previously reported locus sequence (4). A complete description of the IGHD locus has also only been published for a single individual (8). Detailed exploration of genotypic differences of the Ig gene loci between individuals is almost totally lacking.

Although the sequencing of the immunoreceptor gene loci remains a challenge, investigation of allelic variation of a single Ig gene is relatively easy. Many allelic variants were identified in the early years of investigation of the Ig gene loci. However with the first sequencing of the H chain gene locus (4), most exploration of gene variants came to an end. Although hundreds of genes and allelic variants were identified in the 1980s and early 1990s, a mere handful of alleles have been identified since that time. We have recently re-evaluated the reported H chain V region gene sequences and have concluded that many alleles were reported in error (9–11). Using a bioinformatic approach, we also inferred the existence of a number of previously unreported polymorphisms but concluded that the description of genetic diversity in the H chain V region loci was essentially complete. The identification of alleles that are present in the population at low frequency, and the identification of genes that are less commonly rearranged or expressed was not possible from the data sets of our earlier analysis.

We have recently performed high-throughput pyrosequencing of VDJ rearrangements from the genomic DNA of B cell populations in a series of human clinical specimens and have demonstrated the potential of this approach for monitoring of lymphoid malignancies.
and the healthy immune system (12). In this study, we extend these results to show the usefulness of deep sequencing of VDJ rearrangements to determine individual Ig locus gene segment genotypes. Sequencing thousands of VDJ gene rearrangements from the genomic DNA of an individual’s B cells allows the sets of germline IGHV, IGHD, and IGHJ genes that contributed to the rearrangements to be inferred and permits novel allelic variants to be distinguished from somatic hypermutation sequence changes. The rearrangements identified with this approach will include both functional VDJ rearrangements that contribute to the expressed RNA and Ig protein repertoire, as well as some sequences that are nonfunctional.

Large data sets of rearranged IGH sequences from peripheral blood B cells allow accurate genotyping of germline gene segments because of the opportunity to observe a given germline V, D, or J segment in many distinct, independently derived VDJ combinations, even when an individual carries previously unreported alleles in his or her genome. This is true even though pyrosequencing of VDJ genes from genomic DNA of peripheral blood will include many sequences that have accumulated somatic point mutations during an immune response. When an individual carries an unreported polymorphism, the polymorphism is typically identified in data sets by the absence of apparently unmutated rearrangements of a gene that is highly similar to the unreported polymorphism and by the presence of hundreds or even thousands of examples of rearrangements that include shared mismatches when aligned to that similar germline sequence. The analysis reported in this paper shows that the inference of the existence of previously unreported polymorphisms is possible even when the utilization frequency of the allele in VDJ rearrangements is relatively low. Relatively common copy number changes of particular gene segments can also be identified and likely contribute to variation in the IGH repertoire of different individuals.

Our finding that an individual with aplastic anemia has an apparent homozygous deletion of a major part of the IGHD repertoire suggests that investigations of Ig genotypes may provide important insights into the operation of the immune system in health and disease. Whether or not such holes in the available germline gene repertoire are common, or of clinical significance, will require widespread investigation of individual variation in the Ig gene loci.

Materials and Methods

Specimens

Specimens of human peripheral blood and a sample of tonsillar tissue were obtained under a Stanford University Institutional Review Board-approved protocol. Blood samples came from eight healthy individuals, one patient with acute myeloid leukemia, one patient with chronic myeloid leukemia, and one patient with aplastic anemia. The tonsil sample came from a healthy individual.

DNA template preparation

PBMCs were isolated by centrifugation of diluted blood layered over Hypaque 1077 (Sigma-Aldrich, St. Louis, MO). Column purification (Qiagen, Valencia, CA) or magnetic bead-based isolation (Magnapure: Roche Diagnostics, Indianapolis, IN) was used to purify DNA template. Frozen tonsil tissue was homogenized prior to genomic DNA purification.

PCR amplifications and high-throughput pyrosequencing

High-throughput amplico pyrosequencing data obtained with a 454 instrument are presented in this paper and are derived from two independent experiments: one performed using FLX chemistry and one with Titanium chemistry, with amplicons sequencing reads beginning from the “B” primer in the manufacturer’s protocol in both cases (Roche, Branford, CT). PCR amplifications were performed using 200 ng PBMC genomic DNA as template for each reaction. Following PCR, amplicons were pooled in equal amounts and were then purified by 1.5% agarose gel electrophoresis and gel extraction as described previously (12). PCR was performed using the BIOMED-2 protocol and primers that were augmented for sequencing experiments by the addition of 5’ sequencing elements and by the addition of unique 6-, 7-, or 10-nt sample “bar codes” to the IGHJ primers as described previously (12). An additional 10-nt sample bar code was incorporated into the multiplexed IGHV gene segment primers in the analysis of specimens with the 454 Titanium chemistry (12). Sequence bar codes allowed for the pooling and bulk sequencing of many libraries and the subsequent retrieval of sequences that were derived from each sample.

Sequence data analysis

Sequences from each input specimen were sorted based on recognition of a perfect match to the sample bar code as well as a perfect match to the first 3 bases of the IGHJ common primer. Non-Ig sequences were filtered from the sequence sets prior to analysis. Alignment of rearranged IGH sequences to germline V, D, and J segments, and determination of V–D junctions and D–J junctions was performed using the iHMMune-align utility (13). Sequences containing base-pair insertions or deletions in the IGHV or IGHJ gene segments were filtered from the data set, based on the known error properties of pyrosequencing (14, 15). Duplicate sequences were identified in a postprocessing step on the basis of shared IGHV, IGHD, and IGHJ genes as well as shared N regions. Where duplicate sequences were identified that had differing levels of somatic point mutations, the least mutated sequence was retained in the data set.

Genotyping donors

Alignments were reviewed for each individual in the study to identify misaligned sequences. IGHV alignments were considered to be potentially in error where an individual’s data set included alignments to more than one allele of a particular gene. Where unique, unmutated alignments to a second or third allele were seen at a frequency that was >10% of the frequency of a more highly used allele, the alignments were accepted as correct if investigations could exclude PCR-mediated chimerism as an explanation for the sequences (16). Such chimerism was detected by separately aligning 5’ and 3’ portions of the IGHV sequences. Where small numbers of alignments were seen to a single allele, these were accepted as correct alignments if unmutated alignments were seen and if there was no other gene that was present in other VDJ gene rearrangements that could give rise to such a sequence by, in most cases, a single mutation. In a few cases, where hundreds or even thousands of alignments were seen to a particular gene, it was acknowledged that two and rarely three somatic point mutations could give rise to a sequence that aligned better to another gene. In this way, the IGHV genotype of each individual was progressively defined, and finally, misaligned sequences could be reassigned by reference to this genotype. A handful of sequences could not be assigned to another IGHV gene with absolute certainty, and in these cases, the sequences were removed from the analysis. This situation was confined to very rarely rearranged genes, such as IGHV4-30-2. Although this gene was undoubtedly found in a few VDJ rearrangements of some individuals, in others, a single example might be found where a VDJ sequence aligned to IGHV4-30-2 with many mismatches. This was not considered sufficient evidence of the presence of that germline gene in the individual’s genome.

IGHV alignments were reviewed to identify misalignments to allelic variants. For example, where alignments were seen to both IGHV2-8*01 and IGHV2-8*02, the length of alignments, the extent of IGHD mutations, and patterns of 5’ and 3’ exonuclease removals of nucleotides from the IGHD genes were analyzed. It was assumed that data from highly similar alleles would, for example, show similar levels of exonuclease activity. Where differences were seen in comparisons of data sets from the different alleles, misalignments could be identified, and the IGHD genotype of each individual could be determined. iHMMune-align does not attempt to identify IGHJ-2 alleles, because the critical nucleotides that distinguish the three IGHJ-2 alleles are found at the ends of the sequences. Exonuclease removals make it impossible to determine the allelic source of most sequences, whereas N nucleotide addition can very easily make a sequence using one allele align perfectly to another. The large data sets analyzed in this study, however, allowed the unequivocal determination of the IGHJ-2 locus in most cases, because large sets of apparently unprocessed gene ends could be compiled and reviewed. IGHJ alignments were also reviewed, and where small numbers of alignments were seen to variant IGHJ alleles, these were reassigned to the alleles that were predominant in the VDJ gene rearrangements.

Identification of unreported polymorphisms

Levels of mutations in the IGHV, IGHD, and IGHJ genes were analyzed, and where large numbers of alignments to a particular gene carried shared mismatches, further investigations were carried out to identify putative
The IMGT information system (17).

The number of IGHV genes and allelic variants identified in Table I. Although the existence of some of these polymorphisms is supported by hundreds of rearranged sequences, they will not be assigned gene names by the designated authority (the ImMunoGeneTics [IMGT] database) until genomic sequences are amplified from unarranged germline genes. As we believe it is unlikely that all such germline sequences will be quickly reported, and as the availability of the most complete germline repertoire is critical for many studies, we have established a Putative Immunoglobulin Polymorphism database. All the polymorphisms identified in this study, and evidence supporting their existence, are accessible via this database (www.cse.unsw.edu.au/~immune/IgPdb/).

The IGHV genes that were identified in rearrangements can be divided into core genes that were seen in all individuals studied and noncore genes that were seen in some but not all individuals. Table II shows that some core genes make a relatively trivial contribution to the repertoire, despite being a feature of all genomes studied to date. In contrast, some noncore genes make a substantial contribution to the repertoire of certain individuals. Although many genes contributed to widely varying percentages of the overall repertoire of rearranged VDJ genes, there is evidence that this may not solely be a result of selection, but rather may also be a consequence of the varying propensity for the different IGHV alleles to recombine with DJ rearrangements. For example, the allele IGHV7-4-1*01 was seen in from 0.04 to 0.4% of all sequences in the individuals with this allele. The alternative allele IGHV7-4-1*02, which differs from IGHV7-4-1*01 by a single nucleotide but which encodes the identical amino acid sequence, was seen in as many as 3.6% of sequences. Preferential PCR amplification is unlikely to account for this difference in prevalence, given that the nucleotide distinguishing these alleles is distant from the binding sites of the IGHV primer used to amplify the sequences.

Where pairs of alleles were carried in the genomes of multiple individuals, the relative proportions in which the different alleles were seen were remarkably constant, as seen in Table III. This was true for individuals carrying two or even three alleles of a particular gene. In almost all cases, the proportions in which the alleles were observed were similar, and $\chi^2$ analysis showed significant differences between individual levels of rearrangement for only 3 of 23 pairs and sets of three alleles. Interestingly, two of these three pairs included the IGHV3-48*02 allele. Between individuals, there were significantly different patterns of rearrangement of the IGHV3-48*01 and *02 alleles as well as the *02 and *03 alleles. This could result from the existence of variants of the *02 allele that rearrange at both high and low frequency. It is also possible that the unobserved $S'$ region of the V segment upstream of the PCR primer sites used in preparing amplicons for sequencing could contain unreported polymorphisms defining a new allele that has a different propensity for rearrangement, compared with the *02 allele. Some rearrangement of pseudogenes was seen, and despite their extremely low frequency within the data sets, the relative proportions in which the pseudogenes were seen also showed a clear pattern (data not shown).

We have previously suggested that 104 of the 226 reported germline IGHV genes include sequencing errors, or other faults, and should be removed from the available repertoire (11). Analysis of the 108,210 sequences in this study is generally consistent with this conclusion, although 4 of the 104 germline sequences that were absent in analysis of the earlier, smaller data set ($n = 4718$) were seen in this study. These sequences (IGHV3-3-d*01, IGHV7-81*01, IGHV3-35*01, and IGHV4-30-2*02) should therefore be retained in the available repertoire. Some doubt was raised in the earlier study regarding the existence or accuracy of other reported alleles. The existence of nine of these alleles (IGHV1-46*03, IGHV1-69*05, and
IGHV1-c included three genes (IGHV1-c01, IGHV1-c02, and IGHV1-c03), which in this study were not seen to rearrange but have previously been confirmed to be pseudogenes (10). The IGHD repertoire is also relatively small, with few allelic variants. We have previously questioned the existence of IGHD3-3*01, IGHD3-10*02, and IGHD3-16*03 (9), and these alleles were not seen in the current study. A probable new polymorphism of the IGHD6 gene, which we designate as IGHD6*05, was seen in one individual (see Supplemental Table I). There were 902 alignments to this putative polymorphism, of which 588 were perfectly matched to the sequence. An updated version of the iHMMune-align IGHD repertoire is presented as Supplemental Table V.

The IGHD gene repertoire is also relatively small, with few reported allelic variants. We have previously questioned the existence of IGHD3-3*02, IGHD3-10*02, and IGHD3-16*03 (9), and these alleles were not seen in the current study. A probable new polymorphism of the IGHD6 gene, which we designate as IGHD6*05, was seen in one individual (see Supplemental Table I). There were 902 alignments to this putative polymorphism, of which 588 were perfectly matched to the sequence. An updated version of the iHMMune-align IGHD repertoire is presented as Supplemental Table V.

The IGHD gene repertoire is also relatively small, with few reported allelic variants. We have previously questioned the existence of IGHD3-3*02, IGHD3-10*02, and IGHD3-16*03 (9), and these alleles were not seen in the current study. A probable new polymorphism of the IGHD6 gene, which we designate as IGHD6*05, was seen in one individual (see Supplemental Table I). There were 902 alignments to this putative polymorphism, of which 588 were perfectly matched to the sequence. An updated version of the iHMMune-align IGHD repertoire is presented as Supplemental Table V.

The IGHD repertoire is also relatively small, with few reported allelic variants. We have previously questioned the existence of IGHD3-3*02, IGHD3-10*02, and IGHD3-16*03 (9), and these alleles were not seen in the current study. A probable new polymorphism of the IGHD6 gene, which we designate as IGHD6*05, was seen in one individual (see Supplemental Table I). There were 902 alignments to this putative polymorphism, of which 588 were perfectly matched to the sequence. An updated version of the iHMMune-align IGHD repertoire is presented as Supplemental Table V.
gene rearrangement, and the rearrangement frequencies of different IGHD genes in this patient are compared with those of the 11 other individuals in Table IV. The patient with aplastic anemia has an almost complete absence of sequences that seem to use the IGHD1-7, IGHD2-8, IGHD3-3, IGHD4-4/4-11, and IGHD6-6 genes. Analysis of the relatively few alignments that were made to these genes shows that the IGHD alignments are almost invariably short and include more mismatches than alignments to the same genes in other individuals (Supplemental Table III). Virtually all these sequences were found to align equally well to alternative IGHD genes, and no unequivocal alignments could be found to any of the five genes.

It is impossible to differentiate between the presence of IGHD4-4 and IGHD4-11 in VDJ rearrangements, because the coding regions of these genes are identical. IGHD4-11 is defined by IMGT as an ORF that is incapable of recombination because of an unusual 5’ heptamer recombination signaling sequence (19). IGHD4-4 and IGHD4-11 in VDJ rearrangements, because the coding regions of these genes are identical. IGHD4-11 is defined by IMGT as an ORF that is incapable of recombination because of an unusual 5’ heptamer recombination signaling sequence (19). The almost complete absence of alignments to IGHD4-4/4-11 therefore probably reflects an absence of rearrangements using the IGHD4-4 gene. It is also impossible to distinguish between the identical IGHD5-5 and IGHD5-18 genes. The alignment of sequences to the IGHD5-5/5-18 genes could be entirely the result of VDJ rearrangements using the IGHD5-18 gene, and certainly the individual concerned had the second lowest proportion of sequences that aligned to IGHD5-5/5-18. We therefore suggest that this individual may have a deletion of six contiguous genes: IGHD5-18, IGHD5-24, IGHD5-26, IGHD5-28, IGHD5-30, and IGHD5-32.

Table III. The relative proportions of rearrangements that contain different alleles in individuals whose genotype includes multiple alleles.

<table>
<thead>
<tr>
<th>Allele pair</th>
<th>Rearranged (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGHV1-2*02: *04</td>
<td>75:25</td>
</tr>
<tr>
<td>IGHV1-3*01: *02</td>
<td>85:15</td>
</tr>
<tr>
<td>IGHV1-46*01: *03</td>
<td>69:31</td>
</tr>
<tr>
<td>IGHV1-69*01: *02</td>
<td>66:34</td>
</tr>
<tr>
<td>IGHV1-69*01: *06</td>
<td>70:30</td>
</tr>
<tr>
<td>IGHV2-5*01: *p11</td>
<td>60:40</td>
</tr>
<tr>
<td>IGHV2-5*10: *p11</td>
<td>14:86</td>
</tr>
<tr>
<td>IGHV2-70*01: *p14</td>
<td>66:34</td>
</tr>
<tr>
<td>IGHV2-70*01: *11: *p14</td>
<td>36:41:23</td>
</tr>
<tr>
<td>IGHV3-11*01: *p05</td>
<td>72:28</td>
</tr>
<tr>
<td>IGHV3-11*01: *p05</td>
<td>47:39:13</td>
</tr>
<tr>
<td>IGHV3-15*01: *07</td>
<td>39:61</td>
</tr>
<tr>
<td>IGHV3-30*03: *p18</td>
<td>21:79</td>
</tr>
<tr>
<td>IGHV3-30*03: *p04-18</td>
<td>15:55:32</td>
</tr>
<tr>
<td>IGHV3-33*01: *02</td>
<td>82:18</td>
</tr>
<tr>
<td>IGHV3-48*01: *03</td>
<td>47:53</td>
</tr>
<tr>
<td>IGHV3-48*02: *03</td>
<td>38:62</td>
</tr>
<tr>
<td>IGHV3-64*02: *p07</td>
<td>28:72</td>
</tr>
<tr>
<td>IGHV3-66*01: *03</td>
<td>71:29</td>
</tr>
<tr>
<td>IGHV3-66*02: *03</td>
<td>38:62</td>
</tr>
<tr>
<td>IGHV4-4*02: *07</td>
<td>62:38</td>
</tr>
<tr>
<td>IGHV7-4-1*01: *024</td>
<td>74:26</td>
</tr>
</tbody>
</table>

Each column entry for each allele pair represents data derived from a single individual. Data from the tonsillar tissue-derived sequences are not included.

Discussion
The individual variation in the IgH chain V region gene locus that we report in this paper has been suspected (20) but is greatly clarified by examination of large numbers of independent VDJ recombinations from single individuals. IGH sequence data sets from 12 individuals presented in this study each included between 40 and 46 IGHV genes that are known to contribute to the functional, expressed repertoire. This increases to between 45 and 60 IGHV genes when allelic variants are considered. Although all individuals were heterozygous at some gene loci, they were homozygous at most gene loci. Given the extent of reported allelic variation in the H chain IGHV gene repertoire, this is perhaps surprising, but a substantial proportion of these allelic variants may have been reported in error (11). In addition, there has been no data on the frequency with which different alleles are carried in the population. It seems that most gene segments exist in the study population as single common alleles, with relatively rare allelic variants.

Most frequently rearranged IGHV genes are “core” genes that were seen in all individuals. Genetic differences between individuals usually involved “noncore” genes that each make a relatively modest contribution to the overall repertoire of VDJ rearrangements, although together they accounted for 13.5% of all IGHV sequences seen in the rearrangements of one individual (data not shown). IGHV3-30-3, IGHV4-30-2, and IGHV4-30-4 have been reported as a gene cluster that is only seen in the genome of some individuals (21). These genes were seen in 7 of the 12 individuals in this study.

Although copy number polymorphisms have not previously been reported for Ig genes, the presence of more than two alleles of certain genes in the genome of some individuals is not surprising, given the extent of such structural variation within the human genome and the repetitive nature of the IgH chain locus (22). There is certainly no reason to believe that the observation of multiple alleles is a result of sequencing errors. It is possible that some bar codes may have been misread. However all bar codes were designed to differ at two or more base positions so that sequencing errors would be unlikely to convert one bar code into another. Rare chimeric PCR products could also result in swapped bar codes. If, however, contamination happened on the scale that would be required to account for the large numbers of triple allele alignments, it would lead to many genes being represented by more than two alleles in these individuals. In fact only three individuals have more than one gene that was represented by more than two alleles. In one of these individuals, the IGHV1-69 gene and the IGHV2-70 gene were each represented by three alleles, suggesting a single duplication event involving these contiguous genes. It is also noteworthy that three individuals carried the same
three IGHV3-30 alleles, whereas other individuals carried just two of the three alleles. Mutational hot spots and systematic errors, such as the formation of chimeric sequences, cannot therefore account for the observations.

Individual variation was also seen in both the IGHD and IGHJ regions. In the IGHJ region, variation was confined to the IGHJ6 gene, where 7 of 12 individuals were homozygous for IGHJ6*02, one individual was homozygous for IGHJ6*03, three individuals were heterozygous with IGHJ6*02 and IGHJ6*03, whereas one individual produced rearranged sequences using IGHJ6*02 and the putative allele IGHJ6*p05. IGHJ6*02 and IGHJ6*03 define the previously reported “b” and “c” IGHJ haplotypes (23), where these IGHJ alleles are found in association with the IGHJ genes IGHJ1*01, IGHJ2*01, IGHJ3*02, IGHJ4*02, and IGHJ5*02. Variation of the IGHD region was seen in the IGHD2-2, IGHD2-8, IGHD2-21, IGHD3-10, and IGHD3-16 genes, although few individuals were heterozygous at any loci, except IGHD2-2 and IGHD2-21.

The consequences of allelic differences are worthy of further scrutiny. The number of heterozygous loci varied substantially between individuals, leading almost certainly to variation in the diversity of the expressed H chain repertoire between individuals. And although many alleles differ from one another by just a single nucleotide, such differences can be significant. Allelic differences can lead to quite different Ag-binding properties. It has been shown, for example, that although IGHV3-23*01 may form Abs that are critical for defense against encapsulated Haemophilus influenzae, the IGHV3-23*03 allele is unable to do so (24).

This study also demonstrates that highly similar IGHV alleles may have different propensities for rearrangement and therefore for expression. These differences were even seen in alleles such as IGHV1-46*01 and IGHV1-46*03, where the different sequences encode identical amino acid sequences. It therefore seems unlikely that the variable frequencies of alleles in the data sets reflect differences in either self-reactivity or reactivity to foreign Ag. Differences could be the results of variation in the recombination signal sequences that are located at the 3’ end of each IGHV gene. Such variation is known to influence the frequency of rearrangement (25), but because recombination signal sequences have not been reported for all genes and have been reported for very few allelic variants, this cannot yet be determined.

Variations in the translational efficiency of different codons could also contribute to differences between alleles in their propensity to appear in successful VDJ rearrangements in B cells in peripheral blood. There are, however, often 2- or 3-fold differences in the rearrangement frequency, whereas the alleles in question often differ from one another by just a single codon. Translational efficiency is therefore unlikely to be the major determinant of the differences between alleles.

That particular alleles are seen to recombine at similar frequencies in different individuals is unlikely to be a consequence of similar histories of Ag exposure, for similar recombination frequencies were seen in VDI rearrangements when analysis was confined to unmutated sequences (data not shown). These sequences must be derived primarily from naive B cells, and the frequency with which each allele is represented in the data set should therefore be unaffected by Ag selection. Those pseudogenes that are capable of recombination also showed consistency of rearrangement frequency between individuals.

The use of high-throughput sequencing to explore individual differences in the Ig locus is perhaps best illustrated by the certainty with which the genotype of the IGHD locus could be determined in this study. The possible importance of such investigations into individual genotypes is demonstrated by our identification in this study of a major deletion of IGHD genes in an individual with aplastic anemia. Such a deletion would significantly impact the possible expressed repertoire of this individual. It has been argued that variation in the expressed

| Table IV. Mean frequency and range of alignments to IGHD genes in 11 healthy individuals and in an individual with aplastic anemia |
|---------------------------------------------------|------------------------------------------------|-----------------|
| Mean ± SEM (Healthy Individuals) (%) | Range (Healthy Individuals) (%) | AA Sample |
| IGHD1-1 | 0.7 ± 0.1 | 0.5-0.8 | 0.5 |
| IGHD2-2 | 7.4 ± 0.8 | 4.3-13.2 | 3.3 |
| IGHD3-3 | 7.8 ± 0.9 | 2.9-11.3 | 0.3 |
| IGHD4-4/4-11a | 1.0 ± 0.1 | 0.5-1.6 | 0.1 |
| IGHD5-5/5-18b | 3.6 ± 0.3 | 2.4-4.4 | 2.9 |
| IGHD6-6 | 3.6 ± 0.4 | 2.1-4.4 | 0.7 |
| IGHD1-7 | 1.0 ± 0.2 | 0.5-1.0 | 0.1 |
| IGHD2-8 | 1.2 ± 0.2 | 0.4-1.6 | 0.2 |
| IGHD3-9 | 5.3 ± 0.4 | 10.8-5.2 | 4.7 |
| IGHD3-10 | 10.5 ± 1.3 | 9.9-15.4 | 20.8 |
| IGHD5-12 | 2.3 ± 0.2 | 1.4-3.0 | 2.4 |
| IGHD6-13 | 7.5 ± 0.5 | 5.5-12.1 | 13.0 |
| IGHD2-15 | 4.9 ± 0.3 | 3.0-5.6 | 3.6 |
| IGHD3-16 | 2.0 ± 0.1 | 1.2-2.8 | 2.2 |
| IGHD4-17 | 5.0 ± 0.3 | 3.6-6.3 | 4.4 |
| IGHD6-19 | 7.6 ± 0.3 | 6.6-9.1 | 11.8 |
| IGHD1-20 | 0.7 ± 0.1 | 4.2-6.9 | 5.6 |
| IGHD2-21 | 2.4 ± 0.1 | 2.1-4.4 | 1.8 |
| IGHD3-22 | 12.2 ± 0.8 | 11.3-14.8 | 13.0 |
| IGHD4-23 | 1.4 ± 0.1 | 0.8-2.1 | 1.5 |
| IGHD5-24 | 2.1 ± 0.1 | 1.8-2.6 | 1.6 |
| IGHD1-26 | 5.1 ± 0.2 | 4.2-6.9 | 5.6 |
| IGHD7-27 | 1.0 ± 0.1 | 0.6-1.7 | 1.4 |

Data for six IGHD genes that may not be present in the genome of the AA patient are in boldface.
aIGHD4-4 and IGHD4-11 have identical coding regions, but IGHD4-11 is associated with an apparently nonfunctional recombination signal sequence.
bIGHD5-5 and IGHD5-18 have identical coding regions, and both genes are normally functional. The alignments shown for the AA patient could all use the IGHD5-18 gene.

AA, aplastic anemia.
repertoire could lead to the emergence of self-reactive B cell clones (reviewed in Ref. 26). Whether or not an absence of IGHD genes may play a causal role in the development of aplastic anemia is deserving of further study.

Investigation of genotypic differences in the H chain V region gene locus requires a detailed understanding of the allelic variants that are found in the human population. We have previously argued that few common IGHV gene polymorphisms remain to be discovered, other than alleles of rarely expressed genes, and perhaps uncommon alleles of more frequently expressed genes (11). In this study, however, a large number of previously unreported polymorphisms were identified, and some of them were found to be relatively commonly carried in the genomes of the study population. The IGHV3-7*03 allele, for example, was seen in 4 of 12 individuals and was found in between 1.0 and 2.0% of all VDJ rearrangements in those individuals. The IGHV2-5*01 allele that we first inferred in an earlier study (11) was seen in all individuals examined and was found in ∼1.4% of the 108,210 VDJ rearrangements analyzed. Many other putative polymorphisms were present at lower frequency in the sequence data set but were clearly detectable. Nine of the 14 new putative polymorphisms were each seen in just a single individual and were present in as few as 0.2% of the VDJ rearrangements of that individual. Nevertheless, this amounted to between 27 and 69 examples of the sequences, allowing even these putative polymorphisms to be identified with confidence.

The immunoreceptor gene loci are arguably the most complex loci in the mammalian genome, and this complexity has hampered the study of human Ig gene sequences. Nevertheless, insights into both protective and pathogenic immune responses have been gained over a period of 25 y by the reporting of thousands of sequences. Because the diversity of the peripheral blood IgH repertoire is thought to include, at a lower estimate, millions of distinct H chain sequences (12), the combined data of prior studies likely represent only a first glimpse of the available repertoire and of variation within human populations. Further application of high-throughput sequencing to the study of Ig and TCR genes, both in V(D)J rearrangements and via whole genome sequencing efforts now under way, will allow the repertoire of rearranged genes, and even individual variation in the available repertoire, to be explored in detail.

Disclosures
The authors have no financial conflicts of interest.

References
Supplementary Table 1: Putative IGHD, IGHJ and IGHV genes inferred as partial gene sequences from alignments of VDJ gene rearrangements. Sequences are shown with suggested names as ‘p’ or ‘putative’ alleles. For each sequence, the nearest previously-reported allele is shown in brackets, and nucleotide differences with respect to that allele are shown in upper case.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Allele</th>
<th>Nucleotide Differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;IGHD3-16<em>p03 (IGHD3-16</em>02)</td>
<td>gtattatgattacAtttgggggagttatcgttatacc</td>
<td></td>
</tr>
<tr>
<td>&gt;IGHJ6<em>p05 (IGHJ6</em>03)</td>
<td>attactactactactactagggggaagttatcgttatacc</td>
<td></td>
</tr>
<tr>
<td>&gt;humIGHV181*p02 (humIGHV181)</td>
<td>aagggcctgtgggagctcaaccccttaatgataacacacactacAcacagaagtgctgggggcagagtcacat</td>
<td></td>
</tr>
<tr>
<td>&gt;IGHV1-2<em>p05 (IGHV1-2</em>01)</td>
<td>gggcttgagttgatggagctgctcagcccctaatgataacacacactacAcacagaagtgctgggggcagagtcacat</td>
<td></td>
</tr>
<tr>
<td>&gt;IGHV1-2<em>p06 (IGHV1-2</em>02)</td>
<td>gggcttgagttgatggagctgctcagcccctaatgataacacacactacAcacagaagtgctgggggcagagtcacat</td>
<td></td>
</tr>
<tr>
<td>&gt;IGHV1-46<em>p04 (IGHV1-46</em>01)</td>
<td>gggcttgagttgatggagctgctcagcccctaatgataacacacactacAcacagaagtgctgggggcagagtcacat</td>
<td></td>
</tr>
<tr>
<td>&gt;IGHV1-f<em>p03 (IGHV1-f</em>02)</td>
<td>gggcttgagttgatggagctgctcagcccctaatgataacacacactacAcacagaagtgctgggggcagagtcacat</td>
<td></td>
</tr>
<tr>
<td>&gt;IGHV2-70<em>p15 (IGHV2-70</em>01)</td>
<td>ccctgtagttgctctgctcagcccctaatgataacacacactacAcacagaagtgctgggggcagagtcacat</td>
<td></td>
</tr>
<tr>
<td>&gt;IGHV2-70<em>p16 (IGHV2-70</em>11)</td>
<td>ccctgtagttgctctgctcagcccctaatgataacacacactacAcacagaagtgctgggggcagagtcacat</td>
<td></td>
</tr>
<tr>
<td>&gt;IGHV2-70<em>p17 (IGHV2-70</em>01)</td>
<td>ccctgtagttgctctgctcagcccctaatgataacacacactacAcacagaagtgctgggggcagagtcacat</td>
<td></td>
</tr>
<tr>
<td>&gt;IGHV3-20<em>p02 (IGHV3-20</em>01)</td>
<td>aaggggctgagttgctctgctcagcccctaatgataacacacactacAcacagaagtgctgggggcagagtcacat</td>
<td></td>
</tr>
<tr>
<td>&gt;IGHV3-30<em>p20 (IGHV3-30</em>02)</td>
<td>aaggggctgagttgctctgctcagcccctaatgataacacacactacAcacagaagtgctgggggcagagtcacat</td>
<td></td>
</tr>
</tbody>
</table>
aaggggctggagtgggtggcatttatacggtatgatggaagtaataaatactacgcagactccgtgaagggccgattcacatctccagagacaattcacaagaacacgcgtgtatctgcaaatgaaacagctgagctgaggacacggctgtgtattactgtgcagaga

>IGHV3-30*p21 (IGHV3-30*01)
aaggggctggagtgggtggcagttatatcatatgatggaagtaataaatactacgcagactccgtgaagggccgattcacatctccagagacaattcacaagaacacgcgtgtatctgcaaatgaaacagctgagctgaggacacggctgtgtattactgtgcagaga

>IGHV3-74*p04 (IGHV3-74*01)
aaggggctgggtgtcgcagtcgttaataatctgatgggagtagcacaagctgagctgaggacacggctgtgtattactgtgcagaga

>IGHV4-39*p08 (IGHV4-39*07)
aggtggaggtgattggatgccatctatataatgtgggagacactaaccctcctcctcactgagtggtcgtcagcatattacagtagacagtccacagacttggtgcaggcgcAgcggacacggccggtgtattactgtgcagaga

>IGHV4-59*p11 (IGHV4-59*01)
ggactggagtggattggtatatattacagtggaggacacaactacaaccctcctcactgagtggtcgtcagcatattacagtagacagtccacagacttggtgcaggcgcAgcggacacggccggtgtattactgtgcagaga
Supplementary Table 2: The number of individuals, amongst twelve studied, whose genome was inferred to include various IGHV genes whose existence has previously been questioned (11). The total number of VDJ rearranged genes that aligned perfectly to each of those genes is shown.

<table>
<thead>
<tr>
<th></th>
<th>No. of Individuals</th>
<th>Total alignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGHV1-46*03</td>
<td>2</td>
<td>182</td>
</tr>
<tr>
<td>IGHV1-69*05</td>
<td>1</td>
<td>192</td>
</tr>
<tr>
<td>IGHV1-69*10</td>
<td>1</td>
<td>518</td>
</tr>
<tr>
<td>IGHV1-f*01</td>
<td>4</td>
<td>273</td>
</tr>
<tr>
<td>IGHV2-70*12</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>IGHV3-15*07</td>
<td>3</td>
<td>231</td>
</tr>
<tr>
<td>IGHV3-33*05</td>
<td>6</td>
<td>480</td>
</tr>
<tr>
<td>IGHV3-66*02</td>
<td>6</td>
<td>166</td>
</tr>
<tr>
<td>IGHV4-31*01</td>
<td>1</td>
<td>42</td>
</tr>
</tbody>
</table>
Supplementary Table 3: Comparison of IGHD alignment length and mutation frequency seen in VDJ datasets from a healthy individual (S21: n=13043) and an individual with aplastic anaemia (S22: n=15947). Data are shown for six IGHD genes (shaded) that appear to be absent from the genome of the AA patient, as well as for the short IGHD4-17 gene and the longer IGHD3-9 gene.

<table>
<thead>
<tr>
<th></th>
<th>S21 No.</th>
<th>S21 Mean D length</th>
<th>S21 % Mutated</th>
<th>S22 No.</th>
<th>S22 Mean D length</th>
<th>S22 % Mutated</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGHD1-7</td>
<td>88</td>
<td>11.6</td>
<td>8.0</td>
<td>16</td>
<td>9.6</td>
<td>50.0</td>
</tr>
<tr>
<td>IGHD2-8</td>
<td>99</td>
<td>18.2</td>
<td>14.1</td>
<td>35</td>
<td>10.4</td>
<td>62.9</td>
</tr>
<tr>
<td>IGHD3-3</td>
<td>714</td>
<td>19.0</td>
<td>11.5</td>
<td>52</td>
<td>9.5</td>
<td>38.5</td>
</tr>
<tr>
<td>IGHD4-4/11</td>
<td>95</td>
<td>10.1</td>
<td>11.6</td>
<td>23</td>
<td>9.0</td>
<td>21.7</td>
</tr>
<tr>
<td>IGHD5-5/18</td>
<td>586</td>
<td>14.1</td>
<td>14.7</td>
<td>475</td>
<td>14.2</td>
<td>14.5</td>
</tr>
<tr>
<td>IGHD6-6</td>
<td>285</td>
<td>11.1</td>
<td>23.9</td>
<td>120</td>
<td>9.5</td>
<td>26.7</td>
</tr>
<tr>
<td>IGHD4-17</td>
<td>684</td>
<td>11.7</td>
<td>5.7</td>
<td>718</td>
<td>12.0</td>
<td>5.3</td>
</tr>
<tr>
<td>IGHD3-9</td>
<td>695</td>
<td>19.0</td>
<td>7.9</td>
<td>766</td>
<td>19.2</td>
<td>13.1</td>
</tr>
</tbody>
</table>
Supplementary Table 4: The iHMUnne-align IGHV repertoire. Polymorphisms that were first identified in this study are included as partial sequences. Pseudogenes are indicated as (P).

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;humIGHV177 (P)</td>
<td>ATGACTCACCAGCTTTGATAGACTCTCTCTGTTGCCCAGCTCATGGGCTTC ATGGAAGGTCTCCTGCAAAGCTCTGGGATATACACCCTCCAGCTAGCGACAT TCACTGTGCTGACAGGCCCCCTGGATAAAGGTTGTTGAATGATGGTAGAGGA GCTACTCTGGGATATGGTAGTACAGCTGCTAGCACAGAAGTCTGAGGCA GTCACCAGCAGGGACACAGTCACCACGAGCAGCAGCCTACATGAGCTGAG CAGTCCAGAGATCTCGAGACATAGATGTACTGTGACTACTGTGAGGAGA</td>
</tr>
<tr>
<td>&gt; humIGHV181 (P)</td>
<td>CCAAGGTTCAGCTGGTGAGCTGGAGTGCAAGAAGCCTGCGCTCT CAGTGAAGGTCTCCTGCAAAGCTCTCCAGATACACTTCACCCAAATACTTTA CACAGTGGCTGACAGGGCCCTGGGACACAGGCTATGCTGGATGACAT GCATACCCCTCAATTGATAAACACACTACGACAGAAGTCTCCAGGCAC AGTTCACCATTACCAGTAGCAGGCTCCTGAGCACAGCTCAGTACAGCTGAG GCAGCCTGAGATCTGAAAGACATGGCTGTCGTAATCTCTGTGTGAAGA</td>
</tr>
<tr>
<td>&gt; humIGHV181*p02 (partial) (P)</td>
<td>AGGGCATAGTGTTGGGATGCACTCAAACCTCTACATGATACACACTACA CACACAGAAGTCTCCAGGGGCACAGTCACCATTACCACTGACAGGCTCTGTA GCACACCCCTATGAGGACAGCAGCTCAGCTACATGAGCTGAGGAGCAGC TATCTCGTGAGGA</td>
</tr>
<tr>
<td>&gt;IGHV1-2*01</td>
<td>CAGGTGCAAGCTGGTGAGCTGGAGTGAGAGAAGACCTGCGCTCT CAGTGAAGGTCTCCTGCAAAGCTCTCCAGATACACTTCACCCAAATACTTTA CACAGTGGCTGACAGGGCCCTGGGACACAGGCTATGCTGGATGACAT GCATACCCCTCAATTGATAAACACACTACGACAGAAGTCTCCAGGCAC AGTTCACCATTACCAGTAGCAGGCTCCTGAGCACAGCTCAGTACAGCTGAG GCAGCCTGAGATCTGAAAGACATGGCTGTCGTAATCTCTGTGTGAAGA</td>
</tr>
<tr>
<td>&gt;IGHV1-2*02</td>
<td>CAGGTGCAAGCTGGTGAGCTGGAGTGAGAGAAGACCTGCGCTCT CAGTGAAGGTCTCCTGCAAAGCTCTCCAGATACACTTCACCCAAATACTTTA CACAGTGGCTGACAGGGCCCTGGGACACAGGCTATGCTGGATGACAT GCATACCCCTCAATTGATAAACACACTACGACAGAAGTCTCCAGGCAC AGTTCACCATTACCAGTAGCAGGCTCCTGAGCACAGCTCAGTACAGCTGAG GCAGCCTGAGATCTGAAAGACATGGCTGTCGTAATCTCTGTGTGAAGA</td>
</tr>
<tr>
<td>&gt;IGHV1-2*04</td>
<td>CAGGTGCAAGCTGGTGAGCTGGAGTGAGAGAAGACCTGCGCTCT CAGTGAAGGTCTCCTGCAAAGCTCTCCAGATACACTTCACCCAAATACTTTA CACAGTGGCTGACAGGGCCCTGGGACACAGGCTATGCTGGATGACAT GCATACCCCTCAATTGATAAACACACTACGACAGAAGTCTCCAGGCAC AGTTCACCATTACCAGTAGCAGGCTCCTGAGCACAGCTCAGTACAGCTGAG GCAGCCTGAGATCTGAAAGACATGGCTGTCGTAATCTCTGTGTGAAGA</td>
</tr>
<tr>
<td>&gt;IGHV1-2*05 (partial)</td>
<td>GGGCTTGAGTTGGGAGGCAGCAGCCACACCTGAGCAGGACAGCCTCAGTCA GCACAGCGCTCTACATGAGGAGGCTGAGCAGCTCAGCTACATGAGCTGAG GCAGGCTGAGATCTGAGAGACAGCGCGCTGTTACTACTGTGAGGAGA</td>
</tr>
<tr>
<td>&gt;IGHV1-2*06 (partial)</td>
<td>GGGCTTGAGTTGGGAGGCAGCAGCCACACCTGAGCAGGACAGCCTCAGTCA GCACAGCGCTCTACATGAGGAGGCTGAGCAGCTCAGCTACATGAGCTGAG GCAGGCTGAGATCTGAGAGACAGCGCGCTGTTACTACTGTGAGGAGA</td>
</tr>
</tbody>
</table>
ATCACACCTTTCAATGGTAACACCAACTACGCACAGAAATTCCAGGACAGAGTCACCATTACTAGGGACAGGTCTATGAGCACAGCCTACATGGAGCTGA
GCAGCCTGAGATCTGAGGACAGACAGCCATGTATTACTGTGCAAGAXA
>IGHV1-45*02
CAGATGCAGCTGGTGACGTGCTGGGCTGAGGTAAGAAGACGTGGTCTCCT
AGTGAAGTTTTTCTGCAAGGGCTTCCGGAATACACCTCACCCTACCGCTATCCT
GCACGTGGTGCGACAGGGCCGGCAGACAGGGCTTGGATGAGGATGGAATG
ATCACACCCCTAATGGAACACACTACGCAAGAAATTCAGAGGACAG
AGTCAACCATTACCCGAGCAAGTCACTATGAGCAACAGGCTACATGGAGCTGA
GCAGCCTGAGATCTGAGGACAGACAGCCATGTATTACTGTGCAAGAXA
>IGHV1-46*01
CAGGTTGAGCTGGGTCATGCTTGGGGCTGAGGTAAGAAGACGTGGGCTCCT
AGTGAAGGGTTTTCCCTGCAAGGGCTTCCGGAATACACCTCACCCTACCGCTATAT
GCACGTGGTGCGACAGGGCCGGCAGACAGGGCTTGGATGAGGATGGAATATG
ATCAACCCCTAGTGGTGTAGCAACACTACGCAAGAAATTCAGAGGACAG
AGTCAACCATTACCCGAGCAAGTCACTATGAGCAACAGGCTACATGGAGCTGA
GCAGCCTGAGATCTGAGGACAGACAGCCATGTATTACTGTGCAAGAXA
>IGHV1-46*02
CAGGTTGAGCTGGGTCATGCTTGGGGCTGAGGTAAGAAGACGTGGGCTCCT
AGTGAAGGGTTTTCCCTGCAAGGGCTTCCGGAATACACCTCACCCTACCGCTATAT
GCACGTGGTGCGACAGGGCCGGCAGACAGGGCTTGGATGAGGATGGAATATG
ATCAACCCCTAGTGGTGTAGCAACACTACGCAAGAAATTCAGAGGACAG
AGTCAACCATTACCCGAGCAAGTCACTATGAGCAACAGGCTACATGGAGCTGA
GCAGCCTGAGATCTGAGGACAGACAGCCATGTATTACTGTGCAAGAXA
>IGHV1-46*03
CAGGTTGAGCTGGGTCATGCTTGGGGCTGAGGTAAGAAGACGTGGGCTCCT
AGTGAAGGGTTTTCCCTGCAAGGGCTTCCGGAATACACCTCACCCTACCGCTATAT
GCACGTGGTGCGACAGGGCCGGCAGACAGGGCTTGGATGAGGATGGAATATG
ATCAACCCCTAGTGGTGTAGCAACACTACGCAAGAAATTCAGAGGACAG
AGTCAACCATTACCCGAGCAAGTCACTATGAGCAACAGGCTACATGGAGCTGA
GCAGCCTGAGATCTGAGGACAGACAGCCATGTATTACTGTGCAAGAXA
>IGHV1-46*p04 (partial)
GGGCTTGAGTGGATGGGAATAATCAACCCCTAGTGGTGTAGCAACACTACGCAAGAAATTCAGAGGACAG
AGTCAACCATTACCCGAGCAAGTCACTATGAGCAACAGGCTACATGGAGCTGA
GCAGCCTGAGATCTGAGGACAGACAGCCATGTATTACTGTGCAAGAXA
>IGHV1-58*01
CAATGCAAGCTGGTGACGTGCTGGGCTGAGGTAAGAAGACGTGGGCTCCT
AGTGAAGGGTTTTCCCTGCAAGGGCTTCCGGAATACACCTCACCCTACCGCTATAT
GCACGTGGTGCGACAGGGCCGGCAGACAGGGCTTGGATGAGGATGGAATATG
ATCACACCCCTAGTGGTGTAGCAACACTACGCAAGAAATTCAGAGGACAG
AGTCAACCATTACCCGAGCAAGTCACTATGAGCAACAGGCTACATGGAGCTGA
GCAGCCTGAGATCCGAGGACAGACAGCCCTGTAATTACTGTGCCAGAGA
>IGHV1-58*02
CAATGCAAGCTGGTGACGTGCTGGGCTGAGGTAAGAAGACGTGGGCTCCT
AGTGAAGGGTTTTCCCTGCAAGGGCTTCCGGAATACACCTCACCCTACCGCTATAT
GCACGTGGTGCGACAGGGCCGGCAGACAGGGCTTGGATGAGGATGGAATATG
ATCACACCCCTAGTGGTGTAGCAACACTACGCAAGAAATTCAGAGGACAG
AGTCAACCATTACCCGAGCAAGTCACTATGAGCAACAGGCTACATGGAGCTGA
GCAGCCTGAGATCCGAGGACAGACAGCCCTGTAATTACTGTGCCAGAGA
>IGHV1-69*01
CAGGTGCAGCTGTCGACGTCTGCTCTGGGCTGAGGTGAAGAAGCCTGGGTCCTCGG
CACGCTCTCTGCAAGGCTTCTCTGAGGCACTTCTACGAGCTATGCTACTCAGCTG
TAGCTGGATCTGAGGACACGGCCGTGTATTACTGTGCGAGAGA

>IGHV1-69*02
CAGGTGCAGCTGTCGACGTCTGCTCTGGGCTGAGGTGAAGAAGCCTGGGTCCTCGG
CACGCTCTCTGCAAGGCTTCTCTTGAGGCACTTCTACGAGCTATGCTACTCAGCTG
TAGCTGGATCTGAGGACACGGCCGTGTATTACTGTGCGAGAGA

>IGHV1-69*04
CAGGTGCAGCTGTCGACGTCTGCTCTGGGCTGAGGTGAAGAAGCCTGGGTCCTCGG
CACGCTCTCTGCAAGGCTTCTCTTGAGGCACTTCTACGAGCTATGCTACTCAGCTG
TAGCTGGATCTGAGGACACGGCCGTGTATTACTGTGCGAGAGA

>IGHV1-69*05
CAGGTGCAGCTGTCGACGTCTGCTCTGGGCTGAGGTGAAGAAGCCTGGGTCCTCGG
CACGCTCTCTGCAAGGCTTCTCTTGAGGCACTTCTACGAGCTATGCTACTCAGCTG
TAGCTGGATCTGAGGACACGGCCGTGTATTACTGTGCGAGAGA

>IGHV1-69*06
CAGGTGCAGCTGTCGACGTCTGCTCTGGGCTGAGGTGAAGAAGCCTGGGTCCTCGG
CACGCTCTCTGCAAGGCTTCTCTTGAGGCACTTCTACGAGCTATGCTACTCAGCTG
TAGCTGGATCTGAGGACACGGCCGTGTATTACTGTGCGAGAGA

>IGHV1-69*08
CAGGTGCAGCTGTCGACGTCTGCTCTGGGCTGAGGTGAAGAAGCCTGGGTCCTCGG
CACGCTCTCTGCAAGGCTTCTCTTGAGGCACTTCTACGAGCTATGCTACTCAGCTG
TAGCTGGATCTGAGGACACGGCCGTGTATTACTGTGCGAGAGA

>IGHV1-69*09
CAGGTGCAGCTGTCGACGTCTGCTCTGGGCTGAGGTGAAGAAGCCTGGGTCCTCGG
CACGCTCTCTGCAAGGCTTCTCTTGAGGCACTTCTACGAGCTATGCTACTCAGCTG
TAGCTGGATCTGAGGACACGGCCGTGTATTACTGTGCGAGAGA

>IGHV1-69*10
CACTCATTATTGGAAATGATGATAAGCGTACGACCACATCTCTGAAGAGC AGGCTCACCATACCAAGGACACCTCAGAAAAACCAGGATGCTCCTTACAT GACCAACATGGACCCCTGTGGAACAGCAGCCACATATTACTGTGCACACAGAC C
>IGHV2-5*10
CAGATCACCTTGAAAGAGATCGTGTCCCAGATCGTGGAACCCACACAGAC CCTCAGCTGTGACGGGTAACCCGCTGCGAGTTCTCTGGGTCAGCTGAGG
GGGTTGGGCTGGGATCGTCCGCGCAGGAAAGGCGCTGGGAGCTGAGTGGCTTGG CACCTCTTTTCTGTGCATAGGCTGCTGTCTTCGAGGAGAGGAGTGGCTTG GAACCGTACCATCAGCAAGGACACCTCCAAAAACCAACGGTGGGTCTTACAT GACCAACATGGACCCCTGTGGAACAGCAGCCACATATTACTGTGCACACAGAC C
>IGHV2-5*p11
CAGATCACCTTGAAAGAGATCGTGTCCCAGATCGTGGAACCCACACAGAC CCTCAGCTGTGACGGGTAATCCGCTGCGAGTTCTCTGGGTCAGCTGAGG
GGGTTGGGCTGGGATCGTCCGCGCAGGAAAGGCGCTGGGAGCTGAGTGGCTTGG CACCTCTTTTCTGTGCATAGGCTGCTGTCTTCGAGGAGAGGAGTGGCTTG GAACCGTACCATCAGCAAGGACACCTCCAAAAACCAACGGTGGGTCTTACAT GACCAACATGGACCCCTGTGGAACAGCAGCCACATATTACTGTGCACACAGAC C
>IGHV2-26*01
CAGGTTCACCTTGAAAGGAGATCGTGTCCCAGATCGTGGAACCCACACAGAC CCTCAGCTGTGACGGGTAACCCGCTGCGAGTTCTCTGGGTCAGCTGAGG
GGGTTGGGCTGGGATCGTCCGCGCAGGAAAGGCGCTGGGAGCTGAGTGGCTTGG CACCTCTTTTCTGTGCATAGGCTGCTGTCTTCGAGGAGAGGAGTGGCTTG GAACCGTACCATCAGCAAGGACACCTCCAAAAACCAACGGTGGGTCTTACAT GACCAACATGGACCCCTGTGGAACAGCAGCCACATATTACTGTGCACACAGAC C
>IGHV2-70*01
CAGGTTCACCTTGAAAGGAGATCGTGTCCCAGATCGTGGAACCCACACAGAC CCTCAGCTGTGACGGGTAACCCGCTGCGAGTTCTCTGGGTCAGCTGAGG
GGGTTGGGCTGGGATCGTCCGCGCAGGAAAGGCGCTGGGAGCTGAGTGGCTTGG CACCTCTTTTCTGTGCATAGGCTGCTGTCTTCGAGGAGAGGAGTGGCTTG GAACCGTACCATCAGCAAGGACACCTCCAAAAACCAACGGTGGGTCTTACAT GACCAACATGGACCCCTGTGGAACAGCAGCCACATATTACTGTGCACACAGAC C
>IGHV2-70*03
CAGGTTCACCTTGAAAGGAGATCGTGTCCCAGATCGTGGAACCCACACAGAC CCTCAGCTGTGACGGGTAACCCGCTGCGAGTTCTCTGGGTCAGCTGAGG
GGGTTGGGCTGGGATCGTCCGCGCAGGAAAGGCGCTGGGAGCTGAGTGGCTTGG CACCTCTTTTCTGTGCATAGGCTGCTGTCTTCGAGGAGAGGAGTGGCTTG GAACCGTACCATCAGCAAGGACACCTCCAAAAACCAACGGTGGGTCTTACAT GACCAACATGGACCCCTGTGGAACAGCAGCCACATATTACTGTGCACACAGAC C
>IGHV2-70*10
CAGGTTCACCTTGAAAGGAGATCGTGTCCCAGATCGTGGAACCCACACAGAC CCTCAGCTGTGACGGGTAACCCGCTGCGAGTTCTCTGGGTCAGCTGAGG
GGGTTGGGCTGGGATCGTCCGCGCAGGAAAGGCGCTGGGAGCTGAGTGGCTTGG CACCTCTTTTCTGTGCATAGGCTGCTGTCTTCGAGGAGAGGAGTGGCTTG GAACCGTACCATCAGCAAGGACACCTCCAAAAACCAACGGTGGGTCTTACAT GACCAACATGGACCCCTGTGGAACAGCAGCCACATATTACTGTGCACACAGAC C
>IGHV2-70*11
CAGGTTCACCTTGAAAGGAGATCGTGTCCCAGATCGTGGAACCCACACAGAC CCTCAGCTGTGACGGGTAACCCGCTGCGAGTTCTCTGGGTCAGCTGAGG
GGGTTGGGCTGGGATCGTCCGCGCAGGAAAGGCGCTGGGAGCTGAGTGGCTTGG CACCTCTTTTCTGTGCATAGGCTGCTGTCTTCGAGGAGAGGAGTGGCTTG GAACCGTACCATCAGCAAGGACACCTCCAAAAACCAACGGTGGGTCTTACAT GACCAACATGGACCCCTGTGGAACAGCAGCCACATATTACTGTGCACACAGAC C
GTGTGTGAGCTGGATCCGTCAGCCCCCAGGGAAGGCCCTGGAGTGGCTTG
CACGCATTGAGTTGGATGATGATAAATACTACAGCACATCTCTGAAGACC
AGGCTCACCATCTCCAAAGGAACACCTCCTCAAAAACAGAGTGAGTCCTTACAAT
GACCAACATGGACCCCTGTGGACACAGCCACAGATTATTGTCACCGGATAC
>IGHV2-70*12
CAGATCACCTTGAAGGAGTCTGGTCCTACGCTGTGAAAACCCACACAGAC
CTCAACACTGAGCTCGCCCTCTTGTTCTACTCAGCACTAGTGGAAT
GTGTTGAGCTGGATCCCGCTAGCCACGAAAAGGAGGCCCTGGAGTGGCTTG
CACGTAGTTGGATGATGATAAATACTACAGCACATCTCTGAAGACC
AGGCTCACCATCTCCAAAGGAACACCTCCTCAAAAACAGAGTGAGTCCTTACAAT
GACCAACATGGACCCCTGTGGACACAGCCACAGATTATTGTCACCGGATAC
>IGHV2-70*13
CAGGTCACCTTGAAGGAGTCTGGTCCTGCGCTGTGAAAACCCACACAGAC
CTCAACACTGAGCTCGCCCTCTTGTTCTACTCAGCACTAGTGGAAT
GTGTTGAGCTGGATCCCGCTAGCCACGAAAAGGAGGCCCTGGAGTGGCTTG
CACGTAGTTGGATGATGATAAATACTACAGCACATCTCTGAAGACC
AGGCTCACCATCTCCAAAGGAACACCTCCTCAAAAACAGAGTGAGTCCTTACAAT
GACCAACATGGACCCCTGTGGACACAGCCACAGATTATTGTCACCGGATAC
>IGHV2-70*p14
CAGGTCACCTTGAAGGAGTCTGGTCCTGCGCTGTGAAAACCCACACAGAC
CTCAACACTGAGCTCGCCCTCTTGTTCTACTCAGCACTAGTGGAAT
GTGTTGAGCTGGATCCCGCTAGCCACGAAAAGGAGGCCCTGGAGTGGCTTG
CACGTAGTTGGATGATGATAAATACTACAGCACATCTCTGAAGACC
AGGCTCACCATCTCCAAAGGAACACCTCCTCAAAAACAGAGTGAGTCCTTACAAT
GACCAACATGGACCCCTGTGGACACAGCCACAGATTATTGTCACCGGATAC
>IGHV2-70*p15 (partial)
CCCTGGAGTGGCTTGCACTCATTGATCTGAGTGATGAAATACTCAGAC
ACATCTCTGAAGACCAAGGCTACATCTCCCAAGGACACCTCTCAAAAACCA
GGTGGTCCTTACAATGACCAACATGGACCCCTCTGTGGACACAGCCACCGTAT
TACTGTGACCGGATAC
>IGHV2-70*p16 (partial)
GCCCTGGAGTGGCTTGCACTCATTGATCTGAGTGATGAAATACTCAGAC
ACATCTCTGAAGACCAAGGCTACATCTCCCAAGGACACCTCTCAAAAACCA
GGTGGTCCTTACAATGACCAACATGGACCCCTCTGTGGACACAGCCACCGTAT
TACTGTGACCGGATAC
>IGHV2-70*p17 (partial)
GCCCTGGAGTGGCTTGCACTCATTGATCTGAGTGATGAAATACTCAGAC
ACATCTCTGAAGACCAAGGCTACATCTCCCAAGGACACCTCTCAAAAACCA
GGTGGTCCTTACAATGACCAACATGGACCCCTCTGTGGACACAGCCACCGTAT
TACTGTGACCGGATAC
>IGHV3-7*01
GAGGTGCACTCCTGGAGTCTGGAGGTGGAGGAGCCTTGGTCCAGCCTGGGGGCTC
CCTGAAGACTCTCTCGTCCAGCTTGGATGAAATACTCAGCACTAGTCTCGGAT
GAGCTTGCCCGCAGGCCAGCTGAGGAGGGCTGGAGTGATGGCAAGAAC
ATAAAGCAGATGAGATGAGAAATACTTGGAGTGGACCTCGGAAGGCCG
GATCCACACTCTCCAGACAAGCAGCCAAGAACTACGACTGATCTCGAAATAG
AACAGCCTGAGAGCCAGAGGCCACAGGCTGATTACTGTGCGAGAGA
>IGHV3-7*02
GAGGTGCACTCCTGGAGTCTGGAGGTGGAGGAGCCTTGGTCCAGCCTGGGGGCTC
CCTGAAGACTCTCTCGTCCAGCCTGGATGAAATACTCAGCACTAGTCTCGGAT
GAGCTTGCCCGCAGGCCAGCTGAGGAGGGCTGGAGTGATGGCAAGAAC
ATAAAGCAGATGAGATGAGAAATACTTGGAGTGGACCTCGGAAGGCCG
GATCCACACTCTCCAGACAAGCAGCCAAGAACTACGACTGATCTCGAAATAG
AACAGCCTGAGAGCCAGAGGCCACAGGCTGATTACTGTGCGAGAGA
ATGGTACTGCTGGTGACACATACTATCCAGGCTCCGTGAAGGGCGATTCCACCTCGGACAGAATAATGCAAAAGAACATTCACTGGTATCTTCATCAAATGAAACAGCTTGAGAGCCGGCCACGACCGCTGGTATTACTGTGCAAGAGA
>IGHV3-13*03
GAGGTGCACTGGTGGAGTCTGGGAGGAGCTTGGTACAGGGGCTGGGCTGTGTATTACTGTGCAAGAGA
CCTGAGAGCCGGGGACACGGCTGTGTATTACTGTGCAAGAGA
>IGHV3-13*01
GAGGTGCACTGGTGGAGTCTGGGAGGAGCTTGGTACAGGGGCTGGGCTGTGTATTACTGTGCAAGAGA
CCTGAGAGCCGGGGACACGGCTGTGTATTACTGTGCAAGAGA
>IGHV3-15*01
GAGGTGCACTGGTGGAGTCTGGGAGGAGCTTGGTACAGGGGCTGGGCTGTGTATTACTGTGCAAGAGA
CCTGAGAGCCGGGGACACGGCTGTGTATTACTGTGCAAGAGA
>IGHV3-15*02
GAGGTGCACTGGTGGAGTCTGGGAGGAGCTTGGTACAGGGGCTGGGCTGTGTATTACTGTGCAAGAGA
CCTGAGAGCCGGGGACACGGCTGTGTATTACTGTGCAAGAGA
>IGHV3-15*07
GAGGTGCACTGGTGGAGTCTGGGAGGAGCTTGGTACAGGGGCTGGGCTGTGTATTACTGTGCAAGAGA
CCTGAGAGCCGGGGACACGGCTGTGTATTACTGTGCAAGAGA
>IGHV3-20*01
GAGGTGCACTGGTGGAGTCTGGGAGGAGCTTGGTACAGGGGCTGGGCTGTGTATTACTGTGCAAGAGA
CCTGAGAGCCGGGGACACGGCTGTGTATTACTGTGCAAGAGA
>IGHV3-20*p02 (humIGHV154)
GAGGTGCACTGGTGGAGTCTGGGAGGAGCTTGGTACAGGGGCTGGGCTGTGTATTACTGTGCAAGAGA
CCTGAGAGCCGGGGACACGGCTGTGTATTACTGTGCAAGAGA
>IGHV3-21*01
GAGGTGCAGCTGGTGGAGTCTGGGGAGGCCTGGTCAAGCCTGGGGGT
CCCTGAGACTCTCCCTGTCAGCAGCTCTGGATTCACCTTCAGTAGCTATAGCA
TGAACTGGGTCCGCCAGGCTCCAGGGAAGGGGCTGGAGTGGGTCTCATCC
ATTAGTGTAGTAGTACTATATATATACTACGCAAGCTCTAGTGAAAGGCCG
ATTCACCACCACAGACAAACGCCAAGAACTACTGATATCTGCAAATGA
ACAGCCTGAGAGCCAGAGACACACGGGCTGTGATTACTGTGCAGAGAGA
>IGHV3-21*02
GAGGTGCACACTGGTGAGTCTGGGGGAGGCCTGGTCAAGCCTGGGGGT
CCCTGAGACTCTCCCTGTCAGCAGCTCTGGATTCACCTTCAGTAGCTATAGCA
TGAACTGGGTCCGCCAGGCTCCAGGGAAGGGGCTGGAGTGGGTCTCATCC
ATTAGTGTAGTAGTACTATATATATACTACGCAAGCTCTAGTGAAAGGCCG
ATTCACCACCACAGACAAACGCCAAGAACTACTGATATCTGCAAATGA
ACAGCCTGAGAGCCAGAGACACACGGGCTGTGATTACTGTGCAGAGAGA
>IGHV3-22*01 (P)
GAGGTGCATCTGGTGGAGTCTGGGGGAGCCTTGGTACAGCCTGGGGGTCCCTGAGACTCTCCCTGTCAGCAGCTCTGGATTCACCTTCAGTAGCTATAGCA
TGAACTGGGTCCGCCAGGCTCCAGGGAAGGGGCTGGAGTGGGTCTCATCC
ATTAGTGTAGTAGTACTATATATATACTACGCAAGCTCTAGTGAAAGGCCG
ATTCACCACCACAGACAAACGCCAAGAACTACTGATATCTGCAAATGA
ACAGCCTGAGAGCCAGAGACACACGGGCTGTGATTACTGTGCAGAGAGA
>IGHV3-23*01
GAGGTGCAGCTGGTGGAGTCTGGGGGAGGCTTGGTACAGCCTGGGGGTCCCTGAGACTCTCCCTGTCAGCAGCTCTGGATTCACCTTTAGCAGCTATAGCA
TGAACTGGGTCCGCCAGGCTCCAGGGAAGGGGCTGGAGTGGGTCTCATCC
ATTAGTGTAGTAGTACTATATATATACTACGCAAGCTCTAGTGAAAGGCCG
ATTCACCACCACAGACAAACGCCAAGAACTACTGATATCTGCAAATGA
ACAGCCTGAGAGCCAGAGACACACGGGCTGTGATTACTGTGCAGAGAGA
>IGHV3-23*02
GAGGTGCAGCTGGTGGAGTCTGGGGGAGGCTTGGTACAGCCTGGGGGTCCCTGAGACTCTCCCTGTCAGCAGCTCTGGATTCACCTTTAGCAGCTATAGCA
TGAACTGGGTCCGCCAGGCTCCAGGGAAGGGGCTGGAGTGGGTCTCATCC
ATTAGTGTAGTAGTACTATATATATACTACGCAAGCTCTAGTGAAAGGCCG
ATTCACCACCACAGACAAACGCCAAGAACTACTGATATCTGCAAATGA
ACAGCCTGAGAGCCAGAGACACACGGGCTGTGATTACTGTGCAGAGAGA
>IGHV3-23*03
GAGGTGCAGCTGGTGGAGTCTGGGGGAGGCTTGGTACAGCCTGGGGGTCCCTGAGACTCTCCCTGTCAGCAGCTCTGGATTCACCTTTAGCAGCTATAGCA
TGAACTGGGTCCGCCAGGCTCCAGGGAAGGGGCTGGAGTGGGTCTCATCC
ATTAGTGTAGTAGTACTATATATATACTACGCAAGCTCTAGTGAAAGGCCG
ATTCACCACCACAGACAAACGCCAAGAACTACTGATATCTGCAAATGA
ACAGCCTGAGAGCCAGAGACACACGGGCTGTGATTACTGTGCAGAGAGA
>IGHV3-23*04
GAGGTGCAGCTGGTGGAGTCTGGGGGAGGCTTGGTACAGCCTGGGGGTCCCTGAGACTCTCCCTGTCAGCAGCTCTGGATTCACCTTTAGCAGCTATAGCA
TGAACTGGGTCCGCCAGGCTCCAGGGAAGGGGCTGGAGTGGGTCTCATCC
ATTAGTGTAGTAGTACTATATATATACTACGCAAGCTCTAGTGAAAGGCCG
ATTCACCACCACAGACAAACGCCAAGAACTACTGATATCTGCAAATGA
ACAGCCTGAGAGCCAGAGACACACGGGCTGTGATTACTGTGCAGAGAGA
>IGHV3-23*05
>IGHV3-30*01
CAGGTGCACTTGTTGGAGTCTGGGGGAGGCTTGGTACAGCCTGGGGGGTC
CCTGAGACTCTCTGCTGAGCTGGGCTGGATTCACCTTTAGCAGCTATGCCAT
TTTATAGCAGTGTAGGACACATTAGCTATGCTATGCTATGCTATGCTATGCTAT
GCACGTCGGTGTCGCCAGGGAGCTCAGGGCTGGATTCACCTTTAGCAGCTATGCCAT
TATCATATGATGGAAGTAATAAATACTACGCAGACTCCGTGAAGGCCCAGGA
TTCACCATCTCCAGAAGACATTCAGAAGACAGCCTGTATCTGCAAATGCAA
CAGCCTGAGAGCTGAGGACACGGGCTGTTGTTACTGCTGAGAGAAGAAGA
>IGHV3-30*02
CAGGTGCACTTGTTGGAGTCTGGGGGAGGCTTGGTACAGCCTGGGGGGTC
CCTGAGACTCTCTGGCTGAGCTGGGCTGGATTCACCTTTAGCAGCTATGCCAT
TTTATAGCAGTGTAGGACACATTAGCTATGCTATGCTATGCTATGCTATGCTAT
GCACGTCGGTGTCGCCAGGGAGCTCAGGGCTGGATTCACCTTTAGCAGCTATGCCAT
TATCATATGATGGAAGTAATAAATACTACGCAGACTCCGTGAAGGCCCAGGA
TTCACCATCTCCAGAAGACATTCAGAAGACAGCCTGTATCTGCAAATGCAA
CAGCCTGAGAGCTGAGGACACGGGCTGTTGTTACTGCTGAGAGAAGAAGA
>IGHV3-30*03
CAGGTGCACTTGTTGGAGTCTGGGGGAGGCTTGGTACAGCCTGGGGGGTC
CCTGAGACTCTCTCTGCTGAGCTGGGCTGGATTCACCTTTAGCAGCTATGCCAT
TTTATAGCAGTGTAGGACACATTAGCTATGCTATGCTATGCTATGCTATGCTAT
GCACGTCGGTGTCGCCAGGGAGCTCAGGGCTGGATTCACCTTTAGCAGCTATGCCAT
TATCATATGATGGAAGTAATAAATACTACGCAGACTCCGTGAAGGCCCAGGA
TTCACCATCTCCAGAAGACATTCAGAAGACAGCCTGTATCTGCAAATGCAA
CAGCCTGAGAGCTGAGGACACGGGCTGTTGTTACTGCTGAGAGAAGAAGA
>IGHV3-30*04
CAGGTGCACTTGTTGGAGTCTGGGGGAGGCTTGGTACAGCCTGGGGGGTC
CCTGAGACTCTCTCTGCTGAGCTGGGCTGGATTCACCTTTAGCAGCTATGCCAT
TTTATAGCAGTGTAGGACACATTAGCTATGCTATGCTATGCTATGCTATGCTAT
GCACGTCGGTGTCGCCAGGGAGCTCAGGGCTGGATTCACCTTTAGCAGCTATGCCAT
TATCATATGATGGAAGTAATAAATACTACGCAGACTCCGTGAAGGCCCAGGA
TTCACCATCTCCAGAAGACATTCAGAAGACAGCCTGTATCTGCAAATGCAA
CAGCCTGAGAGCTGAGGACACGGGCTGTTGTTACTGCTGAGAGAAGAAGA
>IGHV3-30*10
CAGGTGCACTTGTTGGAGTCTGGGGGAGGCTTGGTACAGCCTGGGGGGTC
CCTGAGACTCTCTCTGCTGAGCTGGGCTGGATTCACCTTTAGCAGCTATGCCAT
TTTATAGCAGTGTAGGACACATTAGCTATGCTATGCTATGCTATGCTATGCTAT
GCACGTCGGTGTCGCCAGGGAGCTCAGGGCTGGATTCACCTTTAGCAGCTATGCCAT
TATCATATGATGGAAGTAATAAATACTACGCAGACTCCGTGAAGGCCCAGGA
TTCACCATCTCCAGAAGACATTCAGAAGACAGCCTGTATCTGCAAATGCAA
CAGCCTGAGAGCTGAGGACACGGGCTGTTGTTACTGCTGAGAGAAGAAGA
>IGHV3-30*14
CAGGTGCACTTGTTGGAGTCTGGGGGAGGCTTGGTACAGCCTGGGGGGTC
CCTGAGACTCTCTCTGCTGAGCTGGGCTGGATTCACCTTTAGCAGCTATGCCAT
TTTATAGCAGTGTAGGACACATTAGCTATGCTATGCTATGCTATGCTATGCTAT
GCACGTCGGTGTCGCCAGGGAGCTCAGGGCTGGATTCACCTTTAGCAGCTATGCCAT
TATCATATGATGGAAGTAATAAATACTACGCAGACTCCGTGAAGGCCCAGGA
TTCACCATCTCCAGAAGACATTCAGAAGACAGCCTGTATCTGCAAATGCAA
CAGCCTGAGAGCTGAGGACACGGGCTGTTGTTACTGCTGAGAGAAGAAGA
>IGHV3-30*18
CAGGTGCAGCTGGTGGAGTCTGGGGGAGGCTTGGTACAGCCTGGGGGGTC
CCTGAGACTCTCTCTCTGACGGCTCTGTGATACCTACCTTGGATTATATACAT
GCACGTGGGTCTCCAGGCTCTCGAGGGAGGCTTGGGAGTTGGTTGATATACAA
ATAGGATGAGCATGAGGACATGGCTGTGTATTATTGTGCAAGA
TTCACCATCTCCAGAGACAATGCCAAGAACTCACTGTATCTGCAAATGAA
CAGCCTGAGAGCCGAGGACACGGCTGTGTATTACTGTGCGAGAGA
>IGHV3-48*02
GAGGTCAGCTCTGGAGTGGAGTACAGCTCGTGAGGCTGGTGCTCAGCTCACTGTGAGA
GAGGTTCACCATCTCCAGAGACGAGGACACGGCTGTGTATTACTGTGCGAGAGA
>IGHV3-48*03
GAGGTCAGCTCTGGAGTGGAGTACAGCTCGTGAGGCTGGTGCTCAGCTCACTGTGAGA
GAGGTTCACCATCTCCAGAGACGAGGACACGGCTGTGTATTACTGTGCGAGAGA
>IGHV3-49*01
GAGGTCACCATCTCCAGAGACGAGGACACGGCTGTGTATTACTGTGCGAGAGA
GAGGTCACCATCTCCAGAGACGAGGACACGGCTGTGTATTACTGTGCGAGAGA
>IGHV3-49*03
GAGGTCACCATCTCCAGAGACGAGGACACGGCTGTGTATTACTGTGCGAGAGA
GAGGTCACCATCTCCAGAGACGAGGACACGGCTGTGTATTACTGTGCGAGAGA
>IGHV3-49*04
GAGGTCACCATCTCCAGAGACGAGGACACGGCTGTGTATTACTGTGCGAGAGA
GAGGTCACCATCTCCAGAGACGAGGACACGGCTGTGTATTACTGTGCGAGAGA
>IGHV3-49*05
GAGGTCACCATCTCCAGAGACGAGGACACGGCTGTGTATTACTGTGCGAGAGA
GAGGTCACCATCTCCAGAGACGAGGACACGGCTGTGTATTACTGTGCGAGAGA
>IGHV3-52*01 (P)
GAGGTGCAGCTGGTGGAGTCTGGGGGAGGCTTGGTCCAGCCTGGAGGGTC
CCTGAAACTCTCTCTGGCAAGCTGGTGTGGTACCTTGATCTCTCTTCTCAT
GGAGCTGGGTCCGCCAGGCTCCAGGGAAGGGGCTGGAGTGGGTTGGCCGT
ACTAGAACAAACTCAACTACCACACAGAAATACCGCAGTCTGTGAA
AGGCAGATTACCACATCTAAAGAGATGATTTCAAAAAGAATCTAGTCTGCA
AAATGAAACAGCCTGAAAACCGAGGACACGGCGGTGTATTACTGTGCTAGA
>IGHV3-73*01
GAGGTGCAGCTGGTGGAGTCTGGGGGAGGCTTGGTCCAGCCTGGAGGGTC
CCTGAAACTCTCTCTGGCAAGCTGGTGTGGTACCTTGATCTCTCTTCTCAT
GGAGCTGGGTCCGCCAGGCTCCAGGGAAGGGGCTGGAGTGGGTTGGCCGT
ACTAGAACAAACTCAACTACCACACAGAAATACCGCAGTCTGTGAA
AGGCAGATTACCACATCTAAAGAGATGATTTCAAAAAGAATCTAGTCTGCA
AAATGAAACAGCCTGAAAACCGAGGACACGGCGGTGTATTACTGTGCTAGA
>IGHV3-73*02
GAGGTGCAGCTGGTGGAGTCCGGGGGAGGCTTGGTCCAGCCTGGAGGGTC
CCTGAAACTCTCTCTGGCAAGCTGGTGTGGTACCTTGATCTCTCTTCTCAT
GGAGCTGGGTCCGCCAGGCTCCAGGGAAGGGGCTGGAGTGGGTTGGCCGT
ACTAGAACAAACTCAACTACCACACAGAAATACCGCAGTCTGTGAA
AGGCAGATTACCACATCTAAAGAGATGATTTCAAAAAGAATCTAGTCTGCA
AAATGAAACAGCCTGAAAACCGAGGACACGGCGGTGTATTACTGTGCTAGA
>IGHV3-73*03
GAGGTGCAGCTGGTGGAGTCCGGGGGAGGCTTGGTCCAGCCTGGAGGGTC
CCTGAAACTCTCTCTGGCAAGCTGGTGTGGTACCTTGATCTCTCTTCTCAT
GGAGCTGGGTCCGCCAGGCTCCAGGGAAGGGGCTGGAGTGGGTTGGCCGT
ACTAGAACAAACTCAACTACCACACAGAAATACCGCAGTCTGTGAA
AGGCAGATTACCACATCTAAAGAGATGATTTCAAAAAGAATCTAGTCTGCA
AAATGAAACAGCCTGAAAACCGAGGACACGGCGGTGTATTACTGTGCTAGA
>IGHV3-73*p04 (partial)
AAGGGGCTGGTGGGTTCACGATTAATACTGATGGGAGTAGCACAAG
CTACGGGACTCCGGAAAGGGGGTACCCCACTCCAGAGACAGCCAGTACTGCAAAATGAA
CAGTCTAGAGCCAGAGACAGGCTGTTGATTACTGTGCAAGA
>IGHV3-74*01
GAGGTGCAGCTGGTGGAGTCTGGGGGAGGCTTGGTCCAGCCTGGAGGGTC
CCTGAAACTCTCTCTGGCAAGCTGGTGTGGTACCTTGATCTCTCTTCTCAT
GGAGCTGGGTCCGCCAGGCTCCAGGGAAGGGGCTGGAGTGGGTTGGCCGT
ACTAGAACAAACTCAACTACCACACAGAAATACCGCAGTCTGTGAA
AGGCAGATTACCACATCTAAAGAGATGATTTCAAAAAGAATCTAGTCTGCA
AAATGAAACAGCCTGAAAACCGAGGACACGGCGGTGTATTACTGTGCTAGA
>IGHV3-74*02
GAGGTGCAGCTGGTGGAGTCTGGGGGAGGCTTGGTCCAGCCTGGAGGGTC
CCTGAAACTCTCTCTGGCAAGCTGGTGTGGTACCTTGATCTCTCTTCTCAT
GGAGCTGGGTCCGCCAGGCTCCAGGGAAGGGGCTGGAGTGGGTTGGCCGT
ACTAGAACAAACTCAACTACCACACAGAAATACCGCAGTCTGTGAA
AGGCAGATTACCACATCTAAAGAGATGATTTCAAAAAGAATCTAGTCTGCA
AAATGAAACAGCCTGAAAACCGAGGACACGGCGGTGTATTACTGTGCTAGA
>IGHV3-74*03
GAGGTGCAGCTGGTGGAGTCTGGGGGAGGCTTGGTCCAGCCTGGAGGGTC
CCTGAAACTCTCTCTGGCAAGCTGGTGTGGTACCTTGATCTCTCTTCTCAT
GGAGCTGGGTCCGCCAGGCTCCAGGGAAGGGGCTGGAGTGGGTTGGCCGT
ACTAGAACAAACTCAACTACCACACAGAAATACCGCAGTCTGTGAA
AGGCAGATTACCACATCTAAAGAGATGATTTCAAAAAGAATCTAGTCTGCA
AAATGAAACAGCCTGAAAACCGAGGACACGGCGGTGTATTACTGTGCTAGA
>IGHV3-74*d01
AAGGGGCTGGTGGGTTCACGATTAATACTGATGGGAGTAGCACAAG
CTACGGGACTCCGGAAAGGGGGTACCCCACTCCAGAGACAGCCAGTACTGCAAAATGAA
CAGTCTAGAGCCAGAGACAGGCTGTTGATTACTGTGCAAGA
>IGHV3-74*04 (partial)
AAGGGGCTGGTGGGTTCACGATTAATACTGATGGGAGTAGCACAAG
CTACGGGACTCCGGAAAGGGGGTACCCCACTCCAGAGACAGCCAGTACTGCAAAATGAA
CAGTCTAGAGCCAGAGACAGGCTGTTGATTACTGTGCAAGA
>IGHV3-74*d01
>IGHV4-b*02
CAGGTGCAGCTGCAGGAGTCCGGGCCAGGGACTGGGAAGCCTTCGGAGA
CCCTGTCCTACCTCACTGACTGGTCTCTGGTTACTCCATCAACAGTGTTTACT
ACTGGGGCTGGTAGCCTCCCAACCCAGGAAGGGGGATGAGGTGATTGATGGAG
GAGTATCTACTATAGTGGAGACACTACTAACCCGCTCCTCAAGATGC
GACTCATTGAGCAGGCGCAGACACCGGCGCTGATTTACTGTCGAGA
>IGHV5-51*01
GAGGTGCAGCTGCTGAGCTGGAGAGCTGGAAAGCCTCCGGGGAGAT
ACTTTCTCCTGATACCTGATACGCTGCTGATACGGTTTGATGAGATGGGTAG
TGACACCTCTAGTGATACAGCTGCTGATACGGTTTGATGAGATGGGTAG
GAGTATCTACTATAGTGGAGACACTACTAACCCGCTCCTCAAGATGC
GACTCATTGAGCAGGCGCAGACACCGGCGCTGATTTACTGTCGAGA
>IGHV5-51*02
GAGGTGCAGCTGCTGAGCTGGAGAGCTGGAAAGCCTCCGGGGAGAT
ACTTTCTCCTGATACCTGATACGCTGCTGATACGGTTTGATGAGATGGGTAG
TGACACCTCTAGTGATACAGCTGCTGATACGGTTTGATGAGATGGGTAG
GAGTATCTACTATAGTGGAGACACTACTAACCCGCTCCTCAAGATGC
GACTCATTGAGCAGGCGCAGACACCGGCGCTGATTTACTGTCGAGA
>IGHV5-51*03
GAGGTGCAGCTGCTGAGCTGGAGAGCTGGAAAGCCTCCGGGGAGAT
ACTTTCTCCTGATACCTGATACGCTGCTGATACGGTTTGATGAGATGGGTAG
TGACACCTCTAGTGATACAGCTGCTGATACGGTTTGATGAGATGGGTAG
GAGTATCTACTATAGTGGAGACACTACTAACCCGCTCCTCAAGATGC
GACTCATTGAGCAGGCGCAGACACCGGCGCTGATTTACTGTCGAGA
>IGHV5-51*04
GAGGTGCAGCTGCTGAGCTGGAGAGCTGGAAAGCCTCCGGGGAGAT
ACTTTCTCCTGATACCTGATACGCTGCTGATACGGTTTGATGAGATGGGTAG
TGACACCTCTAGTGATACAGCTGCTGATACGGTTTGATGAGATGGGTAG
GAGTATCTACTATAGTGGAGACACTACTAACCCGCTCCTCAAGATGC
GACTCATTGAGCAGGCGCAGACACCGGCGCTGATTTACTGTCGAGA
>IGHV5-a*01
GAAGTGCAGCTGCTGAGCTGGAGAGCTGGAAAGCCTCCGGGGAGAT
ACTTTCTCCTGATACCTGATACGCTGCTGATACGGTTTGATGAGATGGGTAG
TGACACCTCTAGTGATACAGCTGCTGATACGGTTTGATGAGATGGGTAG
GAGTATCTACTATAGTGGAGACACTACTAACCCGCTCCTCAAGATGC
GACTCATTGAGCAGGCGCAGACACCGGCGCTGATTTACTGTCGAGA
>IGHV5-a*03
GAAGTGCAGCTGCTGAGCTGGAGAGCTGGAAAGCCTCCGGGGAGAT
ACTTTCTCCTGATACCTGATACGCTGCTGATACGGTTTGATGAGATGGGTAG
TGACACCTCTAGTGATACAGCTGCTGATACGGTTTGATGAGATGGGTAG
GAGTATCTACTATAGTGGAGACACTACTAACCCGCTCCTCAAGATGC
GACTCATTGAGCAGGCGCAGACACCGGCGCTGATTTACTGTCGAGA
>IGHV5-a*04
GAAGTGCAGCTGGTGCAGTCTGGAGCAGAGGTGAAAAAGCCCGGGAGTCTCTGAGGATCTCCTGTAAGGGTTCTGGATACAGCTTTACCAGCTACTGGA
TCAGCTGGTGCCAGCATGCCCCGGGAAAAAGCCTGGAGTGAGTAGGAGGAGATTGATCCTAGTGACTCTTATACCAACTACAGCCCTCGCTTCCAAGGCCA
GCCAGCTCAGCTCGAAGATCTCCTACCTGCAGTGGAGCACAGCTGAAGGCCTCGGACACCGCCACGCTATTACTGTGATGAGAGA
>IGHV6-1*01
CAGGTACAGCTGCAGCAGTCAGGTCCAGGACTGGTGAAGCCCTCGCAGAgg
CTCTCACTCCACTGCTGACACCTGGGAGAGCCTGGGACCTCAAGGGATGG
CAGTCCACATCCACAGGACAGCTGGGACACCCACACACAGCATCAGAGG
ATTCGACATATCGGACAGATGCACGCTATCGTTAGCTGATATAGCAGGAGA
>IGHV7-4-1*01
CAGGTGCAGCTGGTGCAATCTGGGTCTGAGTTGAAGCCTGGGGCCTCAGTGAGGTTTCCTGCAAGGCTTCTGGATACACCTTCACTAGCTATGCTAT
CAGAATTGGGTGCGACAGGCCCCTGGACAAGGGCTTGAGTGGATGGGATGG
AAATGAGCTTTGCTCGAATCTGCTGCGATATCTGCTATTGCTGGGAGA
>IGHV7-4-1*02
CAGGTGCAGCTGGTGCAATCTGGGTCTGAGTTGAAGCCTGGGGCCTCAGTGAGGTTTCCTGCAAGGCTTCTGGATACACCTTCACTAGCTATGCTAT
CAGAATTGGGTGCGACAGGCCCCTGGACAAGGGCTTGAGTGGATGGGATGG
AAATGAGCTTTGCTCGAATCTGCTGCGATATTGCTGGGAGA
>IGHV7-81*01
CAGGTGCAGCTGGTGCAATCTGGGTCTGAGTTGAAGCAGGAGCTGGGGCCTCAGTGAGGTTTCCTGCAAGGCTTCTGGATACACCTTCACTAGCTATGCTAT
CAGAATTGGGTGCGACAGGCCCCTGGACAAGGGCTTGAGTGGGATGGGATGG
AAATGAGCTTTGCTCGAATCTGCTGCGATATTGCTGGGAGA
>IGHV7-81*02
CAGGTGCAGCTGGTGCAATCTGGGTCTGAGTTGAAGCAGGAGCTGGGGCCTCAGTGAGGTTTCCTGCAAGGCTTCTGGATACACCTTCACTAGCTATGCTAT
CAGAATTGGGTGCGACAGGCCCCTGGACAAGGGCTTGAGTGGGATGGGATGG
AAATGAGCTTTGCTCGAATCTGCTGCGATATTGCTGGGAGA
Supplementary Table 5: The iHMune-align IGHJ repertoire

>IGHJ1*01
GCTGAATCTTCCAGCAGTGGGGGACGGCACCTTGTCACCGTCTCCTC
AG

>IGHJ2*01
CTACTGTTACTTACTCTTGCTGGGGACGGCACCTTGTCACCGTCTCCTC
AG

>IGHJ3*01
ATGCTTTGGATATCTGGGGCCAGGACAATGTGTCACCGTCTCCTCAG

>IGHJ3*02
ATGCTTTGGATATCTGGGGCCAGGACAATGTGTCACCGTCTCCTCAG

>IGHJ4*01
ACTACTTTGACTACTGGGGCCAAGGGACCTTGTCACCGTCTCCTCAG

>IGHJ4*02
ACTACTTTGACTACTGGGGCCAAGGGACCTTGTCACCGTCTCCTCAG

>IGHJ5*01
ACAACTGGTTCGTATGACTCCTGGGGCCAGGAAAATGTGTCACCGTCTCCTCAG

>IGHJ5*02
ACAACTGGTTCGTATGACTCCTGGGGCCAGGAAAATGTGTCACCGTCTCCTCAG

>IGHJ6*01
ATTACTACTACTACCTAGGTATGGACGTCTGGGGCCAGGACAACCGGTC
ACCCTCTCCTCA

>IGHJ6*02
ATTACTACTACTACCTAGGTATGGACGTCTGGGGCCAGGACAACCGGTC
ACCCTCTCCTCA

>IGHJ6*03
ATTACTACTACTACCTAGGTATGGACGTCTGGGGCCAGGACAACCGGTC
ACCCTCTCCTCA

>IGHJ6*04
ATTACTACTACTACCTAGGTATGGACGTCTGGGGCCAGGACAACCGGTC
ACCCTCTCCTCA

>IGHJ6*p05
ATTACTACTACTACTACCTAGGTATGGACGTCTGGGGCCAGGACAACCGGTC
ACCCTCTCCTCA
Supplementary Table 6: The iHMune-align IGHD repertoire. IGHD1-14*01 and IGHD6-25*01 are not included in the repertoire, as there is no evidence that they are capable of rearrangement and expression.

| IGHD1-1*01   | GGTACAACTGGAACGAC |
| IGHD1-7*01   | GGTATAACTGGAACGAC |
| IGHD1-20*01  | GGTATAACTGGAACGAC |
| IGHD1-26*01  | GGTATAGTGAGGAGCTACTAC |
| IGHD2-2*01   | AGGATATTGTAGTAGTACCAGCTAGCTATGCC |
| IGHD2-2*02   | AGGATATTGTAGTAGTACCAGCTAGCTATACC |
| IGHD2-2*03   | TGGATATTGTAGTAGTACCAGCTAGCTATGCC |
| IGHD2-8*01   | AGGATATTGTACTAATGGTGATGCTATACC |
| IGHD2-8*02   | AGGATATTGTACTGGTGATGCTATACC |
| IGHD2-15*01  | AGGATATTGTAGTGTTGGTGATGCTACTCC |
| IGHD2-21*01  | AGCATATTGTGGGTGATTGCTATCC |
| IGHD2-21*02  | AGCATATTGTGGGTGATTGCTATCC |
| IGHD3-3*01   | GTATTACGATTTTTGGAGTGGTTATTATCC |
| IGHD3-9*01   | GTATTACGATTTTTGACTGGTTATTATAAC |
| IGHD3-10*01  | GTATTACGATTTTTGACTGGTTATTATAAC |
| IGHD3-10*p03 | GTATTACGATTTTTGACTGGTTATTATAAC |
| IGHD3-16*02  | GTATTATGATTACCATTTGGGGGAGTGGTTATTATAAC |
| IGVD3-16*p03 | GTATTATGATTACCATTTGGGGGAGTGGTTATTATAAC |
| IGHD4-4/11*01 | GTATTACGTCAGTATTGAGTGTATTATAAC |
| IGHD4-17*01  | TGACTACAGTACTAC |
| IGHD4-23*01  | TGACTACGGTGACTAC |
| IGHD5-5/18*01 | GTGGATACAGCTATGGTTAC |
>IGHD5-12*01
GTGGATATAGGGCTACGATTAC
>IGHD5-24*01
GTAGAGATGGCTACAATTAC
>IGHD6-6*01
GAGTATAGCAGCTCGTCC
>IGHD6-13*01
GGGTATAGGGCTGCTGGTAC
>IGHD6-19*01
GGGTATAGGCTGCTGGTAC
>IGHD7-27*01
CTAAGCTGGGGA