



Tissue-Specific Expressed Antibody Variable Gene Repertoires

Bryan S. Briney¹, Jordan R. Willis³, Jessica A. Finn¹, Brett A. McKinney⁵, James E. Crowe Jr^{1,2,4*}

1 Department of Pathology, Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, Tennessee, United States of America, **2** Department of Pediatrics, Vanderbilt University Medical Center, Nashville, Tennessee, United States of America, **3** The Chemical and Physical Biology Program, Vanderbilt University, Nashville, Tennessee, United States of America, **4** The Vanderbilt Vaccine Center, Vanderbilt University Medical Center, Nashville, Tennessee, United States of America, **5** Tandy School of Computer Science and Department of Mathematics, University of Tulsa, Tulsa, Oklahoma, United States of America

Abstract

Recent developments in genetic technologies allow deep analysis of the sequence diversity of immune repertoires, but little work has been reported on the architecture of immune repertoires in mucosal tissues. Antibodies are the key to prevention of infections at the mucosal surface, but it is currently unclear whether the B cell repertoire at mucosal surfaces reflects the dominant antibodies found in the systemic compartment or whether mucosal tissues harbor unique repertoires. We examined the expressed antibody variable gene repertoires from 10 different human tissues using RNA samples derived from a large number of individuals. The results revealed that mucosal tissues such as stomach, intestine and lung possess unique antibody gene repertoires that differed substantially from those found in lymphoid tissues or peripheral blood. Mutation frequency analysis of mucosal tissue repertoires revealed that they were highly mutated, with little evidence for the presence of naïve B cells, in contrast to blood. Mucosal tissue repertoires possessed longer heavy chain complementarity determining region 3 loops than lymphoid tissue repertoires. We also noted a large increase in frequency of both insertions and deletions in the small intestine antibody repertoire. These data suggest that mucosal immune repertoires are distinct in many ways from the systemic compartment.

Citation: Briney BS, Willis JR, Finn JA, McKinney BA, Crowe JE Jr (2014) Tissue-Specific Expressed Antibody Variable Gene Repertoires. PLOS ONE 9(6): e100839. doi:10.1371/journal.pone.0100839

Editor: Sophia N. Karagiannis, King's College London, United Kingdom

Received: October 13, 2013; **Accepted:** May 30, 2014; **Published:** June 23, 2014

Copyright: © 2014 Briney 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: The work was supported by NIH Contract HHSN272200900047C, NIH grant R01 AI106002, and DoD grant HDTRA1-10-1-0067. BSB was supported by NIH T32 HL069765. 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.

* Email: james.crowe@vanderbilt.edu

Introduction

The humoral immune response produces a massively diverse repertoire of antibodies. In order to respond effectively to challenge from a multitude of unfamiliar pathogens, diversity in the primary (or, naïve) B cell repertoire is accomplished by combinatorial diversity that occurs following recombination of germline variable (V), diversity (D) and joining (J) germline genes and pairing of unique heavy and light chains [1–3]. Repertoire diversity is further enhanced in the memory repertoire by several affinity maturation processes including somatic hypermutation, which introduces point mutations and insertions/deletions (indels), and class-switching [4–6].

In studies of the circulating antibody repertoire, pathogenic infections have been shown to induce antibody responses with biased germline antibody variable gene use, and this bias is often maintained in the post-infection memory B cell population [7–12]. Since each individual has experienced a unique set of pathogenic encounters in a unique order, it is logical to expect that each individual might possess a uniquely biased memory repertoire that reflects the enrichment of clones specific for the particular history of pathogens. Surprisingly, however, circulating memory B cell repertoires often appear very similar when compared across individuals at the level of antibody variable gene usage, suggesting the presence of a global mechanism regulating the genetic composition of the peripheral blood antibody repertoire [13–16].

Circulating B cells with diverse surface receptors (that later become secreted antibodies with the same specificity) constitute the primary humoral immune cell type responding to systemic infection, and recent work has described the human peripheral blood antibody repertoire in great detail [13–15,17,18]. Much less is known about the repertoire composition of tissue-resident B cells, however. In the gut mucosa, resident plasma cells secrete almost exclusively IgA, and the presence of IgA-secreting plasma cells depends on the presence of colonizing bacteria in the gut [19]. In contrast to conventional germinal centers in lymph nodes and other lymphoid organs, many mucosal B cells are thought to mature using T cell-independent routes, which likely affects the diversity of the mucosal antibody repertoire [20,21]. Indeed, spectratypic analyses of the mucosal antibody repertoire have provided evidence of increased oligoclonality of the mucosal IgA repertoire [22]. This finding raises the intriguing possibility that mucosal antibody repertoires are distinct from the peripheral blood repertoire, possibly because they are induced in response to site-specific pathogens or using unique maturation processes. Alternatively, there is substantial evidence that the B cell composition of the mucosa is different than peripheral blood, resulting in alterations in the expressed antibody repertoire. The presence of large numbers of commensals in the gut microbiome also could influence the specificity of the mucosal B cell repertoire. In this report, we used high-throughput DNA sequencing

techniques to analyze the expressed antibody gene repertoire in order to determine whether mucosal lymphocytes harbor a unique repertoire. Indeed, a detailed analysis of mucosal and lymphoid repertoires revealed that mucosal antibody repertoires are genetically distinct from the antibody repertoires of both non-mucosal lymphoid tissues and peripheral blood cells.

Materials and Methods

Tissue-specific total RNA and mRNA

Purified polyA⁺ mRNA (lymph node) or total tissue RNA (all other samples) from the tissues of healthy human subjects was obtained from a commercial source (Clontech). Each RNA sample, as provided by Clontech, contains pooled RNA from multiple donors. The number of donors and demographic breakdown for each tissue donor pool is shown in Table 1.

cDNA synthesis and PCR amplification of antibody genes

100 ng of each mRNA or total RNA sample and 10 pmol of each RT-PCR primer (Table 2, adapted from [23]) were used in duplicate 50 μ L RT-PCR reactions using the OneStep RT-PCR system (Qiagen). Thermal cycling was performed in a BioRad DNA Engine PTC-0200 thermal cycler using the following protocol: 50°C for 30 minutes, 95°C for 15 minutes, 35 cycles of (94°C for 45 seconds, 58°C for 45 seconds, 72°C for 2 minutes), 72°C for 10 minutes. cDNA synthesis and amplification was verified by agarose gel electrophoresis before duplicate RT-PCR reactions were pooled. 5 μ L of each pooled RT-PCR reaction was used as template for 100 μ L 454-adaptor PCR reactions, which were carried out in quadruplicate. 20 pmol of 454-adaptor primers (Table 2) and 0.25 units of AmpliTaq Gold polymerase (Applied Biosystems) were used for each reaction. Thermal cycling was performed in a BioRad DNA Engine PTC-0200 thermal cycler using the following protocol: 95°C for 10 minutes, 10 cycles of (95°C for 30 seconds, 58°C for 45 seconds, 72°C for 2 minutes), 72°C for 10 minutes.

Amplicon purification and quantification

Amplicons were purified from the 454-adaptor PCR reaction mix using the Agencourt AMPure XP system (Beckman Coulter Genomics) according to the manufacturer's standard protocol. Purified amplicons were analyzed on a 2% agarose gel to verify complete primer removal and appropriate amplicon size before quantification using a Qubit fluorometer (Invitrogen, High-Sensitivity dsDNA kit) following the manufacturer's standard protocol.

454 method sequence analysis of amplicons

The amplicon libraries were quantified using a Qubit fluorometer (Invitrogen, CA). The size and quality of the DNA libraries were evaluated on an Agilent Bioanalyzer 2100 using the DNA 7500 labchip (Agilent, Palo Alto, CA). The samples then were diluted to a working concentration of 1×10^6 Molecules per μ L. Quality control of the amplicon libraries and emulsion-based clonal amplification and sequencing on the 454 Genome Sequencer FLX Titanium system were performed by the W. M. Keck Center for Comparative and Functional Genomics at the University of Illinois at Urbana-Champaign according to the manufacturer's instructions (454 Life Sciences, Branford, CT). Signal processing and base calling were performed using the bundled 454 Data Analysis Software version 2.5.3 for amplicons. Raw sequencing data will be available at the NCBI Sequence

Table 1. Demographics of pooled tissue sample donors from which the RNA pools were isolated.

Sample	# of donors	Age range (years)
Peripheral leukocytes	39	18–47
Bone marrow	56	22–85
Small intestine	15	29–61
Lung	13	32–61
Stomach	7	20–55
Lymph node	42	NA
Tonsil	34	22–61
Spleen	12	18–54
Thymus	25	NA

NA indicates age range was not available.
doi:10.1371/journal.pone.0100839.t001

Read Archive (<http://www.ncbi.nlm.nih.gov/sra>) on the date of publication.

Antibody Sequence Analysis

The FASTA files resulting from the 454 method sequence analysis were submitted to IMGT HIGH/V-Quest (IMGT, the international ImMunoGeneTics information system; www.imgt.org; founder and director: Marie-Paule LeFranc, Montpellier, France) [24,25]. The sequences were analyzed using the HIGH/V-Quest option to perform codon-based correction of homopolymer-induced insertion/deletion errors, which are relatively common when using pyrosequencing. The IMGT output was parsed into a custom MySQL database for further analysis. Following initial analysis by IMGT, sequences were retained for analysis only if they passed additional antibody-related filters: (1) appropriate read length (>300 bases), (2) identification of V, D, and J segments by IMGT analysis, and (3) presence of an in-frame junctional rearrangement. To reduce the effect of high copy-number plasma cells on the repertoire analysis, identical sequences were collapsed to produce a dataset of unique, high-quality antibody sequences.

Clustering of Antibody Repertoires

We perform agglomerative hierarchical clustering with complete linkage on both VDJ genes and tissue-specific donor pools. First, we perform a filter that removes VDJ genes with low counts of low variation across all samples. Then we calculate pairwise distances between genes and tissue-specific donor pools using Pearson correlation. We standardize the values in the heat map to display in the range -3 to $+3$. Dendrograms and heatmaps were created with Matlab R2010b.

Analysis of Differential Expression of V(D)J Recombinants

We use the edgeR software [26] to calculate differential expression between tissues. EdgeR uses the negative binomial as the appropriate distribution for count data. We obtain dispersion estimates and test differential expression using the generalized linear model (GLM) likelihood ratio test. The columns in the table show the fold change between tissues and the p-value and Benjamini and Hochberg false discovery rate.

Table 2. Primers used in RT-PCR and 454-Adapter PCR.

PRIMER NAME	SEQUENCE	PRIMER TYPE
V _H 1/7-FR1	CCATCAGCGTGTCTCTAGGCCTCAGTGAAGGTCTCTGCAAG	RT-PCR
V _H 2-FR1	CCATCAGCGTGTCTCTAGTCTGGTCTACGCTGGTGAACCC	RT-PCR
V _H 3-FR1	CCATCAGCGTGTCTCTACTGGGGGGTCCCTGAGACTCTCTG	RT-PCR
V _H 4-FR1	CCATCAGCGTGTCTCTACTTCGGAGACCCTGTCCCTCACCTG	RT-PCR
V _H 5-FR1	CCATCAGCGTGTCTCTACGGGGAGTCTCTGAAGATCTCTGT	RT-PCR
V _H 6-FR1	CCATCAGCGTGTCTCTATCGCAGACCCTCTCACTCACCTGTG	RT-PCR
J _H Consensus	CGCTCAGAGCACTGTAGCTTACCTGAGGAGACGGTGACC	RT-PCR
454-Adapter-A	CGTATCGCCTCCCTCGCGCCATCAGCGTGTCTCTA	454 PCR
454-Adapter-B	CTATGCGCCTTGCCAGCCGCTCAGAGCACTGTAG	454 PCR

Multiplex identifiers (MIDs) are underlined. Forward primers use MID7 and reverse primers use MID4 (Roche). Additional complementarity between RT-PCR primers and 454-Adapter PCR primers is identified in italics. All primer sequences are shown in the 5' to 3' orientation.
doi:10.1371/journal.pone.0100839.t002

Results

Antibody variable gene use in peripheral blood, bone marrow or tissue repertoires

We obtained RNA from several pooled tissue samples: peripheral blood leukocytes, bone marrow, small intestine, lung, stomach, lymph node, tonsil, spleen and thymus. The number of donors and the age distribution for each tissue RNA pool are shown in Table 1. The antibody genes were amplified using RT-PCR, sequence adapters required for 454 sequencing and indexing barcodes were added during a second round of PCR, and the resulting amplicons were subjected to high-throughput DNA sequencing. Following initial sequence analysis using the IMG_T High/V-QUEST server, we performed additional antibody-specific quality filtering. Antibody sequences were only retained for analysis if they were of the appropriate length (> 300 bp); contained IMG_T-assigned V, D and J genes; and an in-frame junction without ambiguous nucleotides. In addition, to reduce the repertoire-skewing effect of high-copy plasma cells, we removed redundant sequences. High copy number sequences resulting from plasma cells may still be over-represented, however. Due to sequencing and/or PCR errors, multiple reads of antibody sequences derived from a single plasma cell may not be identical

and thus would not be removed during the filtering process. After filtering, we obtained a total of 1,412,943 unique antibody sequences. Read statistics for each tissue sample are given in Table 3.

We first determined the frequency at which each variable gene family was found in each tissue and discovered substantial differences in the gene family use in tissues compared to peripheral blood (Figure 1). There are too many differences to detail individually, so we focus here on three of the most immediately apparent trends. First, variable gene family 2 (V_H2) use was increased in every tissue except for small intestine, when compared to the peripheral blood repertoire. V_H2 was found in 7.7% of peripheral blood sequences, and the largest increases were found in bone marrow (13.8%), thymus (12.5%) and lymph node (12.3%). Second, while V_H3 was the most common antibody variable gene family in most of the samples, including peripheral blood as has been shown previously [13,16,17], the lung and thymus samples used the V_H4 family more frequently than V_H3. This finding was somewhat surprising, since V_H3 has been shown to be the most frequently used germline gene throughout early B cell development and across peripheral blood B cell subsets [13,16]. The substantial increase in use of V_H4 in select tissue samples suggests strong selection of V_H4 antibodies in these tissues.

Table 3. Read data from high-throughput sequence analysis of tissue RNA.

Sample	Number of reads	High quality reads	% High quality
Peripheral blood leukocytes	149,896	137,986	92.05
Bone marrow	171,111	159,515	93.22
Small intestine	118,044	137,986	93.45
Lung	198,660	183,264	92.25
Stomach	145,163	133,201	91.76
Lymph node	165,091	157,334	95.30
Tonsil	197,846	186,962	94.50
Spleen	185,247	175,689	94.84
Thymus	179,969	168,679	93.73
Total	1,511,027	1,412,943	93.51

Reads are wells that passed all standard 454 sequence analysis filters. Filtered reads are reads that passed additional antibody-related filters, including appropriate read length (>300 bases), identification of V, D, and J segments by IMG_T analysis, and presence of an in-frame junctional rearrangement.
doi:10.1371/journal.pone.0100839.t003

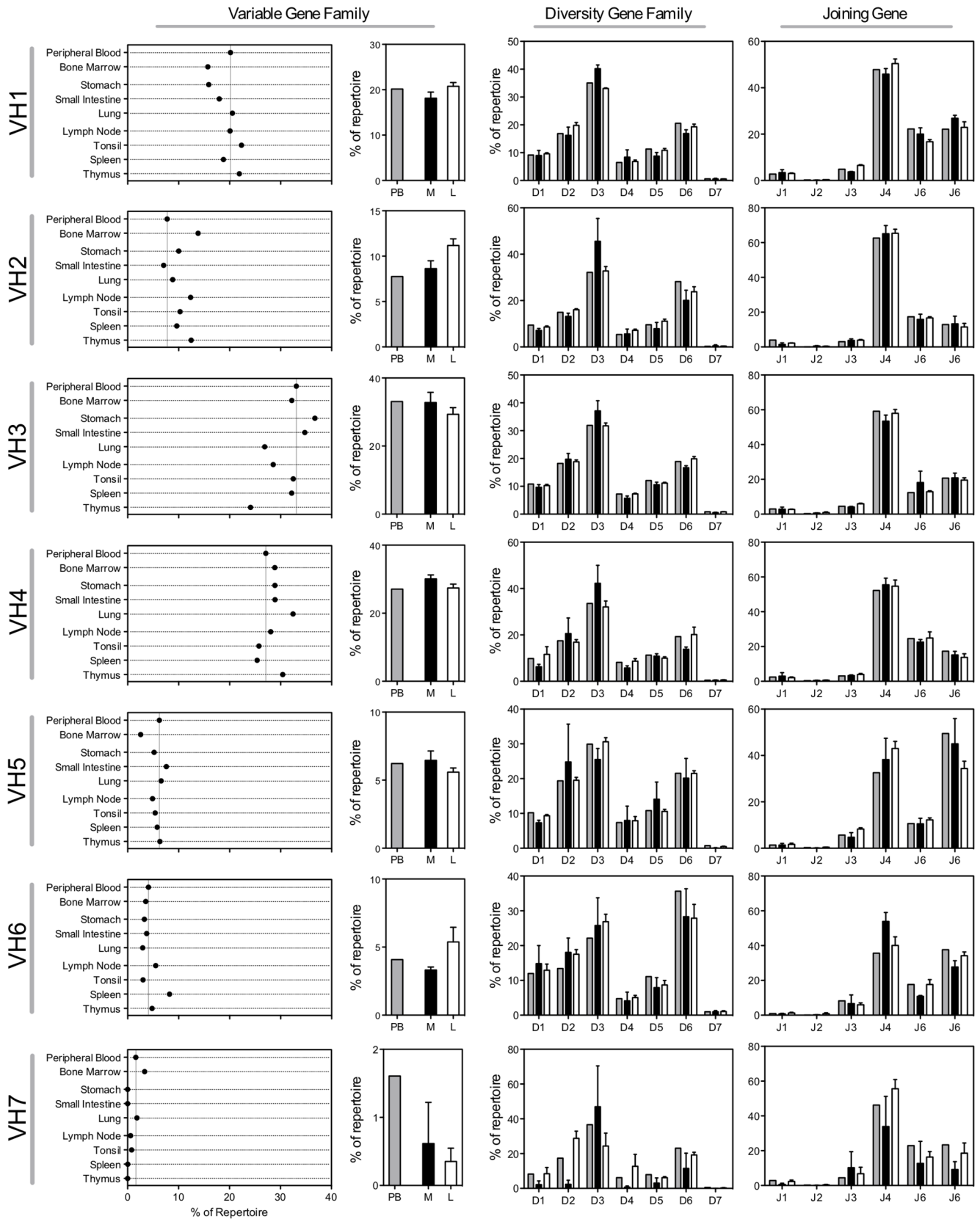


Figure 1. Germline gene family use. Starting from the left, the first column of panels shows the variable gene family use in peripheral blood, bone marrow, mucosal tissues (lung, small intestine, stomach) and lymphoid tissues (lymph node, tonsil, spleen and thymus). For easier comparison, the dashed vertical line in each panel represents the peripheral blood frequency. The second column of panels shows the variable gene use of peripheral blood and the combined variable gene family use of mucosal or lymphoid tissues. Bars indicate mean \pm SEM for each group of tissue

samples. The third column of panels shows the diversity gene family use in peripheral blood (grey bars), mucosal tissues (black bars) and lymphoid tissues (white bars). Bars indicate mean \pm SEM for each group of tissue samples. The final column of panels shows the joining gene use. Colors are the same as the diversity gene family frequency panels. Bars indicate mean \pm SEM for each group of tissue samples.
doi:10.1371/journal.pone.0100839.g001

Finally, we found interesting contrasts when grouping tissue antibody repertoires into mucosal (stomach, lung, small intestine) and lymphoid (lymph node, spleen, tonsil, thymus) groups. All four lymphoid tissue samples showed reduced V_H3 family use compared to peripheral blood, while two of the three mucosal tissue samples (stomach and small intestine) showed increased use of V_H3 family genes. Analysis of the diversity (D) gene family and joining (J) gene use within each variable gene family produced similar usage patterns for both the mucosal and lymphoid groups.

The V(D)J recombinant repertoire of mucosal tissues differed from that of peripheral blood or lymphoid tissues

We next performed a hierarchical clustering analysis on the V(D)J recombinant repertoire for each tissue sample (Figure 2). Interestingly, lymphoid tissues (tonsil, spleen, lymph node and thymus) clustered with each other and with peripheral blood. Mucosal tissues (lung, small intestine, stomach) also clustered together, along with bone marrow. This analysis indicates that the architecture of V(D)J recombinant repertoires of mucosal tissues differs substantially from both peripheral blood and lymphoid tissue repertoires. It is also interesting that the mucosal tissue repertoires, which have been shown to be composed primarily of B

cells encoding antigen-specific antibodies [27], cluster with the bone marrow samples, in which the overwhelming majority of transcription is performed by long-lived plasma cells which are thought to produce much of the circulating antibody proteins in serum. These data suggests that antibody repertoires encoded by tissue-specific B cells may better represent the repertoire of circulating serum antibody proteins than are the antibody repertoires encoded by B cells circulating in the peripheral blood.

To more closely investigate these repertoire differences, we determined the frequency of each V(D)J combination and, for each tissue repertoire sample, calculated the number of V(D)J combinations for which the frequency differed statistically from the peripheral blood sample (Figure 3A and Table S1). We found a trend toward more differences between mucosal tissue samples and peripheral blood than were present between lymphoid tissue samples and blood ($p = 0.079$). We also analyzed the magnitude of the top 50 differences from peripheral blood for each of the mucosal and lymphoid samples, and found that differences in V(D)J frequency between mucosal tissue samples and peripheral blood were significantly larger than differences between lymphoid tissue samples and peripheral blood (Figure 3B; $p = 0.039$). Thus, the genetic composition of mucosal tissue antibody repertoires differs more substantially from peripheral blood than it does from lymphoid tissue repertoires.

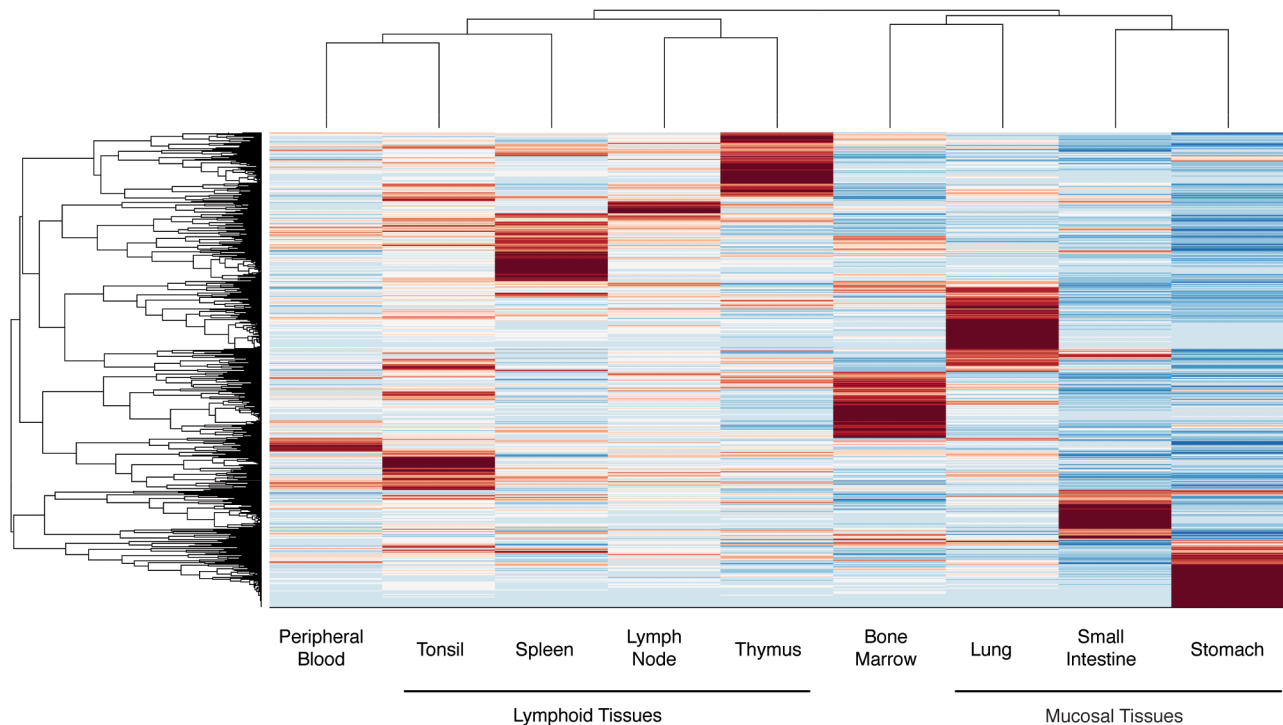


Figure 2. Clustergram of antibody gene repertoires. The frequency of each $V_H(D)J_H$ recombination was determined for each of nine tissues, and a clustergram was created. $V_H(D)J_H$ recombinants were clustered by relative frequency in each tissue-specific repertoire, and the resulting phylogenetic tree is shown on the left. Tissue-specific repertoires were clustered by the overall $V_H(D)J_H$ usage of each repertoire, and the resulting clustering diagram is shown at the top. The frequency variation for each $V_H(D)J_H$ recombination across all tissue-specific repertoires was determined, and standardized to a range of -3 to 3 . A complete list of the frequency variation of all $V_H(D)J_H$ recombinants for each tissue-specific repertoire, along with statistical significance and false discovery rate (FDR) calculations, is available in File S1 (see eight files, each named Suppl Info_PB_vs_tissue).
doi:10.1371/journal.pone.0100839.g002

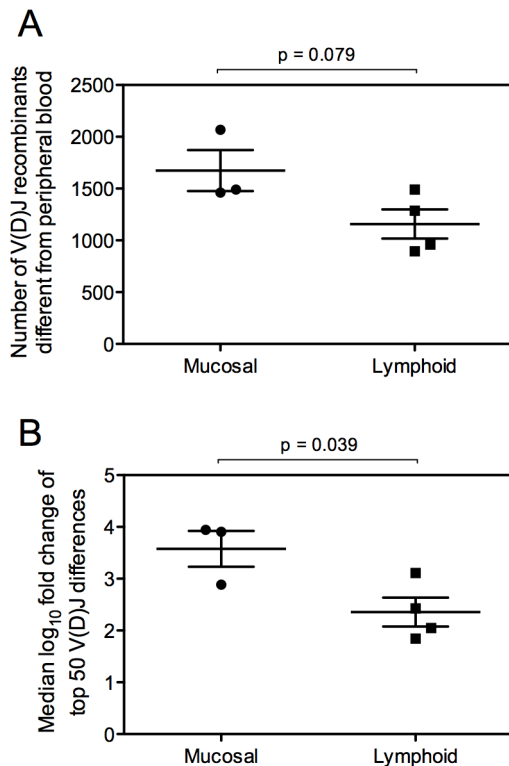


Figure 3. Comparison of $V_H(D)J_H$ use in lymphoid or mucosal tissues to that in peripheral blood. (A) The frequency of each $V_H(D)J_H$ recombination was calculated for each tissue and compared to peripheral blood. The number of $V_H(D)J_H$ recombinants for which the frequency differed significantly from peripheral blood was calculated for each tissue (statistical false discovery rate (FDR) calculations are available in File S1 [see eight files, each named Suppl Info_PB_vs_tissue]). The number of statistically different $V_H(D)J_H$ combinations is shown for each mucosal (lung, small intestine, stomach) and lymphoid (lymph node, tonsil, spleen and thymus) tissue. (B) The frequency of each $V_H(D)J_H$ recombination was determined for each tissue and compared to peripheral blood. The fold change of the 50 most different $V_H(D)J_H$ recombinations is shown in \log_{10} scale for each tissue. doi:10.1371/journal.pone.0100839.g003

Mutation frequency analysis of peripheral blood, bone marrow or tissue repertoires

Sequences from each tissue subset were grouped by mutation frequency. Then, the relative abundance of each mutation frequency group was calculated, and a mutation histogram was created for each tissue sample (Figure 4A). The peripheral blood sample contained a large number of sequences with few or no mutations, as has been shown previously [13,16]. The bone marrow sample contained very few sequences with few or no mutations, which is somewhat surprising, since bone marrow contains many progenitor and precursor B cells, which are presumably un-mutated. The low frequency of un-mutated sequences is likely due to the abundance of plasma cell transcripts among the RNA used as template for antibody gene amplification. Since bone marrow resident long-lived plasma cells transcribe the antibody gene at a much higher rate than immature B cells, it is likely that oversampling of RNA transcripts derived from plasma cells skewed the bone marrow sequence repertoire toward highly mutated sequences. In this sense, the bone marrow antibody sequences reported here are more likely to reflect the composition of plasma antibody repertoires, which are composed mainly of

antibody produced by long-lived plasma cells, than a representation of the early developing B cell population.

All three mucosal tissue samples (small intestine, stomach and lung) showed a dramatically lower frequency of sequences with few or no mutations, suggesting that naïve B cells are less frequent in tissue. Interestingly, the lymphoid tissues (lymph node, tonsil, spleen and thymus) contained a higher frequency of antibody genes with few or no mutations (mean = 7.0%) than mucosal tissues (1.4%; $p = 0.03$), but a much lower frequency than circulating B cells in the peripheral blood (30.6%; Figure 4B). Here again, we observed a similarity between the bone marrow and mucosal tissue repertoires. As with $V(D)J$ gene use, the mutation frequency and abundance of un-mutated sequences in the bone marrow repertoire was similar to each of the mucosal tissue repertoires and differed substantially from repertoires in lymphoid tissue and peripheral blood.

A more detailed breakdown of mutation frequency by antibody gene region (Figure 4C) shows a reduction in mutation frequency in lymphoid tissue repertoires across all framework regions (FRs) and complementarity determining regions (CDRs) when compared to mucosal tissue repertoires ($p = 0.0002$).

Mucosal tissue repertoires encode longer HCDR3s than lymphoid tissue repertoires

Sequences from each tissue repertoire were grouped by HCDR3 length and the frequency of each HCDR3 length group was determined (Figure 5A). To facilitate comparisons, the HCDR3 length histogram for the peripheral blood repertoire is displayed alongside each tissue HCDR3 histogram. Tissue repertoires then were divided into two groups based on HCDR3 length: short HCDR3s (≤ 14 amino acids; shorter than the mean HCDR3 length in the peripheral blood repertoire) and long HCDR3s (≥ 15 amino acids; longer than the mean HCDR3 length in the peripheral blood repertoire) (Figure 5B). The repertoires of lymphoid tissues were approximately evenly split between short (50.7%) and long (49.3%) HCDR3 lengths. In contrast, repertoires from mucosal tissues contained a significantly higher frequency of long HCDR3s (58.4%) and thus a significantly lower frequency of short HCDR3s (41.6%) than lymphoid repertoires ($p = 0.02$). Further, the overall mean HCDR3 length of the mucosal tissue repertoires was significantly longer than the overall mean HCDR3 length of lymphoid tissue repertoires (Figure 5C; $p = 0.035$). This finding was surprising since mucosal repertoires contain a higher fraction of highly mutated sequences than lymphoid repertoires (Figure 4), but the highly mutated memory B cell subsets have been shown to encode shorter HCDR3s than the un-mutated naïve subset [16,28]. Since long HCDR3s tend to have a lower frequency of hydrophobic and charged residues [29–33], we determined the frequency of both hydrophobic and charged HCDR3 residues in mucosal and lymphoid repertoires (Figure 5D). While there was a trend toward reduced frequency of hydrophobic HCDR3 residues in the mucosal tissue repertoires compared to lymphoid repertoires ($p = 0.13$), there was no difference in the frequency of charged HCDR3 residues in mucosal repertoires (23.4%) compared to lymphoid repertoires (23.2%; $p = 0.61$).

Somatic hypermutation-associated insertions and deletions in peripheral blood, mucosal tissue and lymphoid tissue repertoires

Short nucleotide insertions or deletions have been shown to be associated with the somatic hypermutation process and antibodies encoding these somatic hypermutation-associated insertions and

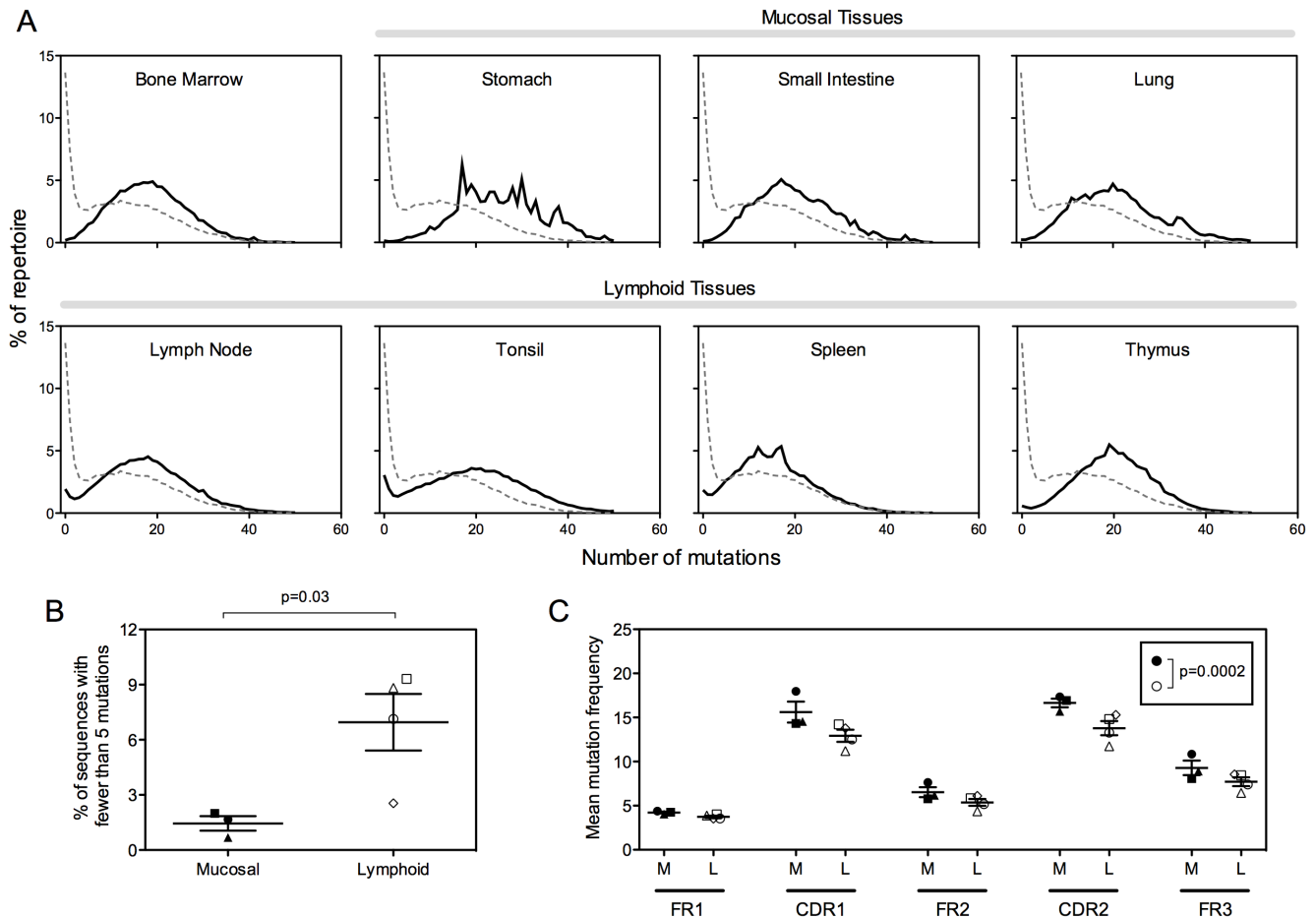


Figure 4. Mutation frequency for peripheral blood, bone marrow and mucosal and lymphoid tissues. (A) Mutation histograms are shown for each sample. Each of the three mucosal tissue samples (small intestine, stomach and lung) shows a complete loss of un-mutated sequences, which constitute a large portion of the peripheral blood repertoire. Repertoires for each of the lymphoid tissues (lymph node, tonsil, spleen and thymus) contained antibody genes with few or no mutations, but at a lower frequency than peripheral blood. For ease of comparison, the mutation distribution for peripheral blood is shown as a dashed line in each tissue plot. (B) The frequency of sequences with fewer than 5 mutations was determined for each mucosal and lymphoid tissue sample. For mucosal samples, lung, small intestine or stomach samples are plotted as filled circles, squares or triangles, respectively. For lymphoid samples, lymph node, spleen, thymus or tonsil are plotted as open circles, squares, triangles and diamonds, respectively. (C) The mean mutation frequency is shown for each genetic region of the variable gene: Framework Regions 1, 2 and 3 (FR1, FR2, FR3) and Complementarity Determining Regions 1 and 2 (CDR1, CDR2). Bars indicate mean \pm SEM for each group of tissue samples. Sample glyphs are as in (B). doi:10.1371/journal.pone.0100839.g004

deletions (SHA indels) have been shown to be critical to the immune response against pathogens that initiate infection at mucosal surfaces [34–40]. The heavy chain sequences of antibodies from all tissue repertoires were analyzed for the presence of codon-length nucleotide insertions or deletions in the antibody variable gene region. Sequences from the peripheral blood repertoire containing SHA indels were analyzed further to determine the position of each SHA indel, and the frequency of insertions (Figure 6) or deletions (Figure 7) at each codon position of the variable gene was calculated. For each mucosal or lymphoid tissue repertoire, the difference in SHA indel frequency, compared to peripheral blood, was calculated at each codon position. Insertions and deletions were both located predominantly in CDRs as opposed to framework regions, and were distributed roughly equally between heavy chain CDR1 and heavy chain CDR2 loops.

A large increase in frequency of both insertions and deletions in the small intestine antibody repertoire was immediately apparent. This observation was even more surprising since the frequency of

SHA indels has been shown to correlate with the frequency of somatic hypermutation events [41], but a corresponding increase in somatic mutations was not seen in the small intestine antibody repertoire (Figure 4A). Thus, the large increase in SHA indel frequency is likely due to specific, antigen-driven enrichment in the small intestine repertoire, not as an indirect result of a more general increase in somatic hypermutation events.

Discussion

The substantial differences discovered in mucosal tissue repertoires compared to peripheral blood and lymphoid tissue repertoires suggest that mucosal tissues contain B cells that express a unique, specialized repertoire of antibody genes, possibly to mount a more effective response to infections that are initiated at mucosal surfaces. The genetic composition of antibody repertoires encoded by circulating naive and memory B cells have been shown to differ, but the repertoires between naive and memory subsets were observed previously to be much more similar than expected.

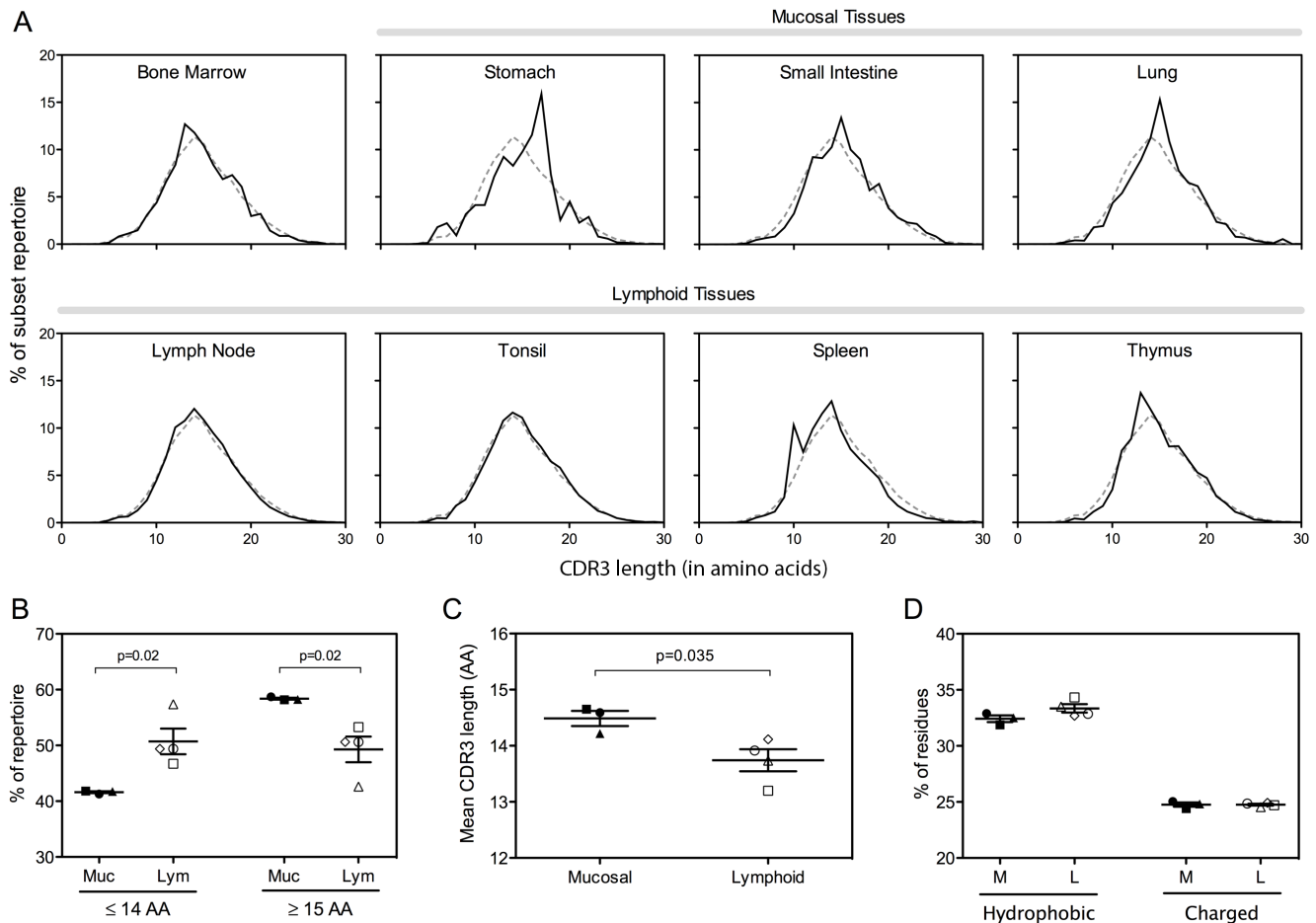


Figure 5. Mucosal tissue antibody gene repertoires encode longer CDR3s and are more mutated than lymphoid tissue repertoires. (A) Heavy chain CDR3 length histograms for each tissue sample. For ease of comparison, the CDR3 length distribution for peripheral blood is shown as a dashed line in each tissue plot. (B) Frequency of sequences containing short (14AA or shorter) or long (15AA or longer) CDR3s. Bars indicate mean \pm SEM for each group of samples. (C) The mean CDR3 length was determined for each tissue-specific repertoire. Bars indicate mean \pm SEM for each group of samples. For mucosal samples, lung, small intestine or stomach samples are plotted as filled circles, squares and triangles, respectively. For lymphoid samples, lymph node, spleen, thymus and tonsil or plotted as open circles, squares, triangles and diamonds, respectively. (D) Frequency of hydrophobic and charged CDR3 residues. Bars indicate mean \pm SEM for each group of samples. Samples glyphs are as in (C). doi:10.1371/journal.pone.0100839.g005

Further, memory repertoires have been shown to be very similar across individuals. While it is unclear how the consistency of the memory repertoire is maintained over the course of multiple pathogenic encounters that each induce biased antibody responses, we hypothesized that some fraction of the pathogen-induced memory population may be segregated into the mucosal tissue that represents the initial site of pathogen contact. To test this hypothesis, we performed a detailed genetic analysis of the expressed antibody repertoires of peripheral blood, bone marrow, and various mucosal and lymphoid tissues. Although we cannot define the root cause of the repertoire differences – likely due a combination of factors, including altered B cell subset composition, exposure to unique pathogenic challenges, and distinct antibody maturation mechanisms – our findings suggest that mucosal tissues harbor unique tissue-specific repertoires.

We found several genetic characteristics, including germline gene use, mutation frequency and HCDR3 length, for which the mucosal tissue repertoires substantially differed from lymphoid tissue and peripheral blood repertoires. When performing hierarchical clustering on the V(D)J use in each sample, we found that all mucosal tissue repertoires clustered together, separate from

peripheral blood and lymphoid tissue repertoires. Interestingly, the two mucosal tissues that would be considered most similar, the stomach and small intestine samples, were identified as the mucosal samples containing the most closely related antibody repertoire. Analysis of mutation frequency in each sample revealed the nearly complete absence of un-mutated sequences in the mucosal tissue repertoires, in stark contrast to the relatively high frequency of sequences with few or no mutations found in lymphoid tissue and peripheral blood repertoires. Finally, HCDR3 analysis revealed striking differences between mucosal and lymphoid tissues. Lymphoid tissues were most similar to the peripheral blood repertoire, with about half of each lymphoid tissue repertoire encoding HCDR3s that were as long or longer than the mean HCDR3 length in the peripheral blood repertoire (15 amino acids). In contrast, nearly 60% of sequences in each of the mucosal tissue repertoires were 15 amino acids in length or longer, indicating strong selection pressure for long HCDR3s in the mucosal repertoires. These data suggest that mucosal antibody repertoires may be tuned specifically to respond effectively to a subset of pathogens likely to be encountered at the mucosal interface.

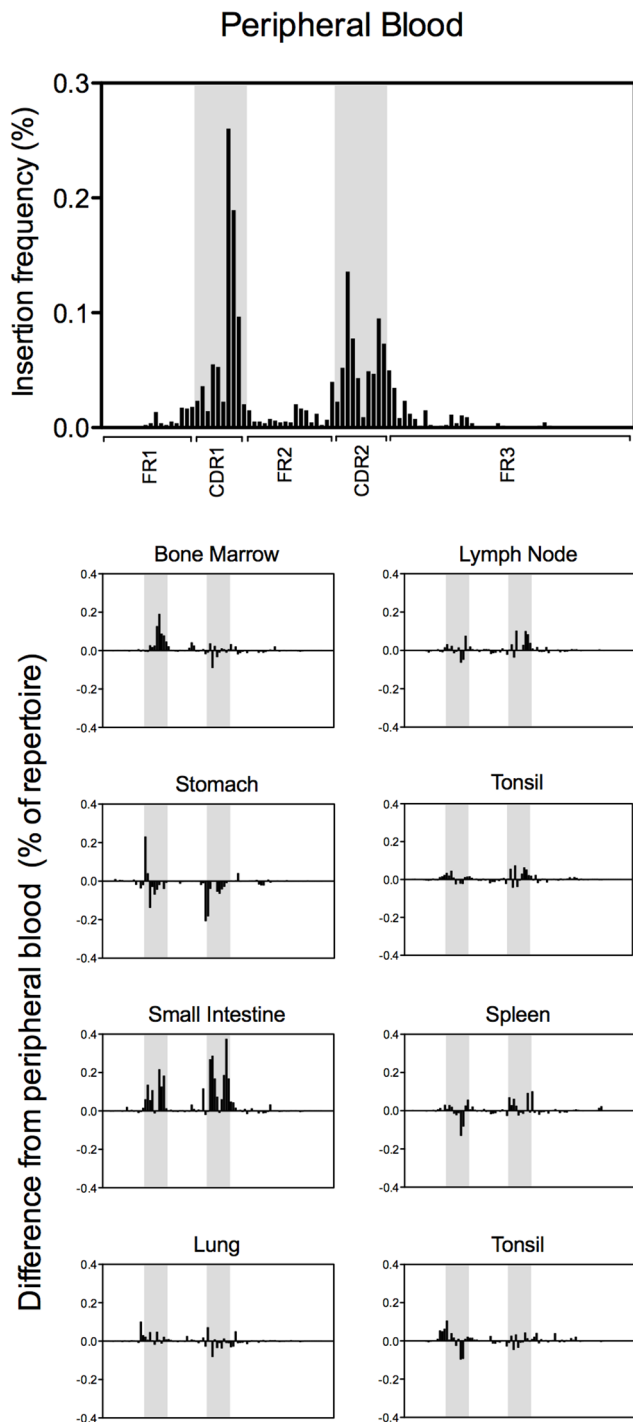


Figure 6. Frequency and position of DNA fragments encoding non-frameshift insertions. (A) The presence and frequency of non-frameshift insertions is shown for the peripheral blood repertoire. The frequency is plotted as the percent of sequences in the repertoire displaying deletions for each codon position in the variable gene. The location of CDR1 and CDR2 are highlighted in grey. (B) The difference in insertion frequency compared to peripheral blood is shown for each tissue. As in (A), CDR1 and CDR2 are highlighted in grey. doi:10.1371/journal.pone.0100839.g006

In some sense, comparing the genetic composition of these repertoires is akin to comparing apples and oranges: peripheral blood primarily consists of mature naïve B cells, while the tissue

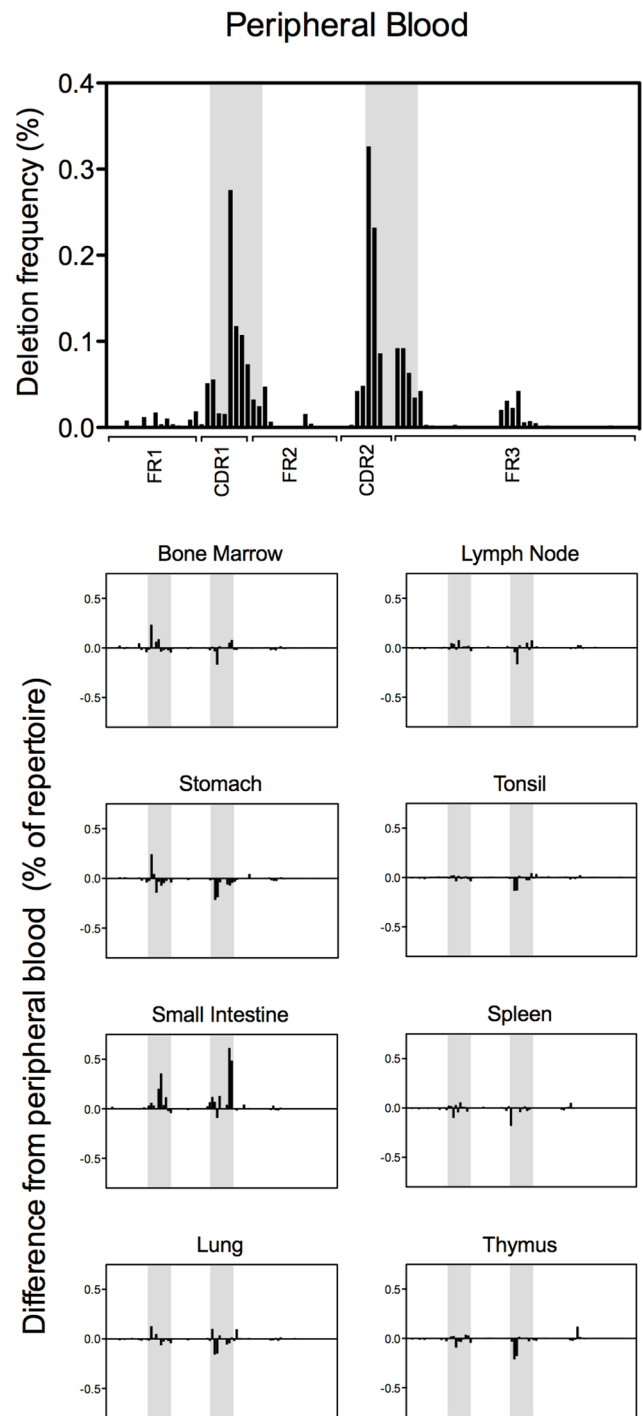


Figure 7. Frequency and position of DNA fragments encoding non-frameshift deletions. (A) The presence and frequency of non-frameshift deletions is shown for the peripheral blood subset. The frequency is plotted as the percent of sequences in the repertoire displaying deletions for each codon position in the variable gene. The location of CDR1 and CDR2 are highlighted in grey. (B) The difference in deletion frequency when compared to that in peripheral blood is shown for each tissue. As in (A), CDR1 and CDR2 are highlighted in grey. doi:10.1371/journal.pone.0100839.g007

samples likely contain a far higher fraction of memory B cells and antibody secreting cells. However, the observed differences cannot entirely be explained by the varying B cell subset proportions. Although variable gene use differed substantially between tissue

and peripheral blood samples, diversity and joining gene use was statistically indistinguishable, indicating the presence of selective pressure on variable gene use that extends beyond B cell subset composition. In addition, the two tissues that would be expected to contain the most similar proportion of B lineage cells, stomach and intestine, show striking differences that defy their presumably similar repertoire composition. If the differences between the tissue and peripheral blood repertoires were due primarily to the altered B cell subset contribution, it would be expected that known differences between the primary components of the peripheral blood repertoire (mature naïve B cells) and the tissue repertoires (memory B cells and plasmablasts) would be similar to those seen between the entire repertoires. For example, it has been repeatedly shown that memory B cells and plasmablasts have shorter HCDR3s than mature naïve B cells [16,27,28]. In the tissue repertoires, however, we saw the opposite: mucosal tissue repertoires, which would be expected to have shorter HCDR3s than peripheral blood due to increased memory B cell and plasmablast frequency, encoded significantly longer HCDR3s than the peripheral blood and lymphoid tissue repertoires. While it is clear that some of the observed differences are due to the dissimilar B cell subset makeup of tissue and peripheral blood samples, it is equally clear that the subset differences are insufficient to explain the entire difference.

While the differences between the mucosal tissue repertoires and peripheral blood were unexpected, more surprising were the observed similarities between mucosal tissue repertoires and the expressed repertoire of the bone marrow. At the cellular level, much of the bone marrow B cell population consists of progenitor and precursor B cells in the early stages of development. In contrast, long-lived plasma cells, which produce orders of magnitude more antibody transcript per cell than developing B cells, contribute a disproportionately large portion of the bone marrow antibody RNA pool. Because long-lived plasma cells are thought to produce the majority of soluble antibody proteins in serum and other body fluids, the genetic composition of the bone

marrow antibody mRNA pool would be expected to provide a reasonable estimation of the soluble plasma antibody repertoire. Our data, combined with recent proteomics evidence showing that the serum antibody repertoire differs from the antibody repertoire encoded by circulating B cells [42], suggest that genetic and functional antibody studies must expand their focus beyond the peripheral blood antibody repertoire. Although further work must be done to more completely define the antigen-specific nature of mucosal antibody repertoires, the work presented here provides a valuable step towards a more complete understanding of the entire humoral immune response.

Supporting Information

Table S1 Number and fold change of V(D)J combinations for each tissue repertoire sample for which the frequency differed statistically from the peripheral blood sample.

(PDF)

File S1 Suppl Info_PB_vs_Bone_marrow.csv.txt. Suppl Info_PB_vs_Lung.csv.txt. Suppl Info_PB_vs_Lymph_node.csv.txt. Suppl Info_PB_vs_Small_intestine.csv.txt. Suppl Info_PB_vs_Spleen.csv.txt. Suppl Info_PB_vs_Stomach.csv.txt. Suppl Info_PB_vs_Thymus.csv.txt. Suppl Info_PB_vs_Tonsil.csv.txt. (ZIP)

Acknowledgments

We would thank Chris L. Wright and Alvaro G. Hernandez at the W. M. Keck Center for Comparative and Functional Genomics at the University of Illinois at Urbana-Champaign for performing the 454 sequencing.

Author Contributions

Conceived and designed the experiments: BSB JEC. Performed the experiments: BSB JRW JAF BAM. Analyzed the data: BSB BAM JEC. Contributed reagents/materials/analysis tools: BAM. Wrote the paper: BSB JEC.

References

1. Tonegawa S (1983) Somatic generation of antibody diversity. *Nature* 302: 575–581.
2. Schatz DG (2004) V(D)J recombination. *Immunol Rev* 200: 5–11. doi:10.1111/j.0105-2896.2004.00173.x.
3. Alt FW, Oltz EM, Young F, Gorman J, Taccioli G, et al. (1992) VDJ recombination. *Immunology Today* 13: 306–314.
4. Neuberger MS (2008) Antibody diversification by somatic mutation: from Burnet onwards. *Immunol Cell Biol* 86: 124–132. doi:10.1038/sj.icb.7100160.
5. Milstein C, Neuberger M (1996) Maturation of the immune response. *Adv Protein Chem* 49: 451–485.
6. Neuberger MS, Milstein C (1995) Somatic hypermutation. *Curr Opin Immunol* 7: 248–254.
7. Adderson EE, Shackelford PG, Quinn A, Carroll WL (1991) Restricted Ig H chain V gene usage in the human antibody response to Haemophilus influenzae type b capsular polysaccharide. *J Immunol* 147: 1667–1674.
8. Tian C, Luskin GK, Dischert KM, Higginbotham JN, Shepherd BE, et al. (2008) Immunodominance of the VH1-46 antibody gene segment in the primary repertoire of human rotavirus-specific B cells is reduced in the memory compartment through somatic mutation of nondominant clones. *J Immunol* 180: 3279–3288.
9. Tian C, Luskin GK, Dischert KM, Higginbotham JN, Shepherd BE, et al. (2007) Evidence for preferential Ig gene usage and differential TdT and exonuclease activities in human naïve and memory B cells. *Mol Immunol* 44: 2173–2183. doi:10.1016/j.molimm.2006.11.020.
10. Weitkamp JH, Kallewaard NL, Bowen AL, LaFleur BJ, Greenberg HB, et al. (2005) VH1-46 is the dominant immunoglobulin heavy chain gene segment in rotavirus-specific memory B cells expressing the intestinal homing receptor alpha4beta7. *J Immunol* 174: 3454–3460.
11. Gorny MK, Wang X-H, Williams C, Volsky B, Revesz K, et al. (2009) Preferential use of the VH5-51 gene segment by the human immune response to code for antibodies against the V3 domain of HIV-1. *Mol Immunol* 46: 917–926. doi:10.1016/j.molimm.2008.09.005.
12. Zouali M (1996) Nonrandom features of the human immunoglobulin variable region gene repertoire expressed in response to HIV-1. *Appl Biochem Biotech* 61: 149–155.
13. Briney B, Willis JR, McKinney BA, Crowe JE (2012) High-throughput antibody sequencing reveals genetic evidence of global regulation of the naïve and memory repertoires that extends across individuals. *Genes Immun*. doi:10.1038/gene.2012.20.
14. Arnaout R, Lee W, Cahill P, Honan T, Sparrow T, et al. (2011) High-resolution description of antibody heavy-chain repertoires in humans. *PLoS ONE* 6: e22365. doi:10.1371/journal.pone.0022365.
15. Dekosky BJ, Ippolito GC, Deschner RP, Lavinder JJ, Wine Y, et al. (2013) High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nature biotechnology*. doi:10.1038/nbt.2492.
16. Wu Y-C, Kipling D, Leong HS, Martin V, Ademokun AA, et al. (2010) High-throughput immunoglobulin repertoire analysis distinguishes between human IgM memory and switched memory B-cell populations. *Blood* 116: 1070–1078. doi:10.1182/blood-2010-03-275859.
17. Boyd SD, Marshall EL, Merker JD, Maniar JM, Zhang LN, et al. (2009) Measurement and clinical monitoring of human lymphocyte clonality by massively parallel VDJ pyrosequencing. *Sci Transl Med* 1: 12ra23.
18. Boyd SD, Gaëta BA, Jackson KJ, Fire AZ, Marshall EL, et al. (2010) Individual variation in the germline Ig gene repertoire inferred from variable region gene rearrangements. *J Immunol* 184: 6986–6992. doi:10.4049/jimmunol.1000445.
19. Cerutti A, Chen K, Chorny A (2011) Immunoglobulin responses at the mucosal interface. *Annu Rev Immunol* 29: 273–293. doi:10.1146/annurev-immunol-031210-101317.
20. Fagarasan S, Honjo T (2003) Intestinal IgA synthesis: regulation of front-line body defences. *Nat Rev Immunol* 3: 63–72. doi:10.1038/nri982.
21. Suzuki K, Maruya M, Kawamoto S, Sitnik K, Kitamura H, et al. (2010) The sensing of environmental stimuli by follicular dendritic cells promotes immunoglobulin A generation in the gut. *Immunity* 33: 71–83. doi:10.1016/j.immuni.2010.07.003.

22. Holtmeier W, Hennemann A, Caspary WF (2000) IgA and IgM V(H) repertoires in human colon: evidence for clonally expanded B cells that are widely disseminated. *Gastroenterology* 119: 1253–1266.
23. van Dongen JJM, Langerak AW, Brüggemann M, Evans PAS, Hummel M, et al. (2003) Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia* 17: 2257–2317. doi:10.1038/sj.leu.2403202.
24. Lefranc M-P, Giudicelli V, Ginesstoux C, Jabado-Michaloud J, Folch G, et al. (2009) IMGT, the international ImMunoGeneTics information system. *Nucl Acids Res* 37: D1006–D1012. doi:10.1093/nar/gkn838.
25. Alamyar E, Giudicelli V, Li S, Duroux P, Lefranc M-P (2012) IMGT/HighV-QUEST: the IMGT web portal for immunoglobulin (IG) or antibody and T cell receptor (TR) analysis from NGS high throughput and deep sequencing. *Immunome Res* 8: 26.
26. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139–140. doi:10.1093/bioinformatics/btp616.
27. Benckert J, Schmolka N, Kreschel C, Zoller MJ, Sturm A, et al. (2011) The majority of intestinal IgA+ and IgG+ plasmablasts in the human gut are antigen-specific. *J Clin Invest* 121: 1946–1955. doi:10.1172/JCI44447.
28. Briney B, Willis JR, Crowe JE (2012) Human peripheral blood antibodies with long HCDR3s are established primarily at original recombination using a limited subset of germline genes. *PLoS ONE* 7: e36750. doi:10.1371/journal.pone.0036750.
29. Aguilera I, Melero J, Nuñez-Roldan A, Sanchez B (2001) Molecular structure of eight human autoreactive monoclonal antibodies. *Immunology* 102: 273–280.
30. Wardemann H, Yurasov S, Schaefer A, Young JW, Meffre E, et al. (2003) Predominant autoantibody production by early human B cell precursors. *Science* 301: 1374–1377. doi:10.1126/science.1086907.
31. Crouzier R, Martin T, Pasquali JL (1995) Heavy chain variable region, light chain variable region, and heavy chain CDR3 influences on the mono- and polyreactivity and on the affinity of human monoclonal rheumatoid factors. *J Immunol* 154: 4526–4535.
32. Haynes BF, Fleming J, St Clair EW, Katinger H, Stiegler G, et al. (2005) Cardiolipin polyspecific autoreactivity in two broadly neutralizing HIV-1 antibodies. *Science* 308: 1906–1908. doi:10.1126/science.1111781.
33. Zemlin M, Schelonka RL, Ippolito GC, Zemlin C, Zhuang Y, et al. (2008) Regulation of repertoire development through genetic control of DH reading frame preference. *J Immunol* 181: 8416–8424.
34. Wilson PC, Liu YJ, Banchemreau J, Capra JD, Pascual V (1998) Amino acid insertions and deletions contribute to diversify the human Ig repertoire. *Immunol Rev* 162: 143–151.
35. Wilson PC, de Bouteiller O, Liu YJ, Potter K, Banchemreau J, et al. (1998) Somatic hypermutation introduces insertions and deletions into immunoglobulin V genes. *J Exp Med* 187: 59–70.
36. Briney B, Willis JR, Crowe JE (2012) Location and length distribution of somatic hypermutation-associated DNA insertions and deletions reveals regions of antibody structural plasticity. *Genes Immun*. doi:10.1038/gene.2012.28.
37. Krause JC, Ekiert DC, Tumpey TM, Smith PB, Wilson IA, et al. (2011) An insertion mutation that distorts antibody binding site architecture enhances function of a human antibody. *MBio* 2: e00345–10. doi:10.1128/mBio.00345-10.
38. Wu X, Yang Z-Y, Li Y, Hogerkorp C-M, Schief WR, et al. (2010) Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science* 329: 856–861. doi:10.1126/science.1187659.
39. Walker LM, Huber M, Doores KJ, Falkowska E, Pejchal R, et al. (2011) Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature* 477: 466–470. doi:10.1038/nature10373.
40. Walker LM, Phogat SK, Chan-Hui P-Y, Wagner D, Phung P, et al. (2009) Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. *Science* 326: 285–289. doi:10.1126/science.1178746.
41. Wilson PC, de Bouteiller O, Liu Y, Potter K, Banchemreau J, et al. (1998) Somatic hypermutation introduces insertions and deletions into immunoglobulin genes. *J Exp Med* 187: 59–70.
42. Wine Y, Boutz DR, Lavinder JJ, Miklos AE, Hughes RA, et al. (2013) Molecular deconvolution of the monoclonal antibodies that comprise the polyclonal serum response. *Proc Natl Acad Sci USA*. doi:10.1073/pnas.1213737110.