# Generation, Diversity Determination, and Application to Antibody Selection of a Human Naïve Fab Library

Sangkyu Kim[1,11], Insoo Park[2,6,11], Seung Gu Park[3,7], Seulki Cho[2,8], Jin Hong Kim[2,9], Nagesh S.Ipper[4,10], Sun Shim Choi[3], Eung Suk Lee[5], and Hyo Jeong Hong[1,4,*]

[1]Department of Systems Immunology, College of Biomedical Science, Kangwon National University, Chuncheon 24341, Korea, [2]Therapeutic Antibody Research Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon 34141, Korea, [3]Department of Medical Biotechnology, College of Biomedical Science, Kangwon National University, Chuncheon 24341, Korea, [4]Institute of Bioscience and Biotechnology, Kangwon National University, Chuncheon 24341, Korea, [5]Scripps Korea Antibody Institute, Chuncheon 24341, Korea, [6]Present address: Molecular Imaging and Therapy Branch, National Cancer Center, Goyang 10408, Korea, [7]Present address: Department of Biomedical Engineering, School of Life Sciences, and The Genomics Institute (TGI), Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, Korea, [8]Present address: Chong Kun Dang pharmaceutical Corp, Yongin 16995, Korea, [9]Present address: PCL Inc., Seoul 08510, Korea, [10]Present address: Boditechmed Inc., Chuncheon 24398, Korea, [11]These authors contributed equally to this work.
*Correspondence: hjhong@kangwon.ac.kr
http://dx.doi.org/10.14348/molcells.2017.0106
www.molcells.org

We constructed a large naïve human Fab library ($3 \times 10^{10}$ colonies) from the lymphocytes of 809 human donors, assessed available diversities of the heavy-chain variable (VH) and κ light-chain variable (VK) domain repertoires, and validated the library by selecting Fabs against 10 therapeutically relevant antigens by phage display. We obtained a database of unique 7,373 VH and 41,804 VK sequences by 454 pyrosequencing, and analyzed the repertoires. The distribution of VH and VK subfamilies and germline genes in our antibody repertoires slightly differed from those in earlier published natural antibody libraries. The frequency of somatic hypermutations (SHMs) in heavy-chain complementarity determining region (HCDR)1 and HCDR2 are higher compared with the natural IgM repertoire. Analysis of position-specific SHMs in CDRs indicates that asparagine, threonine, arginine, aspartate and phenylalanine are the most frequent non-germline residues on the antibody-antigen interface and are converted mostly from the germline residues, which are highly represented in germline SHM hotspots. The amino acid composition and length-dependent changes in amino acid frequencies of HCDR3 are similar to those in previous reports, except that frequencies of aspartate and phenylalanine are a little higher in our repertoire. Taken together, the results show that this antibody library shares common features of natural antibody repertoires and also has unique features. The antibody library will be useful in the generation of human antibodies against diverse antigens, and the information about the diversity of natural antibody repertoires will be valuable in the future design of synthetic human antibody libraries with high functional diversity.

Keywords: diversity, human monoclonal antibody, naïve antibody library, phage display, somatic hypermutation

## INTRODUCTION

The humoral immune response recognizes a wide range of

antigens and generates an immense number of distinct antibodies (Chaplin, 2003; Selsted and Ouellette, 2005). Antibody paratopes that mediate specific binding to different antigens are found at the CDRs of VH and light-chain variable regions (VL) of immunoglobulin (Ig). Each chain contains three CDR loops, and CDR1 and CDR2 are encoded in each V germline gene segment. There are 51 VH domain, 40 Vκ and 30 λ light-chain variable domain (V λ) germline segments in a typical human haplotype (Barbie and Lefranc, 1998; Cox et al., 1994; Matsuda and Honjo, 1996; Pallares et al., 1998; Tomlinson et al., 1992). These human germline VH genes are grouped into seven families and VL genes into six kappa and ten lambda families, based on sequence homologies (Kawasaki et al., 1997; Vargas-Madrazo et al., 1997; Williams et al., 1996). HCDR3 is formed by the junction of V-(D)-J joining and the light-chain complementarity determining region (LCDR)3 by V-J joining (Market and Papavasiliou, 2003; Schatz, 2004). Diversity in each chain is determined by combinatorial V-(D)-J (for the heavy) or V-J (for the light) rearrangements, P and N-addition, junctional flexibility, and SHM of variable domain nucleotides (Muramatsu et al., 1999; Tonegawa, 1983; Wu and Kabat, 1970). In addition, the combinatorial association of heavy and light chains increases diversity by many orders of magnitude (Perelson and Oster, 1979; Trepel, 1974).

SHM introduces changes in the VH and VL genes that give rise to affinity maturation of the antibody (Muramatsu et al., 2000). SHMs are mediated by the activation-induced deaminase (AID) enzyme (Muramatsu et al., 1999). This enzyme is also the key mediator of class switch recombination that replaces the Ig heavy chain constant region gene, resulting in a switch from IgM to IgG, IgE, or IgA (Muramatsu et al., 2000). Thus, the percentages of non-mutated V region sequences associated with IgM are significantly higher than those associated with the other isotypes (Jackson et al., 2014), and the frequency of SHM is highest in the CDRs (Tomlinson et al., 1996). The analysis of SHMs in the structures of antibodies revealed that the solvent-exposed surface positions and antigen-antibody interface are more likely to be targets of productive SHMs but that residues in the VL-VH interface or core which contribute to antibody stability and affinity are also often preferred targets for SHM (Burkovitz et al., 2014; Clark et al., 2006).

Monoclonal antibodies (mAbs) represent the most important class of biologics and have great potential for both diagnostic and therapeutic applications (Ecker et al., 2015; Geng et al., 2015; Sharma et al., 2016). Fully human mAbs are currently generated by hybridoma technology using human IgG transgenic animals and *in vitro* technologies such as phage, ribosome, and yeast display of combinatorial antibody libraries (Boder and Wittrup, 1997; Bruggemann et al., 2015; Hanes and Pluckthun, 1997; McCafferty et al., 1990; Ponsel et al., 2011). Phage display has become the most frequently used display technology, with the advantage of robustness due to high stability of the phage (Hoet et al., 2005). The *in vitro* methods employ a naïve antibody library generated from donor-derived B cells, a synthetic library from synthetically derived diversity, or a semi-synthetic library derived from a combination of these two approaches, which

differ in framework region (FR) composition and CDR diversification (Glanville et al., 2009; Hoogenboom, 2005; Kugler et al., 2015; Lloyd et al., 2009; Tiller et al., 2013; Zhai et al., 2011). The most commonly used antibody formats for phage display are single-chain Fv (scFv) and Fab. ScFv fragments are moderately stable on average and often have a high tendency to form multimers as well as aggregates, whereas Fab fragments have been found to possess comparably higher structural stability by an additional interface of the constant domains (Rothlisberger et al., 2005). For these reasons, Fab has replaced the scFv as a display format in many of recent phage libraries (Ponsel et al., 2011).

In this study, we constructed a large naïve human Fab library ($3 \times 10^{10}$ colonies) from the peripheral blood lymphocytes (PBLs), lymph node, spleen, and bone marrow of 809 human donors and assessed available diversities of the VH and κ VL (VK) repertoires in the entire library by 454 pyrosequencing. In addition, we validated the antibody library by selecting specific Fabs against 10 different therapeutically relevant antigens by phage display. The antibody library will be useful in the generation of human mAbs against diverse antigens, and the information about the diversity of natural antibody repertoires will be valuable in the future design of synthetic human antibody libraries with high functional diversity.

## MATERIALS AND METHODS

### Materials
Restriction enzymes were obtained from New England Biolabs (Hoet et al.). *E. coli* TG1 was from Lucigen (USA). VCSM13 helper phage was from Stratagene (USA). RNAzol reagent was from Tel-Test Inc. (USA). Polyacrylamide gel electrophoresis-purified oligonucleotides were from Thermo Fisher Scientific (USA). *Taq* DNA polymerase was from Takara (Japan). Bovine serum albumin and Tween 20 were from Sigma. Tetramethylbenzidine was from BD Biosciences (USA). Goat anti-mouse IgG-horseradish peroxidase and goat anti-human IgG-horseradish peroxidase were from Thermo Fisher Scientific. Human tissue samples, spleen from two males (C1234161), lymph-node from two males (C1234024), and bone-marrow from two females (C1234246) were obtained from BioChain Institute (USA). Human PBLs were obtained from hospitals in Korea in 2005.

### Construction of a human naïve Fab library
A modified version (pKRIBB-FabD) of the human Fab phage display vector pCOMB3H (Gram et al., 1992) was previously constructed by inserting new cloning sites for a VH or VL repertoire and epitope tags for the detection and purification of expressed Fabs (Kim and Hong, 2012). The Fab display vector contains a bicistronic operon under the control of *LacZ* promoter. The first open-reading frame encoded a polypeptide consisting of the *ompA* signal peptide, a VK, human κ light-chain constant region (Cκ), and a S1 tag (Oh et al., 2003). Two *Bst*XI (5'-CCATTGCAGTGG and 5'-CCAGTACGGTGG) sites were incorporated between the *ompA* and the human κ constant domain sequences for cloning a VK repertoire. The second open-reading frame

encoded a polypeptide consisting of a *pelB* signal peptide, a VH, the first constant domain of human γ1, a *myc* tag, the C-terminal domain of the p3 of M13, and a His-tag. The two *Sfi*I (5'-GGCCCAGCCGGCC and 5'-GGCCAGCACGGCC) sites were incorporated between the *pelB* and the CH1 sequences for cloning a VH repertoire.

Total RNAs were prepared from the PBLs of 33 Korean healthy human donors for a VK repertoire by using RNAzol reagent, and prepared from the PBLs of 770 and commercially available normal tissues (spleen from two donors, lymph node from two donors, and bone marrow from two donors) for a VH repertoire. First-strand cDNA was synthesized from total RNA by using oligo dT primer and reverse transcriptase (Superscript II, Thermo Fisher, USA) and subjected to polymerase chain reaction (PCR) to create a VH or VK repertoire using the PCR primers (www.vbase.mrc-cpe.cam.ac.uk) designed to amplify the VH or VK sequences derived from all VH or VK germline genes, respectively. For a VK repertoire, 13 forward primers (Vk1a, 1b, 1c, 1d, 2a, 2b, 3a, 3b, 3c, 4a, 5a, 6a, and 6b) and 4 backward primers (121, 122, 123, and 124) were used. For a VH repertoire, 14 forward primers (VH1a, 1b, 1c, 1d, 2a, 2b, 3a, 3b, 3c, 4a, 4b, 5a, 6a, and 7a) and 4 backward primers (VHR1, VHR2, VHR3, and VHR4) were used. The VH or VK repertoire was amplified by PCR using an equimolar mixture of the degenerate family primers and *Taq* polymerase at 94℃ for 30 s, 55℃ for 30 s, and 72℃ for 30 s for 30 cycles. The final extension step was carried out at 72℃ for 10 min. The resulting PCR products were subjected to extension PCR for 20 cycles using the VK extension primers containing the *Bst*XI site or the VH extension primers containing the *Sfi*I site under the same PCR conditions. The final PCR products were purified from agarose gels.

To construct a combinatorial Fab library, first the purified DNA of a VK repertoire was cloned into the *Bst*XI sites of the pKRIBB-FabD vector to create pKFabD-VL. Subsequently, the purified DNA of a VH repertoire was cloned into the *Sfi*I sites of the pKFabD-VL to construct pKFabD-VL-VH. *E. coli* TG1 competent cells were transformed with the pKFabD-VL-VH by electroporation using MicroPulser (Bio-Rad). The transformed bacteria were incubated with 2xYT containing 2% glucose and 5 mM $MgCl_2$ at 37℃ for 1 h. An aliquot of 100 µL was used for serial dilution to estimate the transformation efficiency, and the remaining cells were centrifuged, resuspended in 10 ml 2xYT broth, and plated on 245 mm × 245 mm × 25 mm SOBAG agar square plates. The plates were then incubated at 37℃ overnight. The following day, 10 ml of 2xYT containing 2% glucose and 5 mM $MgCl_2$ was poured over the plates and gently scraped into a 50 ml falcon tube. $OD_{600}$ was measured, and the samples were stored with 20% glycerol at -80℃. The quality of the library ($3 \times 10^{10}$ colonies) was evaluated by plasmid preparation and DNA sequencing of 200 randomly selected clones, and the amino acid sequences of both chains were analyzed using the IMGT human antibody database http://imgt.cines.fr/IMGT_vquest/vquest?livret=0&Option=humanIg.

### 454 sequencing of antibody library

The transformants containing the human Fab library that were stored at -80℃ were grown overnight and the plasmid DNA was prepared. Pyrosequencing of the human naive Fab library was carried out by DNA Link (South Korea). Briefly, the VH and VK genes were separately amplified by PCR from the Fab library DNA using unidirectional fusion primers complementary to the constant regions or the vector sequences containing the 454-Titanium adaptor sequences (Supplementary Table 2) to generate an Ig amplicon library. Thermal cycling consisted of initial denaturation at 95℃ for 10 min, followed by 30 cycles of denaturation at 98℃ for 30 s, annealing at 60℃ for 1 min, and extension at 72℃ for 1 min, with a final extension of 10 min at 72℃. Amplicons were visualized on a 1% agarose gel and purified using the QIAquick PCR Purification Kit (Qiagen). Amplicon DNA concentrations were determined using Nanodrop (ND1000). After quantification, the DNA was subjected to 2% agarose gel electrophoresis, and target size was resolved using a Qiagen gel extraction kit. The DNA was pooled, and pyrosequencing was carried out following manufacturer instructions on a 454 Life Sciences Genome Sequencer FLX instrument (Roche, USA) using titanium chemistry. The sequences were assigned to samples according to sample specific barcodes.

### Sequence analysis of antibody library

With the aim of detecting and discarding most of the reads that exhibited sequencing errors or non-functional genes, we developed stepwise data-processing procedures using lab-built Perl scripts (Supplementary Fig. 1). We obtained two data sets of VH or VK gene sequences from forward and reverse sequencing. In the VH sequence set, the forward and the reverse sequencing data had 76,483 and 79,011 sequences, respectively. In the VK sequence set, the forward and the reverse sequencing data had 93,031 and 89,743 sequences, respectively. The VH or VK sequences were translated into three possible reading frames and subjected to a filtering process with the antibody numbering scheme and CDR definitions referred from Dr. Andrew C.R. Martin's Group (http://www.bioinf.org.uk/abs/). The invariant residues adjacent to the CDRs and the cysteine residues required to form disulfide bonds were also included in the filtering process. Finally, 7,373 unique VH sequences and 41,804 unique VK sequences were collected and analyzed for germline classification using the BLASTP program. Difference in amino acid sequence, CDR length, and amino acid frequencies were determined by using EXCEL VISUAL BASIC Macros.

### Selection and generation of human mAbs from a phage-displayed Fab library

A phage-displayed Fab library was prepared from the library glycerol stocks using VCSM13 helper phages, as described previously (Zhu and Dimitrov, 2009). Phages were subjected to 3-4 rounds of panning against 10 antigens such as L1 cell adhesion molecule (L1CAM), angiopoietin 2 (Ang2), extracellular domain (ECD) of Tie-2, activation-inducible TNF receptor (AITR), activation-inducible TNF receptor ligand (AITRL), monokine induced by interferon γ (MIG), brain cytoplasmic 200 (BC200) RNA, transmembrane 4 L6 family

member 5 (TM4SF5), preS1 peptide of hepatitis B virus (HBV), and Notch4 EGF 11-13, using the antigen immobilized on MaxiSorp™ plates (Nunc), biotinylated antigen bound to streptavidin-coated Dynabeads (ThermoFisher), or antigen-Fc fusion protein captured via goat anti-Fc antibody, according to standard protocols (Coomber, 2002). After panning, phage-Fabs were screened by enzyme-linked immunosorbent assay (ELISA), and positive binders were converted into soluble Fabs by digestion of the phagemid plasmid with *Not*I, deleting the gene encoding the cytoplasmic domain of protein III. Soluble Fabs were expressed in *E. coli* TG1 cells at 30°C overnight after addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM, and the culture supernatants were individually subjected to quantitative ELISA and indirect ELISA to quantify the soluble Fab secreted in the culture supernatant and to measure its antigen-binding activity, respectively. The clones showing high affinities were selected, and their VH and VK sequences were determined.

To convert the selected Fabs into whole IgG1 format, the human VH and VK sequences were amplified by PCR and combined with the leader sequences of IgG heavy and light chains, respectively, by recombinant PCR. The VH and VK with leader sequences were sequentially subcloned into the *Eco*RI-*Apa*I and *Hin*dIII-*Bsi*WI sites, respectively, in the antibody expression cassette (pdCMV-dhfrC) containing the human constant region of γ1 heavy chain (Cγ1) and CK gene (Yoon et al., 2006). For transient expression of the IgG1, the resulting expression plasmid was introduced into HEK293T cells using Lipofectamine (Invitrogen) according to the manufacturer's instructions, and the transfected cells were cultured in protein-free medium (CD293, Invitrogen). The culture supernatant was recovered and analyzed for antigen-binding affinity of the IgG1 using quantitative ELISA and indirect ELISA. The IgG1 was purified from the culture su-

pernatant by affinity chromatography on Protein A (Millipore, USA), and its purity was analyzed by SDS-polyacrylamide gel electrophoresis under reducing and non-reducing conditions and by using the Agilent 2100 Bioanalyzer (Agilent Technology, Germany). Protein concentration was determined by UV-spectrophotometry (Nanodrop, Thermo Fisher Scientific, USA).

## RESULTS

### Generation of a human naïve Fab library
A VK repertoire was created by PCR amplification from the PBLs from 33 healthy human donors using the degenerate family primers for all VK germline genes and cloned into the *Bst*XI sites of pKRIBB-FabD. We did not create Vλ repertoire because most of approved therapeutic antibodies are IgG/κ and also we thought that Vλ/CK pair might be less stable compared to VK/CK. Subsequently, a VH repertoire was created by PCR amplification from the PBLs from 803 donors and normal tissue samples (lymph node, spleen, and bone marrow) from healthy human donors using the degenerate family primers for all VH germline genes and cloned into the *Sfi*I sites of the phagemid DNA containing a VK repertoire. The resulting Fab library was introduced into *E. coli* TG1 by electroporation, and approximately $3 \times 10^{10}$ colonies were obtained. To assess the diversity of the library, phagemid DNA was prepared from 200 randomly selected bacterial colonies, then the VH and VK sequences were individually determined by dideoxynucleotide sequencing and analyzed using the IMGT human antibody database. Among these, 94% of VH and 97% of VK sequences have functional open-reading frame without stop codons, while 41 VH germline genes and 32 VK germline genes were identified, indicating that the natural antibody library was successfully constructed.
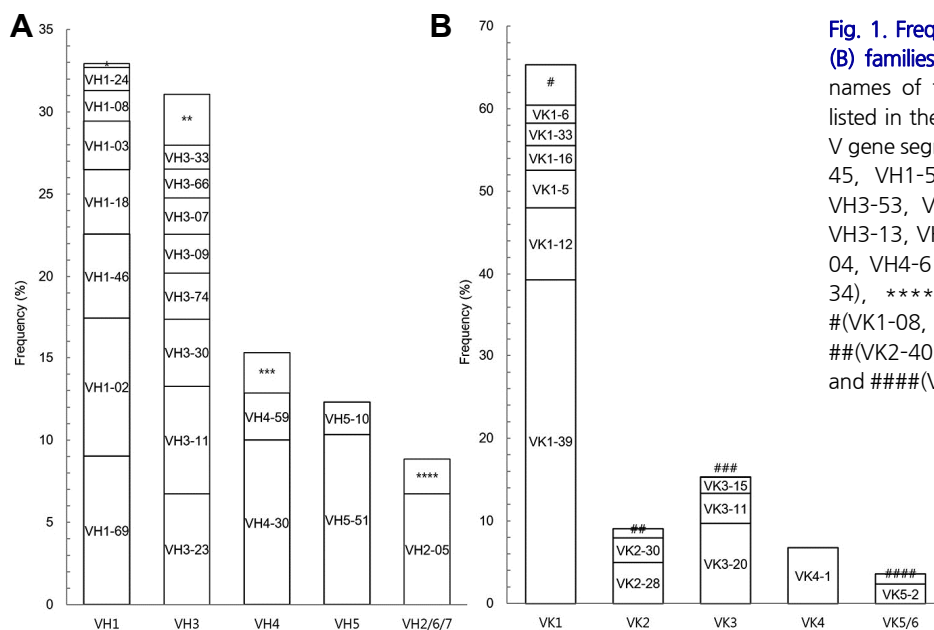


Fig. 1. Frequency distribution of VH (A) and VK (B) families and germline gene fragments. The names of the most frequent V segments are listed in the respective boxes, and less frequent V gene segment groups are indicated by *(VH1-45, VH1-58), **(VH3-48, VH3-21, VH3-20, VH3-53, VH3-72, VH3-73, VH3-64, VH3-43, VH3-13, VH3-15, VH3-49), ***(VH4-28, VH4-04, VH4-61, VH4-39, VH4-31, VH4-38, VH4-34), ****(VH6-01, VH2-26, VH7-4, VH2-70), #(VK1-08, VK1-13, VK1-09, VK1-17, VK1-27), ##(VK2-40, VK2-24, VK2-29), ###(VK3-07), and ####(VK6-41, VK6-21).

## 454 sequencing of the human naïve Fab library

To analyze the diversity of VH and VK repertoires in the human naïve antibody library, phagemid DNA was isolated from the *E. coli* transformants, and the VH and VK PCR amplicon libraries were generated to conduct 454 pyrosequencing. Two (forward and reverse) sequencing runs yielded an average 77,747 VH and 91,387 VK reads. The sequence data were then filtered by stepwise data processing procedures that we developed (Supplementary Fig. 1). Finally, 7,373 VH and 41,804 VK unique sequences, which were perfectly matched with the antibody numbering scheme and CDR definitions, were obtained.

## Germline classification of the antibody library clones

The 7,373 VH and 41,804 VK sequences were classified by Ig V-gene sequence databases such as V BASE (Cambridge, UK) and Kabat numbering (UCL, UK). All sequences were classified into the 48 VH and 32 VK germline segments. As shown in Fig. 1, the frequencies of the VH subfamilies were VH1 32.9%, VH3 31.1%, VH4 14.9%, VH5 12.3%, VH2 7.9%, VH7 0.8%, and VH6 0.2%; those of the VK subfamilies were VK1 65.2%, VK3 15.3%, VK2 9.2%, VK4 6.76%, VK5 2.3%, and VK 6 1.2%.

Within VH subfamilies, the most frequently found gene segments were VH1-69 (9.0%), VH1-02 (8.5%), VH1-46 (5.1%), VH1-18 (3.9%), VH2-05 (6.7%), VH3-23 (6.7%), VH3-11 (6.5%), VH3-30 (4.1%), VH4-30 (10.0%), and VH5-51 (10.3%) (Fig. 1A). Within VK families, the VK1-39 (39.3%) gene segment was the most frequently found,

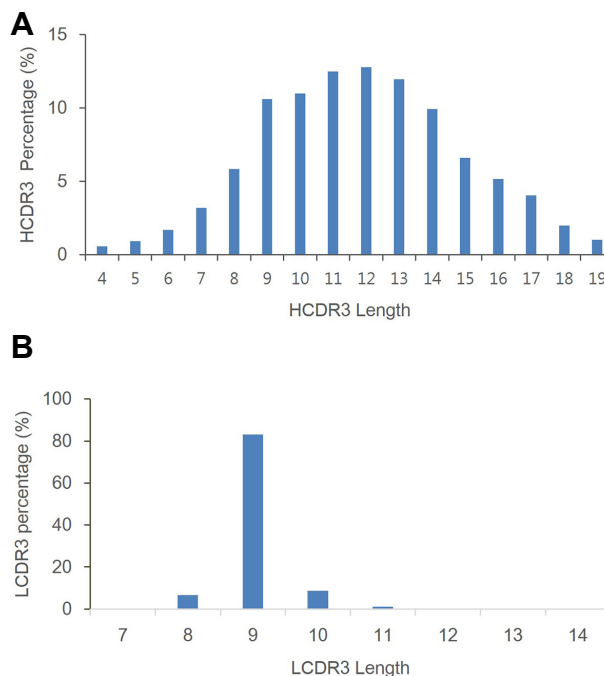followed by VK3-20 (9.7%), VK1-12 (8.7%), and VK4-1 (6.8%), in that order (Fig. 1B).

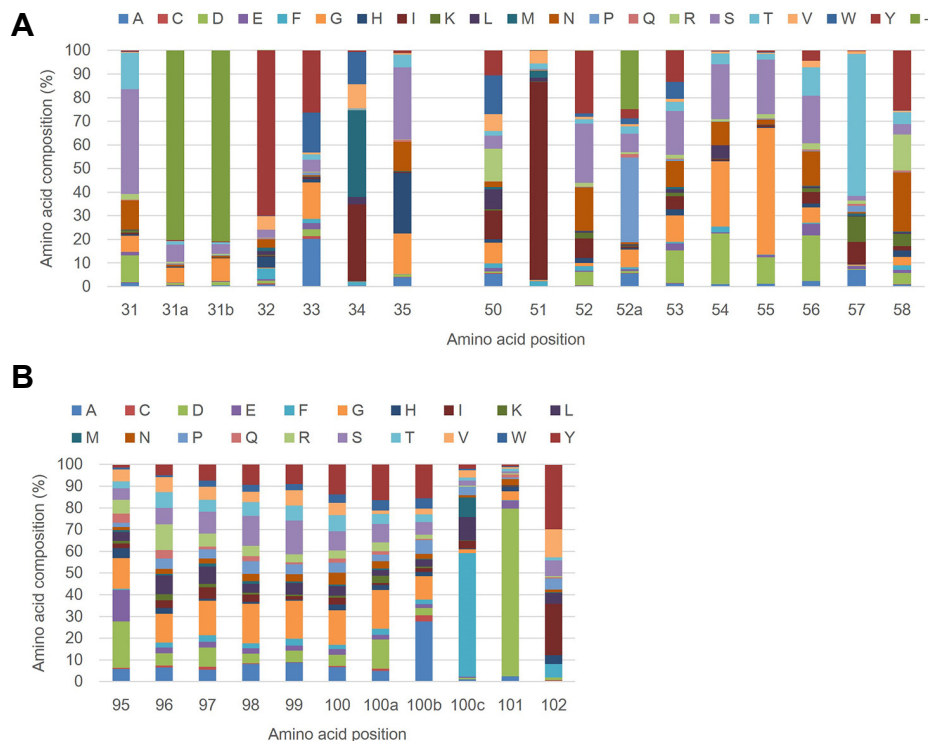Fig. 2. Length diversity of HCDR3 (A) and LCDR3 (B).

Fig. 3. Diversity of HCDR1 and HCDR2 (A) or HCDR3 (B). (A) Dash (-) indicates a percentage of the sequences with no amino acid at position 31a, 31b, or 52a.

## CDR3 length and amino acid distribution of CDRs

Analysis of the 7,373 VH sequences revealed that HCDR3 lengths ranged from 4 to 19 amino acids, while 11 or 12 amino acids was the most frequently occurring length (Fig. 2A). For VH, the amino acid composition at each position in HCDR1 (Kabat number 31-35), HCDR2 (Kabat number 50-58) and HCDR3 (Kabat number 95-102) were determined. The data showed that diversity is greatest at position 33 (H33), H50, H53, H56 and H58, followed by H31, H52, H52a and H54 in HCDR1 and HCDR2 (Fig. 3A). However, diversity in HCDR3 is much greater than that in HCDR1 or HCDR2 (Fig. 3B).

Analysis of the 41,804 VK sequences revealed that LCDR3 length ranged from 8 to 11 amino acids, while a length of 9 amino acids was predominant (Fig. 2B). We determined the amino acid composition at each position in LCDR1 (Kabat number 28-34), LCDR2 (Kabat number 50-58), and LCDR3 (Kabat number 89-97). Diversity at position 30 (L30), L31, L50, L53, L91, L92, L94, or L96 is higher compared with the other positions (Fig. 4A and 4B).

## Distribution of SHMs in CDRs

To analyze the frequency of SHM, the 7,373 VH and 41,804 VK sequences were compared with the closest germline VH or VK gene segments in V BASE. As expected, mutation frequencies within the CDRs were significantly higher than those within FRs (Supplementary Fig. 2). In HCDR1 and HCDR2, 14% of the 7,373 VH sequences had no SHM, 62% of them had between 1 and 6 mutations, and the remaining 24% had between 7 and 14 mutations (Fig. 5A), while an average of four amino acids were mutated.

For VK, in LCDR1 (Kabat number 24-34) and LCDR2 (Kabat number 50-56), only 3% of 41,804 VK sequences had no mutation; 80% of them had between 1 and 6 mutations, and the remaining 17% had between 7 and 14 mutations, while an average of 4 amino acids (two amino acids in LCDR1 or LCDR2) were mutated. Within LCDR3 part (Kabat number 89-95) in VK germline gene segments, 13% had no mutation, and the remaining 87% had between 1 and 6 mutations, while an average of 2 amino acids were mutated. In three LCDRs, only 1% had no mutation, 57% had between 1 and 6 mutations, and 41% had between 7 and 14 mutations, while an average of 6 amino acids were mutated. The results indicate that the three LCDRs have fairly even frequency of SHM, which is consistent with a previous report (Glanville et al., 2009).

The frequencies of mutations at a given position in the CDRs and FRs were also determined (Fig. 6). Mutations occurred with a frequency of 25% or more at H31, H33, or H35 in HCDR1; at H50, H53, H54, H56, or H58 in HCDR2 (Fig. 6A); at L28, L30, or L31 in LCDR1; at L50, L51, or L53 in LCDR2; and at L91-L94 in LCDR3 (Fig. 6B). Among them, H33, H50, H53, H58, L50, L91 and L94 are also highly diversified by germline diversity (Figs. 6A and 6B), which are
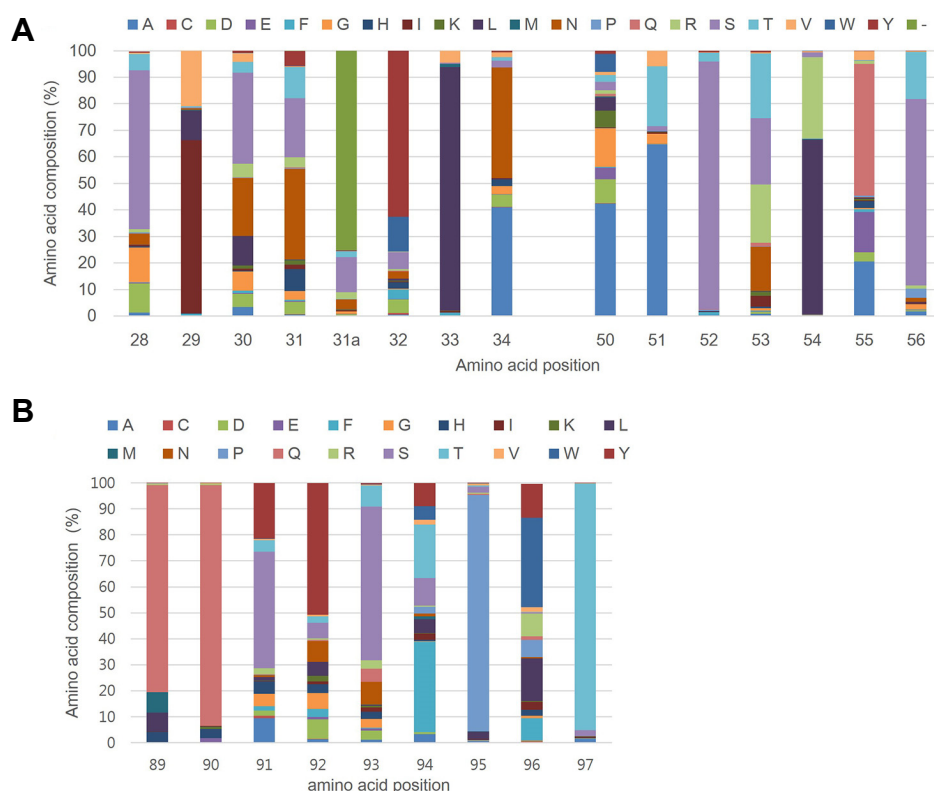


Fig. 4. Diversity of LCDR1 and LCDR2 (A) or LCDR3 (B). (A) Dash (-) indicates a percentage of the sequences with no amino acid at position 31a.
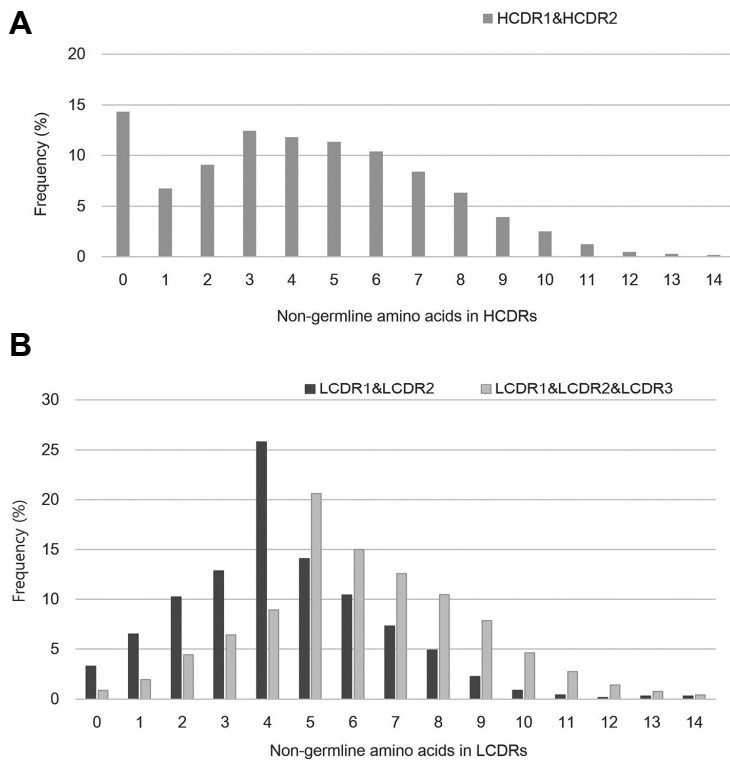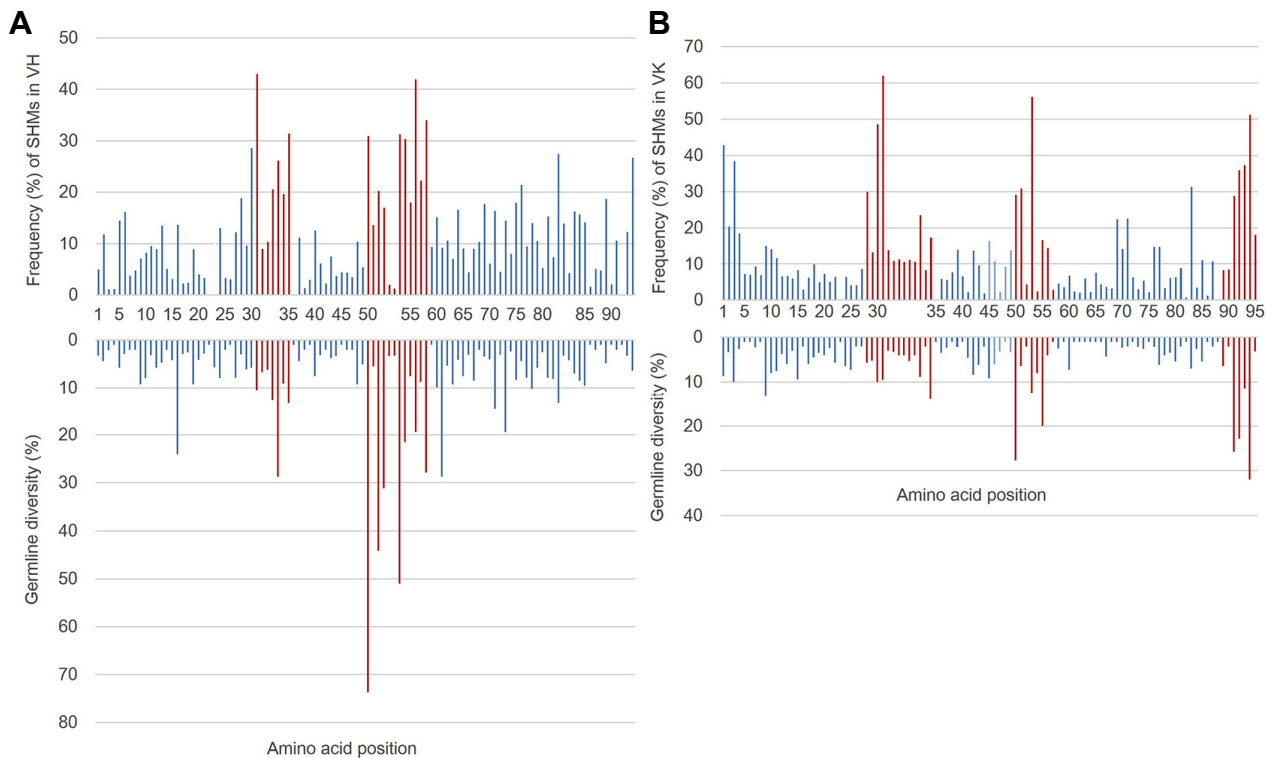
**A**

**B**



**A**



**B**



Fig. 6. The frequency of somatic hypermutation and germline diversity at a given position in the VH (A) and VK (B). Percentage of SHMs
was calculated as the total number of differences between the rearranged sequences and their corresponding V germline gene seg-
ments in a specific position, divided by the number of residues at that position. The CDRs and FRs are shown in red and blue, respective-
ly. Germline diversity equals the total number of different residues at each position divided by the relative frequency of the most com-
mon residue at that position.

rendered the most diversified positions in the CDRs of germline VH and VK, as shown in Figs. 3 and 4. In addition, mutation frequencies of 25% or more were observed in the FR of VH (H30, H82a, or H94) or VK (L1 or L3). The distributions of mutations in the CDRs and FRs of VH and VK were in good agreement with those given in previous reports (Burkovitz et al., 2014; Clark et al., 2006; Raghunathan et al., 2012; Tomlinson et al., 1996), except that at L83 in FR3 of VK, germline phenylalanine is mutated to valine in 42% of abundant VK1-39 sequences.

### Amino acid composition of position-specific SHMs in CDRs

We determined the amino acid composition of non-germline residues at each CDR position in the VH or VK sequences (Fig. 7). Among the highly diversified eight positions in the HCDR1 and HCDR2, asparagine or threonine at H31, asparagine at H35, arginine at H50, and aspartic acid at H54 were frequently found, accounting for 8% or more (Fig. 7A). Among the highly diversified ten positions in the LCDRs, asparagine at L30, asparagine or threonine at L31, threonine at L51, arginine or threonine at L53, threonine at L93, and phenylalanine at L94 accounted for more than 8% (Fig. 7B). Taken together, asparagine and threonine are the most frequent substitutions (H31, H35, L30, L31, L51, L53, and L93), and arginine is the next most frequent (H50 and L53), while

aspartic acid and phenylalanine are over-represented particularly at H54 and L94, respectively.

A previous study reported that some germline residues on the antibody-antigen interface, including serine, tyrosine, glycine and asparagine, have a high tendency to be coded in mutation hotspot motifs, and that their abundance is decreased during antibody maturation by a net conversion to other types: serine to asparagine, threonine, arginine and tryptophan; tyrosine to histidine, glutamine and phenylalanine; glycine to aspartic acid, alanine, tryptophan and arginine; asparagine to aspartic acid and serine (Clark et al., 2006). In the present study, to examine whether the most frequent five non-germline residues (asparagine, threonine, arginine, aspartic acid and phenylalanine) on the antibody-antigen interface are converted from the well represented germline mutation hotspots, we determined their germline counterparts (Fig. 8). In accordance with the intrinsic SHM propensities, the five non-germline residues are converted mostly from the germline serine, glycine, tyrosine, and/or asparagine, with the exception of arginine at H50 and phenylalanine at L94. Arginine at H50 is converted substantially from a germline tyrosine (35%), and phenylalanine at L94 is mostly converted from a germline threonine (83%). These residue type changes were shown to occur with relatively lower intrinsic mutation probabilities (Clark et al., 2006).
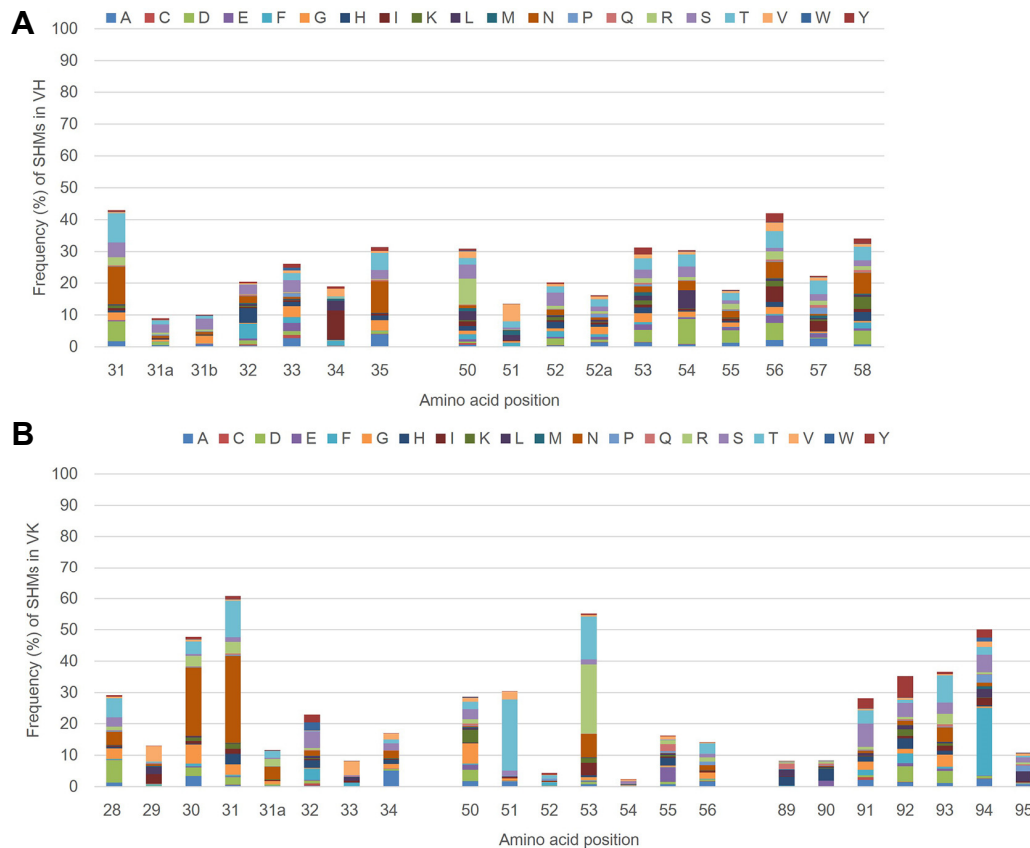


Fig. 7. Position-specific amino acid frequencies of SHMs for each CDR position in VH (A) and VK (B). The amino acid composition of SHMs was calculated as the number of each SHM in a specific position, divided by the total number of residues at that position.

| Kabat number | Non-germline residue | Germline residues (%) | | | | |
|---|---|---|---|---|---|---|
| H31 | N | S 91 | G 4 | D 3 | N 2 | |
| H31 | T | S 89 | D 7 | G 3 | etc 1 | |
| H35 | N | S 85 | H 14 | G 1 | | |
| H50 | R | G 53 | Y 35 | W 7 | I 3 | etc 2 |
| H54 | D | N 78 | G 17 | S 5 | | |
| L30 | N | S 99 | R 1 | | | |
| L31 | N | S 96 | Y 3 | H 1 | | |
| L31 | T | S 92 | N 7 | etc 1 | | |
| L53 | R | S 97 | N 1 | T 1 | etc 1 | |
| L53 | T | S 94 | N 6 | | | |
| L93 | T | S 89 | N 10 | etc 1 | | |
| L94 | F | T 83 | Y 11 | L 3 | etc 3 | |

Fig. 8. Frequency of SHM from a given germline residue to non-germline residue in the HCDRs or LCDRs.

## Amino acid composition of HCDR3

HCDR3 is formed by the junction of V-(D)-J joining and diversity in the HCDR3 contributes most to paratope diversity in antibodies, as shown in Fig. 3B. Looking at the amino acid composition of all HCDR3 length groups (4-19 amino acids),

aspartate, glycine and tyrosine are the most frequent, followed by serine, four hydrophobic amino acid residues (alanine, phenylalanine, leucine and valine), arginine, proline, threonine and isoleucine, in that order, whereas cysteine, lysine, glutamine, methionine, tryptophan, histidine, asparagine and glutamic acid are infrequent, in that order (Fig. 9). The overall relative amino acid frequencies are similar to those in a previous report on 4,751 human HCDR3 sequences (Zemlin et al., 2003), except that frequencies of aspartic acid and phenylalanine are a few percentage higher in our repertoire than that in this previous report.

With respect to length-dependent changes in amino acid frequencies, the frequency of aspartic acid (21.4 → 10.4%) steadily decreased with increasing HCDR3 length, whereas that of tyrosine (10.0 → 15.8%), serine (4.0 → 9.0%), cysteine (0 → 2.9%), and methionine (0.6 → 2.7%) tended to increase with increasing length. The frequency of glycine tended to decrease with increasing length for sequences of 4-9 amino acid residues (15.3 → 11.7%). The frequency of phenylalanine tended to increase with increasing length for sequences of 4-9 amino acid residues (4.9 → 7.5%), was maintained at length 9-14 (7.4%), then decreased for sequences of 14-19 amino acid residues (7.0 → 4.8%). The frequency of alanine, leucine, valine, asparagine, arginine and isoleucine did not change with increasing length. These data are similar to those in a previous report (Zemlin et al., 2003), except that frequency of arginine tended to decrease with increasing HCDR3 length in the earlier study.

## Selection of mAbs from a human naïve Fab library by phage display

To validate the diversity of the naïve Fab library, phages were prepared from the library, and Fab phages were screened by panning against 10 different therapeutically relevant antigens, including eight human recombinant proteins, RNA,
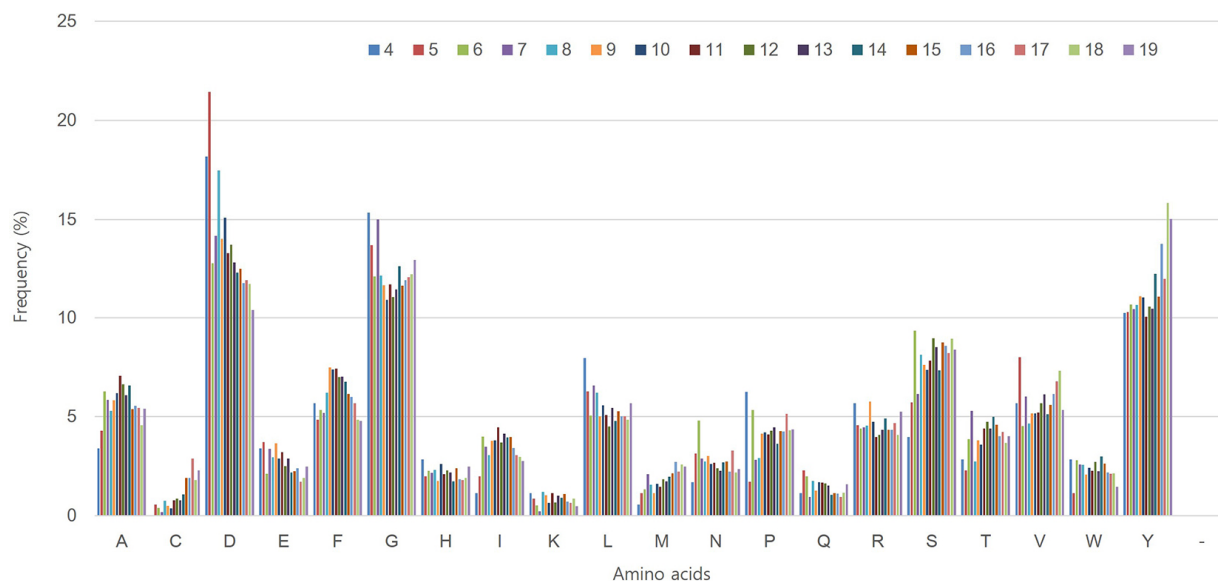


Fig. 9. Amino acid composition and length-dependent changes in amino acid frequencies of HCDR3.

and peptides: L1 cell adhesion molecule (L1CAM), angiopoietin 2 (Ang2), extracellular domain (ECD) of Tie-2, activation inducible TNF receptor (AITR), activation inducible TNF receptor ligand (AITRL), monokine induced by interferon γ (MIG), transmembrane 4 L6 family member 5 (TM4SF5), Notch 4, brain cytoplasmic 200 (BC200) RNA, and preS1 peptide of hepatitis B virus (HBV), as described in the Materials and Methods. After three or four rounds of panning against each antigen, 94 clones were randomly selected, and Fabs were expressed in *E. coli* at 30℃ overnight in a 96-well format. The culture supernatants were then analyzed by quantitative ELISA, and indirect ELISA with bovine serum albumin or skim milk was used as negative control antigen to select antigen-specific Fabs. A positive hit was defined by binding to a target antigen with a signal at least 10-fold above the assay background. The phagemid DNAs isolated from the selected positive hits were subjected to sequence analysis. The results revealed that two to nine unique Fabs were identified per antigen: nine from L1CAM, four from AITR, eight from Ang2, seven from Tie2, five from AITRL, three from MIG, six from TM4SF5, four from Notch4, two from BC200 RNA, and three from preS1 peptide. The unique antibodies used the VH3 (46%), VH1 (38%), VH5 (12%), and VH4 (4%) families, while VK1 was predominant and some were VK3, indicating that most of the dominant germline families were represented in the leads generated after panning.

For further characterization of selected Fabs, three Fabs (3HC, Ab4, and FabBC200-A) specific to L1CAM or BC200 RNA were converted to whole IgG. The resulting 3HC antibody specifically bound to the Ig1 domain of L1CAM in epitope mapping using serially deleted ECD of human L1CAM (Wei et al., 2011). The affinities ($K_D$) of 3H3 antibody for human and mouse L1CAM were 7.3 nM and 14 nM, respectively, as determined by competition ELISA (data not shown). Ab4 antibody bound to the Ig5 domain of L1CAM with the affinities ($K_D$) for human L1CAM (130 nM) and mouse (1.7 nM) (Cho et al., 2016). The IgG form of FabBC200-A exhibited a $K_D$ of 36 nM for human (BC200) RNA (Jung et al., 2014). The results indicate that the diversity of the human Fab library is high enough to isolate specific antibodies against diverse antigens.

## DISCUSSION

In this study, we constructed a large naïve human Fab library, obtained a database of unique 7,373 VH and 41,804 VK sequences by 454 pyrosequencing, and thoroughly determined the diversities of the VH and VK repertoires. Analyses of the frequencies of VH or VK subfamilies in our antibody library showed that VH1 (33%) and VH3 (31%) are almost equally the most frequent, followed by VH4 (15%), VH5 (12%), and VH2 (8%), while VK1 (65%) is predominant among the analyzed 41,804 sequences, followed by VK3 (15%) and VK2 (9%). Another study with natural human IgM repertoires in 18,158 VH or VK sequences showed that VH1 (36%) was the most frequent, followed by VH3 (25%) and VH4 (21%), while VK3 (41%) was the most frequent, followed by VK1 (30%) and VK2 (7%) (Glanville et al.,

2009). In one investigation that combined repertoires of naïve non-mutated IgM and somatically mutated IgM and IgG in 2,464 VH or 1,656 VK sequences, VH3 (48%) was the most frequent, followed by VH4 (27%) and VH1 (18%), while VK1 (44%) was the most frequent, followed by VK3 (36%) and VK2 (10%) (Tiller et al., 2013). In addition, two other studies of natural antibody repertoires showed that VH3 was the most frequent, followed by VH1 and VH4, while VK1 (38-36%) was slightly more frequent than VK3 (27-32%), followed by VK2 (9-20%) (Kugler et al., 2015; Lloyd et al., 2009; Schwimmer et al., 2013). Taken together, the data demonstrate that the distribution of VH and VK subfamilies in our antibody library differs from that of other natural antibody libraries published so far.

Regarding the usage of VH germline genes, VH1-69 (9.0%), VH1-02 (8.5%), VH1-46 (5.1%), VH1-18 (3.9%), VH2-05 (6.7%), VH3-23 (6.7%), VH3-11 (6.5%), VH3-30 (4.1%), VH4-30 (10.0%), and VH5-51 (10.3%) were frequently found in our VH repertoire. The VH germline gene usage of our library is similar to that of the IgM or combined repertoires described above (Glanville et al., 2009; Lloyd et al., 2009; Tiller et al., 2013), except that the frequencies of VH2-05, VH3-11, and VH4-30 were greatly reduced in these other repertoires while those of VH3-7, VH3-9, VH4-34, VH4-39, and VH4-59 were largely increased. It was interesting to note that VH4-34 and VH4-39, which were frequently found in other natural antibody repertoires but greatly reduced after phage selection (Glanville et al., 2009; Lloyd et al., 2009; Tiller et al., 2013), were infrequent in our antibody library. For VK germline gene usage, VK1-39 (39%) was the most frequently found in the VK repertoire, followed by VK3-20 (10%), VK1-12 (9%), and VK4-1 (7%), in that order. These data are similar to the frequencies of VK germline gene segments in other antibody repertoires (Glanville et al., 2009; Tiller et al., 2013) except that the frequency of VK1-12 was greatly reduced and that of VK3-11 and VK3-15 was moderately increased in these other antibody repertoires. Taken together, the germline gene data indicate that seven VH germline genes (VH1-69, VH1-02, VH1-46, VH1-18, VH3-23, VH3-30, and VH5-51) and three VK germline genes (VK1-39, VK3-20, and VK4-1) are frequently found in different natural antibody libraries, while the frequencies of the other VH and VK germline are differential among antibody libraries. In agreement with this, VH1-69, VH3-23, VH3-30, VK1-39 and VK3-20 are commonly used in synthetic antibody libraries (Tiller et al., 2013; Zhai et al., 2011).

Analysis of SHMs in the HCDR1 and HCDR2 indicated that 14% of the 7,373 VH sequences had no mutation, 62% of them had between 1 and 6 mutations, and 24% of them had more than 6 mutations, while an average of 4 SHMs were found. In contrast, in case of the natural IgM repertoire, 17% of them had no SHM, 78% had between 1 and 6 mutations, and 5% had more than 6 mutations, while an average of 4 SHMs was found (Glanville et al., 2009). Thus, the frequency of more than 6 SHMs in the HCDR1 and HCDR2 was higher in our library than in the IgM repertoire. Given that our naïve Fab library has combined repertoires from IgM and other Ig isotypes, the data are consistent with the fact that more SHMs occur in IgG than in IgM because AID is the

key mediator of SHM and class switching (Muramatsu et al., 2000).

SHMs were found with a frequency of 25% or more at eight positions (H31, H33, H35, H50, H53, H54, H56, and H58) in the HCDR1 and HCDR2 and at ten positions (L28, L30, L31, L50, L51, L53, and L91-L94) in the LCDRs. Among these positions, five non-germline residues (asparagine, threonine, arginine, aspartic acid and phenylalanine) were shown to be the most frequently occurring and are converted mostly from germline serine, glycine, tyrosine and/or asparagine, which are on the antibody-antigen interface and have a high tendency to be present in germline mutation hotspots, in accordance with the intrinsic SHM propensities (Clark et al., 2006). However, arginine at H50 and phenylalanine at L94 are also converted from germline tyrosine and threonine, respectively, and these residue type changes were shown to occur with relatively lower intrinsic mutation probabilities (Clark et al., 2006). Considering the previous report that the residue type mutation patterns are mainly determined by biases in the germline, location of mutation hotspots and functional pressures during selection (Burkovitz et al., 2014), the SHMs at H50 and L94 may have occurred as a result of biases in the germline and functional pressures during selection.

In conclusion, we generated a large naïve human Fab library and determined the diversities of the VH and VK repertoires. The results suggest that our library not only shares common features of natural antibody repertoires but also has unique features compared to other natural antibody libraries reported previously. The information about the diversity of natural antibody repertoires will provide more understanding of human antibody repertoires and insights into the future design of synthetic human antibody libraries with high functional diversity. Given that we successfully demonstrated the generation of specific mAbs from the antibody library and that Fab fragments possess comparably higher structural stability compared to scFvs, this phage antibody library will be useful in the generation of human mAbs against diverse antigens.

*Note: Supplementary information is available on the Molecules and Cells website (www.molcells.org).*

## REFERENCES

Barbie, V., and Lefranc, M.P. (1998). The human immunoglobulin kappa variable (IGKV) genes and joining (IGKJ) segments. Exp. Clin. Immunogenet. *15*, 171-183.

Boder, E.T., and Wittrup, K.D. (1997). Yeast surface display for screening combinatorial polypeptide libraries. Nat. Biotechnol. *15*, 553-557.

Bruggemann, M., Osborn, M.J., Ma, B., Hayre, J., Avis, S., Lundstrom, B., and Buelow, R. (2015). Human antibody production in transgenic animals. Arch. Immunol. Ther. Exp. (Warsz.) *63*, 101-108.

Burkovitz, A., Sela-Culang, I., and Ofran, Y. (2014). Large-scale analysis of somatic hypermutations in antibodies reveals which structural regions, positions and amino acids are modified to improve affinity. The FEBS journal *281*, 306-319.

Chaplin, D.D. (2003). 1. Overview of the immune response. J. Allergy Clin. Immunol. *111*, S442-459.

Cho, S., Park, I., Kim, H., Jeong, M.S., Lim, M., Lee, E.S., Kim, J.H., Kim, S., and Hong, H.J. (2016). Generation, characterization and preclinical studies of a human anti-L1CAM monoclonal antibody that cross-reacts with rodent L1CAM. mAbs *8*, 414-425.

Clark, L.A., Ganesan, S., Papp, S., and van Vlijmen, H.W. (2006). Trends in antibody sequence changes during the somatic hypermutation process. J. Immunol. *177*, 333-340.

Coomber, D.W. (2002). Panning of antibody phage-display libraries. Standard protocols. Methods Mol. Biol. *178*, 133-145.

Cox, J.P., Tomlinson, I.M., and Winter, G. (1994). A directory of human germ-line V kappa segments reveals a strong bias in their usage. Eur. J. Immunol. *24*, 827-836.

Ecker, D.M., Jones, S.D., and Levine, H.L. (2015). The therapeutic monoclonal antibody market. mAbs *7*, 9-14.

Geng, X., Kong, X., Hu, H., Chen, J., Yang, F., Liang, H., Chen, X., and Hu, Y. (2015). Research and development of therapeutic mAbs: An analysis based on pipeline projects. Hum. Vaccin. Immunother. *11*, 2769-2776.

Glanville, J., Zhai, W., Berka, J., Telman, D., Huerta, G., Mehta, G.R., Ni, I., Mei, L., Sundar, P.D., Day, G.M., et al. (2009). Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. Proc. Natl. Acad. Sci. USA. *106*, 20216-20221.

Gram, H., Marconi, L.A., Barbas, C.F., 3rd, Collet, T.A., Lerner, R.A., and Kang, A.S. (1992). In vitro selection and affinity maturation of antibodies from a naive combinatorial immunoglobulin library. Proc. Natl. Acad. Sci. USA. *89*, 3576-3580.

Hanes, J., and Pluckthun, A. (1997). In vitro selection and evolution of functional proteins by using ribosome display. Proc. Natl. Acad. Sci. USA. *94*, 4937-4942.

Hoet, R.M., Cohen, E.H., Kent, R.B., Rookey, K., Schoonbroodt, S., Hogan, S., Rem, L., Frans, N., Daukandt, M., Pieters, H., et al. (2005). Generation of high-affinity human antibodies by combining donor-derived and synthetic complementarity-determining-region diversity. Nat. Biotechnol. *23*, 344-348.

Hoogenboom, H.R. (2005). Selecting and screening recombinant antibody libraries. Nat. Biotechnol. *23*, 1105-1116.

Jackson, K.J., Wang, Y., and Collins, A.M. (2014). Human immunoglobulin classes and subclasses show variability in VDJ gene mutation levels. Immunol. Cell Biol. *92*, 729-733.

Jung, E., Lee, J., Hong, H.J., Park, I., and Lee, Y. (2014). RNA recognition by a human antibody against brain cytoplasmic 200 RNA. RNA *20*, 805-814.

Kawasaki, K., Minoshima, S., Nakato, E., Shibuya, K., Shintani, A., Schmeits, J.L., Wang, J., and Shimizu, N. (1997). One-megabase sequence analysis of the human immunoglobulin lambda gene locus. Genome Res. *7*, 250-261.

Kim, S.J., and Hong, H.J. (2012). Humanization by guided selections. Methods Mol. Biol. *907*, 247-257.

Kugler, J., Wilke, S., Meier, D., Tomszak, F., Frenzel, A., Schirrmann, T., Dubel, S., Garritsen, H., Hock, B., Toleikis, L., et al. (2015). Generation and analysis of the improved human HAL9/10 antibody phage display libraries. BMC Biotechnol. *15*, 10.

Lloyd, C., Lowe, D., Edwards, B., Welsh, F., Dilks, T., Hardman, C., and Vaughan, T. (2009). Modelling the human immune response: performance of a 1011 human antibody repertoire against a broad panel of therapeutically relevant antigens. Protein Eng. Des. Sel. *22*, 159-168.

Market, E., and Papavasiliou, F.N. (2003). V(D)J recombination and the evolution of the adaptive immune system. PLoS Biol. *1*, E16.

Matsuda, F., and Honjo, T. (1996). Organization of the human immunoglobulin heavy-chain locus. Adv. Immunol. *62*, 1-29.

McCafferty, J., Griffiths, A.D., Winter, G., and Chiswell, D.J. (1990). Phage antibodies: filamentous phage displaying antibody variable domains. Nature *348*, 552-554.

Muramatsu, M., Sankaranand, V.S., Anant, S., Sugai, M., Kinoshita, K., Davidson, N.O., and Honjo, T. (1999). Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. J. Biol. Chem. *274*, 18470-18476.

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell *102*, 553-563.

Oh, M.S., Kim, K.S., Jang, Y.K., Maeng, C.Y., Min, S.H., Jang, M.H., Yoon, S.O., Kim, J.H., and Hong, H.J. (2003). A new epitope tag from hepatitis B virus preS1 for immunodetection, localization and affinity purification of recombinant proteins. J. Immunol. Methods *283*, 77-89.

Pallares, N., Frippiat, J.P., Giudicelli, V., and Lefranc, M.P. (1998). The human immunoglobulin lambda variable (IGLV) genes and joining (IGLJ) segments. Exp. Clin. Immunogenet. *15*, 8-18.

Perelson, A.S., and Oster, G.F. (1979). Theoretical studies of clonal selection: minimal antibody repertoire size and reliability of self-non-self discrimination. J. Theor. Biol. *81*, 645-670.

Ponsel, D., Neugebauer, J., Ladetzki-Baehs, K., and Tissot, K. (2011). High affinity, developability and functional size: the holy grail of combinatorial antibody library generation. Molecules *16*, 3675-3700.

Raghunathan, G., Smart, J., Williams, J., and Almagro, J.C. (2012). Antigen-binding site anatomy and somatic mutations in antibodies that recognize different types of antigens. J. Mol. Recognit. *25*, 103-113.

Rothlisberger, D., Honegger, A., and Pluckthun, A. (2005). Domain interactions in the Fab fragment: a comparative evaluation of the single-chain Fv and Fab format engineered with variable domains of different stability. J. Mol. Biol. *347*, 773-789.

Schatz, D.G. (2004). V(D)J recombination. Immunol. Rev. *200*, 5-11.

Schwimmer, L.J., Huang, B., Giang, H., Cotter, R.L., Chemla-Vogel, D.S., Dy, F.V., Tam, E.M., Zhang, F., Toy, P., Bohmann, D.J., et al. (2013). Discovery of diverse and functional antibodies from large human repertoire antibody libraries. J. Immunol. Methods *391*, 60-71.

Selsted, M.E., and Ouellette, A.J. (2005). Mammalian defensins in the antimicrobial immune response. Nat. Immunol. *6*, 551-557.

Sharma, S., Byrne, H., and O'Kennedy, R.J. (2016). Antibodies and antibody-derived analytical biosensors. Essays Biochem. *60*, 9-18.

Tiller, T., Schuster, I., Deppe, D., Siegers, K., Strohner, R., Herrmann, T., Berenguer, M., Poujol, D., Stehle, J., Stark, Y., et al. (2013). A fully synthetic human Fab antibody library based on fixed VH/VL framework pairings with favorable biophysical properties. mAbs *5*, 445-470.

Tomlinson, I.M., Walter, G., Marks, J.D., Llewelyn, M.B., and Winter, G. (1992). The repertoire of human germline VH sequences reveals about fifty groups of VH segments with different hypervariable loops. J. Mol. Biol. *227*, 776-798.

Tomlinson, I.M., Walter, G., Jones, P.T., Dear, P.H., Sonnhammer, E.L., and Winter, G. (1996). The imprint of somatic hypermutation on the repertoire of human germline V genes. J. Mol. Biol. *256*, 813-817.

Tonegawa, S. (1983). Somatic generation of antibody diversity. Nature *302*, 575-581.

Trepel, F. (1974). Number and distribution of lymphocytes in man. A critical analysis. Klin. Wochenschr. *52*, 511-515.

Vargas-Madrazo, E., Lara-Ochoa, F., Ramirez-Benites, M.C., and Almagro, J.C. (1997). Evolution of the structural repertoire of the human V(H) and Vkappa germline genes. Int. Immunol. *9*, 1801-1815.

Wei, C.H., Lee, E.S., Jeon, J.Y., Heo, Y.S., Kim, S.J., Jeon, Y.H., Kim, K.H., Hong, H.J., and Ryu, S.E. (2011). Structural mechanism of the antigen recognition by the L1 cell adhesion molecule antibody A10-A3. FEBS Lett. *585*, 153-158.

Williams, S.C., Frippiat, J.P., Tomlinson, I.M., Ignatovich, O., Lefranc, M.P., and Winter, G. (1996). Sequence and evolution of the human germline V lambda repertoire. J. Mol. Biol. *264*, 220-232.

Wu, T.T., and Kabat, E.A. (1970). An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. J. Exp. Med. *132*, 211-250.

Yoon, S.O., Lee, T.S., Kim, S.J., Jang, M.H., Kang, Y.J., Park, J.H., Kim, K.S., Lee, H.S., Ryu, C.J., Gonzales, N.R., et al. (2006). Construction, affinity maturation, and biological characterization of an anti-tumor-associated glycoprotein-72 humanized antibody. J. Biol. Chem. *281*, 6985-6992.

Zemlin, M., Klinger, M., Link, J., Zemlin, C., Bauer, K., Engler, J.A., Schroeder, H.W., Jr., and Kirkham, P.M. (2003). Expressed murine and human CDR-H3 intervals of equal length exhibit distinct repertoires that differ in their amino acid composition and predicted range of structures. J. Mol. Biol. *334*, 733-749.

Zhai, W., Glanville, J., Fuhrmann, M., Mei, L., Ni, I., Sundar, P.D., Van Blarcom, T., Abdiche, Y., Lindquist, K., Strohner, R., et al. (2011). Synthetic antibodies designed on natural sequence landscapes. Journal of molecular biology *412*, 55-71.

Zhu, Z., and Dimitrov, D.S. (2009). Construction of a large naive human phage-displayed Fab library through one-step cloning. Methods Mol. Biol. *525*, 129-142, xv.