RESEARCH ARTICLE

Rhesus Macaque B-Cell Responses to an HIV-1 Trimer Vaccine Revealed by Unbiased Longitudinal Repertoire Analysis

Kaifan Dai,[a] Linling He,[a] Salar N. Khan,[a] Sijy O'Dell,[b] Krisha McKee,[b] Karen Tran,[c] Yuxing Li,[d] Christopher Sundling,[e] Charles D. Morris,[a] John R. Mascola,[b] Gunilla B. Karlsson Hedestam,[e] Richard T. Wyatt,[a,c,f] Jiang Zhu[a,f,g]

Department of Immunology and Microbial Science, The Scripps Research Institute, La Jolla, California, USA[a]; Vaccine Research Center, National Institutes of Health, Bethesda, Maryland, USA[b]; IAVI Neutralizing Antibody Center, The Scripps Research Institute, La Jolla, California, USA[c]; Institute for Bioscience and Biotechnology Research, University of Maryland, Rockville, Maryland, USA[d]; Department of Microbiology and Tumor Cell Biology, Karolinska Institutet, Stockholm, Sweden[e]; Center for HIV/AIDS Vaccine Immunology and Immunogen Discovery, The Scripps Research Institute, La Jolla, California, USA[f]; Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, California, USA[g]

K.D. and L.H. contributed equally to this work

ABSTRACT **Next-generation sequencing (NGS) has been used to investigate the diversity and maturation of broadly neutralizing antibodies (bNAbs) in HIV-1-infected individuals. However, the application of NGS to the preclinical assessment of human vaccines, particularly the monitoring of vaccine-induced B-cell responses in a nonhuman primate (NHP) model, has not been reported. Here, we present a longitudinal NGS analysis of memory B-cell responses to an HIV-1 trimer vaccine in a macaque that has been extensively studied by single B-cell sorting and antibody characterization. We first established an NHP antibodyomics pipeline using the available 454 pyrosequencing data from this macaque and developed a protocol to sequence the NHP antibody repertoire in an unbiased manner. Using these methods, we then analyzed memory B-cell repertoires at four time points of NHP immunization and traced the lineages of seven CD4-binding site (CD4bs)-directed monoclonal antibodies previously isolated from this macaque. Longitudinal analysis revealed distinct patterns of B-cell lineage development in response to an HIV-1 trimer vaccine. While the temporal B-cell repertoire profiles and lineage patterns provide a baseline for comparison with forthcoming HIV-1 trimer vaccines, the newly developed NHP antibody NGS technologies and antibodyomics tools will facilitate future evaluation of human vaccine candidates.**

IMPORTANCE **The nonhuman primate model has been widely used in the preclinical assessment of human vaccines. Next-generation sequencing of B-cell repertoires provides a quantitative tool to analyze B-cell responses to a vaccine. In this study, the longitudinal B-cell repertoire analysis of a rhesus macaque immunized with an HIV-1 trimer vaccine revealed complex B-cell lineage patterns and showed the potential to facilitate the evaluation of future HIV-1 vaccines. The repertoire sequencing technologies and antibodyomics methods reported here can be extended to vaccine development for other human pathogens utilizing the nonhuman primate model.**

Received 10 August 2015 **Accepted** 12 October 2015 **Published** 3 November 2015

Citation Dai K, He L, Khan SN, O'Dell S, McKee K, Tran K, Li Y, Sundling C, Morris CD, Mascola JR, Karlsson Hedestam GB, Wyatt RT, Zhu J. 2015. Rhesus macaque B-cell responses to an HIV-1 trimer vaccine revealed by unbiased longitudinal repertoire analysis. mBio 6(6):e01375-15. doi:10.1128/mBio.01375-15.

Editor Michael G. Katze, University of Washington

Address correspondence to Richard T. Wyatt, wyatt@scripps.edu, or Jiang Zhu, jiang@scripps.edu.

In a rational vaccine strategy proposed for human immunodeficiency virus type 1 (HIV-1), identification of broadly neutralizing antibodies (bNAbs) plays a guiding role in structure-based immunogen design (1–3). Owing to the advances in experimental methods of B-cell isolation, a large panel of bNAbs has been obtained from HIV-1-infected individuals (4–15), defining multiple sites of HIV-1 vulnerability (16–18). Accompanying the discovery of bNAbs is the burgeoning use of next-generation sequencing (NGS) for antibody repertoire analysis, which has significantly advanced our understanding of the diversity and maturation of bNAbs (4, 10, 11, 19–23). As the antibody NGS technology matures, new applications are emerging in the field of vaccine development. One such application is to monitor the antibody responses in nonhuman primates (NHPs) in the preclinical assessment of human vaccine candidates.

Rhesus macaques can recapitulate many salient features of HIV-1 infection in humans and have been widely used as an NHP model in HIV-1 vaccine research (24–29). In the antibody-based vaccine paradigm, evaluation of envelope glycoprotein (Env)-based vaccine candidates in an NHP model can accelerate the selection of promising immunogens that merit advancement into clinical trials. However, much of the NHP B-cell biology has not been well understood until relatively recently. Sundling et al. investigated the memory and plasma B-cell responses of rhesus macaques to a soluble HIV-1 trimer vaccine using a regimen of five inoculations followed by a heterologous challenge (30). This study established a baseline of HIV-1 Env-specific B-cell responses in NHPs. In follow-up studies, Sundling et al. analyzed the rhesus immunoglobulin (Ig) loci and designed an experimental protocol, as well as gene-specific primers, for the cloning of antibody V(D)J

sequences from sorted B cells, which enabled the identification of a set of CD4-binding site (CD4bs)-directed monoclonal antibodies (MAbs) from a vaccinated macaque designated F128 (31, 32). The findings from this study complement the extensive investigation of the VRC01-class bNAbs targeting the CD4bs in HIV-1-infected human donors (4, 7, 10, 13, 23, 33). Sundling et al. then analyzed IgG-switched heavy chains of three V gene families (VH1, VH3, and VH4) as well as the D and J gene usage by 454 pyrosequencing of peripheral blood mononuclear cells (PBMCs) and compared the gene usage to that derived from single B-cell sorting of total and antigen-specific IgG-switched memory B cells (34). Most recently, Guo et al. investigated the dynamics of VH1, VH3, and VH4 gene diversity and somatic hypermutation (SHM) in NHPs following SIVmac239 infection using a paired-end sequencing method (35). Collectively, these studies provide a wealth of information on primate B-cell responses to trimeric Envs that is highly relevant in light of the recent progress in the development of improved Env-based HIV-1 vaccines such as the soluble, cleaved BG505 SOSIP.664 trimer and other forthcoming well-folded Env trimers (36–40).

In this study, we first incorporated the current rhesus macaque germline gene database (32) into the framework of a human antibodyomics pipeline to create an equivalent method for NHP repertoire analysis. We calibrated the new NHP pipeline and examined how the choice of germline gene database affected repertoire and lineage analyses using available 454 sequencing data from the well-studied NHP F128 (30, 32, 34). We then combined a 5′ RACE (rapid amplification of cDNA ends) PCR protocol and the long-read Ion Torrent personal genome machine (PGM) platform (41) to capture F128 memory B-cell repertoires after 2, 3, 4, and 5 immunizations. We determined the general repertoire profiles and traced the lineages of seven previously isolated CD4bs-directed MAbs (32) in the repertoires of four time points, providing a temporal view of the NHP memory B-cell responses to an HIV-1 trimer vaccine. The distinct lineage patterns observed highlight the complexity of Env-specific B-cell responses in an NHP model. We also conducted additional sequencing to determine the effect of experimental factors such as library variation, sequencing variation, and sequencing depth on the derived lineage patterns. Our study thus presents enabling technologies as well as a robust template for longitudinal analysis of vaccine-induced B-cell responses in NHPs, with important implications for the preclinical assessment of human vaccines.

RESULTS

F128 MAbs and an NHP antibodyomics pipeline. In the previous study of rhesus macaques immunized five times with a YU2 gp140 foldon (gp140-F) trimer (30), F128 was one of the animals that displayed the most potent neutralizing titers against a panel of cross-clade tier 1 isolates. A total of eight functional MAbs directed to the conserved CD4bs were isolated from F128. Among these MAbs, GE136, GE140, and GE143 neutralized SF162 and MN, while GE136 and GE143 also neutralized non-clade B viruses. Sequence analysis based on the current database of rhesus macaque germline genes (32) revealed a degree of SHM ranging from 1.4 to 8.4% at the nucleotide level, with the heavy chain complementarity-determining region 3 (HCDR3) varying between 15 and 21 amino acids (aa). In the current study, seven MAbs (GE121, GE136, GE137, GE140, GE143, GE147, and GE148) were selected to query the NGS-derived F128 antibody repertoires, with GE125 being excluded due to its similarity to GE121.

We developed an NHP antibodyomics pipeline based on the human antibodyomics pipeline that has been widely used to study HIV-1 bNAbs (4, 10, 19–21, 41). Rhesus H, κ, and λ germline gene segments—including the variable (V), diverse (D), and joining (J) segments—reported by Sundling et al. (32) (termed the "CS germline gene database" hereafter) were incorporated into the pipeline to facilitate germline gene assignment, error correction, and determination of H/LCDR3 and variable region boundaries (see Materials and Methods). Sixty-three V_H genes along with 60 D_H and 7 J_H genes were included for heavy chain analysis, while 62 V_K genes with 5 J_K genes and 50 V_L genes with 6 J_L genes were included for κ and λ light chain analysis, respectively. The resulting NHP antibodyomics pipeline was first validated using the seven F128 MAbs, resulting in an SHM level of 1.4 to 7.1% for heavy chains and 1.8 to 12.2% for light chains based on the CS germline gene database, consistent with the results from manual alignment (32). A pipeline based on the IMGT rhesus macaque germline gene database (http://www.imgt.org/), which contains 23 V_H, 32 D_H, and 8 J_H gene segments, was implemented for purposes of comparison.

Antibodyomics analysis of the 454 pyrosequencing data. We reanalyzed the 454 sequencing data (34) using both NHP antibodyomics pipelines (see Fig. S1 in the supplemental material). Of 846,736 reads, 797,794 and 798,333 can be assigned to NHP germline V_H genes defined within the CS and IMGT germline gene databases, respectively. After error correction using the respective germline genes as templates (21), the sequences were then compared to the seven MAb heavy chains at both nucleotide and amino acid levels. After pipeline processing, an additional filter was used to eliminate sequences containing PCR errors, resulting in 657,713 and 657,603 full-length heavy chain variable region sequences for the two germline gene databases, respectively. From the data processed by the CS germline gene database, we identified 163,921 (24.9%), 220,529 (33.5%), and 201,647 (30.6%) heavy chains of VH1, VH3, and VH4 origins, with VH2 and VH7 genes accounting for 1.7 and 9.2% of the population, respectively (Fig. 1A, top). Of note, the VH1, VH3, and VH4 germline gene families obtained from this NHP antibodyomics pipeline are 2.4- to 8.1-fold larger than reported previously (34), providing a more complete data set for subsequent analysis. From the data processed by the IMGT germline gene database, a highly skewed usage of IgHV1 (24.9%) and IgHV4 (30.7%) was observed, likely due to incorrect gene assignment (Fig. 1A, bottom left). The germline divergence distributions calculated from the CS and IMGT germline gene databases showed a marked difference, with averages of 7.1 and 11.7%, respectively (Fig. 1A, bottom middle). In contrast, the choice of germline gene database appeared to have no effect on the HCDR3 distribution, with an average loop length of 12.4 aa (Kabat definition) for both data sets (Fig. 1A, bottom right).

We then searched the reprocessed 454 sequencing data to identify the somatic variants of seven MAbs using an HCDR3 identity cutoff of 95% or greater. Such analysis is a "trademark" of human bNAb studies (4, 10, 19–23, 41) but not yet demonstrated for NHP MAbs. A two-dimensional (2D) identity-divergence analysis was performed to visualize the F128 heavy chain population with the identified somatic variants shown on the plots (Fig. 1B). Irrespective of the germline gene database used, the same somatic variants were identified for GE136 ($n = 1$), GE137 ($n = 3$), GE140 ($n =$

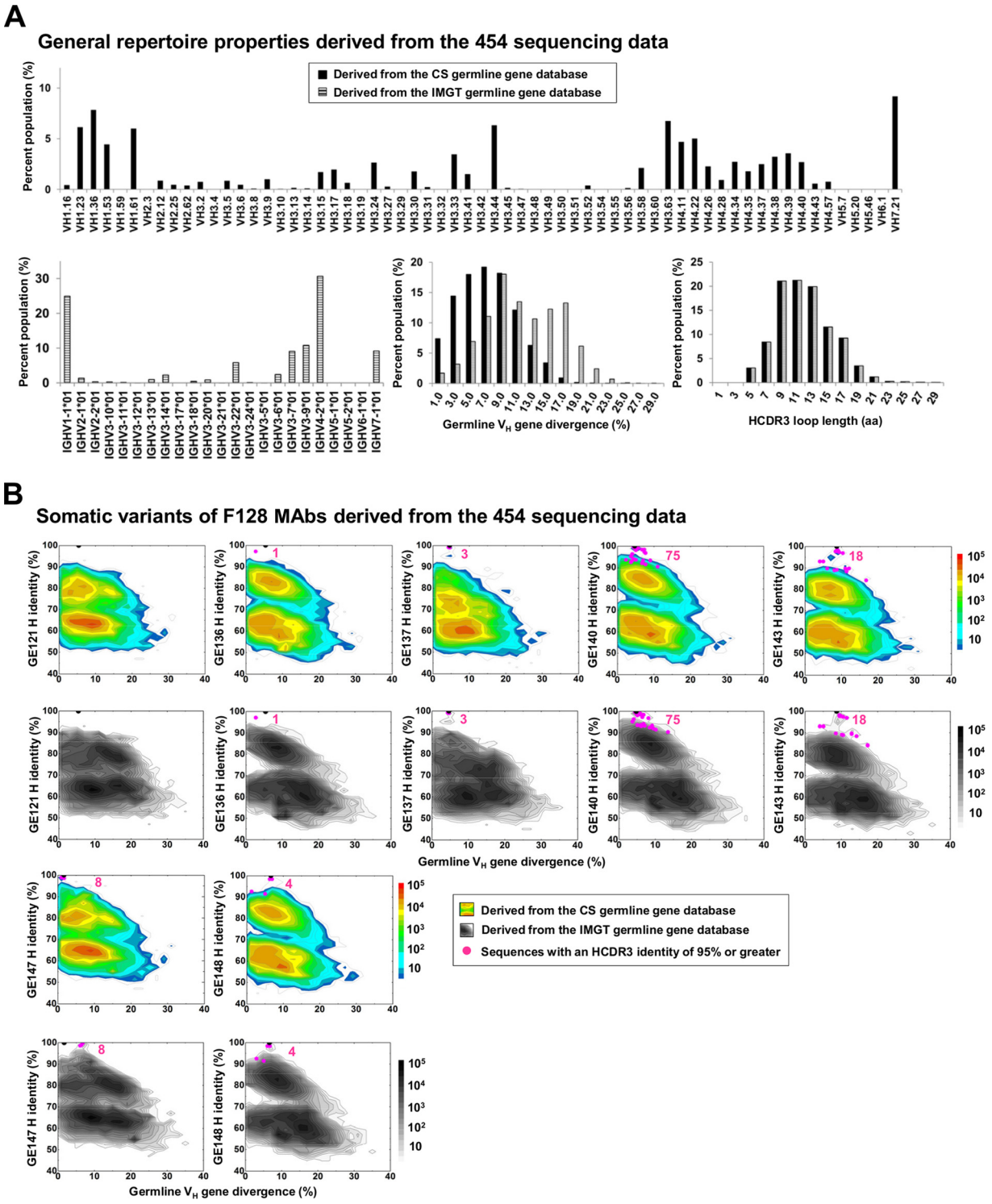

FIG 1 Analysis of reported 454 sequencing data of F128 heavy chains. (A) General repertoire properties determined from the 454-derived F128 heavy chain population, including germline gene usage, germline gene divergence, and the length of heavy chain complementarity determining region 3 (HCDR3). The 454 sequencing data were reanalyzed by the NHP antibodyomics pipeline using either the CS germline gene database (black histogram) or the IMGT germline gene database (gray histogram). (B) Two-dimensional (2D) identity-divergence analysis of the 454-derived F128 heavy chain population. Seven CD4-binding site (CD4bs)-directed monoclonal antibodies (MAbs) isolated from F128 were used as the template in the sequence identity calculation. The heavy chains are plotted as a function of sequence identity to the template and sequence divergence from putative germline V genes. The template and the somatic variants identified by an HCDR3 identity cutoff of 95% or greater are shown as black and magenta dots, respectively, with the number of somatic variants provided on the 2D plots. For contour plots derived from the CS and IMGT germline gene databases, the density of sequence distribution is indicated by color coding and darkness of the shading, respectively.

75), GE143 ($n = 18$), GE147 ($n = 8$), and GE148 ($n = 4$), with no relatives found for GE121. Of note, the 2D plots calculated from the IMGT germline gene database appeared to be elongated along the divergence ($X-$) axis, consistent with the shift of germline divergence distribution (Fig. 1A, bottom middle). Overall, the identity-divergence plots validated the NHP antibodyomics pipeline and provided initial insight into the somatic population of these F128 MAbs. However, the ontogeny of these MAb lineages and their development during the course of vaccination are currently unclear, and will likely require a longitudinal repertoire analysis.

Taken together, the results of reanalysis of 454 sequencing data demonstrated a robust performance of the NHP antibodyomics pipeline, as indicated by the output during data processing as well as the distributions and 2D plots derived from the processed data (Fig. 1). More importantly, the relative completeness of the NHP germline gene database showed no effect on the HCDR3 calculation, which is the basis of antibody lineage analysis. Therefore, the NHP antibodyomics pipeline, in conjunction with the most current CS germline gene database, provides a reliable platform for both NHP B-cell repertoire profiling and lineage tracing, which will be used hereafter to analyze immune responses following F128 immunization.

Utility of 5′ RACE PCR for unbiased capture of NHP antibody repertoire. The unbiased analysis of human B-cell repertoires has been reported (41, 42). Such methods can in principle be adapted to monitor the vaccine-induced B-cell responses in NHPs, as illustrated schematically in Fig. 2A. For F128, blood samples were collected 1 and 2 weeks after each inoculation with an additional blood draw 3 weeks after the second and fourth inoculations. Given the F128 sample availability, four time points (after immunizations 2, 3, 4, and 5; referred to as post-2, -3, -4, and -5) were chosen to investigate the memory B-cell responses to an HIV-1 Env immunogen (Fig. 2B, NGS). In template preparation, the use of 5′ RACE PCR is critical to unbiased repertoire analysis but has not been demonstrated for NHPs. Here, we designed an experimental protocol to prepare NHP antibody libraries from purified PBMCs (Fig. 2C). In brief, a 5′ RACE adaptor was attached to the 5′ end of the antibody cDNA molecules by reverse transcription (RT). Three 3′ reverse primers, the IgH inner primer, the Igκ outer primer, and the Igλ outer primer, were adapted from the nested PCR protocol (31) to amplify the 5′ RACE adaptor-tagged H, κ, and λ chains in PCRs (see Table S1 in the supplemental material). These primers have been used to isolate hundreds of NHP MAbs in previous studies (30, 32, 34). The use of a single 3′ reverse primer ensured unbiased repertoire capture for each antibody chain type. As shown previously, the 5′ RACE PCR products of human antibody transcripts are ~600 bp, exceeding the optimal sequencing length of most NGS platforms (41). The combined use of isothermal amplification (IA), a high-quality (Hi-Q) sequencing enzyme, and a modified PGM protocol has enabled a more accurate analysis of human antibody repertoires (41). However, it is unclear whether this advance is applicable to the NHP antibody libraries. To address this issue, we characterized the human and NHP antibody libraries generated by different PCR methods using gel electrophoresis and PGM sequencing (Fig. 2D). For a human donor, the multiplex PCR products showed three bands ranging from 400 to 450 bp, with an average read length of 428 bp from PGM sequencing. In contrast, the 5′ RACE PCR products of human H, κ, and λ chains yielded bands of similar size (600 bp) when prepared from an aliquot of the same PBMC sample, with an average read length of 571 bp from PGM sequencing. For F128, the multiplex PCR products showed three bands with sizes similar to those of the human donor (400 to 450 bp), with an average read length of 437 bp. The 5′ RACE PCR products of NHP H, κ, and λ chains showed three bands of 600 to 700 bp on the gel, with an average read length of 635 bp from PGM sequencing. The longer reads from the 5′ RACE PCR-prepared NHP antibody library can be attributed to the use of further downstream primers, a library barcode (10 bp), a full-length PGM P1 adaptor (18 bp longer than the truncated P1 adaptor), and an increased flow number in PGM sequencing. Nonetheless, our results highlight the importance of long-read NGS capability for an unbiased analysis of human and NHP B-cell repertoires alike.

Analysis of the F128 repertoire during immunization. PBMCs were obtained from F128 at four different time points to prepare H, κ, and λ chain libraries from the 5′ RACE PCR products (Fig. 2C; see also Table S1 in the supplemental material) and to confirm sample-to-sample reproducibility. We pooled the libraries from two time points on a single chip, 316 or 318 v2, and repeated the PGM sequencing twice to reduce potential experimental noise. In total, four PGM experiments generated over 22 million reads (Tables 1 and 2). Each of the four PGM data sets was divided into three subsets (H, κ, and λ) based on germline V gene assignment, with each subset being processed by a chain-specific NHP antibodyomics pipeline (see Fig. S2 in the supplemental material), yielding an average read length of 606 to 645 bp. Following correction for potential PCR errors, 80 to 90% of the reads were of sufficient length to cover the entire V(D)J reading frame (Table 1). The error-corrected, annotated sequences were merged into four data sets, each representing one time point and containing the H, κ, and λ libraries (Table 2). Greater sequencing depth and unbiased repertoire capture were expected to enable a more accurate comparison of repertoire properties across multiple time points.

We first analyzed the germline gene usage of F128 memory B-cell repertoires at four time points spanning the course of immunization (Fig. 3A). Since total PBMCs were taken 2 or 3 weeks after each inoculation per the original experimental design, the germline gene usage is mostly indicative of total memory B-cell responses. Overall, the germline gene usage appeared to be relatively constant during the immunization regimen. The heavy chain repertoire exhibited more fluctuations than the two light chain repertoires, suggesting that vaccine-induced changes in heavy chain gene usage and SHM might be detectable above contributions from the total resting memory B-cell compartment. In particular, the VH4 usage increased from 44.9 to 52.4% between the post-2 and -3 time points (see Fig. S3 in the supplemental material), while VH3.44 showed a notable decrease during the same 3-week period (Fig. 3), potentially resulting from oligoclonal B-cell expansion. The frequencies of VH1, VH3, and VH4 after 5 immunizations appeared to be different than those obtained from the 454 sequencing (Fig. 1A), in which the antibody libraries were prepared from three separate gene-specific primer reactions (34). For light chains, VK1 constituted ~60% of the κ chain repertoires, whereas VL1, VL2, and VL3 each accounted for 18.3 to 37.0% of the λ chain repertoires (see Fig. S3). Of note, the κ chains showed less than a 2.5% change between any two consecutive time points, compared to maximal changes of 6.9 and 3.0% for VL3 and VL5, respectively. Given the unbiased nature of this analysis, the germ-

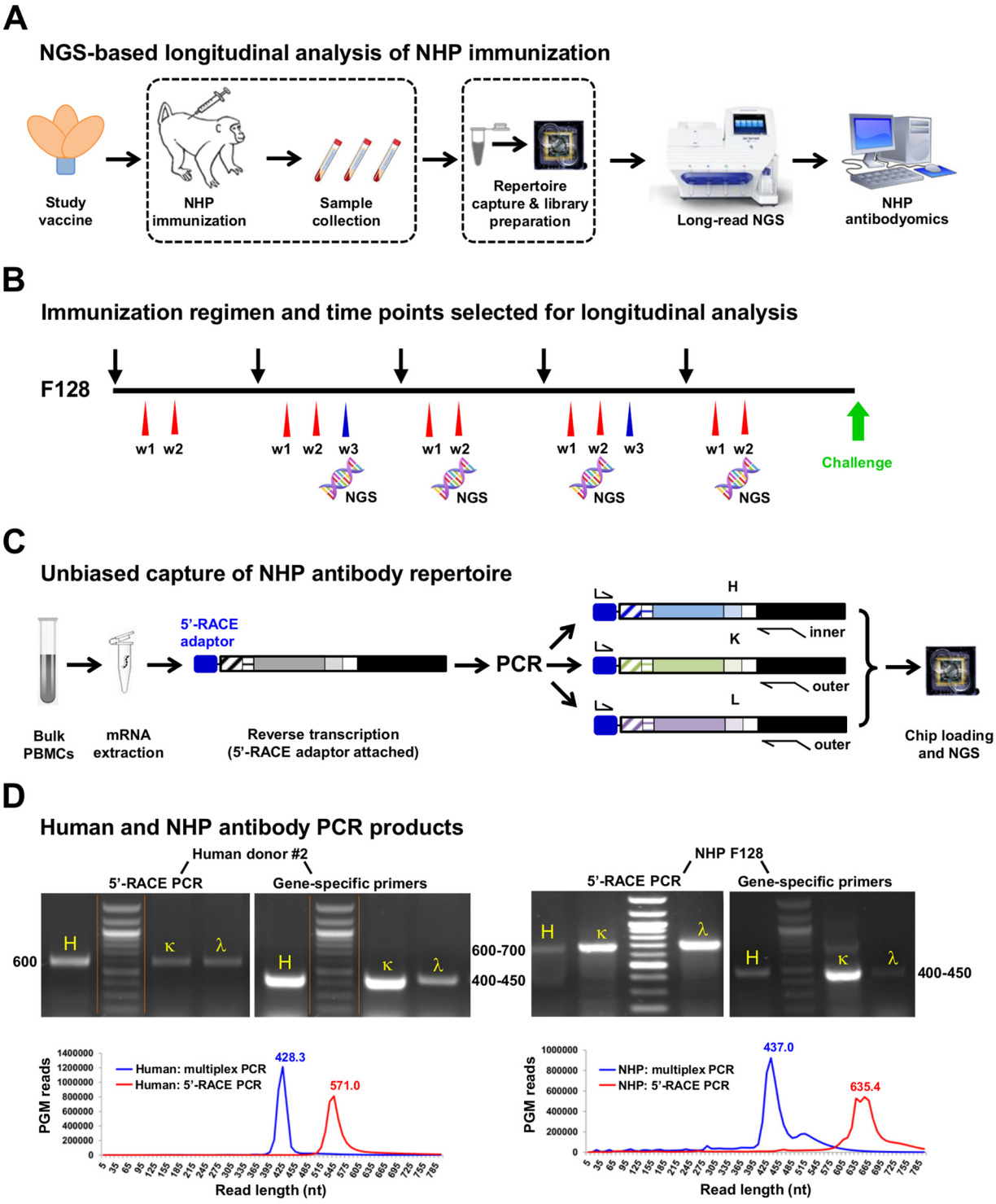

FIG 2 Experimental strategies for unbiased longitudinal B-cell repertoire analysis of NHP immunization. (A) Schematic flow of an NGS-based vaccine strategy consisting of vaccine immunogen design, NHP immunization, and sample collection, unbiased repertoire capture and library preparation, NGS profiling of B-cell response, and antibodyomics analysis. (B) F128 immunization regimen and time points selected for longitudinal analysis of memory B-cell response. Four time points—3 weeks (w3) after 2 immunizations and 2 weeks (w2) after 3, 4, and 5 immunizations—were selected for NGS profiling and are labeled with "NGS." (C) Schematic flow of unbiased capture of NHP antibody repertoire starting from peripheral blood mononuclear cells (PBMCs) to loaded sequencing chip. (D) Comparison of PCR products of human (left) and NHP (right) antibody transcripts obtained from gene-specific primers and 5′ RACE PCR. The PCR products are characterized with gel electrophoresis (top) and PGM sequencing (bottom). For the purpose of formatting, the gel of human antibody libraries has been rearranged with splicing (labeled with orange lines). The average read length from PGM sequencing, without the default 3′ trimming in the base-calling process, is provided on the read length distribution.

TABLE 1 PGM sequencing of 5'-RACE PCR products derived from the B-cell transcripts of macaque F128 over four time points of immunization[a]

Expt	PGM chip	Total no. of raw reads	Time point and no. of raw reads	Chain	No. of Ab chains	Avg read length (nt)	% usable sequences[a]
1	318 v2	6,284,962	Post-2; 4,510,364	H	1,124,796	613.9	80.0
				κ	1,143,809	622.3	84.8
				λ	686,230	634.5	86.6
			Post-4; 1,774,598	H	514,441	608.4	82.0
				κ	335,327	626.4	88.9
				λ	355,740	633.7	86.5
2	318 v2	6,868,200	Post-3; 3,698,440	H	1,478,758	619.5	85.1
				κ	842,188	628.2	88.4
				λ	893,693	651.8	89.7
			Post-5; 3,169,760	H	1,102,282	611.9	84.6
				κ	814,721	624.0	86.8
				λ	637,974	646.5	89.2
3	318 v2	5,521,542	Post-2; 3,091,616	H	720,229	611.5	80.3
				κ	737,689	621.4	84.8
				λ	449,051	634.9	85.0
			Post-4; 2,429,926	H	656,724	606.2	81.9
				κ	431,952	626.2	88.8
				λ	476,752	635.4	84.9
4	316 v2	3,649,734	Post-3; 1,771,957	H	707,489	621.1	84.6
				κ	415,917	631.0	86.2
				λ	415,793	649.9	90.5
			Post-5; 1,877,777	H	656,061	613.4	84.0
				κ	501,232	627.5	84.7
				λ	369,740	644.7	90.3

[a] Listed items include the index of PGM sequencing experiment, PGM sequencing chip, total number of raw reads, two time points from which the libraries were prepared and the number of raw reads for each time point, antibody chain type, number of antibody chains, average read length, and percentage of usable sequences after pipeline processing.

line gene frequencies likely reflect the general patterns of the NHP resting memory, class-switched, B-cell repertoire. We next determined the germline gene divergence and CDR3 length distributions. Consistent with the pattern of gene usage, the germline divergence showed limited changes over time (Fig. 3B). Overall, heavy chains yielded an average divergence of 9.1 to 9.7%, while κ and λ chains showed lower divergence values of 6.4 to 7.0% and 5.7 to 6.7%, respectively. These were lower than the divergence values observed for HIV-1-infected individuals but more closely resembled those of uninfected human donors (41). We also observed an oscillating pattern of divergence over the four time points, with more mature heavy chains (≥13%) appearing after the second and fourth inoculations and low-divergence light chains (<5%) increasing by up to 5% at the other two time points. The NHP CDR3 length distributions exhibited human-like patterns (Fig. 3C), with an average HCDR3 length of 12.5 aa compared to an average of 12.5 to 15.6 aa for human donors (Kabat definition) (41), supporting the notion that NHP can produce long HCDR3 loops similar to those displayed by human MAbs.

TABLE 2 Unbiased memory B-cell repertoire analysis using combined pipeline-processed PGM data sets from four time points of F128 immunization[a]

Time point	Chain	N_{seq}
Post-2	H	1,477,433
	κ	1,596,001
	λ	975,882
Post-3	H	1,856,896
	κ	1,102,935
	λ	1,177,747
Post-4	H	959,677
	κ	681,948
	λ	712,479
Post-5	H	1,483,858
	κ	1,131,668
	λ	903,272

[a] Listed items include the time point, antibody chain type, and total number of usable sequences after combining multiple data sets.

The unbiased analysis thus provided a global view of NHP memory B-cell repertoires during the course of vaccination. The population-based survey revealed a rather stable profile of general properties such as germline gene usage, degree of SHM, and CDR3 length with relatively small but notable changes. To better understand the vaccine-induced NHP B-cell responses, we traced the lineages of three CD4bs-directed neutralizing MAbs isolated from F128 (GE136, GE140, and GE143) in the unbiased antibody repertoires of four time points.

GE136-like transient memory B-cell response. One GE136-like heavy chain was identified in the 454 sequencing data obtained at the completion of this immunization regimen (Fig. 1B). Here we sought to determine the approximate "birth time" and lineage development of GE136 in the context of unbiased repertoires. Using GE136 as a template, we interrogated the F128 repertoire from each of the four time points (Fig. 4A). GE136-like heavy chains were not identified from the last time point, perhaps indicating a low representation of the GE136 lineage in the

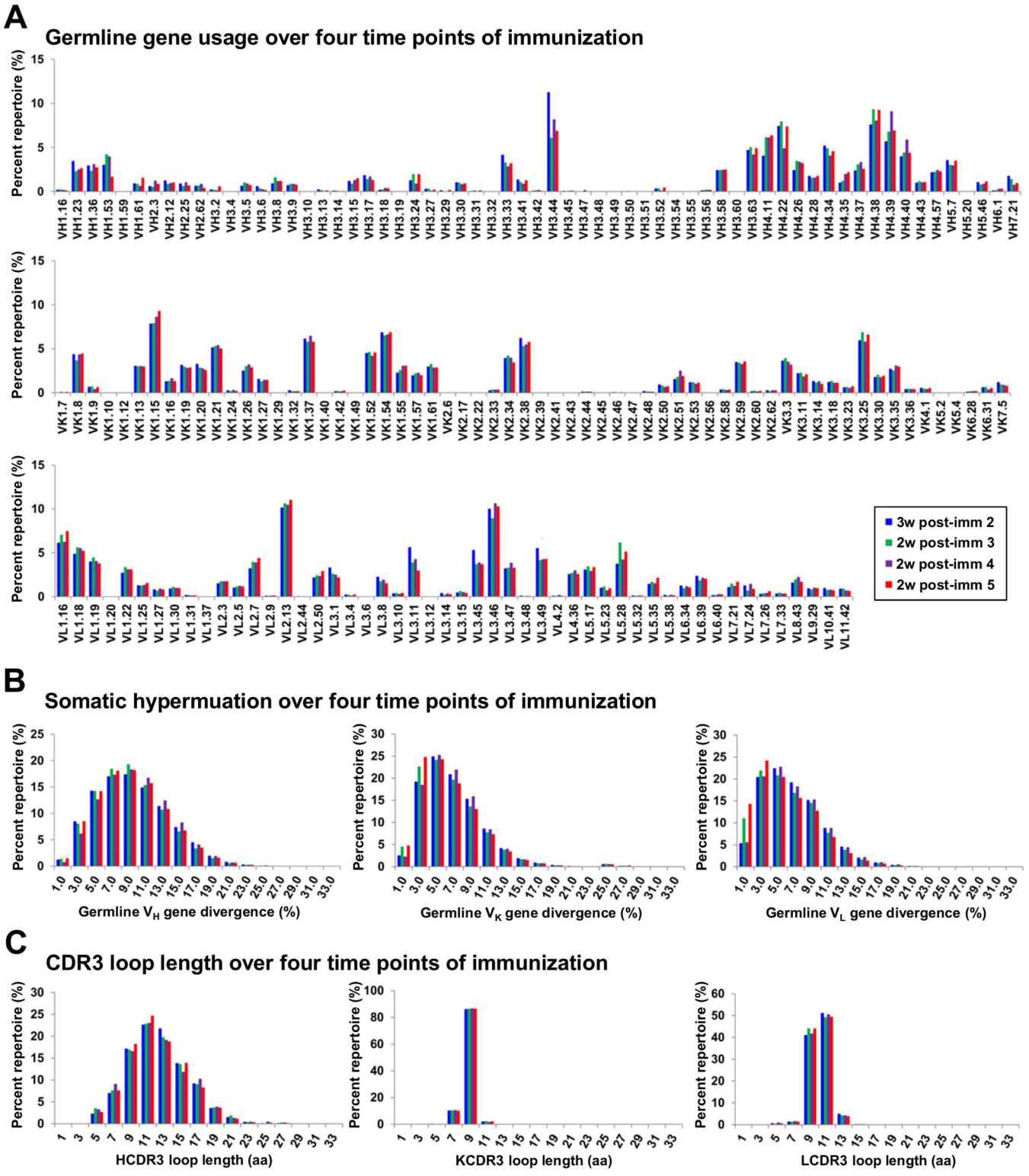

FIG 3 Unbiased memory B-cell repertoires at four time points of F128 immunization. For each time point, unbiased heavy (H) and light chain (κ and λ) libraries were obtained using a 5′ RACE PCR protocol. PGM sequencing was performed using Ion 318 (or 316) v2 chips, and the sequencing data were processed with the NHP antibodyomics pipeline (Tables 1 and 2). The processed antibody chain sequences were used to determine general repertoire properties such as germline gene usage (A), germline gene divergence (B), and complementarity determining region 3 (CDR3) length (C) for H, κ, and λ chains.

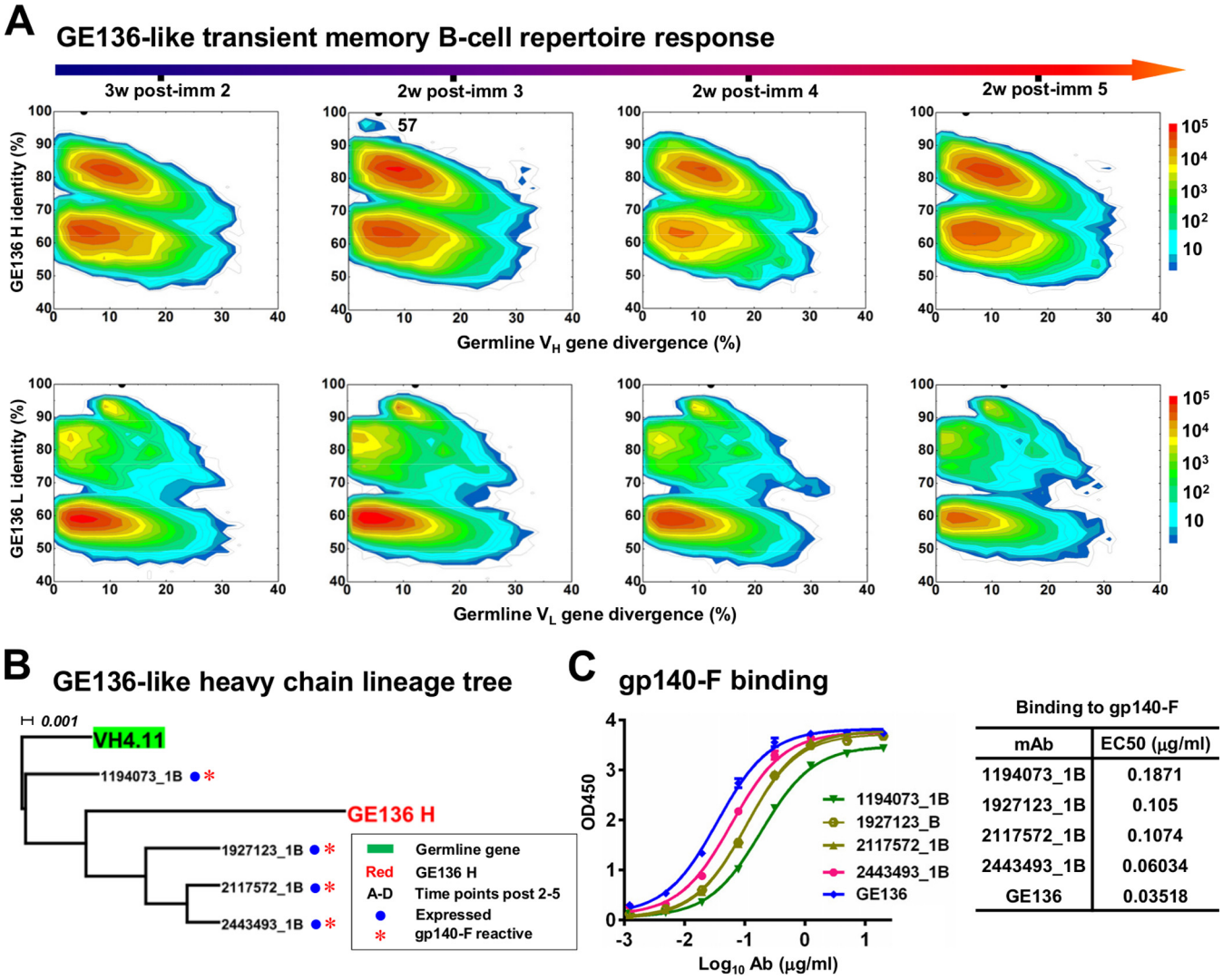

FIG 4 Longitudinal analysis of the GE136 lineage development. (A) Identity-divergence analysis of the unbiased heavy (H) and light (λ) chain repertoires. The sequences are plotted as a function of sequence identity to GE136 and sequence divergence from putative germline V genes. Color coding denotes sequence density. (B) Maximum-likelihood (ML) tree of selected GE136 heavy chain variants rooted by the putative germline V gene, VH4.11. The bar represents a 0.001 change per nucleotide site. A bioinformatics procedure consisting of HCDR3 comparison at an identity cutoff of 95%, motif filtering, and clustering at a full-length identity cutoff of 99% was used to select sequences for phylogenetic analysis and experimental validation. Four selected sequences are displayed in the ML tree. Reconstituted antibodies are labeled with blue dots if expressed and red asterisks if reactive with the HIV-1 trimer vaccine antigen. (C) (Left) Enzyme-linked immunosorbent assay (ELISA) analysis of antibody binding to the vaccine antigen, with standard deviation (SD) bars. (Right) Averaged EC_{50}s.

repertoire in general, consistent with the 454 sequencing data (34). From the repertoires of the first three time points, we could find only 57 GE136-like heavy chains within the post-3 repertoire. These results indicate that the clonal lineage of GE136 notably expanded following the third immunization but fell below the level of detection afterward, perhaps due to competing CD4bs-directed antibody lineages or lineages of other epitope specificities.

We next examined the biological function of NGS-derived GE136-like heavy chains. The bioinformatics selection (see Materials and Methods) resulted in seven sequences, four of which were subjected to phylogenetic analysis and functional validation (see Table S2A in the supplemental material). In the maximum-likelihood (ML) tree rooted by the germline gene VH4.11, the GE136 heavy chain gene was sandwiched by 1194073_1B and three other sequences (Fig. 4B). GE136-like heavy chains were paired with the GE136 light chain and tested for their binding to the gp140-F trimer. By enzyme-linked immunosorbent assays (ELISA), the reconstituted antibodies bound to the vaccine antigen with an ~10-fold range of 50% effective concentrations (EC_{50}s) (Fig. 4C). The ability of the reconstituted antibodies to neutralize HIV-1 was confirmed using a cross-clade 6-virus panel (Table 3). Similar to the parental GE136, three of four GE136 somatic variants neutralized three of six HIV-1 isolates.

GE140-like evolving memory B-cell response. A group of diverse somatic variants ($n = 75$) was identified for the GE140 heavy chain from the 454 sequencing of a sample after 5 immunizations (Fig. 1B). Given the transient presence of the GE136 lineage, we asked whether we could detect a similar pattern of lineage development for GE140. We first visualized the unbiased heavy and

TABLE 3 Neutralization titers of 29 chimeric antibodies derived from NHP F128 against 6 HIV-1 Env-pseudoviruses[a]

Antibody or variant	Time point	Neutralization IC_{50} (μg/ml)					
		JRFL.JB	HxB2.DG	MN.3	SF162.LS	DJ263.8	MW965.26
Antibodies							
GE136	Post-5	>50	4.59‖	1.36‖	28.20‡	>50	>50
GE140	Post-5	>50	1.11‖	>50	>50	>50	>50
GE143	Post-5	>50	1.74‖	1.44‖	12.70‡	21.70‡	>50
VRC01		0.08*	0.08*	0.07*	0.16¶	0.10¶	0.08*
GE136 heavy chain somatic variants							
2443493_1B/GE136L	Post-3	>50	>50	>50	35.30‡	>50	>50
1194073_1B/GE136L	Post-3	>50	5.46‖	1.31‖	19.30‡	>50	>50
2117572_1B/GE136L	Post-3	>50	4.73‖	0.92¶	40.20‡	>50	>50
1927123_1B/GE136L	Post-3	>50	5.68‖	1.56‖	38.40‡	>50	>50
GE140 heavy chain variants identified from longitudinal tracing and light chain variants from post-5							
1095891_2B/GE140L	Post-3	>50	0.24¶	>50	7.57‖	>50	>50
2356992_1B/GE140L	Post-3	>50	>50	>50	>50	>50	>50
1021989_2C/GE140L	Post-4	>50	1.19‖	>50	>50	>50	>50
1000200_1D/GE140L	Post-5	>50	2.10‖	>50	>50	>50	>50
665503_2D/GE140L	Post-5	>50	1.47‖	>50	>50	>50	>50
1195826_1D/GE140L	Post-5	>50	4.44‖	>50	>50	>50	>50
GE140H/1528713_2D	Post-5	>50	>50	>50	>50	>50	>50
GE140H/1657151_2D	Post-5	>50	5.15‖	>50	>50	>50	>50
GE140H/1530613_2D	Post-5	>50	>50	>50	>50	>50	>50
GE143-like heavy chains of VH4.34 and VH4.40 origins identified from longitudinal tracing							
2723042_1A/GE143L[b]	Post-2	>50	>50	>50	27.00‡	>50	>50
1772170_2A/GE143L[c]	Post-2	>50	>50	>50	27.80‡	>50	>50
2948381_1A/GE143L[b]	Post-2	>50	8.60‖	>50	13.50‡	>50	>50
3064746_1A/GE143L[b]	Post-2	>50	>50	>50	>50	>50	>50
1105550_2B/GE143L[b]	Post-3	>50	1.62‖	1.87‖	10.20‡	>50	>50
1016368_1B/GE143L[b]	Post-3	>50	4.61‖	>50	>50	>50	>50
1065596_1B/GE143L[c]	Post-3	>50	9.28‖	>50	22.50‡	>50	>50
1189841_1B/GE143L[b]	Post-3	>50	>50	>50	23.60‡	>50	>50
1331376_1B/GE143L[b]	Post-3	>50	1.88‖	2.47‖	1.710	>50	>50
2330877_1B/GE143L[c]	Post-3	>50	2.96‖	4.02‖	12.20‡	>50	>50
714249_2B/GE143L[b]	Post-3	>50	4.29‖	4.84‖	23.90‡	>50	>50
928105_2B/GE143L[c]	Post-3	>50	1.99‖	3.70‖	15.30‡	>50	>50
1527642_2C/GE143L[b]	Post-4	>50	>50	>50	25.60‡	>50	>50
552786_2C/GE143L[b]	Post-4	>50	>50	>50	29.20‡	>50	>50
1037809_1D/GE143L[b]	Post-5	>50	1.26‖	1.00‖	8.24‖	>50	>50
681893_2D/GE143L[b]	Post-5	>50	2.19‖	1.75‖	23.70‡	>50	>50

[a] The chimeric antibodies were expressed using the 29 heavy chains and 3 light chains derived from longitudinal repertoire analysis of NHP F128 with their native partner chains from the GE136, GE140, and GE143. The wild-type GE136, GE140, and GE143 were included as controls. A human broadly neutralizing antibody VRC01 was added for comparison. *, $IC_{50} < 0.01$ μg/ml; ¶, $IC_{50} = 0.01$ to ~1 μg/ml; ‖, $IC_{50} = 1$ to ~10 μg/ml; ‡, $IC_{50} = 10$ to ~50 μg/ml.

[b] GE143-like heavy chain assigned to the putative germline V_H gene VH4.34.

[c] GE143-like heavy chain assigned to the putative germline V_H gene VH4.40.

light chain repertoires with respect to GE140 using the 2D plots (Fig. 5A). Interestingly, GE140-like heavy chains were found from all time points except for the one after 2 immunizations. This finding suggests that the GE140 lineage formed within a time frame similar to that of GE136 but undertook a different developmental pathway, such that it remained detectable over the remaining course of the vaccination regimen (Fig. 5A, row 1). It should also be noted that two GE140-like MAbs were isolated from a different post-2 sample by single B-cell sorting (our unpublished data), suggesting an earlier birth time of the GE140 lineage. GE140-like heavy chains possessed a lower divergence after the third immunization yet continued to mature and form a separate lineage over the following 2 months, as indicated by the more divergent sequences along the 100% identity line and a wedge-shaped density departing from the main population (Fig. 5A, row 1). To quantify the GE140 lineage development, we identified all heavy chains of the VH4.11 origin and determined the fraction of heavy chains with an HCDR3 identity of 95% or more to GE140 within the VH4.11 family (Fig. 5A, row 2). The GE140 lineage evolved rapidly between post-3 and -4, with the fraction increasing from 0.83 to 6.07‰ before reaching a plateau of 5.59‰ between post-4 and -5. The GE140 lineage continued to diversify during maturation, as indicated by more sequences with higher divergence but lower GE140 identity (Fig. 5A, row 2). In contrast,

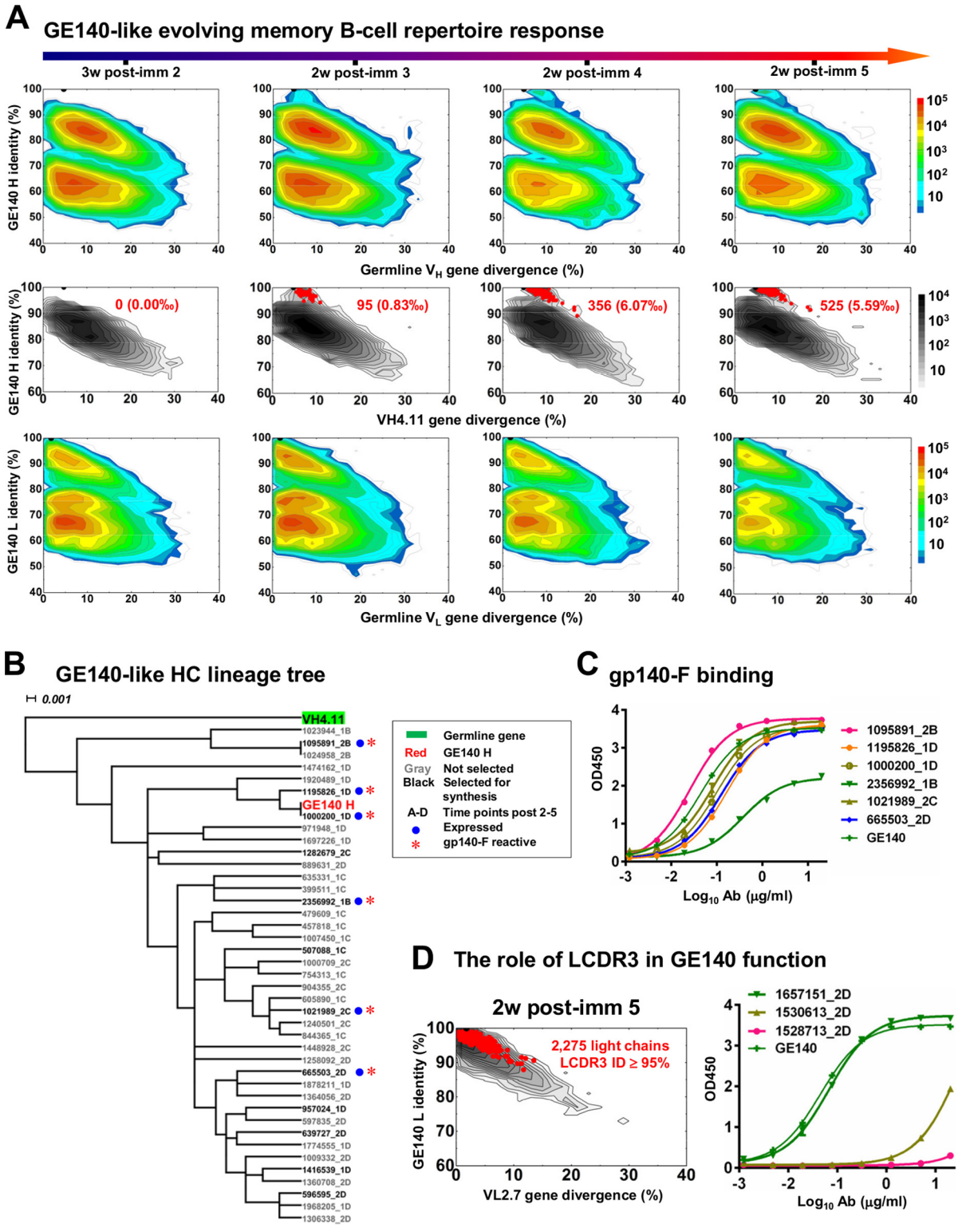

FIG 5 Longitudinal analysis of the GE140 lineage development. (A) Identity-divergence analysis of the unbiased heavy (H) and light (λ) chain repertoires. The sequences are plotted as a function of sequence identity to GE140 and of sequence divergence from putative germline V genes. Color coding denotes sequence density. The VH4.11 heavy chain family at each time point is visualized by black contour lines (middle row). The GE140-like heavy chains with an HCDR3 identity of over 95% are shown as red dots on the plots; the numbers are numbers of sequences. (B) ML tree of selected GE140 heavy chain variants rooted by the putative germline VH4.11 gene (see Fig. 4 for sequence selection and labeling). The bar represents a 0.001 change per nucleotide site. (C) ELISA analysis of antigen binding is shown for six antibodies reconstituted from GE140 heavy chain variants with SD bars. (D) Analysis of GE140-like light chains. (Left) The VL2.7 light chain family at the last time point is visualized with black contours. Light chains with an LCDR3 identity of 95% or greater are shown as red dots on the 2D plot. (Right) ELISA analysis of HIV-1 trimer vaccine antigen binding (right) is shown for three antibodies reconstituted from GE140 light chain variants, with SD bars.

the light chain repertoires contained sequences that were nearly 100% identical to the GE140 light chain, displaying little change over time (Fig. 5A, row 3).

We next investigated the phylogeny of GE140-like heavy chains. A bioinformatics selection (see Materials and Methods) yielded 40 heavy chains, which are displayed in an ML tree rooted by VH4.11 (Fig. 5B). Remarkably, the tree branching pattern was consistent with the chronological order of the maturation events: the post-3 sequences were found only from the top branches (marked as "B"), the post-4 sequences were found from the middle branches ("C"), and the distant branches were occupied solely by the post-5 sequences ("D"). Of note, some post-5 heavy chains were clustered with post-3 sequences on the top branch, suggestive of a recall of post-3 memory B cells after the last inoculation. The GE140 heavy chain itself was sandwiched by two post-5 sequences, consistent with the finding that a larger number of GE140 MAbs (28) were seen from the last time point (post-5) versus the two early time points (post-3 and -4). Overall, phylogenetic analysis revealed sustained GE140 lineage development over a 2-month period. We then selected 12 NGS-derived GE140 heavy chains for pairing with the GE140 light chain (see Table S2B in the supplemental material). Six of the reconstituted antibodies were expressed and showed gp140-F binding comparable to that of GE140 by ELISA (Fig. 5C), except for 2356992_1B. Of note, the four post-5 heavy chains on the most distant phylogenetic branch were not expressed, suggesting that they are likely remote variants of the GE140 heavy chain and might not be compatible with the GE140 light chain.

We identified 2,275 light chains with the VL2.7 germline origin and an LCDR3 identity of 95% or greater to the GE140 light chain from the last time point (Fig. 5D, left). This finding agrees with the low diversity expected of light chains due to the absence of the D gene segment. We then estimated the contribution of LCDR3 to GE140-antigen interactions by testing three light chains selected from this population: one with the same LCDR3 as GE140 and two possessing an insertion in LCDR3 (see Table S2B in the supplemental material). Interestingly, although all three light chain variants were expressed when paired with the GE140 heavy chain, only the one with an identical LCDR3 behaved like GE140 in ELISA binding (Fig. 5D, right), suggesting that LCDR3 plays a critical role in CD4bs recognition. Overall, a large number of light chains with the same germline gene composition and LCDR3 will facilitate the rapid production of a pool of antigen-binding B cell receptors (BCRs) capable of further antigen selection and maturation.

All expressed GE140 variants were tested on a 6-virus panel to assess their neutralizing ability (Table 3). As expected, the gp140-F-binding heavy chain variant (1095891_2B) and the light chain variant with identical LCDR3 (1657151_2D) could neutralize the same tier 1 isolates as GE140 itself when paired with respective GE140 partner chains. In summary, the clonal lineage of GE140 exhibited a pattern of sustained development in response to the CD4bs presented on an HIV-1 Env trimer, in contrast to the transient GE136 lineage of the same epitope specificity and germline V_H origin (but a different germline V_L origin, VL5.28). Recently, Tran et al. reported a less favorable angle of approach for GE136 (43). Therefore, the rapid expansion of the GE140 lineage in comparison to GE136 can likely be explained by GE140 having more favorable Env interactions, which would lead to greater fitness of the GE140 lineage in B-cell selection against GE136 and other antibody lineages.

GE143-like oscillating memory B-cell response. Among all the CD4bs-directed MAbs isolated from F128, GE143 showed the highest degree of SHM at 8.4% (32). Compared to the GE136 and GE140 lineages, which first became detectable after 3 immunizations, a larger set of GE143-like heavy chains can be found in the repertoires of all four time points studied (Fig. 6A, row 1), suggesting that the GE143 lineage originated from naïve B cells responding to the first two inoculations. In addition, the higher degree of SHM and larger somatic population observed for GE143 could be the result of a more extensive maturation process. Specifically, the distribution of GE143-like heavy chains on the 2D plots exhibited an oscillating pattern (Fig. 6A, row 1), suggesting—within the limits of sampling error—a unique lineage development pertaining to the regimen, vaccine antigen, and antibody-antigen interactions.

Further analysis revealed that GE143-like heavy chains comprised the sequences of two putative germline origins: VH4.34 and VH4.40 (Fig. 6A, rows 2 and 3). Although both families waxed and waned during the immunization regimen, VH4.34 dominated the oscillating pattern. VH4.34 is different from VH4.40—the putative GE143 germline gene—by 18 nucleotides (6%) and 8 amino acids (8%), suggesting that multiple sites in the sequence may contribute to the ambiguity in gene assignment (see Fig. S4 in the supplemental material). To investigate the lineage structure of two putative germline families, we plotted sequences with an HCDR3 identity of over 92.0% (Fig. 6A, rows 2 and 3, blue dots) and over 95.0% (red dots) relative to GE143. Such hierarchical HCDR3 analysis will likely reveal hidden lineage patterns that would otherwise be overlooked with a single cutoff. Within the VH4.34 family (Fig. 6A, row 2), the frequency of GE143-like sequences defined by a lower cutoff (92.0%) appeared to increase over time (from 0.95 to 2.88‰). In contrast, the more stringent cutoff (95%) identified GE143-like sequences only after 3 and 5 immunizations, with frequencies of 1.13 and 1.81‰, respectively. Within the VH4.40 family (Fig. 6A, row 3), the lower cutoff showed a response peaking after 3 immunizations (0.79‰), whereas the more stringent cutoff—similar to the case of the VH4.34 family—identified GE143-like sequences only after 3 and 5 immunizations, with a notable decrease at the later time point. Our analysis thus uncovered a memory B-cell repertoire evolving toward GE143 driven by the HCDR3-mediated interactions. The activated memory B cells, upon acquiring the GE143-like HCDR3s, as defined by the 95% cutoff, were no longer detectable in the repertoire, perhaps due to antigen adsorption. The light chain repertoire remained consistent over the four time points studied (Fig. 6A, row 4), suggesting a lesser role of light chains in GE143 antigen recognition and maturation.

We next combined phylogenetic analysis, antibody synthesis, ELISA binding, and HIV-1 neutralization to interrogate the function of GE143-like heavy chains of two putative germline origins. The ML tree of VH4.34-originated heavy chains showed three major branches, each containing sequences from multiple time points (Fig. 6B). The bioinformatics selection identified 15 heavy chains (see Materials and Methods; see also Table S2C in the supplemental material). Twelve of these heavy chains could be expressed when paired with the GE143 light chain and bind to the vaccine antigen with ELISA curves similar to those of both GE143 and a human bNAb (VRC01), with the exception of 3064746_1A

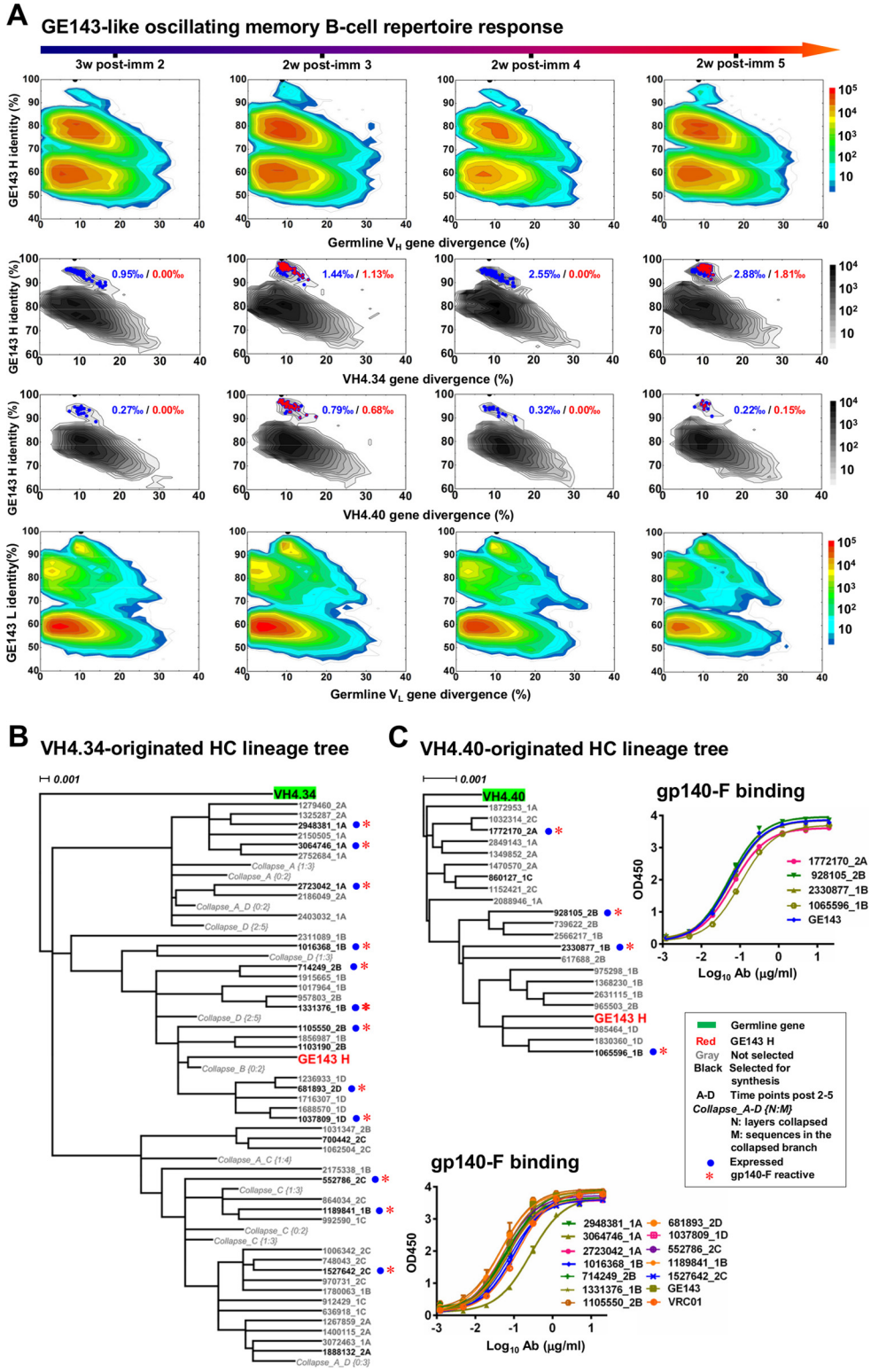

FIG 6 Longitudinal analysis of the GE143 lineage development. (A) Identity-divergence analysis of the unbiased heavy (H) and light (λ) chain repertoires. The sequences are plotted as a function of sequence identity to GE143 and sequence divergence from putative germline V genes. Color coding denotes sequence density. The VH4.34 and VH4.40 families are visualized by black contour lines (second and third rows). Heavy chains with an HCDR3 identity of over 92% and 95% to GE143 are shown as blue and red dots, respectively. (B) ML tree of selected GE143-like heavy chains of VH4.34 origin (left). ELISA analysis of HIV-1 vaccine antigen binding is shown for 12 antibodies reconstituted from GE143-like heavy chains of VH4.34 origin with SD bars (right). (C) (Left) ML tree of selected GE143-like heavy chains of VH4.40 origin. (Right) ELISA analysis of HIV-1 vaccine antigen binding for four antibodies reconstituted from GE143-like heavy chains of VH4.40 origin, with SD bars. The procedure used for sequence selection and labeling of ML tree is described in the legend to Fig 4. The bar represents a 0.001 change per nucleotide site.

(Fig. 6B, bottom right). The antibody reconstituted from 681893_2D, which has the shortest genetic distance to the GE143 heavy chain on the ML tree, bound to gp140-F and neutralized three isolates on a 6-virus panel (Table 3). Our results thus confirmed that these heavy chains, although assigned to a different germline V_H gene, are functionally GE143-like. The ML tree of VH4.40-originated heavy chains was consistent with the oscillating pattern observed from the 2D plots, with sequences from after 2 and 4 immunizations grouped closer to the germline gene and the sequences from after 3 and 5 immunizations occupying a more distant branch (Fig. 6C). Five heavy chains were selected from the phylogenetic tree and paired with the GE143 light chain for experimental validation (see Table S2C). Four reconstituted antibodies were expressed that showed binding to gp140-F (Fig. 6C, top right). When assessed for neutralization on a cross-clade 6-virus panel, they could neutralize up to 3 isolates (Table 3).

In summary, longitudinal tracing of the GE143 lineage revealed an unexpected complexity of F128 memory B-cell responses, with two related sub-lineages likely originating from the same germline gene before diverging into two distinct groups during maturation. Such intralineage diversity is not uncommon for HIV-1 bNAbs, as previously reported for the VRC01-class bNAbs targeting the same site (4, 10, 19, 23), which can develop different HCDR3 loops and up to 30% difference in sequence during maturation. However, this is the first time that such B-cell lineage diversification, albeit at a lower level than that of VRC01, was observed in NHP immunization. It is possible that the oscillating pattern may be related to the short interval in vaccination, although a more detailed comparative study is warranted to confirm this possibility. Therefore, our results suggest that unbiased longitudinal repertoire analysis may assist in the optimization of injection interval, adjuvant, and other critical variables in a vaccine regimen.

Validation of the NHP B-cell lineage patterns. We have performed a similar lineage tracing analysis for other F128 MAbs. For GE121, an HCDR3 identity cutoff of 95% or greater did not yield any somatic variants in the unbiased repertoires of four time points except for the post-3 repertoire. Six GE121-like heavy chains were found after 3 immunizations, accounting for 0.2‰ of the VH3.8 family, similar to the GE136 lineage (see Fig. S5A in the supplemental material). No GE137 somatic variants were identified in the unbiased repertoire at any time point, indicating a low frequency of this MAb lineage (see Fig. S5B in the supplemental material). For GE147, we observed a rapid lineage expansion between the post-3 and post-4 time points, with the frequency increasing from 4.51 to 53.87‰ in the VH3.24 family (see Fig. S6A in the supplemental material). The GE147 lineage showed a notable decline after 5 immunizations, accounting for only 5.57‰ of the VH3.24 family. Overall, the lineage pattern of GE147 is reminiscent of GE140, even though these two MAbs evolved from different germline origins. For the GE148 lineage, an HCDR3 identity cutoff of 92% revealed an oscillating pattern resembling that of the GE143 lineage despite the different germline origins (see Fig. S6B). Taken together, the lineage patterns observed for MAbs GE136, GE140, and GE143 represent the general patterns of CD4bs-directed memory B-cell responses during F128 immunization.

We next evaluated how experimental factors such as library variation, sequencing variation, and sequencing depth influence the lineage patterns by sequencing different aliquots of the heavy chain cDNA libraries from four time points (see Table S3 in the supplemental material). We first examined the correlation of repertoire properties derived from two sets of PGM experiments, since the unbiased repertoires form the basis of lineage tracing (see Fig. S7A in the supplemental material). A correlation coefficient (R^2) of 0.94 or greater was observed for germline gene usage, germline divergence, and HCDR3 loop length. However, we did notice that the average germline divergence (8.3 to 8.7%) is ~1% lower than that obtained previously, likely due to different library pooling strategies and upgraded PGM technologies. We then traced the GE136, GE140, and GE143 lineages in the resequenced heavy chain repertoires (see Fig. S7B to D in the supplemental material). The sequencing depth, with a reduction of 13 to 45%, showed a clear effect on the lineage analysis. The distribution of somatic variants appeared to be sparser on the 2D plots, with no somatic relatives identified for GE136 from the post-3 repertoire, consistent with a 42% reduction in the repertoire size for this time point. For GE140 and GE143, although the lineage patterns remained consistent between the two sets of PGM experiments, the frequencies became less reliable due to the reduced data size. In summary, additional NGS experiments not only confirmed the B-cell lineage patterns but also demonstrated the importance of experimental design (e.g., library pooling and sequencing depth) in antibody lineage analysis.

DISCUSSION

The impact of NGS-based immune repertoire analysis has been highlighted in the context of vaccine development (44–46). It has been envisioned that this technology will soon expand from bNAb characterization to other key aspects of vaccine research, such as longitudinal analysis of B-cell responses in NHP immunization (Fig. 2). The antibody repertoire profiles derived from NGS may facilitate quantitative assessment of vaccine-induced B-cell responses and provide a predictive indicator of vaccine efficacy even at the preclinical stage (47). As an example, here we present a finely detailed antibody NGS analysis of F128, an HIV-1 Env-immunized rhesus macaque, the subject of recent B-cell analyses (30, 32, 34).

Bioinformatics plays a critical role in the interpretation of antibody NGS data (47); however, little has been done to develop bioinformatics tools for NHP repertoire analysis in a systematic manner. In this study, we incorporated the most current NHP germline gene database (32) into the framework of a well-established human antibodyomics pipeline (4, 10, 19–21, 41) to create a set of consistent tools for NHP antibody repertoire analysis. Since the antibodyomics methods are independent of the animal species studied, the NHP pipeline should produce results with quality equivalent to that of the human pipeline (4, 10, 19–21, 41), allowing a comparison of NHP and human B-cell repertoires. Of particular note, the error correction method used in the human pipeline that was originally developed for the 454 platform (21) has been validated for the PGM platform (41, 48). Therefore, after benchmarking against the available 454 sequencing data from macaque F128, the NHP antibodyomics pipeline was used to analyze the longitudinal repertoire data obtained from PGM sequencing of F128 post-immunization samples.

Previously, we reported a significant bias associated with the use of gene-specific primers and demonstrated the utility of 5′ RACE PCR in the unbiased human B-cell repertoire analysis (41). The sequencing of 5′ RACE PCR-derived human antibody library

was challenging, as the read length approached 600 bp. It is not surprising that we encountered a similar challenge with the 5′ RACE PCR-derived NHP antibody libraries: the further downstream reverse primers, the addition of a library-specific barcode, and the use of a full-length P1 primer led to a library length of 600 to 700 bp. This problem was solved by using a single-end PGM platform consistent with the one-directional nature of 5′ RACE PCR and the long-read capability developed based on this platform (Fig. 2). Other NGS platforms, however, might not be suited for this purpose due to either limited throughput (Roche 454 and PacBio RS II) or short read length (Illumina MiSeq) (35, 42). In addition to the long-read capability, the combined use of 5′ RACE PCR and reverse primers optimized for single B-cell isolation provided a self-consistent protocol for repertoire analysis and lineage tracing of sorted MAbs.

Antigen-specific single B-cell sorting and antibody cloning have provided a panel of HIV-1-neutralizing MAbs that can be used as references to query the F128 antibody repertoires (32). The NHP antibodyomics pipeline in combination with the improved NGS technologies further enabled an in-depth analysis of F128 memory B-cell repertoires. The longitudinal tracing of seven MAbs within the context of unbiased repertoires at four time points revealed distinct patterns of lineage development, uncovering a dynamic interplay between the NHP B-cell repertoire and the CD4bs present on an HIV-1 trimer vaccine antigen. While the rapidly evolving GE140 lineage gives hope for eliciting effective neutralizing antibodies toward this conserved site, the oscillating GE143 lineage with two sub-lineages demonstrates the complexity of B-cell responses that may be expected from forthcoming HIV-1 trimer vaccines. Furthermore, the comparison of pre- and post-immunization repertoires will lend insight into which germline genes have been activated by the vaccination. However, due to limited sample availability, the pre-immunization and post-1 samples were not included in this study. The use of the pre-immunization sample as a baseline to dissect temporal B-cell response is critical and was outlined in a recent review (47).

While bioinformatics analysis provided crucial information on the NHP B-cell response to an HIV-1 trimer vaccine antigen, functional characterization of NGS-derived somatic variants and additional PGM experiments were performed to validate the biological findings. In our study, the unusually long reads (600 to 700 bp) led to lower quality toward the 5′ end of the sequences. The less commonly used motifs in framework 1 (FR1) of the variable regions (see Table S2 in the supplemental material) were suggestive of sequencing errors, although some might be biological. In principle, such sequence ambiguity can be eliminated if a single-molecule barcode (41) can be incorporated into the 5′ RACE PCR-based template preparation. Chimeric antibodies reconstituted from the NGS-derived sequences and the wild-type partner chains were assessed for antigen binding and HIV-1 neutralization, confirming the function of these putative somatic variants. However, some of the chimeric antibodies showed no detectable expression, suggesting either uncorrected sequencing errors or mismatched heavy and light chains. Unfortunately, the paired repertoire sequencing technology (49), although promising, has not been widely used due to its technical complexity. Nonetheless, phylogenetic analysis in conjunction with functional validation of putative somatic variants offered valuable insight into the MAb lineage structure, as previously demonstrated for HIV-1 bNAbs (4, 10, 19–21, 41). In this study, we also provided an example of how to validate B-cell repertoire profiles and lineage patterns by varying experimental parameters (see Fig. S7 in the supplemental material). Our results highlight the need for cautious interpretation of lineage patterns, which are usually more sensitive to experimental factors such as sequencing depth than repertoire profiles.

In conclusion, the B-cell repertoire profiles and lineage patterns observed for F128 provide useful insights for the evaluation of HIV-1 trimer vaccines in an NHP model. We expect that the NHP antibodyomics methods and antibody NGS technologies established in this study will find many applications in the preclinical assessment of human vaccine candidates.

MATERIALS AND METHODS

NHP specimens. Peripheral blood mononuclear cells (PBMCs) were obtained from F128, a rhesus macaque that has been the subject of intensive study of NHP B-cell response to an HIV-1 trimer vaccine (30, 32, 34, 43). The animal was housed at the animal facility of the Astrid Fagraeus Laboratory at the Swedish Institute for Infectious Disease Control. Housing and care procedures were in compliance with the guidelines of the Swedish Board of Agriculture. The facility has been assigned an Animal Welfare Assurance number by the Office of Laboratory Animal Welfare at the National Institutes of Health. All procedures were approved by the Local Ethical Committee on Animal Experiments. A detailed description of the immunization experiment was previously reported (30). Essentially, the regimen consists of purified gp140-F trimers in adjuvant inoculated five times at monthly intervals.

Sample preparation for PGM sequencing. The 5′ RACE PCR protocol developed for library preparation from human PBMCs was modified to prepare NHP antibody libraries (41). Briefly, total RNA was extracted from 10 to 20 million PBMCs into 30 μl of water with an RNeasy minikit (Qiagen). For unbiased antibody repertoire analysis, 5′ RACE was performed with a SMARTer RACE cDNA amplification kit (Clontech). The immunoglobulin (Ig) PCRs were set up with Platinum *Taq* high-fidelity DNA polymerase (Life Technologies) in a total volume of 50 μl, with 5 μl of cDNA as the template, 1 μl of 5′ RACE primer, and 1 μl of 10 μM reverse primer. The 5′ RACE primer contained a PGM P1 adaptor, while the reverse primer contained both a PGM A adaptor and a time point-specific barcode (see Table S1 in the supplemental material). Two Ion Xpress barcodes (Life Technologies), 1 and 2, were used to tag post-2 and -4 samples (or post-3 and -5 samples), respectively. Twenty-five cycles of PCRs were performed, and the expected PCR products (600 to 800 bp) were gel purified (Qiagen).

Ion Torrent PGM sequencing of NHP antibody libraries. The PGM sequencing protocol used for NHP antibody libraries was similar to that for human antibody libraries (41). Briefly, the antibody heavy (H) and light (κ and λ) chain libraries were quantitated using a Qubit 2.0 fluorometer with a Qubit double-stranded DNA (dsDNA) HS (high sensitivity) assay kit and mixed at a ratio of 1:1:1, except for the validation runs, in which only heavy chains were sequenced to confirm the lineage patterns (see Table S3 in the supplemental material). Next, post-2 and -4 (or post-3 and -5) libraries with different barcodes were pooled at a 1:1 ratio. The pooled library was diluted to a concentration of 50 pM prior to the template preparation, which was performed with the isothermal amplification (IA) kit obtained from the Early Access Program. The sequencing experiment was performed on the Ion Torrent PGM sequencer with the PGM Hi-Q 400 kit using an Ion 316 or 318 v2 chip for a total of 1,200 nucleotide flows. Note that an Ion 314 chip was used to sequence the post-4 heavy chain library in the validation run 3 (see Table S3 in the supplemental material). Raw data were processed without 3′-end trimming in the base-calling process to extend the read length (41). An antibody-specific bioinformatics pipeline (see below) was used to filter, correct, and annotate the full-length reads, with a number of metrics measured to ensure the sequence quality (see Fig. S1 and S2 in the supplemental material).

Bioinformatics analysis of antibody NGS data. An NHP antibodyomics pipeline has been developed based on the framework of a human antibodyomics pipeline (4, 10, 19–21, 41). The NHP antibodyomics pipeline consists of five consecutive steps. Given a data set of NGS-derived NHP antibody sequences, each sequence was (i) reformatted and labeled with a unique index number; (ii) assigned to the V, D (for heavy chain only), or J gene family using the current rhesus macaque germline gene database and an in-house implementation of IgBLAST, with sequences possessing an *E* value of $>10^{-3}$ for V gene assignment removed from the data set; (iii) subjected to a template-based error correction procedure, in which insertion-and-deletion (indel) errors of fewer than three nucleotides in the V gene segment were detected by alignment to their respective germline gene sequences and corrected; (iv) compared to the template antibody sequences at both the nucleotide level and the amino acid level using a global alignment module in CLUSTALW2 (50); (v) subjected to a multiple-sequence-alignment (MSA)-based procedure to determine complementarity-determining region 3 (CDR3), which was further compared to a set of template CDR3 sequences at the nucleotide level, and to determine the sequence boundary of the V(D)J coding region. For this purpose, the MSAs of representative germline V genes (truncated after the CXR motif) and J genes (truncated before the F/WGXG motif) have been precalculated and calibrated. Using heavy chain analysis as an example, a PGM-derived read would be included in the sequence alignment with 63 truncated V_H genes and separately with 7 truncated J_H genes to determine HCDR3, which should lie between the last column of aligned V_H genes and the first column of aligned J_H regions, and to determine the variable region, which should lie between the first column of aligned V_H genes and the last column of aligned J_H genes. This implicit pattern-matching method was specifically designed to take into account the sequence variation in the CDR3 flanking regions and the termini of variable regions (20). At the end of pipeline processing, a bioinformatics filter was applied to detect and remove erroneous sequences that may contain swapped gene segments due to PCR errors. Specifically, a full-length read would be removed from the data set if the V-gene alignment was less than 250 bp (41). The resulting antibody chain sequences were then subjected to a more in-depth repertoire analysis.

The NHP antibodyomics pipeline is a suite of Perl scripts that can be run on a Linux system with the option for parallel computing to reduce the data processing time. It requires pre-installed BLAST (51) and CLUSTALW2 (50) modules to facilitate germline gene assignment and global sequence alignment, respectively. Pre-compiled NHP germline gene databases are required for heavy and light chain-specific pipeline processing. The software suite can be obtained by request from J. Zhu, with source codes provided for main and auxiliary functions such as sequence quality assessment (see Fig. S1 and S2 in the supplemental material) and repertoire profiling (Fig. 3).

A bioinformatics procedure was devised to select heavy chain somatic variants for phylogenetic analysis, antibody synthesis, and characterization. From the NGS data obtained for each time point, we identified heavy chains with an HCDR3 identity of 95% or greater to a given MAb and subjected them to a filter that selects only for heavy chains possessing two conserved cysteines in the framework and the CXR-WGXG motif that flanks the HCDR3 region. Due to the complexity of GE143 lineage, a lower HCDR3 identity of 92% was used to include remote somatic variants. After filtering, a clustering procedure with an identity cutoff of 99.0% was used to further reduce sequence redundancy. The representative sequences of resulting clusters from four time points, if available, were combined into a single data set for further manual inspection.

The putative germline gene sequence was added to each data set, and a multiple sequence alignment (MSA) generated by CLUSTALW2 (50) was used as the input to construct a phylogenetic tree using DNAMLK (http://evolution.genetics.washington.edu/phylip/doc/dnamlk.html) in the PHYLIP package v3.69 (http://evolution.genetics.washington.edu/phylip.html). The displayed phylogenetic trees for the three MAb lineages (Fig. 4 to 6) were generated using Dendroscope (52), ordered to ladderize right, and rooted by their respective germline genes.

Antibody expression. Antibody production was performed as previously described (43). Briefly, the antibody sequences selected by bioinformatics were synthesized (Invitrogen, Inc.) and cloned into the cytomegalovirus R (CMV/R) expression vector containing the constant regions of NHP IgG. The heavy and light chains identified from PGM sequencing of F128 repertoires at multiple time points were paired with their respective partner chain DNAs. Full-length IgGs were expressed by transient transfection of 293F cells and purified using a recombinant protein A column (Pierce). The expression and other information for PGM-derived antibodies are summarized in Table S2 in the supplemental material.

HIV-1 neutralization assays. Neutralization assays were performed as previously described (32). Briefly, neutralization was measured using HIV-1 Env pseudoviruses and TZM-bl reporter cells (53–55). A six-virus panel was used in the current study. Neutralization curves were fit by a nonlinear regression analysis using a 5-parameter Hill slope equation (55). The 50% inhibitory concentration (IC_{50}) is defined as the antibody concentration required to inhibit HIV-1 infection by 50%.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.01375-15/-/DCSupplemental.

Figure S1, TIF file, 1.8 MB.
Figure S2, TIF file, 1.3 MB.
Figure S3, TIF file, 0.4 MB.
Figure S4, TIF file, 1.2 MB.
Figure S5, TIF file, 2.5 MB.
Figure S6, TIF file, 3 MB.
Figure S7, TIF file, 2.9 MB.
Table S1, DOCX file, 0.04 MB.
Table S2, DOCX file, 0.05 MB.
Table S3, DOCX file, 0.05 MB.

ACKNOWLEDGMENTS

This work was supported by grants from the Scripps Center for HIV/AIDS Immunology & Immunogen Discovery (CHAVI-ID UM1 AI 100663), the International AIDS Vaccine Initiative (IAVI) and its generous donors (who can be found at http://www.iavi.org), and NIH/NIAID grants HIVRAD P01 AI104722 and R01AI102766.

R. T. Wyatt and J. Zhu conceived the project. K. Dai, L. He, and J. Zhu performed the research, with the NHP primers designed by K. Dai, Y. Li, and C. Sundling, sequencing protocol devised by L. He, the cDNA, PCR, and library preparation and PGM sequencing performed by L. He, the NHP antibodyomics pipeline implemented by J. Zhu, bioinformatics analysis performed by J. Zhu, reconstitution of synthesized NHP antibodies and purification carried out by K. Tran and S. N. Khan, HIV-1 neutralization performed and analyzed by S. O'Dell, K. McKee, and J. R. Mascola; G. B. K. Hedestam contributed F128 material and provided valuable suggestions. J. Zhu and R. Wyatt wrote the paper, with all principal investigators providing comments or revisions.

REFERENCES

1. **Walker LM, Burton DR**. 2010. Rational antibody-based HIV-1 vaccine design: current approaches and future directions. Curr Opin Immunol **22:**358–366. http://dx.doi.org/10.1016/j.coi.2010.02.012.
2. **Haynes BF, Kelsoe G, Harrison SC, Kepler TB**. 2012. B-cell-lineage immunogen design in vaccine development with HIV-1 as a case study. Nat Biotechnol **30:**423–433. http://dx.doi.org/10.1038/nbt.2197.
3. **Burton D, Ahmed R, Barouch D, Butera S, Crotty S, Godzik A, Kaufmann D, McElrath M, Nussenzweig M, Pulendran B, Scanlan C, Schief W, Silvestri G, Streeck H, Walker B, Walker L, Ward A, Wilson I, Wyatt R**. 2012. A blueprint for HIV vaccine discovery. Cell Host Microbe **12:**396–407. http://dx.doi.org/10.1016/j.chom.2012.09.008.
4. **Wu X, Zhou T, Zhu J, Zhang B, Georgiev I, Wang C, Chen X, Longo NS, Louder M, McKee K, O'Dell S, Perfetto S, Schmidt SD, Shi W, Wu L, Yang Y, Yang Z, Yang Z, Zhang Z, Bonsignori M, Crump JA, Kapiga**

SH, Sam NE, Haynes BF, Simek M, Burton DR, Koff WC, Doria-Rose NA, Connors M, Mullikin JC, Nabel GJ, Roederer M, Shapiro L, Kwong PD, Mascola JR, Progra NCS. 2011. Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. Science **333:** 1593–1602. http://dx.doi.org/10.1126/science.1207532.

5. **Walker LM, Huber M, Doores KJ, Falkowska E, Pejchal R, Julien J-P, Wang S-K, Ramos A, Chan-Hui P-Y, Moyle M, Mitcham JL, Hammond PW, Olsen OA, Pham P, Fling S, Wong C-H, Phogat S, Wrin T, Simek MD, Koff WC, Wilson IA, Burton DR, Poignard P, Protocol GPI.** 2011. Broad neutralization coverage of HIV by multiple highly potent antibodies. Nature **477:**466–470.
6. **Walker LM, Phogat SK, Chan-Hui P-Y, Wagner D, Phung P, Goss JL, Wrin T, Simek MD, Fling S, Mitcham JL, Lehrman JK, Priddy FH, Olsen OA, Frey SM, Hammond PW, Kaminsky S, Zamb T, Moyle M, Koff WC, Poignard P, Burton DR, Protocol GPI.** 2009. Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. Science **326:**285–289.
7. **Wu X, Yang Z-Y, Li Y, Hogerkorp C-M, Schief WR, Seaman MS, Zhou T, Schmidt SD, Wu L, Xu L, Longo NS, McKee K, O'Dell S, Louder MK, Wycuff DL, Feng Y, Nason M, Doria-Rose N, Connors M, Kwong PD, Roederer M, Wyatt RT, Nabel GJ, Mascola JR.** 2010. Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. Science **329:**856–861. http://dx.doi.org/10.1126/science.1187659.
8. **Huang J, Ofek G, Laub L, Louder MK, Doria-Rose NA, Longo NS, Imamichi H, Bailer RT, Chakrabarti B, Sharma SK, Alam SM, Wang T, Yang Y, Zhang B, Migueles SA, Wyatt R, Haynes BF, Kwong PD, Mascola JR, Connors M.** 2012. Broad and potent neutralization of HIV-1 by a gp41-specific human antibody. Nature **491:**406–412. http://dx.doi.org/10.1038/nature11544.
9. **Falkowska E, Le K, Ramos A, Doores K, Lee J, Blattner C, Ramirez A, Derking R, van Gils M, Liang C, McBride R, von Bredow B, Shivatare S, Wu C, Chan-Hui P, Liu Y, Feizi T, Zwick M, Koff W, Seaman M, Swiderek K, Moore JP, Evans D, Paulson JC, Wong C-H, Ward AB, Wilson IA, Sanders RW, Poignard P, Burton DR.** 2014. Broadly neutralizing HIV antibodies define a glycan-dependent epitope on the prefusion conformation of gp41 on cleaved envelope trimers. Immunity **40:** 657–668. http://dx.doi.org/10.1016/j.immuni.2014.04.009.
10. **Zhou T, Zhu J, Wu X, Moquin S, Zhang B, Acharya P, Georgiev I, Altae-Tran H, Chuang G, Joyce M, Do Kwon Y, Longo N, Louder M, Luongo T, McKee K, Schramm C, Skinner J, Yang Y, Yang Z, Zhang Z, Zheng A, Bonsignori M, Haynes BF, Scheid JF, Nussenzweig MC, Simek M, Burton DR, Koff WC, Mullikin JC, Connors M, Shapiro L, Nabel GJ, Mascola JR, Kwong PD, Program NCS.** 2013. Multidonor analysis reveals structural elements, genetic determinants, and maturation pathway for HIV-1 neutralization by VRC01-class antibodies. Immunity **39:**245–258. http://dx.doi.org/10.1016/j.immuni.2013.04.012.
11. **Doria-Rose NA, Schramm CA, Gorman J, Moore PL, Bhiman JN, DeKosky BJ, Ernandes MJ, Georgiev IS, Kim HJ, Pancera M, Staupe RP, Altae-Tran HR, Bailer RT, Crooks ET, Cupo A, Druz A, Garrett NJ, Hoi KH, Kong R, Louder MK, Longo NS, McKee K, Nonyane M, O'Dell S, Roark RS, Rudicell RS, Schmidt SD, Sheward DJ, Soto C, Wibmer CK, Yang Y, Zhang Z, Sequencing NC, Mullikin JC, Binley JM, Sanders RW, Wilson IA, Moore JP, Ward AB, Georgiou G, Williamson C, Abdool Karim SS, Morris L, Kwong PD, Shapiro L, Mascola JR.** 2014. Developmental pathway for potent V1V2-directed HIV-neutralizing antibodies. Nature **509:**55–62. http://dx.doi.org/10.1038/nature13036.
12. **Huang J, Kang BH, Pancera M, Lee JH, Tong T, Feng Y, Imamichi H, Georgiev IS, Chuang G, Druz A, Doria-Rose NA, Laub L, Sliepen K, van Gils MJ, de la Peña AT, Derking R, Klasse P, Migueles SA, Bailer RT, Alam M, Pugach P, Haynes BF, Wyatt RT, Sanders RW, Binley JM, Ward AB, Mascola JR, Kwong PD, Connors M.** 2014. Broad and potent HIV-1 neutralization by a human antibody that binds the gp41-gp120 interface. Nature **515:**138–142. http://dx.doi.org/10.1038/nature13601.
13. **Scheid JF, Mouquet H, Ueberheide B, Diskin R, Klein F, Oliveira TYK, Pietzsch J, Fenyo D, Abadir A, Velinzon K, Hurley A, Myung S, Boulad F, Poignard P, Burton DR, Pereyra F, Ho DD, Walker BD, Seaman MS, Bjorkman PJ, Chait BT, Nussenzweig MC.** 2011. Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. Science **333:**1633–1637. http://dx.doi.org/10.1126/science.1207227.
14. **Liao H, Lynch R, Zhou T, Gao F, Alam SM, Boyd SD, Fire AZ, Roskin KM, Schramm CA, Zhang Z, Zhu J, Shapiro L, Mullikin J, Gnanakaran S, Hraber P, Wiehe K, Kelsoe G, Yang G, Xia S-M, Montefiori DC, Parks R, Lloyd KE, Scearce RM, Soderberg KA, Cohen M, Kamanga G, Louder MK, Tran LM, Chen Y, Cai F, Chen S, Moquin S, Du X, Joyce MG, Srivatsan S, Zhang B, Zheng A, Shaw GM, Hahn BH, Kepler TB, Korber BTM, Kwong PD, Mascola JR, Haynes BF, NISC Comparative Sequencing Program.** 2013. Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. Nature **496:**469–476. http://dx.doi.org/10.1038/nature12053.
15. **Sok D, van Gils MJ, Pauthner M, Julien J, Saye-Francisco KL, Hsueh J, Briney B, Lee JH, Le KM, Lee PS, Hua Y, Seaman MS, Moore JP, Ward AB, Wilson IA, Sanders RW, Burton DR.** 2014. Recombinant HIV envelope trimer selects for quaternary-dependent antibodies targeting the trimer apex. Proc Natl Acad Sci U S A **111:**17624–17629. http://dx.doi.org/10.1073/pnas.1415789111.
16. **Kwong PD, Mascola JR, Nabel GJ.** 2013. Broadly neutralizing antibodies and the search for an HIV-1 vaccine: the end of the beginning. Nat Rev Immunol **13:**693–701. http://dx.doi.org/10.1038/nri3516.
17. **Kwong P, Mascola J.** 2012. Human antibodies that neutralize HIV-1: identification, structures, and B cell ontogenies. Immunity **37:**412–425. http://dx.doi.org/10.1016/j.immuni.2012.08.012.
18. **Klein F, Mouquet H, Dosenovic P, Scheid JF, Scharf L, Nussenzweig MC.** 2013. Antibodies in HIV-1 vaccine development and therapy. Science **341:**1199–1204. http://dx.doi.org/10.1126/science.1241144.
19. **Zhu J, Wu X, Zhang B, McKee K, O'Dell S, Soto C, Zhou T, Casazza JP, Mullikin JC, Kwong PD, Mascola JR, Shapiro L, NISC Comparative Sequencing Program.** 2013. De novo identification of VRC01 class HIV-1-neutralizing antibodies by next-generation sequencing of B-cell transcripts. Proc Natl Acad Sci U S A **110:**E4088–E4097. http://dx.doi.org/10.1073/pnas.1306262110.
20. **Zhu J, Ofek G, Yang Y, Zhang B, Louder MK, Lu G, McKee K, Pancera M, Skinner J, Zhang Z, Parks R, Eudailey J, Lloyd KE, Blinn J, Alam SM, Haynes BF, Simek M, Burton DR, Koff WC, Mullikin JC, Mascola JR, Shapiro L, Kwong PD, Progra NCS.** 2013. Mining the antibodyome for HIV-1-neutralizing antibodies with next-generation sequencing and phylogenetic pairing of heavy/light chains. Proc Natl Acad Sci U S A **110:** 6470–6475. http://dx.doi.org/10.1073/pnas.1219320110.
21. **Zhu J, O'Dell S, Ofek G, Pancera M, Wu X, Zhang B, Zhang Z, Mullikin JC, Simek M, Burton DR, Koff WC, Shapiro L, Mascola JR, Kwong PD, NISC Comparative Sequencing Program.** 2012. Somatic populations of PGT135-137 HIV-1-neutralizing antibodies identified by 454 pyrosequencing and bioinfornnatics. Front Microbiol **3:**315.
22. **Sok D, Laserson U, Laserson J, Liu Y, Vigneault F, Julien J, Briney B, Ramos A, Saye KF, Le K, Mahan A, Wang S, Kardar M, Yaari G, Walker LM, Simen BB, John EP, Chan-Hui P, Swiderek K, Kleinstein SH, Alter G, Seaman MS, Chakraborty AK, Koller D, Wilson IA, Church GM, Burton DR, Poignard P.** 2013. The effects of somatic hypermutation on neutralization and binding in the PGT121 family of broadly neutralizing HIV antibodies. PLoS Pathog **9:**e1003754. http://dx.doi.org/10.1371/journal.ppat.1003754.
23. **Wu X, Zhang Z, Schramm C, Joyce M, Do Kwon Y, Zhou T, Sheng Z, Zhang B, O'Dell S, McKee K, Georgiev I, Chuang G, Longo N, Lynch R, Saunders K, Soto C, Srivatsan S, Yang Y, Bailer R, Louder M, Mullikin JC, Connors M, Kwong PD, Mascola JR, Shapiro L, Program NCS.** 2015. Maturation and diversity of the VRC01-antibody lineage over 15 years of chronic HIV-1 infection. Cell **161:**470–485. http://dx.doi.org/10.1016/j.cell.2015.03.004.
24. **Hatziioannou T, Evans DT.** 2012. Animal models for HIV/AIDS research. Nat Rev Microbiol **10:**852–867. http://dx.doi.org/10.1038/nrmicro2911.
25. **Ambrose Z, KewalRamani VN, Bieniasz PD, Hatziioannou T.** 2007. HIV/AIDS: in search of an animal model. Trends Biotechnol **25:**333–337. http://dx.doi.org/10.1016/j.tibtech.2007.05.004.
26. **Sui Y, Gordon S, Franchini G, Berzofsky JA.** 2013. Nonhuman primate models for HIV/AIDS vaccine development. Curr Protoc Immunol **102.** 12/14/11 **12:**14.30.
27. **Joag SV.** 2000. Primate models of AIDS. Microb Infect **2:**223–229.
28. **Nath BM, Schumann KE, Boyer JD.** 2000. The chimpanzee and other non-human-primate models in HIV-1 vaccine research. Trends Microbiol **8:**426–431. http://dx.doi.org/10.1016/S0966-842X(00)01816-3.
29. **Morgan C, Marthas M, Miller C, Duerr A, Cheng-Mayer C, Desrosiers R, Flores J, Haigwood N, Hu S, Johnson RP, Lifson J, Montefiori D, Moore J, Robert-Guroff M, Robinson H, Self S, Corey L.** 2008. The use of nonhuman primate models in HIV vaccine development. PLoS Med **5:**1200–1204. http://dx.doi.org/10.1371/journal.pmed.0050173.
30. **Sundling C, Forsell MNE, O'Dell S, Feng Y, Chakrabarti B, Rao SS,**

Lore K, Mascola JR, Wyatt RT, Douagi I, Hedestam GB. 2010. Soluble HIV-1 Env trimers in adjuvant elicit potent and diverse functional B cell responses in primates. J Exp Med **207:**2003–2017. http://dx.doi.org/10.1084/jem.20100025.

31. **Sundling C, Phad G, Douagi I, Navis M, Hedestam GB**. 2012. Isolation of antibody V(D)J sequences from single cell sorted rhesus macaque B cells. J Immunol Methods **386:**85–93. http://dx.doi.org/10.1016/j.jim.2012.09.003.
32. **Sundling C, Li Y, Huynh N, Poulsen C, Wilson R, O'Dell S, Feng Y, Mascola JR, Wyatt RT, Hedestam GB**. 2012. High-resolution definition of vaccine-elicited B cell responses against the HIV primary receptor binding site. Sci Transl Med **4:**142ra196. http://dx.doi.org/10.1126/scitranslmed.3003752.
33. **Zhou T, Georgiev I, Wu X, Yang Z-Y, Dai K, Finzi A, Do Kwon Y, Scheid JF, Shi W, Xu L, Yang Y, Zhu J, Nussenzweig MC, Sodroski J, Shapiro L, Nabel GJ, Mascola JR, Kwong PD**. 2010. Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. Science **329:**811–817. http://dx.doi.org/10.1126/science.1192819.
34. **Sundling C, Zhang Z, Phad GE, Sheng Z, Wang Y, Mascola JR, Li Y, Wyatt RT, Shapiro L, Hedestam GB**. 2014. Single-cell and deep sequencing of IgG-switched macaque B cells reveal a diverse Ig repertoire following immunization. J Immunol **192:**3637–3644. http://dx.doi.org/10.4049/jimmunol.1303334.
35. **Guo K, Halemano K, Schmitt K, Katuwal M, Wang Y, Harper MS, Heilman KJ, Kuwata T, Stephens EB, Santiago ML**. 2015. Immunoglobulin V-H gene diversity and somatic hypermutation during SIV infection of rhesus macaques. Immunogenetics **67:**355–370. http://dx.doi.org/10.1007/s00251-015-0844-3.
36. **Julien J-P, Cupo A, Sok D, Stanfield RL, Lyumkis D, Deller MC, Klasse P-J, Burton DR, Sanders RW, Moore JP, Ward AB, Wilson IA**. 2013. Crystal structure of a soluble cleaved HIV-1 envelope trimer. Science **342:**1477–1483. http://dx.doi.org/10.1126/science.1245625.
37. **Lyumkis D, Julien J-P, de Val N, Cupo A, Potter CS, Klasse P-J, Burton DR, Sanders RW, Moore JP, Carragher B, Wilson IA, Ward AB**. 2013. Cryo-EM structure of a fully glycosylated soluble cleaved HIV-1 envelope trimer. Science **342:**1484–1490. http://dx.doi.org/10.1126/science.1245627.
38. **Sanders RW, Derking R, Cupo A, Julien J, Yasmeen A, de Val N, Kim HJ, Blattner C, de la Peña AT, Korzun J, Golabek M, de los Reyes K, Ketas TJ, van Gils MJ, King CR, Wilson IA, Ward AB, Klasse PJ, Moore JP**. 2013. A next-generation cleaved, soluble HIV-1 Env trimer, BG505 SOSIP.664 gp140, expresses multiple epitopes for broadly neutralizing but not non-neutralizing antibodies. PLoS Pathog **9:**e1003618. http://dx.doi.org/10.1371/journal.ppat.1003618.
39. **Pancera M, Zhou T, Druz A, Georgiev IS, Soto C, Gorman J, Huang J, Acharya P, Chuang G, Ofek G, Stewart-Jones GBE, Stuckey J, Bailer RT, Joyce MG, Louder MK, Tumba N, Yang Y, Zhang B, Cohen MS, Haynes BF, Mascola JR, Morris L, Munro JB, Blanchard SC, Mothes W, Connors M, Kwong PD**. 2014. Structure and immune recognition of trimeric pre-fusion HIV-1 Env. Nature **514:**455–461. http://dx.doi.org/10.1038/nature13808.
40. **Sharma S, de Val N, Bale S, Guenaga J, Tran K, Feng Y, Dubrovskaya V, Ward A, Wyatt R**. 2015. Cleavage-independent HIV-1 Env trimers engineered as soluble native spike mimetics for vaccine design. Cell Rep **11:**539–550. http://dx.doi.org/10.1016/j.celrep.2015.03.047.
41. **He L, Sok D, Azadnia P, Hsueh J, Landais E, Simek M, Koff WC, Poignard P, Burton DR, Zhu J**. 2014. Toward a more accurate view of human B-cell repertoire by next-generation sequencing, unbiased repertoire capture and single-molecule barcoding. Sci Rep **4:**6778. http://dx.doi.org/10.1038/srep06778.
42. **Bashford-Rogers R, Palser AL, Idris SF, Carter L, Epstein M, Callard RE, Douek DC, Vassiliou GS, Follows GA, Hubank M, Kellam P**. 2014. Capturing needles in haystacks: a comparison of B-cell receptor sequencing methods. BMC Immunol **15:**9. http://dx.doi.org/10.1186/s12865-014-0029-0.
43. **Tran K, Poulsen C, Guenaga J, de Val N, Wilson R, Sundling C, Li Y, Stanfield RL, Wilson IA, Ward AB, Hedestam GB, Wyatt RT**. 2014. Vaccine-elicited primate antibodies use a distinct approach to the HIV-1 primary receptor binding site informing vaccine redesign. Proc Natl Acad Sci U S A **111:**E738–E747. http://dx.doi.org/10.1073/pnas.1319512111.
44. **Koff WC, Burton DR, Johnson PR, Walker BD, King CR, Nabel GJ, Ahmed R, Bhan MK, Plotkin SA**. 2013. Accelerating next-generation vaccine development for global disease prevention. Science **340:**1232910. http://dx.doi.org/10.1126/science.1232910.
45. **Nabel GJ**. 2013. Designing tomorrow's vaccines. N Engl J Med **368:**551–560. http://dx.doi.org/10.1056/NEJMra1204186.
46. **Koff WC, Gust ID, Plotkin SA**. 2014. Toward a human vaccines project. Nat Immunol **15:**589–592. http://dx.doi.org/10.1038/ni.2871.
47. **He L, Zhu J**. 2015. Computational tools for epitope vaccine design and evaluation. Curr Opin Virol **11:**103–112. http://dx.doi.org/10.1016/j.coviro.2015.03.013.
48. **Xie J, Yea K, Zhang H, Moldt B, He L, Zhu J, Lerner R**. 2014. Prevention of cell death by antibodies selected from intracellular combinatorial libraries. Chem Biol **21:**274–283. http://dx.doi.org/10.1016/j.chembiol.2013.12.006.
49. **DeKosky BJ, Ippolito GC, Deschner RP, Lavinder JJ, Wine Y, Rawlings BM, Varadarajan N, Giesecke C, Dörner T, Andrews SF, Wilson PC, Hunicke-Smith SP, Willson CG, Ellington AD, Georgiou G**. 2013. High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. Nat Biotechnol **31:**166–169. http://dx.doi.org/10.1038/nbt.2492.
50. **Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG**. 2007. Clustal W and Clustal X version 2.0. Bioinformatics **23:**2947–2948. http://dx.doi.org/10.1093/bioinformatics/btm404.
51. **Altschul S, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ**. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25:**3389–3402. http://dx.doi.org/10.1093/nar/25.17.3389.
52. **Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, Rupp R**. 2007. Dendroscope: an interactive viewer for large phylogenetic trees. BMC Bioinformatics **8:**460. http://dx.doi.org/10.1186/1471-2105-8-460.
53. **Li M, Gao F, Mascola JR, Stamatatos L, Polonis VR, Koutsoukos M, Voss G, Goepfert P, Gilbert P, Greene KM, Bilska M, Kothe DL, Salazar-Gonzalez JF, Wei X, Decker JM, Hahn BH, Montefiori DC**. 2005. Human immunodeficiency virus type 1 env clones from acute and early subtype B infections for standardized assessments of vaccine-elicited neutralizing antibodies. J Virol **79:**10108–10125. http://dx.doi.org/10.1128/JVI.79.16.10108-10125.2005.
54. **Wu X, Zhou T, O'Dell S, Wyatt RT, Kwong PD, Mascola JR**. 2009. Mechanism of human immunodeficiency virus type 1 resistance to monoclonal antibody b12 that effectively targets the site of CD4 attachment. J Virol **83:**10892–10907. http://dx.doi.org/10.1128/JVI.01142-09.
55. **Seaman MS, Janes H, Hawkins N, Grandpre LE, Devoy C, Giri A, Coffey RT, Harris L, Wood B, Daniels MG, Bhattacharya T, Lapedes A, Polonis VR, McCutchan FE, Gilbert PB, Self SG, Korber BT, Montefiori DC, Mascola JR**. 2010. Tiered categorization of a diverse panel of HIV-1 Env pseudoviruses for assessment of neutralizing antibodies. J Virol **84:**1439–1452. http://dx.doi.org/10.1128/JVI.02108-09.