

RESEARCH ARTICLE

# Guinea pig immunoglobulin VH and VL naïve repertoire analysis

Shun Matsuzawa<sup>1,2</sup>, Masaharu Isobe<sup>1</sup>, Nobuyuki Kurosawa<sup>1\*</sup>

**1** Laboratory of Molecular and Cellular Biology, Graduate School of Science and Engineering for Research, University of Toyama, Toyama-shi, Toyama, Japan, **2** Medical & Biological Laboratories Co., Ltd., Ina-shi, Nagano, Japan

\* [kurosawa@eng.u-toyama.ac.jp](mailto:kurosawa@eng.u-toyama.ac.jp)



## Abstract

The guinea pig has been used as a model to study various human infectious diseases because of its similarity to humans regarding symptoms and immune response, but little is known about the humoral immune response. To better understand the mechanism underlying the generation of the antibody repertoire in guinea pigs, we performed deep sequencing of full-length immunoglobulin variable chains from naïve B and plasma cells. We gathered and analyzed nearly 16,000 full-length  $V_H$ ,  $V_K$  and  $V_\lambda$  genes and analyzed V and J gene segment usage profiles and mutation statuses by annotating recently reported genome data of guinea pig immunoglobulin genes. We found that approximately 70% of heavy, 73% of kappa and 81% of lambda functional germline V gene segments are integrated into the actual V(D)J recombination events. We also found preferential use of a particular V gene segment and accumulated mutation in CDRs 1 and 2 in antigen-specific plasma cells. Our study represents the first attempt to characterize sequence diversity in the expressed guinea pig antibody repertoire and provides significant insight into antibody repertoire generation and Ig-based immunity of guinea pigs.

## OPEN ACCESS

**Citation:** Matsuzawa S, Isobe M, Kurosawa N (2018) Guinea pig immunoglobulin VH and VL naïve repertoire analysis. PLoS ONE 13(12): e0208977. <https://doi.org/10.1371/journal.pone.0208977>

**Editor:** Sebastian D. Fugmann, Chang Gung University, TAIWAN

**Received:** September 26, 2018

**Accepted:** November 28, 2018

**Published:** December 13, 2018

**Copyright:** © 2018 Matsuzawa et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the manuscript and its Supporting Information files.

**Funding:** This research was supported in part by grants from the Hokuriku Industrial Advancement Center (HIAC) (H27HIAC122) for Kurosawa, N., and the Basic Science and Platform Technology Program for Innovative Biological Medicine (15657579), and the Emerging/Re-emerging Infectious Diseases Project of Japan (16768665) from Japan Agency for Medical and Development (AMED) for Isobe, M., and MEXT: Regional

## Introduction

Immunoglobulins (Igs), a major component of the adaptive immune response, are heterodimeric proteins composed of two heavy (H) and two light (L) chains. Each Ig chain consists of variable (V) domains that have great genetic diversity and specifically bind antigen epitopes [1]. V gene diversity is generated by various strategies that differ in different species [2]. Primary V gene diversity in mice and humans is generated by V(D)J recombination, which rearranges multiple germline V, diversity (D) and junction (J) gene segments to the functional V gene [3]. During V(D)J recombination, junctional diversity is generated by recombination-associated nucleotide deletion as well as addition of a series of palindromic, 'P' nucleotides followed by the addition of random 'N' nucleotides [1]. The complementarity-determining region 3 of  $V_H$  (CDRH3) is the most diverse region because it is composed of V, D and J gene segments [3]. After V(D)J recombination, further V gene sequence diversity is produced through antigen-driven somatic hypermutation (SHM) [4]. In cattle and sheep, V(D)J

Innovation Strategy Support Program for Matsuzawa, S. and Isobe, M. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Medical & Biological Laboratories Co., Ltd. provided support in the form of salaries for the author [SM], but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of the author is articulated in the 'author contributions' section.

**Competing interests:** We have the following interests. Shun Matsuzawa is employed by Medical & Biological Laboratories Co., Ltd. There are no patents, products in development or marketed products to declare. This does not alter our adherence to all the PLOS ONE policies on sharing data and materials, as detailed online in the guide for authors.

combinatorial diversity is limited because only a few V genes rearrange [5, 6]. Therefore, they use SHM to generate their primary V gene diversity. After V(D)J recombination, cattle and sheep B cells migrate to ileal Peyer's patches for additional diversification of the limited V(D)J combination by SHM [7, 8]. Cattle have exceptionally long CDRH3s, which have multiple cysteine residues mostly produced by SHM. The ultralong CDRH3s fold into a plethora of mini-domains and create additional V gene diversity [9]. Furthermore, interior codons of D gene segments of the ultralong CDRH3s are deleted with high frequency, and their length and cysteine positions are altered to further diversify their sequences [10]. In chickens and rabbits, the combinatorial diversity is also limited. Therefore, they use gene conversion of homologous sequence from number of 5' variable genes, which can be rearranged, and can donate diverse genetic fragments into the rearranged V domains [11–13].

The guinea pig (*Cavia porcellus*) is a species of rodent belonging to the family Caviidae. This animal shares similarity with humans with regard to hormonal and immunologic responses and has been used as an experimental model of various human infectious diseases, such as tuberculosis [14, 15], Zika virus infection [16] and Cytomegalovirus infection [17]. Guinea pigs have also been immunized for polyclonal antibody (pAb) production because they are phylogenetically distant from mice and rats. [18–22]. Recently, we developed a single cell-based monoclonal antibody (mAb) development method that enables guinea pig as a host animal for mAb production [23, 24]. Despite the importance of guinea pigs as laboratory and immunized animals, their immune system has not been well understood. Recently, Guo *et al.* characterized guinea pig  $Ig_H$  and  $Ig_L$  loci and found that the germline genetic components of the guinea pig antibody repertoire are large compared to those of humans, mice, and other vertebrates. Data from this study show that the  $Ig_H$  locus is composed of 507  $V_H$  gene segments, including 94 potentially functional genes and 413 pseudogenes, 41  $D_H$  gene segments, and six  $J_H$  gene segments; the  $Ig_K$  locus is composed of 349  $V_K$  gene segments (111 potentially functional genes and 238 pseudogenes), three  $J_K$  gene segments; and  $Ig_\lambda$  is composed of 142  $V_\lambda$  gene segments (58 potentially functional genes and 84 pseudogenes) and 11  $J_\lambda$  gene segments. The authors suggested that pseudogenes may serve as donor pools for gene conversion and contribute to Ig diversity in guinea pigs [25]. However, neither the gene usage in the guinea pig Ig repertoire nor the extent of nucleotide diversity from the germline sequence is known.

To address these questions, we amplified full-length V transcripts from naïve B cells obtained from spleens and lymph nodes from 2 guinea pigs and sequenced them using the 454 GS-FLX Plus next-generation sequencer (NGS). Our results indicate that the primary V gene diversity of guinea pigs is mainly produced via V(D)J recombination. Most of the germline V and J gene segments of the  $V_H$ ,  $V_K$  and  $V_\lambda$  loci annotated as functional are integrated into the actual recombination events; therefore, combinatorial diversity is vast. We also compared V gene usage and mutation distribution between naïve B and antigen-specific plasma cells (PCs) and found that guinea pigs generate antibody diversity by both broader usage of V gene segments and an increased number of somatic mutations. Our study represents the first attempt to characterize sequence diversity in the expressed guinea pig antibody repertoire.

## Materials and methods

### Naïve guinea pig V gene library preparation and high-throughput sequencing

All animal studies were approved by the Committee for Laboratory Animal Care and Use at University of Toyama, and the experiments were carried out in accordance with approved guidelines (Protocol Number: A2016ENG-3). Guinea pigs purchased from Japan SLC, Inc. were immunized four times intramuscularly at the tail base with a 200  $\mu$ L of 50:50 water-in-oil

TiterMax Gold adjuvant emulsion containing 100 µg of antigens. Iliac lymph nodes and spleens were surgically removed from two individual naïve female Hartley guinea pigs under euthanasia with a barbiturate overdose. After the removal, the cells were dispersed, and B cells were stained with anti-guinea pig IgG (H + L) labeled with DyLight 650 (Abcam) followed by sorting with J-SAN cell sorter (BayBioscience). Approximately  $2 \times 10^5$  B cells were subjected to 5'-end homopolymer-tailed cDNA synthesis to amplify full-length  $V_H$  ( $\mu$ ),  $V_\kappa$  and  $V_\lambda$  genes, as described previously [23, 24]. The first round of PCR was performed with a universal 5' RACE primer (Nhe polyC S) and a mixture of reverse primers specific for Ig constant region 1 (Gpig IgM AS1, Gpig IgK AS1 and Gpig IgL AS1). PCR was performed using PrimeStar DNA polymerase in  $1 \times$  PrimeStar GC buffer with the BIO-RAD MyCycler (35 cycles with denaturation at 98°C for 10 s, annealing and strand elongation at 55°C for 15 s and a final extension at 72°C for 60 s). The resulting PCR mixtures were diluted 1:8 with water, and 12 µL of each was used for the second round of PCR. The second round of PCR was performed with a forward primer designed for identifying each sample origin (Nhe-Eco47-S1, Nhe-Eco47-L1, Nhe-Eco47-S2 and Nhe-Eco47-L2) and a respective nested reverse primer (Gpig IgM AS2 FLX-Rv, Gpig IgK AS2 FLX-Rv, and Gpig IgL AS2 FLX-Rv). A unique 10-nt multiplex identifier (MID) for each sample origin was included in each forward primer between the adapter and the forward primer. The primers used are summarized in [S1 Table](#). The PCR products were purified using a QIAquick PCR purification kit (Qiagen) and sequenced with 454 GS-FLX Plus NGS (Roche).

### Guinea pig IgBLAST database preparation

The potentially functional germline V and J gene segments of guinea pig  $Ig_H$ ,  $Ig_\kappa$  and  $Ig_\lambda$  annotated by Guo *et al.* [25] were obtained from the Broad Institute, which conducted genome sequencing and assembly (cavPor3, 6.79x coverage, Jul 2008). To achieve an accurate assignment of the V gene segment, we retrieved the V leader sequence and joined it in silico to its V gene segment to produce a custom guinea pig V database. The sequences are presented in the [S1 File](#).

### Data preprocessing

All NGS data were first processed using the sequence quality of the 454 pipeline and filtered by nucleotide length >400 bp. The libraries were aligned to germline V gene segments in the custom databases described above, and aligned amino acid (aa) sequences containing stop codons were removed. The selected in-frame libraries were separated according to the MIDs for each sample origin. The preprocessed read sequences are presented in the [S2 File](#).

### Ig gene analysis

The preprocessed reads were analyzed as follows:

- i. V and J gene segment combination frequency: Germline V and J gene segment assignments were derived from IgBLAST alignments against the custom IgBLAST database described above. V-J gene segment combination percent frequencies of all possible combinations were calculated. (ii) CDR3 length distribution: CDR3s were defined according to International ImMunoGeneTics (IMGT) unique numbering [26, 27]. Relative percent frequencies of CDR3 aa length of  $V_H$ ,  $V_\kappa$  and  $V_\lambda$  were calculated. (iii) Mutation frequency in the V gene segment: The number of nucleotide base substitutions, insertions and deletions from the aligned germline V gene segment sequence was counted to determine the number of mutations. Percent mutation frequencies were calculated by comparing the fraction of mutations

in a given nucleotide to the fraction of that nucleotide in the germline sequences. Phylogenetic trees were compiled using Clustal X.

## Guinea pig monoclonal antibodies

The nucleotide sequences of guinea pig mAbs against CHK2 peptide, p53 peptide, ROR $\gamma$ T peptide and human insulin prepared from antigen-specific PCs were used as described previously [23, 24].

## Results

### Data sets

To analyze antibody repertoires, we analyzed Ig transcripts from sorted B cell populations to obtain an overview of antibody repertoires in individual tissue. Full-length V gene transcripts of Ig<sub>H</sub>, Ig<sub>κ</sub> and Ig<sub>λ</sub> chains were amplified from B cell libraries by 5' RACE PCR with universal forward primer containing unique 10-nt MID and reverse primers specific for the guinea pig Ig constant region of Ig<sub>H(μ)</sub>, Ig<sub>κ</sub> and Ig<sub>λ</sub>. The amplified DNAs were sequenced by 454 pyrosequencing. The raw-read sequences were sorted into Ig<sub>H</sub>, Ig<sub>κ</sub> and Ig<sub>λ</sub> groups and checked for quality. Useful 42,771 Ig<sub>H</sub>, 60,308 Ig<sub>κ</sub> and 45,643 Ig<sub>λ</sub> sequencing reads were obtained (Table 1). Each group was then subjected to bioinformatics analysis as follows: (i) Length cutoff: sequences shorter than 400 nucleotide were removed; (ii) Translation: sequences containing stop codons were removed, and in-frame sequence containing homologies to conserved framework regions were collected; (iii) Custom database arrangement: Sequences containing productive V(D)J junctions were collected; (iv) MID sorting: Antibody repertoires between independent organs were analyzed (Fig 1). When we assigned each of the (i)- and (ii)-passed sequences to the best-scoring guinea pig germline V gene segment reference sequence, we found that 93.3% of transcribed V<sub>H</sub> clone sequences, 83.9% of transcribed V<sub>κ</sub> clone sequences and 95.9% of transcribed V<sub>λ</sub> clone sequences showed more than 95% nucleotide identity with their germline sequences. These results suggest that gene conversion events are rare in rearranged guinea pig Ig diversification.

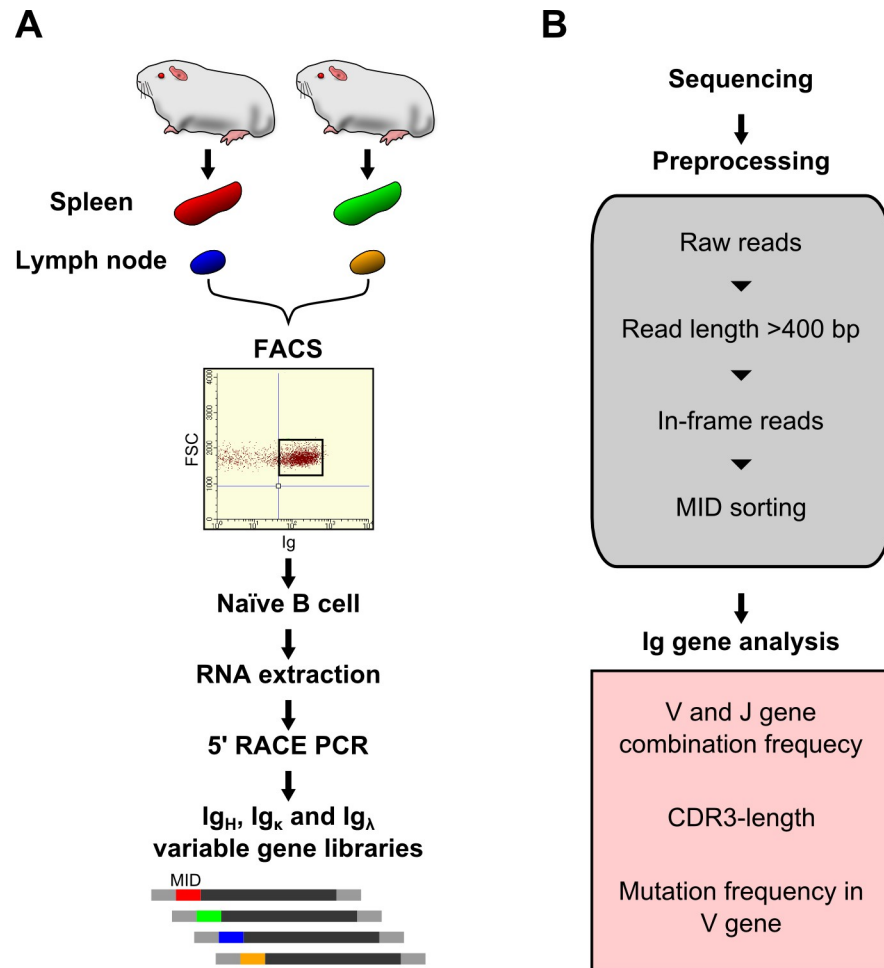
### V and J gene segment usage in naïve guinea pig

We next analyzed V and J gene segment usage of transcribed V<sub>H</sub>, V<sub>κ</sub> and V<sub>λ</sub> genes in naïve B cells. The preprocessed reads were analyzed using a custom V gene database to assign the best-scoring germline V and J gene segment reference sequence, and the overall view of antibody repertoires in individual guinea pigs was visualized on 2D meshes in which the x-axis represented V gene segments, and the y-axis represented J gene segments (Fig 2).

**Table 1. Sequencing read counts from two individual naïve guinea pigs.**

		V <sub>H</sub>	V <sub>κ</sub>	V <sub>λ</sub>
<b>Raw reads</b>		42771	60308	45643
<b>In-frame reads</b>		5675	9360	1230
<b>V identity &gt;95%</b>	<b>Total</b>	5296	7852	1180
	<b>Spleen 1</b>	2388	1864	404
	<b>Lymph node 1</b>	526	1694	151
	<b>Spleen 2</b>	1366	2270	313
	<b>Lymph node 2</b>	1016	2024	312

<https://doi.org/10.1371/journal.pone.0208977.t001>

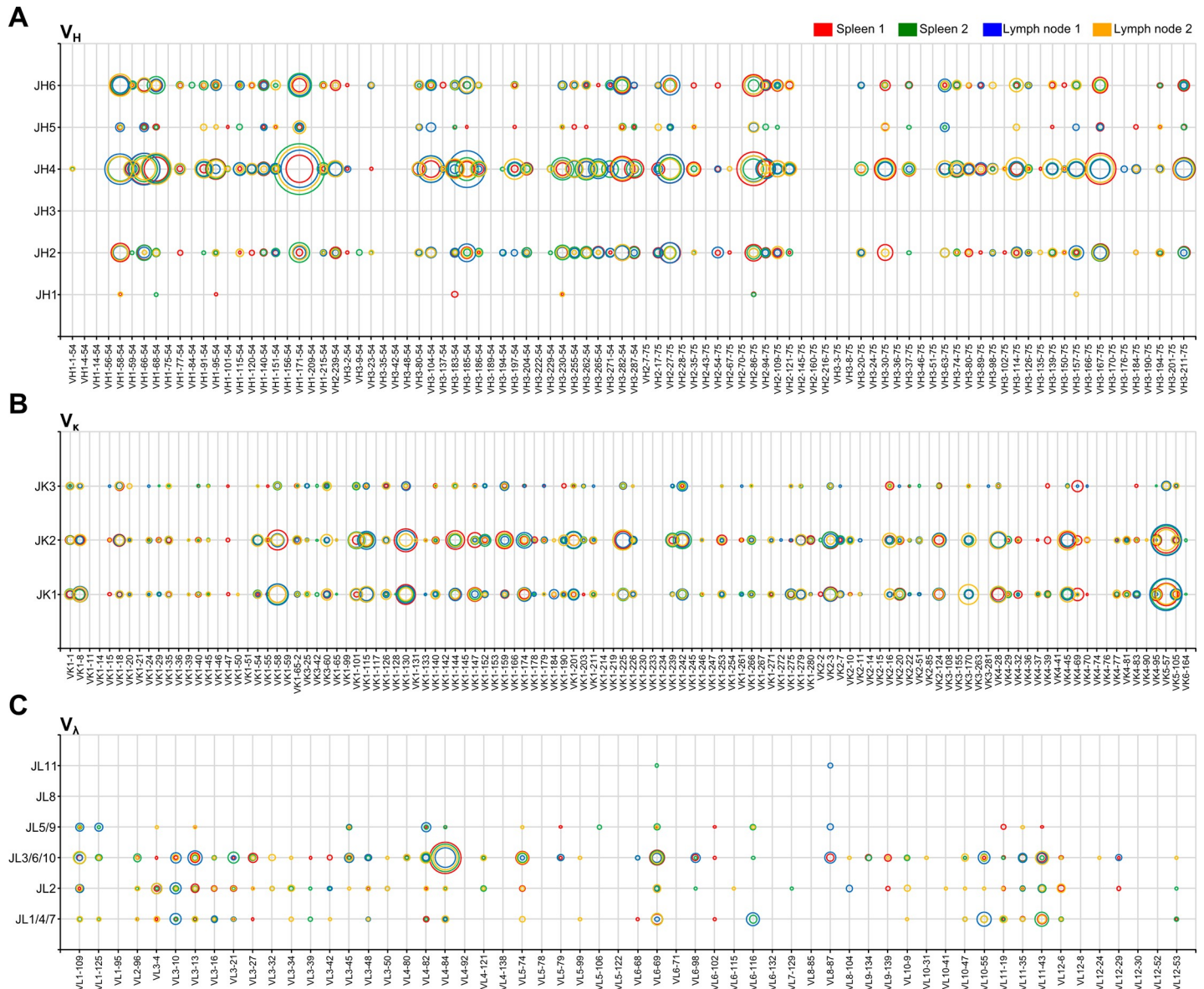


**Fig 1. Workflow of NGS-based guinea pig naïve V gene analysis.** (A) Naïve B cells in spleens and lymph nodes were isolated from two individual naïve guinea pigs by FACS. Each of the  $V_{H(\mu)}$ ,  $V_{\kappa}$  and  $V_{\lambda}$  gene libraries was prepared by 5' RACE PCR. Multiplex identifier (MID), which identifies the sample origin, is added with the sequencing adaptor. (B) The libraries are first sequenced with the 454 GS-FLX Plus system. Raw data are first processed using the sequence quality, and in-frame reads with lengths greater than 400 bp are pooled. After partitioning each V(D)J combinatorial match into its own subset, IgBLAST alignments are performed using custom IgBLAST databases for guinea pigs. Output data are sorted by MID and analyzed for V gene usage frequency, CDR analysis, and mutation frequency.

<https://doi.org/10.1371/journal.pone.0208977.g001>

There are 94 functional  $V_H$  and six functional  $J_H$  germline gene segments in a guinea pig, and our transcribed naïve  $Ig_H$  clone sequences were annotated to 66 (70.2%) of  $V_H$  and five (83.3%) of  $J_H$  gene segments, which represent 39.5% of the theoretical germline combination. The  $V_H$  usage appeared to have unequal frequencies, with ten  $V_H$  gene segments dominating 51.5% of the naïve  $Ig_H$  repertoire, in which  $VH1-171-54$  (9.2%) was the most common  $V_H$  gene segment. The most common  $J_H$  gene segment was  $JH4$  (64.0%); thus, the  $VH1-171-54 \times JH4$  combination shared 6.7% of the naïve  $Ig_H$  repertoire.  $JH3$  was not used in the repertoire analyzed here.

There are 114 functional  $V_{\kappa}$  and three functional  $J_{\kappa}$  germline gene segments in a guinea pig, and our transcribed naïve  $Ig_{\kappa}$  clone sequences were annotated to 83 (72.8%) of  $V_{\kappa}$  and three (100%) of  $J_{\kappa}$  gene segments, which represent 63.7% of the theoretical germline combination. The  $V_{\kappa}$  usage also showed unequal frequencies, with ten  $V_{\kappa}$  gene segments dominating 49.6% of the naïve  $Ig_{\kappa}$  repertoire, in which  $VK5-57$  (13.3%) was the most common  $V_{\kappa}$  gene



**Fig 2. V and J gene segment usages in naïve B cells from individual guinea pigs.** The x-axis represents the germline V gene segment, and the y-axis represents the germline J gene segment. Each point in the x-y axis represents all possible V-J gene segment combinations. The bubble size at each point corresponds to the percent frequency of reads from each sample matching a particular V-J gene segment combination. The bubble color indicates the sample origin, spleen (red) and lymph node (blue) from guinea pig no. one and spleen (green) and lymph node (orange) from guinea pig no. two.

<https://doi.org/10.1371/journal.pone.0208977.g002>

segment. The most common  $J_{\kappa}$  gene segment was JK2 (50.8%); thus, the VK5-57 x JK2 combination shared 6.3% of the naïve  $Ig_{\kappa}$  repertoire.

It is interesting to note that guinea pigs use  $\lambda$  light chains, while  $\kappa$  light chains dominate over  $\lambda$  in mice [28]. There are 58 functional  $V_{\lambda}$  germline gene segments in a guinea pig, and our transcribed naïve  $Ig_{\lambda}$  clone sequences were annotated to 47 (81.0%)  $V_{\lambda}$  gene segments. Among the 11 functional  $J_{\lambda}$  germline gene segments, the nucleotide sequences of JL1, 4 and 7, JL3, 6 and 10, and JL5 and 9 are identical; therefore, we annotated our transcribed naïve  $Ig_{\lambda}$  clone sequences to six  $J_{\lambda}$  gene segment groups, including J1/4/7, J2, J3/6/10, J5/9, J8 and J11. Five out of six  $J_{\lambda}$  gene segment groups were used. The  $V_{\lambda}$  gene segment usage was also unequal,

and the five  $V_{\lambda}$  gene segments dominated 52.8% of the entire naïve repertoire. The most common  $V_{\lambda}$  gene segment was VL4-84 (25.0%), and the most common  $J_{\lambda}$  segment was JL3/6/10 (62.0%). The VL4-84 x JL3/6/10 combination shared 23.7% of the naïve  $Ig_{\lambda}$  repertoire. No significant difference in V-J gene segment usage was observed in four MID samples, suggesting that each organ of two naïve guinea pigs performs a similar V(D)J recombination.

### CDR3 length distribution of $V_H$ , $V_K$ and $V_{\lambda}$ genes in naïve guinea pig

CDR3 is the result of V(D)J recombination and largely determines the diversity of the Ig repertoire; additionally, its length distribution differs among species [1, 29–32]. To understand the guinea pig Ig repertoire structure, we analyzed the naïve CDR3 length distributions of the  $V_H$ ,  $V_K$  and  $V_{\lambda}$  genes (Fig 3 and S1 Fig). The  $V_H$  CDR3 length in guinea pigs ranged from five to 18 aa (98.9%), with the highest frequency at 11 aa (15.1%). Thus, the guinea pig  $V_H$  CDR3 length profile is similar to that of mice, which ranged from five to 26 aa with the highest frequency at ten aa [31].  $V_K$  CDR3 length in guinea pigs ranged from four to 12 aa with the highest frequency at nine aa (74.8%).  $V_{\lambda}$  CDR3 length in guinea pigs was not normally distributed, ranging from seven to 13 aa, with nine aa (55.2%) and 11 aa (28.4%) as frequently occurring lengths. The length distribution profile of naïve CDRs 1 and 2 of the  $V_H$ ,  $V_K$  and  $V_{\lambda}$  genes closely mimicked those of the germline repertoire.

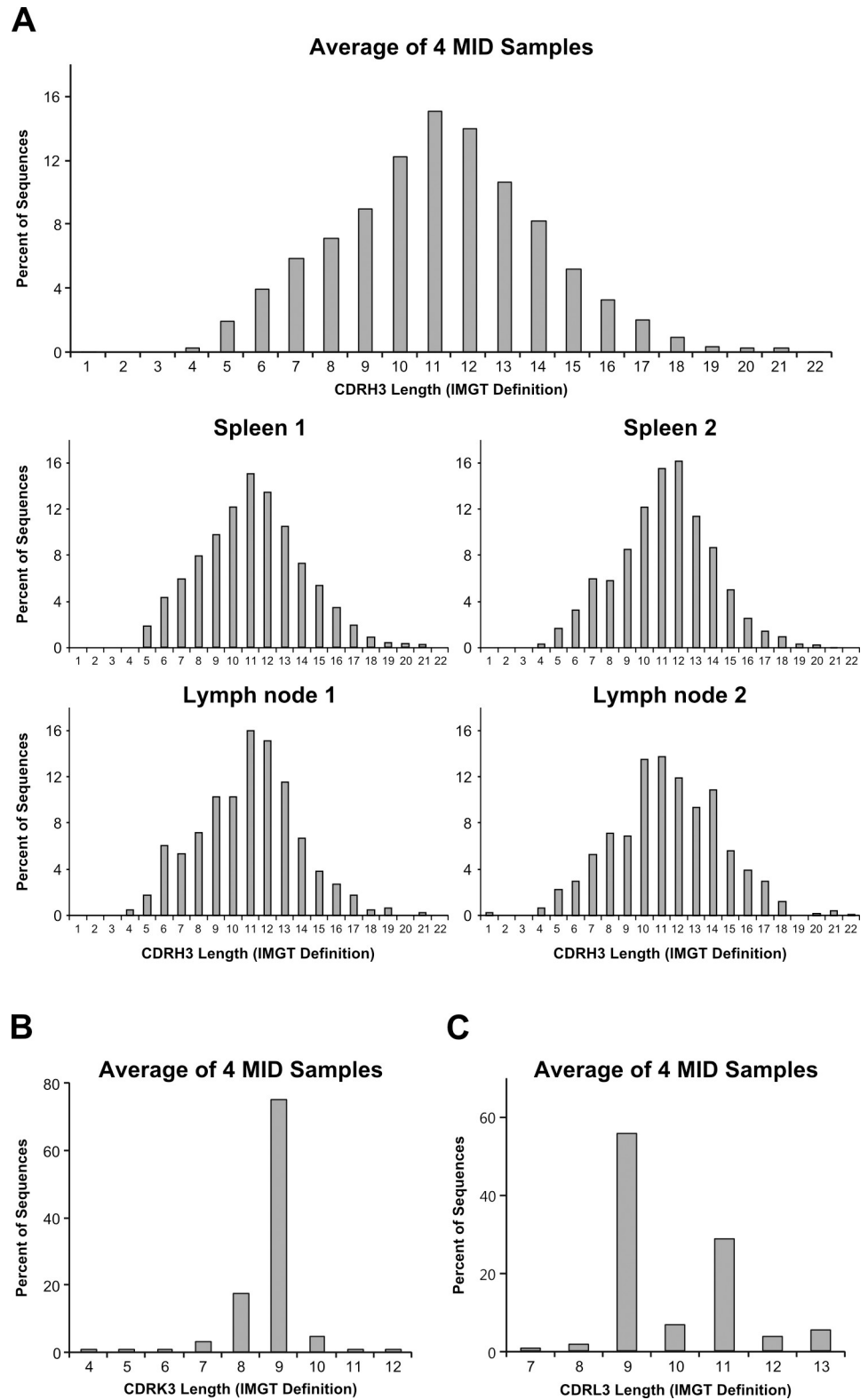
### Preferential use of a particular V gene segment with accumulated mutation in plasma cells

After antigen encounters, B cell clones with affinity to specific antigens undergo clonal selection and expansion, then terminally differentiated into PCs [1]. We previously isolated guinea pig mAbs against CHK2 peptide, p53 peptide, ROR $\gamma$ T peptide and human insulin by directly cloning V genes from antigen-specific PCs. To understand the effect of immunization with foreign antigens on the guinea pig Ig repertoire, we sequenced 86 for  $V_H$ , 68 for  $V_K$ , and 25 for  $V_{\lambda}$  genes from those mAb clones. Comparison of Ig repertoires between naïve B cells (Fig 4: red bubbles) and PCs (Fig 4: black bubbles) revealed striking bias in  $Ig_H V$  and  $Ig_K V/Ig_{\lambda} V$  gene usage in PCs. For example, the VH3-229-54 x JH6 gene segment combination appears to be frequently expressed in PCs, whereas the combination in naïve B cells is rare.

Next, we selected V gene segments that were expressed in both naïve B and PCs: VH3-183-54 (naïve = 100 and PC = 14), VH3-157-75 (naïve = 88 and PC = 8), VK4-95 (naïve = 100 and PC = 23), VK4-45 (naïve = 100 and PC = 16), VL11-35 (naïve = 30 and PC = 10), and VL4-82 (naïve = 44 and PC = 6). The selected reads were aligned to their germline V gene segments, and the relative mutation frequency was plotted on each germline V gene segment nucleotide sequence (Fig 5). We found higher mutations in all PC repertoires than in naïve B repertoires, which preferentially accumulated in CDRs 1 and 2 (Table 2). The higher mutation frequency may thus have been involved in the selection of the cells producing high affinity antibodies. Phylogenetic analysis of mAb clones of PCs derived from insulin-immunized guinea pigs (VH3-183-54 x JH6) revealed that these clones originate from the expansion and somatic hypermutation of a single B cell encoding an unmutated VH3-183-54 x JH6 antibody (S2 Fig).

## Discussion

In this study, we analyzed the expressed B cell receptor repertoire of splenic and lymph node naïve B cells that were isolated from two naïve guinea pigs. We found that primary repertoire diversity is developed mainly via V(D)J recombination using a large number of functional Ig gene segments, and additional diversity is introduced with SHM.

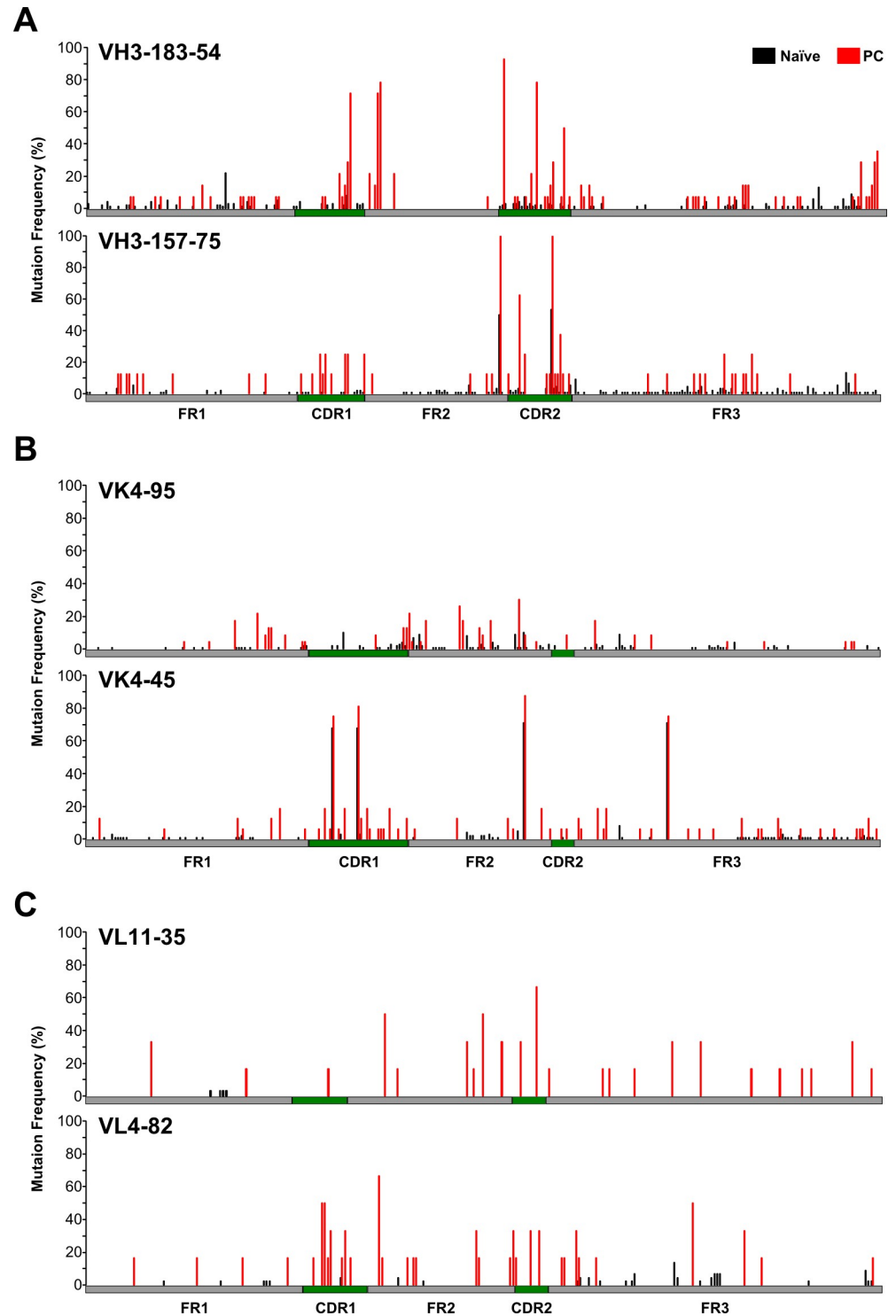


**Fig 3. CDR3 length distributions of  $V_H$ ,  $V_K$  and  $V_L$  in naïve B cells.** The relative frequencies of the CDR3 amino acid lengths of the  $V_H$  (A),  $V_K$  (B) and  $V_L$  (C) genes are shown. Data are from four MID samples ( $N = 5675$  for  $V_H$ , 9360 for  $V_K$  and 1230 for  $V_L$ ). CDR3s are defined according to IMGT unique numbering.

<https://doi.org/10.1371/journal.pone.0208977.g003>







**Fig 5. Distribution of mutational status in naïve and immunized guinea pigs.** The x-axis represents nucleotide position defined according to IMGT unique numbering. The mutation frequencies in V gene segments (VH3-183-54, VH3-157-75, VK4-95, VK4-45, VL11-35 and VL4-82) of B cells in naïve guinea pigs (black) and plasma cells in immunized guinea pigs (red) are determined. Analyzed Ig sequence counts: VH3-183-54, naïve = 100, PC = 14; VH3-157-75, naïve = 88, PC = 8; VK4-95, naïve = 100, PC = 23; VK4-45, naïve = 100, PC = 16; VL11-35, naïve = 30, PC = 10; VL4-82, naïve = 44, PC = 6.

<https://doi.org/10.1371/journal.pone.0208977.g005>

**Table 2. Mutation frequencies in naïve and immunized guinea pigs.**

		Mutation Frequency (%)			CDR/FR Ratio
		FR	CDR	V	
VH3-183-54	naïve	0.7	1.3	0.8	1.8
	PC	3.0	8.3	3.9	2.8
VH3-157-75	naïve	1.1	1.9	1.2	1.8
	PC	1.9	10.2	3.3	5.3
VK4-95	naïve	0.6	0.8	0.6	1.4
	PC	1.4	1.0	1.3	0.7
VK4-45	naïve	1.0	3.2	1.4	3.2
	PC	2.1	7.1	2.9	3.4
VL11-35	naïve	0.1	0.0	0.1	0.0
	PC	1.3	2.6	1.5	1.9
VL4-82	naïve	0.4	0.1	0.4	0.2
	PC	2.0	8.3	3.0	4.1

FR, framework region; CDR, complementarity-determining region; V, V gene segment; PC, plasma cell

<https://doi.org/10.1371/journal.pone.0208977.t002>

approximately 81.0% of them are rearranged with  $J_\lambda$  genes to express functional light chains. This feature may contribute to greater  $Ig_H$  and  $Ig_L$  combinatorial diversity in guinea pigs than in other rodents.

The number of VH genes and their subgroup composition varies by species, but placental mammalian VH segments can be classified into three clusters, clans I, II and III [34, 35]. Several species express only one or several closely related VH subgroups belong to either clan II (horse and cattle) or clan III (rabbit, dog, and pig) [32, 36–40]. Contrary, VH repertoires expressed in the mouse and human are distributed across all three mammalian VH clans, in which 22% of 181 functional VH segments are expressed in mouse, and most of 36–49 functional VH segments are expressed in human [41]. Guinea pig VH gene segments are divided into three subgroups (families 1, 2 and 3), in which families 1 and 2 belong to clan II, and family 3 belongs to clan III [25]. We have shown that at least 13 out of 22 functional VH genes (59%) in family 1, 10 out of 17 functional VH genes (58%) in family 2, and 35 out of 55 functional VH genes in family 3 were expressed. Taking the large number of functional VH genes and their high usage into account, expressed VH repertoire in guinea pig is larger than those found in mouse and human.

Another distinctive feature of the guinea pig germline  $V_H$  repertoire is that the number of pseudo- $V_H$  gene segments is four times more than the potentially functional  $V_H$  gene segments. It is not known whether these pseudo-V gene segments serve as donor pools for gene conversion and contribute to Ig diversity in guinea pigs. In this study, we did not find clear evidence of a gene conversion event; however, it might be detected among Ig genes with a favorable arrangement of V genes and an antigenic stimulation that selects cells with conversions [42–44]. Because our data depth was not sufficient to cover the relatively rare repertoire, deeper read depth is needed to provide precise determination of naïve and antigen-stimulated guinea pig Ig repertoire structure and size.

Taking advantage of the evolutionary distance of guinea pigs between mice and humans, guinea pigs have been used as host animals for pAb production [18–22]. We recently developed a method for generating mAb from guinea pigs and showed that this animal can produce high affinity and specificity antibodies [23, 24]. To understand the effect of immunization with foreign antigens on guinea pig Ig repertoire, we analyzed V and J gene segment usage and

mutation state of  $V_H$ ,  $V_K$  and  $V_L$  genes of PCs extracted from immunized guinea pigs. The V and J gene segment usage of the PCs suggested that immunization skewed the usage, and even relatively rare V-J gene segment combinations could be selected and amplified. We also found higher mutation frequencies in all PC repertoires with preferential accumulation in CDRs 1 and 2. These results imply a high frequency and stepwise occurrence of somatic point mutations in the expressed V genes and substantial clonal expansion of B cells in the guinea pigs and support the notion that guinea pigs can be a promising animal for antibody production. It has been shown that V-segment usage is influenced by a variety of factors including genetics, age and environment [45–47]. Further studies will be needed to understand the mechanism underlining the variance in V-segment usage in guinea pig.

Our study represents the first attempt to characterize sequence diversity in the expressed guinea pig antibody repertoire. These results provide significant insight into antibody repertoire generation and Ig-based immunity of guinea pigs.

## Supporting information

### **S1 Table. The primers used in this study.**

(TIF)

### **S1 File. The sequences used in the custom guinea pig IgBLAST database.**

(XLSX)

### **S2 File. The NGS read sequences obtained in this study.**

(XLSX)

**S1 Fig. CDR3 length distributions of  $V_K$  and  $V_L$  in naïve guinea pig B cells.** The relative frequencies of CDR3 amino acid length of  $V_K$  (A) and  $V_L$  (B) of each four MID samples are shown.

(TIF)

**S2 Fig. Phylogenetic tree of guinea pig naïve and immunized  $V_H$  genes of VH3-183-54 x JH6 gene segment combination.** FR1 to FR4 nucleotide sequences of  $V_H$  genes were arranged in a phylogenetic tree denoting sequence similarity. Sequences of mAbs derived from insulin-immunized guinea pigs are shown in bold red. Sequences derived from naïve guinea pigs are shown in black.

(TIF)

## Acknowledgments

We thank past and current members of our laboratory for fruitful discussions. We also thank Y. Nohara for technical support.

## Author Contributions

**Funding acquisition:** Masaharu Isobe, Nobuyuki Kurosawa.

**Investigation:** Shun Matsuzawa.

**Supervision:** Masaharu Isobe, Nobuyuki Kurosawa.

**Writing – original draft:** Shun Matsuzawa.

**Writing – review & editing:** Nobuyuki Kurosawa.

## References

1. Schroeder HW, Cavacini L. Structure and Function of Immunoglobulins. *J Allergy Clin Immunol*. 2010 Feb; 125(2 0 2):S41–52.
2. de los Rios M, Criscitiello MF, Smider VV. Structural and genetic diversity in antibody repertoires from diverse species. *Curr Opin Struct Biol*. 2015 Aug; 33:27–41. <https://doi.org/10.1016/j.sbi.2015.06.002> PMID: 26188469
3. Tonegawa S. Somatic generation of antibody diversity. *Nature*. 1983 Apr; 302(5909):575–81. PMID: 6300689
4. Teng G, Papavasiliou FN. Immunoglobulin Somatic Hypermutation. *Annual Review of Genetics*. 2007; 41(1):107–20.
5. Jenne CN, Kennedy LJ, Reynolds JD. Antibody repertoire development in the sheep. *Developmental & Comparative Immunology*. 2006 Jan; 30(1):165–74.
6. Zhao Y, Jackson SM, Aitken R. The bovine antibody repertoire. *Developmental & Comparative Immunology*. 2006 Jan; 30(1):175–86.
7. Yasuda M, Jenne CN, Kennedy LJ, Reynolds JD. The sheep and cattle Peyer's patch as a site of B cell development. *Vet Res*. 2006 May 1; 37(3):401–15. <https://doi.org/10.1051/vetres:2006008> PMID: 16611555
8. Liljavirta J, Ekman A, Knight JS, Pernthaner A, Iivanainen A, Niku M. Activation-induced cytidine deaminase (AID) is strongly expressed in the fetal bovine ileal Peyer's patch and spleen and is associated with expansion of the primary antibody repertoire in the absence of exogenous antigens. *Mucosal Immunology*. 2013 Sep; 6(5):942–9. <https://doi.org/10.1038/mi.2012.132> PMID: 23299615
9. Wang F, Ekiert DC, Ahmad I, Yu W, Zhang Y, Bazirgan O, et al. Reshaping Antibody Diversity. *Cell*. 2013 Jun 6; 153(6):1379–93. <https://doi.org/10.1016/j.cell.2013.04.049> PMID: 23746848
10. Deiss TC, Vadnais M, Wang F, Chen PL, Torkamani A, Mwangi W, et al. Immunogenetic factors driving formation of ultralong VH CDR3 in *Bos taurus* antibodies. *Cell Mol Immunol*. 2017 Dec;
11. Becker RS, Knight KL. Somatic diversification of immunoglobulin heavy chain VDJ genes: Evidence for somatic gene conversion in rabbits. *Cell*. 1990 Nov; 63(5):987–97. PMID: 2124176
12. Reynaud C-A, Anquez V, Grimal H, Weill J-C. A hyperconversion mechanism generates the chicken light chain preimmune repertoire. *Cell*. 1987 Feb 13; 48(3):379–88. PMID: 3100050
13. Reynaud C-A, Dahan A, Anquez V, Weill J-C. Somatic hyperconversion diversifies the single VH gene of the chicken with a high incidence in the D region. *Cell*. 1989 Oct 6; 59(1):171–83. PMID: 2507167
14. Ordway D, Palanisamy G, Henao-Tamayo M, Smith EE, Shanley C, Orme IM, et al. The Cellular Immune Response to *Mycobacterium tuberculosis* Infection in the Guinea Pig. *The Journal of Immunology*. 2007 Aug 15; 179(4):2532–41. PMID: 17675515
15. Padilla-Carlin DJ, McMurray DN, Hickey AJ. The Guinea Pig as a Model of Infectious Diseases. *Comp Med*. 2008 Aug; 58(4):324–40. PMID: 18724774
16. Kumar M, Krause KK, Azouz F, Nakano E, Nerurkar VR. A guinea pig model of Zika virus infection. *Virology Journal*. 2017 Apr; 14:75. <https://doi.org/10.1186/s12985-017-0750-4> PMID: 28399888
17. McGregor A, Choi KY. Cytomegalovirus Antivirals and Development of Improved Animal Models. *Expert Opin Drug Metab Toxicol*. 2011 Oct; 7(10):1245–65. <https://doi.org/10.1517/17425255.2011.613824> PMID: 21883024
18. Yalow RS, Berson SA. Radioimmunoassay of gastrin. *Gastroenterology*. 1970 Jan 1; 58(1):1–14. PMID: 5460746
19. Odell W, Skowsky R, Abraham G, Hescox M, Fisher D, Grover PK. Production of Antisera for Polypeptide and Steroid Radioimmunoassays. *Biol Reprod*. 1972 Jun 1; 6(3):427–42. PMID: 5036869
20. Hanly WC, Artwohl JE, Bennett BT. Review of Polyclonal Antibody Production Procedures in Mammals and Poultry. *ILAR J*. 1995 Jan 1; 37(3):93–118. PMID: 11528030
21. Lent SV, Vos WHD, Creasy HH, Marques PX, Ravel J, Vanrompuy D, et al. Analysis of Polymorphic Membrane Protein Expression in Cultured Cells Identifies PmpA and PmpH of *Chlamydia psittaci* as Candidate Factors in Pathogenesis and Immunity to Infection. *PLOS ONE*. 2016 Sep 15; 11(9): e0162392. <https://doi.org/10.1371/journal.pone.0162392> PMID: 27631978
22. D'Erchia AM, Gissi C, Pesole G, Saccone C, Arnason U. The guinea-pig is not a rodent. *Nature*. 1996 Jun; 381(6583):597–600. <https://doi.org/10.1038/381597a0> PMID: 8637593
23. Kurosawa N, Yoshioka M, Fujimoto R, Yamagishi F, Isobe M. Rapid production of antigen-specific monoclonal antibodies from a variety of animals. *BMC Biology*. 2012; 10:80. <https://doi.org/10.1186/1741-7007-10-80> PMID: 23017270

24. Kurosawa N, Wakata Y, Inobe T, Kitamura H, Yoshioka M, Matsuzawa S, et al. Novel method for the high-throughput production of phosphorylation site-specific monoclonal antibodies. *Scientific Reports*. 2016 Apr 29; 6:25174. <https://doi.org/10.1038/srep25174> PMID: 27125496
25. Guo Y, Bao Y, Meng Q, Hu X, Meng Q, Ren L, et al. Immunoglobulin Genomics in the Guinea Pig (*Cavia porcellus*). *PLOS ONE*. 2012 Jun 22; 7(6):e39298. <https://doi.org/10.1371/journal.pone.0039298> PMID: 22761756
26. Lefranc M-P. Unique database numberings system for immunogenetic analysis. *Immunology Today*. 1997 Nov 1; 18(11):509. PMID: 9386342
27. Lefranc M-P, Pommié C, Ruiz M, Giudicelli V, Foulquier E, Truong L, et al. IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Developmental & Comparative Immunology*. 2003 Jan; 27(1):55–77.
28. Sun Y, Liu Z, Ren L, Wei Z, Wang P, Li N, et al. Immunoglobulin genes and diversity: what we have learned from domestic animals. *J Anim Sci Biotechnol*. 2012 Jun 20; 3(1):18. <https://doi.org/10.1186/2049-1891-3-18> PMID: 22958617
29. Prabakaran P, Chen W, Singarayan MG, Stewart CC, Streaker E, Feng Y, et al. Expressed antibody repertoires in human cord blood cells: 454 sequencing and IMGT/HighV-QUEST analysis of germline gene usage, junctional diversity, and somatic mutations. *Immunogenetics*. 2012 May 1; 64(5):337–50. <https://doi.org/10.1007/s00251-011-0595-8> PMID: 22200891
30. Rettig TA, Ward C, Bye BA, Pecaup MJ, Chapes SK. Characterization of the naïve murine antibody repertoire using unamplified high-throughput sequencing. *PLOS ONE*. 2018 Jan 10; 13(1):e0190982. <https://doi.org/10.1371/journal.pone.0190982> PMID: 29320559
31. Wu L, Oficjalska K, Lambert M, Fennell BJ, Darmanin-Sheehan A, Shúilleabháin DN, et al. Fundamental Characteristics of the Immunoglobulin VH Repertoire of Chickens in Comparison with Those of Humans, Mice, and Camelids. *The Journal of Immunology*. 2012 Jan 1; 188(1):322–33. <https://doi.org/10.4049/jimmunol.1102466> PMID: 22131336
32. Lavinder JJ, Hoi KH, Reddy ST, Wine Y, Georgiou G. Systematic Characterization and Comparative Analysis of the Rabbit Immunoglobulin Repertoire. *PLOS ONE*. 2014 Jun 30; 9(6):e101322. <https://doi.org/10.1371/journal.pone.0101322> PMID: 24978027
33. Schroeder HW. Similarity and divergence in the development and expression of the mouse and human antibody repertoires. *Developmental & Comparative Immunology*. 2006 Jan; 30(1):119–35.
34. Ota T, Nei M. Divergent evolution and evolution by the birth-and-death process in the immunoglobulin VH gene family. *Mol Biol Evol*. 1994 May 1; 11(3):469–82. <https://doi.org/10.1093/oxfordjournals.molbev.a040127> PMID: 8015440
35. Sitnikova T, Su C. Coevolution of immunoglobulin heavy- and light-chain variable-region gene families. *Mol Biol Evol*. 1998 Jun 1; 15(6):617–25. <https://doi.org/10.1093/oxfordjournals.molbev.a025965> PMID: 9615443
36. Sun Y, Wang C, Wang Y, Zhang T, Ren L, Hu X, et al. A comprehensive analysis of germline and expressed immunoglobulin repertoire in the horse. *Developmental & Comparative Immunology*. 2010 Sep; 34(9):1009–20.
37. Saini SS, Hein WR, Kaushik A. A single predominantly expressed polymorphic immunoglobulin VH gene family, related to mammalian group, I, clan, II, is identified in cattle. *Molecular Immunology*. 1997 Jun; 34(8):641–51.
38. Kodangattil S, Huard C, Ross C, Li J, Gao H, Mascioni A, et al. The functional repertoire of rabbit antibodies and antibody discovery via next-generation sequencing. *MAbs*. 2014 May 1; 6(3):628–36. <https://doi.org/10.4161/mabs.28059> PMID: 24481222
39. Bao Y, Guo Y, Xiao S, Zhao Z. Molecular characterization of the VH repertoire in *Canis familiaris*. *Veterinary Immunology and Immunopathology*. 2010 Sep; 137(1):64–75.
40. Butler JE, Sun J, Kacskovics I, Brown WR, Navarro P. The VH and CH immunoglobulin genes of swine: implications for repertoire development. *Veterinary Immunology and Immunopathology*. 1996 Nov; 54(1):7–17.
41. Lefranc M-P. IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res*. 2001 Jan 1; 29(1):207–9. PMID: 11125093
42. Tsai H-F, D'Avirro N, Selsing E. Gene conversion-like sequence transfers in a mouse antibody transgene: antigen selection allows sensitive detection of V region interactions based on homology. *Int Immunol*. 2002 Jan 1; 14(1):55–64. PMID: 11751752
43. D'Avirro N, Truong D, Xu B, Selsing E. Sequence Transfers between Variable Regions in a Mouse Antibody Transgene Can Occur by Gene Conversion. *The Journal of Immunology*. 2005 Dec 15; 175(12):8133–7. PMID: 16339551

44. Darlow JM, Stott DI. Gene conversion in human rearranged immunoglobulin genes. *Immunogenetics*. 2006 Jul 1; 58(7):511–22. <https://doi.org/10.1007/s00251-006-0113-6> PMID: 16705406
45. Corcoran MM, Phad GE, Bernat NV, Stahl-Hennig C, Sumida N, Persson MAA, et al. Production of individualized V gene databases reveals high levels of immunoglobulin genetic diversity. *Nature Communications*. 2016 Dec 20; 7:13642. <https://doi.org/10.1038/ncomms13642> PMID: 27995928
46. Boyd SD, Crowe JE. Deep sequencing and human antibody repertoire analysis. *Curr Opin Immunol*. 2016 Jun; 40:103–9. <https://doi.org/10.1016/j.coi.2016.03.008> PMID: 27065089
47. Martin V, Wu Y-C (Bryan), Kipling D, Dunn-Walters D. Ageing of the B-cell repertoire. *Phil Trans R Soc B*. 2015 Sep 5; 370(1676):20140237. <https://doi.org/10.1098/rstb.2014.0237> PMID: 26194751