OPEN

Clinical & Translational Immunology (2016) 5, e93; doi:10.1038/cti.2016.42
Official journal of the Australasian Society for Immunology Inc.
www.nature.com/cti

ORIGINAL ARTICLE

Repertoire comparison of the B-cell receptor-encoding loci in humans and rhesus macaques by next-generation sequencing

Vladimir Vigdorovich[1,4], Brian G Oliver[1,4], Sara Carbonetti[1], Nicholas Dambrauskas[1], Miles D Lange[1,5], Christina Yacoob[2], Will Leahy[3], Jonathan Callahan[3], Leonidas Stamatatos[2] and D Noah Sather[1]

Rhesus macaques (RMs) are a widely used model system for the study of vaccines, infectious diseases and microbial pathogenesis. Their value as a model lies in their close evolutionary relationship to humans, which, in theory, allows them to serve as a close approximation of the human immune system. However, despite their prominence as a human surrogate model system, many aspects of the RM immune system remain ill characterized. In particular, B cell-mediated immunity in macaques has not been sufficiently characterized, and the B-cell receptor-encoding loci have not been thoroughly annotated. To address these gaps, we analyzed the circulating heavy- and light-chain repertoires in humans and RMs by next-generation sequencing. By comparing V gene segment usage, J-segment usage and CDR3 lengths between the two species, we identified several important similarities and differences. These differences were especially notable in the IgM$^+$ B-cell repertoire. However, the class-switched, antigen-educated B-cell populations converged on a set of similar characteristics, implying similarities in how each species responds to antigen. Our study provides the first comprehensive overview of the circulating repertoires of the heavy- and light-chain sequences in RMs, and provides insight into how they may perform as a model system for B cell-mediated immunity in humans.

Clinical & Translational Immunology (2016) **5,** e93; doi:10.1038/cti.2016.42; published online 22 July 2016

Rhesus macaques (RMs) are the most widely used non-human primate (NHP) model for biomedical research, and are commonly used for the study of microbial pathogenesis, host–pathogen interaction and drug or vaccine efficacy.[1–8] Research focused on human diseases such as Parkinson's, Alzheimer's, cancers and diabetes, as well as infectious diseases such as malaria, tuberculosis, Dengue virus, Ebola virus and HIV-1/AIDS utilizes RMs as a model organism because they are believed to closely approximate biological systems found in human beings.[1,3,5,9–13] In particular, RMs have become a widely used model to study disease progression and host–pathogen interactions in HIV-1, as they are naturally susceptible to infection by Simian Immunodeficiency Virus (SIV), the closely related progenitor of HIV-1.[14,15] The development of chimeric HIV-1/SIV-1 viruses (SHIVs), which encode an HIV-1 Env glycoprotein in an SIV genetic backbone, also have allowed the model to be used for comprehensive vaccination and viral challenge studies using the exact HIV-1 Env immunogens that would ultimately advance to human clinical trials.[16–24] However, the use of RMs for preclinical vaccine studies remains somewhat problematic, as it is unclear how well the observed adaptive responses predict human immunity. In particular, B cell-mediated immunity is ill characterized in RMs. Thus, it is critical to understand how similar RM B cell-mediated immunity is to its human counterpart in order to assess the model's suitability for the preclinical evaluation of vaccines. Furthermore, it is important to identify where the two differ, so that the translation of preclinical experimental results in NHP to human clinical trials becomes more predictable.

The first completed RM draft genome was published in 2007 and enabled comprehensive comparisons between the macaque and human genomes, particularly those of the B-cell receptor (BCR)-encoding loci of each species.[25] Macaques share 93% overall sequence identity and, like humans, maintain three BCR-encoding loci (on chromosomes 7, 13 and 10).[26] However, significant changes have taken place since RMs and humans shared the last common ancestor—the IgH locus in humans and the IgK locus in RMs have expanded[25]—potentially altering the number of gene segments available for recombination. The use of the available RM genomic data is complicated by the incomplete annotation of the BCR loci and the presence of numerous sequence gaps (runs of N's).[25,27] In contrast, the human BCR-encoding loci (located on chromosomes 14,

[1]Center for Infectious Disease Research (formerly Seattle BioMed), Seattle, WA, USA; [2]Fred Hutchinson Cancer Research Center, Viral and Infectious Disease Division, Seattle, WA, USA and [3]Mazama Science, Seattle, WA, USA
[4]These two authors contributed equally to this work.
[5]Current address: Harry K. Dupree Stuttgart National Aquaculture Research Center, USDA-ARS, Stuttgart, AR, USA.
Correspondence: Dr DN Sather, Center for Infectious Disease Research, Seattle, WA 98109, USA.
E-mail: noah.sather@CIDResearch.org
Received 29 March 2016; revised 15 June 2016; accepted 16 June 2016

2 and 22) are well annotated, allowing for straightforward lineage prediction for expressed immunoglobulin (Ig) sequences. In both organisms, the Ig heavy-chain (IgH) locus is organized as a tandem array of variable (V), diversity (D) and joining (J) gene segments. These segments recombine during B-cell maturation into a mature V–(D)–J rearrangement that expresses the mature heavy chain[28] of the molecule. Similarly, the light-chain Ig kappa and lambda (IgK and IgL, respectively) loci are organized as arrays of V and J segments (there are no D segments in the light-chain loci) that recombine into a mature V–J rearrangement.[28]

The lack of coverage within the BCR-encoding genomic loci in RM has hampered efforts to understand the potential human context of vaccine-elicited B-cell responses, although there has been recent progress in identifying and evaluating potential BCR gene segments.[29–31] One drawback of these studies is the lack of unity within the published segment nomenclature, which leads to redundancies and can be a source of confusion. A set of heavy- and light-chain gene segment sequences is available through the 'Rhesus macaque Immunoglobulin gene database' (hosted by the King's College London School of Medicine), which is a collection of sequences from several sources and consists of 88 IGHV, 62 IGKV, 44 IGLV, 30 IGHD and 7 IGHJ segments. More recently, Sundling *et al.*[26,30] created an annotated list under a different nomenclature system, which consists of 63 IGHV, 62 IGKV, 50 IGLV, 30 IGHD, and 6 IGHJ, 5 IGKJ, 6 IGLJ gene segments. The international ImMunoGeneTics information system (IMGT) has begun to independently annotate the RM IgH locus, but there are currently 19 IGHV, 83 IGKV, 86 IGLV, 24 IGHD, 7 IGHJ, 4 IGKJ and 6 IGLJ gene segments available through this database.[32] Finally, a provisional set containing 98 RM IGHV alleles with a new nomenclature was recently reported.[33] Although it is uncertain that the available sets of reference sequences are exhaustive in coverage, they provide an indispensable foundation for annotating the circulating BCR repertoires in RM.

Until very recently, little was known about the makeup of the circulating B-cell repertoire in outbred RMs. Several recent studies utilized 454 pyrosequencing to define the circulating IGHV (heavy chain) gene family usage in IgG^+ B-cell repertoires after vaccination or infection in a small number of animals.[31,33] These studies provided unique insight into the heavy-chain IgG^+ BCR repertoires, but did not evaluate light-chain repertoires or the IgM heavy-chain repertoires. RM light-chain repertoires have only recently been described, but only in a single animal.[34] The lack of available data describing IgM^+ BCR repertoire is significant, as these B cells constitute the majority of the circulating BCR repertoire (~90% of circulating B cells in RMs), and are the cells that will initially respond to antigen after vaccination to give rise to the subsequent B-cell populations.[35] These studies represented significant advances in our understanding of RM BCR repertoires, but the broad applicability of such findings to larger outbred RM populations and to human BCR repertoires remains unclear. This information can only be gained through an unbiased and systematic evaluation of the BCR-encoding sequences in a larger number of humans and RM, as we report here.

Understanding the genetic constitution of the circulating RM BCR repertoire is even more critical for vaccine studies aimed at eliciting antibody responses with specific genetic characteristics, such as studies aimed at eliciting broadly neutralizing antibodies (bNAbs) against HIV-1. These antibodies are well characterized in humans and are known to have common features and genetic requirements.[36–43] For example, a class of anti-CD4-binding site (CD4-BS) antibodies exemplified by the mAb VRC01 has a strong genetic restriction to the human IGHV1-2*02 allele and utilizes a CDRL3 of five amino acids (AAs).[41,44,45] In addition, many bNAbs have long insertions in the third complementarity determining regions (CDR3's) in the heavy chains, which are thought to be assembled during VDJ rearrangement and are present in the naive circulating repertoires in humans.[46–48] Both of these examples are characteristics that enable the BCR to bind to complex epitopes on the HIV Env glycoprotein and would need to be present in the circulating BCR repertoire in RMs in order to elicit such antibodies during preclinical vaccine studies.

We sought to compare the BCR repertoires in humans and RM in order to assess the utility and potential drawbacks of the RM model system for the study of B cell-mediated immunity. To accomplish this, we developed an NGS and bioinformatics platform based on Illumina sequencing technology. We sequenced both the IgM^+ and IgG^+ heavy-chain repertoires and both the IgK and IgL light-chain repertoires in five outbred RMs and five healthy human subjects using a 5' random amplification of cDNA ends (RACE) method. To minimize the overrepresentation by clonally-expanded sequences and to eliminate errors introduced by amplification and sequencing steps, we clustered sequences by V-segment family, J-segment family and CDR3 amino-acid sequences and used these data sets to represent the circulating repertoires. Side by side comparisons of the repertoires showed that the heavy-chain repertoires were similar in segment family usage between the IgG^+ and IgM^+ subsets within species, although some notable differences were observed in cross-species comparisons. Our comparisons of the IgK and IgL light-chain repertoires revealed that humans and RMs had moderately discordant gene family usage. Overall, our study provides the first comparative, comprehensive evaluation of BCR repertoires in RMs and humans that covers both the IgM (which are mostly naive B cells) and class-switched B-cell populations and includes all three BCR-encoding sequences. These analyses highlight important similarities and differences in the circulating BCR repertoires between humans and RMs that are relevant to their continued use as a preclinical model system for human B cell-mediated immunity research.

RESULTS

A next-generation sequencing platform to assess BCR repertoires

The technologies to evaluate BCR repertoires by NGS have been advancing quickly over the last several years.[49] And yet, a robust, easily deployed NGS platform to sequence full-length BCRs is not widely available. As such, we set out to develop a comprehensive pipeline for BCR repertoire analysis based on Illumina sequencing technology that could be easily adapted for use in a variety of species. This sequencing platform has several advantages that make it well suited for use in BCR sequencing, including long read lengths (up to 600 nucleotides) and relatively low error rates during the sequencing phase.[50,51]

Our final platform utilized a modified 5' RACE method that enabled us to amplify and sequence the entire variable portions of the heavy and light chains in humans and RMs (Figure 1a). Of the various methods used for Ig library construction, 5' RACE is thought to produce the least biased libraries.[34,52] This is because it uses a single primer set (one universal and one gene-specific primer) for the amplification steps, rather than a multiplex or pooled primer method, in which a large number of primers may compete with one another during amplification. Therefore, it is likely that 5' RACE does not suffer from the amplification bias that accompanies the use of a large number of primer pairs.

For each of the three genetic loci in both humans and RM, we developed gene-specific primers for the initial cDNA first-strand synthesis, rather than using a polydeoxythymidine primer (oligo-dT).

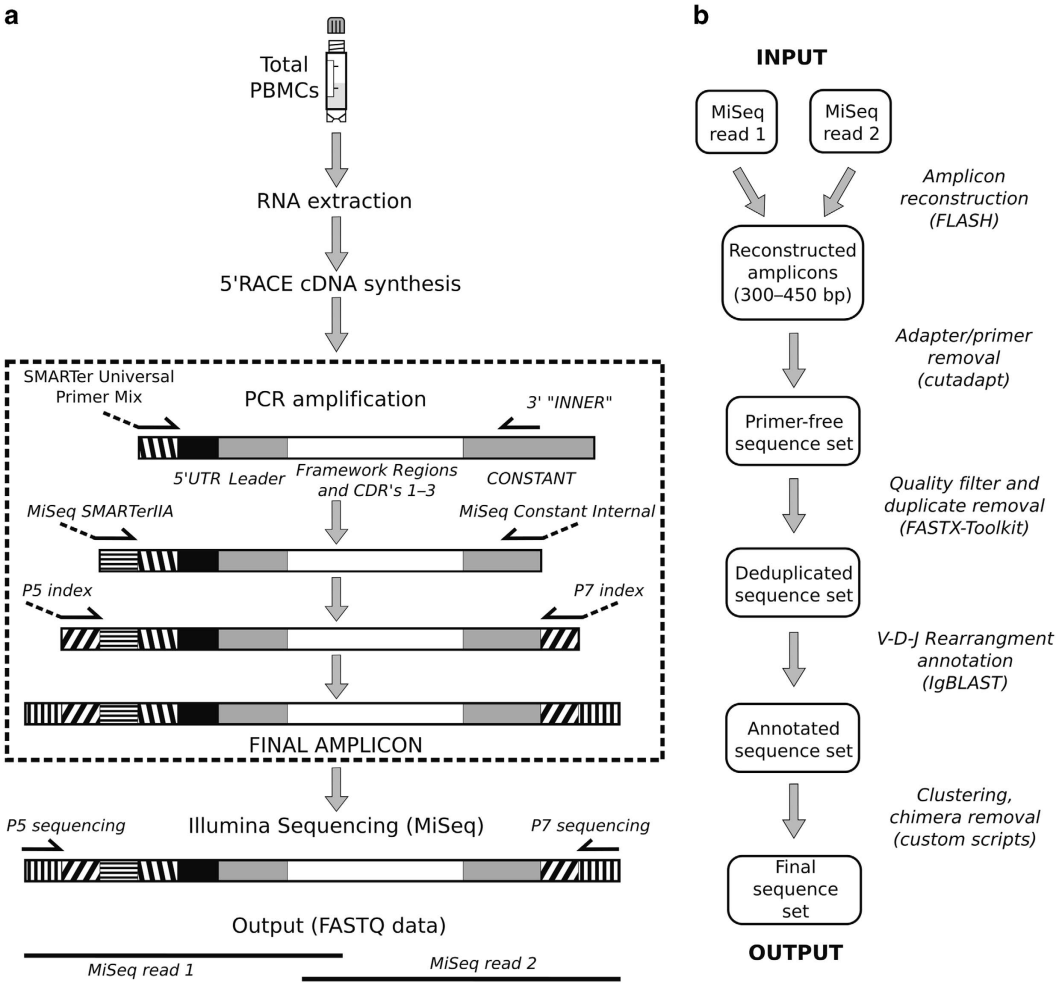

Figure 1 Sample preparation/analysis pipeline. Schematics depicting sample treatment and the PCR amplification steps required for library generation (**a**), and the computational pipeline used to filter and analyze the Illumina MiSeq output sets (**b**).

For PCR amplification steps, we developed custom reverse primers that were nested inside the synthesis primer (Supplementary Table S1). For the forward PCR primer, we utilized a universal primer in all reactions. The final step in library production was the addition of Illumina compatible linker sequences by a PCR refurbishing step that allowed capture of PCR products on the sequencing chip. Our reverse Illumina adapter primer was further nested inside the amplification primer, creating a protocol with two semi-nested amplification steps (Figure 1a). Within the IgH sequences, the reverse primers provided specificity between the IgM^+ and IgG^+ subsets. For IgM and IgG, the final PCR amplicons were ~550–650 bp in length, whereas the IgK and IgL amplicons were 500–600 bp long. Each amplicon covered the entire V–(D)–J rearrangement (all of framework regions (1–4) and CDRs (1–3)) and a portion of the 5′ UTR. The sequencing primers used by MiSeq generated forward and reverse reads ('MiSeq read 1' and 'MiSeq read 2', respectively, in Figure 1a) that overlapped in the middle of the amplicon, and allowed subsequent amplicon sequence elucidation (see below).

The raw FASTQ sequence data harvested from a sequencing run underwent a number of processing steps before comparative repertoire analysis. In the absence of a readily available informatics toolkit, we assembled an informatics pipeline that used currently available academic programs to facilitate the processing, finishing and quality control of each data set (detailed in Materials and Methods, and in Figure 1b). Raw sequences from forward and reverse reads were first matched to creating a single extended (forward-orientation) sequence, followed by the elimination of any reads in which the primer sequences could not be identified. Additional quality control steps discarded low-quality sequences, and coalesced duplicate sequences. Finally, the sequence sets were annotated with V, D and J segments usage and grouped into sequence clusters (based on CDR3 amino-acid sequence, V-segment family and J-segment family annotations, see below) in order to (1) identify and remove suspected chimeric sequences, (2) eliminate potential overrepresentation by dominant B-cell clones and (3) eliminate errors introduced by amplification and sequencing steps. The population diversity found within these clustered data sets is described in Supplementary Figure S1.

A comprehensive Ig gene segment reference set for RM was not available, necessitating the construction of a sequence database containing all of the previously-identified V, D and J segments. The published *Macaca mulatta* genome, as well as other sequencing efforts, has resulted in several publically available compilations of Ig heavy-chain sequences. The majority of the current genome sequence was obtained from one or two animals;[25] with the remaining identified sequences originating from a collection of different genomic scaffolds,[53] locus sequencing efforts[33] as well as sequences documented by the Rhesus macaque Immunoglobulin database (accession numbers DQ437773–855, AY161053–81, AF173903–32, AF417167–96,

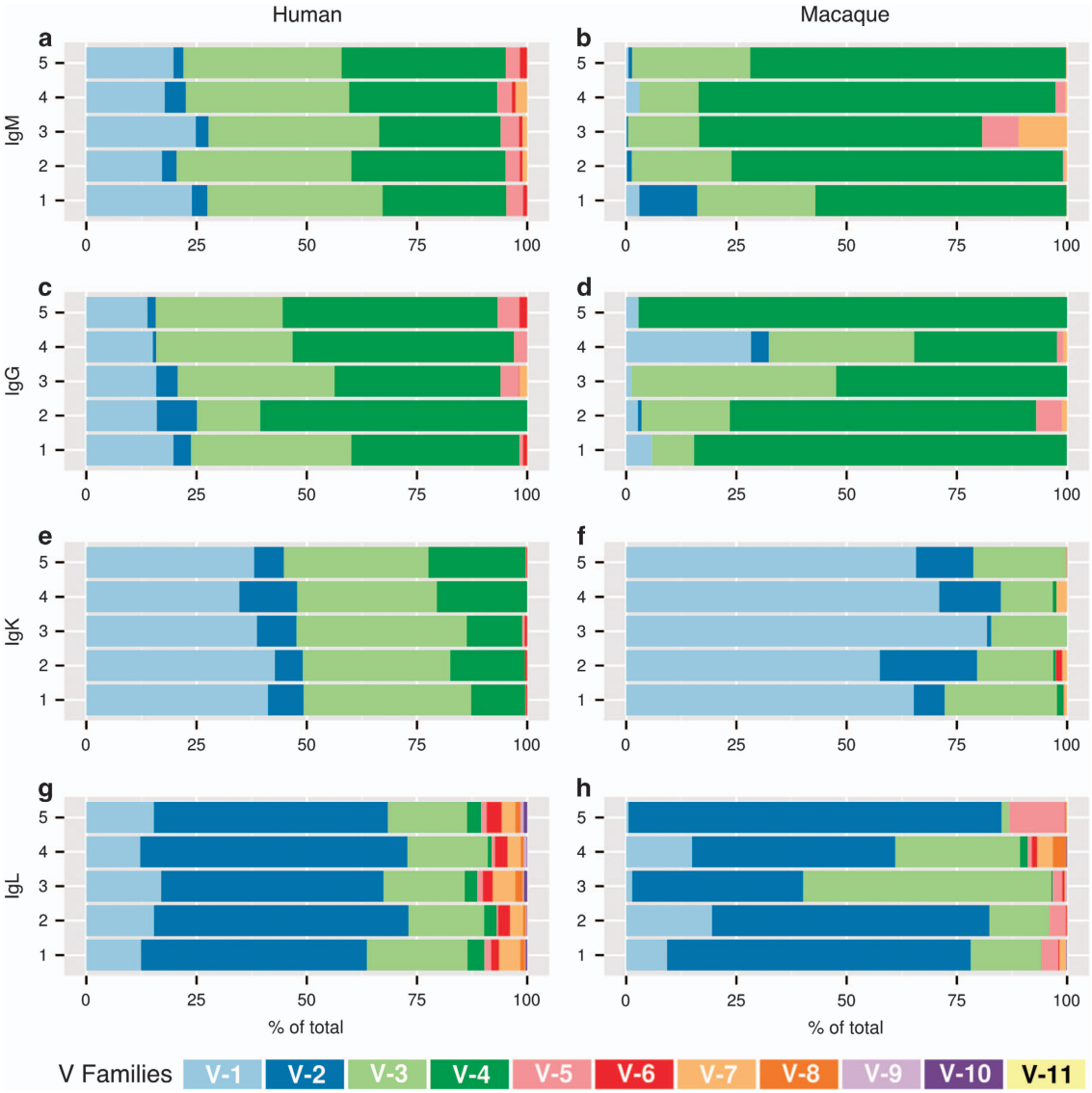

Figure 2 Distribution of V-family genes in the expressed IgM$^+$ (**a**,**b**), IgG$^+$ (**c**,**d**), IgK$^+$ (**e**,**f**), and IgL$^+$ (**g**,**h**) BCR repertoires of healthy human subjects (**a**,**c**,**e**,**g**) and untreated macaques (**b**,**d**,**f**,**h**). Productively rearranged sequences amplified from the expressed BCR repertoires for each human subject ($n=5$) or animal ($n=5$) were grouped into sequence clusters and are represented by horizontal bars. V-family assignments are color-coded (color assignments are described below the graphs) and scaled according to the proportion present in each data set.

U57580–7, AY057983, AY963709–73 and AM055970–6022). In absence of fully annotated genomic loci, it is currently difficult to reliably estimate the total number of germline gene segments. However, in order to utilize all available RM Ig information, we compiled all of the available non-redundant germline heavy- and light-chain gene segments for our analyses.[30,31,33] A list of germline gene segments identified according to their original designation, accession number, alternative nomenclature and closest human germline gene segment homolog is shown in Supplementary Table S2.

IGHV gene segment family usage in the circulating IgM and IgG BCR repertoires

We evaluated the assigned IGHV segment family usage within the IgM sequence libraries and compared the frequencies of segment family usage between humans and RMs (Figures 2a and b). To rule out an overestimation of IGHV gene segment usage within the different families due to clonal expansion, we focused our analysis on sequence clusters, rather than including all non-redundant sequences. Our clustering strategy grouped sequences with closely related V–(D)–J segment rearrangements; so that sequences with identical CDR3 amino-acid sequence and using V and J segments belonging to the same families were coalesced to form a single sequence cluster. Only clusters containing five or more sequences were included in the repertoire data sets. In this way, our system is able to correct for errors and reduce the potential bias in gene family usage that could be caused by the presence of a large number of closely related sequences, resulting from a few nucleotide differences due to PCR or sequencing error, or from somatic mutation in highly stimulated B-cell lineages. However, we also analyzed the non-redundant, non-clustered sequence sets to evaluate for sequence bias before clustering (Supplementary Figure S2). In all of the non-redundant sequence sets, the gene family usage approximated that of the clustered sequence sets, indicating that our results were likely an accurate representation of the actual circulating BCR repertoires in both species and that the

gene family distributions were not skewed due to over-abundant, closely related sequences that could have arisen from clonal expansion or sequencing errors.

In humans, the IGHV3 and IGHV4 gene families were the predominantly expressed segment families, present at ~38 and 32%, respectively (Figure 2a, Supplementary Figure S4A). The IGHV1 family was the next most abundant (~21%), with the remainder of the gene families combining to make up <9% of the repertoires (Figure 2a). In contrast, the RM IgM compartment was dominated by IGHV4 gene family, making up ~70% of the expressed IgM sequences (Figure 2b). IGHV3 was the next most abundant, averaging ~21%, whereas the remaining gene families IGHV1, 2, 5 and 7 combined to make up only ~9% of the expressed repertoires in macaques. We did not detect any IGHV6-family segment usage in macaques using 5' RACE amplification. However, IGHV6 transcripts were detectable in a standard RT-PCR containing IGHV6-specific primers (data not shown). Thus, it is likely that IGHV6-containing transcripts in macaques are present, but at an extremely low abundance. Comparatively, human repertoires contained significantly more IGHV1 ($P<0.0001$) and IGHV3 ($P<0.0001$) than RM repertoires, whereas RMs contained more IGHV4 ($P<0.0001$) (Supplementary Figure S4). Thus, the circulating IgM$^+$ repertoires in humans and RMs differed significantly in the three most prevalent gene families.

We also evaluated the IGHV segment family usage in circulating IgG$^+$ B cells in both humans and RMs. The IgG$^+$ repertoire represents a population of B cells that have experienced antigen stimulation, have undergone additional development, class-switching and proliferation, and are poised to rapidly respond to secondary antigenic stimulation during repeat infection.[54] In comparing the IgG$^+$ compartments, humans and RMs expressed very similar IGHV gene family repertoires, with the exception of IGHV4, which was more abundant in RM IgG repertoires ($P=0.013$) (Supplementary Figure S4). Within each

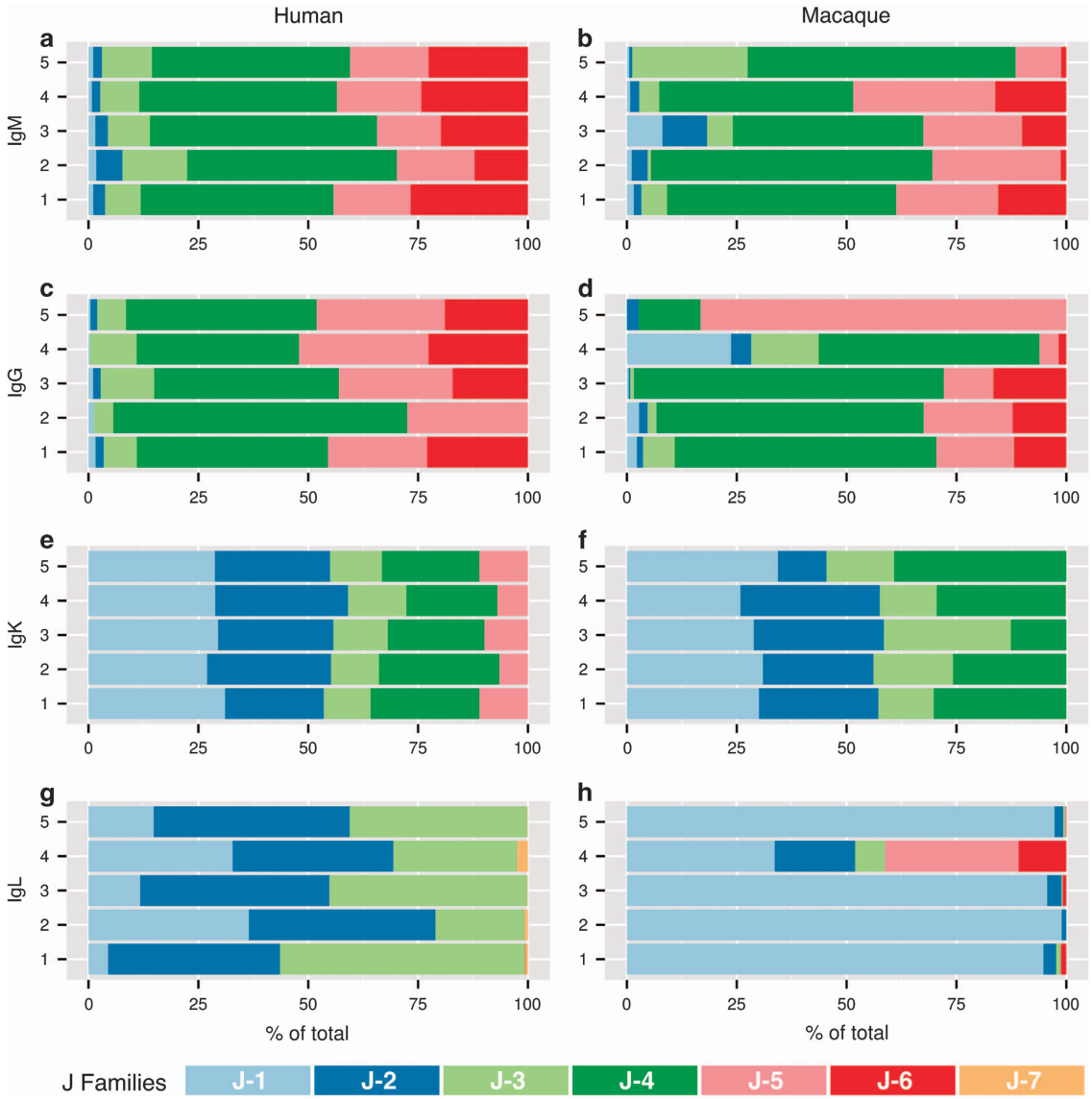

Figure 3 Distribution of J-family genes in the expressed IgM$^+$ (**a**,**b**), IgG$^+$ (**c**,**d**), IgK$^+$ (**e**,**f**), and IgL$^+$ (**g**,**h**) BCR repertoires of healthy human subjects (**a**,**c**,**e**,**g**) and untreated macaques (**b**,**d**,**f**,**h**). Productively rearranged sequences amplified from the expressed BCR repertoires for each human subject ($n=5$) or animal ($n=5$) were grouped into sequence clusters and are represented by horizontal bars. J-family assignments are color-coded (color assignments are described below the graphs) and scaled according to the proportion present in each data set.

species, we also compared the IGHV gene family distribution between the IgG and IgM repertoires. In humans, the IgG repertoires contained more IGHV4 ($P<0.0001$) and less IGHV3 ($P=0.0058$) than the IgM repertoires, indicating that IGHV4 may be more prominently stimulated into B-cell memory (Supplementary Figure S4). In macaques, the IgG and IgM repertoires expressed very similar distributions of IGHV families, with no statistically significant differences noted between the two B-cell compartments (Supplementary Figure S4).

Divergent V-segment family usage in the IgK, but not IgL, repertoires between RMs and humans

A mature BCR consists of one heavy chain encoded by the IgH locus and one light chain encoded by either the IgK or IgL locus. In humans, the IgK repertoires predominantly consisted of the IGKV1 and IGKV3 gene segment families, which combined accounted for nearly 75% of the repertoires in most of the subjects (Figure 2e). The IGKV4 and IGKV2 gene families accounted for ~16 and 8% of the human repertoires, respectively. The remaining gene families accounted for the remaining 1%, although we did not detect expression of IGKV7 genes in humans. In contrast, all IGKV gene families were detected in RMs (Figure 2f). The IgK repertoires in RM were predominantly IGKV1 (68%). The IGKV3 and IGKV2 families accounted for 18 and 11% of the repertoires, respectively, with the remaining gene families accounting for the remaining 3%. Comparatively, the human IgK repertoires expressed more IGKV3 ($P>0.0001$) and IGKV4 ($P>0.0001$) than RMs, whereas the RM repertoires contained significantly more IGKV1 ($P>0.0001$) (Supplementary Figure S4).

Interestingly, the IgL repertoires were remarkably similar between the two species. Both repertoires predominantly expressed the IGLV2, accounting for ~55% in humans and 60% in RM (Figures 2g and h). IGLV3 was the next most prevalent gene family in both species (19% in humans and 23% in RM), followed by IGLV1 (14% in humans and 9% in RMs). We did not detect IGLV11 gene family expression in humans, and detected it in only two out of five RMs at very low levels (<0.1%). The remaining seven IGLV gene families were present at <1–3% in the repertoires. Comparatively, we noted no statistically significant differences in the IgL gene family distributions between humans and RM (Supplementary Figure S4).

Humans and RMs express similar J-gene segment family distributions, with the exception of IgL

We also compared J-family usage for all three BCR-encoding sequences between humans and RMs. The J-gene segment is assembled into the V–(D)–J (or V–J) rearrangement during B-cell maturation and comprises part of the CDR3 at the D/J or V/J boundary. We found a high degree of similarity in the IGHJ gene segment family distributions between humans and RMs in three of the four repertoires we analyzed (Figure 3; compare with Supplementary Figure S3 for unclustered data). Heavy-chain sequences of both the IgG and IgM repertoires were dominated by IGHJ4 in humans and RMs (Figures 3a–d). IGHJ5 and IGHJ6 were the next two most commonly expressed segment families, although IGHJ6 was more heavily expressed in humans than IGHJ5 (Figures 3a and b), whereas RMs expressed more IGHJ5 than IGHJ6 (Figures 3b and d). The remaining IGHJ segment families together constituted an average of <10% of the overall repertoires. Comparatively, the only difference observed was that humans expressed more IGHJ6 than RM ($P=0.0086$) (Supplementary Figure S5). Thus, although significant differences in IGHV gene usage were observed between humans and RMs, these differences were not reflected in the incorporation of J-gene segments.

Similarly, the IGKJ gene segment family usage was remarkably concordant between humans and RMs (Figures 3e and f), despite the major differences we observed in IGKV segment usage (Figures 2e and f). IGKJ families 1, 2 and 4 were the most commonly expressed gene segment families in both species. Overall, the distribution of IGKJ gene usage was statistically similar between species with one exception: humans expressed more IGKJ5 than RMs ($P=0.0276$) (Supplementary Figure S5). In contrast, IGLJ segment family expression between RMs and humans was discordant (Figures 3g and h), despite the observation that the IGLV gene usage between species was statistically similar (Figures 2g and h). The IGLJ2 family was the most abundantly expressed segment family in humans, followed by the IGLJ1 and IGLJ3 families (Figure 3g). In RM, the IGLJ repertoires were predominantly IGLJ1 (Figure 3h). Comparatively, RMs expressed more IGLJ1 than humans ($P<0.0001$), whereas humans expressed more IGLJ2 ($P<0.0001$) and IGLJ3 ($P<0.0001$) than RMs (Supplementary Figure S5).

Analysis of CDR3 sequence lengths

For heavy-chain repertoires, the average length of CDRH3s in humans and RMs in both the IgM and IgG sequence sets was similar, averaging 13–15 AAs in length (Figures 4a–d). However, humans and RMs did not generate long CDRH3s (⩾20 AAs in length) with similar frequencies. Long CDRH3s in the IgM sequences were more frequently detected in humans than in RMs ($P=0.0005$, Supplementary Figure S6), with an average of 15.5% (±1.32%) of human IgM repertoires consisting of sequences with long CDRH3s, compared with 2.49% (±1.92%) in RMs. However, this is likely an overestimation for RMs, as one animal out of five had an unusually large number of sequences with long CDRH3s (~10% in the IgM) (Figure 4b, Supplementary Figure S6). Excluding this one animal, RMs averaged only 0.52% (±0.16%) long CDRH3. Interestingly, within in the IgG sequence sets the prevalence of long CDRH3 was not statistically different between humans and RMs ($P=0.58$), with the human sequence sets containing an average of 14.79% (±2.18%) and RMs an average of 16.72% (±2.5%). Thus, although CDRH3 lengths of ⩾20 AA were rare in the IgM^+ B-cell compartment in RMs, they appeared to populate the class-switched B-cell compartment at a similar rate to that of humans.

In the light-chain repertoires, humans and RMs expressed BCRs with the same average CDRL3 lengths (Figures 4e–h, Supplementary Figure S6). In both species, the average CDRL3 length in the IgK populations was 9 AAs, and 10 AAs in the IgL repertoires. Long CDRL3s (⩾13 AAs in length) were rare in both humans and RMs in the IgK sequence sets, making up <1% of the repertoires (Figures 4e and f, Supplementary Figure S6). In the IgL sequences, long CDRL3 was slightly more common, averaging ~2% of the repertoires (Figures 4g and h, Supplementary Figure S6). Short CDRL3s (⩽7 AAs in length) were also rare in the IgK and IgL repertoires, making up <0.3–0.6% of sequence populations. Interestingly, an exception was the set of human IgK repertoires, which averaged 1.74% (±0.36%) short CDRL3 (Figure 4e, inset, Supplementary Figure S6). Comparatively, humans expressed significantly more short CDRL3s than RMs ($P=0.0062$, Supplementary Figure S6).

Key sequence features within the BCR repertoires relevant to HIV-1 vaccine development

As noted above, some classes of anti-HIV-1 bNAbs have common defining characteristics that are essential to their broadly neutralizing

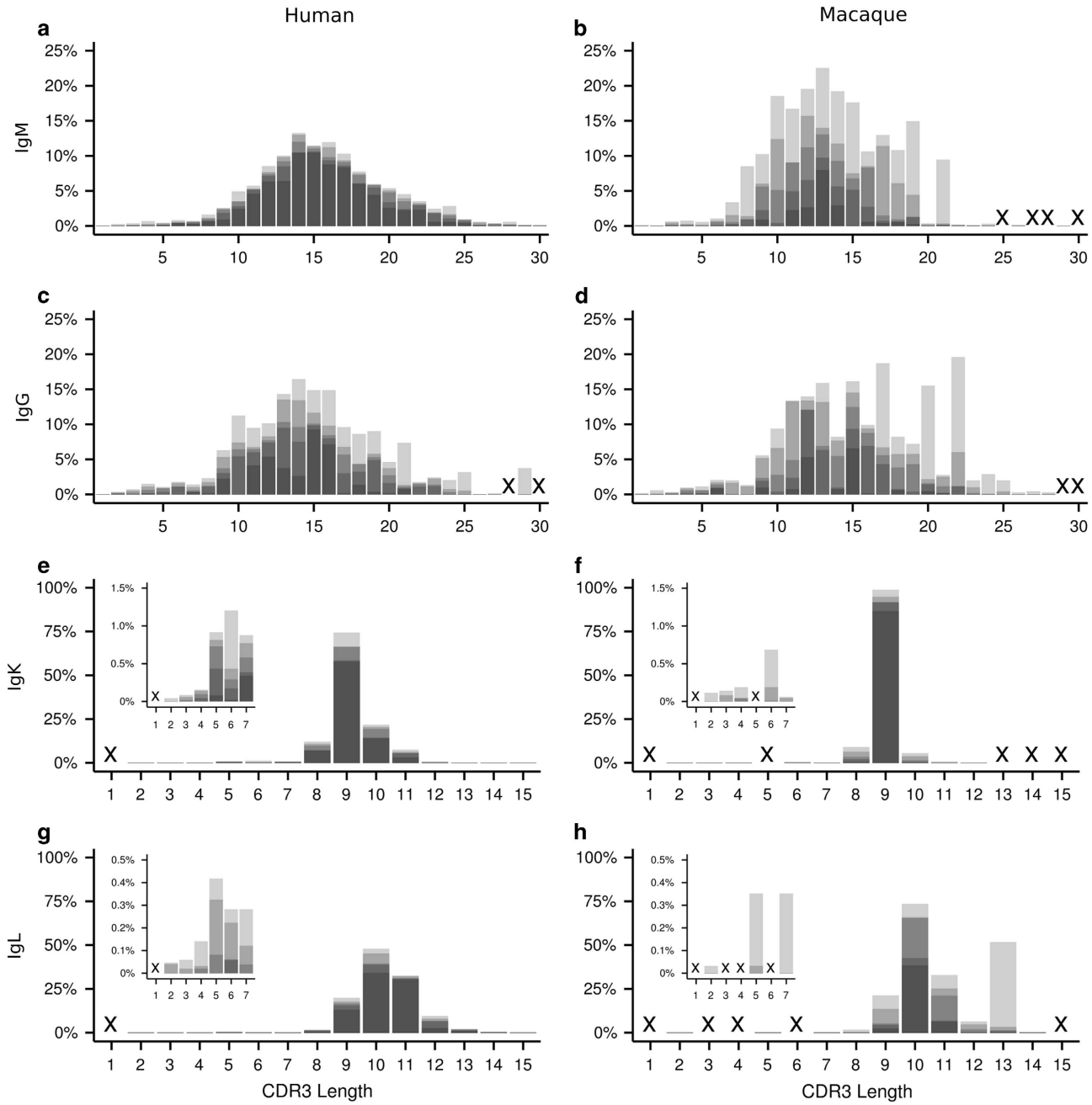

Figure 4 Distribution of CDR3 lengths in the expressed IgM$^+$ (**a**,**b**), IgG$^+$ (**c**,**d**), IgK$^+$ (**e**,**f**), and IgL$^+$ (**g**,**h**) BCR repertoires of healthy human subjects (**a**,**c**,**e**,**g**) and untreated macaques (**b**,**d**,**f**,**h**). Lengths of CDR3 amino-acid sequences were compiled for each of the human ($n=5$) and macaque ($n=5$) data sets and plotted as stacked transparent gray bars. Darker regions of the plots indicate occurrence in multiple data sets. Very low abundance outlier sequences (CDR3 lengths >30 AAs for VH, and >15 AAs for VL) were excluded from the plots. Short (<8 AAs) CDRL3 sequences present at lower frequencies are shown as insets (**e**–**h**). An 'x' denotes that no sequences were found containing CDR3s of that length.

activities, namely CDR3 length and genetic restriction. The collection of comparative NGS data sets on the BCR-encoding loci in humans and macaques provides an opportunity to evaluate the prevalence of these characteristics in the circulating BCR repertoires. As described above, we evaluated the BCR repertoires for the presence of long CDRH3s, which are a common characteristic of many bNAbs. We found that RMs express long CDRH3s in the IgM BCR populations with low frequency, whereas they are much more frequent in humans (around 15% of a repertoire, Supplementary Figure S6). However, in the IgG repertoires, RMs expressed long CDRH3s at a rate similar to humans, around 15% of the total IgG repertoires. This suggests that the rarity of IgM$^+$ B cells with long CDRH3s does not preclude their significant contribution to the memory compartment in RM. Thus, RMs are well suited to produce secreted antibodies with long CDRH3s. Indeed, monoclonal antibodies with long CDRH3s have been isolated from RMs previously.[33,55]

We also evaluated the BCR repertoires for two key features reported to be important for the development of the VRC01 class of anti-CD4-BS bNAbs: IGHV allelic restriction and an IgK CDRL3 of five AAs in length.[41,44,45] Five-AA CDRL3s in the IgK repertoires were detected in

all of the humans that we analyzed, making up 0.6% (95% CI [0.17,1.02]) of the total IgK repertoire (Figure 4e, inset). In contrast, we did not detect 5-AA CDRL3s in the IgK repertoires in any of the macaques (Figure 4f, inset). Thus, IgK 5-AA CDRL3s are likely so extraordinarily rare that they are not readily captured in routine unbiased analysis.

Additionally, VRC01-class bNAbs are known to be restricted to a specific human IgH allele, IGHV1-2*02.[41,44,45] This is thought to be due to the presence of three key AAs encoded in the germline sequence that interact with the HIV-1 Envelope protein, amino-acid positions W50, N58 and R71.[44] The IGHV1-2*02 allele was found to be expressed in four of five humans in this study at an average of 4.43% (95% CI [3.26,5.61]) of the total BCR heavy-chain repertoires (Supplementary Figure S7). This allele was also detected in the remaining human subject, but at extremely low frequency (0.0041%). IGHV1-Kl (Supplementary Table S2), which is also known as VH1.23,[56] is the closest homologous sequence in RMs to the human IGHV1-2*02 at 92% amino-acid sequence identity. Despite its close sequence relatedness, IGHV1-Kl encodes for only two of the three key AAs needed for VRC01 binding, and lacks a tryptophan at position 50. IGHV1-Kl was detected in 4/5 animals at low levels, with an average frequency of 0.62% (95% CI [−0.66,1.91]) (Supplementary Figure S7). These data imply that humans are well equipped within their IgM^+ B-cell compartments to develop VRC01-like antibodies with these characteristics. However, the RM homolog of IGHV1-2*02, which is missing a key amino-acid position, was far less abundant, and RM expression of IgK chains with five-AA CDRL3s was not detectable. Thus, RMs likely are not competent to develop antibodies with the characteristics typical of VRC01-class antibodies.

DISCUSSION

Given the importance of RMs for biomedical research, we sought to characterize the expressed BCR heavy- and light-chain repertoires in RMs and systematically compare them to human expressed BCR repertoires. Our goal was to provide foundational data that can be used by researchers to make rational assessments about the utility of RMs as a model system to study B cell-mediated immune response in humans. We developed an Illumina sequencing technology-based NGS B-cell analysis pipeline based on the bulk amplification of the expressed V–(D)–J (or V–J) rearrangements of the three BCR-encoding loci (IgH, IgK and IgL) from the IgM and IgG heavy-chain repertoires, and examined samples from five humans and five RMs in parallel. Concurrent sequencing of the four BCR chain repertoires separately in each subject enabled us to conduct a quantitative, population-based comparison. Our analyses included both V- and J-segment usage in the heavy- and light-chain repertoires, as well as CDR3 length comparisons. Our study provides the first comprehensive, comparative evaluation of the B cell heavy- and light-chain repertoires in humans and in outbred RMs, and provides insights about the translational potential of experimental results to human B cell-mediated immunity.

Characterization of BCR repertoires with unprecedented depth is now possible due to advances in NGS. Here, we report one of the most broad and detailed analyses of BCR repertoires to date in either humans or RMs, which allows us to compare repertoires among the individuals and between the species. In humans, the frequency of V- and J-gene segment usage among human subjects in the IgM, IgG, IgK and IgL repertoires was in good agreement, as were the distributions of CDR3 lengths. Similarly, with a few exceptions, the repertoires and CDR3 lengths were concordant among the individual RMs, especially in the IgM, IgK and IgL repertoires. The relative similarity of the BCR repertoires in individual subjects between and within species is remarkable, considering that our data represent snapshots of continuously changing populations of B cells that arose in unique immunological circumstances in each individual. In addition, the repertoires that we report here are in good agreement with those reported in previous studies for both humans and RMs.[34,40,52,57–61] Taken together, these findings provide confidence that the BCR repertoires we report here are representative of actual average circulating repertoires in both species, although the collection of more repertoire data will be needed to confirm this.

Comparative analyses of the IgM, IgG, IgK and IgL sequences in RMs and humans revealed some significant differences in the overall makeup of the circulating B-cell repertoires. Namely, gene segment family usage and the frequencies of CDR3s with lengths at the extremes of the size distribution differ moderately between RMs and humans in the IgM^+ B-cell compartment. Nevertheless, it may be that these differences are too subtle to have significant functional consequences for the utility of RM as a model system. Importantly, we found that (1) RMs express diverse BCR repertoires that contained sequences utilizing nearly all of the V- and J-gene families in both the heavy and light chains, (2) they possessed a significant range of CDR3 lengths and (3) these parameters were generally comparable to those we found in humans. Given the observed recombinatorial complexity and diversity of the B-cell repertoires in RMs and the close genetic relationship between the two species, it is likely that RM B-cell populations are poised to respond to antigen in ways similar to those of humans.

Indeed, the differences in V- and J-family distributions in the IgM^+ repertoires between humans and RMs largely resolved into concordant gene family usage in the class-switched (IgG) compartment (Supplementary Figure S4B). This convergence was not due to a remodeling in the gene family distribution in RMs, but rather reflected changes in gene family frequencies between the IgM and IgG repertoires in humans (Supplementary Figures S4E and F). Additionally, the low abundance of long CDRH3s in RM IgM^+ B cells did not preclude their relative expansion in the IgG^+ population, and they were present to a similar degree to human IgG repertoires (Supplementary Figure S6A). Thus, both the human and RM IgM populations underwent some degree of repertoire remodeling after antigen education, but converged on a set of similar characteristics that defined the IgG^+ B-cell populations. This level of convergence is remarkable, considering that the changes between the IgM and IgG cell populations in each species are driven by different natural histories of antigen exposure. Taken together, these observations imply that RMs possess the capacity to recapitulate a class-switched memory B-cell compartment similar to that of humans and are likely suitable model organisms for the study of B cell-mediated immunity.

Despite the general similarity, there are likely rare cases in which the RM model will not provide an adequate facsimile of human immunity, as described above for the VRC01-class of anti-HIV-1 antibodies. As such cases arise, in which specific sequence characteristics are necessary to gain the desired immune response, the competency of RMs to recapitulate these responses should not be simply assumed and should be evaluated directly. If these characteristics are genetic in nature, then continued sequencing efforts targeting the RM BCR-encoding loci will be required to evaluate the suitability of RM as a model. Nevertheless, such shortcomings do not imply that RMs are an inadequate model system for the study of HIV-1 bNAbs targeting in or around the CD4-BS in general, as bNAbs that do not use IGHV1-2*02 or do not require 5-AA CDRL3s have been described

in HIV-1-positive humans.[44,45,62,63] Further, the comparable abundance of BCRs with long CDRH3s in the RM IgG populations indicates that RMs can recapitulate that particular aspect of many anti-HIV-1 bNAbs. Indeed, bNAbs with similar specificities to those of humans have been isolated from RMs in previous studies, supporting the continued use of RMs as a model for the study of HIV-1 bNAbs.[31,34,56,64,65]

However, there are caveats associated with B-cell repertoire analysis by NGS. Although we have evaluated both the heavy- and light-chain repertoires, we did not have the ability to pair up natively-matched heavy- and light-chain pairs. Thus, we cannot say what fraction of the IgK and IgL repertoires comes from the memory B-cell compartment and we cannot evaluate the important dimension of the frequency of specific heavy- and light-chain pairings. Methods to do such analyses are now becoming available, but they are not widely used and their use for comprehensive, high-throughput analyses is not yet practical. Further, the humans included in this study were a mix of male and females, but they are all of North American origin. Recent evidence suggests far more geographic IGHV allelic variation than was previously thought.[66] Additionally, the makeup of B-cell repertoires can be influenced by sex and age, and although the animals used in this study were approximately the same age and life stage of the humans, the differences in actual age may confound our analysis. Reporting the repertoires as gene family (as we have done here) likely mitigates these variations, but inclusion of geographically disparate samples and larger age ranges in subsequent studies will be a necessary future step. Finally, lack of allelic coverage in RMs continues to be an issue in NGS analyses, although efforts are under way to further characterize the IgH locus.

Overall, our studies provide a comprehensive overview of the circulating IgM, IgG, IgK and IgL repertoires in outbred RMs, allowing for the first time an extensive global comparison with the circulating B-cell repertoires in humans. Our study revealed many striking similarities, indicating that RMs are a reasonable model system for human B cell-mediated immunity. However, our study also revealed several differences that may be relevant to the performance of RMs in specific cases. Further characterization of the BCR-encoding loci and functional analyses of B-cell repertoires will serve to clarify exact relationships between the humans and RMs, and allow for more extensive analyses of the translatability of experimental results to humans.

METHODS

Sample acquisition

Whole blood samples were obtained by venipuncture from five human subjects under CIDR human subjects protocol HS029, which was reviewed and approved by the Western Institutional Review Board. In addition, samples were obtained by venipuncture from five RMs of Indian origin under IACUC-approved protocols. The human subjects ranged in age from 22 to 32 years and were a mix of males and females. The macaques were all sexually mature males aged 4–6 years. Adjusted for relative lifespan, the humans and macaques in this study were roughly of the same life stage. Peripheral blood mononuclear cells (PBMCs) were isolated from whole blood by Ficoll density gradient separation. After two washes in RPMI containing 10% fetal bovine serum, PBMCs were slowly frozen in fetal bovine serum (Sigma, St Louis, MO, USA) supplemented with 10% DMSO in Mr. Frosty (Thermo Fisher Scientific, Waltham, MA, USA) cryo-containers. The RMs were housed at the Washington National Primate Research Center in Seattle, WA, USA. The care and all procedures were carried out under approved protocols monitored by the University of Washington Institutional Animal Care and Use Committee (IACUC).

Cell preparation

Macaque PBMCs were briefly thawed at 37 °C, diluted 10-fold in RPMI and centrifuged for 10 min at 400 *g*. Cells were washed once in phosphate-buffered saline, centrifuged again, resuspended in the FACS buffer (phosphate-buffered saline w/ 2% fetal bovine serum) and enumerated using a Countess II FL cell counter (Thermo Fisher Scientific). Cells were then resuspended in Buffer RLT (Qiagen, Germantown, MD, USA) and stored at −80 °C until RNA was extracted.

Library preparation for Illumina MiSeq

B-cell libraries for Illumina MiSeq sequencing were prepared, as follows. Total RNA was extracted from 5 to 10 million PBMCs using the AllPrep DNA/RNA Mini Kit (Qiagen). In an effort to generate unbiased B-cell libraries, cDNA synthesis was subsequently performed using the Takara Clontech SMARTer RACE cDNA Amplification Kit using primers with specificity to IgG, IgM, IgK and IgL. The subsequent RACE-ready cDNA was diluted in Tricine-EDTA according to the manufacturer's recommended protocol. First-round Ig-encoding sequence amplification was performed using AccuPrime *Pfx* Supermix (Invitrogen, Waltham, MA, USA), containing gene-specific primers (120 nM) and 1× concentration of Takara/Clontech 10× Universal primer mix. Amplicons were purified using FlashGels (Lonza, Allendale, NJ, USA) and used as templates for second-round PCR amplification. A second-round PCR amplification (10 cycles) was performed in order to add MiSeq adapter sequences to both ends of the amplicon. After re-purification, a final 5-cycle amplification was performed by adding P5 and P7 index sequences for Illumina sequencing. Purified, indexed libraries were quantitated using the KAPA library quantification kit (Kapa Biosystems, Wilmington, MA, USA) performed on an Applied Biosystems 7500 Fast real-time PCR machine (Applied Biosystems, Foster City, CA, USA).

Illumina MiSeq operation

Libraries were denatured and loaded onto Illumina 600-cycle V3 cartridges, according to the manufacturer's suggested workflow. Briefly, libraries were combined at equimolar concentrations, denatured and neutralized according to the manufacturer's protocol for 5 min at room temperature with freshly prepared 0.2 N NaOH. After incubation, the reaction was diluted with ice-cold, HT1 Hybridization Buffer (Illumina, San Diego, CA, USA). Illumina PhiX Control (12.5 pM) was used as an internal quality control according to the manufacturer's protocol at a concentration of 5% of the final, combined library volume.

Sequence analysis of human and RM IgH, IgL and IgK libraries

Raw data obtained from the forward and reverse MiSeq reads were used to reconstruct the amplicon using FLASH.[67] The resulting sequence sets were filtered to select only sequences that contained the amplification primers (a procedure during which the primer sequences themselves were removed) using cutadapt;[68] sequences containing low-confidence base calls (N's) were then removed from the set using FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit). Annotation of the resulting sequence sets was carried out using IgBLAST[69] against a customized database of macaque gene segment sequences listed in Supplementary Table S2 or against a similar human data set.

The IgBLAST annotation was used for identification of the closest-matching V and J segments for each sequence. D-segment annotations were excluded from this analysis due to the low confidence of alignment of such short sequences (frequently, this problem manifested as multiple D-segment matches with identical scores). CDR3 sequences were identified in the heavy-chain amino-acid sequences using the following strategy: sequences were scanned for an anchor sequence Y(Y/H/F)C (indicating the end of Framework 3) followed by an anchor sequence WGxG (with 'x' denoting any residue), and, failing that, a W alone; the sequence between the anchor sequences was harvested as CDRH3. Similarly, light-chain sequences were scanned for Y(Y/H/F)C and FG as anchor sequences, and the intervening sequence was harvested as CDRL3. Annotated sequences that were determined to be productively rearranged (that is, lacking stop codons and in-frame with the J-segment sequence) and containing the consensus motifs of Framework 4 (at least 30 000 sequences per data set) were grouped together based on identical CDR3 amino-acid

sequences and defined as sequence clusters. This procedure was further used to identify and filter away artifactual chimeric sequences, which differed in V-family and/or J-family assignment despite sharing the CDR3 sequence with their putative clusters. Furthermore, clustering was used to counteract potential dominance of expanded B-cell clones, and to eliminate sequence errors that arose during amplification and sequencing. We used the R-package *alakazam*[70] in order to characterize clustered populations by calculating the Hill diversity indices, as well as Shannon's entropy and evenness metrics (Supplementary Figure S1).

Statistical analyses

All statistical analyses were performed in GraphPad Prism 6.07 statistical software package (GraphPad Software Inc., La Jolla, CA, USA). The comparison of IgV and IgJ gene family frequency between humans and RMs was carried out using two-way ANOVA. Multiple comparisons were carried out using Sidak's multiple comparisons test (Alpha = 0.05) with Sidak's correction for multiple comparisons. Reported *P*-values are corrected for multiple comparisons. Characteristics of the CDRH3 and CDRL3 populations were compared by unpaired *t* test, assuming equal standard deviations (parametric). *P*-values of <0.05 were considered as significant. In those cases where averages are reported from the *t* test groups or two-way ANOVA, we also report the standard error of the mean (SEM), as denoted by the addition of ($\pm X\%$). In the figures describing statistical analyses, the error bars are the standard deviation. Simple averages in which the means between two populations were not compared statistically are reported as the average percent and the 95% confidence interval.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We gratefully acknowledge the veterinary staff at the Washington National Primate Research Center, Seattle WA, USA, including J Ogle and M Gough. We thank Dr F Matsen IV for his insightful critique of this manuscript. This study was funded by NIH/NIAID R33 AI089405 to DNS and R01 AI104384 to LS.

1 Evans DT, Silvestri G. Nonhuman primate models in AIDS research. *Curr Opin HIV AIDS* 2013; **8**: 255–261.
2 Mangus LM, Dorsey JL, Laast VA, Ringkamp M, Ebenezer GJ, Hauer P *et al.* Unraveling the pathogenesis of HIV peripheral neuropathy: insights from a simian immunodeficiency virus macaque model. *ILAR J* 2014; **54**: 296–303.
3 Peña JC, Ho W-Z. Monkey models of tuberculosis: lessons learned. *Infect Immun* 2015; **83**: 852–862.
4 Potts LF, Wu H, Singh A, Marcilla I, Luquin MR, Papa SM. Modeling Parkinson's disease in monkeys for translational studies, a critical analysis. *Exp Neurol* 2014; **256**: 133–143.
5 Pound LD, Kievit P, Grove KL. The nonhuman primate as a model for type 2 diabetes. *Curr Opin Endocrinol Diabetes Obes* 2014; **21**: 89–94.
6 Vallender EJ, Miller GM. Nonhuman primate models in the genomic era: a paradigm shift. *ILAR J* 2013; **54**: 154–165.
7 Vierboom M, Breedveld E, 't Hart BA. New drug discovery strategies for rheumatoid arthritis: a niche for nonhuman primate models to address systemic complications in inflammatory arthritis. *Expert Opin Drug Discov* 2012; **7**: 315–325.
8 Wang F. Nonhuman primate models for Epstein-Barr virus infection. *Curr Opin Virol* 2013; **3**: 233–237.
9 Clark KB, Onlamoon N, Hsiao H-M, Perng GC, Villinger F. Can non-human primates serve as models for investigating dengue disease pathogenesis? *Front Microbiol* 2013; **4**: 305.
10 Martins K, Cooper C, Warren T, Wells J, Bell T, Raymond J *et al.* Characterization of clinical and immunological parameters during Ebola virus infection of rhesus macaques. *Viral Immunol* 2015; **28**: 32–41.
11 Shen S, Pyo C-W, Vu Q, Wang R, Geraghty DE. The essential detail: the genetics and genomics of the primate immune response. *ILAR J* 2013; **54**: 181–195.
12 Sodora DL, Allan JS, Apetrei C, Brenchley JM, Douek DC, Else JG *et al.* Toward an AIDS vaccine: lessons from natural simian immunodeficiency virus infections of African nonhuman primate hosts. *Nat Med* 2009; **15**: 861–865.
13 Weinfurter JT, May GE, Soma T, Hessell AJ, León EJ, Macnair CE *et al.* Macaque long-term nonprogressors resist superinfection with multiple CD8+ T cell escape variants of simian immunodeficiency virus. *J Virol* 2011; **85**: 530–541.
14 Schmitz JE, Korioth-Schmitz B. Immunopathogenesis of simian immunodeficiency virus infection in nonhuman primates. *Curr Opin HIV AIDS* 2013; **8**: 273–279.
15 Zhou Y, Bao R, Haigwood NL, Persidsky Y, Ho WZ. SIV infection of rhesus macaques of Chinese origin: a suitable model for HIV infection in humans. *Retrovirology* 2013; **10**: 89.
16 Chen Z, Zhao X, Huang Y, Gettie A, Ba L, Blanchard J *et al.* CD4+ lymphocytopenia in acute infection of Asian macaques by a vaginally transmissible subtype-C, CCR5-tropic Simian/Human Immunodeficiency Virus (SHIV). *J Acquir Immune Defic Syndr* 2002; **30**: 133–145.
17 Gautam R, Nishimura Y, Lee WR, Donau O, Buckler-White A, Shingai M *et al.* Pathogenicity and mucosal transmissibility of the R5-tropic simian/human immunodeficiency virus SHIV(AD8) in rhesus macaques: implications for use in vaccine studies. *J Virol* 2012; **86**: 8516–8526.
18 Harouse JM, Gettie A, Tan RC, Blanchard J, Cheng-Mayer C. Distinct pathogenic sequela in rhesus macaques infected with CCR5 or CXCR4 utilizing SHIVs. *Science* 1999; **284**: 816–819.
19 Hessell AJ, Hangartner L, Hunter M, Havenith CE, Beurskens FJ, Bakker JM *et al.* Fc receptor but not complement binding is important in antibody protection against HIV. *Nature* 2007; **449**: 101–104.
20 Hessell AJ, Poignard P, Hunter M, Hangartner L, Tehrani DM, Bleeker WK *et al.* Effective, low-titer antibody protection against low-dose repeated mucosal SHIV challenge in macaques. *Nat Med* 2009; **15**: 951–954.
21 Hsu M, Harouse JM, Gettie A, Buckner C, Blanchard J, Cheng-Mayer C. Increased mucosal transmission but not enhanced pathogenicity of the CCR5-tropic, simian AIDS-inducing simian/human immunodeficiency virus SHIV(SF162P3) maps to envelope gp120. *J Virol* 2003; **77**: 989–998.
22 Humbert M, Rasmussen RA, Song R, Ong H, Sharma P, Chenine AL *et al.* SHIV-1157i and passaged progeny viruses encoding R5 HIV-1 clade C env cause AIDS in rhesus monkeys. *Retrovirology* 2008; **5**: 94.
23 Luciw PA, Pratt-Lowe E, Shaw KE, Levy JA, Cheng-Mayer C. Persistent infection of rhesus macaques with T-cell-line-tropic and macrophage-tropic clones of simian/human immunodeficiency viruses (SHIV). *Proc Natl Acad Sci USA* 1995; **92**: 7490–7494.
24 Siddappa NB, Watkins JD, Wassermann KJ, Song R, Wang W, Kramer VG *et al.* R5 clade C SHIV strains with tier 1 or 2 neutralization sensitivity: tools to dissect env evolution and to develop AIDS vaccines in primate models. *PLoS One* 2010; **5**: e11689.
25 Rhesus Macaque Genome Sequencing and Analysis Consortium, Gibbs RA, Rogers J, Katze MG, Bumgarner R, Weinstock GM *et al.* Evolutionary and biomedical insights from the rhesus macaque genome. *Science* 2007; **316**: 222–234.
26 Sundling C, Li Y, Huynh N, Poulsen C, Wilson R, O'Dell S *et al.* High-resolution definition of vaccine-elicited B cell responses against the HIV primary receptor binding site. *Sci Transl Med* 2012; **4**: 142ra96.
27 Zimin AV, Cornish AS, Maudhoo MD, Gibbs RM, Zhang X, Pandey S *et al.* A new rhesus macaque assembly and annotation for next-generation sequencing analyses. *Biol Direct* 2014; **9**: 20.
28 Market E, Papavasiliou FN. V(D)J recombination and the evolution of the adaptive immune system. *PLoS Biol* 2003; **1**: E16.
29 Sundling C, Martinez P, Soldemo M, Spångberg M, Bengtsson KL, Stertman L *et al.* Immunization of macaques with soluble HIV type 1 and influenza virus envelope glycoproteins results in a similarly rapid contraction of peripheral B-cell responses after boosting. *J Infect Dis* 2013; **207**: 426–431.
30 Sundling C, Phad G, Douagi I, Navis M, Karlsson Hedestam GB. Isolation of antibody V(D)J sequences from single cell sorted rhesus macaque B cells. *J Immunol Methods* 2012; **386**: 85–93.
31 Sundling C, Zhang Z, Phad GE, Sheng Z, Wang Y, Mascola JR *et al.* Single-cell and deep sequencing of IgG-switched macaque B cells reveal a diverse Ig repertoire following immunization. *J Immunol* 2014; **192**: 3637–3644.
32 Lefranc M-P, Ehrenmann F, Ginestoux C, Giudicelli V, Duroux P. Use of IMGT(®) databases and tools for antibody engineering and humanization. *Methods Mol Biol* 2012; **907**: 3–37.
33 Francica JR, Sheng Z, Zhang Z, Nishimura Y, Shingai M, Ramesh *et al.* Analysis of immunoglobulin transcripts and hypermutation following SHIV(AD8) infection and protein-plus-adjuvant immunization. *Nat Commun* 2015; **6**: 6565.
34 Dai K, He L, Khan SN, O'Dell S, McKee K, Tran K *et al.* Rhesus macaque b-cell responses to an HIV-1 trimer vaccine revealed by unbiased longitudinal repertoire analysis. *MBio* 2015; **6**: e01375–01315.
35 Neumann B, Klippert A, Raue K, Sopper S, Stahl-Hennig C. Characterization of B and plasma cells in blood, bone marrow, and secondary lymphoid organs of rhesus macaques by multicolor flow cytometry. *J Leukoc Biol* 2015; **97**: 19–30.
36 Scheid JF, Mouquet H, Feldhahn N, Seaman MS, Velinzon K, Pietzsch J *et al.* Broad diversity of neutralizing antibodies isolated from memory B cells in HIV-infected individuals. *Nature* 2009; **458**: 636–640.
37 Walker LM, Huber M, Doores KJ, Falkowska E, Pejchal R, Julien JP *et al.* Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature* 2011; **477**: 466–470.
38 McLellan JS, Pancera M, Carrico C, Gorman J, Julien J-P, Khayat R *et al.* Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. *Nature* 2011; **480**: 336–343.
39 Walker LM, Phogat SK, Chan-Hui PY, Wagner D, Phung P, Goss JL *et al.* Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. *Science* 2009; **326**: 285–289.
40 Scanlan CN, Pantophlet R, Wormald MR, Ollmann Saphire E, Stanfield R, Wilson IA *et al.* The broadly neutralizing anti-human immunodeficiency virus type 1 antibody

2G12 recognizes a cluster of alpha1−>2 mannose residues on the outer face of gp120. *J Virol* 2002; **76**: 7306–7321.
41 Wu X, Yang ZY, Li Y, Hogerkorp CM, Schief WR, Seaman MS *et al.* Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science* 2010; **329**: 856–861.
42 Huang J, Ofek G, Laub L, Louder MK, Doria-Rose NA, Longo NS *et al.* Broad and potent neutralization of HIV-1 by a gp41-specific human antibody. *Nature* 2012; **491**: 406–412.
43 Bonsignori M, Hwang K-K, Chen X, Tsao C-Y, Morris L, Gray E *et al.* Analysis of a clonal lineage of HIV-1 envelope V2/V3 conformational epitope-specific broadly neutralizing antibodies and their inferred unmutated common ancestors. *J Virol* 2011; **85**: 9998–10009.
44 West AP, Diskin R, Nussenzweig MC, Bjorkman PJ. Structural basis for germ-line gene usage of a potent class of antibodies targeting the CD4-binding site of HIV-1 gp120. *Proc Natl Acad Sci USA* 2012; **109**: E2083–E2090.
45 Scheid JF, Mouquet H, Ueberheide B, Diskin R, Klein F, Oliveira TY *et al.* Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science* 2011; **333**: 1633–1637.
46 Briney BS, Willis JR, Finn JA, McKinney BA, Crowe JE. Tissue-specific expressed antibody variable gene repertoires. *PLoS One* 2014; **9**: e100839.
47 Briney BS, Willis JR, Hicar MD, Thomas JW, Crowe JE. Frequency and genetic characterization of V(DD)J recombinants in the human peripheral blood antibody repertoire. *Immunology* 2012; **137**: 56–64.
48 Doria-Rose NA, Schramm CA, Gorman J, Moore PL, Bhiman JN, DeKosky BJ *et al.* Developmental pathway for potent V1V2-directed HIV-neutralizing antibodies. *Nature* 2014; **509**: 55–62.
49 Georgiou G, Ippolito GC, Beausang J, Busse CE, Wardemann H, Quake SR. The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat Biotechnol* 2014; **32**: 158–168.
50 Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* 2013; **79**: 5112–5120.
51 Schirmer M, Ijaz UZ, D'Amore R, Hall N, Sloan WT, Quince C. Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Res* 2015; **43**: e37.
52 He L, Sok D, Azadnia P, Hsueh J, Landais E, Simek M *et al.* Toward a more accurate view of human B-cell repertoire by next-generation sequencing, unbiased repertoire capture and single-molecule barcoding. *Sci Rep* 2014; **4**: 6778.
53 Yan G, Zhang G, Fang X, Zhang Y, Li C, Ling F *et al.* Genome sequencing and comparison of two nonhuman primate animal models, the cynomolgus and Chinese rhesus macaques. *Nat Biotechnol* 2011; **29**: 1019–1023.
54 Eibel H, Kraus H, Sic H, Kienzler A-K, Rizzi M. B cell biology: an overview. *Curr Allergy Asthma Rep* 2014; **14**: 434.
55 Yamamoto T, Lynch RM, Gautam R, Matus-Nicodemos R, Schmidt SD, Boswell KL *et al.* Quality and quantity of TFH cells are critical for broad antibody development in SHIVAD8 infection. *Sci Transl Med* 2015; **7**: 298ra120.
56 Navis M, Tran K, Bale S, Phad GE, Guenaga J, Wilson R *et al.* HIV-1 receptor binding site-directed antibodies using a VH1-2 gene segment orthologue are activated by Env trimer immunization. *PLoS Pathog* 2014; **10**: e1004337.
57 O'Connell AE, Volpi S, Dobbs K, Fiorini C, Tsitsikov E, de Boer H *et al.* Next generation sequencing reveals skewing of the T and B cell receptor repertoires in patients with wiskott-Aldrich syndrome. *Front Immunol* 2014; **5**: 340.
58 Galson JD, Clutterbuck EA, Trück J, Ramasamy MN, Münz M, Fowler *et al.* BCR repertoire sequencing: different patterns of B-cell activation after two Meningococcal vaccines. *Immunol Cell Biol* 2015; **93**: 885–895.
59 Bagnara D, Squillario M, Kipling D, Mora T, Walczak AM, Da Silva L *et al.* A Reassessment of IgM Memory Subsets in Humans. *J Immunol* 2015; **195**: 3716–3724.
60 Wang C, Liu Y, Cavanagh MM, Le Saux S, Qi Q, Roskin KM *et al.* B-cell repertoire responses to varicella-zoster vaccination in human identical twins. *Proc Natl Acad Sci USA* 2015; **112**: 500–505.
61 Cortina-Ceballos B, Godoy-Lozano EE, Téllez-Sosa J, Ovilla-Muñoz M, Sámano-Sánchez H, Aguilar-Salgado *et al.* Longitudinal analysis of the peripheral B cell repertoire reveals unique effects of immunization with a new influenza virus strain. *Genome Med* 2015; **7**: 124.
62 Liao HX, Lynch R, Zhou T, Gao F, Alam SM, Boyd SD *et al.* Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature* 2013; **496**: 469–476.
63 Georgiev IS, Doria-Rose NA, Zhou T, Kwon YD, Staupe RP, Moquin S *et al.* Delineating antibody recognition in polyclonal sera from patterns of HIV-1 isolate neutralization. *Science* 2013; **340**: 751–756.
64 Walker LM, Sok D, Nishimura Y, Donau O, Sadjadpour R, Gautam R *et al.* Rapid development of glycan-specific, broad, and potent anti-HIV-1 gp120 neutralizing antibodies in an R5 SIV/HIV chimeric virus infected macaque. *Proc Natl Acad Sci USA* 2011; **108**: 20125–20129.
65 Phad GE, Vázquez Bernat N, Feng Y, Ingale J, Martinez Murillo PA, O'Dell S *et al.* Diverse antibody genetic and recognition properties revealed following HIV-1 envelope glycoprotein immunization. *J Immunol* 2015; **194**: 5903–5914.
66 Scheepers C, Shrestha RK, Lambson BE, Jackson KJL, Wright IA, Naicker D *et al.* Ability to develop broadly neutralizing HIV-1 antibodies is not restricted by the germline Ig gene repertoire. *J Immunol* 2015; **194**: 4371–4378.
67 Magoč T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 2011; **27**: 2957–2963.
68 Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 2011; **17**: 10.
69 Ye J, Ma N, Madden TL, Ostell JM. IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res* 2013; **41**: W34–W40.
70 Gupta NT, Vander Heiden JA, Uduman M, Gadala-Maria D, Yaari G, Kleinstein SH. Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. *Bioinformatics* 2015; **31**: 3356–3358.

The Supplementary Information that accompanies this paper is available on the Clinical and Translational Immunology website (http://www.nature.com/cti)