Immunogenetics (2011) 63:259–265
DOI 10.1007/s00251-010-0510-8

ORIGINAL PAPER

Genomic screening by 454 pyrosequencing identifies a new human IGHV gene and sixteen other new IGHV allelic variants

Yan Wang · Katherine J. Jackson · Bruno Gäeta · William Pomat · Peter Siba · William A. Sewell · Andrew M. Collins

Received: 4 November 2010 / Accepted: 31 December 2010 / Published online: 20 January 2011

Abstract Complete and accurate knowledge of the genes and allelic variants of the human immunoglobulin gene loci is critical for studies of B cell repertoire development and somatic point mutation, but evidence from studies of VDJ rearrangements suggests that our knowledge of the available immunoglobulin gene repertoire is far from complete. The reported repertoire has changed little over the last 15 years. This is, in part, a consequence of the inefficiencies involved in searching for new members of large, multigenic gene families by cloning and sequencing. The advent of high-throughput sequencing provides a new avenue by which the germline repertoire can be explored. In this report, we describe pyrosequencing studies of the heavy chain IGHV1, IGHV3 and IGHV4 gene subgroups in ten Papua New Guineans. Thousands of 454 reads aligned with complete identity to 51 previously reported functional IGHV genes and allelic variants. A new gene, IGHV3-NL1*01, was identified, which differs from the nearest previously reported gene by 15 nucleotides. Sixteen new IGHV alleles were also identified, 15 of which varied from previously reported functional IGHV genes by between one and four nucleotides, while one sequence appears to be a functional variant of the pseudogene IGHV3-25. BLAST searches suggest that at least six of these new genes are carried within the relatively well-studied populations of North America, Europe or Asia. This study substantially expands the known immunoglobulin gene repertoire and demonstrates that genetic variation of immunoglobulin genes can now be efficiently explored in different human populations using high-throughput pyrosequencing.

Keywords Immunoglobulin heavy chain · IGHV · Allelic variants · Mutation analysis · Polymorphism

Electronic supplementary material The online version of this article (doi:10.1007/s00251-010-0510-8) contains supplementary material, which is available to authorized users.

Y. Wang · K. J. Jackson · A. M. Collins (✉)
School of Biotechnology and Biomolecular Sciences,
University of New South Wales,
Sydney NSW 2052, Australia
e-mail: a.collins@unsw.edu.au

K. J. Jackson · B. Gäeta
School of Computer Science and Engineering,
University of New South Wales,
Sydney NSW 2052, Australia

W. Pomat · P. Siba
Papua New Guinea Institute of Medical Research,
Goroka, Eastern Highlands Province, Papua New Guinea

W. A. Sewell
The Garvan Institute of Medical Research and St Vincent's Clinical School, University of NSW,
Sydney, Australia

Introduction

The diversity of lymphocyte receptor molecules is made possible by multiple sets of highly similar germline genes that recombine to form functional heavy-chain and light-chain genes (Sakano et al. 1980; Tonegawa 1983). Over a period of about 20 years, intense efforts to describe the many germline genes that were thought to contribute to this diversity led to the reporting of hundreds of human genes and pseudogenes. However, in the late 1990s, the sequencing of the complete immunoglobulin heavy-chain IGHV gene locus (Cook et al. 1994; Matsuda et al. 1998) and IGHD gene locus (Corbett et al. 1997) brought the search for new human germline antibody genes to an almost complete halt.

For some years, we have been reevaluating the reported immunoglobulin gene repertoire. We have identified many reported immunoglobulin allelic variants that were reported in error as a consequence of sequencing errors (Wang et al. 2008) and have inferred the existence of a number of previously unreported polymorphisms, using a bioinformatic approach (Lee et al. 2006; Wang et al. 2008). Confirmation of the existence of these putative alleles by cloning and sequencing, however, is problematic. The use of degenerate primers that anneal to the 5′ leader sequences and to 3′ recombination signal sequences inevitably amplifies previously reported genes, pseudogenes and immunoglobulin orphon genes from chromosomes 15 and 16 (Tomlinson et al. 1994.). It is therefore inevitable that large numbers of genes must be cloned and sequenced if the existence of a putative polymorphism is to be confirmed. Using this approach, we have confirmed the existence of three new IGHV genes (Wang et al. 2008) and two new IGKV genes (Collins et al. 2008); however, the inefficiency of the approach has deterred any further investigations.

The advent of massively parallel pyrosequencing provides a new avenue for the exploration of multigenic loci, and the technology has already proven its worth in investigations of MHC polymorphisms (Galan et al. 2010). In order to explore the potential of this approach for the identification of new genes and alleles, in this study, we explored immunoglobulin IGHV gene diversity in a human population that is genetically far-removed from the populations that have been the principal source of immunoglobulin sequences reported to date. Melanesians have not previously been investigated for immunoglobulin gene polymorphisms, though genetic studies have shown that they carry many other unique polymorphisms (Velickovic et al. 2009). Different populations within Papua New Guinean also show distinct genetic features (Redd and Stoneking 1999), for the island is believed to have been populated in two main waves. The highlanders are believed to be descended from the first wave of migration that settled the island around 32,000 BP, while the genetics of coastal communities is strongly influenced by a second wave of migration in 3,500 BP (Vilar et al. 2008). Individuals from both coastal and highland communities were therefore included in this study to maximise the likelihood that polymorphisms would be identified.

We describe the results of 454 pyrosequencing of immunoglobulin genes from ten Papua New Guineans. We show that around a quarter of all IGHV genes in these individuals have never been described before. We describe 15 new allelic variants of functional IGHV genes, which differ from previously reported alleles by from one to four nucleotides, a variant of a pseudogene that appears to be functional, and a new IGHV gene (IGHV3-NL1*01), which differs from the most similar, previously reported germline sequence by 15 nucleotides. Basic local alignment search tool (BLAST) searches show that 6 of these 19 new IGHV genes have complete identity to IGHV genes within VDJ gene rearrangements that have previously been deposited in public sequence databases. The identification of so many new alleles in such a small group of individuals highlights the fact that our understanding of the available germline gene repertoire within the human population is far from complete, while the presence of many of these sequences in public sequence databases suggests that these variants are likely to have a worldwide distribution.

Materials and methods

DNA isolation and amplification

After informed consent, and with the approval of both the UNSW Human Research Ethics Committee and the Papua New Guinea Medical Advisory Council, buccal smears were collected with cotton swabs from ten Papua New Guinean adults. DNA was extracted from the buccal smears by QIAamp® DNA Mini Kit (QIAGEN) and was then divided into four samples (PNG1–PNG4). PNG1 and PNG2 each included DNA from just one highland individual. PNG3 was a mixture of DNA samples from four highlanders, while PNG4 was a mixture of DNA samples from four people from coastal communities. Equal amounts of DNA from each individual were included in PNG3 and PNG4. PCR amplification was performed using the FastStart High Fidelity PCR System (Roche) with additional 10 mM PCR grade Nucleotide Mix (Roche). Master mixes were made with 200 ng DNA, 0.4 μM of each primer (IDT), 200 μM of each dNTP (Roche), 1.25 units FastStart High Fidelity Enzyme Blend (Roche) and a buffer supplied by the manufacturer. GS FLX Titanium Primer A and Primer B were added to the 5′ ends of the forward and reverse IGHV subgroup-specific primers, respectively. The specific primers for the IGHV1, IGVH3 and IGVH4 gene subgroups (Table 1) were used in separate reactions. The PCR conditions were 95°C for 3 min, followed by 30 cycles of 95°C for 30 s, 60°C for 30 s, 72°C for 30 s, and then a final extension at 72°C for 5 min. The PCR products were then cleaned by gel extraction using QIAquick Gel Extraction Kits (QIAgen). The cleaned samples were then sent to the Ramaciotti Centre for Gene Function Analysis, UNSW for sequencing using an FLX Titanium Genome Sequencer (454 Life Sciences).

Germline gene identification

The sequence datasets resulting from the four samples were separately analysed to identify IGHV genes and to quantify the copy number of identical sequences. IGHV gene

Table 1 IGHV subgroup-specific primers

Primer name	Sequence (5′- 3′)	Description
IGHV1F	TCTCCACAGGTGYCCACTCC	IGHV1 forward primer
IGHV1R	AGGATGTGGKTTTTCACACTGTG	IGHV1 reverse primer
IGHV3F	BTGTTTGCAGGTGTCCAGTGT	IGHV3 forward primer
IGHV3R	CAHTGACYTCCCCTCACTGTG	IGHV3 reverse primer
IGHV4F	GGCTCACTGTGKGTYTCTCTGTT	IGHV4 forward primer
IGHV4R	CTCACACTCACCTCCCCTCA	IGHV4 reverse primer

identifications were made by way of BLAST searches against a combined repertoire of reported IGHV germline genes. The database consisted of the full germline repertoires defined by IMGT (Pallares et al. 1999) and NCBI's IgBLAST, with the addition of the unconfirmed polymorphisms reported from our previous studies (Boyd et al. 2010; Wang et al. 2008). BLAST searches were performed with a match reward score of 2 and mismatch penalty of 5. Insertions and deletions in alignments were permitted, with a gap open penalty of 16 and gap extension penalty of 4. BLAST results were limited to a single top scoring alignment, and only alignment results that achieved a minimum alignment length of 285 nucleotides were retained for further analysis.

Copy numbers of identical sequences that aligned to each IGHV gene within each sample were quantified. Sequences were considered to be identical if a shorter alignment to an IGHV gene, based on the BLAST result, shared 100% identify with a longer alignment to the same IGHV gene. A single sequence representing the longest unique alignment to an IGHV gene was selected for retention in the final datasets for PNG1–PNG4, and the copy number of both sense and anti-sense strand alignments was noted for each of the datasets. A minimum of six identical sequences, with a minimum of three sense strand reads and three anti-sense strand reads that aligned with high similarity to an IGHV gene, was used to identify correctly sequenced genomic DNA.

Results

In order to investigate germline IGHV genes in the Papua New Guinean population, DNA was prepared from ten individuals, and a total of 57,951 sequences were generated. Details of the datasets generated from each of the samples are summarised in Table 2. The sequences were aligned against reported immunoglobulin genes, open reading frames, pseudogenes and orphons and analysed to identify identical sequences. In order to exclude sequences that included 454 sequencing errors, which are overwhelmingly insertions and deletions near homopolymer tracts (Droege and Hill 2008), the dataset was filtered to exclude any sequence that was not present within a sample as at least three forward and three reverse complement reads. This filter was found to be sufficient to almost completely eliminate gaps from the final dataset. The filtering resulted in the exclusion of 24,058 unique sequences, and the retention of just 332 unique sequences. The average copy numbers of the excluded and retained sequences were 1.2 and 70.7 copies, respectively. The discarded sequences had, on average, 1.8 mismatches, 2.5 nucleotide gaps and were 280 nucleotides long when the sequences were aligned against their most similar, previously reported immunoglobulin germline gene, pseudogene or orphon. The 332 retained sequences had, on average, 0.7 nucleotide mismatches, 0.06 gaps and were 295 nucleotides long. When the 332 retained sequences of the four samples were combined, just 115 unique sequences remained with an average of 203 reads per sequence. The high copy number and the inclusion of bidirectional reads in this study allowed germline sequences to be identified with great confidence.

Seventeen sequences that were observed as at least three identical forward reads and at least three identical reverse complement reads, differed from previously reported functional alleles by from one to four nucleotides. The amino acid sequences encoded by 9 of the 17 nucleotide sequences differed from their most similar reported allele by at least one amino acid. These 17 sequences were accepted as new IGHV polymorphisms, and details of their representation in the four sequence datasets are presented as Table 3. Six representative sequences (3 forward and 3 reverse complement) derived from these genes were submitted to NCBI, and were assigned the gene names shown in Table 3.

One sequence was seen as 24 replicates in the sample derived from four Papua New Guinean highlanders (PNG3) and as 9 replicates in the sample derived from four coastal donors (PNG4). This sequence aligned most closely to IGHV3-30*18, but included 15 mismatches. The differences between the PNG sequence and the existing germline sequence were mainly concentrated in the CDR2 region, where the PNG sequence is identical to IGHV3-23*03 (Fig. 1). Seven of the 15 nucleotide differences lead to amino acid differences in the translated gene product. Given the high identity to portions of both IGHV3-23*03 and IGHV3-30*18, chimerism was explored as a potential source of the sequence. By separately aligning the 5′ and 3′

Table 2 Summary of 454 pyrosequencing reads and their alignments to germline genes, pseudogenes and orphons

	PNG1	PNG2	PNG3	PNG4
Reads≥285 nts	8,436	8,743	12,185	18,304
Germline sequences identified	73	64	95	100
Functional IGHV (previously reported)	34	30	42	48
'Functional IGHV' (previously unreported)[a]	10	9	13	12
Pseudogenes (previously reported)	7	8	15	13
Orphons (previously reported)	16	14	16	16
Other immunoglobulin-like sequences[b]	6	3	9	11
Mean reads per sequence identified	56.1	63.2	65.1	91.3
Mean reverse complement reads	36.4	39.7	41.6	60.8
Total germline full-length reads	4,096	4,042	6,187	9,128

Data is presented from DNA prepared from two individuals (PNG1 and PNG2) as well as from DNA pooled from two groups of four individuals from highland populations (PNG3) and coastal populations (PNG4).

[a] Germline sequences are reported as 'functional' if they encode an open reading frame and are variants of a previously reported functional gene.

[b] Other immunoglobulin-like sequences were most similar to previously reported pseudogenes, ORFs and orphons.

segments of the sequence against the full germline gene database, chimerism could be excluded. The sequence represents a new gene of the IGHV3 subgroup, and has been assigned the name IGHV3-NL1*01.

Thirteen sequences were seen in the PNG3 sample that differed from the IGHV3-25*03 allele at four nucleotides. The IGHV3-25 gene is considered a pseudogene, as all three reported alleles encode a stop codon at codon 103.

Table 3 Alignments of 454 pyrosequencing reads to previously unreported, germline genes

Gene	Most similar gene	Diffs[a]	Alignments				Total RC[b]	Total reads
			PNG1	PNG2	PNG3	PNG4		
IGHV1-8*02[d]	IGHV1-8*01	1	56	56	104	60	182	276
IGHV1-18*03[c]	IGHV1-18*01	1	69	99	169	79	267	416
IGHV3-7*03[d]	IGHV3-7*01	1				52	39	52
IGHV3-9*02	IGHV3-9*01	1	16	10	23	10	26	59
IGHV3-11*05[d]	IGHV3-11*01	1	64	75	100	28	131	267
IGHV3-13*04[d]	IGHV3-13*01	1				18	12	18
IGHV3-21*03	IGHV3-21*01	1				32	20	32
IGHV3-25*04	IGHV3-25*03	4			13		8	13
IGHV3-33*06	IGHV3-33*03	1	16	14	7		18	37
IGHV3-48*04[d]	IGHV3-48*01	1	47	47	26		71	120
IGHV3-53*04[d]	IGHV3-53*01	3	30	36	44	70	116	180
IGHV3-NL*01	IGHV3-30*18	15			24	9	18	33
IGHV4-28*06	IGHV4-28*01	3			32		25	32
IGHV4-30-2*05	IGHV4-30-2*01	4				7	3	7
IGHV4-59*11[c,d]	IGHV4-59*01	3	56	27	43	69	133	195
IGHV7-4-1*04	IGHV7-4-1*02	1	66		58	27	99	151
IGHV7-4-1*05	IGHV7-4-1*02	2	69	154	81		196	304
Total new alleles			10	9	13	12	1,364	2,192

Data is presented from DNA prepared from two individuals (PNG1 and PNG2) as well as from DNA pooled from two groups of four individuals from coastal populations (PNG3) and highland populations (PNG4)

[a] The number of nucleotide differences between the new polymorphisms and the most similar, previously reported gene

[b] Reverse complement reads

[c] These sequences are not the same as the previously reported, putative polymorphisms, IGHV1-18*p03 and IGHV3-11*p05 (11) and IGHV4-59*p11 (12)

[d] Rearranged sequences were identified by NCBI's IgBlast that appear to include these sequences

Fig. 1 Alignment of IGHV3-NL1*01 against the most similar IGHV3-30 allele and the most similar IGHV3-23 allele. Identity between the sequences is indicated with *dashes*, and the framework (FR) and complementarity determining regions (CDR) used in this study are shown

The variant sequences reported here lead to the encoding of tyrosine at this codon, and the gene may therefore be functional.

BLAST searches of VDJ rearrangements in the Genbank sequence database identified numerous IGHV genes within these rearrangements that were more closely matched to six of the newly identified polymorphisms than to other previously reported sequences. One perfect match was seen to IGHV1-8*02, two perfect matches were seen to IGHV3-48*04, three such matches were seen to IGHV3-7*03, IGHV3-13*04 and IGHV3-53*04, while numerous reported VDJ rearrangements appear to include the unmutated IGHV4-59*11. Reported sequences come from many parts of the developed world, including the USA (Kolar et al. 2004) and Germany (Brauninger et al. 2001). Unpublished sequences also came from Japan (e.g., AB067272) and Italy (e.g., AJ627231).

Sequences that most closely matched previously reported functional IGHV genes are presented as Supplementary Table 1 and are also summarised in Table 2. In each case, at least three identical forward reads and three identical reverse complement reads were seen. All functional genes of the IGHV1 subgroup, other than IGHV1-45 and IGHV1-f, were represented in one or more of the datasets. In addition, at least one full-length alignment was seen with complete identity to IGHV1-24*01, IGHV1-46*01, IGHV1-69*02, IGHV1-69*04, IGHV1-69*09 and IGHV1-69*12, but insufficient forward and reverse complement sequences were seen to unequivocally accept their presence in the genomes of the individuals studied. All functional genes of the IGHV3 and IGHV4 subgroups were seen, except for IGHV3-64, IGHV4-30-1 and IGHV4-b, and two functional alleles of the IGHV7 subgroup were also seen.

Previously reported IGHV pseudogenes and orphons that were identified are presented as Supplementary Table 2. Immunoglobulin-like sequences that appear to be previously unreported IGHV pseudogenes and orphons are presented as Supplementary Table 3. Perfect matches were seen to IGHV1-c*01, an apparently non-functional open reading frame that has previously been reported with a 34 nucleotide 5′ truncation. The full-length open reading frame was seen here, and is also shown in Supplementary Table 3.

All new genes and alleles identified in this study, as well as immunoglobulin-like sequences that appear to be either pseudogenes or orphons, are presented in FASTA format as Supplementary Table 4 and have been submitted to the GenBank database (HM855272 - HM855948).

Discussion

A detailed understanding of the variability of immunoglobulin genes is critical if we are to understand the generation of diversity in the human antibody gene repertoire. We recently identified many new IGHV polymorphisms by the bioinformatic analysis of VDJ rearrangements in public sequence databases (Wang et al. 2008), as well as by analysis of VDJ rearrangements generated using high-throughput sequencing technology (Boyd et al. 2010). However, the 23 resulting putative alleles remain unrecognised by IMGT, the most widely used database of germline genes. Genomic sequences are required if the reality of these putative alleles are to be widely accepted. This study therefore investigated the potential of 454 pyrosequencing for genomic screening of the immunoglobulin locus.

Although it was expected that some new polymorphisms would be identified in this study, the extent of variation between the immunoglobulin genes of this population and the IMGT repertoire was surprising. One quarter of all sequences seen in the two individuals who were screened are previously unreported IGHV alleles. Importantly, these alleles do not appear to be confined to Papua New Guineans. BLAST searches of public sequence databases

suggest that 6 of the 17 new genes and allelic variants are commonly carried in distant communities, for rearranged VDJ genes in public sequence databases were found to align perfectly to these genes. It is likely that at least some of the other 11 new genes are also present, if less commonly, in populations of the developed world, and that much diversity remains to be uncovered in all human populations.

Particularly striking was the identification of a new member of the IGHV3 gene subgroup—the first new IGHV gene described since 1994 (Cook et al. 1994). The IGHV3-NL1*01 gene was identified in both the highlands and coastal group samples, which suggests that this gene may be relatively common within the PNG population. We can confirm its functionality for a search of IgE and IgG VDJ rearrangements that were cloned and sequenced from Papua New Guinean samples as part of a separate study (Wang and Collins, unpublished manuscript), identified seven rearrangements that utilise this sequence (Accession numbers HM773984, HM774108, HM774124, HM774201, HM774302, HM774729, and HM774738).

The present study focused on the amplification of functional genes of the three largest IGHV subgroups: IGHV1, IGHV3 and IGHV4. Most, but not all genes of these families, were successfully amplified and sequenced. The reason why some sequences failed to amplify is likely to be the design of the primers, rather than because of the absence of the genes in the PNG population. The primers used in this study were designed to anneal to the 5′ leader sequences and 3′ recombination signal sequences (RSS); however, these sequences are not known for all previously reported genes and allelic variants. There could also be leader sequence and RSS polymorphisms within the Papua New Guinean population. It may therefore be that the primers were unsuited to the amplification of some genes. In fact, the PCR-based approach used here may never ensure that complete genotypes are always generated. Alternative enrichment methods for immunoglobulin gene-like sequences, as a prelude to pyrosequencing, may be necessary if all variants are to be found (Teer et al. 2010). Such a systematic exploration of immunoglobulin gene diversity will be required if we are to understand the evolution of the immunoglobulin gene loci (Romo-Gonzalez and Vargas-Madrazo 2005), and to properly investigate immunoglobulin genes usage and gene mutation in health and disease (Hamblin et al. 1999). The large number of new IGHV genes identified in this study suggests that efforts to identify further variants must now be undertaken in different populations as a matter of urgency

References

Boyd SD, Gaëta BA, Jackson KJ, Fire AZ, Marshall EL, Merker JD, Maniar JM, Zhang LN, Sahaf B, Jones CD, Simen BB, Hanczaruk B, Nguyen KD, Nadeau KC, Egholm M, Miklos DB, Zehnder JL, Collins AM (2010) Individual variation in the germline Ig gene repertoire inferred from variable region gene rearrangements. J Immunol 184:6986–6992

Brauninger A, Yang W, Wacker HH, Rajewsky K, Kuppers R, Hansmann ML (2001) B-cell development in progressively transformed germinal centers: similarities and differences compared with classical germinal centers and lymphocyte-predominant Hodgkin disease. Blood 97:714–719

Collins AM, Wang Y, Singh V, Yu P, Jackson KJ, Sewell WA (2008) The reported germline repertoire of human immunoglobulin kappa chain genes is relatively complete and accurate. Immunogenetics 60:669–676

Cook GP, Tomlinson IM, Walter G, Riethman H, Carter NP, Buluwela L, Winter G, Rabbitts TH (1994) A map of the human immunoglobulin VH locus completed by analysis of the telomeric region of chromosome 14q. Nat Genet 7:162–168

Corbett SJ, Tomlinson IM, Sonnhammer ELL, Buck D, Winter G (1997) Sequence of the human immunoglobulin diversity (D) segment locus: a systematic analysis provides no evidence for the use of DIR segments, inverted D segments, “minor” D segments or D-D recombination. J Mol Biol 270:587–597

Droege M, Hill B (2008) The Genome Sequencer FLX System–longer reads, more applications, straight forward bioinformatics and more complete data sets. J Biotech 136:3–10

Galan M, Guivier E, Caraux G, Charbonnel N, Cosson JF (2010) A 454 multiplex sequencing method for rapid and reliable genotyping of highly polymorphic genes in large-scale studies. BMC Genomics 11:296

Hamblin TJ, Davis Z, Gardiner A, Oscier DG, Stevenson FK (1999) Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. Blood 94:1848–1854

Kolar GR, Yokota T, Rossi MI, Nath SK, Capra JD (2004) Human fetal, cord blood, and adult lymphocyte progenitors have similar potential for generating B cells with a diverse immunoglobulin repertoire. Blood 104:2981–2987

Lee CEH, Gaëta B, Malming HR, Bain ME, Sewell WA, Collins AM (2006) Reconsidering the human immunoglobulin heavy-chain locus. 1. An evaluation of the expressed human IGHD gene repertoire. Immunogenetics 57:917–925

Matsuda F, Ishii K, Bourvagnet P, Kuma K, Hayashida H, Miyata T, Honjo T (1998) The complete nucleotide sequence of the human immunoglobulin heavy chain variable region locus. J Exp Med 188:2151–2162

Pallares N, Lefebvre S, Contet V, Matsuda F, Lefranc MP (1999) The human immunoglobulin heavy variable genes. Exp Clin Immunogenet 16:36–60

Redd AJ, Stoneking M (1999) Peopling the Sahul: mtDNA variation in aboriginal Australia and Papua New Guinean populations. Am J Hum Genet 65:808–828

Romo-Gonzalez T, Vargas-Madrazo E (2005) Structural analysis of substitution patterns in alleles of human immunoglobulin VH genes. Mol Immunol 42:1085–1097

Sakano H, Maki R, Kurosawa Y, Roeder W, Tonegawa S (1980) Two types of somatic recombination are necessary for the generation of complete immunoglobulin heavy-chain genes. Nature 286:676–683

Teer JK, Bonnycastle LL, Chines PS, Hansen NF, Aoyama N, Swift AJ, Abaan HO, Albert TJ, Program NCS, Margulies EH, Green ED, Collins FS, Mullikin JC, Biesecker LG (2010) Systematic comparison of three genomic enrichment methods for massively parallel DNA sequencing. Genome Res 20:1420–1431

Tomlinson IM, Cook GP, Carter NP, Elaswarapu R, Smith S, Walter G, Buluwela L, Rabbitts TH, Winter G (1994) Human immuno-

globulin VH and D segments on chromosomes 15q11.2 and 16p11.2. Hum Mol Genet 3:853–860

Tonegawa S (1983) Somatic generation of antibody diversity. Nature 302:575–581

Velickovic M, Velickovic Z, Panigoro R, Dunckley H (2009) Diversity of killer cell immunoglobulin-like receptor genes in Indonesian populations of Java, Kalimantan, Timor and Irian Jaya. Tissue Antigens 73:9–16

Vilar MG, Kaneko A, Hombhanje FW, Tsukahara T, Hwaihwanje I, Lum JK (2008) Reconstructing the origin of the Lapita Cultural Complex: mtDNA analyses of East Sepik Province, PNG. J Hum Genet 53:698–708

Wang Y, Jackson KJL, Sewell WA, Collins AM (2008) Many human immunoglobulin heavy-chain IGHV gene polymorphisms have been reported in error. Immunol Cell Biol 86:111–115