3 Evolution of immunoglobulin heavy chain repertoires throughout SIV infection

3.1 Introduction and aims

The antigen-specific humoral response to a pathogen develops rapidly after infection, and through the iterative processes of somatic hypermutation and affinity maturation, generates a highly specific pool of antibody secreting cells and memory B cells. Next-generation sequencing of antibodies offers the ability to sample and sequence tens of thousands of B cells throughout infection[194], and combined with serial sampling across tissues yields a detailed view of the evolution of the antigen-specific antibody repertoire. Since the discovery of the first broadly neutralizing antibodies (bnAbs) to HIV in the early 1990s, hundreds of HIV-specific antibodies ranging in breadth and potency have been catalogued and studied. Certain features are highly represented within the bnAbs: prominent use of germline V genes including VH1-2, VH1-69 and VH1-4, long HCDR3 regions (24-32 residues), and high levels of mutation from germline V genes ranging from 20-30% or higher)[132,219,228]. High-throughput sequencing of the antibody repertoire and phylogenetic analysis have also been used to identify VRC01 class antibodies that bind epitopes in the CD4 binding site of HIV envelope[338]. Here, we use the SIV-rhesus macaque model to identify features of the maturing memory B cell response to SIV during on-going viral infection, examining V-D-J gene usage, mutations and features of heavy chain CDR3 regions, and the diversity and clonality of the memory response.

The aims of this chapter are to:

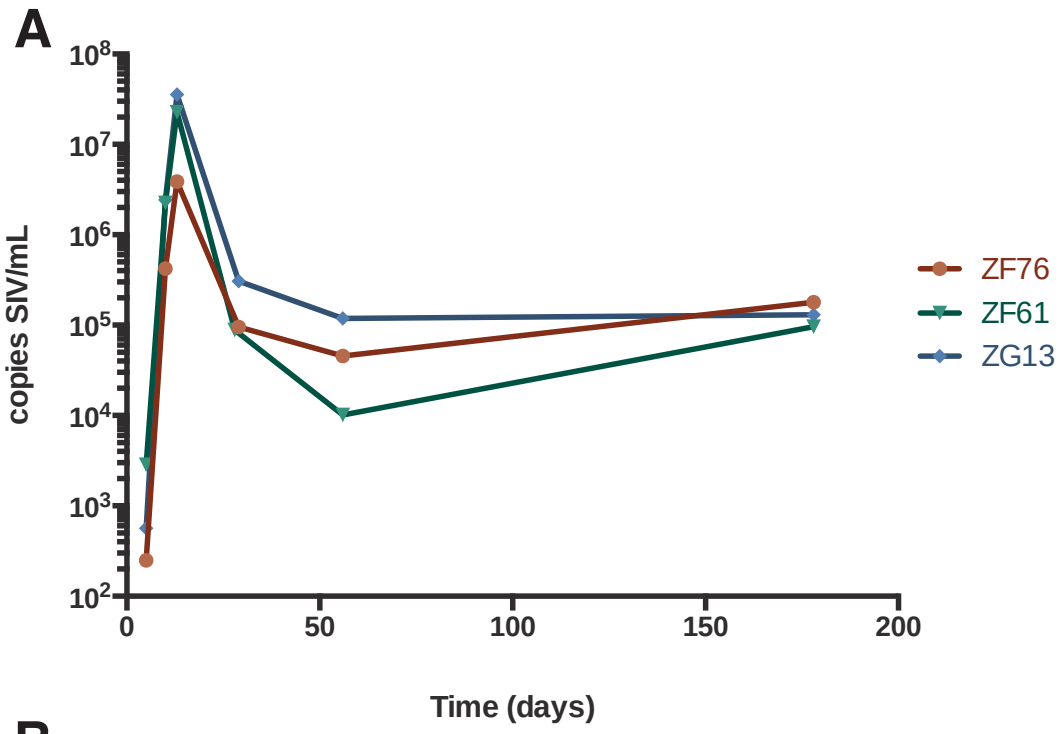
- Measure the proportion of SIV-specific B cells in lymph nodes, bone marrow, and peripheral blood throughout infection
- Track the changes in genetic properties of SIV-specific B cells change over the course of infection, and compare IgG and IgM SIV-specific antibodies
- Compare the physiochemical features of SIV-specific and bulk memory B cells
- Trace antibody linages across tissues and over time, and measure public versus private repertoires
- Compare 5’ RACE repertoire sequencing to single cell primer based methods

3.2 Results

3.2.1 Infection Model and plasma neutralization

Three rhesus macaques (Indian, colony-bred) were infected with intravenous dose of SIV mac251 at a 1:3000 dilution. Blood draws and lymph node and bone marrow biopsies were taken at 4 weeks, 10 weeks, and 24 weeks post infection. Plasma viremia peaked at 13 days post infection ($3.9x10^6$, $2.3x10^7$, and $3.5x10^7$ copies/mL) (Fig 3.1 A). Viral loads were steady at the three time points sampled, with $9.5x10^4$, $8.8x10^4$, and $3.1x10^5$ copies/mL at 4 weeks; $4.5x10^4$, $1.0x10^4$, and $1.2x10^5$ copies/mL at 10 weeks, $1.8x10^5$, $9.6x10^4$, and $1.3x10^5$ copies/mL at 24 weeks. These values are within the expected ranges of plasma viremia during peak infection, viral setpoint, and on-going replication in the absence of antiretroviral therapy[339].

Plasma neutralization was measured using a TZM-BL assay[223] with serial dilutions of plasma, with ID_{50} values representing reciprocal inhibitory dilution (Fig. 3.1 B). The SIVsmE660 isolate used here is a tier 1A envelope (meaning it is highly sensitive to neutralization) while the SIVmac251 envelope in this assay is Tier 2 (moderate resistance to neutralization). Three animals used for immunoglobulin repertoire analysis showed moderate neutralization of SIVsmE660 at 4 weeks post infection (ID_{50} of $5.8x10^3$ -$1.9x10^4$) and no neutralization of SIVmac251 (the virus used for infection) at 4 weeks. By 24 weeks post infection, all animals showed higher neutralization of SIVsmE660 (ID_{50} of $1.7x10^5$ to above $3.3x10^5$, the limit of detection in this assay) and low-level neutralization of SIVmac251 (ID_{50} of 58-746). Previous studies have shown that SIVmac251 founder viruses induce autologous neutralizing antibodies at 5 to 8 months post infection[273], while neutralization of sensitive viruses such as smE660 typically develops earlier[340]. Thus, the levels of plasma neutralization of SIVmac251 and smE660 at 24 weeks are evidence of a typical humoral response in the animals in this study.



B

Animal ID	Infection period (weeks)	ID_{50} SIVsmE660.CP3C-P-A8 (tier 1A)	ID_{50} SIVmac251.30 (tier 2)
ZF61 4 weeks	4	5,888	<20
ZF76 4 weeks	4	6,010	<20
ZG13 4 weeks	4	19,493	<20
ZF61 24 weeks	24	>327,680	273
ZF76 24 weeks	24	176,279	58
ZG13 24 weeks	24	>327,680	746

Values represent reciprocal inhibitory dilution at 50% (ID_{50}) or 80% (ID_{80}) neutralization

Note: Neutralization curves for SIVmac251.30 exhibited a plateau at approx. 40% neutralization

Fig. 3.1 Plasma viremia and neutralization. (A) Plasma viral loads for three animals (ZF76, ZF61, and GZ13) throughout infection. (B) Plasma neutralization of SIVsmE660 and SIVmac251 at 4 weeks and 24 weeks post infection. Good neutralization (high reciprocal inhibitory dilution) indicated in red, moderate neutralization in yellow, and no detectable neutralization in white.

3.2.2 Sorting of SIV-specific B cell throughout infection

To evaluate the SIV-specific repertoire of immunoglobulins generated in response to infection, we used a trimeric SIVgp140 protein probe to sort SIV-specific B cells from live cell samples of PBMCs, lymph nodes, spleen, and bone marrow cells[328]. The SIVmac239 gp140 foldon trimer probe used in this assay to isolate SIV-specific B cells with FACS has been previously validated and published [328]. All cells were sorted based on staining as Aqua (live/dead cell marker) CD14 (monocytes) CD3 (T cells) negative and CD19 CD20 positive B cells (Fig. 3.2 A). SIV gp140 protein probe-positive and protein probe-negative class switched IgG memory B cells (IgM- IgD- IgG+ CD27+) and probe-positive and probe-negative memory IgM B cells (IgM+ IgG- CD27+) were sorted at 4, 10 and 24 weeks post infection from all tissues (Fig 3.2 B).

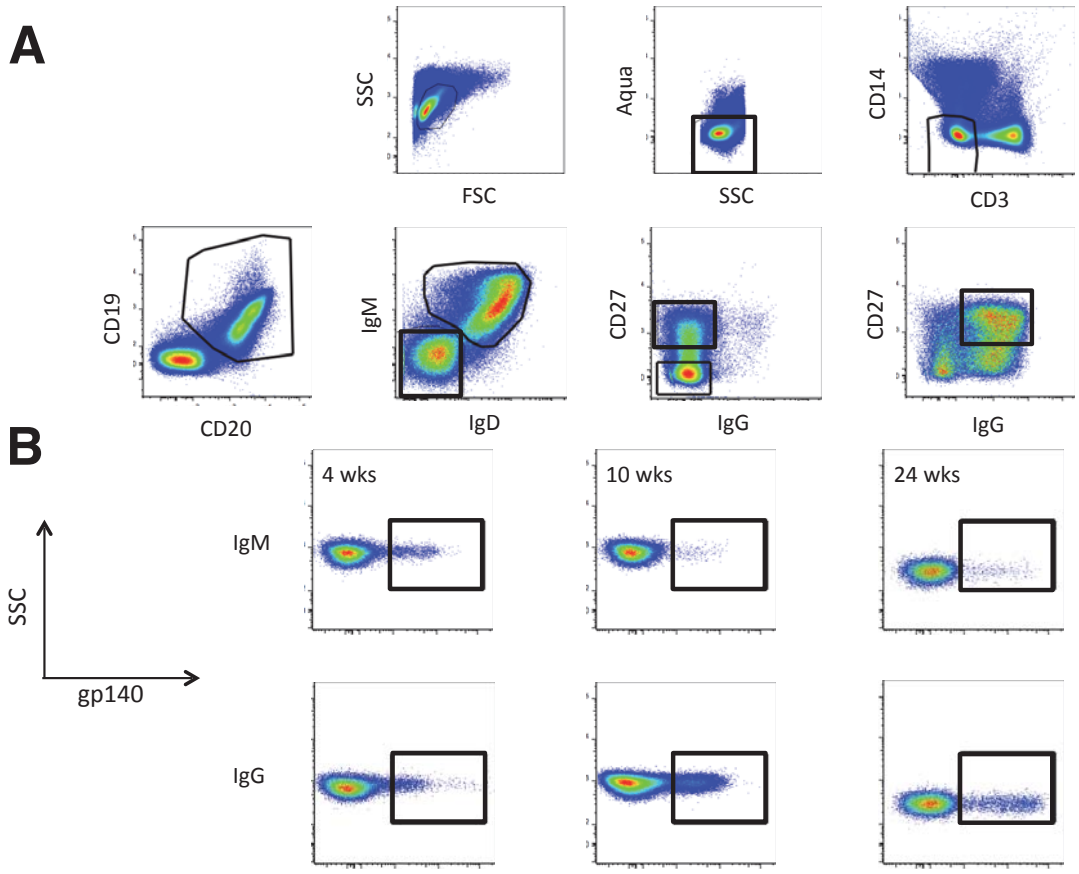


Fig 3.2 Sorting schematic and SIVgp140 probe staining (A) Gating strategy for isolating B cells from lymph nodes, PBMC, bone marrow, and spleen. SIV-specific and non-specific B cells were sorted after gating on lymphocytes, live (Aqua negative), CD14- CD3- CD19+ CD20+ B cells, and either IgM+IgD+CD27+ (memory IgM) or IgM-IgD-CD27+IgG+ (memory IgG). (B) SIVgp140 probe staining on memory IgM (IgM+IgD+IgG-CD27+) top trow, and class switched IgG memory B cells (IgM-IgD-IgG+CD27) bottom row, at 4 weeks, 10 weeks, and 24 weeks post infection.

The frequencies of naïve B cells measured as a percentage of total B cells were unchanged over the 24 weeks of infection studied here in bone marrow and in PBMC (Fig. 3.3 A, B), but in lymph nodes dropped significantly in all animals at 10 weeks post infection to a mean of 27% before increasing at 24 weeks to 44% (Fig. 3.3). IgM CD27+ memory cells in PBMC decreased in all animals, from a mean of 29% to 15% by 24 weeks (Fig. 3.3 A). The percentage of IgM CD27+ memory cells was unchanged in lymph nodes (7%-15%) and declined, though not statistically significantly, in bone marrow from 5.6% to 0.39% of total B cells. The proportion of IgG CD27+ B cells increased in PBMC at 10 weeks post infection of 2 of the three animals (to a mean of 45% of total B cells), but declined by 24 weeks post infection (Fig. 3.3 A). Lymph node IgG CD27+ B cells were statistically unchanged over the infection period studied here, while bone marrow IgG CD27+ cells declined from 7.3% to 0.15% of all B cells.

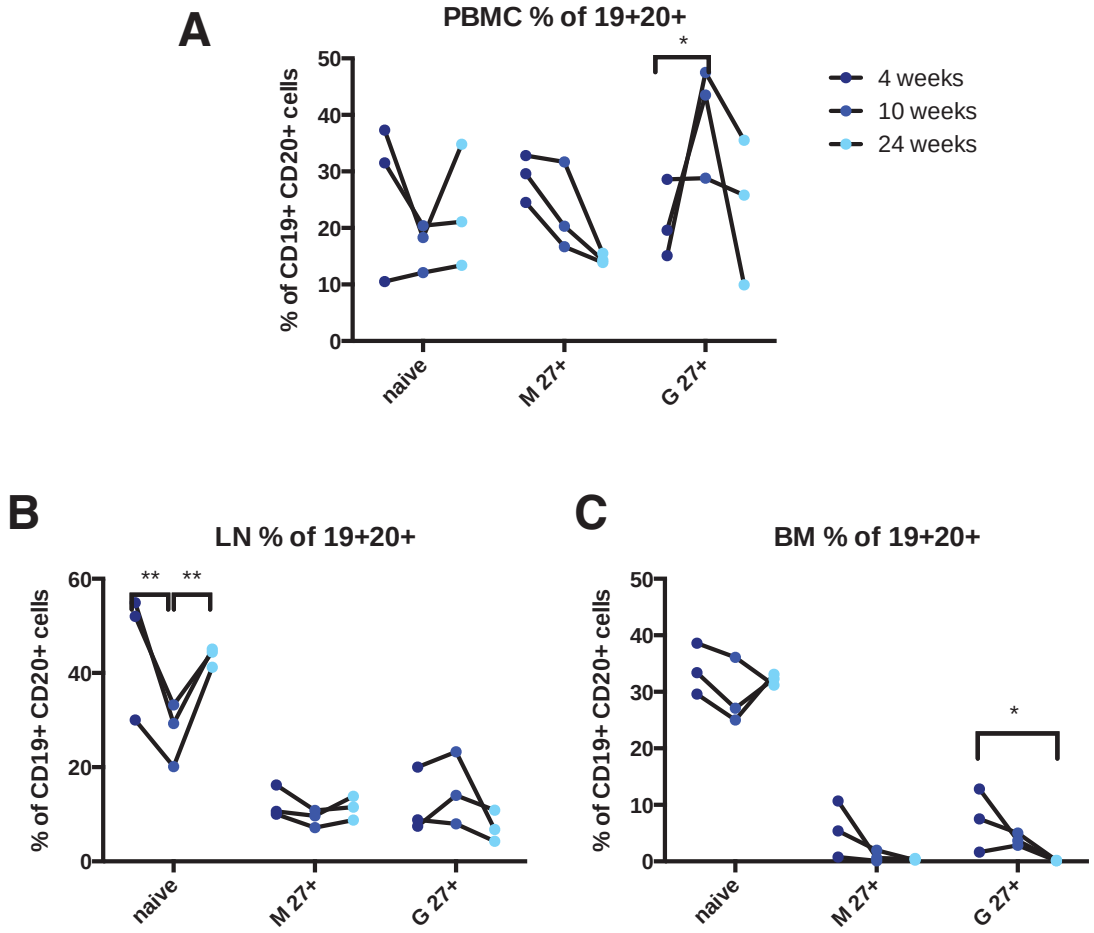


Fig 3.3 Proportions of B cell subpopulations. (A) Naïve, CD27+ IgM memory, and CD27+ IgG memory B cells in PBMC at 4 weeks, 10 weeks, and 24 weeks post infection as a proportion of all B cells. (B) Naïve, CD27+ IgM memory, and CD27+ IgG memory B cells in inguinal lymph nodes at 4 weeks, 10 weeks, and 24 weeks post infection as a proportion of all B cells (C) Naïve, CD27+ IgM memory, and CD27+ IgG memory B cells in bone marrow at 4 weeks, 10 weeks, and 24 weeks post infection as a proportion of all B cells. Significant differences were determined using a paired Student's T-test.

The proportion of SIV-specific B cells increased throughout infection. Measured as a percentage of IgG memory B cells, SIV protein probe positive cells increased (though not statistically significantly) in PBMC and lymph nodes in all three animals, from a mean of 2.5% to 8.6% in PBMC and 5.3% to 12.3% in lymph nodes (Fig. 3.4 A). In bone marrow, one animal had an increase in IgG SIV-specific cells, one animal showed no change, and another animal had an increase in SIV-specific cells at 10 weeks but a lower proportion at 24 weeks than at 4 weeks. The proportion of IgM SIV-specific cells was unchanged in lymph nodes and PBMC (ranging from 0.2% to 1.2% in PBM and 0.9%-1.8% in lymph nodes), while in bone marrow one animal had an increase in SIV-specific cells at 10 weeks followed by a decrease at 24 weeks (the same animal that had a peak in IgG SIV-specific cells) (Fig. 3.4 B)

When measured as a percentage of total B cells, SIV-specific IgG cells were measured at a significantly higher proportion of PBMC in two of three animals (Fig 3.4 C), with a mean of 0.5% at 4 weeks and 2.5% at 24 weeks. In lymph nodes, the pattern of the proportion of SIV-specific cells was the same in all three animals, increasing from a mean of 0.5% at 4 weeks to 1.7% at 10 weeks before decreasing to 0.8% at 24 weeks. The proportion of SIV specific IgG cells in bone marrow declined in all animals, from a mean of 0.6% at 4 weeks to 0.02% at 24 weeks. IgM SIV-specific cells declined slightly in all animals, from a mean of 0.31% at 4 weeks to 0.17% at 24 weeks (with a dip at 10 weeks to 0.04%) in PBMC (Fig. 3.4 D). Lymph node ISV-specific IgM cells declined slightly in two animals (from 0.1% to 0.06%) but increased in one animal to 0.6% at 24 weeks. Bone marrow SIV-specific IgM B cells similarly declined in two animals to 0.004%, but increased to 0.05% in one animal (the same animal that had an increase in lymph node IgM SIV-specific cells. Overall, the SIV gp140-specific response is dominated by IgG B cells when measured by absolute quantity of B cells and proportion of memory populations but has contributions, especially in bone marrow, from IgM memory cells.

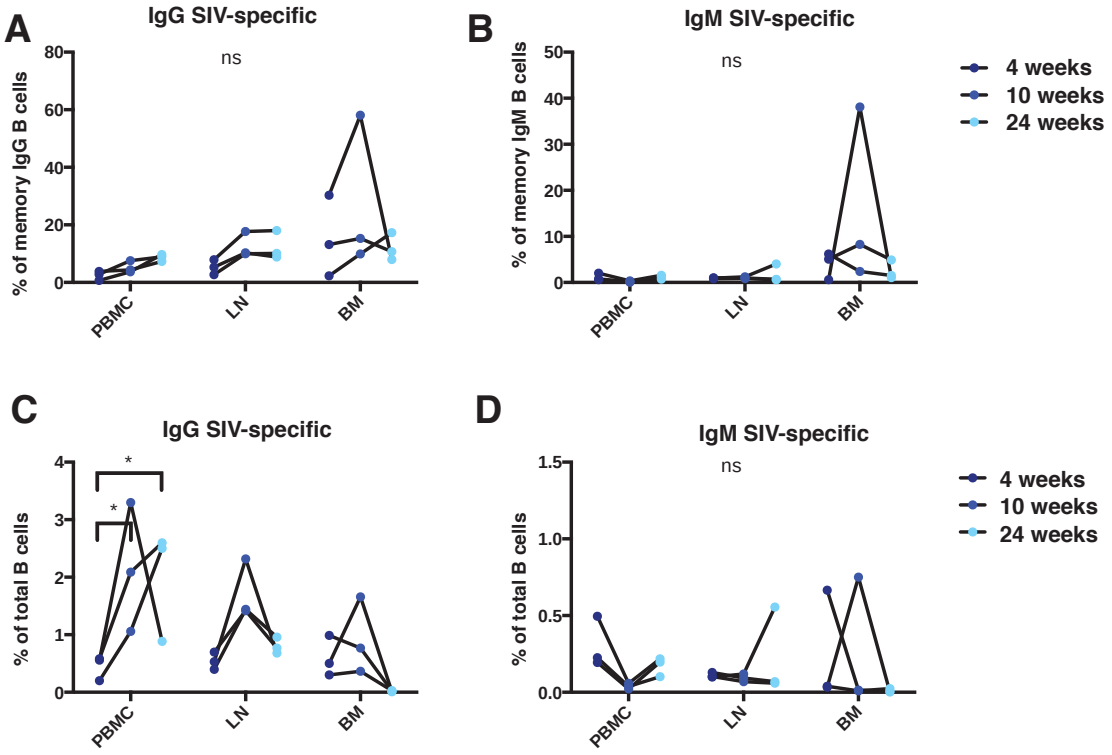


Fig 3.4 Antigen-specific B cells throughout infection. (A) Antigen-specific IgG B cells in lymph nodes, bone marrow and PBMC as a proportion of memory IgG B cells at 4, 10, and 24 weeks post infection. (B) Antigen-specific IgM B cells in lymph nodes, bone marrow and PBMC as a percentage of memory IgM B cells at 4, 10, and 24 weeks post infection. (C) Antigen-specific IgG B cells in lymph nodes, bone marrow and PBMC as a percentage of total B cells at 4, 10, and 24 weeks post infection. (D) Antigen-specific IgM B cells in lymph nodes, bone marrow and PBMC as a percentage of total B cells at 4, 10, and 24 weeks post infection. Significant differences were determined using a paired Student's T-test.

3.2.3 Next-generation sequencing of memory, naïve, and antigen-specific B cells

Quantity alone of antigen-specific cells does not determine the quality (including the breath, specificity, and diversity of epitopes seen) of the humoral response. To further evaluate the immunoglobulin heavy chain repertoire generated in response to infection and to compare the SIV-specific response to the non-SIV-specific memory cells, we extracted RNA from sorted cell populations and amplified immunoglobulin heavy chains using 5' RACE PCR. Briefly, after total cDNA synthesis from mRNA and IgH-specific amplification with constant region primers (either IgM, or IgG), samples were sequenced on the Illumina MiSeq platform using 2x300 base paired end reads. To analyse the sequencing results, a bioinformatics pipeline was adapted from a published pipeline[332] used for BCR analysis of blood disorders and adapted for 5' RACE sequencing methods and rhesus macaque genes (Fig 3.5). After quality control with QUASR7.0[330] and primer trimming, reads were joined using overlapping regions to obtain full length heavy chain sequences. Joined sequences of minimum length 200 underwent filtering for V-gene and J-genes identification, productive open reading frames, and were subsequently collapsed into identical reads to yield a final set of unique heavy chain sequences. Reads were only retained if they had a median read quality (Phred score) of 34, overlapped with the matched paired read by at least 30 identical base pairs, matched a reference database of V genes by nucleotide BLAST with a e-value of 1 and J genes with an e-value of 0.1, and had a productive open reading frame. V, D, and J genes were mapped to a recently generated draft rhesus macaque Ig heavy chain reference database [341]. E-values describe random background noise, with higher e-values (closer to 1) allowing a lower threshold for a match, while high e-values (close to 0) ensure more stringent matches. In this case, higher e-values (1 and 0.1) were used to allow highly mutated J genes to be matched to the nearest reference sequence without discarding highly mutated sequences. A median of 45% of reads retained after quality control, joining, primer matching, and open reading frame filtering, though there was some variability in the quality of the MiSeq runs due to low-quality kits. In high-quality sequencing runs, samples had a median of 45% of reads retained (interquartile range of 38%-51%), while in low-quality sequencing runs, a median of 3.1% of reads were retained (IQR of 0.5%-6.6%). If the number of retained reads did not exceed the number of cells sorted, samples were re-amplified from cDNA or mRNA and re-sequenced to obtain coverage of the entire sample. A list of the samples sorted, the number of cells from each sample, the raw number of sequencing reads obtained for that sample, and the number of reads retained after each step of filtering is detailed in the Materials and Methods Tables 2.14 and 2.15.

The number of cells sorted from each population varied widely, with several hundred SIV-specific IgG and IgM B cells isolated from acute PBMC and lymph node samples, and over 200,000 cells sorted from certain non-specific memory B cell samples. Low cell number samples often required additional PCR cycles to meet the DNA threshold for sequencing, and in many cases had a lower percentage of reads retained in the final analysis. In calibrating the read processing and quality control pipeline, there is a balance between retaining what could be genuine reads that happen to fall below a strict quality threshold and discarding lower-quality reads that may in fact be biologically relevant. Even with strict quality control measures, there were still samples from which the final number of unique reads exceeded the absolute input number of cells. Other groups have noted this in their immunoglobulin sequencing[342], which could be attributed to contamination, chimeric PCR products formed during amplification, or imprecision in the cell sorter counting. However, the vast majority of samples had equal or fewer unique heavy chain sequencing after filtering and quality control. Nonetheless, caution must be used in interpreting the results when only small numbers of cells are isolated from a given population, with the knowledge that the sequences represent a small subset of the total repertoire.

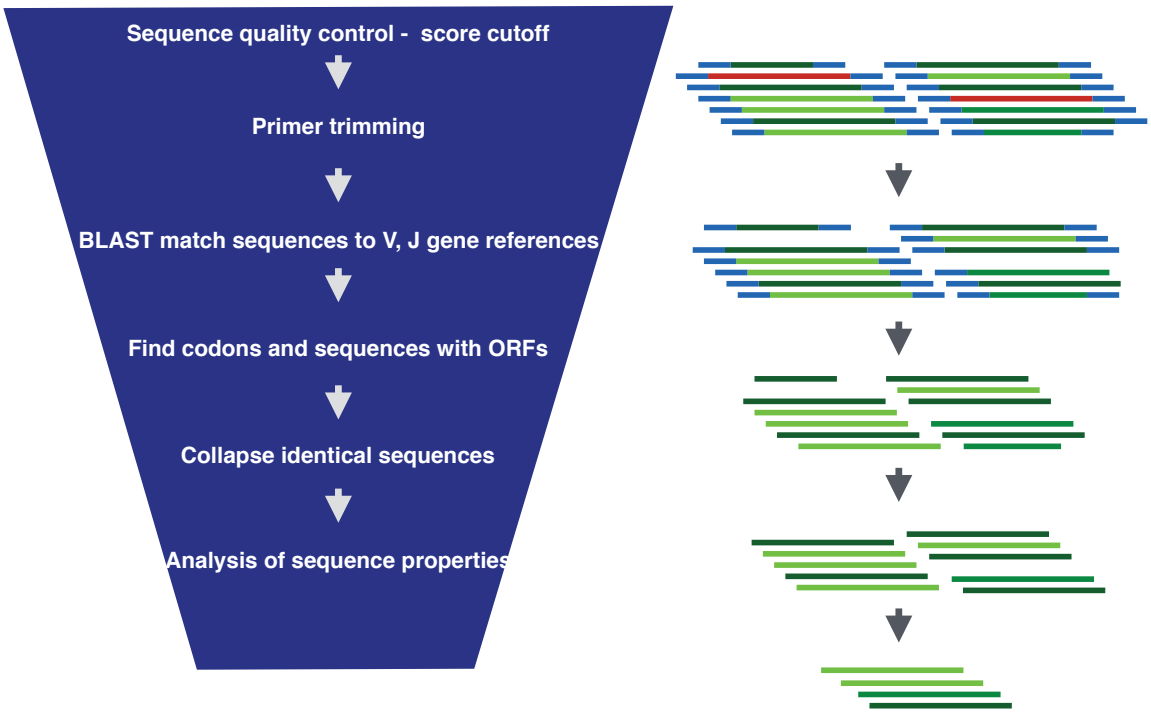


Fig. 3.5 Schematic of BCR Bioinformatics Pipeline for Sequence Analysis. Sequencing reads were filtered from the analysis if they did not meet a minimum PHRED (quality of sequencing) score, had primers trimmed, were BLAST matched to reference genomes (and eliminated if there was not a match, allowing for mutation from germline), and checked for open reading frames (with truncated sequences and sequences with early stop codons eliminated). Identical sequences were collapsed to yield a set of unique, high quality immunoglobulin heavy chains from each sample.

3.2.4 Diversity and evolution of IgG B cells

To examine the diversity and breadth of the response to infection, we analysed several properties of heavy chain immunoglobulin sequences. Long heavy chain CDR3 regions have been strongly associated with several broadly neutralizing antibodies to HIV, interacting with the binding pocket in the CD4 binding site and penetrating the glycan shield to access conserved epitopes[343]. In total IgG memory B cells, median CDR3 length was the statistically unchanged at all three time points, with median CDR3 length of 12-14 amino acids (standard deviation of 5.2-5.7) in bone marrow, 11-12 amino acids (standard deviation of 5.0-5.4) in lymph nodes, and 12 amino acids (standard deviation of 4.7-5.2) in PBMC (Fig. 3.6 A,B, C). In PBMC, there was no difference between SIV-specific and total CDR3 sequences at 4 weeks ($p = 0.09$), while at 10 weeks and 24 weeks SIV-specific and total memory median CDR3 lengths were significantly different ($p<0.0001$) (Table 3.1). Median length in SIV-specific sequences in PBMC was 11 amino acids at 4 weeks and 10 weeks (standard deviation of 4.6 and 4.8, respectively) but increased significantly ($p<0.0001$) to 13 amino acids (standard deviation 4.9) by 24 weeks (Fig. 3.6 F). In lymph nodes and in bone marrow, CDR3 lengths of SIV-specific sequences were significantly different from total IgG memory sequences at all three time points ($p <0.0001$) (Fig. 3.6 D, E; Table 3.1).

In comparing SIV-specific sequences at 4, 10, and 24 weeks, we measured the proportion of long CDR3s (defined as 18 amino acids or longer). In lymph nodes, PBMC, and bone marrow, long CDR3s (18 AA and longer) made up 0.5%-1.1% of all sequences with no significant difference between time points, while antigen-specific sequences had anywhere from 0.2%-1.9% of CDR3s 18 AA or longer (Table 3.1). Previous studies have shown that there is no difference between the total memory and Env-specific median CDR3 length, which is in accordance with the data presented here[344].

Table 3.1 CDR3 Lengths and VH mutation in IgG samples

Time Point (weeks)	Tissue	Population	IGH	**Median CDR3**	**CDR3 standard deviation**	**Median VH mutation**	**VH mutation standard deviation**	**Percentage of CDR3s 18AA or longer**
4	BM	gp140	IgG	11	4.70	3.47	4.30	0.23
10	BM	gp140	IgG	12	4.85	3.78	3.60	1.12
24	BM	gp140	IgG	12	4.77	6.94	3.09	0.48
4	LN	gp140	IgG	11	5.09	3.44	3.01	0.93
10	LN	gp140	IgG	11	5.31	2.75	2.32	1.92
24	LN	gp140	IgG	11	4.76	6.8	2.32	0.34
4	PBMC	gp140	IgG	11	4.47	3.82	2.90	0.46
10	PBMC	gp140	IgG	11	4.81	3.82	2.58	1.02
24	PBMC	gp140	IgG	13	4.89	6.25	2.60	0.61
4	BM	mem	IgG	12	5.27	6.6	4.49	0.67
10	BM	mem	IgG	13	5.19	4.83	4.08	1.11
24	BM	mem	IgG	12	5.67	6.87	2.30	0.78
4	LN	mem	IgG	12	5.32	5.17	3.39	0.80
10	LN	mem	IgG	11	5.02	5.15	3.78	0.57
24	LN	mem	IgG	12	5.36	5.9	3.71	0.55
4	PBMC	mem	IgG	12	4.77	3.74	4.09	0.69
10	PBMC	mem	IgG	12	4.92	3.18	3.49	0.88
24	PBMC	mem	IgG	12	5.21	5.14	2.74	0.91

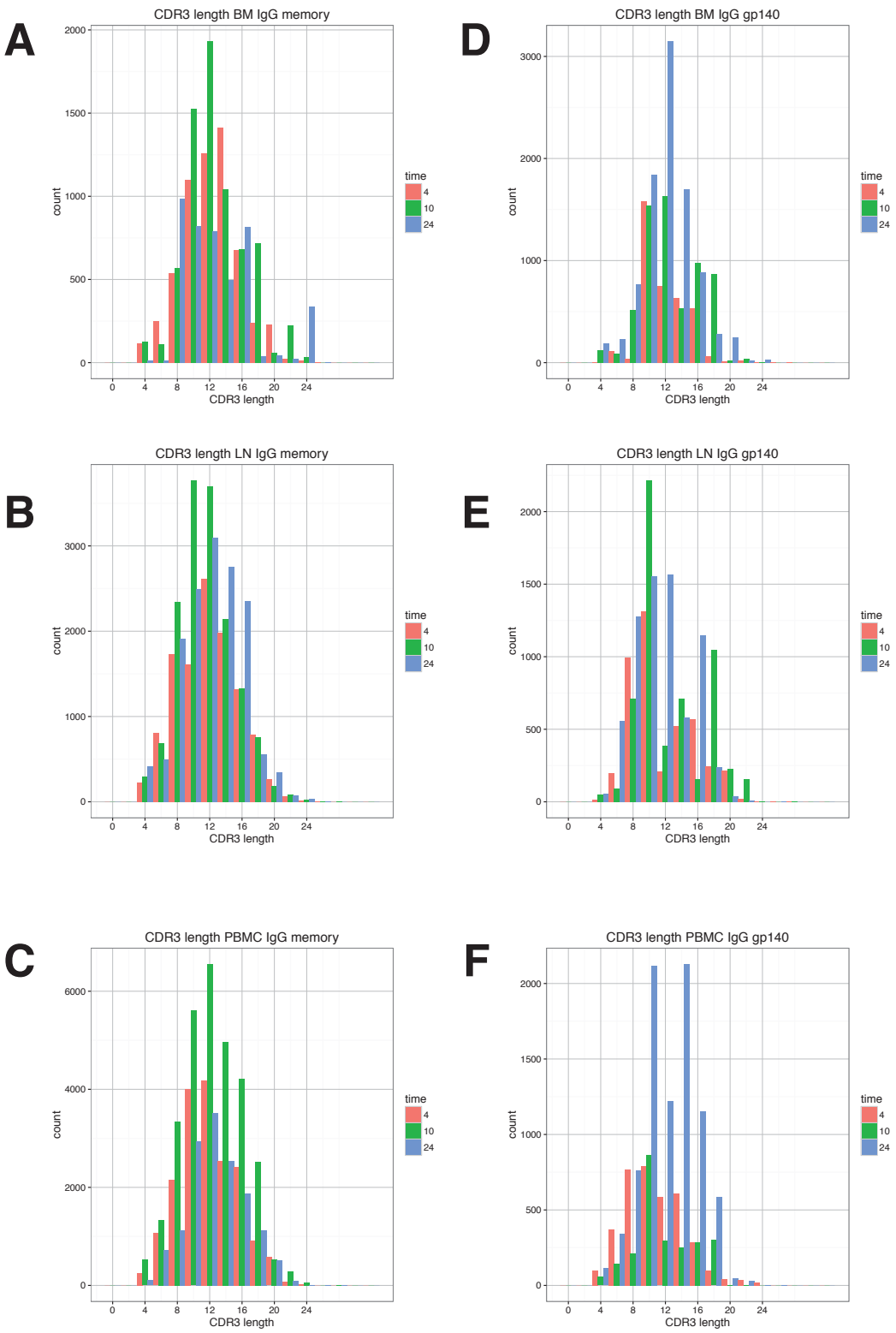


Fig 3.6 CDR3 length in SIV-specific and non-specific IgG heavy chain sequences. Histogram of CDR3 length in heavy chain sequences from non-specific IgG memory B cells by time point (4, 10, 24 weeks) in bone marrow (A), lymph nodes (B), and PBMC (C). Histogram of CDR3 length in heavy chain sequences from SIV-specific IgG memory B cells by time point (4, 10, 24 weeks) in bone marrow (D), lymph nodes (E), and PBMC (F).

Many neutralizing antibodies to HIV are also heavily mutated, with a subset of HIV bnAbs with 60-90 mutations in the V gene (10-30% mutation)[345]. Divergence from germline, measured using the genetic distance between the closest germline reference sequence of the variable region of the heavy chains, was unchanged in total lymph node and PBMC memory B cells, with a divergence of 4.8% to 6.9% in bone marrow, 5.2%-5.9% in lymph nodes, and 3.7%-5.1% in PBMC (Fig. 3.7 A-C). In all three compartments, median V mutation of SIV-specific sequences was unchanged at 4 and 10 weeks (from 3.5% to 3.8% in bone marrow, from 3.4% to 2.8% in lymph nodes, and 3.8% at both time points in PBMC) but increased at 24 weeks to 6.9%, 6.8%, and 6.3% in each tissue (Fig. 3.7 D-F). This was true in VH mutation in individual animals, with a statistically significant increase in the mean VH mutation between 4 weeks and 24 weeks in lymph nodes and PBMC in all three animals (Table 3.2). Two of the three animals had higher VH mutation at 24 weeks than at 4 weeks in bone marrow, while one animal had a slight decrease in the mean VH mutation between 4 weeks and 24 weeks. This shift in median divergence from germline likely reflects the on-going mutation and adaptation of SIV-specific B cells in the milieu of chronic infection and persistent viremia.

Table 3.2 VH mutation and CDR3 length in SIV-specific IgG sequences within individual animals

			CDR3 Length				VH mutation			
			4 weeks		24 weeks		4 weeks		24 weeks	
			median	SD	median	SD	median	SD	median	SD
ZF61	IgG gp140	PBMC	8	0.77	14	4.02	5.63	0.95	6.97	2.75
ZF76	IgG gp140	PBMC	12	4.63	10	3.52	4.51	2.73	4.86	2.39
ZG13	IgG gp140	PBMC	10	3.94	13	3.91	1.76	2.67	5.56	2.29
ZF61	IgG gp140	LN	15	3.76	10	3.86	7.72	2.38	6.8	2.16
ZF76	IgG gp140	LN	9	2.95	11	3.99	2.41	0.69	6.6	2.04
ZG13	IgG gp140	LN	10	2.87	12	3.27	1.72	2.45	6.94	2.23
ZF61	IgG gp140	BM	11	1.49	13	3.38	1.37	0.85	6.53	3.09
ZF76	IgG gp140	BM	14	3.07	12	3.85	11.5	2.60	6.87	3.03
ZG13	IgG gp140	BM	11	2.78	12	2.54	1.04	1.86	7.56	1.42

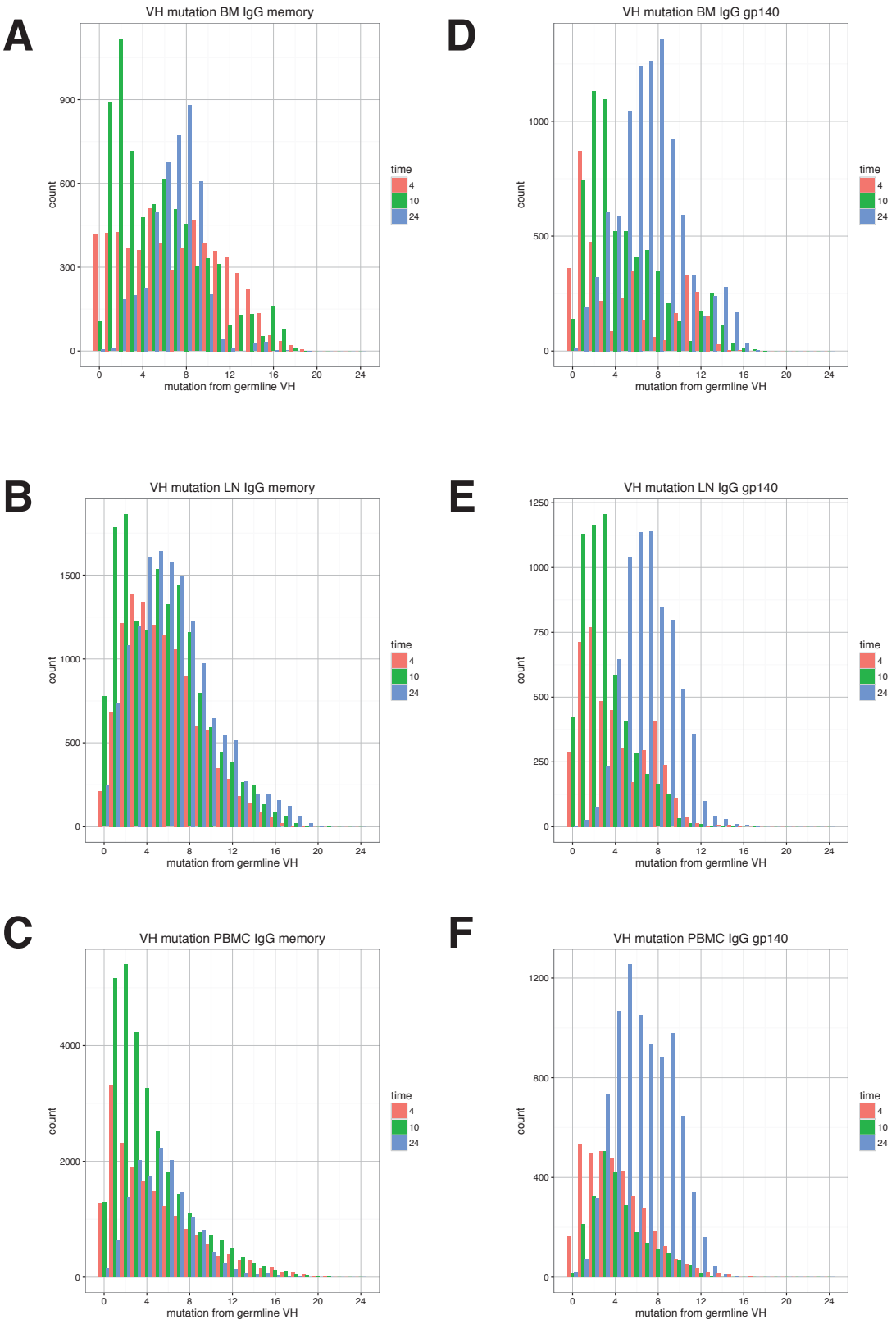


Fig 3.7 Percent mutation from predicted VH germline in SIV-specific and non-specific IgG heavy chain sequences. Percent divergence from V gene germline in heavy chains from non-specific IgG memory B cells in bone marrow (A), lymph nodes (B), and PBMC (C) by time point (4 weeks, 10 weeks, 24 weeks). Percent divergence from V gene germline in heavy chains from SIV-specific IgG memory B cells in bone marrow (D), lymph nodes (E), and PBMC (F) by time point (4 weeks, 10 weeks, 24 weeks).

3.2.5 Gene usage in IgG B cells

Of the hundreds of known human broadly neutralizing antibodies, certain V gene families are over-represented relative to their proportion in the pool of memory B cells[346]. To examine the early usage of V and J gene families in the SIV-specific response, each sequence was assigned a V and J gene based on BLAST nucleotide specificity, and the relative frequencies of each family were compared. While there was some variation in the proportion of the seven V gene families within total IgG memory bone marrow, lymph node, and PBMC cells at the three time points measured (Fig 3.8 A), the distributions largely followed previously established proportional distributions, with VH1, VH3, and VH4 dominating the total V gene usage[346]. IGHV2 made up only a small fraction of bone marrow and lymph node IgG memory sequences (up to 3.5%), but was slightly higher in PBMC (up to 10%). VH5, VH6, and VH7 generally represented less than 10% combined of all VH gene sequences in total memory IgG cells, but in lymph node SIV-specific sequences at four weeks and ten weeks, make up over 25% of all heavy chains (Fig. 3.8 B). Lymph node and PBMC VH gene distributions differed significantly ($p<0.0001$) from the total memory VH gene distributions at all time points. SIV-specific IgG bone marrow sequences at four weeks are dominated by VH4 and are significantly different from total IgG memory ($p<0.00001$) but by 24 weeks have frequencies of VH1, VH2, and VH3 that not significantly different from total bone marrow IgG memory ($p=0.10$). The dominance of VH4, VH3, and VH1 we observe is in agreement with data published by Sundling et al (2014) of IgG repertoires after immunization of macaques with HIV-1 Env[347]. Our work is also in accordance with studies showing no bias toward particular gene segments for Env-specificity[348].

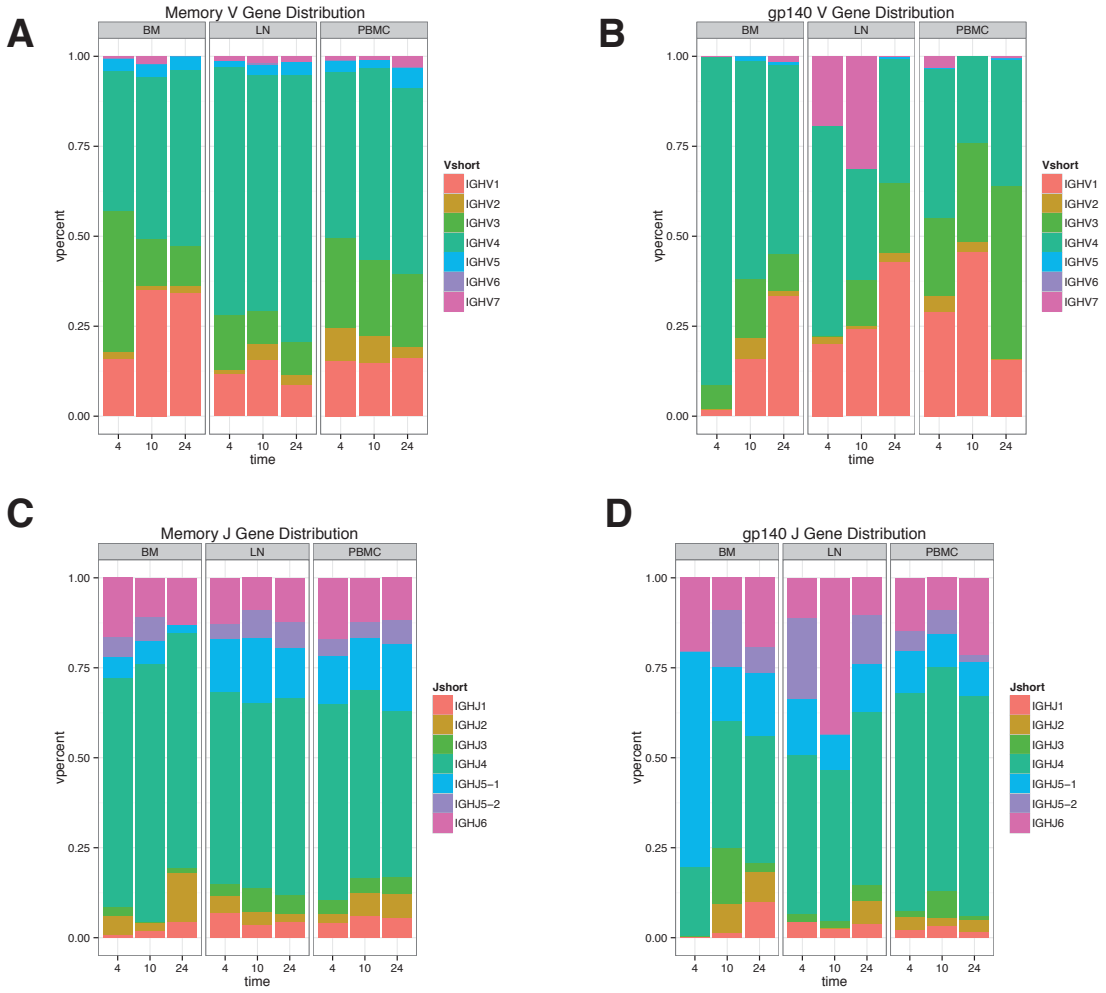


Fig 3.8 V and J gene distributions in IgG B cells. (A) Proportions of heavy chain V gene families in non-specific IgG memory B cells in bone marrow, lymph nodes, and PBMC. (B) Proportions of heavy chain V gene families in SIV-specific IgG memory B in bone marrow, lymph nodes, and PBMC. (C) Proportions of heavy chain J gene families in non-specific IgG memory B cells in bone marrow, lymph nodes, and PBMC. (D) Proportions of heavy chain J gene families in SIV-specific IgG memory B cells in bone marrow, lymph nodes, and PBMC.

IgG memory sequence J genes were dominated by JH4 (45%-60%), in agreement with published data[346]. In total IgG B cells, bone marrow sequences had fewer JH1 and JH5 genes than lymph node or PBMC, but all non-SIV specific sequences had similar frequencies of J genes (Fig. 3.8 C). J genes in SIV-specific sequences in lymph nodes were significantly different than total IgG B cells at all time points ($p = 0.004$ at 4 weeks, $p=0.02$ at 10 weeks and $p<0.0001$ at 24 weeks), largely due to fewer JH4 genes and more JH5 and JH6 genes (Fig 3.8 B). The SIV-specific PBMC J gene frequencies were not significantly different from total PBMC J gene frequencies at 4 weeks ($p=0.64$) but by 10 weeks ($p=0.0006$) and 24 weeks ($p<0.0001$) were significantly different, with increases in the proportion of JH4 genes in SIV-specific cells. Bone marrow J genes showed the greatest differences between SIV-specific and total memory, with significantly more JH5-1 genes at 4 weeks in SIV-specific sequences

(p<0.0001) and fewer JH4 sequences at 10 weeks and 24 weeks (both p<0.0001). However, there did not appear to be a unique J gene signature of SIV-specific sequences.

Of the long CDR3 sequences (18 amino acids or greater), the V and J gene distributions were significantly altered when compared to the total distribution in SIV-specific sequences (Fig. 3.9). Long CDR3 sequences in bone marrow were almost entirely from VH1 (95% at 4 weeks) and VH4 (77% at 10 weeks and 78% at 24 weeks) (Fig. 3. A). Lymph node sequences with CDR3s of 18 amino acids or longer had significantly more VH1 genes at 4 and 10 weeks, and significantly fewer VH7 genes at all time points. PBMC sequences with CDR3s of 18 amino acids or longer contained higher frequencies of VH2 genes at 4 weeks and 10 weeks and of VH3 and VH5 at 24 weeks, and fewer VH4 genes at 10 weeks. J genes were also significantly different in sequences with long CDR3s (Fig. 3.9 B). JH1 genes were overrepresented in bone marrow at four weeks, and JH2 were present at a significantly higher proportion at 24 weeks. In lymph nodes, there were higher frequencies of JH6 genes at 4 and 10 weeks in long CDR3 samples, and higher JH5-1 at 24 weeks. PBMC samples also had higher frequencies of JH1 (4 weeks and 10 weeks) and JH6 (24 weeks) than total SIV-specific sequences.

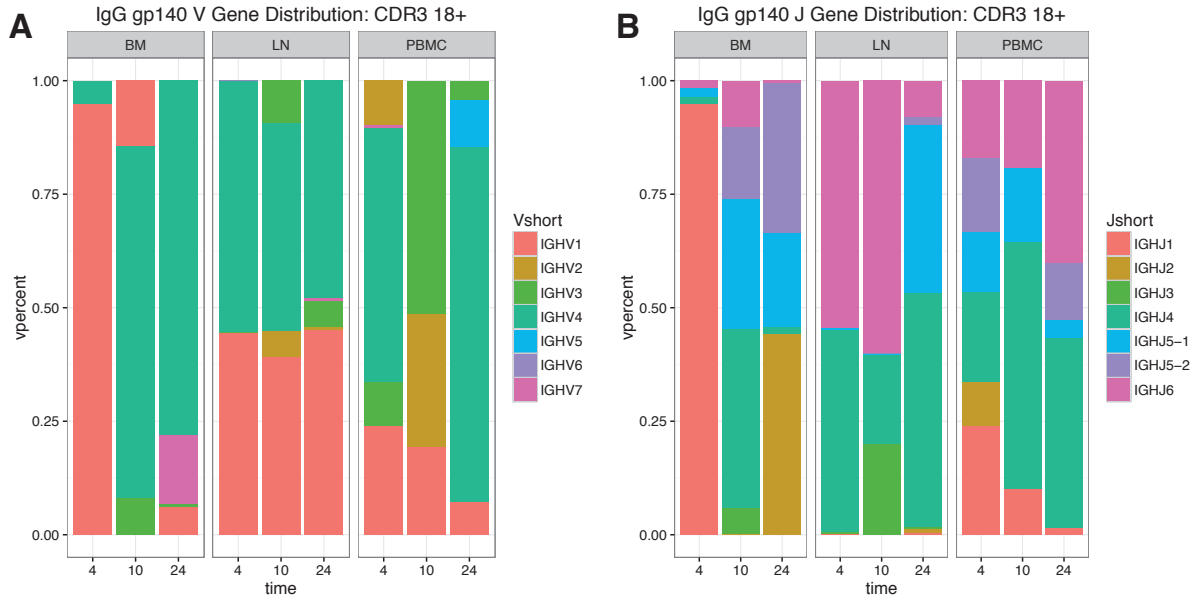


Fig 3.9 V and J gene distributions in SIV-specific IgG B cells with CDR3 length 18 amino acids or longer. (A) Proportions of heavy chain V gene families in bone marrow, lymph nodes, and PBMC at 4, 10, and 24 weeks post infection. (B) Proportions of heavy chain J gene families in bone marrow, lymph nodes, and PBMC at 4, 10, and 24 weeks post infection.

3.2.6 IgM memory B cells: CDR3, mutation, and V/J gene usage

While the humoral response to HIV/SIV is dominated by IgG, there are low numbers of antigen binding IgM memory B cells circulating and in lymphoid tissues[349]. Nonspecific IgM memory B cells in PBMC and lymph nodes had median CDR3 lengths of 10-12 amino acids (standard deviation 3.8-4.9), while SIV-specific IgM memory B cells had median CDR3 lengths of 11-12 amino acids (SD 3.2-5.0) (Fig 3.10). Although the median CDR3 lengths were similar, there was a significant difference between the distributions of SIV-specific and total lymph node and PBMC IgM CDR3 lengths at each time point ($p<0.0001$) (Table 3.3). No significant differences were observed in the percentage of long CDR3s (over 18 AA), ranging from 0% to 0.26% of SIV-specific sequences and 0.09%-0.44% of bulk IgM memory sequences (Table 3.4). However, IgM sequences had fewer long CDR3s than their IgG counterparts, with an average of 3.2 times fewer CDR3s of length 18 AA or longer.

Table 3.3 Median CDR3 length and VH gene mutation in IgM B cells throughout infection.

Time Point (weeks)	Tissue	Population	IGH	Median CDR3	CDR3 standard deviation	Median VH mutation	VH mutation standard deviation	Percentage of CDR3s 18AA or longer
4	**LN**	**gp140**	**IgM**	11	3.3	5.9	2.83	0.00
10	**LN**	**gp140**	**IgM**	11	4.0	2.79	3.15	0.24
24	**LN**	**gp140**	**IgM**	11	4.1	4.53	3.32	0.18
4	**PBMC**	**gp140**	**IgM**	10	4.9	4.48	3.42	0.29
10	**PBMC**	**gp140**	**IgM**	11	5.0	5.78	2.24	0.11
24	**PBMC**	**gp140**	**IgM**	11	4.0	3.33	3.09	0.27
4	**LN**	**mem**	**IgM**	10	4.3	4.81	2.87	0.13
10	**LN**	**mem**	**IgM**	10	3.8	4.23	3.38	0.09
24	**LN**	**mem**	**IgM**	11	4.9	5.54	2.80	0.44
4	**PBMC**	**mem**	**IgM**	11	4.2	3.94	3.06	0.23
10	**PBMC**	**mem**	**IgM**	11	4.1	3.44	2.87	0.33
24	**PBMC**	**mem**	**IgM**	12	3.8	4.53	2.91	0.18

Table 3.4 VH mutation and CDR3 length in SIV-specific IM sequences within individual animals

			CDR3 Length				**VH mutation**			
			4 weeks		**24 weeks**		**4 weeks**		**24 weeks**	
			median	**SD**	**median**	**SD**	**median**	**SD**	**median**	**SD**
ZF61	IgM gp140	PBMC	10	5.8	10	4.00	5.44	3.62	3.78	3.03
ZF76	IgM gp140	PBMC	11	4.0	11	3.73	4.51	2.91	3.04	2.93
ZG13	IgM gp140	PBMC	9	3.8	17	3.14	2.38	3.73	11.81	1.87
ZF61	IgM gp140	LN	13	2.0	10	3.33	7.22	1.61	6.97	2.87
ZF76	IgM gp140	LN	12	2.9	11	4.13	5.84	3.48	3.47	3.25
ZG13	IgM gp140	LN	10	2.1	10	4.74	5.84	1.58	5.26	3.51

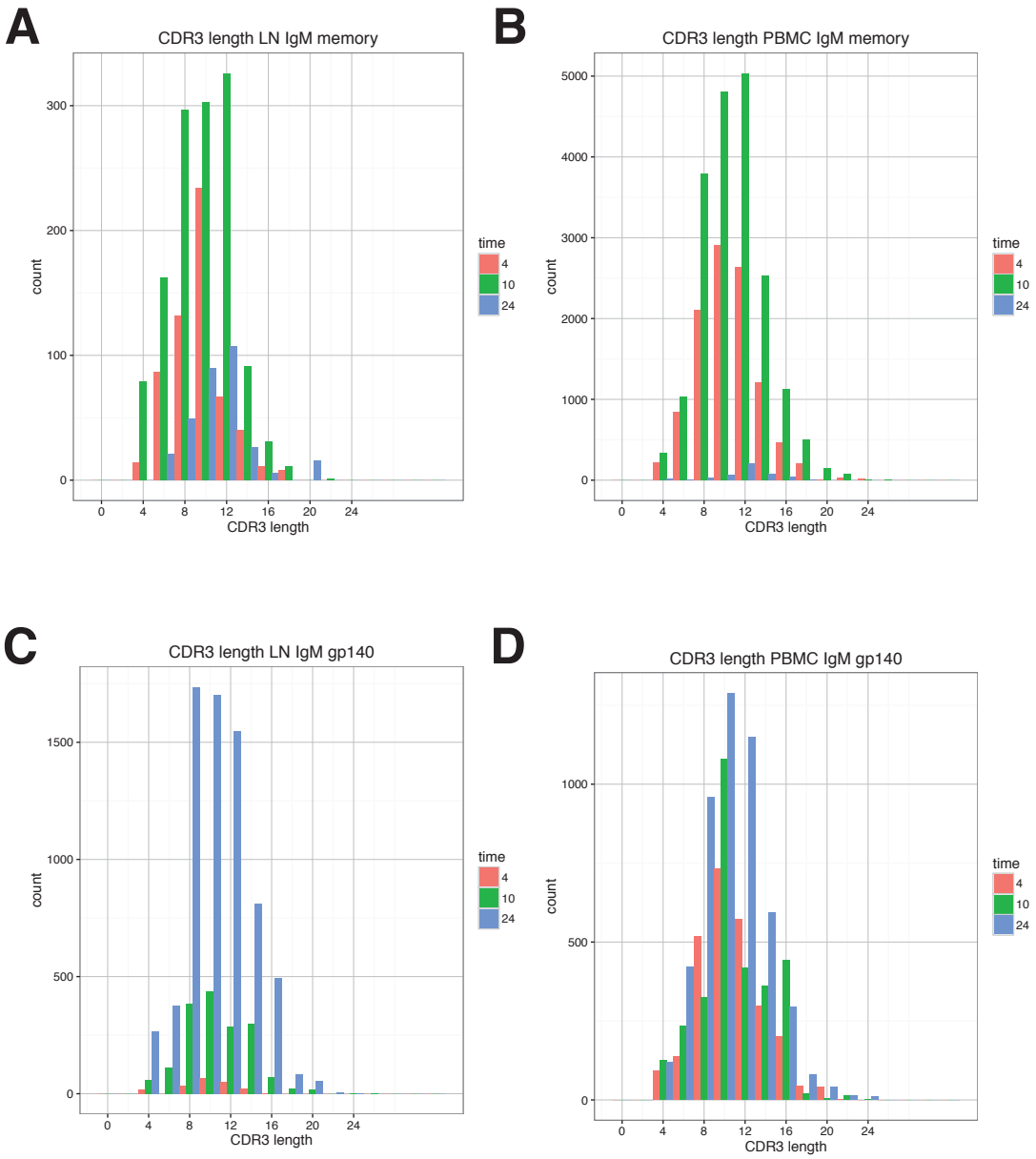


Fig 3.10 CDR3 length in SIV-specific and non-specific IgM heavy chain sequences. (A) CDR3 length in heavy chains from non-specific IgM memory B cells in lymph nodes by time point (4 weeks, 10 weeks, 24 weeks). (B) CDR3 length in heavy chains from non-specific IgM memory B cells in PBMC by time point (4 weeks, 10 weeks, 24 weeks). (C) CDR3 length in heavy chains from SIV-specific IgM memory B cells in lymph nodes by time point (4 weeks, 10 weeks, 24 weeks). (D) CDR3 length in heavy chains from SIV-specific IgM memory B cells in PBMC by time point (4 weeks, 10 weeks, 24 weeks).

SIV-specific IgM memory V genes did not show accumulation of mutations over the course of infection. Sequences from 4 weeks post infection in lymph nodes had a median VH gene mutation of 5.9% and standard deviation of 2.8 but by 24 weeks was 4.5% with a standard deviation of 3.3 (Fig 3.11 C). In PBMC, median VH gene mutation similarly declined from 4.5% at 4 weeks to 3.3% at 24 weeks (Fig. 3.11 D). In the non-SIV specific IgM memory cells, median VH gene mutation ranged from a median of 4.2% to 5.5% in lymph nodes and 3.9%-4.5% in PBMC (Fig 3.11 A, B). Thus, in contrast to IgG SIV-specific memory B cells,

SIV-specific IgM cells did not show the accumulation of mutations over the course of infection.

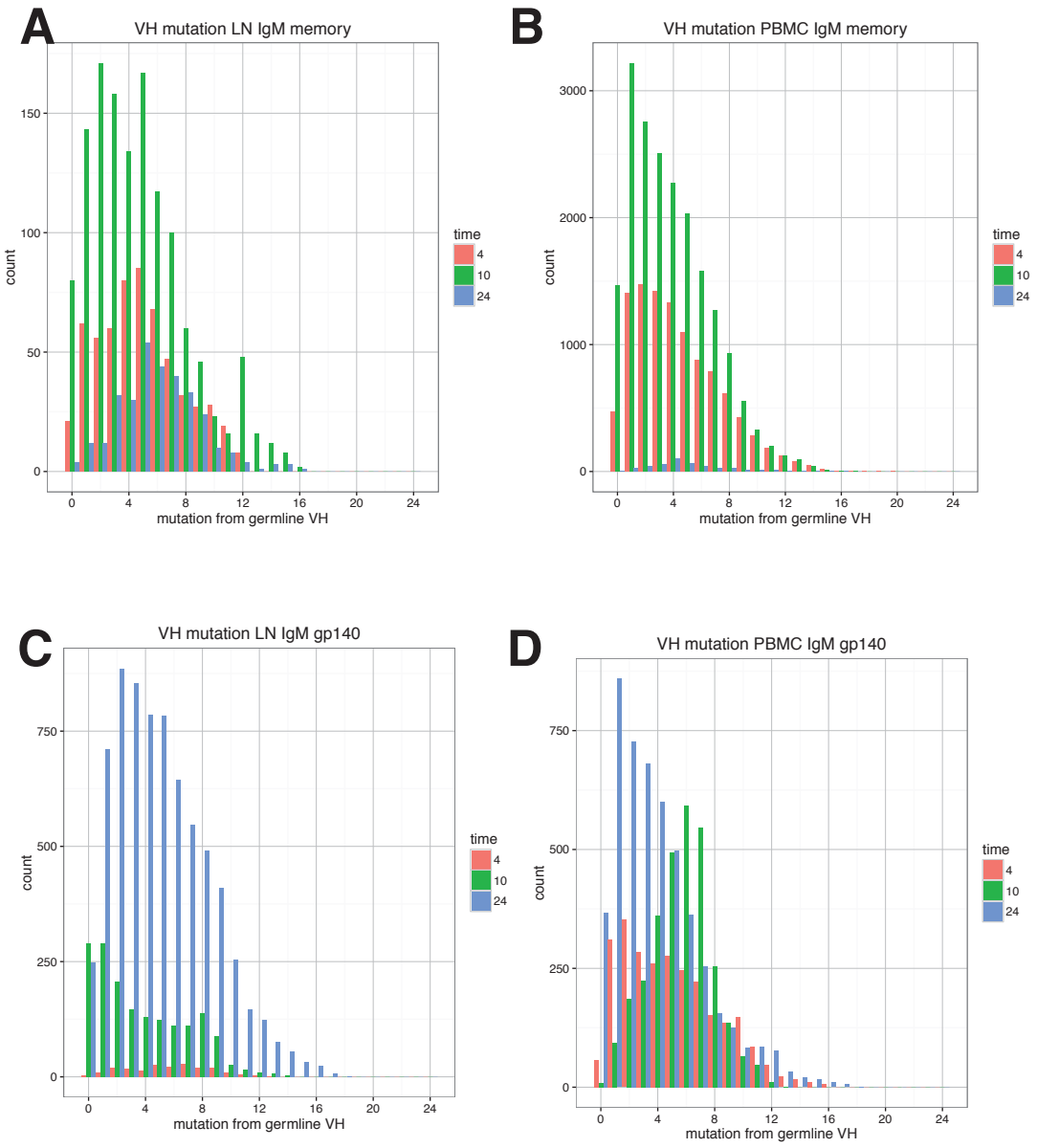


Fig 3.11 Percent mutation from predicted V Gene germline in SIV-specific and non-specific IgM heavy chain sequences. (A) Percent divergence from V gene germline in heavy chains from non-specific IgM memory B cells in lymph nodes by time point (4 weeks, 10 weeks, 24 weeks). (B) Percent divergence from V gene germline in heavy chains from non-specific IgM memory B cells in PBMC by time point (4 weeks, 10 weeks, 24 weeks). (C) Percent divergence from V gene germline in heavy chains from SIV-specific IgM memory B cells in lymph nodes by time point (4 weeks, 10 weeks, 24 weeks). (D) Percent divergence from V gene germline in heavy chains from SIV-specific IgM memory B cells in PBMC by time point (4 weeks, 10 weeks, 24 weeks).

VH and JH gene distributions in the non-specific IgM memory sequences in lymph node and PBMC showed greater variation between time points than the IgG sequences. VH gene frequencies were dominated by VH4 and VH3, except in PBMC at 24 weeks post infection when the majority of sequences had a VH1 gene assignment (Fig 3.12). There was also a

substantially higher fraction of VH2 sequences in total IgM lymph node samples at all time points, in contrast to total IgG sequences where VH2 sequences were only detected in as a small fraction of sequences. The SIV-specific IgM sequences had higher frequencies of VH5 and VH7 in PBMC at 4 weeks and 10 weeks, and in both tissues had low to undetectable fractions of VH1, in contrast to the IgG SIV-specific sequences (Fig 3.12 B). Both lymph node and PBMC sequences had significantly different distributions of V genes between SIV-specific and total IgM cells at 4, 10, and 24 weeks (all $p<0.0001$). Lymph node SIV-specific sequences were dominated by VH2 and VH3 at 4 weeks and 10 weeks (65% and 67% of all J genes). J gene sequences in non-specific IgM memory B cells were similar to the IgG memory distributions with the exception of both lymph node and bone marrow at 24 weeks post infection, where JH6 genes dominated and made up nearly 75% of the sequences (Fig. 3.12 C). This bias was not observed as strongly in the SIV-specific J gene sequences, which had nearly 40% JH6 genes in lymph nodes at 24 weeks and PBMC at 10 weeks (Fig 3.12 D).

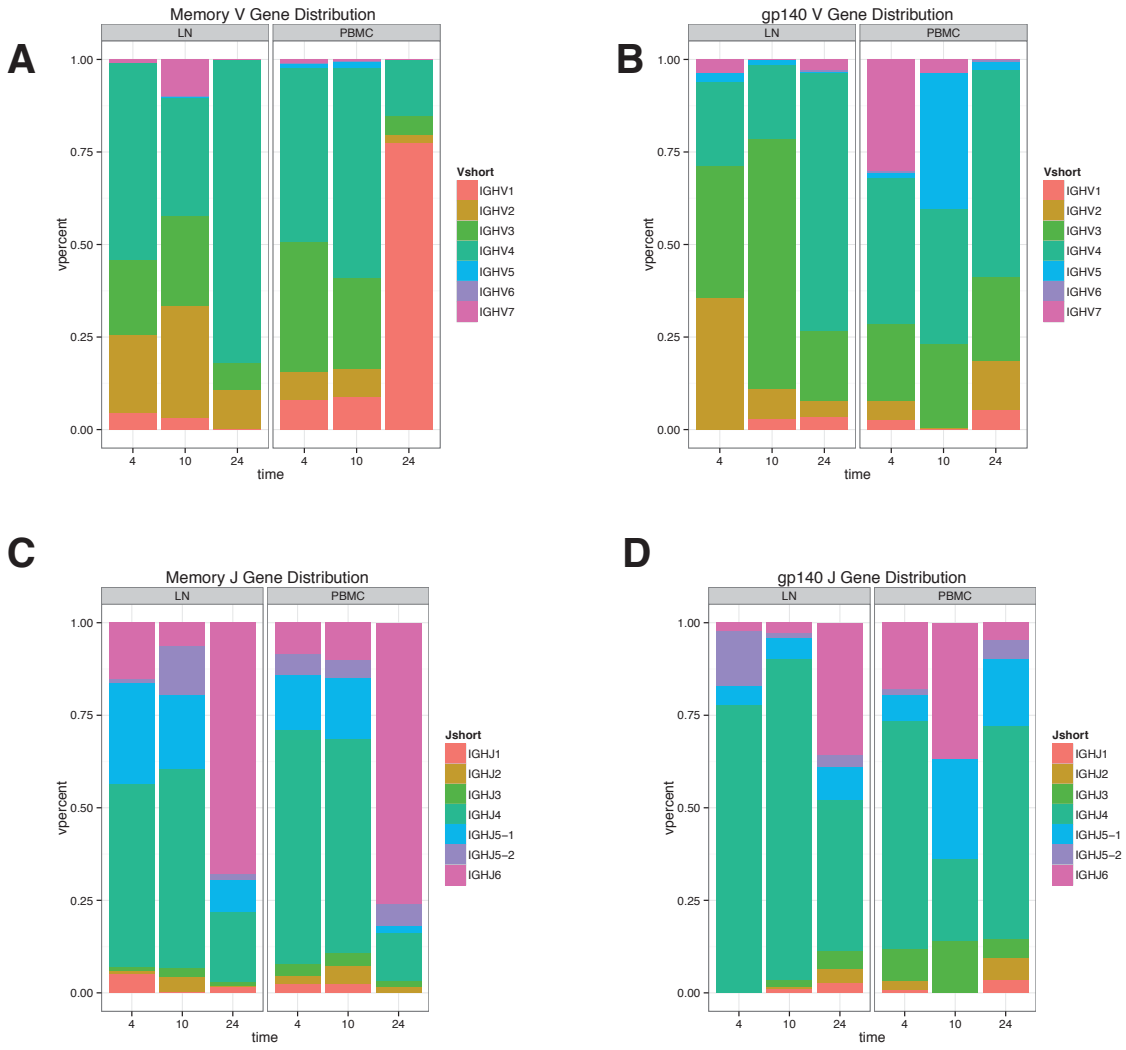


Fig 3.12 V and J gene distributions in IgM B cells. (A) Proportion of V gene families in non-specific IgM memory B cells in lymph nodes and PBMC at 4, 10, and 24 weeks post infection. (B) Proportion of V gene families in SIV-specific IgM memory B cells in lymph nodes and PBMC at 4, 10, and 24 weeks post infection. (C) Proportion of V gene families in non-specific IgM memory B cells in lymph nodes and PBMC at 4, 10, and 24 weeks post infection. (D) Proportion of V gene families in SIV-specific IgM memory B cells in lymph nodes and PBMC at 4, 10, and 24 weeks post infection.

3.2.7 SIV-specific heavy chains become more diverse throughout infection

While the number of sequencing reads per unique heavy chain in next-generation bulk sequencing is not a direct measure of biological clone size, it nonetheless offer a qualitative insight into the distribution of clones and expansion of lineages. We used a clustering algorithm, CD-HIT[350], to analyse closely related sequences and study the size of clusters and the connectivity of sequences within a sample. These network plots show expanded clones and the diversity of sequences within a sample, and allow the calculation of several diversity measures using the size and number of connections within networks[332]. Non-specific memory populations typically show a large number of unique sequences with some smaller networks of expanded clones, while the SIV-specific samples are generally dominated by a small number of clones that are closely related to many other sequences within the sample (Fig. 3.13). This is reflected in the percentage of vertices in the largest cluster, where memory

samples have an average of 2.1% of all vertices (unique sequences) in the largest cluster in the sample, whereas SIV-specific samples have an average of 10% of vertices in the largest cluster (Fig 3.14). While there was not a statistically significant difference between the total and SIV-specific samples, early samples (4 weeks) had the greatest mean percentage of vertices in the largest cluster with 20%, indicating that the early response was dominated by a small number of clones. In contrast, in later infection (24 weeks) only 4.6% of vertices were in the largest cluster, meaning a single clonal family did not dominate the SIV-specific response. In the IgM sequences, the non-specific sequences had a mean 2.6% of vertices in the largest cluster and did not differ significantly across time-points (Fig 3.14 B). The SIV-specific IgM sequences had an average of 6.6% of vertices in the largest cluster, and did not show the dominance of a small number of clones in the early time point as the IgG sequences did.

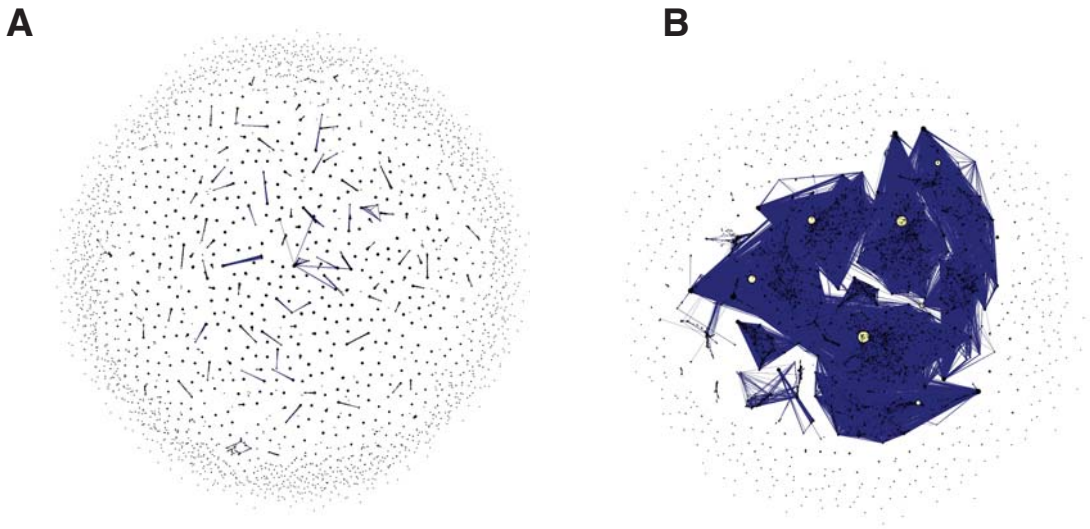


Fig 3.13 Representative network plots of non-specific memory and SIV-specific heavy chain sequences. Network plots of non-SIV specific (A) and SIV-specific (B) IgG memory heavy chain sequences. Network plots visualize the relatedness of sequences in a sample by drawing each unique sequence as a point and lines connecting closely related sequences. The size of each point is determined by the number of identical sequencing reads for that sequence.

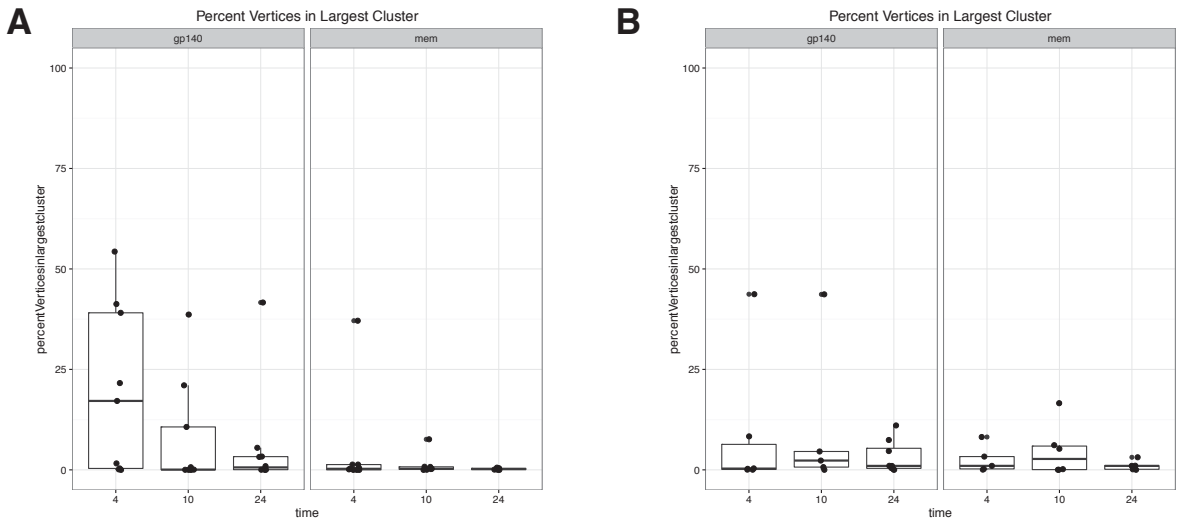


Fig 3.14 Vertices in largest cluster (A) Percentage of vertices in the largest cluster in the network plots of bone marrow, lymph node, PBMC, and spleen SIV-specific and total memory IgG B cells at 4, 10, and 24 weeks. (B) Percentage of vertices in the largest cluster in the network plots of bone marrow, lymph node, PBMC, and spleen SIV-specific and total memory IgM B cells at 4, 10, and 24 weeks.

We use two diversity measures with two parameters to further examine the patterns in the SIV-specific and nonspecific memory heavy chains. The Gini index is a measure of the inequality of a system, while the Renyi index is a measure of evenness in a system, both on a scale of 0 to 1[332]. When applied to BCR repertoires, inequality and evenness can be used to quantify whether a group of sequences contain equal numbers of diverse sequences (low inequality and high evenness), or whether the sample is dominated by a small number of closely related sequences (high inequality and low evenness). A low Gini index (close to 0) indicates low inequality, and a low Renyi index indicates high evenness. In this system, the Gini and Renyi indices can be calculated both for the clusters distributions (based on the total size of each cluster) and the vertices (how many related sequences are in each cluster). The cluster Gini index indicates the level of somatic hypermutation and families of related sequences, while the vertex Gini index describes how clonal, or dominated by a sequence or small group of sequences, the sample is. The vertex Gini for non-specific IgG B cells had a median of 0.67 and a range of 0.41 to 0.79 (Fig 3.15 A), and was statistically unchanged between 4, 10, and 24 weeks (p=0.15, p=0.38, and p=.90). The SIV-specific IgG B cells had a median of 0.39 and range 0.13 of to 0.74 and was statistically unchanged between 4, 10, and 24 weeks (p=0.33, p=0.53, and p=0.70). However, in comparing SIV-specific to non-specific vertex Gini indices, SIV-specific samples were significantly lower at all three time points (p=0.02, p=0.0003, and p=0.005). In nonspecific memory IgG B cells, the cluster Gini index was not significantly different (p=0.36, p=0.84, p=0.35) between sampling time points, with a median of 0.84 and range of 0.71 to 0.88 (Fig. 3.15 B). The SIV-specific IgG sequences had a

lower median Gini index of 0.61, and a range of 0.41 to 0.83, and were not significantly different between time points. In comparing SIV-specific to non-specific cluster Gini indices, SIV-specific samples were significantly lower at all three time points ($p=0.003$, $p=0.0001$, and $p=0.0006$). Thus, the SIV-specific sequences were more clonal (a small fraction of clones made up a large proportion of the sample) than the non-specific IgG memory sequences. Thus, the SIV-specific IgG B cells tend to have more somatic hypermutation within clonally related families, indicated by the higher number of vertices connecting related sequences within clusters and representing closely related sequences.

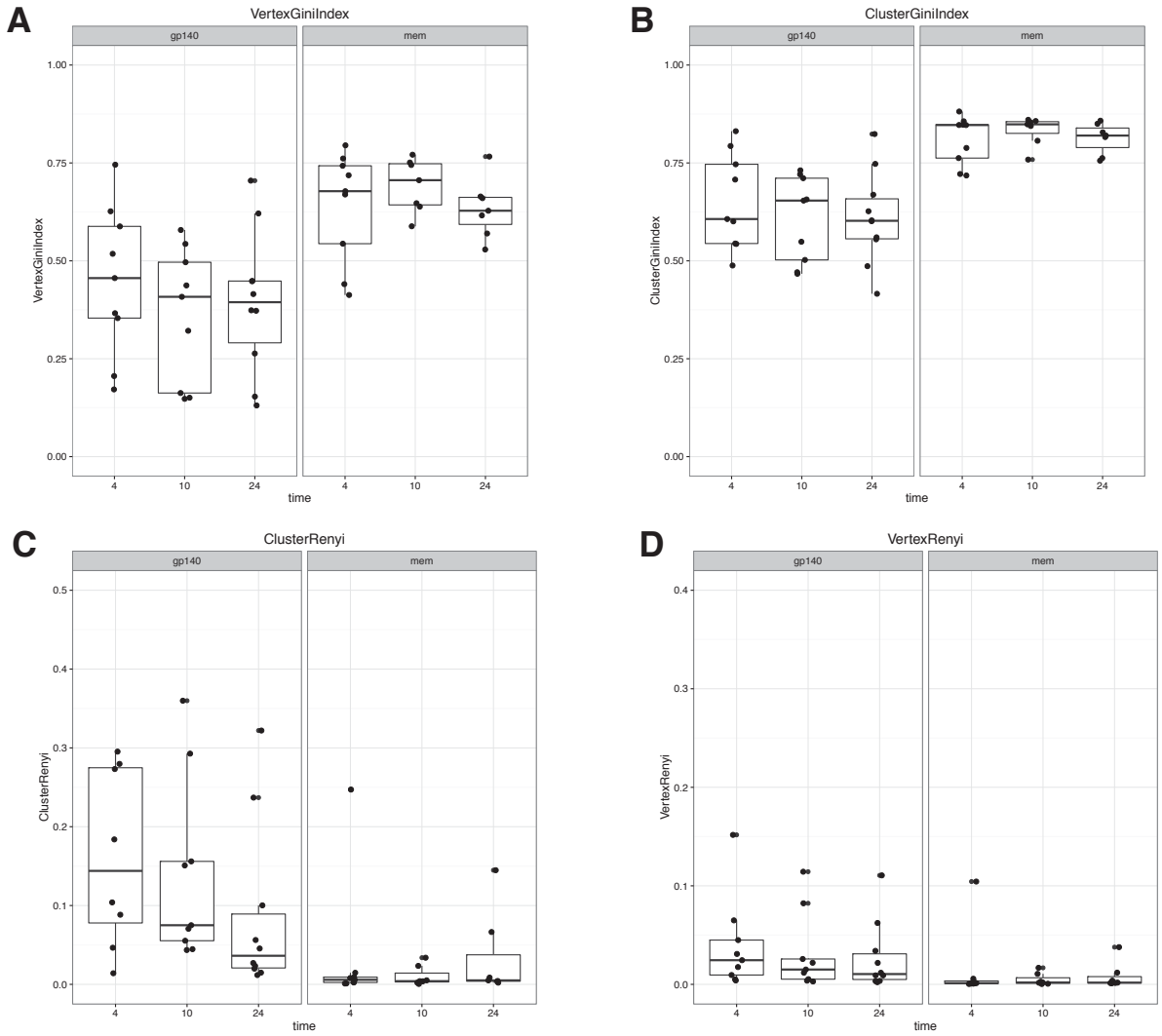


Fig 3.15 Diversity measures in IgG B cells (A) Vertex and (B) cluster Gini index for bone marrow, lymph node, PBMC, and spleen SIV-specific and total memory IgG B cells at 4, 10, and 24 weeks. (C) Vertex and (D) cluster Renyi index for bone marrow, lymph node, PBMC, and spleen SIV-specific and total memory IgG B cells at 4, 10, and 24 weeks.

The Renyi index measures evenness, and can similarly be applied to the clusters and vertices of heavy chain sequences. A low cluster Renyi index indicates an even sample, with most clones and families of similar size and read count. A low vertex Renyi indicates relatively low levels of somatic hypermutation and less expansion of closely related B cells. The IgG non-specific B cells had a median cluster Renyi index of $5.8x10^{-3}$, with a range of $5.5x10^{-4}$ to 0.24 (although only 2 values were over 0.07), and no significant difference between Renyi index at 4, 10, and 24 weeks (p=0.29-0.97) (Fig. 3.15 C). SIV-specific B cells had a median cluster Renyi index of 0.073, with a range of 0.012 to 0.57, and were unchanged between 4, 10, and 24 weeks (p=0.08-034). SIV-specific B cell cluster Renyi indices were significantly lower at 4 weeks (p=0.015) and 10 weeks (p=0.012), but not at 24 weeks (p=0.25). This indicates that total IgG memory B cell heavy chain sequences had a more even distribution of clone sizes, while the SIV-specific samples had an uneven distribution of clone sizes at 4 and 10 weeks. The vertex Renyi values for non-specific sequences had a median of $1.6x10^{-3}$ and range of $3.6x10^{-4}$ to 0.104 and none were significantly different between 4, 10, and 24 weeks post infection (p=0.54-0.73). (Fig 3.15 D). SIV-specific IgG heavy chains had a median vertex Renyi index of 0.18 and a range of $2.2x10^{-3}$to 0.015 and were not significantly different between 4, 10, and 24 weeks post infection (p=0.26-0.70). Though the vertex Renyi indices were higher in SIV-specific memory B cells (indicating greater unevenness of somatic hypermutation and expanded clonal families), there was not a statistically significant difference between total and SIV-specific B cells. Lymph node samples had a higher median vertex and cluster Gini Index (0.50 and 0.71) than bone marrow and PBMC (0.37 and 0.60, 0.34 and 0.54). Vertex Renyi index was also highest in lymph nodes, 0.022) but not cluster Renyi index (0.075).

Unlike the IgG samples, the differences between Gini and Renyi diversity indices in SIV-specific and non-specific IgM memory sequences were not statistically significant (Fig 3.16 A-D). The median cluster Gini index for SIV-specific sequences was 0.52 (range 0.40 to 0.68) whereas for non-specific IgM sequences it was 0.54 (range 0.39 to 0.73). The median vertex Gini index for SIV-specific sequences was 0.43 (range 0.10 to 0.60), while the median non-specific IgM vertex Gini index was 0.34 (range 0.14 to 0.56). PBMC sequences had consistently higher Gini indices in SIV-specific sequences at 4 weeks (0.51 and 0.66 for vertex and cluster) than lymph nodes (0.16 and 0.50), but by 24 weeks were not substantially different (0.44 and 0.53 for PBMC, 0.43 and 0.59 for lymph nodes). This indicates early IgM SIV-specific heavy chains in PBMC are less diverse than later sequences, but overall the IgM

sequences do not show the same expansion and diversification as IgG SIV-specific sequences. With the exception of two samples, the vertex Renyi indices for IgM samples were virtually identical across tissue types and time points, with median values of $7.3x10^{-3}$ and $9.7x10^{-3}$for SIV-specific and nonspecific and ranges of $1.4x10^{-3}$- 0.17 and $1.3x10^{-3}$-0.11 respectively (Fig. 3.16 C). Cluster Renyi indices were similarly low, with median values of 0.058 and 0.029 for SIV-specific and non-specific IgM heavy chains (Fig. 3.16 D).

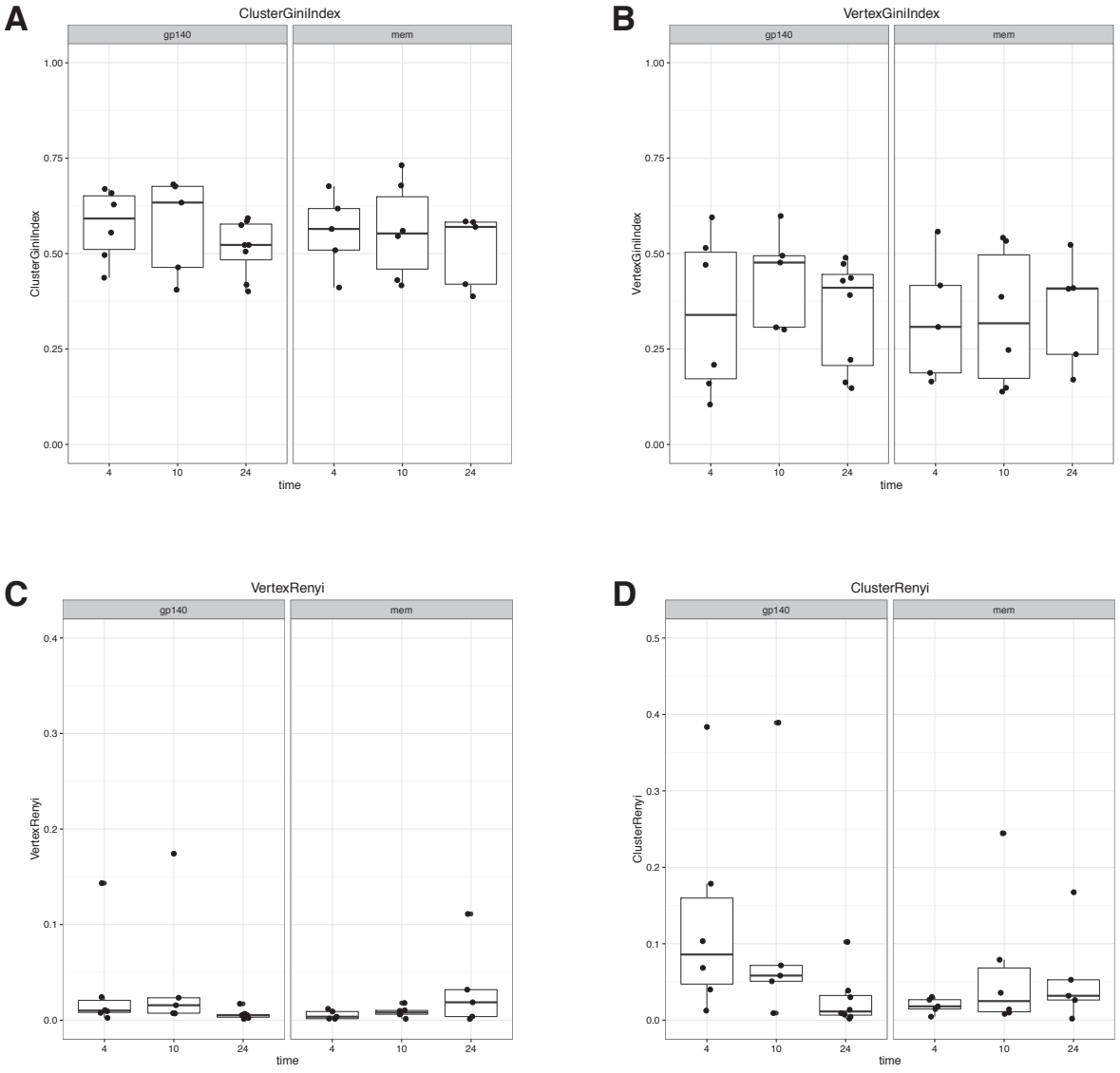


Fig 3.16 Diversity measures of IgM heavy chains. (A) Cluster and (B) vertex Gini index for lymph node, PBMC, and spleen SIV-specific and total memory IgM B cells at 4, 10, and 24 weeks. (C) Cluster and (D) vertex Renyi index for lymph node, PBMC, and spleen SIV-specific and total memory IgM B cells at 4, 10, and 24 weeks.

3.2.8 Shared sequences across time and space

Antibody lineages are derived from a single common ancestor and represent generations of antibodies that target specific epitopes and have undergone successive rounds of somatic

hypermutation. To track the persistence of related sequences over time and in different tissue compartments, we defined sequences as identical if they had the same V and J genes and identical CDR3 regions. We also looked for sequences in different compartments and time points that had were closely (same V and J genes but CDR3 regions that differed by one or two amino acids, either insertions, deletions or substitutions). We wanted to look at the features of the gp140-specific response within individual animals as between animals: identical or similar heavy chain sequences in multiple animals.

Table 3.5 Sharing of CDR3 sequences between tissues.

SIV-specific			**Identical**	**SD**	**1AA**	**SD**	**2AA**	**SD**	**3AA**	**SD**
LNvsPBMC	**IgG**	**gp140**	3.1%	1.8%	14.2%	3.3%	25.6%	6.4%	34.4%	9.2%
LNvsBM	**IgG**	**gp140**	1.8%	2.5%	18.4%	14.2%	22.7%	15.6%	25.7%	16.8%
BMvsPBMC	**IgG**	**gp140**	0.8%	1.5%	3.4%	4.2%	6.7%	7.7%	12.3%	11.2%
LNvsPBMC	**IgM**	**gp140**	0.3%	0.4%	1.7%	1.3%	5.8%	3.4%	17.6%	7.4%
Total Memory			**Identical**	**SD**	**1AA**	**SD**	**2AA**	**SD**	**3AA**	**SD**
LNvsPBMC	**IgG**	**gp140**	1.0%	0.4%	3.5%	1.6%	7.2%	2.4%	16.3%	3.3%
LNvsBM	**IgG**	**gp140**	0.5%	0.3%	8.2%	12.5%	11.1%	14.4%	17.5%	14.1%
BMvsPBMC	**IgG**	**gp140**	0.4%	0.3%	3.5%	5.1%	5.3%	5.3%	12.0%	6.1%
LNvsPBMC	**IgM**	**gp140**	0.5%	0.6%	1.8%	1.5%	6.7%	6.2%	20.0%	11.2%

Of all the SIV-specific IgG sequences within an individual animal, an average of 3.1% of sequences were identical between lymph nodes and PBMC from all time points (Table 3.5), though not significantly different than the proportion of identical sequences in total memory (1.0%). An average of 14.2% of all lymph node and PBMC sequences were within one amino acid apart, 25.6% were within 2 amino acids, and 34.4% were within three amino acids from a sequence in the other compartment – all significantly more than in total memory lymph node vs. PBMC sequences ($p= 0.007$, $p=0.009$, $p=0.03$). Between lymph nodes and bone marrow, 1.8% of sequences were identical, rising to 25.7% of sequences within a 3 amino acid difference in the CDR3 region, though there were not significantly more sequences shared than in total memory. Bone marrow and PBMC shared fewer sequences, with an average of 0.8% identical sequences, 3.4% of sequences within a single amino acid difference in the CDR3 region, and 12.3% of sequences within three amino acids, again with no significant difference between SIV-specific and total memory sharing. For IgM SIV-specific sequences, only lymph node compared to PBMC had sufficient sequences for statistical comparison.

Between these compartments, an average of 0.3% of sequences were identical, rising to 17.6% of sequences within a 3 amino acid difference in the CDR3, but were not significantly different from total memory Thus, the greatest sharing of sequences is between lymph nodes and PBMC and there is significantly more sharing within SIV-specific sequences than in total memory, though shared sequences are observed across all compartments throughout infection.

Table 3.6 Sharing of CDR3 sequences across time points

SIV-specific			**Identical**	**SD**	**1AA**	**SD**	**2AA**	**SD**	**3AA**	**SD**
4vs10	**IgG**	**LN**	7.9%	6.5%	31.6%	33.1%	32.8%	33.3%	38.6%	30.9%
10vs24	**IgG**	**LN**	0.5%	0.5%	2.3%	2.9%	5.8%	4.6%	13.7%	10.3%
4vs24	**IgG**	**LN**	0.2%	0.2%	5.5%	6.4%	6.3%	7.5%	9.4%	9.7%
4vs10	**IgM**	**LN**	0.0%	0.0%	0.0%	0.0%	2.3%	5.0%	6.3%	7.4%
10vs24	**IgM**	**LN**	0.3%	0.6%	1.1%	1.3%	3.6%	3.6%	13.3%	8.4%
4vs24	**IgM**	**LN**	0.0%	0.0%	0.0%	0.0%	3.8%	5.5%	9.8%	12.7%
Total Memory			**Identical**	**SD**	**1AA**	**SD**	**2AA**	**SD**	**3AA**	**SD**
4vs10	**IgG**	**LN**	0.3%	0.3%	1.3%	0.8%	3.4%	1.5%	9.1%	3.0%
10vs24	**IgG**	**LN**	0.2%	0.2%	1.0%	1.1%	2.7%	1.6%	8.4%	3.9%
4vs24	**IgG**	**LN**	0.6%	0.2%	2.1%	1.1%	4.9%	2.1%	11.4%	3.5%
4vs10	**IgM**	**LN**	0.0%	0.0%	0.0%	0.0%	1.2%	1.5%	9.4%	6.4%
10vs24	**IgM**	**LN**	0.0%	0.0%	0.0%	0.0%	0.6%	0.8%	5.0%	5.0%
4vs24	**IgM**	**LN**	0.0%	0.0%	0.0%	0.0%	0.8%	1.4%	6.3%	8.1%

Within the lymph nodes, certain heavy chain sequences persisted over the twenty-four weeks of infection studied here (Table 3.6). There was no statistically significant difference between the percentages of sequences shared over time in SIV-specific versus total memory at any time point. Between 4 weeks and 10 weeks, an average of 7.9% of SIV-specific IgG sequences were identical, while between 10 weeks and 24 weeks 0.5% of sequences were identical, and between 4 weeks and 24 weeks only 0.2% of sequences were identical. There were more closely related sequences shared between 4 and 10 weeks (an average of 38.6% of sequences were within 3 amino acids) than there were between 10 to 24 weeks (13.7% of sequences within 3 amino acids), or 4 to 24 weeks (9.4% of sequences within 3 amino acids). This likely reflects the different time intervals between sampling (6 weeks vs. 14 weeks), however, the persistence of identical and closely related sequences indicates the on-going evolution of antibody lineages in addition to the constant generation of new SIV-specific B cells from naïve B cells. There were fewer IgM unique sequences isolated from lymph nodes, however sharing of closely related sequences between time-points was observed in many of the samples. Identical sequences were recorded only between 10 and 24 weeks post infection

(average of 0.3%), but there were an average of 6.3% of sequences within a three amino acid difference between 4 and 10 weeks, 13.3% between 10 and 24 weeks, and 9.8% between 4 and 24 weeks. Even in cases where few sequences were recorded in a sample, a fraction of those sequences or their close relatives were recorded at different time points. Thus, there is greater persistence of lineages in IgG memory B cells, but IgM lineages are also observed over the time frame of this study.

Table 3.7 Sharing of CDR3 sequences between animals

SIV-specific			**Identical**	**SD**	**1AA**	**SD**	**2AA**	**SD**	**3AA**	**SD**
LN	**IgG**	**gp140**	2.2%	1.7%	22.8%	23.4%	25.8%	25.9%	29.4%	26.0%
LN	**IgM**	**gp140**	5.8%	9.0%	6.7%	9.4%	10.5%	9.7%	23.4%	9.5%
PBMC	**IgG**	**gp140**	1.7%	2.6%	13.4%	27.6%	16.2%	27.0%	23.2%	24.9%
PBMC	**IgM**	**gp140**	2.5%	4.8%	3.9%	5.3%	6.5%	6.8%	16.0%	9.3%
Total memory			**Identical**	**SD**	**1AA**	**SD**	**2AA**	**SD**	**3AA**	**SD**
LN	**IgG**	**gp140**	14.2%	8.0%	28.9%	7.2%	32.1%	7.6%	39.5%	7.5%
LN	**IgM**	**gp140**	0.0%	0.0%	0.1%	0.2%	4.2%	1.9%	18.1%	6.0%
PBMC	**IgG**	**gp140**	1.5%	1.8%	6.1%	8.5%	8.8%	8.8%	15.5%	8.4%
PBMC	**IgG**	**gp140**	2.3%	2.6%	4.4%	4.4%	9.6%	5.6%	23.9%	9.6%

Sequencing studies of both IGH and TCR have shown that there are both private, i.e. specific to an individual; and public, i.e. shared between individual, repertoires[351–356]. In spite of the small number of animals sampled here, we were able to identify numerous sequences within the SIV-specific IgG and IgM heavy chains that were similar between individual animals (Table 3.7). An average of 2.2% of all IgG lymph nodes sequences were identical between animals, and an average of 5.8% of IgM lymph node sequences were identical between animals at all time points. The proportion of identical or closely related sequences between animals did not differ significantly between SIV-specific and total memory repertoires. The mean CDR3 length of IgG similar sequences was 13.6, significantly higher ($p<0.0001$) than the total CDR3 length of 12.5 (Fig 3.17 A). However, IgM similar CDR3s were significantly shorter ($p<0.0001$), with a mean of 10.0 versus 11.2 in total IgM sequences (Fig 3.17 B). The mean VH mutation in similar IgG sequences was slightly lower than in total sequences (5.24 vs. 5.54, $p=0.003$). IgM similar sequences had significantly higher ($p<0.0001$) VH mutation than total IgM sequences, with a mean of 5.9% compared to 4.9% in total IgM. Similar IgG sequences showed accumulation of mutations in V genes that mirrored the total sequences, increasing from a mean of 5.25% at 4 weeks to 6.5% at 24 week. However, similar IgM sequences also showed an increase in VH mutation, from 5.5% at 4 weeks to 6.6% at 24 weeks in IgM. Total IgM sequences did not show this increase, with an average of 5.1%

mutation t 4 weeks and 4.7% at 24 weeks. The V and J gene proportional distributions were significantly different in the common repertoires for both IgM and IgG sequences (Fig 3.18). IgG similar sequences had a higher proportion of VH1 and VH5 genes, and JH1 and JH6 genes, than the total IgG repertoire (Fig 3.18) IgM similar sequences has significantly higher VH4, JH6, and fewer V2, V5, and J5-1 than total IgM (Fig 3.18). When we expanded our filtering to include sequences within a three amino acid difference in the CDR3 region, an average of 29.4% of IgG and 23.4% of IgM sequences were similar between animals. In comparing sequences in the PBMC, an average of 1.7% of IgG and 2.5% of IgM sequences were identical, expanding to 23.2% and 16.0% of sequences within 3 amino acids. These data indicate common features of the SIV-specific response between animals.

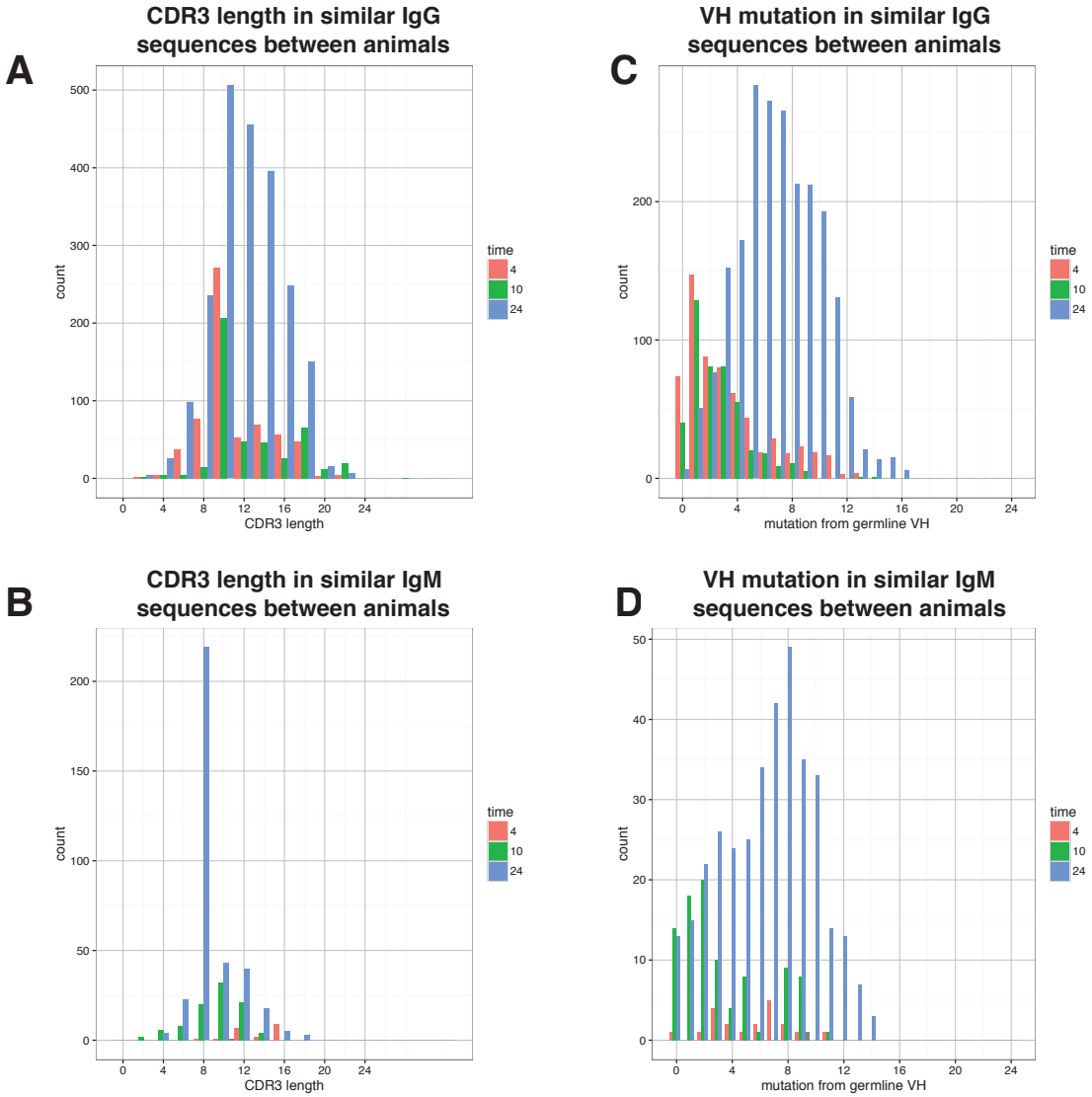


Fig 3.17 CDR3 Length and VH mutation in similar sequences between animals. (A) CDR3 length in similar IgG sequences (B) CDR3 length in similar IgM sequences (C) Percent divergence from V gene germline in heavy chains in similar IgG sequences (D) Percent divergence from V gene germline in heavy chains in similar IgM sequences

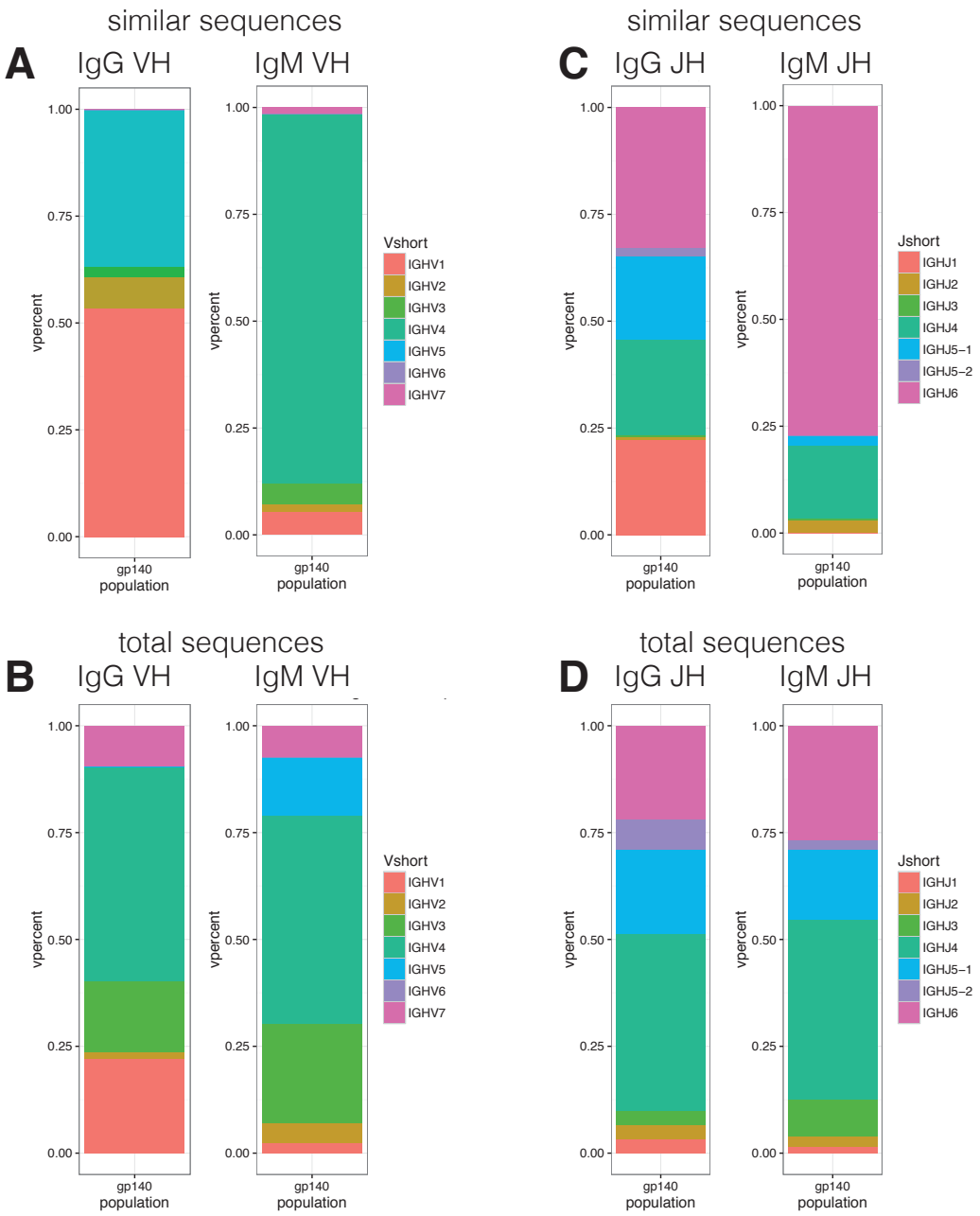


Fig 3.18– VH and JH gene frequencies in similar sequences. (A) VH gene frequencies in similar IgG and IgM sequences. (B) JH gene frequencies in similar IgG and IgM sequences. (B) VH gene frequencies in total IgG and IgM sequences. (D) JH gene frequencies in total IgG and IgM sequences.

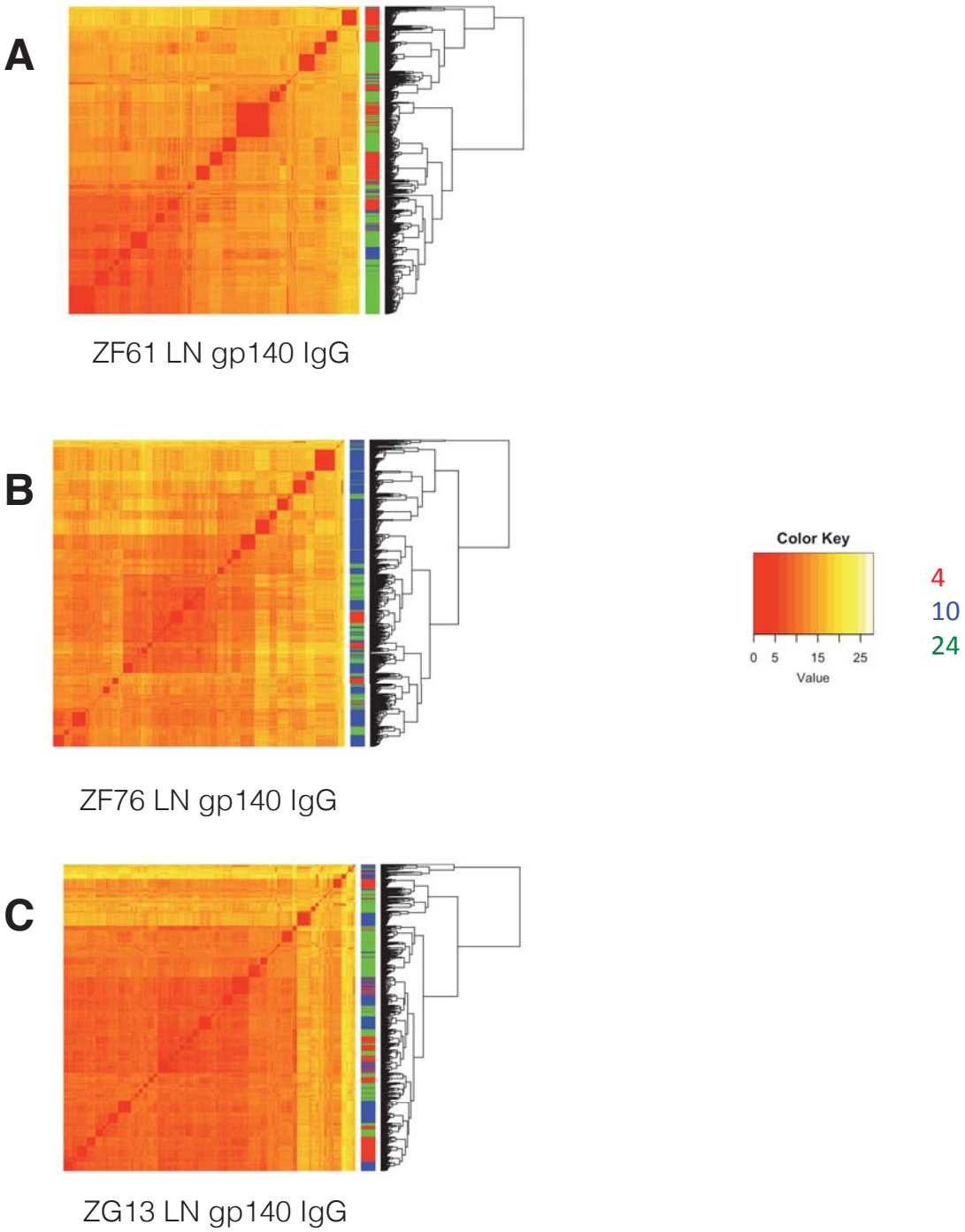


Fig 3.19 Sharing of CDR3 and V-J gene identical sequences between samples in lymph nodes. Hierarchical clustering of lymph node SIV-specific IgG heavy chain sequences from 4, 10, and 24 weeks post infection in animals (A) ZF61, (B) ZF76, and (C) ZG13. Colored bar on right indicates whether sequences are from week 4 (red), 10 (blue), or 24 (green).

Hierarchical clustering of IgG CDR3 sequences from lymph node samples taken at 4, 10, and 24 weeks shows the relative genetic distance between sequences. In all three animals, clusters of closely related lymph node sequences (indicated by blocks of red and orange) are made up of sequences from all three time points, indicated by sidebars of red, blue, and green (Fig 3.19). These clusters represent lineages of heavy chain sequences that persist and evolve over

the course of infection, and that related sequences are located at different times and not restricted to a single sampling point. PBMC IgG sequences showed similar clustering, except in cases where there were significantly lower numbers of heavy chain sequences recovered from a sample (eg.ZF61 IgG PBMC week 10, Fig. 3.20 A). In the remaining two animals, sequences from all three time points were evenly distributed in small clusters, with no blocks of sequences restricted to a single time point (Fig 3.20 B, C). These plots illustrate the presence of individual lineages that persist over the period of sampling, and show that the sequences sampled here are not primarily clustered by time point. Most samples cluster with similar sequences from other time points and indicate closely related B cell linages. While these samples represent only a small fraction of the circulating SIV-specific B cells, these plots show that we are still able to detect the presence of related B cells that persist throughout infection.

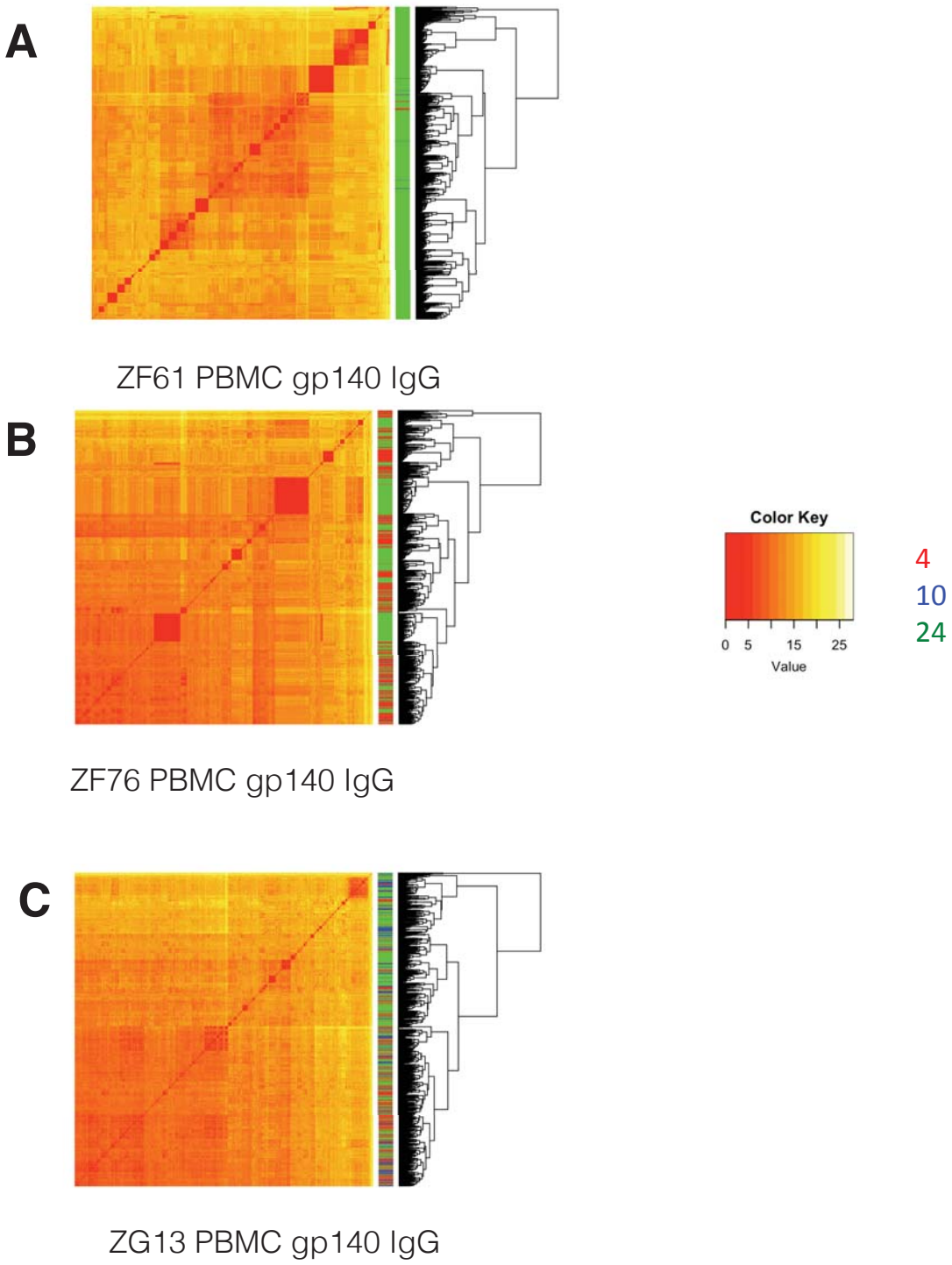


Fig 3.20 Sharing of CDR3 and V-J gene identical sequences between samples in PBMC. Hierarchical clustering of PBMC SIV-specific IgG heavy chain sequences from 4, 10, and 24 weeks post infection in animals (A) ZF61, (B) ZF76, and (C) ZG13. Colored bar on right indicates whether sequences are from week 4 (red), 10 (blue), or 24 (green).

3.2.9 Isoelectric point and charge distinguish HCDR3 in antigen-specific memory cells

The physicochemical properties of the CDR3 regions can influence the binding of the antibody to its target[357]. Furthermore, they can be used to indicate if the antibody response to

an antigen is targeted beyond engaging a set of VDJ germline genes, and distinguish IgG and IgM memory responses[235,358]. We performed a computational analysis of four summary properties of the HCDR3 amino acid sequences in SIV-specific and non-specific IgG and IgM memory B cells in PBMC, LN and bone marrow. Hydrophobicity is computed using the GRAVY (grand average of hydropathy) using the Kidera method[359], where scores above zero are hydrophobic and scores below zero are more hydrophilic. Isoelectric point is the pH at which the CDR3 region carries no net charge: proteins with basic amino acids will have a high pI, whereas proteins with many acidic amino acids will have a low pI. It is calculated using the pKa of individual amino acids. The net charge of the CDR3 region was calculated using the EMBOSS method[360]. Finally, the aliphatic index calculates the relative volume occupied by alanine, valine, leucine, and isoleucine[361]. High aliphatic index indicates higher thermostability of globular proteins.

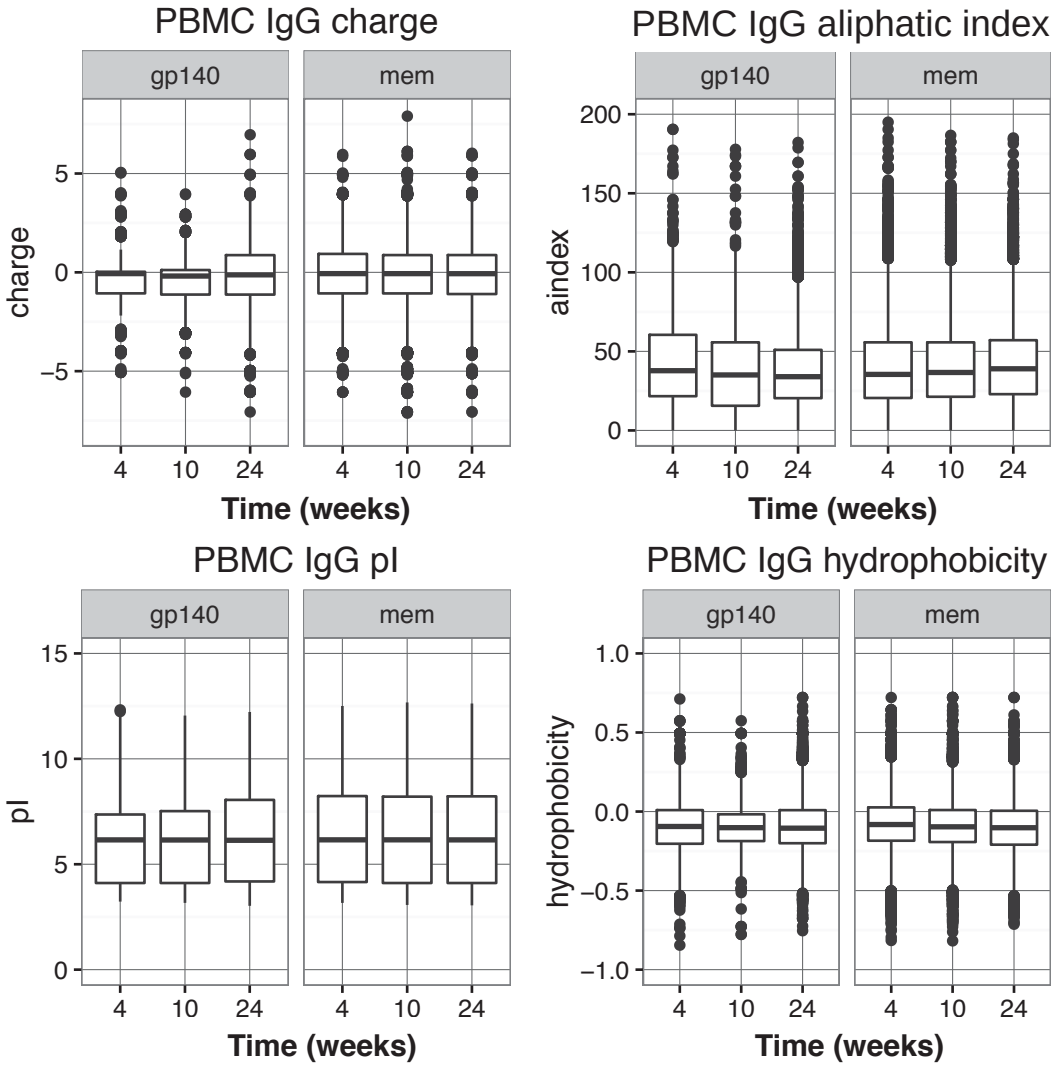


Fig 3.21 Physiochemical features of CDR3 sequences in PBMC IgG B cells. Net charge, aliphatic index, isoelectric point, and GRAVY hydrophobicity score from PBMC IgG samples. All plots show SIV-specific and non-specific samples at 4 weeks, 10 weeks, and 24 weeks post infection.

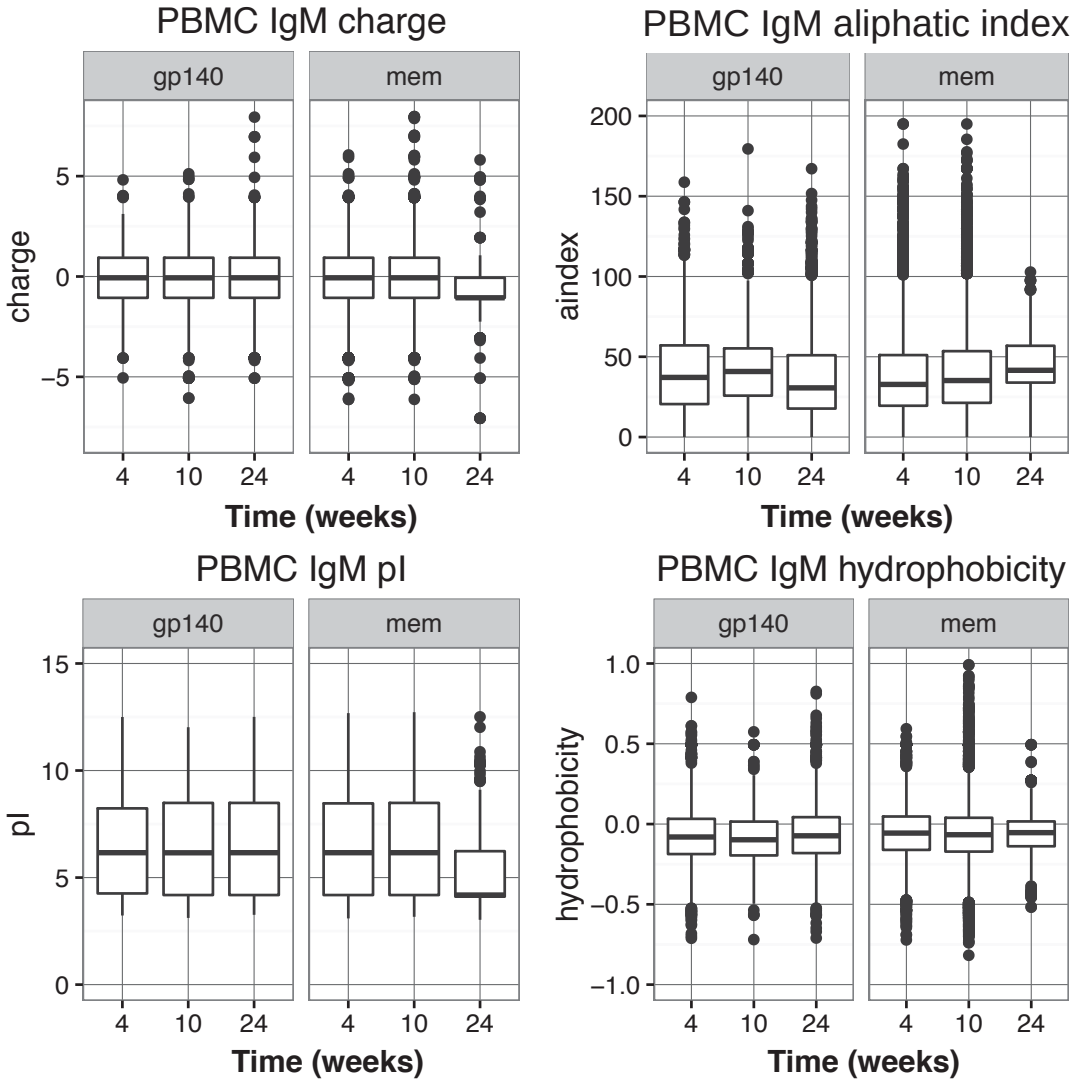


Fig 3.22 Physiochemical features of CDR3 sequences in PBMC IgM B cells. Net charge, aliphatic index, isoelectric point, and GRAVY hydrophobicity score from PBMC IgM samples. All plots show SIV-specific and non-specific samples at 4 weeks, 10 weeks, and 24 weeks post infection.

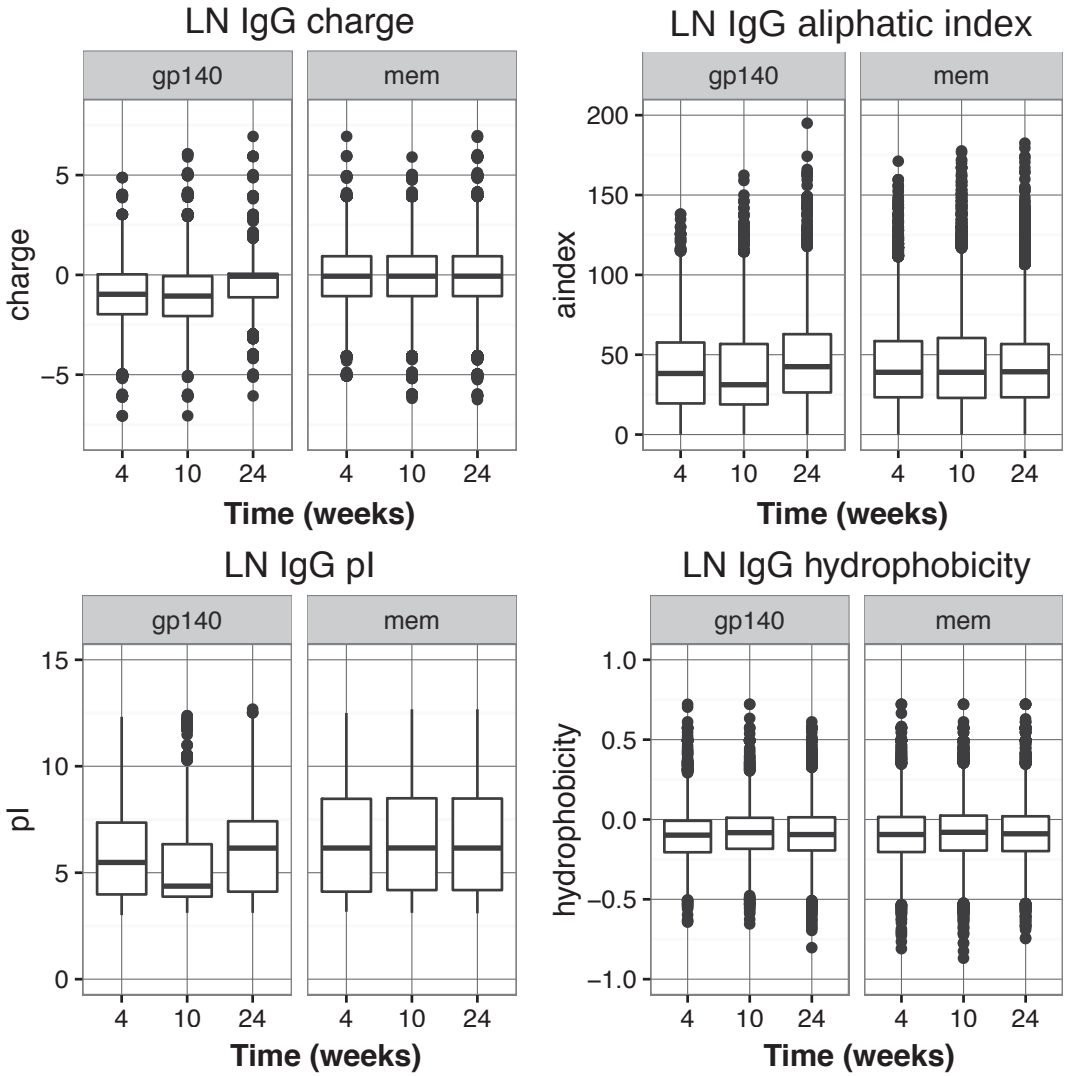


Fig 3.23 Physiochemical features of CDR3 sequences in lymph node IgG B cells. Net charge, aliphatic index, isoelectric point, and GRAVY hydrophobicity score from lymph node IgG samples. All plots show SIV-specific and non-specific samples at 4 weeks, 10 weeks, and 24 weeks post infection.

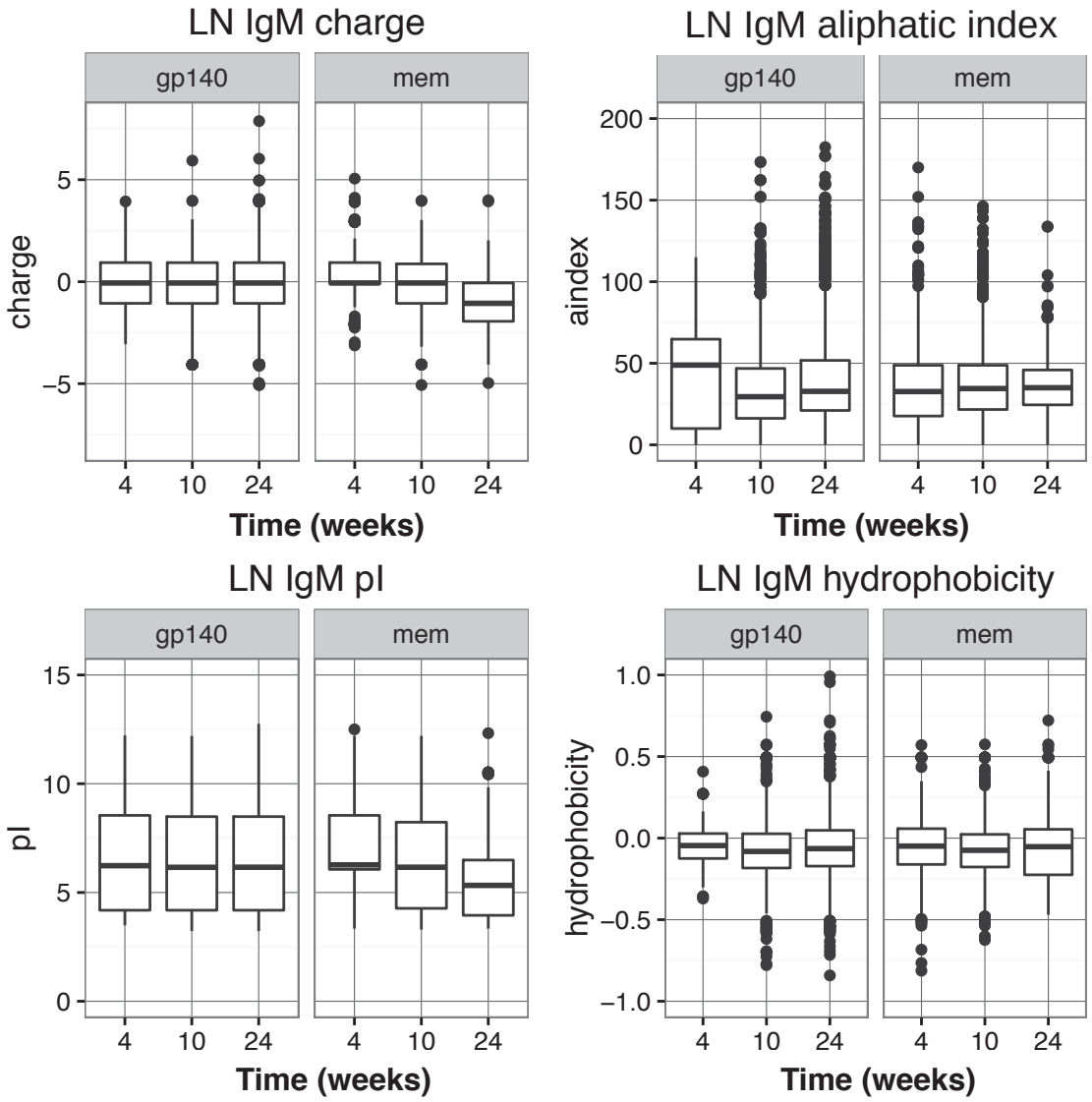


Fig 3.24 Physiochemical features of CDR3 sequences in lymph node IgM B cells. Net charge, aliphatic index, isoelectric point, and GRAVY hydrophobicity score from lymph node IgM samples. All plots show SIV-specific and non-specific samples at 4 weeks, 10 weeks, and 24 weeks post infection.

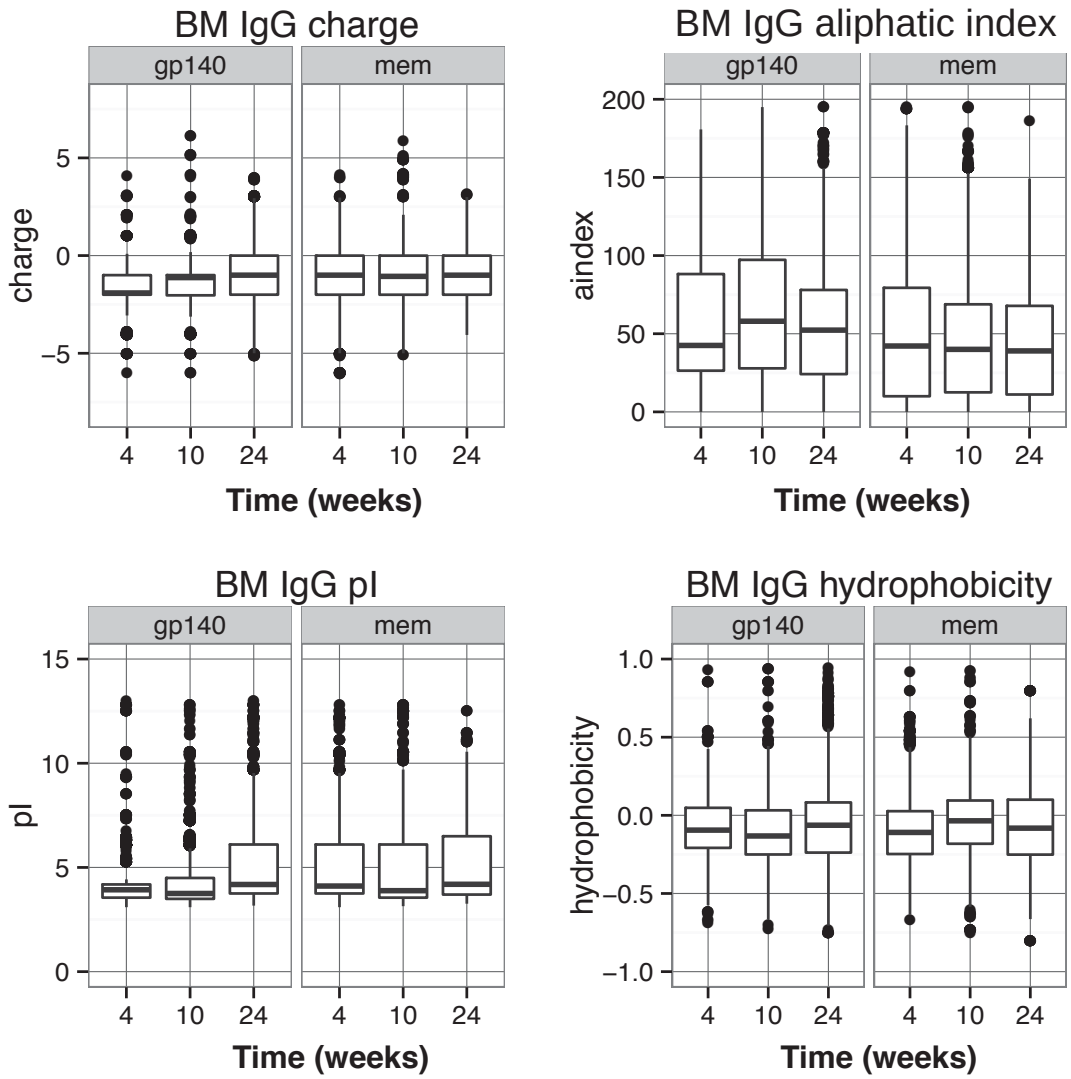


Fig 3.25 Physiochemical features of antigen-specific CDR3 sequences in antigen-specific B cells. Net charge, aliphatic index, isoelectric point, and GRAVY hydrophobicity score from bone marrow IgG samples. All plots show SIV-specific and non-specific samples at 4 weeks, 10 weeks, and 24 weeks post infection.

Sequences from non-specific IgG PBMC were unchanged in hydrophobicity, net charge, and isoelectric point. There was a small shift in aliphatic index (from 35 at 4 weeks to 39 at 24 weeks). SIV-specific IgG PBMC sequences had no different in hydrophobicity and isoelectric point at the three time points sampled (and were equivalent to non-specific sequences), but had an increase in net charge over the course of infection (from -0.065 at 4 weeks to -0.13 at 24 weeks) and a decrease in median aliphatic index (from 38 at 4 weeks to 34 at 24 weeks) (Fig. 3.21) Median aliphatic index in IgM memory PBMC sequences increased over the course of infection, from 33 at 4 weeks to 42 at 24 weeks, while in SIV-specific sequences aliphatic index decreased from 37 at 4 weeks to 31 at 24 weeks (Fig. 3.22). SIV specific sequences had slightly lower hydrophobicity than non-specific sequences, and net charge and isoelectric point were largely unchanged between the two groups (except at 24 week non-

specific IgM sequences, which had lower median isoelectric point and lower median net charge).

The aliphatic index of IgG sequences in lymph nodes was unchanged in non-specific sequences (median 39) but varied in SIV-specific IgG sequences, with a median of 38 at 4 weeks, 31 at 10 weeks, and 43 at 24 weeks (Fig. 3.23). Hydrophobicity was unchanged between SIV-specific and non-specific sequences at all time points, with medians between -0.080 and -0.098. No difference in hydrophobicity between Env-specific and total memory IgG B cells was previously reported by Wang et al (2016)[362] and our analysis confirms this finding. Net charge was unchanged in non-specific sequences, but increased in SIV-specific sequences from -0.97 at 4 weeks to -0.068 at 24 weeks and was indistinguishable from non-specific sequences at 24 weeks (-0.065). Isoelectric point was similar unchanged in non-specific sequences but varied in SIV-specific sequences, with a median of 5.5 at 4 weeks, 4.4 at 10 weeks and 6.2 at 24 weeks (compared to 6.2 in all non-specific sequences). IgM lymph node sequences showed no variation in aliphatic index or hydrophobicity (Fig. 3.24), but at 24 weeks lower median net charge (-1.06 vs. -0.064 at 4 weeks and 10 weeks) and median isoelectric point (5.3 vs. 6.3 and 6.2). In contrast, SIV-specific IgM lymph node heavy chain sequences, aliphatic index decreased from 49 at 4 weeks to 30 at 10 weeks and 33 at 24 weeks, while hydrophobicity, charge and isoelectric points remained constant throughout infection.

In bone marrow IgG sequences, isoelectric point was unchanged in non-specific heavy chains but increased from a median of 4.4 to median 6.1 in SIV-specific sequences (Fig 3.25). Median charge also increased in SIV-specific sequences, from median -1.06 at 4 weeks to -0.19 at 24 weeks. Hydrophobicity and aliphatic index were unchanged from SIV-specific to non-specific and throughout infection. Both SIV-specific and non-specific IgM bone marrow samples had a lower isoelectric point than IgM and IgG lymph node and PBMC samples at all time points. Bone marrow IgM SIV-specific samples also had a lower charge than both lymph node samples and PBMC samples, with IgM samples (both PBMC and lymph node) having higher charge than IgG samples.

3.2.10 Single cell sorting of B cells and comparison with bulk sorting

To evaluate the efficiency of the 5’ RACE bulk PCR, we re-sorted SIV-specific IgG memory B single cells from a PBMC sample and used a primer-based method to sequence their heavy chains. After sorting four plates (384 cells) and amplifying heavy chains using a nested PCR

with V-gene specific primers, 181 wells showed amplified sequences by gel electrophoresis. These wells were Sanger sequenced, and 165 had full-length sequences, of which 154 were productive rearrangements. There were 102 unique V-D-J combinations and 146 unique CDR3s for a total of 147 unique VDJ-CDR3 combinations. Cells from the same sample were sorted in bulk and amplified using the 5' RACE method, with a total of 1974 cells sorted yielding 145 unique VDJ combinations and 608 unique CDR3s (Table 3.8).

Table 3.8 Shared sequences between bulk and single cell sequencing.

Sample	Cells Sorted	Unique numseqs	Unique V	Unique J	Unique D	Unique VDJ	Number of CDR3lengths	Unique CDR3aaseqs	Unique vdj-CDR3
PBMCsc	288	165	12	8	26	102	19	146	147
PBMCbulk	1974	9651	15	9	30	144	18	603	697

Though the numbers of cells sequenced here represent only a small fraction of the SIV-specific repertoire, the distributions of V, D, and J genes not significantly different between the two methods of sequencing (Fig 3.15). Sequences were dominated by VH1, VH3, and VH4, with a small number of VH2, VH5, and VH7 sequences. This is in accordance with published comparisons of bulk and single cells sequencing[347]. The JH gene distribution had a bias toward J4 in the single cell that was not observed in the bulk samples. DH genes had high frequencies of families DH3, DH4, and DH6 in all samples, with single cell sequences also having higher frequencies of DH1 and DH2 than the bulk samples. There was a good correlation (p-0.0079 and p=0.0085) between the frequencies of the VH genes and DH genes in the two samples.

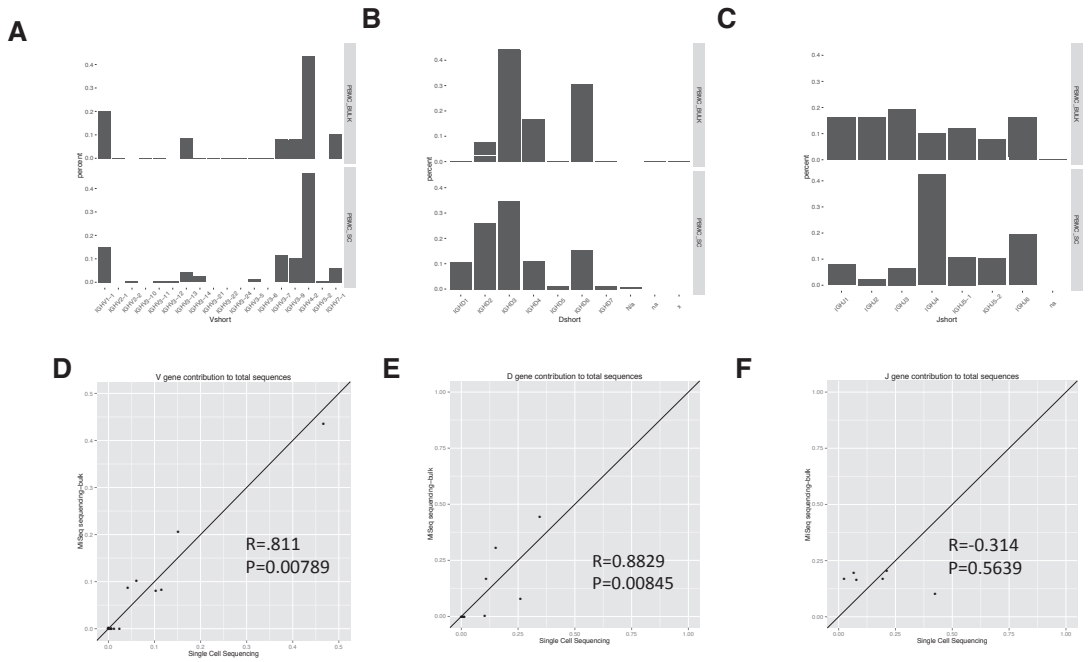


Fig. 3.26 V, D, and J gene frequencies in bulk and single cell sequencing. (A) VH gene proportion distribution from bulk (top) and single cell (bottom) SIV-specific PBMC. (B) DH gene proportion distribution from bulk (top) and single cell (bottom) SIV-specific PBMC. (C) JH gene proportion distribution from bulk (top) and single cell (bottom) SIV-specific PBMC. (D) Correlation between proportion distributions of VH genes in bulk and single cell sequencing. (E) Correlation between proportion distributions of DH genes in bulk and single cell sequencing. (F) Correlation between proportion distributions of JH genes in bulk and single cell sequencing.

The depth of sequencing in the bulk 5' RACE method yields data on the relative frequencies of individual clones, with identical sequences from multiple cells occurring at higher frequencies than reads from unique single cell BCR heavy chains. The sequences with the highest read counts showed some overlap with the single cell sequences, although there were some sequences that were detected in the bulk that were not detected in the single cell, and vice versa. This was partially due to mismatches between the V gene primers (used in single cell sequencing) and sequences in the bulk samples, with the result that the V gene primers were unable to properly amplify BCR heavy chains that were detected in the bulk samples (Fig 3.27). Three of the heavy chain sequences present at high frequencies in the bulk samples (16%, 10% and 8%) had at least one mismatch between the VH sequence and the inner and/or outer VH primers used in the single cell method. Thus, while the single cell method has the advantage of downstream pairing of heavy and light chains, primer mismatches can result in mutated sequences not being detected.

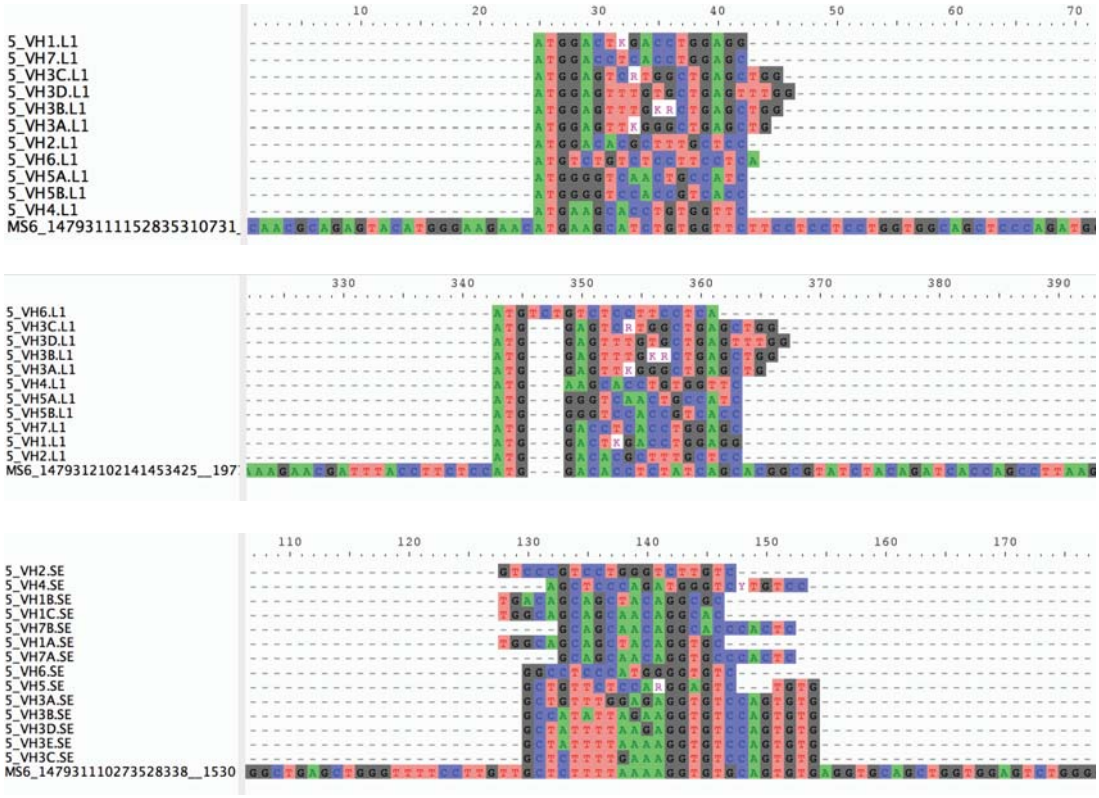


Fig 3.23 Single cell primer mismatches in sequences from bulk sequencing. Mismatches in the VH primer sets and the V gene sequences from heavy chains present at high frequencies in the bulk sequencing data.

3.3 Conclusions

This chapter summarizes the sequencing of SIV-specific and non-specific memory B cells throughout SIV infection of three rhesus macaques. One of the advantages of the macaque SIV infection model is the ability to serially sample compartments that are inaccessible in human studies, most prominently bone marrow and lymph nodes, to better capture the development of the humoral response to SIV. We identified IgG and IgM memory B cells that bound an SIV gp140 envelope protein probe and sequenced the heavy chain repertoires in blood, lymph nodes, and bone marrow. SIV-specific IgG B cells increased as fraction of all memory B cells over the course of infection in blood and lymph nodes, while SIV-specific IgM B cells were present at a steady and lower proportion. Median CDR3 length was unchanged in SIV-specific IgG and IgM heavy chains over the course of infection and was not significantly different than non-SIV specific memory B cells. However, the average mutation in the VH gene of SIV-specific IgG heavy chains shifted over the course of infection with an accumulation of more highly mutated heavy chains later in infection. While we observed some differences in the proportion distributions of VH and JH genes, we did not identify a particular signature of the SIV-specific response. We used several diversity measures to track the proportion and size of related antibody lineages, finding that SIV-specific IgG B cells were dominated by a small number of clones and had low diversity and low evenness early in infection, but resembled the non-specific memory repertoire later in

infection. Identical and closely related sequences were found across tissues and over the three time points sampled in this study, and a small percentage of similar IgG SIV-specific sequences were found in multiple animals. CDR3 sequences from SIV-specific IgG heavy chains in lymph nodes had small differences in net charge and isoelectric point when compared with non-SIV specific sequences. Finally, in a limited comparison of bulk 5' RACE PCR based sequencing versus single cell VH-primer directed sequencing, we identified efficiency of the single cell method and primer mismatches as primary reasons for differences in repertoires sequenced using the two methods.