Contribution of $V_H$ Replacement Products in Mouse Antibody Repertoire

Lin Huang1,*, Miles D. Lange1,*, Yangsheng Yu1, Song Li1, Kaihong Su1,2,3, Zhixin Zhang1,2,*

1Department of Pathology and Microbiology, University of Nebraska Medical Center, Omaha, Nebraska, United States of America, 2The Eppley Cancer Institute, University of Nebraska Medical Center, Omaha, Nebraska, United States of America, 3Department of Internal Medicine, University of Nebraska Medical Center, Omaha, Nebraska, United States of America

Abstract

$V_H$ replacement occurs through RAG-mediated recombination between the cryptic recombination signal sequence (cRSS) near the 3’ end of a rearranged $V_H$ gene and the 23-bp RSS from an upstream unrearranged $V_H$ gene. Due to the location of the cRSS, $V_H$ replacement leaves a short stretch of nucleotides from the previously rearranged $V_H$ gene at the newly formed V-D junction, which can be used as a marker to identify $V_H$ replacement products. To determine the contribution of $V_H$ replacement products to mouse antibody repertoire, we developed a Java-based $V_H$ Replacement Footprint Analyzer (VHrFA) program and analyzed 17,179 mouse IgH gene sequences from the NCBI database to identify $V_H$ replacement products. The overall frequency of $V_H$ replacement products in these IgH genes is 5.29% based on the identification of pentameric $V_H$ replacement footprints at their V-D junctions. The identified $V_H$ replacement products are distributed similarly in IgH genes using most families of $V_H$ genes, although different families of $V_H$ genes are used differentially. The frequencies of $V_H$ replacement products are significantly elevated in IgH genes derived from several strains of autoimmune prone mice and in IgH genes encoding autoantibodies. Moreover, the identified $V_H$ replacement footprints in IgH genes from autoimmune prone mice or IgH genes encoding autoantibodies preferentially encode positively charged amino acids. These results revealed a significant contribution of $V_H$ replacement products to the diversification of antibody repertoire and potentially, to the generation of autoantibodies in mice.

Citation: Huang L, Lange MD, Yu Y, Li S, Su K, et al. (2013) Contribution of $V_H$ Replacement Products in Mouse Antibody Repertoire. PLoS ONE 8(2): e57877. doi:10.1371/journal.pone.0057877

Editor: Sebastian D. Fugmann, Chang Gung University, Taiwan

Received June 13, 2012; Accepted January 30, 2013; Published February 28, 2013

Copyright: © 2013 Huang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study was supported by National Institutes of Health grants AI074948 (ZZ), AI076475 (ZZ), and AR059351 (KS), and by faculty developmental fund for ZZ and KS from University of Nebraska Medical Center, Eppley Cancer Institute, Omaha, Nebraska. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: zhangj@unmc.edu
1 These authors contributed equally to this work.

Introduction

The variable region exons of the immunoglobulin (Ig) genes are generated through sequential rearrangement of previously separated $V_H$, $D_H$, and $J_H$ gene segments catalyzed by the recombination activating gene products (RAG1 and RAG2) [1–5]. The specific joining of VH,D H, and JH gene catalyzed by the recombination activating gene products (RAG1 and RAG2) [1–5]. The specific joining of VH,D H, and JH gene segments to the DJ H joint in late pro B cells [4,5]. If the rearrangement is non functional, pro B cells will start to rearrange the second IgH allele [4,5]. Functionally rearranged IgH genes will be expressed as the $\mu$ heavy chains to form pre-B cell receptors with the non-rearranged components, Vpre-B and lambda 5 [10–15]. Signaling from the pre-BCR will stimulate pre B cell proliferation and subsequent IgL gene rearrangement [14,15]. The IgL gene variable region exon is generated by a one step rearrangement between a $V_L$ segment and a $J_L$ segment in the small precursor (pre-) B cells [4,5,16]. Due to the random recombination process, two thirds of the VDJ rearrangement products might be out of reading frame and cannot express functional Ig peptides. Even if the IgH gene rearrangements are productive, they might fail to pair with the surrogate or conventional light chains. B cells lacking functional pre-B cell receptors (pre-BCRs) or B cell receptors (BCRs) cannot develop further along the B lineage pathway [14,17]. Moreover, functionally expressed BCRs may be self-reactive. In all these cases, early B lineage cells retain the abilities to initiate secondary RAG-mediated recombination to alter the rearranged Ig genes, a process known as receptor editing [18–20].

Editing of rearranged IgL genes can occur through RAG-mediated secondary recombination between any upstream $V_L$ gene to a downstream $J_L$ gene [21–26]. The intervening DNA fragment containing the previously rearranged $V_LJ_L$ joint is...
accumulated studies indicated that non-functional or autoreactive IgH gene rearrangements can be edited through a VH replacement process [27–33]. VH replacement occurs through RAG-mediated recombination between a cryptic RSS embedded at the 3’ end of the rearranged VH gene with the 23 bp RSS from a upstream VH gene [31]. VH replacement was originally observed in murine pre-B cell leukemia cells, which generated functional IgH genes from non-functional IgH rearrangements [27,28]. The potential biological function of VH replacement in editing IgH genes encoding anti-DNA antibodies was demonstrated in a series of studies using engineered mouse models carrying knocked-in IgH V(D)J rearrangements encoding anti-DNA antibodies [29,34,35]; Later studies also provided evidence that VH replacement was employed to diversify the antibody repertoire in mouse carrying knocked-in IgH genes encoding anti-NP antibodies [30,36] and to rescue B cells with two alleles of non-functional IgH rearrangements [32,33]. Despite of these findings in engineered mice, evidence for ongoing VH replacement during B cell development in normal mouse and contribution of VH replacement products to the mouse antibody repertoire were lacking for a long time [37,38].

Due to the location of the cRSS at the 3’ end of VH germline gene, VH replacement renews almost the entire VH coding region but leaves a short stretch of nucleotides from the previously rearranged VH gene at the newly formed V-D junction [28,31]. These remnants can be used as VH replacement footprints to trace the occurrence of VH replacement and to identify potential VH replacement products through analyzing IgH gene sequences [31]. Our previous analysis of 412 human IgH gene sequences estimated that VH replacement products contribute to about 5% of the primary B cell repertoire in human [31]. A recent analysis of IgH genes generated from knock-in mice expressing IgH genes encoding anti-DNA antibodies showed that 7.5% of the newly generated IgH genes contain pantemeric VH replacement footprints [39]. Similar frequency of VH replacement products were also found in IgH genes obtained from the wild type B6 mice [39].

To explore the contribution of VH replacement products to the diversification of mouse IgH repertoire, we developed a Java based VH replacement footprint analyzer (VHRFA) program and analyzed 17,179 mouse IgH gene sequences from the National Center for Biotechnology Information (NCBI) database to identify VH replacement products. These results revealed a significant contribution of VH replacement products to the murine IgH repertoire and the enrichment of VH replacement products in several strains of autoimmune prone mice.

Results

The Mouse IgH Sequence Repertoire

To analyze a large number of IgH gene sequences and to identify potential VH replacement products, we developed a Java based VH Replacement Footprint Analyzer (VHRFA) program. Using the VHRFA program, we analyzed 17,179 mouse IgH gene sequences from the NCBI databases to identify VH replacement products. First, the potential VH, DH, and JH germline gene usage were assigned using the IMGT/V-QUEST program by sending batches of sequences using the VHRFA program (shown in Table S1). Based on the IgH CDR3 region sequences, clonally identical sequences were stripped out. There are 11309 unique IgH gene sequences; 10159 of them have clearly identifiable VH, DH, and JH families of VH genes; 9774 of them are productive and 373 of them are non-productive IgH rearrangements. In these IgH genes, different families of VH genes are used differentially (Fig. 1). There are 63683 (65%) functional IgH genes using the IGHV1/VHJ558 family of VH genes; 911 (or 9.3%) functional IgH genes using the IGHV5/VHJ183 family of VH genes. The other families of VH genes, including IGHV4/X-24, IGHV11/CP3, IGHV12/CH27, IGHV13/3609N, and IGHV15/VH15A, are used at much lower frequencies (Fig. 1A). Among the non-functional IgH rearrangements, the usage of most VH gene families are similar to those in functional IgH genes, but the usages of the IGHV3/VHJ183 and IGHV3/36-60 gene families are increased (Fig. 1A). Among different DH genes, the IGHD1-1 gene is used the most frequent in almost 39% of the IgH sequences (Fig. 1B). For the JH genes, the IGHJ2 gene is used the most frequent in 43% of IgH genes (Fig. 1C). It should be noted that these 17179 mouse IgH sequences were derived from about 861 published reports (Table S2), presumably from more than 861 experiments with different mice. This analysis represents a comprehensive view of the IgH repertoire of the current available mouse IgH gene sequences in the NCBI database.

Identification of VH Replacement Products

In the initial test, we use the VHRFA program to identify potential VH replacement products in 271 mouse IgH gene sequences described previously [40]. Among them, 252 unique IgH genes have clearly identifiable VH, DH, and JH germline genes. Then, we searched for VH replacement footprint motifs with 3, 4, 5, 6, or 7 nucleotides within the VH-DH junction (N1) regions of these IgH genes. VH replacement can only introduce VH replacement footprint in the N1 region. As an internal control, we searched for similar VH replacement footprint motifs in the DH-JH junction (N2) regions of these IgH genes, which are likely generated by random nucleotide addition. The frequencies of 3, 4, and 5-mer VH replacement footprint motifs in the N1 regions are significantly higher than those in the N2 regions (Table 1, top), suggesting that the distribution of such motifs in the N1 region is not due to random nucleotide addition. Based on the identification of the pantemeric VH replacement footprints within the N1 regions, we estimate that the frequency of VH replacement products is 5.5% in these 252 mouse IgH gene sequences (Table 1, Top). If we consider the 4- or 3-mer of VH replacement footprints in the N1 regions, the frequencies of VH replacement products in these 252 IgH genes will be 21.2% or 38%, respectively (Table 1, top and the identified VH replacement products with 4-mer VH replacement footprints are shown in Table S5).

Further analysis of the 14 identified VH replacement products validated the assignment of VH replacement footprints by the VHRFA program (Table 2). Theoretically, VH replacement occurs through an upstream VH gene replacing a downstream rearranged VH gene. Among these 14 identified potential VH replacement products, 11 of them were likely generated through upstream VH genes replacing downstream VH genes; 3 of them did not follow such order (Table 2).

Contribution of VH Replacement Products to the Mouse IgH Repertoire

Next, we analyzed the 11,309 unique mouse IgH gene sequences from the NCBI database using the VHRFA program to search for VH replacement products. We performed separated analyses to identify VH replacement footprint with 3, 4, 5, 6, and 7 nucleotides in the VH-DH junction (N1) regions. As internal controls, we also searched for the similar motifs in the DH-JH junction (N2) regions. The frequencies of identified VH re-
placement footprints with 3, 4, 5, 6, or 7 nucleotides in the N1 regions are significantly higher than those in the N2 regions (Table 1, bottom). These results indicate that the presence of these motifs at the N1 region is not due to random nucleotide addition. With a stringent setting to search for the pentameric V_H replacement footprints at the N1 regions, 5.29% of the IgH genes contain such motifs and can be assigned as potential V_H replacement products. If we consider V_H replacement footprints with 4 or 3 nucleotides, 15.95% or 33.55% of the IgH genes, respectively, contain such motifs and can be assigned as potential V_H replacement products (Table 1, bottom). These results revealed a significant contribution of V_H replacement products to the diversification of the murine IgH repertoire.

Distribution of V_H Replacement Products in IgH Genes Using Different Families of V_H Genes

As we showed earlier, different V_H gene families are used at different frequencies in the 10159 mouse IgH gene sequences. Next, we analyzed the distribution of the identified V_H replacement products with 5-mer footprint motifs in IgH genes using different V_H gene families. Among all the IgH genes using different families of V_H genes, the frequency of V_H replacement products in IgH genes using the VH2/Q52 genes is significantly higher than that in the overall mouse IgH sequences (Table 1, bottom). The frequencies of V_H replacement products in IgH genes using the other V_H gene families are quite similar. For example, although the IGHV1/VHJ558 and IGHV5/V_H7183 families are used most frequently and the IGHV4/X-24, IGHV12/CH27, and IGHV14/SM7 families are used at very low frequencies, the frequencies of V_H replacement products in IgH genes using the IGHV1/VHJ558, IGHV5/V_H7183, IGHV4/X-24, IGHV12/CH27, and IGHV14/SM7 families are similar (Table 3). These results indicate that although different families of V_H genes are used differentially during the primary V(D)J recombination, they are similarly targeted for secondary recombination during V_H replacement. As an internal negative control, we analyzed the N1 regions of IgH genes using the DH proximal VH5-2/7183.2 gene. Among the 56 functional IgH genes using the VH5-2/7183.2 gene, there is no pentameric V_H replacement footprints in the N1 regions. Such result provides supporting evidence that the presence of pentameric footprints in the N1 regions of mouse IgH genes is contributed by V_H replacement.

Enrichment of V_H Replacement Products in IgH Genes Derived from Different Strains of Autoimmune Prone Mice and IgH Genes Encoding Autoantibodies

To explore the biological significance of V_H replacement in mouse, we analyzed the distribution of V_H replacement products in IgH genes correlating with different keywords in the NCBI database. Based on the identification of 5-mer V_H replacement footprints within the N1 regions, the frequencies of V_H replacement products in IgH genes derived from C57BL/6 and
**Table 1.** Frequencies of V\_H replacement footprint motifs with different length in the N1 and N2 regions of the mouse IgH genes.

| Number of unique Sequences\(^a\) | Number of Sequences with V, D, J genes\(^b\) | Minimal Length of V\_H replacement footprint | V\_H replacement footprint motifs in the N1 region\(^c\) | V\_H replacement footprint motifs in the N2 region\(^d\) | Frequency of V\_H replacement products (%)\(^e\) | p-value\(^*\) |
|----------------------------------|---------------------------------------------|---------------------------------------------|---------------------------------------------|---------------------------------------------|---------------------------------------------|------------------|
| Test IgH genes\(^g\)            | 271                                         | 252                                         | 3                                          | 101                                         | 65                                          | 0.0001           | 40.1            |
|                                  | 4                                           | 55                                          | 23                                         | 23                                         | 0.0001                                      | 21.8             |
|                                  | 5                                           | 14                                          | 4                                          | 0.0308                                     | 5.5                                         |                  |
|                                  | 6                                           | 2                                           | 0                                          | 0.4786                                     | 0.79                                        |                  |
|                                  | 7                                           | 1                                           | 0                                          | 0.3168                                     | 0.39                                        |                  |
| NCBI IgH genes\(^h\)            | 11309                                       | 10159                                       | 3                                          | 3384                                       | 2622                                        | 0.0001           | 33.55           |
|                                  | 4                                           | 1609                                        | 979                                         | 979                                        | 0.0001                                      | 15.95            |
|                                  | 5                                           | 534                                         | 256                                         | 256                                        | 0.0001                                      | 5.29             |
|                                  | 6                                           | 179                                         | 50                                          | 50                                         | 0.0001                                      | 1.77             |
|                                  | 7                                           | 45                                          | 8                                           | 8                                          | 0.0001                                      | 0.45             |

\(^a\)Unique sequences were identified after removal of IgH sequences with identical CDR3 regions.

\(^b\)Total number of IgH gene sequences with clearly identifiable V, D, H, a, n, and J genes.

\(^c\)N1 region refers to the V-D junction.

\(^d\)N2 region refers to the D-J junction.

\(^e\)The frequencies of potential V\_H replacement footprint motifs in the N1 and N2 regions were compared by two-tailed Chi-square with Yate’s correction. p<0.05 was considered significant and p<0.0001 is considered extremely significant.

\(^f\)Numbers of IgH gene sequences with V\_H replacement “footprint” motifs in the N1 regions were divided by the total number of IgH gene sequences with V, D, J gene assignment.

\(^g\)Mouse IgH gene sequences were previously described.

\(^h\)The mouse IgH gene sequences were downloaded from the NCBI database on May 7, 2011. The GI numbers of these sequences were included in Table S1.

DOI:10.1371/journal.pone.0057877.t001
BALB/c strains of mice are 3.17% and 5%, respectively (Fig. 2A and Table S6). Such numbers may serve as the basal levels of VH replacement products in these mice. Comparing IgH genes derived from several strains of mice, the frequencies of VH replacement products are highly elevated in IgH genes derived from different strains of autoimmune prone mice (Fig. 2A). In particular, the frequencies of VH replacement product are elevated in IgH genes derived from lupus prone NZB/NZW F1, NZM2410, MRL/lpr, and SLE1/SLE3 mice. In IgH genes derived from mice carrying the spontaneous Faslpr mutation (MRL/MpJ-Lpr/Lpr), the frequency of VH replacement products is 15.38%. In IgH genes from the Sle1/Sle3 mice, the frequency of VH replacement product is 3.16.

Table 2. List of potential VH replacement products in the test IgH sequences.

| Sequence ID | VH gene       | 3’ VH | P     | N1        | Dint | Potential footprint donor | Position |
|-------------|---------------|-------|-------|-----------|------|---------------------------|----------|
| FJ816520    | IGHV1S132     | tgtgcaaga | gggaggccct | IGHD2-14 | IGHV8-10, IGHV8-14, IGHVS2 | Y        |
| FJ150867    | IGHV1-3       | tgtgcaaga | gggaggggggcccgtagtc | IGHD1-1 | IGHV3-3, IGHV10-3, IGHV13-1 | Y        |
| FJ150854    | IGHV1S132     | tgtgcaaga | gggaggccct | IGHD2-12 | IGHV7-1 | Y                        |
| GU907018    | IGHV1-9       | tgtgcaaga | gggaggccct | IGHD1-1 | IGHV8-10, IGHV8-14, IGHVS2 | Y        |
| FJ816537    | IGHV1-74      | tgtgcaaga | gggaggccct | IGHD2-12 | IGHV3-3, IGHV10-3, IGHV13-1 | Y        |
| FJ816495    | IGHV1-47      | tgtgcaaga | gggaggccct | IGHD1-1 | IGHV3-3, IGHV10-3, IGHV13-1 | Y        |
| GU907010    | IGHV1-5       | tgtgcaaga | gggaggccct | IGHD2-1 | IGHV10-1, IGHV12-3 | Y        |
| GU907038    | IGHV1-4       | tgtgcaaga | tcggaggccct | IGHD2-3 | IGHV3-1 | Y                        |
| FJ816546    | IGHV1-4       | tgtgcaaga | gggaggccct | IGHD1-1 | IGHV8-12, IGHV11-1, IGHV12-3 | Y        |
| FJ816592    | IGHV14-1      | tgtgcaaga | gggaggccct | IGHD2-14 | IGHV2-6-7 | Y                        |
| FJ816442    | IGHV14-1      | tgtgcaaga | gggaggccct | IGHD1-1 | IGHV2-3, IGHV2-6-6 | Y        |
| FJ816522    | IGHV2-9-1     | tgtgcaaga | tcggaggccct | IGHD2-14 | IGHV7-3 | N                        |
| GU906999    | IGHV14-3      | tgtgcaaga | gggaggccct | IGHD1-1 | IGHV8-10, IGHV8-14, IGHVS2 | N        |
| GU906995    | IGHV14-3      | tgtgcaaga | gggaggccct | IGHD1-1 | IGHV8-10, IGHV8-14, IGHVS2 | N        |

The identified VH replacement footprints in the N1 regions are underlined.

*aThe relative positions of the potential donors and recipient VH genes in the identified VH replacement product were analyzed to determine if the VH replacement occurred through an upstream VH gene replacing a downstream VH gene (Y) or a downstream VH gene replacing an upstream gene (N). Only functional VH germline genes were used in this analysis.

doi:10.1371/journal.pone.0057877.t002

Table 3. Frequencies of VH replacement products in IgH genes using different families of mouse VH genes.

| VH family   | Number of IgH gene sequences | Motifs in the N1 region | Frequency of VH replacement products (%)a |
|-------------|------------------------------|-------------------------|-----------------------------------------|
| VH1/J558    | 6530                         | 314                     | 4.81                                    |
| VH2/Q52     | 665                          | 55                      | 8.27                                    |
| VH3/36-60   | 565                          | 30                      | 5.31                                    |
| VH4/X-24    | 57                           | 3                       | 5.26                                    |
| VH5/7183    | 998                          | 68                      | 6.81                                    |
| VH6/3606    | 131                          | 6                       | 4.58                                    |
| VH7/1017    | 253                          | 8                       | 3.16                                    |
| VH8/3609    | 139                          | 9                       | 6.47                                    |
| VH9/VGAM3-8 | 144                          | 11                      | 7.64                                    |
| VH10/VH10   | 127                          | 4                       | 3.15                                    |
| VH11/CP3    | 37                           | 0                       | 0                                       |
| VH12/CH27   | 43                           | 3                       | 6.98                                    |
| VH13/3609N  | 7                            | 1                       | 14.29                                   |
| VH14/SM7    | 459                          | 26                      | 5.66                                    |
| VH15/VH15A  | 4                            | 0                       | 0                                       |
| VH5-2/7183  | 56                           | 0                       | 0                                       |

*aNumber of IgH gene sequences with VH replacement “footprint” motifs in the N1 regions divided by the total number of IgH gene sequences assigned to a VH gene family.

bFunctional IgH genes using the VH5-2/7183.2 gene were analyzed for potential VH replacement footprints in the N1 regions.

CThe frequency of VH replacement products using VH2/Q52 family of VH genes is significantly higher than the overall frequency of VH replacement products in mouse IgH genes.

doi:10.1371/journal.pone.0057877.t003
products is 30%. These frequencies are significantly higher than that in the BALB/c or C57BL/6 mice (p<0.05, two tailed Chi-square test) (Fig. 2A). The elevated levels of \( \text{V}_{\text{H}} \) replacement products in autoimmune prone mice suggest that \( \text{V}_{\text{H}} \) replacement products contribute to the generation of autoantibodies. Indeed, further analyses of the IgH genes encoding different antibodies showed that the frequencies of \( \text{V}_{\text{H}} \) replacement products are 12.1% in IgH genes encoding ANA antibody and 9.34% in IgH genes encoding anti-DNA antibodies. These levels are significantly higher than those in the BALB/c or C57BL/6 mice. As a negative control, the frequency of \( \text{V}_{\text{H}} \) replacement products in IgH genes obtained from mice immunized with NP is 3.66%, which is similar to that in the C57BL/6 mice. Taken together, these results provide the first information that \( \text{V}_{\text{H}} \) replacement products are highly enriched in IgH genes derived from different strains of autoimmune prone mice and in IgH genes encoding anti-DNA and ANA autoantibodies.

Using the VHRFA program, we also analyzed the frequencies of \( \text{V}_{\text{H}} \) replacement products based on the 4- or 3-mer of \( \text{V}_{\text{H}} \) replacement footprints in IgH genes derived these diseased subcategories. Extending the assignment of \( \text{V}_{\text{H}} \) replacement products with considering the 4- and 3-mer \( \text{V}_{\text{H}} \) replacement footprints clearly increases the frequencies of \( \text{V}_{\text{H}} \) replacement products in IgH genes from all subcategories. With considering the 4-mer \( \text{V}_{\text{H}} \) replacement footprints, the frequencies of \( \text{V}_{\text{H}} \) replacement products in IgH genes derived from NZB/NZW, NZM2410, MRL/lpr, SLE1, SLE1/SLE3 and IgH genes encoding anti-DNA and ANA antibodies are significantly higher than that in the BALB/c mice (p<0.05, two tailed Chi-square test) (Fig. 2B); with considering the 3-mer \( \text{V}_{\text{H}} \) replacement footprints, the frequencies of \( \text{V}_{\text{H}} \) replacement products in IgH genes derived from NZB/NZW, NZM2410, MRL/lpr, SLE1, SLE1/SLE3, NOD/NOR and IgH genes encoding auto antibodies, anti-DNA antibodies, and ANA antibodies are significantly higher than that in the BALB/c mice (p<0.05, two tailed Chi-square test) (Fig. 2C). Taken together, these results showed that \( \text{V}_{\text{H}} \) replacement products are enriched in IgH genes derived from different strains of autoimmune prone mice and in IgH genes encoding autoantibodies.

The Identified \( \text{V}_{\text{H}} \) Replacement Footprints Preferentially Encode Charged Amino Acids

Our previous analysis of the identified \( \text{V}_{\text{H}} \) replacement products in human IgH genes showed that the \( \text{V}_{\text{H}} \) replacement footprints preferentially encode charged amino acids into the IgH CDR3 regions [31]. Here, analysis of the identified \( \text{V}_{\text{H}} \) replacement products from mouse IgH genes showed that 64% of the amino acids encoded by the identified \( \text{V}_{\text{H}} \) replacement footprints contribute charged amino acids, including K, R, D, E, N, and Q. Such frequency is significantly higher than the overall frequency of charged amino acids in the N1 regions (p<0.0001) (Fig. 3A). Moreover, the frequencies of charged amino acids, including E, K, and R, encoded by the identified \( \text{V}_{\text{H}} \) replacement footprints are significantly higher than those encoded by the N1 regions of non-\( \text{V}_{\text{H}} \) replacement products (p<0.0001) (Fig. 3B). The preferential contribution of charged amino acids by the \( \text{V}_{\text{H}} \) replacement footprints seems to be predetermined by the sequences at the 3' end of \( \text{V}_{\text{H}} \) germline genes following the cRSS sites. The frequencies of charged amino acids encoded by the 3' ends of \( \text{V}_{\text{H}} \) germline gene, including K, R, D, E, N, and Q, are significantly higher than those encoded by the D4h germline genes (p<0.0001) (Fig. 3C). In non-functional IgH genes, the identified \( \text{V}_{\text{H}} \) replacement footprints also preferentially encode charged amino acids, although the usages of different charged residues are slightly different from those in the functional \( \text{V}_{\text{H}} \) replacement footprints in human IgH genes derived from different strains of autoimmune prone mice and in IgH genes encoding anti-DNA and ANA autoantibodies.

Figure 2. Enrichment of \( \text{V}_{\text{H}} \) replacement products in IgH genes derived from different strains of autoimmune prone mice and IgH genes encoding autoantibodies. The frequencies of \( \text{V}_{\text{H}} \) replacement products in IgH genes derived from different strains of mice were analyzed using the VHRFA program based on the keyword linked to each IgH gene in the NCBI database. \( \text{V}_{\text{H}} \) replacement products were assigned based on the identification of (A) 5-mer \( \text{V}_{\text{H}} \) replacement footprints, (B) 4-mer \( \text{V}_{\text{H}} \) replacement footprints, or (C) 3-mer \( \text{V}_{\text{H}} \) replacement footprints within the \( \text{V}_{\text{H}} \)-D\(_{\text{H}} \) junctions (N1 regions). The frequencies of \( \text{V}_{\text{H}} \) replacement products in different subcategories were compared with that in the BALB/c mice. n, number of IgH sequences in each subcategory. Statistical significance was determined using a two-tailed Chi square test with Yate’s correction. p<0.05 (*) is considered...
significant and \( p<0.0001 \) (***) is considered extremely significant. The detailed sequence analysis and the identified VH replacement products with 5-mer VH replacement footprints correlating with keywords are included in Table S6. doi:10.1371/journal.pone.0057877.g002

products (Fig. 3D). Such results are consistent with previous findings that the VH replacement footprints identified in human or mouse VH replacement products preferentially encoded charged residues [31,39].

The 3-mer VH Replacement Footprints are Less Likely to Contribute Charged Amino Acids to the CDR3 Regions

VH replacement was considered as a receptor editing process to change non-functional IgH rearrangements or IgH genes encoding autoantibodies [29,41]. Finding that the 5-mer VH replacement footprints preferentially encoded charged amino acids, especially R and K residues, is contrast to the original goal of VH replacement to eliminate autoreactive IgH genes. Because charged residues within the IgH CDR3 might contribute to autoreactivity. Interestingly, when we analyzed the amino acids encoded by the identified 3-mer VH replacement footprints, the usages of charged residues, including R, K, and E, are significantly reduced; meantime, the usages of several neutral residues, including H, L, and Y, are significantly increased (Fig. 4A). These results showed that shorter VH replacement footprints are less likely to encode charged residues.

VH Replacement Products have Longer CDR3 Lengths

During VH replacement products, a short stretch of nucleotides from previously rearranged VH genes were left within the newly generated VH-DH joints [31]. Comparison of the IgH CDR3 lengths of the identified VH replacement products showed that the average CDR3 length of VH replacement products with 5-mer footprints is significantly longer than that of VH replacement products with 3-mer footprints; the average CDR3 length of VH replacement products with 3-mer footprints is significantly longer than that of the total functional IgH genes in the NCBI database \( (p<0.0001, \) unpaired \( t \) test) (Fig. 4B). These results indicate that elongation of IgH CDR3 region is one of the intrinsic features of VH replacement.

Selection of VH Replacement Footprints Encoding Positively Charged Residues in Autoantibodies

The preferential contribution of charged amino acids by VH replacement footprints is likely predetermined by the 3’ end sequences of VH germline genes. Based on the 3’ end sequences of VH germline genes, VH replacement footprints can contribute almost equal numbers of positively or negatively charged residues (Fig. 5A). Indeed, in the identified VH replacement products from IgH genes derived from BALB/c or C57BL/6 mice, the frequencies of positively and negatively charged amino acids encoded by the VH replacement products are similar (Fig. 5A). However, in the identified VH replacement products in IgH genes from autoimmune prone mice, including MRL/lpr and Sle1/Sle3 mice, the frequencies of positively charged residues encoded by the VH replacement footprints are significantly higher than that in the control mice. Meanwhile, the frequencies of negatively charged residues encoded by the VH replacement footprints are significantly lower than that in the control mice (Fig. 5A). The frequencies of negatively charged residues encoded by the identified VH replacement footprints are significantly lower in IgH genes derived from C3HBL/6/lpr mice and in IgH genes encoding anti-DNA or ANA antibodies (Fig. 5A). Detailed analysis of the functional versus non-functional IgH genes derived from MRL/lpr mice showed that the frequencies of positively charged residues encoded by the identified VH replacement footprints were elevated in functional but not in non-functional IgH genes (Fig. 5B). These results indicate that the positively charged residues encoded by VH replacement products were positively selected in these autoimmune prone mice.

The Identified VH Replacement Products are Mutated

The accumulation of VH replacement products in IgH genes derived from different strains of autoimmune prone mice and IgH genes encoding different autoantibodies suggested that VH replacement products contribute to the generation of autoantibodies in mice. Analyses of the mutation status of these identified VH replacement products showed that the enriched VH replacement products in autoimmune prone mice or IgH genes encoding anti-DNA or ANA autoantibodies are mutated (Fig. 5C), indicating that these VH replacement products are positively selected in these autoimmune prone mice.

Discussion

In the current report, we analyzed 17,179 mouse IgH gene sequences available from the NCBI database and provided a comprehensive view of the VH, DH, and JH gene usages of these mouse IgH genes. Based on the identification of the pentameric VH replacement footprints in the N1 regions, we estimated that the frequency of VH replacement products in the 11309 unique mouse IgH gene sequences with identifiable DH genes is 5.29%. Such result indicates a significant contribution of VH replacement products to the diversification of murine antibody repertoire. This result is consistent with the previously estimated frequencies of VH replacement products in human and mouse IgH genes [31,39]. It should be pointed out that such estimation is based on the identification of VH replacement footprints with a minimal length of 5 nucleotides. In comparison to human VH germline genes, many mouse VH germline genes have fewer nucleotides following the cRSS sites. Out of the 150 functional mouse VH germline genes with cRSS sites, 60 of them have only 5 nucleotides following the cRSS sites. If there is any exo-nuclease activity to remove one nucleotide at either the 3’ or the 5’ end of the VH replacement footprint during primary VH to DH recombination or VH replacement recombination, respectively, the remaining VH replacement footprints will have less than 5 nucleotides and cannot be identified from this analysis. Based on this consideration, assigning VH replacement footprints with 4 or 3 nucleotides might be a reasonable and accurate method to identify potential VH replacement products in mouse IgH genes. If we consider the 4- or 3-mer VH replacement footprints at the N1 regions to assign VH replacement products, the frequencies of VH replacement products in the mouse IgH gene sequences should be 16% or 32%, respectively.

It has been shown previously that in mice carrying two non-functional alleles of IgH genes, VH replacement occurs efficiently to generate almost normal number of B cells with a diversified repertoire [32,33]. All these functional IgH genes in this mouse are generated through VH replacement. However, only about 20% of the IgH gene sequences contain potential VH replacement footprints (>3 mer). The other 80% of IgH gene sequences have no identifiable VH replacement footprints [32,33]. This result indicates that most of the VH replacement footprints are deleted during VH replacement recombination. Thus, even if using the minimal length of VH replacement footprints with 4 or 3

PLOS ONE | www.plosone.org 7 February 2013 | Volume 8 | Issue 2 | e57877

VH Replacement Products in Mouse

Selection of VH Replacement Footprints Encoding Positively Charged Residues in Autoantibodies

The preferential contribution of charged amino acids by VH replacement footprints is likely predetermined by the 3’ end sequences of VH germline genes. Based on the 3’ end sequences of VH germline genes, VH replacement footprints can contribute almost equal numbers of positively or negatively charged residues (Fig. 5A). Indeed, in the identified VH replacement products from IgH genes derived from BALB/c or C57BL/6 mice, the frequencies of positively and negatively charged amino acids encoded by the VH replacement products are similar (Fig. 5A). However, in the identified VH replacement products in IgH genes from autoimmune prone mice, including MRL/lpr and Sle1/Sle3 mice, the frequencies of positively charged residues encoded by the VH replacement footprints are significantly higher than that in the control mice. Meanwhile, the frequencies of negatively charged residues encoded by the VH replacement footprints are significantly lower than that in the control mice (Fig. 5A). The frequencies of negatively charged residues encoded by the identified VH replacement footprints are significantly lower in IgH genes derived from C3HBL/6/lpr mice and in IgH genes encoding anti-DNA or ANA antibodies (Fig. 5A). Detailed analysis of the functional versus non-functional IgH genes derived from MRL/lpr mice showed that the frequencies of positively charged residues encoded by the identified VH replacement footprints were elevated in functional but not in non-functional IgH genes (Fig. 5B). These results indicate that the positively charged residues encoded by VH replacement products were positively selected in these autoimmune prone mice.

The Identified VH Replacement Products are Mutated

The accumulation of VH replacement products in IgH genes derived from different strains of autoimmune prone mice and IgH genes encoding different autoantibodies suggested that VH replacement products contribute to the generation of autoantibodies in mice. Analyses of the mutation status of these identified VH replacement products showed that the enriched VH replacement products in autoimmune prone mice or IgH genes encoding anti-DNA or ANA autoantibodies are mutated (Fig. 5C), indicating that these VH replacement products are positively selected in these autoimmune prone mice.

Discussion

In the current report, we analyzed 17,179 mouse IgH gene sequences available from the NCBI database and provided a comprehensive view of the VH, DH, and JH gene usages of these mouse IgH genes. Based on the identification of the pentameric VH replacement footprints in the N1 regions, we estimated that the frequency of VH replacement products in the 11309 unique mouse IgH gene sequences with identifiable DH genes is 5.29%. Such result indicates a significant contribution of VH replacement products to the diversification of murine antibody repertoire. This result is consistent with the previously estimated frequencies of VH replacement products in human and mouse IgH genes [31,39]. It should be pointed out that such estimation is based on the identification of VH replacement footprints with a minimal length of 5 nucleotides. In comparison to human VH germline genes, many mouse VH germline genes have fewer nucleotides following the cRSS sites. Out of the 150 functional mouse VH germline genes with cRSS sites, 60 of them have only 5 nucleotides following the cRSS sites. If there is any exo-nuclease activity to remove one nucleotide at either the 3’ or the 5’ end of the VH replacement footprint during primary VH to DH recombination or VH replacement recombination, respectively, the remaining VH replacement footprints will have less than 5 nucleotides and cannot be identified from this analysis. Based on this consideration, assigning VH replacement footprints with 4 or 3 nucleotides might be a reasonable and accurate method to identify potential VH replacement products in mouse IgH genes. If we consider the 4- or 3-mer VH replacement footprints at the N1 regions to assign VH replacement products, the frequencies of VH replacement products in the mouse IgH gene sequences should be 16% or 32%, respectively.

It has been shown previously that in mice carrying two non-functional alleles of IgH genes, VH replacement occurs efficiently to generate almost normal number of B cells with a diversified repertoire [32,33]. All these functional IgH genes in this mouse are generated through VH replacement. However, only about 20% of the IgH gene sequences contain potential VH replacement footprints (>3 mer). The other 80% of IgH gene sequences have no identifiable VH replacement footprints [32,33]. This result indicates that most of the VH replacement footprints are deleted during VH replacement recombination. Thus, even if using the minimal length of VH replacement footprints with 4 or 3
nucleotides, we may still underestimate the actual frequency of VH replacement products in the murine IgH repertoire. Theoretically, 66.7% of the IgH rearrangements generated during V(D)J recombination will be out of reading frame and cannot produce functional IgH proteins; about 44% of the pro B cells undergoing V(D)J recombination should carry non-functional rearrangements.

Figure 3. VH replacement footprints preferentially contribute charged amino acids to the CDR3 regions. (A) The frequencies of charged amino acids encoded by the identified pentameric VH replacement footprints or the N1 regions of non-VH replacement products were compared. Detailed amino acid sequences of the IgH CDR3 regions are listed in Table S6. (B) The frequencies of individual amino acid encoded by the identified VH replacement footprints or the N1 regions of non-VH replacement products. (C) The frequencies of individual amino acid encoded by the 3' end of VH germline genes and Dk regions were compared. (D) Usages of different amino acids encoded by the identified VH replacement footprints in functional VH replacement products and non-functional VH replacement products. n, amino acids encoded by the VH gene 3' ends or DH regions. Statistical significance was determined using a two-tailed Chi square test with Yate's correction. n, number of amino acid residues encoded by indicated sequences. p<0.05 (*) is considered significant and p<0.0001 (**) is considered extremely significant.

doi:10.1371/journal.pone.0057877.g003
on both IgH alleles. If VH replacement can efficiently rescue these pro B cells, at least 44% of the expressed IgH genes should be generated by VH replacement.

We should also point out that this sequence analysis based approach in identification of VH replacement footprints may have false positive calls. Theoretically, there are no VH replacement footprints in the N2 regions. In some of the IgH sequences, we identified similar 3, 4, or 5 mer VH replacement footprint motifs in the N2 regions, although the frequencies of such motifs in the N2 regions are significantly lower than those in the N1 regions. The presence of such VH replacement footprint motifs in the N2 regions could be due to random nucleotide addition during V(D)J recombination. In this regard, a low frequency of identified footprints might be false positive.

If we use the 5-mer VH replacement footprints to assign VH replacement products, the frequencies of VH replacement products in IgH genes derived from BALB/C or C57BL/6 mice are about 5% or 3.2%, respectively, which may represent the basal level of VH replacement product in these two strains of mice. Interestingly, the frequencies of VH replacement products are significantly elevated in IgH genes derived from different strains of autoimmune prone mice, including MRL/Lpr and Sle1/Sle3 mice. It has been well demonstrated that these mice spontaneously produce anti-DNA or anti-ANA antibodies and develop lupus like symptom [42–49]. Indeed, VH replacement products are significantly elevated in IgH genes encoding anti-DNA antibodies or ANA autoantibodies derived from mice with lupus glomerular nephritis. These results suggested a potential contribution of VH replacement products to the generation of autoantibodies. When we consider the 4- or 3-mer VH replacement footprints to assign VH replacement products, the frequencies of VH replacement products are elevated in all the sub-categories of IgH genes. Nevertheless, the frequencies of VH replacement products in IgH genes derived from different strains of autoimmune prone mice and IgH genes encoding anti-DNA and ANA antibodies are significantly higher than that in the BALB/c mice.

Due to the location of the cRSS, VH replacement will leave a short stretch of VH replacement footprints to elongate the IgH CDR3 region [31,41]. Strikingly, the identified pentameric VH replacement footprints preferentially encode charged amino acids in the newly formed CDR3 regions. Such features are commonly found in VH replacement footprints preferentially encoded amino acids in the newly formed CDR3 regions. Such features are commonly found in VH replacement footprints. These results further support the notion that VH replacement footprints preferentially encode charged amino acids in the newly formed CDR3 regions.

Figure 4. Comparison of the amino acids encoded by VH replacement footprints and the IgH CDR3 lengths of VH replacement products. (A) The usages of different amino acids encoded by VH replacement footprints with 5, 4, or 3 nucleotides were compared. n, number of amino acid residues encoded by the identified VH replacement footprints with different lengths. Statistical significance was determined using a two-tailed Chi square test with Yate's correction. p<0.05 (*) is considered significant and p<0.0001 (**) is considered extremely significant. (B) Comparison of the IgH CDR3 lengths of VH replacement products containing the 5-mer or the 3-mer VH replacement products with the CDR3 length of the total functional IgH genes. n, number of IgH sequences or VH replacement products with 3- or 5-mer VH replacement footprints. Statistical significance was determined using unpaired t test. p<0.05 (*) is considered significant and p<0.0001 (**) is considered extremely significant.

doi:10.1371/journal.pone.0057877.g004
are frequently encoding autoantibodies or anti-viral antibodies [51]. Here, our results showed that the frequencies of V_H replacement products are significantly elevated in IgH genes encoding autoantibodies. Theoretically, the VH replacement footprints can encode either positively or negatively charged residues. Analysis of the amino acids encoded by the identified VH replacement products from different strains of autoimmune prone mice and IgH genes encoding autoantibodies showed that the frequencies of positively charged residues encoded by VH replacement footprints are significantly elevated; while the frequencies of negatively charged residues encoded by VH replacement footprints are significantly reduced. Previous studies have shown that positively charged residue like Arg within the IgH CDR3 is critical for DNA binding [52–54]. These results suggested that the identified VH replacement products from autoimmune prone mice have been positively selected. Such notion is also supported by the accumulated mutations in these identified VH replacement products.

V_H replacement was originally recognized as a receptor editing process to change either non-functional IgH genes or IgH genes encoding autoreactive antibodies [20,55]. The enrichment of VH replacement products in IgH genes from different strains of autoimmune prone mice and in IgH genes encoding autoantibodies are surprising findings from this study. Currently, it is not clear why VH replacement products are accumulated in autoimmune prone mice. Like any recombination process, VH replacement is a random process that can generate non-functional IgH genes or IgH genes encoding autoreactive antibodies.

Previous studies have shown that VH replacement products generated through replacing the knocked-in anti-DNA IgH genes can produce high affinity anti-DNA antibodies during chronic graft-versus-host (cGVH) response [56]. Theoretically, after VH replacement recombination, the newly generated IgH genes should be subjected to strict negative selection again to eliminate B cells expressing autoreactive BCRs. The observed accumulation of VH replacement products in autoimmune prone mice could be due to the defective negative selection processes in these mice. In autoimmune prone mice, the newly generated VH replacement products encoding autoreactive antibodies cannot be efficiently eliminated, but are rather positively selected and contribute to the generation of autoantibodies. To this extend, the different strains of autoimmune prone mice will be excellent experimental models to dissect how the VH replacement products are selected and enriched during early B cell development.

Our analyses of the amino acid residues encoded by the identified VH replacement footprints also uncovered an interesting finding that short VH replacement footprints, especially the 3-mer footprints, encode less charged residues. These results suggested that if the VH replacement footprints were trimmed down to 3-mer during primary or secondary recombination, they will be less likely to contribute charged amino acids into the IgH CDR3 regions. Given the fact that 33.55% of IgH genes contain 3-mer VH replacement footprints at their N1 regions, it is reasonable to conclude that the majority of these VH replacement products successfully edited the IgH genes without introducing of extra charged residues into the newly formed CDR3 regions. The
observed accumulation of V<sub>H</sub> replacement products based on the identification of 5-mer footprints in the N1 regions in IgH genes derived from autoimmune prone mice may represent the failed V<sub>H</sub> replacement attempts either due to defects in negative selection or defects in trimming down the V<sub>H</sub> replacement footprints during primary or secondary recombination. Such findings raised several interesting questions that require further studies.

In conclusion, analysis of large number of mouse IgH gene sequences from the NCBI database provides a comprehensive view of the IgH repertoire of the available mouse IgH genes in the NCBI database and reveals a significant contribution of V<sub>H</sub> view of the IgH repertoire of the available mouse IgH genes in the primary or secondary recombination. Such findings raised several questions that require further studies.

Materials and Methods

Mouse IgH Sequences

Entrez IDs of mouse IgH sequences were provided by Igblast (http://www.ncbi.nlm.nih.gov/projects/igblast/) on May 07, 2011, which were used to download GenBank records of the sequences from NCBI. There were total 17,179 mouse IgH gene sequences retrieved at that time. The IDs of these IgH genes and their V<sub>H</sub>, D<sub>H</sub>, and J<sub>H</sub> gene assignments are included in Table S1. After assignment of the potential germline V<sub>H</sub>, D<sub>H</sub>, J<sub>H</sub> genes, clonally redundant sequences were stripped out based on their identical CDR3 regions. The resulting 11,308 unique sequences were further analyzed. Clonally related sequences with mutations within the N1 or N2 region sequences.

The V<sub>H</sub>RFA Program

We developed a Java-based V<sub>H</sub>RFA program to incorporate assignments of the V<sub>H</sub>, D<sub>H</sub>, and J<sub>H</sub> gene segments using the V-QUEST program (http://www.imgt.org/IMGT_vquest), identification of V<sub>H</sub> replacement footprints with different lengths, analysis of amino acids encoded by the identified V<sub>H</sub> replacement footprints, calculation of the amino acid usage encoded by the identified V<sub>H</sub> replacement footprints, and correlation of the identified V<sub>H</sub> replacement products with different keywords and publications associated with the sequences in the NCBI database.

V<sub>H</sub>, D<sub>H</sub>, and J<sub>H</sub> Germline Gene Assignment

Mouse IgH sequences in the GenBank format were converted to FASTA format and submitted to IMGT/V-QUEST (http://www.imgt.org/IMGT_vquest/share/sequences/) for assign potential germline V<sub>H</sub>, D<sub>H</sub>, J<sub>H</sub> genes, allowing 1 mutation at the 3' end of V<sub>H</sub> genes and at the 5' end of J<sub>H</sub> genes. All the IgH gene sequences were analyzed in batches containing 50 sequences each batch and the results were downloaded to a local computer as Excel files. These processes were conducted using the V<sub>H</sub>RFA program.

Identification of V<sub>H</sub> Replacement Footprint

All the rest steps were conducted on a local computer by the V<sub>H</sub>RFA program. First, a library file was generated, which contains all the potential V<sub>H</sub> replacement footprints derived from functional V<sub>H</sub> germline reference genes from the IMGT database (Table S3). Basically, the 3' end segments following the cRSS sites from functional mouse V<sub>H</sub> genes were sliced into different groups with 3, 4, 5, 6, 7, 8, 9, 10, and 11 nucleotides in length (Table S4). The V<sub>H</sub>RFA program will use this library to search the N1 (V<sub>H</sub>-D<sub>H</sub> junction [N1] or D<sub>H</sub>-J<sub>H</sub> junction (N2, as negative control) regions of the IgH genes to identify matched footprint motifs. For each IgH gene, the V<sub>H</sub>RFA program started by searching the longest footprint motif (11 mer) from the 5' to 3' of the DNA sequences and then goes to search footprints with one nucleotide shorter. The identified footprints were listed if it does not overlap with any previously identified footprint within this region. For examples, the end results of footprint analyses of with specified 5 mer included all the footprints with 5, 6, 7, 8, 9, 10, and 11 mer from the V<sub>H</sub> replacement footprint library. The end result was exported as a CVS file that contains the gene ID, functionality, V<sub>H</sub>, D<sub>H</sub>, J<sub>H</sub> gene assignment, V<sub>H</sub> replacement footprint in N1 (N1 signatures) or N2 (N2 signatures), together with other information from the original Excel file provided by the IMGT V-QUEST program. The identified footprints were shown in parenthesis within the N1 or N2 region sequences.

Analysis of the Amino Acid Encoded by V<sub>H</sub> Replacement Footprints, Keyword and Publication Linked to Each Gene, and Mutation

After identification of the V<sub>H</sub> replacement footprints within the N1 regions, the V<sub>H</sub>RFA program further analyzed the amino acids encoded by the V<sub>H</sub> replacement footprints and the usages of different amino acid. Each result was exported as an individual Excel file. The V<sub>H</sub>RFA program can also analyze the original GenBank file to correlate the keywords and publication information with each IgH gene sequence. Basically, the V<sub>H</sub>RFA program correlates the GenBank file for keywords in the KEYWORDS and FEATURES sections of each entry sequence and output the keyword list in correlation with the sequence IDs, VDJ assignments, N1 footprints, and N2 footprints. Through this analysis, we can determine the distribution of V<sub>H</sub> replacement products in different diseases.

For mutation analysis, the V<sub>H</sub>RFA program only calculated the mutation rate of IgH V<sub>H</sub> genes with >80% similarities to the assigned germline V<sub>H</sub> genes.

Statistical Analysis

Statistical significance was determined by using either the two tailed Chi-square test with Yates' correction or non paired student t test. Significant difference was determined if the p value <0.05.

Supporting Information

Table S1 Analyses of mouse IgH genes and identification of VH replacement products. (XLSX)

Table S2 Number of sequences from each publication. (XLSX)

Table S3 Mouse VH genes containing the TACTGTG cRSS. (DOCX)

Table S4 Potential mouse VH replacement footprint motifs with different length. (DOCX)

Table S5 Identification of 4-mer VH replacement footprint motifs in mouse IgH sequences. (DOCX)
Table S6 Identification of V_{H} replacement products in IgH genes correlating with different keywords.

(ODOCX)

References

1. Schatz DG, Baltimore D (1983) Stable expression of immunoglobulin gene V(D)J recombinease activity by gene transfer into 3T3 fibroblasts. Cell 33: 107–115.
2. Schatz DG, Oettinger MA, Baltimore D (1989) The VDJ recombination activating gene, RAG-1. Cell 59: 1035–1048.
3. Oettinger MA, Schatz DG, Gorst C, Baltimore D (1990) RAG-1 and RAG-2, adjacent genes that synergistically activate VDJ recombination. Science 240: 1517–1523.
4. Bassin CH, Swat W, Alt FW (2002) The mechanism and regulation of chromosomal VDJ recombination. Cell 109: 845–855.
5. Jung D, Gilbourcins C, Mozoleski R, Alt FW (2006) Mechanism and control of VDJ recombination at the immunoglobulin heavy chain locus. Annu Rev Immunol 24: 541–570.
6. Lewis SM (1994) The mechanism of VDJ joining: lessons from molecular, immunological, and comparative analysis. Annu Rev Immunol 12: 27–50.
7. Schatz DG, Swanson PC (2011) VDJ recombination: mechanisms of initiation. Annu Rev Genet 45: 167–202.
8. Ramden DA, Baertz K, Wu GE (1994) Conservation of sequence in recombination signal sequence spacers. Nucleic Acids Res 22: 1765–1768.
9. Swanson PC, Desiderio S (1998) VDJ recombination signal recognition: distinct, overlapping DNA-protein contacts in complexes containing RAG1 with and without RAG2. Immunity 9: 115–123.
10. Karasuyama H, Kudo A, Melchers F (1990) The proteins encoded by the Vpre-B and lambda 3 pre-B cell-specific genes can associate with each other and with mu heavy chain. J Exp Med 172: 899–972.
11. Karasuyama H, Rolink A, Melchers F (1993) A complex of glycoproteins is associated with Vpre-B/lambda 3 surrogate light chain on the surface of mu heavy-chain-negative early precursor B cell lines. J Exp Med 178: 469–478.
12. Karasuyama H, Rolink A, Shinak B, Young F, Alt FW, et al. (1994) The expression of Vpre-B/lambda 3 surrogate light chain in early bone marrow (pre-B) cells stimulates normal and B cell-deficient mutant mice. Cell 77: 133–143.
13. Lassoued K, Nunez CA, Bills L, Kabiagawa H, Monteiro RC, et al. (1993) Expression of surrogate light chain receptors is restricted to a late stage in pre-B cell differentiation. Cell 73: 73–86.
14. Burrows PD, Stenad RP, Wang YH, Lassoued K, Zhang Y, et al. (2002) The transient expression of pre-B cell receptors governs B cell development. Semin Immunol 14: 343–349.
15. Karasuyama H, Nakamura T, Nagata K, Kurumochi T, Kitamura F, et al. (1997) The roles of pre-B cell receptor in early B cell development and its signal transduction. Immunol Cell Biol 75: 209–216.
16. Jung D, Alt FW (2004) Unraveling V(D)J recombination: insights into gene rearrangement. Annu Rev Genet 45: 167–202.
17. Jung D, Alt FW (2004) Unraveling V(D)J recombination: insights into gene rearrangement. Annu Rev Immunol 12: 27–50.
18. Tiegs SL, Russell DM, Nemazee D (1993) Receptor editing in self-reactive bone marrow B cells. J Exp Med 177: 1009–1020.
19. Rogosch T, Kerzel S, Sikula L, Gentil K, Liebetruth M, et al. (2010) Plasma cells of the quasimonoclonal mouse. J Immunol 184: 4947–4953.
20. Morel L (2010) Genetics of SLE: evidence from mouse models. Nat Rev Rheumatol 6: 348–357.
21. Radic MZ, Mackle J, Erikson J, Mol C, Anderson WF, et al. (1993) Residues that mediate DNA binding of autoimmune antibodies. J Immunol 150: 4966–4972.
22. Zhang Z, Wang YH, Zemlin M, Findley H, Bridges SL, et al. (2003) Contribution of receptor editing to the antibody repertoire. Science 291: 1541–1544.
23. Rech M, Gehramm P, Petrac E, Wiese P (1986) A novel V_{H} to V_{J}DJ_{H} joining mechanism in heavy-chain-negative (null) pre-B cells results in heavy-chain production. Nature 322: 840–842.
24. Kleinfield R, Hardy RR, Tarlinton D, Dang J, Herzenberg LA, et al. (1991) Recombination between an expressed immunoglobulin heavy-chain gene and a germline variable gene segment in a Ly 1+ B-cell lymphoma. Nature 352: 843–846.
25. Chen C, Nagy Z, Prak EL, Weigert M (1995) Immunoglobulin heavy chain gene replacement: a mechanism of receptor editing. Immunity 3: 747–753.
26. Cassalino M, Weng J, Wahl M (1997) V_{H} gene replacement in hyperexpressed B cells of the quasimonoclonal mouse. J Immunol 159: 5795–5801.
27. Zhang Z, Zemlin M, Wang Y-H, Munfs D, Hoye LE, et al. (2003) Contribution of V_{H} gene replacement to the primary B cell repertoire. Immunity 19: 21–31.
28. Yazici ZA, Behrendt M, Goodfield M, Partridge LJ, Lindsey NJ (1998) Does the contribution of receptor editing to the antibody repertoire. Science 291: 1541–1544.
29. Sun Y, Liu Z, Li Z, Lian Z, Zhao Y (2012) Phylogenetic conservation of the 3’ cryptic recombination signal sequence (cRSS) in the V_{H} genes of jawed vertebrates. Front Immunol 3: 392.
30. O’Keefe TL, Datta SK, Imanishi-Kari T (1992) Cationic residues in pathogenic anti-DNA idiotypes from NZB x SWR model of lupus nephritis. Ann N Y Acad Sci 675: 47–58.
31. Datta SK, Gavalcin J (1986) Origins of pathogenic anti-DNA idiotypes in the NZB x SWR model of lupus nephritis. Ann N Y Acad Sci 675: 47–58.
32. Cai L, Zhao J, Huang P, Zhang C, Li C, et al. (2002) Contribution of receptor editing to the antibody repertoire. Science 291: 1541–1544.
33. Rech M, Gehramm P, Petrac E, Wiese P (1986) A novel V_{H} to V_{J}DJ_{H} joining mechanism in heavy-chain-negative (null) pre-B cells results in heavy-chain production. Nature 322: 840–842.
34. Kleinfield R, Hardy RR, Tarlinton D, Dang J, Herzenberg LA, et al. (1991) Recombination between an expressed immunoglobulin heavy-chain gene and a germline variable gene segment in a Ly 1+ B-cell lymphoma. Nature 352: 843–846.
35. Chen C, Nagy Z, Prak EL, Weigert M (1995) Immunoglobulin heavy chain gene replacement: a mechanism of receptor editing. Immunity 3: 747–753.