Supporting Information (SI) Appendix for

De novo identification of VRC01-class HIV-1 broadly neutralizing antibodies by next-generation sequencing analysis of B cell transcripts

Jiang Zhu[a,1], Xueling Wu[a,2], Baoshan Zhang[a], Krisha McKee[a], Sijy O'Dell[a], Cinque Soto[a], Tongqing Zhou[a], Joseph Casazza[a], NISC Comparative Sequencing Program[c], James C. Mullikin[c], Peter D. Kwong[a,3], John R. Mascola[a,3], and Lawrence Shapiro[b,3]

[a] Vaccine Research Center and [c]NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA.

[b] Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, USA.

[1]Present address: The Department of Immunology and Microbial Science, The Scripps Research Institute, 10550 N Torrey Pines Rd La Jolla, CA 92037

[2] Present address: The Aaron Diamond AIDS Research Center, 455 First Avenue, 7th Floor, New York, NY 10016, USA.

[3] To whom correspondence should be addressed:

PDK: Phone (301)594-8439; Email- pdkwong@nih.gov

JRM: Phone (301)496-1852; Email- jmascola@nih.gov

LS: Phone (212)851-5381; Email- lss8@columbia.edu

This Appendix includes:

Materials and Methods

Human Specimens. The sera and peripheral blood mononuclear cells (PBMCs) were obtained from HIV-1 infected donors (**Table S1**) enrolled in investigational review board approved clinical protocols at the National Institute of Allergy and Infectious Diseases. At the time of sampling, all HIV-1 infected donors were off anti-retroviral treatment.

Antibodies, Plasmids, Antibody and Protein Expression and Purification.
Antibody production was similar to what has been described before (1, 2). Briefly, the sequences were selected using the bioinformatics procedures described in the main text and checked for sequencing errors using an automatic error correction procedure (1, 2) (and see below) followed by manual inspection. The corrected antibody sequences were synthesized (GenScript, Inc) and cloned into the CMV/R expression vector containing the constant regions of IgG1 (3). The heavy chains identified from donor C38 antibodyomes by lineage rank and cross-donor phylogenetic analysis were paired with VRC01 light chain DNA for transfection, while C38 light chains identified by signature matching were initially paired with VRC01 heavy chain and then with functional C38 heavy chain DNAs for transfection. Full-length IgGs were expressed from transient transfection of 293F cells and purified using a recombinant protein-A column (Pierce).

HIV-1 Neutralization and Protein Competition Assays. Neutralization was measured using HIV-1 Env-pseudoviruses to infect TZM-bl cells as described (4-6). Neutralization curves were fit by nonlinear regression using a 5-parameter hill slope equation as described(5). The 50% inhibitory concentrations (IC_{50}) were reported as the antibody concentrations required to inhibit infection by 50%. Competition of serum neutralization (7) was assessed by adding a fixed concentration (25 μg/ml) of RSC3 or ΔRSC3 to serial dilutions of serum for 15 min prior to the addition of virus. The resulting IC_{50} values were compared to the control with mock phosphate buffered saline (PBS) added. The neutralization blocking effect of the proteins was calculated as the percent reduction in the ID_{50} (50% inhibitory dilution) value of the serum in the presence of protein compared to PBS.

Sample Preparation for 454 Pyrosequencing. Samples for 454 pyrosequencing were prepared as described (1, 2) with minor modifications. Briefly, mRNA was extracted from 20 million PBMC into 200 µl of elution buffer (Oligotex kit, Qiagen), then concentrated to 10-30 µl by centrifuging the buffer through a 30 kD micron filter (Millipore). The reverse-transcription was performed in one or multiple 35 µl-reactions, each composed of 13 µl of mRNA, 3 µl of oligo$(dT)_{12\text{-}18}$ at 0.5 µg/µl (invitrogen), 7 µl of 5x first strand buffer (Invitrogen), 3 µl of RNase Out (Invitrogen), 3 µl of 0.1M DTT (Invitrogen), 3 µl of dNTP mix (each at 10 mM), and 3 µl of SuperScript II (Invitrogen). The reactions were incubated at 42^{o}C for 2 hours. The cDNAs from each sample were combined, cleaned up and eluted in 20 µl of elution buffer (NucleoSpin Extract II kit, Clontech). In this way, 1 µl of the cDNA was equivalent of transcripts from 1 million PBMC. The immunoglobulin gene-specific PCRs were set up in a total volume of 50 µl, using 5 µl of the cDNA as template (equivalent of transcripts from 5 million PBMC). The DNA polymerase systems used were either the Platinum *Taq* High-Fidelity (HiFi) DNA Polymerase system (Invitrogen) or the Phusion HiFi DNA Polymerase system (Finnzymes). According to the instructions of the manufacturer, the reaction mix was composed of water, the appropriate buffer and 1 µl of supplied MgSO4 if required, 2 µl of dNTP mix (each at 10 mM), 1-2 µl of primers or primer mixes (**Table S2**) at 25 µM, and 1 µl of DNA polymerase. For IGHV1 amplification, the forward primers were either primer mix of pooled equal parts of four VH1 primers (H1) or seven VH1 primers (G1) published by Scheid et al (8) (**Table S2**). The primers each contained the appropriate adaptor sequences (XLR-A or XLR-B) for subsequent 454 pyrosequencing. For the Platinum *Taq* HiFi polymerase, the PCRs were initiated at 95^{o}C for 30 sec, followed by 25 cycles of 95^{o}C for 30 sec, 58^{o}C for 30 sec, and 72^{o}C for 1 min, followed by 72^{o}C for 10 min. For the Phusion HiFi polymerase, the PCRs were initiated at 98^{o}C for 30 sec, followed by 25 cycles of 98^{o}C for 10 sec, 57^{o}C for 30 sec, and 72^{o}C for 30 sec, followed by 72^{o}C for 10 min. The PCR products at the expected size (~500bp) were gel purified (Qiagen), followed by phenol/chloroform extraction.

454 Library Preparation and Pyrosequencing. The 454 pyrosequencing was carried out as described (1, 2). Briefly, PCR products were quantified using Qubit (Life Technologies, Carlsbad, CA). Library concentrations were determined using the KAPA Biosystems qPCR system (Woburn, MA) with 454 standards provided in the KAPA system. 454 pyrosequencing of the PCR products was performed on a GS FLX sequencing instrument (Roche-454 Life Sciences, Bradford, CT) using the manufacturer's suggested methods and reagents. Initial image collection was performed on the GS FLX instrument and subsequent signal processing, quality filtering, and generation of nucleotide sequence and quality scores were performed on an off-instrument linux cluster using 454 application software (version 2.5.3). The amplicon quality filtering parameters were adjusted based on the manufacturer's recommendations (Roche-454 Life Sciences Application Brief No. 001-2010). Quality scores were assigned to each nucleotide using methodologies incorporated into the 454 application software to convert flowgram intensity values to Phred-based quality scores and as described (9). The quality of each run was assessed by analysis of internal control sequences included in the 454 sequencing reagents. Reports were generated for each region of the PicoTiterPlate (PTP) for both the internal controls and the samples.

Bioinformatics Pipeline for Primary Analysis of 454-Pyrosequencing-Determined Antibodyomes. The general bioinformatics pipeline developed in our previous study (1, 2) has been further revised to take all germline genes (heavy chain, κ- and λ-light chains) into consideration and to calculate parameters required for antibodyome analysis. The current pipeline, termed *Antibodyomics 1.0*, consists of five steps. Given a 454-pyrosequencing-determined antibodyome, each sequence read was (1) reformatted and labeled with a unique index number; (2) assigned to variable (V), diverse (D) (for heavy chain only), and joining (J) gene families and alleles using an in-house implementation of IgBLAST (http://www.ncbi.nlm.nih.gov/igblast/), and sequences with E-value > 10^{-3} for V-gene assignment were rejected; (3) subjected to a template-based error correction scheme where 454 homopolymer errors in V, D (for heavy chains only) and J regions were detected and corrected based on the alignment to respective germline sequence (D or J gene was corrected only when the gene assignment was reliable, indicated by E-value < 10^{-3}); (4) compared with the a set of template antibody sequences at both nucleotide level and amino-acid level using a global alignment module in CLUSTALW2 (10), which provides the basis for the identity/divergence analysis; (5) subjected to a multiple sequence alignment (MSA)-based scheme to determine the third complementarity-determining region (CDR H3 or L3), which was further compared with a set of template CDR H3 or L3 sequences at nucleotide level, and to determine the sequence boundary of the variable domain. In this scheme, the multiple alignments of representative germline V genes (truncated before the CDR3 region) and J genes (truncated after the CDR3 region) were pre-calculated and used for determination of variable domain sequences. For example, in heavy chain analysis, a 454-pyrosequencing-derived sequence will be added to the multiple alignment of 52 representative IGHV genes (truncated at the second amino acid after a cysteine near the C-terminus) and 6 representative IGHJ genes (truncated before the WGXG motif) to determine the CDR H3 region, which lies between the last column of aligned V genes and first column of aligned J regions, as well as the variable domain, which lies between the first column of aligned V genes and last column of aligned J genes. Only full-length variable domain sequences were retained in the final data set for subsequent analysis.

Lineage assignment and lineage rank analysis of heavy chain antibodyomes. The lineage rank analysis consisted of two consecutive steps, to identify CDR H3 groups and to determine heavy-chain lineages based on identified CDR H3 groups. An iterative procedure has been developed to identify the largest CDR H3 group in a given antibodyome, which was removed from the antibodyome and the rest of sequences were considered a new antibodyome and subjected to the next iteration of analysis until the resulting CDR H3 group contained no more than 300 sequences. In each iteration, the CDR H3 sequences present in the given antibodyome were compiled into a BLAST database using “makeblastdb” in the NCBI-BLAST package (11), and then each CDR H3 sequence was used to BLAST search against the database, allowing one-nucleotide variation in CDR H3 length (usually caused by 454 sequencing error) and no more than 5-nucleotide difference in CDR H3 alignment. Sequences found using this criterion were defined as a “CDR H3 group”. Sequences in the largest CDR H3 group were extracted from the antibodyome and subjected to a clustering analysis to determine the representative CDR H3. In the second step, multiple CDR H3 groups were merged into one lineage if their CDR H3 sequences were of the same length and shared over 80% amino-acid sequence identity, and if the V gene variation was not increased significantly upon group merging. Of note, the variation here was calculated as averaged nucleotide difference of all sequences in the CDR H3 group from the consensus V gene sequence.

Cross-donor phylogenetic analysis of donor C38 heavy chain antibodyomes. A revised procedure was used for the cross-donor phylogenetic analysis, which consists of an iterative analysis based on the neighbor-joining (NJ) method (12) implemented in CLUSTALW2 (10) and a second-step analysis based on the maximum likelihood (ML) method with molecular clock implemented in DNAMLK (http://evolution.genetics.washington.edu/phylip/doc/promlk.html) in the PHYLIP package v3.69 (http://evolution.genetics.washington.edu/phylip.html).

In the NJ-based analysis, the donor heavy chain variable domain (V_H) sequences of IGHV1-2 origin were randomly shuffled and divided into subsets of no more than 5,000 sequences. For each subset, the V_H sequences of 13 template VRC01-like antibodies – VRC01, VRC02, VRC03, VRC-PG04, VRC-PG04b, VRC-CH30, VRC-CH31, VRC-CH32, 12A12, 12A21, 3BNC60, 3BNC117 and NIH45-46 – and their germline V gene – IGHV1-2*02 – were added exogenously. A NJ tree was then constructed for each subset using the "Phylogenetic trees" option in CLUSTALW2 (10). The donor sequences clustered in the smallest branch that contains all template VRC01-like antibodies were extracted from each NJ tree and combined into a new data set for the next round of cross-donor NJ analysis. The analysis was repeated until convergence, where all the donor sequences resided within a VRC01-like sub-tree containing all 13 template VRC01-class antibodies and no other sequences resided between this sub-tree and the root, and where further repeat of the analysis did not change the NJ tree. Of note, the use of sequence shuffling has been found to improve the convergence efficiency, and the use of V gene instead of inferred reverted unmutated ancestors of VRC01 (or VRC-PG04) for tree rooting can eliminate potential biases for particular sets of D and J genes.

ML-based analysis was used to confirm the cross-donor dendrogram derived from the NJ-based analysis. Starting from the data set obtained from the last iteration of NJ analysis, the multiple sequence alignment generated by CLUSTALW2 (10) was provided as input to construct a phylogenetic tree using DNAMLK (for DNA Maximum Likelihood program with Molecular Clock) (http://cmgm.stanford.edu/phylip/dnamlk.html) in the PHYLIP package v3.69 (http://evolution.genetics.washington.edu/phylip.html). The calculation was done with default parameters (empirical base frequencies, a transitions to transversions ratio of 2.0, and an overall base substitution model with A 0.24, C 0.28, G 0.27, T 0.21). The output unrooted tree was visualized using Dendroscope (13), then ordered to ladderize right and rooted by the IGHV1-2*02 germline gene. Any sequences outside the ML-defined VRC01-like sub-tree were removed and the remaining sequences were used to construct the final cross-donor dendrogram (**Fig. 3**).

The robustness of cross-donor phylogenetic analysis was examined from three aspects. First, we asked if the cross-donor identification was driven by the high divergence of template antibodies. Our calculations showed that the germline divergence of 191 cross-donor-identified V_H sequences from the donor C38 G1-primer data set ranged from 19.9 to 35.7%, which overlapped with the divergence of 13 template antibodies. However, they only accounted for 0.4% of IGHV1-2-originated sequences within the same divergence range, suggesting that divergence is not a deterministic factor in the cross-donor analysis. Second, we examined if cross-donor analysis could be applied to sequences originated from other germline genes. To assess this possibility, we performed a similar cross-donor analysis on IGHV1-8, IGHV1-18, IGHV1-46 and IGHV1-69 families. Most analyses converged rapidly after a single iteration with no sequence identified except for 591 sequences from cross-donor analysis of IGVH1-46 family, among which 265 are unique. 3 sequences were selected from the cross-donor dendrogram for experimental validation, however, with no detectable neutralizing activity (**Table S10**). Finally, we tested whether cross-donor identification was restricted by the J genes of template antibodies. Most of the cross-donor-identified heavy chain variable domain sequences from G1-primer data set used JH5 and JH6 genes, with the exception of two sequences clustered with VRC01-03, which shared the IGHJ1*01 allelic origin, confirming that cross-donor analysis was not restricted to a set of particular J genes or by the J gene usage of template antibodies.

Sequence-specific motif search of donor C38 light chain antibodyomes. The CDR L3 sequences were extracted from the light chain antibodyomes after pipeline processing. Two sequence-specific features - a CDR L3 length of 5 amino acids and glutamine (Q) or glutamate (E) at position 96 (Kabat numbering) or position 4 within the CDR L3 sequence - were used to search for VRC01-class antibody light chains. Four κ-chains were identified from the 454-pyrosequencing data set generated from amplification with both κ and λ primers, and nine κ-chains were found from the second data set generated from amplification with κ primers only.

References

1. Zhu J, *et al.* (2012) Somatic Populations of PGT135-137 HIV-1-Neutralizing Antibodies Identified by 454 Pyrosequencing and Bioinformatics. *Frontiers in microbiology* 3:315-315.
2. Wu X, *et al.* (2011) Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. (Translated from eng) *Science* 333(6049):1593-1602 (in eng).
3. Barouch DH & Nabel GJ (2005) Adenovirus vector-based vaccines for human immunodeficiency virus type 1. *Hum Gene Ther* 16(2):149-156.
4. Li M, *et al.* (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(16):10108-10125.
5. Seaman MS, *et al.* (2010) Tiered categorization of a diverse panel of HIV-1 Env pseudoviruses for neutralizing antibody assessment. (Translated from Eng) *J Virol* 84(3):1439-1452 (in Eng).
6. Wu X, *et al.* (2009) Mechanism of human immunodeficiency virus type 1 resistance to monoclonal antibody B12 that effectively targets the site of CD4 attachment. (Translated from eng) *J Virol* 83(21):10892-10907 (in eng).
7. Wu X, *et al.* (2010) Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. (Translated from eng) *Science* 329(5993):856-861 (in eng).
8. Scheid JF, *et al.* (2011) Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. (Translated from eng) *Science* 333(6049):1633-1637 (in eng).
9. Brockman W, *et al.* (2008) Quality scores and SNP detection in sequencing-by-synthesis systems. *Genome Research* 18(5):763-770.
10. Larkin MA, *et al.* (2007) Clustal W and Clustal X version 2.0. (Translated from eng) *Bioinformatics* 23(21):2947-2948 (in eng).
11. Altschul SF, Gish W, Miller W, Myers EW, & Lipman DJ (1990) BASIC LOCAL ALIGNMENT SEARCH TOOL. *Journal of molecular biology* 215(3):403-410.
12. Kuhner MK & Felsenstein J (1994) A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. (Translated from eng) *Mol Biol Evol* 11(3):459-468 (in eng).
13. Huson DH, *et al.* (2007) Dendroscope: An interactive viewer for large phylogenetic trees. *BMC Bioinformatics* 8:460.

```
Heavy chain ---------FR1------------------_____CDR1_____-----FR2------_____CDR2_________-----------------FR3--------------------______CDR3______----FR4----
IGHV1-2*02  QVQLVQSGAEVKKPGASVKVSCKASGYTFTG.........YYMHWVRQAPGQGLEWMGWINPNSGGTNY.AQKFQGRVT..MTR.......DTSISTAYMELSRLRSDDTAVYYCAR..........................

Template VRC01-class heavy chains:
VRC01 H     QVQLVQSGGQMKKPGESMRISCRASGYEFID.........CTLNWIRLAPGKRPEWMGWLKPRGGAVNY.ARPLQGRVT..MTR.......DVYSDTAFLELRSLTVDDTAVYFCTRGKNCD....YNWDFEHWGRGTPVIVSS
VRC02 H     QVQLVQSGGQMKKPGESMRISCQASGYEFID.........CTLNWVRLAPGRRPEWMGWLKPRGGAVNY.ARPLQGRVT..MTR.......DVYSDTAFLELRSLTADDTAVYYCTRGKNCD....YNWDFEHWGRGTPVTVSS
VRC03 H     QVQLVQSGAVIKTPGSSVKISCRASGYNFRD.........YSIHWVRLIPDKGFEWIGWIKPLWGAVSY.ARQLQGRVS..MTRQLSQDPDDPDWGVAYMEFSGLTPADTAEYFCVRRGSCD..YCGDFPWQYWGQGTVVVVSS
VRC-PG04 H  QVQLVQSGSGVKKPGASVRVSCWTSEDIFER........TELIHWVRQAPGQGLEWIGWVKTVTGAVNFGSPDFRQRVS..LTR.......DRDLFTAHMDIRGLTQGDTATYFCARQKFYTGG..QGWYFDLWGRGTLIVVSS
VRC-PG04b H QVQLVQSGSGVKKPGASVRVSCWTSEDIFER........TELIHWVRQAPGQGLEWIGWVKTVTGAVNFGSPNFRHRVS..LTR.......DRDLFTAHMDIRGLTQGDTATYFCARQKFERGG..QGWYFDLWGRGTLIVVSS
VRC-CH30 H  QVQLVQSGAAVRKPGASVTVSCKFAEDDDYSPHWVNPAPEHYIHFLRQAPGQQLEWLAWMNPTNGAVNY.AWQLHGRLT..ATR.......DGSMTTAFLEVRSLRSDDTAVYYCARAQKRGR...SEWAYAHWGQGTPVAVSS
VRC-CH31 H  QVQLVQSGAAVRKPGASVTVSCKFAEDDDYSPYWVNPAPEHFIHFLRQAPGQQLEWLAWMNPTNGAVNY.AWYLNGRVT..ATR.......DRSMTTAFLEVKSLRSDDTAVYYCARAQKRGR...SEWAYAHWGQGTPVVVSS
VRC-CH32 H  QVQLVQSGAAVRKPGASVTVSCKFAEDDDFSPHWVNPAPEHYIHFLRQAPGQQLEWLAWMKPTNGAVNY.AWQLQGRVT..VTR.......DRSQTTAFLEVKNLRSDDTAVYYCARAQKRGR...SEWAYAHWGQGTPVVISA
3BNC60 H    QVHLSQSGAAVTKPGASVRVSCEASGYKISD.........HFIHWWRQAPGQGLQWVGWINPKTGQPNN.PRQFQGRVS..LTR...QASWDFDTYSFYMDLKAVRSDDTAIYFCARQRS......DFWDFDVWGSGTQVTVSS
3BNC117 H   QVQLLQSGAAVTKPGASVRVSCEASGYNIRD.........YFIHWWRQAPGQGLQWVGWINPKTGQPNN.PRQFQGRVS..LTR...HASWDFDTFSFYMDLKALRSDDTAVYFCARQRS......DYWDFDVWGSGTQVTVSS
12A12 H     SQHLVQSGTQVKKPGASVRISCQASGYSFTD.........YVLHWWRQAPGQGLEWMGWIKPVYGARNY.ARRFQGRIN..FDR.......DIYREIAFMDLSGLRSDDTALYFCARDGSGDDT...SWHLDPWGQGTLVIVSA
12A21 H     SQHLVQSGTQVKKPGASVRVSCQASGYTFTN.........YILHWWRQAPGQGLEWMGLIKPVFGAVNY.ARQFQGRIQ..LTR.......DIYREIAFLDLSGLRSDDTAVYYCARDESGDDL...KWHLHPWGQGTQVIVSP
NIH45-46 H  QVRLSQSGGQMKKPGESMRLSCRASGYEFLN.........CPINWIRLAPGRRPEWMGWLKPRGGAVNY.ARKFQGRVT..MTR.......DVYSDTAFLELRSLTSDDTAVYFCTRGKYCTARDYYNWDFEHWGRGAPVTVSS
```

Fig. S1. Sequence alignment of 13 template VRC01-class antibody heavy chain variable domain (VH) sequences. The sequence of inferred germline gene, IGHV1-2*02, is used as reference in alignment. Amino acids in their V genes that differ from IGHV1-2*02 are highlighted in red.

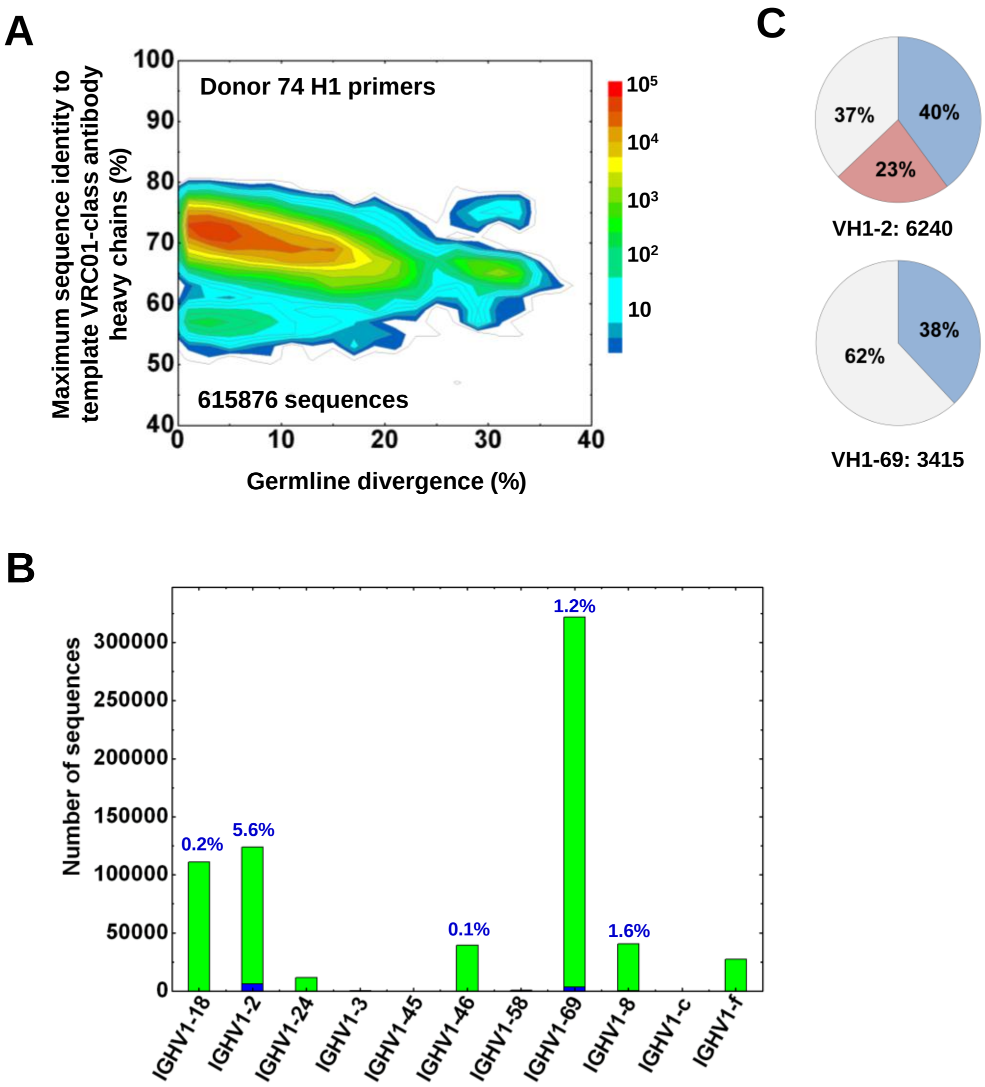


Fig. S2. Analysis of the repertoire of heavy chain sequences from donor 74 (2008 sample) generated with H1 primers. (A) Heavy chain sequences are plotted as a function of maximal sequence identity to the heavy chains of all template VRC01-class antibodies except VRC-PG04 and 04b, which were isolated from donor 74, and of sequence divergence from inferred germline genes. Color coding indicates the number of sequences; (B) The distribution of heavy chain sequences from VH1 germline families (in green) and sequences with a divergence of 20% or greater (in blue, percentage labeled for VH1-18, VH1-2, VH-46, VH1-69 and VH1-8 families); (C) Lineage rank analysis of divergent heavy chain sequences (>20%) from VH1-2 and VH1-69 families (other VH1 families do not have lineages with more than 1000 sequences). Note that the two VH1-2 lineages correspond to the broadly neutralizing VRC-PG04-like antibodies designated classes 7 and 8 in our earlier paper[20].

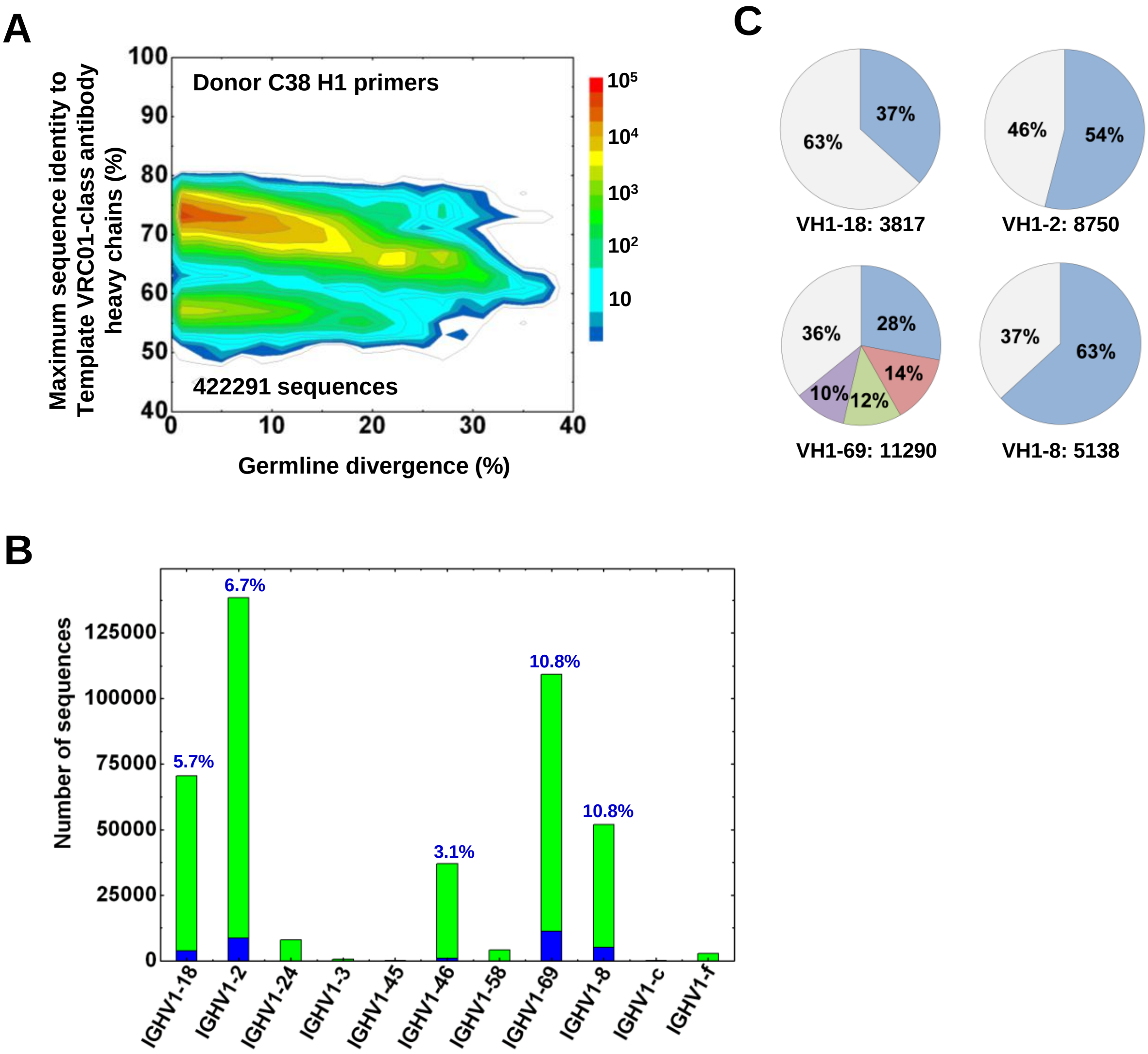
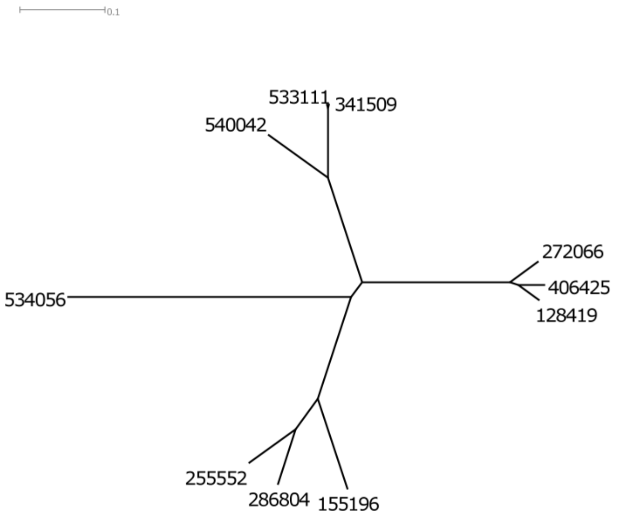


Fig. S3. Analysis of the repertoire of heavy chain sequences from donor C38 (2008 sample) generated with H1 primers. (A) Heavy chain sequences are plotted as a function of maximal sequence identity to the heavy chains of 13 template VRC01-class antibodies and of sequence divergence from putative germline genes. Color coding indicates the number of sequences; (B) The distribution of heavy chain sequences from VH1 germline families (in green) and sequences with a divergence of 20% or greater (in blue, percentage labeled for VH1-18, VH1-2, VH-46, VH1-69 and VH1-8 families); (C) Lineage rank analysis of divergent heavy chain sequences (>20%) from VH1-18, VH1-2, VH1-69 and VH1-8 families (VH1-46 does not have lineages with more than 1000 sequences).

A **Maximum-likelihood tree of 10 C38 heavy chains**



B **Germline gene family analysis of 10 functional C38 heavy chains.**

Seq ID	Donor	IGHV	IGHD	IGHJ	IMGT CDRH3 length (amino acids)
534056	C38	1-2*02	2-21*01	1*01	18
255552	C38	1-2*02	2-21*02	5*02	12
286804	C38	1-2*02	2-21*02	5*02	12
155196	C38	1-2*02	2-21*02	5*02	12
128419	C38	1-2*02	7-27*01	5*02	12
406425	C38	1-2*02	7-27*01	5*02	12
272066	C38	1-2*02	7-27*01	5*02	12
540042	C38	1-2*02	5-24*01	5*02	11
533111	C38	1-2*02	5-24*01	5*02	11
341509	C38	1-2*02	5-24*01	5*02	11

Fig. S4. Phylogenetic analysis and V(D)J gene family analysis of 10 neutralizing heavy chain sequences identified from C38 by cross-donor phylogenetic analysis. (A) Maximum likelihood (ML) tree of 10 sequences rooted at IGHV1-2*02 for visualization. The tree revealed 4 major branches, indicating 4 distinct maturation patterns in the V-gene. (B) The V(D)J usage of 10 sequences inferred using a combination of IMGT (http://www.imgt.org/IMGT_vquest/share/textes/), IgBLAST (http://www.ncbi.nlm.nih.gov/igblast/) and JoinSolver (http://joinsolver.niaid.nih.gov/), supporting 4 distinct lineages, which are consistent with the DNAML tree. The correspondence between sequence indexes and names (gVRC-H1-10$_{dC38}$) for these 10 heavy chains can be found in **Fig. 3**.

A

	IGHV1-2*02													IGHJ1*01															
	GCG	A~~GA~~	~~GA~~										~~GCT~~	~~GAA~~	~~T~~AC	TTC	CAG	CAC	TGG	GGC	CAG	GGC	ACC	CTG	GTC	ACC	GTC	TCC	TCA
	A	~~R~~			IGHD2-21*01								~~A~~	~~E~~	~~Y~~	F	Q	H	W	G	Q	G	T	L	V	T	V	S	S
			~~A~~	~~GCA~~	TAT	TGT	GGT	GGT	GAT	~~TGC~~	~~TAT~~	~~TCC~~																	
				~~A~~	Y	C	G	G	D	~~C~~	~~Y~~	~~S~~																	
534056	GCC	ATG	AGA	GAT	TAT	TGT	CGT	GAT	GAT	AAT	TGT	AAT	AGA	TGG	GAC	CTC	GGT	CAC	TGG	GGC	CAG	GGC	AGC	CTC	ATC	GTC	GTC	TCC	GCG
	A	M	R	D	Y	C	R	D	D	N	C	N	R	W	D	L	G	H	W	G	Q	G	S	L	I	V	V	S	A
														IGHJ2*01															
												~~C~~	~~TAC~~	TGG	TAC	TTC	GAT	CTC	TGG	GGC	CGT	GGC	ACC	CTG	GTC	ACT	GTC	TCC	TCA
													~~Y~~	W	Y	F	D	L	W	G	R	G	T	L	V	T	V	S	S
534056	GCC	ATG	AGA	GAT	TAT	TGT	CGT	GAT	GAT	AAT	TGT	AAT	AGA	TGG	GAC	CTC	GGT	CAC	TGG	GGC	CAG	GGC	AGC	CTC	ATC	GTC	GTC	TCC	GCG
	A	M	R	D	Y	C	R	D	D	N	C	N	R	W	D	L	G	H	W	G	Q	G	S	L	I	V	V	S	A
																		IGHJ5*02											
													~~AC~~	~~AAC~~	~~TGG~~	~~TTC~~	~~GAC~~	CCC	TGG	GGC	CAG	GGA	ACC	CTG	GTC	ACC	GTC	TCC	TCA
														~~N~~	~~W~~	~~F~~	~~D~~	P	W	G	Q	G	T	L	V	T	V	S	S
534056	GCC	ATG	AGA	GAT	TAT	TGT	CGT	GAT	GAT	AAT	TGT	AAT	AGA	TGG	GAC	CTC	GGT	CAC	TGG	GGC	CAG	GGC	AGC	CTC	ATC	GTC	GTC	TCC	GCG
	A	M	R	D	Y	C	R	D	D	N	C	N	R	W	D	L	G	H	W	G	Q	G	S	L	I	V	V	S	A

Fig. S5. CDR H3 analysis of 10 454-pyrosequencing-identified neutralizing heavy chain sequences from donor C38. (A) CDR H3 analysis of gVRC-H2$_{dC38}$ with a sequence index of 534056. The nucleotide and amino acid sequences of the CDR H3 and J region of the heavy chain sequence 534056 were aligned to the putative V, D and J germline genes. Putative nucleotide excisions are indicated with strikethrough lines. In blue are the putative TdT N additions in V-D and D-J junctions. In red are mutations from the putative germline genes and the TdT N additions. Two alternative but unfavorable J genes (IGHJ2*01 and IGHJ5*02) are also indicated.

B

```
             IGHV1-2*02                                        IGHJ5*02
             GCG AGA GA             ~~AC AAC TGG TTC GAC CCC~~ TGG GGC CAG GGA ACC CTG GTC ACC GTC TCC TCA
             A   R      IGHD2-21*02   ~~N   W   F   D   P~~   W   G   Q   G   T   L   V   T   V   S   S
                  ~~A GCA TAT TGT~~ GGT GGT GAC TG~~C TAT TCC~~
                     ~~A   Y   C~~   G   G   D   ~~C   Y   S~~
255552       GCG AGA GGA TTT GGG GGT TCT GAC TGG AGT TTC CTG TGG GGT CAG GGA ACC CTC ATA ATA GTC TCG TCT
             A   R   G   F   G   G   S   D   W   S   F   L   W   G   Q   G   T   L   I   I   V   S   S
286804       GCG AGA GGA TTT GGG GGT TCT GAC TGG AAT TTC GTG TGG GGT CAA GGA ACC CGA ATT ACA GTC TCG GCT
             A   R   G   F   G   G   S   D   W   N   F   V   W   G   Q   G   T   R   I   T   V   S   A
155196       GCG AGA GGA TTT GCC GGT TAT GAG TGG AGT TTC CTC TGG GGT CAG GGA ACT CTG GTC ATA GTC TCC TCT
             A   R   G   F   A   G   Y   E   W   S   F   L   W   G   Q   G   T   L   V   I   V   S   S


             IGHV1-2*02                                        IGHJ5*02
             ~~GCG AGA GA~~         ~~AC AAC TGG TTC GA~~C CCC TGG GGC CAG GGA ACC CTG GTC ACC GTC TCC TCA
             ~~A   R~~    IGHD7-27*01     ~~N   W   F   D~~   P   W   G   Q   G   T   L   V   T   V   S   S
                     ~~CTA~~ ACT GGG ~~GA~~
                     ~~L~~   T   G
128419       GTC AAG GGG ACT GGG GGC AAT GAA TGG GGT TTC GTC TGG GGC CAG GGA TCC CTG GTC GTC GTC TCG CCA
             V   K   G   T   G   G   N   E   W   G   F   V   W   G   Q   G   S   L   V   V   V   S   P
272066       GTC AAG GGG ACT GGA GGC AAT GAA TGG GGT TTC TCG TGG GGC CAG GGA ACC GTG GTC GTC GTC TCG CCA
             V   K   G   T   G   G   N   E   W   G   F   S   W   G   Q   G   T   V   V   V   V   S   P
406425       GTC AAG GGG ACT GGA GGC AAT GAG TGG GGT TTC GTC TGG GGC CAG GGT ACG GTG GTC GTC GTC TCG CCA
             V   K   G   T   G   G   N   E   W   G   F   V   W   G   Q   G   T   V   V   V   V   S   P


             IGHV1-2*02                                        IGHJ5*02
             GCG AGA GA             ~~AC AAC TGG TTC GA~~C CCC TGG GGC CAG GGA ACC CTG GTC ACC GTC TCC TCA
             A   R       IGHD3-16*02      ~~N   W   F   D~~   P   W   G   Q   G   T   L   V   T   V   S   S
      ~~GT ATT ATG ATT ACG~~ TTT GGG GGA ~~GTT ATC GTT ATA CC~~
          ~~I   M   I   T~~   F   G   G   ~~V   I   V   I~~
155196       GCG AGA GGA TTT GCC GGT TAT GAG TGG AGT TTC CTC TGG GGT CAG GGA ACT CTG GTC ATA GTC TCC TCT
             A   R   G   F   A   G   Y   E   W   S   F   L   W   G   Q   G   T   L   V   I   V   S   S
255552       GCG AGA GGA TTT GGG GGT TCT GAC TGG AGT TTC CTG TGG GGT CAG GGA ACC CTC ATA ATA GTC TCG TCT
             A   R   G   F   G   G   S   D   W   S   F   L   W   G   Q   G   T   L   I   I   V   S   S
286804       GCG AGA GGA TTT GGG GGT TCT GAC TGG AAT TTC GTG TGG GGT CAA GGA ACC CGA ATT ACA GTC TCG GCT
             A   R   G   F   G   G   S   D   W   N   F   V   W   G   Q   G   T   R   I   T   V   S   A
                             IGHD3-16*02
~~G TAT TAT GAT TAC GTT TGG~~ GGG AGT TAT ~~CGT TAT ACC~~
   ~~Y   Y   D   Y   V   W~~   G   S   Y   ~~R   Y   T~~
128419       GTC AAG GGG ACT GGG GGC AAT GAA TGG GGT TTC GTC TGG GGC CAG GGA TCC CTG GTC GTC GTC TCG CCA
             V   K   G   T   G   G   N   E   W   G   F   V   W   G   Q   G   S   L   V   V   V   S   P
272066       GTC AAG GGG ACT GGA GGC AAT GAA TGG GGT TTC TCG TGG GGC CAG GGA ACC GTG GTC GTC GTC TCG CCA
             V   K   G   T   G   G   N   E   W   G   F   S   W   G   Q   G   T   V   V   V   V   S   P
406425       GTC AAG GGG ACT GGA GGC AAT GAG TGG GGT TTC GTC TGG GGC CAG GGT ACG GTG GTC GTC GTC TCG CCA
             V   K   G   T   G   G   N   E   W   G   F   V   W   G   Q   G   T   V   V   V   V   S   P
```

Fig. S5 (continued) (B) CDR H3 analysis of 6 heavy chain sequences with equal CDR H3 length, gVRC-H1$_{dC38}$ and gVRC-H3$_{dC38}$ to gVRC-H7$_{dC38}$ (index numbers in **Fig. 2**), supporting two separate groups. The nucleotide and amino acid sequences of the CDR H3 and J region of 6 heavy chains were aligned to the putative V, D and J germline genes. Putative nucleotide excisions are indicated with strikethrough lines. In blue are the putative TdT N additions in V-D and D-J junctions. In red are mutations from the putative germline genes and the TdT N additions. One alternative but unfavorable D gene (IGHD3-16*02) is also indicated. Note that even when the same D gene was analyzed, the alignment and the reading frame of the D-gene is different between the two groups.

C

	IGHV1-2*02														IGHJ5*02						
	GC~~G~~	~~AGA~~	~~GA~~			~~AC~~	~~AAC~~	~~TGG~~	~~TTC~~	~~GAC~~	~~CC~~C	TGG	GGC	CAG	GGA	ACC	CTG	GTC	ACC	GTC	TCC	TCA
	A	~~R~~					~~N~~	~~W~~	~~F~~	~~D~~	~~P~~	W	G	Q	G	T	L	V	T	V	S	S
						IGHD5-24*01																
				~~GT~~	~~AGA~~	GAT	GGC	TAC	AAT	~~TAC~~												
					~~R~~	D	G	Y	N	Y												
533111	GCC	TTT	GGG	GCG	GGA	GAT	GGT	TGG	GAT	CTT	GTC	TGG	GGC	CAG	GGA	ACC	CTG	GTC	ATC	GTC	TCC	TCA
	A	F	G	A	G	D	G	W	D	L	V	W	G	Q	G	T	L	V	I	V	S	S
341509	GCC	TTT	GGG	GCG	GGA	GAT	GGT	TGG	GAT	CTT	GTC	TGG	GGC	CAG	GGA	ACC	CTG	GTC	ATC	GTC	TCC	TCA
	A	F	G	A	G	D	G	W	D	L	V	W	G	Q	G	T	L	V	I	V	S	S
540042	ACA	TAT	GGG	GCG	GGA	GAT	GGA	TGG	AAT	CTT	GTC	TGG	GGC	CAG	GGA	ACT	CTG	GTC	ATC	GTC	TCC	GCC
	T	Y	G	A	G	D	G	W	N	L	V	W	G	Q	G	T	L	V	I	V	S	A

															IGHJ1*01						
						~~GCT~~	~~GAA~~	~~TAC~~	~~TTC~~	~~CAG~~	~~CA~~C	TGG	GGC	CAG	GGC	ACC	CTG	GTC	ACC	GTC	TCC	TCA
						~~A~~	~~E~~	~~Y~~	~~F~~	~~Q~~	~~H~~	W	G	Q	G	T	L	V	T	V	S	S
533111	GCC	TTT	GGG	GCG	GGA	GAT	GGT	TGG	GAT	CTT	GTC	TGG	GGC	CAG	GGA	ACC	CTG	GTC	ATC	GTC	TCC	TCA
	A	F	G	A	G	D	G	W	D	L	V	W	G	Q	G	T	L	V	I	V	S	S
341509	GCC	TTT	GGG	GCG	GGA	GAT	GGT	TGG	GAT	CTT	GTC	TGG	GGC	CAG	GGA	ACC	CTG	GTC	ATC	GTC	TCC	TCA
	A	F	G	A	G	D	G	W	D	L	V	W	G	Q	G	T	L	V	I	V	S	S
540042	ACA	TAT	GGG	GCG	GGA	GAT	GGA	TGG	AAT	CTT	GTC	TGG	GGC	CAG	GGA	ACT	CTG	GTC	ATC	GTC	TCC	GCC
	T	Y	G	A	G	D	G	W	N	L	V	W	G	Q	G	T	L	V	I	V	S	A

Fig. S5 (continued) (C) CD RH3 analysis of gVRC-H8$_{dC38}$ – gVRC-H10$_{dC38}$ (index number in **Fig. 3**), supporting that all 3 sequences belong to the same CDR H3 group. The nucleotide and amino acid sequences of CDR H3 and J region of the 3 heavy chains are aligned to the putative V, D and J germline genes. Putative nucleotide excisions are indicated with strikethrough lines. In blue are the putative TdT N additions in V-D and D-J junctions. In red are mutations from the putative germline genes and the TdT N additions. An alternative but unfavorable J gene (IGHJ1*01) is also indicated. Compared to the favorable IGHJ5*02 gene, IGHJ1*01 has one additional mutation. The correspondence between sequence indexes and names (gVRC-H1-10$_{dC38}$) for these 10 heavy chains can be found in **Fig. 3**.

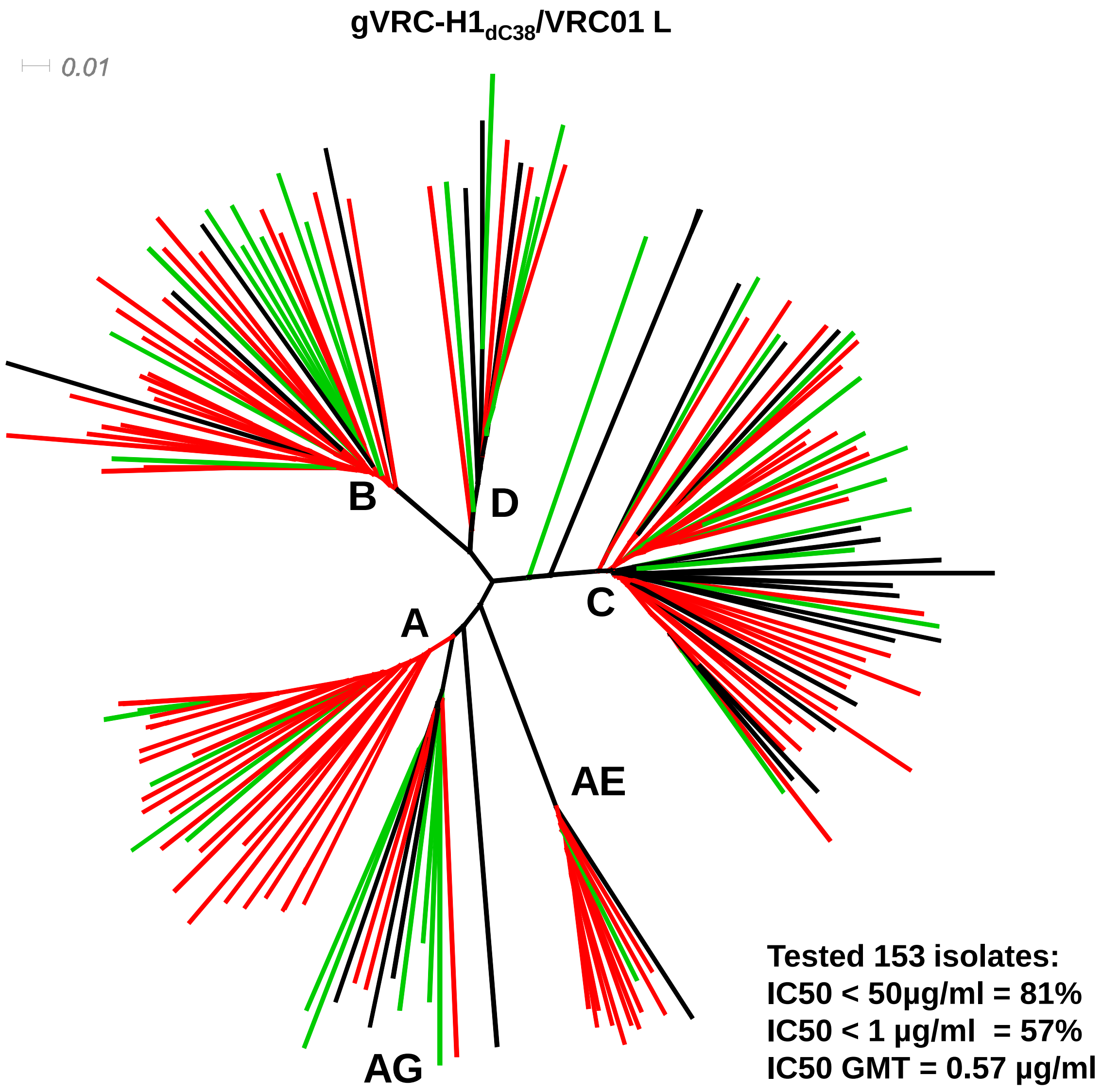


Fig. S6. Neutralization dendrogram. gVRC-H1$_{dC38}$/VRC01L was tested against genetically diverse Env-pseudoviruses representing the major HIV-1 clades. Neighbor-joining trees display the protein distance of gp160 sequences from 153 HIV-1 isolates tested against VRC-H1$_{dC38}$/VRC01L. A scale bar denotes 1% distance in amino acid sequence. Tree branches are colored by the neutralization potencies of gVRC-H1$_{dC38}$/ VRC01L against each particular virus.

Table S1. Information on study donors.

Donor	Gender, age, ethnicity	Clinical classification	Year of diagnosis	Sample time	CD4 count (cells/µl)	Plasma viral load (copies/ml)	Type of analysis
HIV-1 infected donors							
C38	Male, 54, African American	Slow progressor	1987	6/4/2008	323	46,800	Neutralization, 454 pyrosequencing
1	Male, 45, African American	Slow progressor	1985	11/8/2006	545	19,260	Neutralization, 454 pyrosequencing
N32	Male, 60,Caucasian	Slow progressor	1984	10/15/2007	663	7,531	Neutralization, 454 pyrosequencing
84	-	-	-	11/20/2007	-	-	454 pyrosequencing

"-" indicates "not applicable".

Table S2. PCR primers and DNA polymerase systems used to prepare donor C38 samples for 454 pyrosequencing.

Primer or data set	Primer sequence or DNA polymerase

a. Heavy chain sequencing

Primer	Primer sequence (5’ → 3’)
Forward H1 primers	
XLR-A_5’L-VH1	CCATCTCATCCCTGCGTGTCTCCGACTCAG ACAGGTGCCCACTCCCAGGTGCAG
XLR-A_5’L-VH1#2	CCATCTCATCCCTGCGTGTCTCCGACTCAG GCAGCCACAGGTGCCCACTCC
XLR-A_5’L-VH1-24	CCATCTCATCCCTGCGTGTCTCCGACTCAG CAGCAGCTACAGGCACCCACGC
XLR-A_5’L-VH1-69	CCATCTCATCCCTGCGTGTCTCCGACTCAG GGCAGCAGCTACAGGTGTCCAGTCC
Forward G1 primers	
XLR-A_VH1 LEADER-A	CCATCTCATCCCTGCGTGTCTCCGACTCAG ATGGACTGGACCTGGAGGAT
XLR-A_VH1 LEADER-B	CCATCTCATCCCTGCGTGTCTCCGACTCAG ATGGACTGGACCTGGAGCAT
XLR-A_VH1 LEADER-C	CCATCTCATCCCTGCGTGTCTCCGACTCAG ATGGACTGGACCTGGAGAAT
XLR-A_VH1 LEADER-D	CCATCTCATCCCTGCGTGTCTCCGACTCAG GGTTCCTCTTTGTGGTGGC
XLR-A_VH1 LEADER-E	CCATCTCATCCCTGCGTGTCTCCGACTCAG ATGGACTGGACCTGGAGGGT
XLR-A_VH1-LEADER-F	CCATCTCATCCCTGCGTGTCTCCGACTCAG ATGGACTGGATTTGGAGGAT
XLR-A_VH1-LEADER-G	CCATCTCATCCCTGCGTGTCTCCGACTCAG AGGTTCCTCTTTGTGGTGGCAG
Reverse primers	
XLR-B_3CγCH1#2	CCTATCCCCTGTGTGCCTTGGCAGTCTCAG GG GGA AGA CCG ATG GGC CCT TGG T
XLR-B_3CμCH1	CCTATCCCCTGTGTGCCTTGGCAGTCTCAG GGGAATTCTCACAGGAGACGA
Sequence data set	**DNA polymerase system**
H1 primers	Invitrogen Platinum *Taq*HiFi DNA polymerase
G1 primers	FinnzymesPhusionHiFi DNA polymerase

b. Light chain sequencing

Primer	Primer sequence (5’ → 3’)
Forward κ primers	
XLR-A_5'L-VK1/2	CCATCTCATCCCTGCGTGTCTCCGACTCAG ATGAGGSTCCCYGCTCAGCTCCTGGG
XLR-A_5'L-VK3	CCATCTCATCCCTGCGTGTCTCCGACTCAG CTCTTCCTCCTGCTACTCTGGCTCCCAG
XLR-A_5'L-VK4	CCATCTCATCCCTGCGTGTCTCCGACTCAG ATTTCTCTGTTGCTCTGGATCTCTG
Reverse κ primers	
XLR-B_3'CK1	CCTATCCCCTGTGTGCCTTGGCAGTCTCAG CAGCAGGCACACAACAGAGGCAGTTCC
Forward λ primers	
XLR-A_5'L-VL1/2	CCATCTCATCCCTGCGTGTCTCCGACTCAG GCACAGGGTCCTGGGCCCAGTCTG
XLR-A_5'L-VL3	CCATCTCATCCCTGCGTGTCTCCGACTCAG GCTCTGTGACCTCCTATGAGCTG
XLR-A_5'L-VL4/5	CCATCTCATCCCTGCGTGTCTCCGACTCAG GGTCTCTCTCSCAGCYTGTGCTG
XLR-A_5'L-VL6	CCATCTCATCCCTGCGTGTCTCCGACTCAG GTTCTTGGGCCAATTTTATGCTG
XLR-A_5'L-VL7/8	CCATCTCATCCCTGCGTGTCTCCGACTCAG GAGTGGATTCTCAGACTGTGGTG
XLR-A_5'L-VL1#2	CCATCTCATCCCTGCGTGTCTCCGACTCAG GCTCACTGCACAGGGTCCTGGGCC
XLR-A_5'L-VL3-1	CCATCTCATCCCTGCGTGTCTCCGACTCAG GCTTACTGCACAGGATCCGTGGCC
XLR-A_5'L-VL3-19	CCATCTCATCCCTGCGTGTCTCCGACTCAG ACTCTTTGCATAGGTTCTGTGGTT
XLR-A_5'L-VL3-21	CCATCTCATCCCTGCGTGTCTCCGACTCAG TCTCACTGCACAGGCTCTGTGACC
XLR-A_5'L-VL7-43	CCATCTCATCCCTGCGTGTCTCCGACTCAG ACTTGCTGCCCAGGGTCCAATTC
Reverse λ primers	
XLR-B_3'CL	CCTATCCCCTGTGTGCCTTGGCAGTCTCAG CACCAGTGTGGCCTTGTTGGCTTG

Table S3. Output information from pipeline processing of 454 pyrosequencing data sets obtained for donor C38.[a]

	Step 1[b]			Step 2[b]	Step 3[b]	
Chain (primers)	<Length>	N_{Total}	$Perc_{>400bp}$ (%)	N_{germ} (Germline)	CC	<$Perc_{Imp}$> (%)
H (IgVH1-H1)	414.8	460,706	90.0	138,523 (IgHV1-2)	0.69	18.7
H (IgVH1-G1)	422.8	574,027	89.5	168,365 (IgHV1-2)	0.70	14.5
κ (IgVK)[c]	401.1	257,910	82.6	48,347 (IgKV3-20)[d]	0.65	17.0
λ (IgVL)[c]	–	–	–	– (–)	0.71	19.0
κ (IgVK3)	401.2	448,125	83.3	90,764 (IgKV3-20)[d]	0.66	15.3

[a] The items listed in the table include antibody chain sequenced and germline family amplified by gene-specific primers, average read length, total number of reads, percentage of long reads (over 400bp), number of reads with the same germline origin as the confirmed VRC01-class antibodies from donor C38, the correlation coefficient between number of insertion/deletion errors in 454-derived sequence and quality improvement after error correction (P-value <0.001 throughout), and the averaged percent improvement of sequence quality after error correction.

[b] The steps from which the output information was obtained. Note that only results from the first three steps are presented in this table. The identities to known VRC01-class antibody chains from the same donor were calculated in step 4 and the CDR H3- or L3-specific analysis was carried out in step 5. The results from steps 4 and 5 were used to generate the identity/divergence plots and phylogenetic dendrograms. A detailed description of the computational pipeline, *Antibodyomics 1.0*, can be found in SI.

[c] In the first sequencing experiment of donor C38 light chains, since the germline gene of VRC01-class antibodies was unknown, we used primers to amplify both κ and λ germline V genes. After assigning germline V genes, the sequencing reads were divided into κ and λ data sets, which were processed separately.

[d] Due to the high similarity between IgKV3-11, IgKV3-20, IgKV3-NL1 and IgKV3-NL5, as well as the ambiguity caused by the high level of somatic hypermutations in VRC01-class light chains, sequences might have been misassgined. As a precaution of missing relevant light chains, sequences in these four germline families were combined in the bioinformatics analysis.

Table S4. CDR H3 groups identified from donor 74 454-pyrosequencing data derived from H1 primer amplification.[a]

Germline				
Nseq	Group[b]	Nseq (Perc %)	CDR H3	Most represented
(Nseq>20%)			length	CDR H3 sequence
	Group 1	2490 (39.9%)	14	QKFARGDQGWFFDL
IGHV1-2	Group 2	1424 (23.0%)	17	QKFESRYTGGQGWYFDL
111692	Group 3	908 (14.6%)	17	KTKGDVSGDGRGFFFDL
(6240)	Group 4	248 (4.0%)	16	QKFEKYTGGQGWYFDL
	Total	5070 (81.3%)		
IGHV1-8	Group 1	388 (69.7%)	12	GRGGSDQNHFDP
34454	Group 2	110 (19.7%)	15	GLRLSFGDLFNWFDP
(557)	Total	498 (89.4%)		
IGHV1-18	Group 1	40 (16.3%)	14	QKFARGDQGWFFDL
100086	Total	40 (16.3%)		
(246)				
IGHV1-46	Group 1	14 (34.1%)	23	EAELTYRFVGGEHDYDRSLWFDA
35478	Total	14 (34.1%)		
(41)				
	Group 1	489 (14.3%)	17	DLLIRGVTWKLQNHFDP
	Group 2	459 (13.4%)	14	GMGPTAAMENYFDS
IGHV1-69	Group 3	458 (13.4%)	17	DPLVRGRTWKMLHYFDS
295499	Group 4	335 (9.8%)	14	EVGRGVNDVSGMDV
(3415)	Group 5	331 (9.7%)	17	DPLVQGRTWKTLNYFDP
	Group 6	249 (7.3%)	17	DGRGYYDNSGYLGDFFH
	Total	2321 (68.0%)		

[a] The listed items include name of IGHV1 gene family together with total number of sequences (N_{seq}) and number of sequences with a divergence of 20% or greater (Nseq>20%), group index, number of sequences within a particular group (and corresponding percentage), length of the heavy chain complementarity-determining region 3 (CDR H3) and the most representative CDR H3 within the group.

[b] An iterative clustering procedure was utilized to define unique CDR H3 groups within an IGHV1 family. In each iteration, an all-to-all sequence comparison was carried out to determine the largest CDR H3 group using a 5-nucleotide cutoff, and then the remaining data set was subjected to the next iteration of clustering analysis. This procedure was terminated when a CDR H3 group with less than 300 sequences was identified.

Table S5. Lineage rank analysisof donor 74 454-pyrosequencing data derived from H1 primer amplification.[a]

GermlineNseq (Nseq>20%)	Lineage[b]	Nseq (Perc %)	CDR H3 length	Most represented CDR H3 sequence	<Divg>/ <Var> (%)
IGHV1-2	Lineage 1	2490 (39.9%)	14	QKFARGDQGWFFDL	30.0%/3.6%
111692	Lineage 2	1424 (23.0%)	17	QKFESRYTGGQGWYFDL	32.0%/3.6%
(6240)	Total	3914 (62.7%)			
IGHV1-8	—	—	—	—	—
34454	—	—	—	—	—
(557)	Total	0 (0.0%)			
IGHV1-18	—	—	—	—	—
100086	—	—	—	—	—
(246)	Total	0 (0.0%)			
IGHV1-46	—	—	—	—	—
35478	—	—	—	—	—
(41)	Total	0 (0.0%)			
IGHV1-69	Lineage 1	1279 (37.5%)	17	DLLIRGVTWKLQNHFDP	21.4%/7.8%
295499	Total	1279 (37.5%)			
(3415)					

[a] The listed items include name of VH1 gene family together with total number of sequences (N_{seq}) and number of sequences with a divergence of 20% or greater (Nseq>20%), lineage index, number of sequences within a particular lineage (and corresponding percentage), length of the heavy chain complementarity-determining region 3 (CDR H3) and the most representative CDR H3 within the lineage.

[b] The lineages were derived from the CDR H3 groups determined in Supplementary table X. An iterative clustering procedure was utilized to define unique CDR H3 groups within anIGHV1 family. In each iteration, an all-to-all sequence comparison was carried out to determine the largest CDR H3 group using a 5-nucleotide cutoff, and then the remaining data set was subjected to the next iteration of clustering analysis. This procedure was terminated when a CDR H3 group with less than 300 sequences was identified. Individual CDR H3 groups were merged into lineages when they shared visible similarity (amino acid composition and loop length) in CDR H3 region and the V gene variation was not significantly increased upon group merging.

Table S6. CDR H3 groups identified from donor C38 454-pyrosequencing data derived from H1 primer amplification.[a]

Germline Nseq (Nseq>20%)	Group[b]	Nseq (Perc %)	CDR H3 length	Most represented CDR H3 sequence
IGHV1-2 131108 (8750)	Group 1	4397 (50.3%)	17	ALTRTVSMNNGRAALDL
	Group 2	630 (7.2%)	9	DVWEKALDI
	Group 3	451 (5.2%)	9	DVQDGPLDL
	Group 4	326 (3.7%)	17	GLTRTLSLNNGRAAFDL
	Group 5	265 (3.0%)	11	EDYRMGAPFDY
	Total	6069 (69.4%)		
IGHV1-8 47678 (5138)	Group 1	2072 (40.3%)	18	AASRHCDKRRCYWGTQNL
	Group 2	886 (17.2%)	18	AASKHCGRRRCFWGTQNL
	Group 3	487 (9.5%)	17	GPISHRDYALRRSWLDP
	Group4	341 (6.6%)	14	GQGGGDVTWDGMDV
	Group5	288 (5.6%)	18	AAPRKCDKRRCYWGTENL
	Total	4074 (79.3%)		
IGHV1-18 67304 (3817)	Group 1	1146 (30.0%)	9	DQQRVNFDY
	Group 2	430 (11.3%)	12	GMEGQTMTAFDI
	Group 3	312 (8.2%)	13	GLQAPLETTPGGY
	Group 4	255 (6.7%)	9	DHLRLNFDY
	Total	2143 (56.1%)		
IGHV1-46 34565 (1057)	Group 1	343 (32.5%)	17	ALTRTLSMNNGRAALDL
	Group 2	184 (17.4%)	13	GHLRVFDWGHFDA
	Total	527 (49.9%)		
IGHV1-69 104848 (11290)	Group 1	2859 (25.3%)	14	GGGHVVVAVGTFDV
	Group 2	1571 (13.9%)	13	REKAYDNSGLFDY
	Group 3	1336 (11.8%)	14	VAGAYIVGKFYFQS
	Group 4	1195 (10.6%)	16	RPKYSHPSEAHLPFDY
	Group 5	358 (3.2%)	13	DGGTAATLYVFQS
	Group 6	286 (2.5%)	14	GGAHVVLSAGNFDV
	Total	7605 (67.4%)		

[a] The listed items include name of VH1 gene family together with total number of sequences (N_{seq}) and number of sequences with a divergence of 20% or greater (Nseq>20%), group index, number of sequences within a particular group (and corresponding percentage), length of the heavy chain complementarity-determining region 3 (CDR H3) and the most representative CDR H3 within the group.

[b] An iterative clustering procedure was utilized to define unique CDR H3 groups within anIGHV1 family. In each iteration, an all-to-all sequence comparison was carried out to determine the largest CDR H3 group using a 5-nucleotide cutoff, and then the remaining data set was subjected to the next iteration of clustering analysis. This procedure was terminated when a CDR H3 group with less than 300 sequences was identified.

Table S7. Lineage rank analysis of donor C38 454-pyrosequencing data derived from H1 primer amplification.[a]

GermlineNseq (Nseq>20%)	Lineage[b]	Nseq (Perc %)	CDR- H3 length	Most represented CDR-H3 sequence	<Divg>/ <Var> (%)
IGHV1-2	Lineage 1	4723 (54.0%)	17	ALTRTVSMNNGRAALDL	26.9%/6.2%
131108	Total	4723 (54.0%)			
(8750)					
IGHV1-8	Lineage 1	3246 (63.2%)	18	AASRHCDKRRCYWGTQNL	24.5%/8.3%
47678	Total	3246 (63.2%)			
(5138)					
IGHV1-18	Lineage 1	1401 (36.7%)	9	DQQRVNFDY	23.5%/10.3%
67304	Total	1401 (36.7%)			
(3817)					
IGHV1-46	—	—	—	—	—
34565	—	—	—	—	—
(1057)	Total	0 (0.0%)			
	Lineage 1	3145 (27.9%)	14	GGGHVVVAVGTFDV	21.2%/4.0%
IGHV1-69	Lineage 2	1571 (13.9%)	13	REKAYDNSGLFDY	22.0%/2.8%
104848	Lineage 3	1336 (11.8%)	14	VAGAYIVGKFYFQS	23.4%/0.6%
(11290)	Lineage 4	1195 (10.6%)	16	RPKYSHPSEAHLPFDY	22.6%/4.7%
	Total	7247 (64.2%)			

[a] The listed items include name of VH1 gene family together with total number of sequences (N_{seq}) and number of sequences with a divergence of 20% or greater (Nseq>20%), lineage index, number of sequences within a particular lineage (and corresponding percentage), length of the heavy chain complementarity-determining region 3 (CDR H3) and the most representative CDR H3 within the lineage.

[b] The lineages were derived from the CDR H3 groups determined in Supplementary table X. An iterative clustering procedure was utilized to define unique CDR H3 groups within anIGHV1 family. In each iteration, an all-to-all sequence comparison was carried out to determine the largest CDR H3 group using a 5-nucleotide cutoff, and then the remaining data set was subjected to the next iteration of clustering analysis. This procedure was terminated when a CDR H3 group with less than 300 sequences was identified. Individual CDR H3 groups were merged into lineages when they shared visible similarity (amino acid composition and loop length) in CDR H3 region and the V gene variation was not significantly increased upon group merging.

Table S8. CDR H3 groups identified from donor C38 454-pyrosequencing data derived from G1 primer amplification.[a]

Germline Nseq (Nseq>20%)	Lineage	Nseq(Perc %)	CDR H3 length	Most represented CDR H3 sequence
IGHV1-2 163108 (32317)	Lineage 1	9126 (28.2%)	17	ALTRTVSMNNGRAALDL
	Lineage 2	3470 (10.7%)	20	STSITLFGFIVGHYYYAMDV
	Lineage 3	2238 (6.9%)	20	TTAVTMFSMIVGHYYYAMDI
	Lineage 4	1747 (5.4%)	9	DVWEKALDI
	Lineage 5	1425 (4.4%)	11	EDYRMGAPFDY
	Lineage 6	1424 (4.4%)	9	DVQDGPLDL
	Lineage 7	1308 (4.1%)	17	GLTRTSSVNHGRAAFDL
	Lineage 8	1290 (4.0%)	12	ECEYDILSGCLM
	Lineage 9	932 (2.9%)	17	ALTRTVSVNHGRAALDL
	Lineage 10	656 (2.0%)	17	ALTRTISANHGRAAFDL
	Lineage 11	653 (2.0%)	13	GPVVGVGGDWFDP
	Lineage 12	459 (1.4%)	16	ALTRTLSMNNEGCLDL
	Lineage 13	402 (1.2%)	20	STSVTLFGFILGHYYYAMDI
	Lineage 14	366 (1.1%)	17	ALTRTISSNHGRAALDL
	Lineage 15	326 (1.0%)	9	DAQEGDLES
	Lineage 16	302 (0.9%)	20	TTAITMFGLIIAHHYYAMDV
	Lineage 17	265 (0.8%)	20	TTAITMFGLIVGHHYYAMDI
	Total	26389(81.7%)		6 non-redundant lineages
IGHV1-8 49151 (9907)	Lineage 1	3505(35.4%)	18	AASRHCDKRRCYWGTQNL
	Lineage 2	1833(18.5%)	14	GQGGGDVTWDGMDV
	Lineage 3	1206 (12.2%)	18	AASKHCGRRRCFWGTQNL
	Lineage 4	713 (7.2%)	17	GPISHRDYALRRSWLDP
	Lineage 5	501 (5.1%)	18	AAPRKCDKRRCYWGTENL
	Lineage 6	295 (3.0%)	13	ARGGDVTWDSLDV
	Total	8053 (81.3%)		4 non-redundant lineages
IGHV1-18 90097 (16836)	Lineage 1	7180 (42.6%)	14	RQYRGTSSGGWFDP
	Lineage 2	2268 (13.5%)	9	DQQRVNFDY
	Lineage 3	972 (5.8%)	12	GMEGQTMTAFDI
	Lineage 4	949 (5.6%)	14	RQYRGTSGAGWFDP
	Lineage 5	932 (5.5%)	16	IWPSSQNPVRLLEWSS
	Lineage 6	658 (3.9%)	16	IWPSPHPPVRFLEWSH
	Lineage 7	509 (3.0%)	9	DHLRLNFDY
	Lineage 8	337 (2.0%)	19	TAYLGYCSSFSCSTYYFDL
	Lineage 9	314 (1.9%)	19	TPYLGYCASLTCPTYYLDE
	Lineage 10	167 (1.0%)	13	GLQAPLETTPGGY
	Total	14286 (84.9%)		6 non-redundant lineages
IGHV1-46 71270 (24550)	Lineage 1	2897 (11.8%)	13	GHLRVFDWGHFDA
	Lineage 2	1873 (7.6%)	20	TTAVTMFGLIVGHHYYAMDV
	Lineage 3	1488 (6.1%)	13	GDGSHYYNTYMDV
	Lineage 4	1262 (5.1%)	20	TTAVTMFAMIVGHYYYAMDI
	Lineage 5	1136 (4.6%)	17	ALTRTVSTNQGRAALDL
	Lineage 6	1006(4.1%)	18	DQGVPDATSRSTLEYFQY
	Lineage 7	966 (3.9%)	20	SPGITMFGYIVGHRHFALDV
	Lineage 8	856 (3.5%)	20	TTAVTMFGLIIAHHYYAMDI
	Lineage 9	853 (3.5%)	13	GDGNWFYSFYMDV
	Lineage 10	786 (3.2%)	28	DVFERVPRPGAKSLGDVQTKEDGSYMDV
	Lineage 11	636 (2.6%)	20	SPAVTMFGLIVGHRHYALDV

Table S8. Continued.

	Lineage 12	619 (2.5%)	12	VFSWGEGTYFDR
	Lineage 13	574 (2.3%)	11	GVGPGTPPFDY
	Lineage 14	534 (2.2%)	20	TTAITMFSLIVGHYYYAMDV
	Lineage 15	475 (1.9%)	20	SPGVTFFGYIVGHRHRALDV
	Lineage 16	468 (1.9%)	20	STAVTLFGLIVGHYYYAMDV
	Lineage 17	453 (1.8%)	20	SPAITMFGYIVGHRYFALDV
	Lineage 18	400 (1.6%)	20	TTAVTMFGLILAHHYYAFDV
	Lineage 19	278 (1.1%)	20	TTAVTMFALIVGHYYYAMDV
	Total	17560 (71.5%)		9 non-redundant lineages
IGHV1-69 118368 (23545)	Lineage 1	5676 (24.1%)	14	GGGHVVVAVGTFDV
	Lineage 2	3835 (16.3%)	16	RPKYSHASEAHLPFDY
	Lineage 3	3266 (13.9%)	14	VAGAYIVGKFYFQS
	Lineage 4	3155 (13.4%)	13	REKAYDNSGLFDY
	Lineage 5	1392 (5.9%)	13	DGGTAATLYVFQS
	Lineage 6	471 (2.0%)	14	GGAHVVLSAGNFDV
	Lineage 7	371 (1.6%)	14	GGDMSWWPWGTFDV
	Lineage 8	323 (1.4%)	20	GRKASSGWFQAVVYHYPMDA
	Lineage 9	317 (1.3%)	13	ALGGGYHDVVAGH
	Lineage 10	295 (1.3%)	18	LGYSYGADGFLQTSNDDL
	Total	19101 (81.1%)		7 non-redundant lineages

[a] The listed items include name of VH1 gene family together with total number of sequences (N_{seq}) and number of sequences with a divergence of 20% or greater (Nseq>20%), group index, number of sequences within a particular group (and corresponding percentage), length of the heavy chain complementarity-determining region 3 (CDR H3) and the most representative CDR H3 within the group.

[b] An iterative clustering procedure was utilized to define unique CDR H3 groups within anIGHV1 family. In each iteration, an all-to-all sequence comparison was carried out to determine the largest CDR H3 group using a 5-nucleotide cutoff, and then the remaining data set was subjected to the next iteration of clustering analysis. This procedure was terminated when a CDR H3 group with less than 300 sequences was identified.

Table S9. Lineage rank analysis of donor C38 454-pyrosequencing data derived from G1 primer amplification.[a]

GermlineNseq (Nseq>20%)	Lineage[b]	Nseq (Perc %)	CDR-H3 length	Most represented CDR-H3 sequences[c]	<Divg>/ <Var> (%)
IGHV1-2 163108 (32317)	Lineage 1	12847 (39.8%)	17	ALTRTVSMNNGRAALDL	26.1%/9.2%
	Lineage 2	6677 (20.7%)	20	STSITLFGFIVGHYYYAMDV	24.6%/15.3%
	Lineage 3	1747 (5.4%)	9	DVWEKALDI	22.8%/7.9%
	Lineage 4	1425 (4.4%)	11	EDYRMGAPFDY	24.4%/5.7%
	Lineage 5	1424 (4.4%)	9	DVQDGPLDL	24.7%/13.5%
	Lineage 6	1290 (4.0%)	12	ECEYDILSGCLM	24.3%/7.5%
	Total	25736 (78.6%)			
IGHV1-8 49151 (9907)	Lineage 1	5212 (52.6%)	18	AASRHCDKRRCYWGTQNL	24.1%/7.7%
	Lineage 2	1833 (18.5%)	14	GQGGGDVTWDGMDV	30.3%/3.3%
	Total	7045 (71.1%)			
IGHV1-18 90097 (16836)	Lineage 1	8129 (48.3%)	14	RQYRGTSSGGWFDP	26.8%/12.3%
	Lineage 2	2777 (16.5%)	9	DQQRVNFDY	23.3%/9.9%
	Lineage 3	1590 (7.1%)	16	IWPSSQNPVRLLEWSS	31.6%/9.2%
	Total	12496 (74.2%)			
IGHV1-46 71270 (24550)	Lineage 1	5671 (23.1%)	20	TTAVTMFGLIVGHHYYAMDV	25.8%/14.7%
	Lineage 2	2897 (11.8%)	13	GHLRVFDWGHFDA	30.8%/8.5%
	Lineage 3	2530 (10.3%)	20	SPGITMFGYIVGHRHFALDV	22.6%/9.2%
	Lineage 4	1488 (6.1%)	13	GDGSHYYNTYMDV	22.6%/15.8%
	Lineage 5	1136 (4.6%)	17	ALTRTVSTNQGRAALDL	27.7%/5.2%
	Lineage 6	1006(4.1%)	18	DQGVPDATSRSTLEYFQY	23.6%/6.8%
	Total	14728 (60.0%)			
IGHV1-69 118368 (23545)	Lineage 1	6518 (27.7%)	14	GGGHVVVAVGTFDV	20.9%/3.6%
	Lineage 2	3835 (16.3%)	16	RPKYSHASEAHLPFDY	22.3%/3.7%
	Lineage 3	3266 (13.9%)	14	VAGAYIVGKFYFQS	23.2%/0.3%
	Lineage 4	3155 (13.4%)	13	REKAYDNSGLFDY	21.8%/2.6%
	Lineage 5	1392 (5.9%)	13	DGGTAATLYVFQS	23.1%/5.5%
	Total	18166 (77.2%)			

[a] The listed items include name of VH1 gene family together with total number of sequences (N_{seq}) and number of sequences with a divergence of 20% or greater (Nseq>20%), lineage index, number of sequences within a particular lineage (and corresponding percentage), length of the heavy chain complementarity-determining region 3 (CDR H3) and the most representative CDR H3 within the lineage.

[b] The lineages were derived from the CDR H3 groups determined in Supplementary table X. An iterative clustering procedure was utilized to define unique CDR H3 groups within anIGHV1 family. In each iteration, an all-to-all sequence comparison was carried out to determine the largest CDR H3 group using a 5-nucleotide cutoff, and then the remaining data set was subjected to the next iteration of clustering analysis. This procedure was terminated when a CDR H3 group with less than 300 sequences was identified. Individual CDR H3 groups were merged into lineages when they shared visible similarity (amino acid composition and loop length) in CDR H3 region and the V gene variation was not significantly increased upon group merging.

Table S10. Expression of antibodies with selected donor C38 heavy chains paired with VRC01 light chain

No.	Donor	Sequence ID	Yield* (mg/L culture sup)	Neutralization#	Amino acid sequence of heavy chain V domain
a. 2 selected sequences from cross-donor phylogenetic analysis of VH1-2 family from the H1-primer data set					
1	C38	304943	14.16	N	RVQLTQVWAQMRKPGASMRVSCETSGFRRFTDSKIGWVRQAPGQPFEWMGLMESYWGRVHYAAQFRDRVTMTRDVDVETAFLELSGLTLADTAIYYCVTAAGTNEWAFEWGQGTRVIVSP
2	C38	255552	18.48	Y	QVTLVQSGNQLKRPGASVRISCETSGYNFMDHFIHWVRQVPGHGPEWLGWVNPRGGGVNYSRKFQGRFSMTRDVYMETAYLDVTGLSPADTAVYYCARGFGGSDWSFLWGQGTLIIVSS
b. 10 selected sequences from cross-donor phylogenetic analysis of VH1-2 family from the G1-primer data set					
3	C38	540042	12.00	Y	QVHLVQSGTQVKKPGASVRVSCETSGFKFLDSIIHWFRQAPGEGLFWMGWIKPYTGSVNYVRRYQGRVSMTRDVYSDTAYMDLSGLNSDDTAVYFCTYGAGDGWNLVWGQGTLVIVSA
4	C38	533111	17.40	Y	RVHLVQSGTQVKKPGASVKVSCETSGFKFLDSLIHWVRQAPGQGLYWMGWIKPFRGSVNYDGYFRGRVSMTRDIYTDTAYMELSGLRSDDTAIYYCAFGAGDGWDLVWGQGTLVIVSS
5	C38	341509	15.24	Y	RVHLVQSGTQVKKPGASVKVSCETSGFKFLDSLIHWVRQAPGQGLYWMGWIKPYRGSVNYDGYFRGRVSMTRDIYTDTAYLELSGLRSDDTAIYYCAFGAGDGWDLVWGQGTLVIVSS
6	C38	286804	7.44	Y	QVSLVQSGNQLKKPGASVRISCETSGYNFLNHFIHWVRQVPGHGLEWLGWINPRGGGVNYSRNFQGKVSLTRNIDMETVYLDVRGLTPGDTAVYYCARGFGGSDWNFVWGQGTRITVSA
7	C38	155196	4.80	Y	QVRLVQSGNQVRKPGASVRISCEASGYKFIDHFIHWVRQVPGHGLEWLGWINPRGGGVNYSRGFQGKLSMTMTRDNFEETAYLDLSRLNPGDTAVYYCARGFAGYEWSFLWGQGTLVIVSS
8	C38	406425	3.00	Y	RINLDQSGSQVKKSGASVRISCETSGFKFMDSHLHWVRQVAGQPFEWMGWIFTSGGGVNYARQFQGRLTLTRDVFSETVFMDLSGLNAGDTGVYFCVKGTGGNEWGFVWGQGTVVVVSP
9	C38	128419	15.36	Y	RIELHQSGSQVKKSGASVRISCETSGFKFMDSHLHWVRQVAGQRFEWMGWIFTSGGGVNYARQFQGRLRLTRDVFSESVFMDLSGLNSGDTGVYYCVKGTGGNEWGFVWGQGSLVVVSP
10	C38	272066	10.80	Y	RINLHQSGSQVKRSGASVRISCETSGFKFMDSHLHWVRQVAGQGFEWMGWIFTSGGGVNYARQFQGRLTLTRDVFTDTVFMDLSGVNVGDTGVYYCVKGTGGNEWGFSWGQGTVVVVSP
11	C38	240171	15.36	N	QVRLIQSGTQMKKPGSSVKISCDTSGYKFVDFLIYWFRHVPGREIEWIGWLKPYGGGVNFNGNFRDRVTLTRKSDDTDRGTVYMEISGLRAADTAVYYCTRRGLCDHCSKWTFEHWGQGTPVIVSS
12	C38	534056	21.00	Y	QALVQSGSQMKKPGDSVRLSCQTSDSAITKYFIHWIRQAPGKGLEWIAWISPYGGRVNYGWQVRDRATLTRNIHMETVYMDLRGLRPDDTATYYCAMRDYCRDDNCNRWDLGHWGQGSLIVVSA
c. 35 selected sequences from 22 lineages in the lineage rank analysis of G1-primer data set					
13	C38	100588	17.40	N	QVQLLQSGSEIKRPGTSVTMSCTASGYNFNNYFINWVRQAPGQGPEWMGWISPSTWGTSLAPRFHGRVALTRATASNTIYLFLSNLRPDDTAMYFCARALTRTVSMNNGRAALDLWGQGTDLTVST
14	C38	100096	15.00	N	QVRLVQSGPEVKRPGASVTVSCTASGYNLNNYFINWIRQAPGQGPEWMGWINPVNGNTSYAQKFNDRVALTRATSFDRIYLFVSRLRPDDTAVYFCARGLTRTSSVNHGRAAFDLWGQGTLLSVAA

15	C38	102341	17.76	N	QERLTQSGSKLARPGTSLKMSCKASGYTFTSYYVHWVRQAPGHALEWLGEINPRSGGTNYAQKFQGRVAMTSDTSINTVYLDLKGLTSDDTAIYYCSRSTSITLFGFIVGHYYYAMDVWGQGTAVVVSS
16	C38	101970	12.60	N	QGHLVQSESAVTKPGASVRLACATSGYSFTSYYLHWVRQAPGQHFEWMGEINPLTGGTTYAQTFQGRVVMTTDTSTNTVYLDLKRVSVDDTAVYFCTRTTAVTMFSMIVGHYYYAMDIWGQGTTLTVSP
17	C38	100669	10.56	N	QVQLQQSAAEVKKPGTSVRVSCGTSSFSVNPIYVNWVRLVPGQGLQWIGWIKPETGATKYAQKFQDRVTMTTNTSISTAYMELGGLRYDDTAIYYCASDVWEKALDIWGPGTILTVSS
18	C38	104123	7.32	N	QVHLEQSGTEVKKPGTSVRVSCRASSSIINHMYVNWVRLVPREGLEWMGWINPKSGATNSARRLHDRVTMTTNTSINTVYMELRGLLYDDTGLYFCASDLWEKALDVWGPGTVVTVSS
19	C38	104936	10.20	N	QVQLLQSESEMREPGSSLRLSCQASNYSFSAYYIHWLRQVPGQGLEWIGYINPHNGDTNYAQDFQGRVTLTTDTSINTAYMEFKKMTFDDTAIYLCAREDYRMGAPFDYWGQGTLVAISS
20	C38	100996	17.52	N	QAQLKQSGADVKTPGASIKLSCKASGYSFVRHYVHWLRQVPGKRLEWMGWISPQNGGTFFGHNFRGRVAMSRDMSTSTFHLHLFNLTFDDTARYFCARDVQDGPLDLWGQGSLVTVSS
21	C38	101323	12.12	N	QPQLKQSGADLKRPGASMNVSCTASGYSFVRFYVHWVRQVPGKRLEWMGWINPQNGGTLYGQNFRGRVALSRDMSTSTVSLQLFNLTSDDTALYYCARDVQDGPLDLWGQGSLVTVSS
22	C38	100261	9.24	N	QAELAQSGAELRKTGTSIKVSCKASGYVFTGHYIHWVRQAPGEGLTWMGWINPNTGVAKYSPNFEGHVIFTRDTSINTAYVEFASLTVDDTAIYYCARECEYDILSGCLMWGPGSLVTVSS
23	C38	105510	12.48	N	QLQLAQSGAEVTKTGTTVKVSCKTNEYAFTGHYIHWVRQAPGQGLKWMGWINPNSGDTKYSSDFEGRTIITRDTSINTTYLELTGLTLDDTAMYFCAQECEYDVLSGCLIWGQGSLITVSS
24	C38	101517	No expression	n/a	QGQLVQSGAAVQEPGATLRVSCRAFGDTLRNSDVHWVRQAPGQGLQWMGWMNPSSGNTGFQYKFEDRIIMTWDTSTMTAHLEMTRLTSADTALYFCARAASRHCDKRRCYWGTQNLWGQGIHVTVSS
25	C38	101769	12.60	N	QGQLVQSEAAVEEPGATLSVSCRASGDTLTNHDVHWVRQAPGQGLEFMGWMNPSSGNAGIQHKFQGKISMTWDTFTATAYLSVTRLTSADTAVYFCARAASKHCGRRRCFWGTQNLWGQGTKVTVLS
26	C38	100309	17.88	N	QVELLQSGPEVKTPGDSVNVSCQAVGYKFIKHDIVWVRQAPGQGLEWMGWLNPETGKKGFAEKFQDRMMTSRDVSRAMVILTLSHVSFDDTAVYFCARGQGGGDVTWDGMDVWGRGTTVAVKT
27	C38	105014	16.56	N	QVQLVQSGGEVKKPGASVKVSCKASGYPFLNHDVTWVRWSPGQGLEWMAWINVHDDTTKFAERFQGRISLTADTSTATVYLELRSLTSDDTGIYFCGRRQYRGTSSGGWFDPWGQGSLIIVSS
28	C38	10107	16.68	N	QVQLVQSGGEVKRPGASVKVSCKASGYPFLNHDVTWVRQSPGQGLEWMAWINVHGDTTKYADRFQDRITLTADTSTATVYLELRSLTSDDTGIYFCARRQYRGTSSGGWFDPWGQGSLIIVSS
29	C38	102116	18.60	N	QIQLVQSGLEVRKPGASVKISCKATGFIFTSVGYSWVRQAPGKGFEWIGWVNPYNGVRVPTQRLQDRLTLAADTSTNTVYMELKNLRPDDTAIYYCARDQQRVNFDYWGQGTLVTVTS
30	C38	111905	12.48	N	QIQLIQSGPEVKKPGDSVKLSCQSSGFIFTSVGYSWVRQAPGQPFEWIAWVNPYNGVARPAQKLEDRMFLAKDASTSTVYMELRDLRPDDTAVYYCARDHLRLNFDYWGQGTQVTVSS
31	C38	105534	8.40	N	EIELLQSGPEMREPGASIRVSCKTSGYEFDKYSITWVRQTPEKGLEWVGWAGVPDGKLQYSNSEGRVTMTIDKFTKTAYLELKNLRVDDTATYFCVKIWPSSQNPVRLLEWSSWGQGTQIIVSS
32	C38	110795	13.80	N	QIQLLQSGPEMREPGTSIRISCKTSNYDFDKYSITWVRQAPGKGLEWVGWAGLSDGKMKYSQNSQGRVTMTIDKFTKTAFLDLKNLRPEDTAIYYCVRIWPSPHPPVRFLEWSHWGQGTQVTVAT
33	C38	102146	17.52	N	REQLLQSGTQVSLKPGASVKLSCKASGYNFNSYYVHWVRQAPGQGFEWMGEINPFTGGTTYAEKFHDRVAMTSDTSTNTVSVDLSRLTVDDTAVYYCTRTTAVTMFGLIVGHHYYAMDVWGQGTTVIVSA
34	C38	102542	12.36	N	QVQLVQSGAAVRKPGASVQLACTTSGYTFTSYYLHWVRQVPGHRFEWIGEINPLTGGTTLAQAFKERVVMTSDTSAHTVYLDLRSLSLDDTAVYYCTRTTAVTMFAMIVGHYYYAMDIWGRGTPLIVSP
35	C38	101345	10.80	N	QVQLLQSGPEMKNPGTTLEVSCRTSGFIFDDFFLHWVRQTPDHRLQYLGSISPFGGAVDVARRFEGRVSVTRDVSTSTLHMEMRHLRYEDTATYFCARGHLRVFDWGHFDAWGQGTLITVSA

36	C38	101143	12.60	N	QVQLVQSGPEKKNPGASVKVSCQISGFLLGDFFLHWLRQEPDQRLLYVGGMNPFNGRAKVARTFESRVSVTRDMSTSTFHMEMVDLRYGDTATYFCARGHYRVLDWGPLDEWGQGTLITVSA
37	C38	101794	19.20	N	RVQFVQSGTEMKKPGASVTVSCKTSGFSFTSFYLHWARLVPGRGLEWLGEINPNNGATTYAQAFQDRLSLTADKSSNTVSMDLGRLTVDDTAVYYCTRSPGITMFGYIVGHRHFALDVWGQGTTVIVSA
38	C38	101685	20.52	N	RVEFLQSGAEVRKPGTSVKVSCRASGYTFTSYYLHWVRQAPGQGLEWMGEINPNTGGTTYAQRFRDRVSMTADTSTNTGFLDLSRLTVDDTAVYYCTRSPAVTMFGLIVGHRHYALDVWGPGTAVTVSS
39	C38	109675	19.20	N	EVLLVQSGAEVKTPGASLKVSCQGIGYIFTSNYVHWVRQAPGQGLQWMAFINPSDGKVNYAETFKGRLSVTRDTSINSVYMDLRSLTFEDTATYYCARGDGSHYYNTYMDVWGKGTTVTVSP
40	C38	100794	18.00	N	LVQLDQSGLEVKKPGASVRISCKTSGYIFTTHYIHWVRQAPGQGLVWMGFINPDGGSANFTQSFQGRATVTRDVPRNMVLLELKNLQIDDTATYYCARGDGNHYYNYYMDVWGKGTTVAVSP
41	C38	101019	18.12	N	QGQLLQSGTEIKRPGASVTVSCTASGYNFNSYFINWVRQAPGQGPEWMGWINPNTWNTTLAPKFHGRVALTRATTSNIIYLYLSFLRPDDTAVYFCARALTRTVSTNQGRAALDLWGQGTVLIVST
42	C38	152185	3.00	N	QIQLVQSGAELKQAGTSVSISCKTSGYTFTAFFIHWLRRAPGQGLEWMGIINPSGGTATYSPGFQGRLLMTRDSSRNIVYMDLTNLTPQDTAVYFCARDQGVPDATSRSTLEYFQYWGQGTLISVFP
43	C38	100169	18.24	N	QVQLMQSGPEVKRPGSSVRVSCQASGVKLKFHAISWIRQAPGQGLEWLGSILPALASTKYGRNFQGRVSFTADTSQNTVYMELTNLKPDDTAVFYCARGGGHVVVAVGTFDVWGQGTSVTVSP
44	C38	100767	No expression	n/a	QVQLVQSGTEVKKPGSSVKLSCRASGGTFSHYAIHWVRQAPRHGLEWMGGIIFGFNVANNAETFQGRVALTADVSTSTASLTLSSLTLADTALYYCVLRPKYSHASEAHLPFDYWGLGTLVTVAS
45	C38	100163	16.20	N	QERLVQSGAELAKPGSSVRVSCQASGDSFSSDPIAWVRQSPAGGLAWIGAHIPVFDMSRYSQRFQGRVTFTADRSSRTAYMELSHVQSEDTAVYYCAKVAGAYIVGKFYFQSWGQGTLVSVSS
46	C38	101382	18.96	N	QVTVMQSEAVVKKPGSSVRVSCKSAGGSFSHYAMNWVRQAPGQGFEWMGGIIPAFGVVNYAQRFQGRITISANKETSTDFLDLRSLTSGDTAVYYCARREKAYDNSGLFDYWGQGTLVTVSS
47	C38	103129	12.48	N	QVHLLQSGAEVKKPGSSVRLSCALAGGTFGKSGIFWLRQTPMRGLEWMGTLIPMLGTPNYAQPFIGRLTIDADKSTNTAHMELRGLTLDDTAIYYCARDGGTAATLYVFQSWGPGTLVTVSA

d. 3 selected sequences from cross-donor phylogenetic analysis of VH1-46 family from the G1-primer data set using 13 VH1-2 originated founder VRC01-like antibodies

48	C38	249368	8.52	N	HDQLTQTETAVARPGATLNVSCATSGYTFVDYQIHWVRQAPGQGLQWMGLINPADGTTVFSSLFQGRLDLARDMSTKTVYMKLSRLTSDDSATYYCCKVFSWGEGTYFDRWGRGTHAVVSS
49	C38	237270	9.96	N	HDQLTQTETAVTRPGATLNVSCATSGYTFVDYQIHWVRQAPGQGLQWMGLINPSDGTTVFSSLFQGRLDLARDMSTKTIYMILSRLTSDDSATYYCCKVFSWGEGTYFDRWGRGTHAVVSS
50	C38	513421	13.32	N	QVTLLQSGTEVRKTGASVTISCKTSGYNFENYFIHWVRHSPRHGLDFLGTINPPSHRPSYADLLKGRLTLTSDTSTATVSMGLTNLTSDDAAVYYCVRDKIMTTFGDFIKSRYLQHWGQGTHIVVSS

*no expression,yield was less than 0.60mg/L.
#Y, yes; N, no; n/a, not available.

Table S11. Neutralization titers of 10 chimeric antibodies derived from 454 pyrosequencing of donor C38 against 20 HIV-1 clade A, B and C Env-pseudoviruses.[a]

	Neutralization IC50 titers (μg/ml)[b]										
Clade and virus	**VRC01**	**gVRC-H1$_{dC38}$ /VRC01L**	**gVRC-H2$_{dC38}$ /VRC01L**	**gVRC-H3$_{dC38}$ /VRC01L**	**gVRC-H4$_{dC38}$ /VRC01L**	**gVRC-H5$_{dC38}$ /VRC01L**	**gVRC-H6$_{dC38}$ /VRC01L**	**gVRC-H7$_{dC38}$ /VRC01L**	**gVRC-H8$_{dC38}$ /VRC01L**	**gVRC-H9$_{dC38}$ /VRC01L**	**gVRC-H10$_{dC38}$ /VRC01L**
Clade A (n =7)											
Q23.17	0.046	0.086	>50	0.033	0.082	0.062	0.052	0.082	0.107	0.051	0.041
Q842.d12	0.019	0.031	13.3	0.016	0.032	0.028	0.025	0.046	0.071	0.025	0.025
UG037.8	0.082	0.070	>50	0.050	0.085	0.054	0.052	0.090	0.175	0.089	0.088
Q168.a2	0.129	0.204	>50	0.119	0.426	0.287	0.128	0.223	0.453	0.279	0.450
KER2018.11	0.675	1.2	>50	0.491	0.812	0.692	0.612	0.490	1.8	1.1	1.3
KER2008.12	1.0	0.206	4.6	3.9	8.7	4.0	5.9	3.7	3.5	0.917	1.3
Clade B (n=7)											
Yu2	0.091	0.077	>50	0.061	0.168	0.161	0.087	0.172	0.294	0.050	0.063
JR-FL	0.029	0.026	>50	0.007	0.026	0.020	0.009	0.015	0.054	0.006	0.007
7165.18	>50	>50	>50	13.5	>50	28.7	>50	>50	>50	30.1	24.0
BG1168.01	0.392	1.2	>50	0.126	0.504	0.195	>50	2.4	0.909	0.456	0.600
JR-CSF	0.238	0.044	>50	0.066	0.121	0.097	0.051	0.042	0.111	0.040	0.057
PVO.04	0.435	0.178	>50	0.064	0.488	0.214	0.073	0.128	0.294	0.141	0.277
TRO.11	0.381	0.177	>50	0.043	0.143	0.075	0.077	0.158	0.388	0.225	0.302
CAAN.A2	1.3	1.2	10.0	2.0	6.4	3.1	1.9	2.2	4.6	1.1	1.5
Clade C (n=6)											
DU156.12	0.109	0.109	3.3	0.030	0.214	0.118	0.129	0.176	0.238	0.156	0.158
ZM109.4	0.102	0.159	>50	0.809	4.73	9.9	0.520	0.121	0.102	0.058	0.053
ZM106.9	0.268	0.156	>50	0.080	0.443	0.163	0.250	0.199	0.478	0.300	0.309
ZM176.66	0.029	>50	>50	>50	>50	>50	>50	0.048	0.097	0.031	0.025
SO18.18	0.058	0.110	>50	0.052	0.279	0.156	0.072	0.082	0.250	0.044	0.074
TV1.29	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50
Breadth (n = 20)	90%	85%	20%	90%	85%	90%	80%	90%	90%	95%	95%
GMT[c]	0.151	0.151	6.7	0.132	0.338	0.288	0.137	0.177	0.314	0.159	0.187
MuLV[d]	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50	>50

[a] The chimeric antibodies were expressed using the 10 heavy chains derived from donor C38 and the VRC01 light chain. The wild-type mAb VRC01 was included as a control.

[b] The IC$_{50}$ values < 1 μg/ml are highlighted in red; values 1 – 50 μg/ml are in green.

[c] GMT stands for geometric mean titer, which was calculated for neutralization sensitive viruses with an IC$_{50}$ value < 50 μg/ml.

[d] MuLV stands for murine leukemia virus, which was included as a negative control.

Table S12. Neutralization values (μg/ml) of antibody gVRC-H1$_{dC38}$/VRC01L against 153 HIV-1 Env-pesudoviruses.

Virus	Clade	IC_{50}	IC_{80}	Virus	Clade	IC_{50}	IC_{80}	Virus	Clade	IC_{50}	IC_{80}
0260.v5.c36	A	1.440	4.75	3988.25	B	>50	>50	25925-2.22	C	0.821	2.84
0330.v4.c3	A	0.151	0.715	5768.04	B	1.51	6.57	26191-2.48	C	0.400	1.53
0439.v5.c1	A	0.781	2.89	6101.10	B	0.211	1.1	3637_V5_C3	C	10.9	>50
3365.v2.c20	A	0.121	0.468	6535.3	B	2.87	23.4	3873_V1_C24	C	>50	>50
3415.v1.c1	A	0.533	1.61	7165.18	B	>50	>50	6322_V4_C1	C	>50	>50
3718.v3.c11	A	8.570	>50	89.6.DG	B	0.458	1.64	6471_V1_C16	C	>50	>50
398-F1_F6_20	A	0.616	2.42	AC10.29	B	2.02	8.12	6631_V3_C10	C	>50	>50
BB201.B42	A	0.832	3.36	ADA	B	0.870	3.05	6644_V2_C33	C	0.378	2.08
BB539.2B13	A	15.3	>50	BaL.01	B	0.129	0.483	BR025.9	C	>50	>50
BI369.9A	A	0.199	0.961	BG1168.01	B	7.36	>50	CAP210.E8	C	>50	>50
BS208.B1	A	0.037	0.204	BL01.DG	B	>50	>50	CAP244.D3	C	0.622	3.14
KER2008.12	A	0.957	3.31	BR07.DG	B	1.82	7.63	CAP45.G3	C	7.63	>50
KER2018.11	A	3.550	16.2	CAAN.A2	B	4.35	13.7	CNE30	C	2.35	7.69
KNH1209.18	A	0.382	1.12	HO86.8	B	>50	>50	CNE53	C	0.198	0.812
MB201.A1	A	0.510	2.25	HT593.1	B	1.10	3.94	DU123.06	C	>50	>50
MB539.2B7	A	1.020	3.55	HXB2	B	0.099	0.429	DU151.02	C	0.934	5.03
MI369.A5	A	0.208	0.971	JR-CSF	B	0.320	1.57	DU156.12	C	0.249	1.08
MS208.A1	A	0.184	0.84	JR-FL	B	0.026	0.132	DU172.17	C	>50	>50
Q168.a2	A	0.378	1.29	MN.3	B	0.068	>50	DU422.01	C	>50	>50
Q23.17	A	0.192	0.61	PVO.04	B	0.591	3.12	MW965.26	C	0.144	0.913
Q259.17	A	0.471	2.47	QH0515.01	B	0.726	5.44	SO18.18	C	0.274	1.09
Q461.e2	A	0.918	4.29	QH0692.42	B	3.27	11	TV1.29	C	>50	>50
Q769.h5	A	0.075	0.411	REJO.67	B	0.189	0.823	TZA125.17	C	>50	>50
Q842.d12	A	0.045	0.18	RHPA.7	B	0.098	0.568	TZBD.02	C	0.088	0.411
QH209.14M.A2	A	0.033	0.421	SC422.8	B	0.196	0.948	ZA012.29	C	0.422	1.7
RW020.2	A	0.359	1.46	SF162.LS	B	0.109	0.529	ZM106.9	C	0.371	1.16
UG037.8	A	0.226	0.715	SS1196.01	B	0.269	1.08	ZM109.4	C	0.361	2.59
620345.c1	AE	>50	>50	THRO.18	B	18.3	>50	ZM135.10a	C	0.598	2.71
C1080.c3	AE	1.250	14.3	TRJO.58	B	0.370	1.26	ZM176.66	C	>50	>50
C2101.c1	AE	0.335	1.4	TRO.11	B	0.710	2.07	ZM197.7	C	1.41	4.36
C3347.c11	AE	0.140	0.8	WITO.33	B	0.306	1.25	ZM214.15	C	3.16	13.6
C4118.09	AE	0.137	0.918	Yu2	B	0.121	0.783	ZM215.8	C	0.113	0.924
CNE59	AE	0.659	5.14	CNE10	B'	0.584	2.39	ZM233.6	C	>50	>50
M02138	AE	0.561	1.95	CNE14	B'	0.196	0.68	ZM249.1	C	0.349	1.54
R1166.c1	AE	0.612	2.1	CNE4	B'	0.461	3.39	ZM53.12	C	2.53	10.1
R2184.c4	AE	0.077	0.383	CNE57	B'	0.297	1.02	ZM55.28a	C	0.359	1.21
R3265.c6	AE	0.201	4.71	CH038.12	BC	39.0	>50	231965.c1	D	0.840	4.68
TH966.8	AE	0.629	2.42	CH070.1	BC	>50	>50	247-23	D	>50	>50
TH976.17	AE	0.271	1.39	CH117.4	BC	0.124	9.16	3016.v5.c45	D	>50	>50
235-47	AG	0.102	0.499	CH181.12	BC	0.447	1.67	57128.vrc15	D	>50	>50
242-14	AG	>50	>50	CNE40	B'C	0.333	1.84	6405.v4.c34	D	1.80	5
T250-4	AG	>50	>50	286.36	C	1.21	4.72	A03349M1.vrc4a	D	6.73	36.2
T251-18	AG	2.31	8.49	288.38	C	0.224	2.66	NKU3006.ec1	D	0.930	3.22
T255-34	AG	1.35	15.6	0013095-2.11	C	1.270	10.2	UG021.16	D	0.807	5.94
T257-31	AG	4.96	14.9	001428-2.42	C	0.046	0.175	UG024.2	D	12.7	>50
263-8	AG	0.396	1.33	0077_V1_C16	C	>50	>50	X2088_c9	G	>50	>50
T266-60	AG	0.430	1.63	16055-2.3	C	0.327	1.03	6540.v4.c1	AC	>50	>50
271-11	AG	1.350	>50	16845-2.22	C	9.75	>50	6545.v3.c13	AC	>50	>50
T278-50	AG	>50	>50	16936-2.21	C	0.220	1.1	6095_V1_C10	ACD	1.51	8.12
928-28	AG	2.46	7.8	25710-2.43	C	0.513	2.13	3337_V2_C6	CD	0.205	10.6
DJ263.8	AG	1.15	12.7	25711-2.4	C	1.71	6.69	3817.v2.c59	CD	15.7	>50

Table S13. Sequence identity of neutralizing heavy chains and light chains identified from 454 pyrosequencing data of donor C38 to the corresponding chains of 13 template VRC01-like antibodies.[a]

1. Heavy chain sequence identity

				NIH45-	VRC-	VRC-	VRC-	VRC-	VRC-				
HC code[b]	VRC01	VRC02	VRC03	46	PG04	PG04b	CH31	CH32	CH33	3BNC60	3BNC117	12A12	12A21
H1	62.8/57.0	63.9/57.9	62.1/50.8	61.9/53.6	65.1/55.2	65.3/56.0	59.0/48.9	58.8/49.6	59.3/45.0	62.6/52.8	63.1/52.8	66.4/55.7	66.7/56.6
H2	62.5/45.5	62.5/45.5	59.7/42.3	61.6/44.8	62.1/49.6	62.4/50.4	60.1/42.7	57.5/42.7	57.3/42.0	59.3/46.3	60.2/48.0	64.2/55.7	64.5/56.6
H3	59.5/48.8	59.5/50.4	59.5/42.3	59.5/47.2	58.4/48.8	58.7/48.0	61.1/38.9	59.5/38.2	59.3/40.5	57.5/47.2	60.7/48.0	64.2/54.9	65.0/56.6
H4	61.2/51.2	60.6/52.1	59.2/44.6	59.5/48.0	58.9/50.4	59.2/49.6	60.8/39.7	60.3/38.9	59.3/41.2	59.3/47.2	59.9/48.0	65.8/54.9	64.8/55.7
H5	61.4/49.6	60.9/50.4	59.7/44.6	58.4/45.6	58.1/48.8	59.5/48.0	59.5/38.9	58.0/38.2	57.3/40.5	58.3/47.2	59.9/46.3	65.0/52.5	64.5/54.9
H6	61.7/52.1	62.8/54.5	62.1/49.2	61.1/52.0	64.5/56.0	64.8/57.6	59.5/48.9	59.8/49.6	60.6/46.6	62.3/56.1	61.8/56.1	64.8/54.1	64.8/57.4
H7	62.8/58.7	62.3/59.5	62.3/50.0	61.1/54.4	65.3/53.6	65.1/54.4	58.3/51.1	58.8/51.1	59.0/47.3	63.1/56.1	63.1/56.9	68.3/58.2	68.3/59.0
H8	63.9/58.7	62.5/57.0	56.7/50.0	61.9/57.6	63.5/55.2	64.3/54.4	59.8/44.3	59.0/45.0	60.6/48.1	61.8/58.5	63.7/58.5	68.9/66.4	68.6/63.9
H9	64.2/53.7	64.2/54.5	61.3/50.0	63.5/53.6	65.9/56.0	66.1/56.8	63.9/47.3	62.3/48.9	63.1/48.9	64.5/59.3	66.7/57.7	69.4/64.8	71.3/63.9
H10	64.2/54.5	64.2/55.4	60.8/49.2	63.5/54.4	65.3/55.2	65.6/56.0	63.6/48.1	62.1/49.6	63.1/49.6	64.2/58.5	66.4/56.9	68.9/63.9	71.6/64.8

2. Light chain sequence identity

				NIH45-	VRC-	VRC-	VRC-	VRC-	VRC-				
LC code[b]	VRC01	VRC02	VRC03	46	PG04	PG04b	CH31	CH32	CH33	3BNC60	3BNC117	12A12	12A21
L1	72.9/58.4	72.3/57.4	72.9/56.9	71.9/59.4	70.9/55.9	71.9/55.4	59.2/44.7	59.5/44.7	59.9/44.7	61.6/45.5	61.6/46.5	60.8/47.6	58.6/42.7
L2	75.9/64.4	74.6/63.4	74.5/63.7	76.2/67.3	73.5/66.7	74.9/67.3	64.4/50.5	64.7/48.5	65.4/50.5	63.3/46.5	65.0/47.5	63.8/53.4	62.8/50.5
L3	74.3/64.4	72.9/63.4	73.2/64.7	72.6/63.4	70.9/56.9	71.9/56.4	58.9/43.7	58.9/43.7	57.9/43.7	61.3/44.4	61.3/45.5	64.1/52.4	59.9/45.6
L4	72.6/52.5	71.6/50.5	69.9/54.9	72.6/52.5	70.3/51.0	71.6/50.5	62.1/48.5	62.5/48.5	61.8/47.6	63.3/45.5	64.3/45.5	61.5/45.6	60.8/46.6
L5	69.6/56.4	69.0/55.4	70.3/58.8	69.0/56.4	65.4/49.0	66.3/48.5	68.0/56.3	66.3/56.3	66.7/56.3	66.7/56.6	67.0/57.6	71.5/67.0	67.6/59.2
L6	76.2/67.3	75.6/66.3	75.2/64.7	76.6/67.3	74.5/67.6	75.6/67.3	64.4/51.5	65.4/50.5	65.7/51.5	64.0/47.5	64.3/47.5	66.3/52.4	63.8/48.5

[a] Sequence identity was calculated at both nucleotide level and amino-acid level, and both values are shown for each pair of antibody chains, separated by "/".

[b] Chain code is abbreviation for the nomeclature of neutralizing antibody chains identified from 454 pyrosequencing data of donor C38, with Hn for gVRC-Hn_{dC38} and Ln for gVRC-Ln_{dC38}.