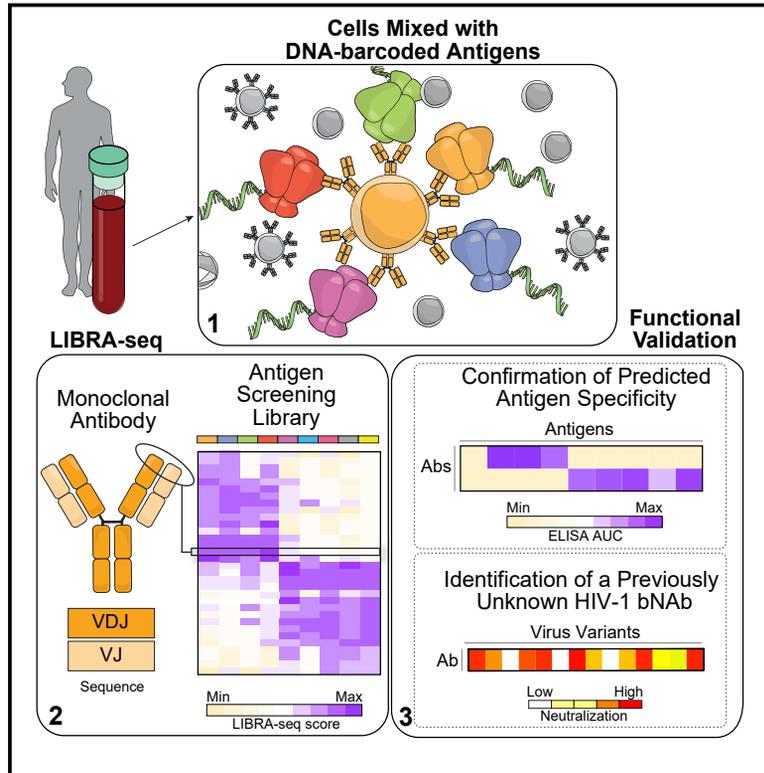


High-Throughput Mapping of B Cell Receptor Sequences to Antigen Specificity

Graphical Abstract



Authors

Ian Setliff, Andrea R. Shiakolas, Kelsey A. Pilewski, ..., Priyamvada Acharya, Lynn Morris, Ivelin S. Georgiev

Correspondence

ivelin.georgiev@vanderbilt.edu

In Brief

LIBRA-seq enables high-throughput mapping of B cell receptor sequence to antigen specificity at the single-cell level.

Highlights

- LIBRA-seq: high-throughput mapping of BCR sequence to antigen specificity
- Identified HIV- and influenza-specific B cells in two HIV-infected subjects
- Predicted antigen reactivity for thousands of single B cells
- Identified a previously unknown broadly neutralizing HIV antibody

High-Throughput Mapping of B Cell Receptor Sequences to Antigen Specificity

Ian Setliff,^{1,2,16} Andrea R. Shiakolas,^{1,3,16} Kelsey A. Pilewski,^{1,3} Aryn A. Murji,^{1,3} Rutendo E. Mapengo,⁴ Katarzyna Janowska,⁵ Simone Richardson,^{4,11} Charissa Oosthuysen,^{4,11} Nagarajan Raju,^{1,3} Larance Ronsard,⁷ Masaru Kanekiyo,⁸ Juliana S. Qin,¹ Kevin J. Kramer,^{1,3} Allison R. Greenplate,¹ Wyatt J. McDonnell,^{3,9,17} Barney S. Graham,⁸ Mark Connors,¹⁰ Daniel Lingwood,⁷ Priyamvada Acharya,^{5,6} Lynn Morris,^{4,11,12} and Ivelin S. Georgiev^{1,3,13,14,15,18,*}

¹Vanderbilt Vaccine Center, Vanderbilt University Medical Center, Nashville, TN 37232, USA

²Program in Chemical and Physical Biology, Vanderbilt University Medical Center, Nashville, TN 37232, USA

³Department of Pathology, Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, TN 37232, USA

⁴National Institute for Communicable Diseases of the National Health Laboratory Service, Johannesburg 2131, South Africa

⁵Division of Structural Biology, Duke Human Vaccine Institute, Duke University School of Medicine, Durham, NC 27710, USA

⁶Department of Surgery, Duke University School of Medicine, Durham, NC 27710, USA

⁷Ragon Institute of Massachusetts General Hospital, Harvard and Massachusetts Institute of Technology, Cambridge, MA 02139, USA

⁸Vaccine Research Center, National Institute of Allergy and Infectious Diseases, NIH, Bethesda, MD 20892, USA

⁹Vanderbilt Center for Translational and Clinical Immunology, Vanderbilt University Medical Center, Nashville, TN 37232, USA

¹⁰National Institute of Allergy and Infectious Diseases, NIH, Bethesda, MD 20892, USA

¹¹Faculty of Health Sciences, University of the Witwatersrand, Johannesburg 2000, South Africa

¹²Centre for the AIDS Programme of Research in South Africa (CAPRISA), University of KwaZulu-Natal, Durban 4041, South Africa

¹³Vanderbilt Institute for Infection, Immunology and Inflammation, Vanderbilt University Medical Center, Nashville, TN 37232, USA

¹⁴Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN 37232, USA

¹⁵Center for Structural Biology, Vanderbilt University, Nashville, TN 37232, USA

¹⁶These authors contributed equally

¹⁷Present address: 10X Genomics, Pleasanton, CA, USA

¹⁸Lead Contact

*Correspondence: ivelin.georgiev@vanderbilt.edu

<https://doi.org/10.1016/j.cell.2019.11.003>

SUMMARY

B cell receptor (BCR) sequencing is a powerful tool for interrogating immune responses to infection and vaccination, but it provides limited information about the antigen specificity of the sequenced BCRs. Here, we present LIBRA-seq (linking B cell receptor to antigen specificity through sequencing), a technology for high-throughput mapping of paired heavy- and light-chain BCR sequences to their cognate antigen specificities. B cells are mixed with a panel of DNA-barcoded antigens so that both the antigen barcode(s) and BCR sequence are recovered via single-cell next-generation sequencing. Using LIBRA-seq, we mapped the antigen specificity of thousands of B cells from two HIV-infected subjects. The predicted specificities were confirmed for a number of HIV- and influenza-specific antibodies, including known and novel broadly neutralizing antibodies. LIBRA-seq will be an integral tool for antibody discovery and vaccine development efforts against a wide range of antigen targets.

INTRODUCTION

The antibody repertoire—the collection of antibodies present in an individual—responds efficiently to invading pathogens because of its exceptional diversity and ability to fine-tune antigen specificity via somatic hypermutation (Briney et al., 2019; Rajewsky, 1996; Soto et al., 2019). This antibody repertoire is a rich source of potential therapeutic agents, but its size makes it difficult to examine more than a small cross-section of the total repertoire (Brekke and Sandlie, 2003; Georgiou et al., 2014; Wang et al., 2018; Wilson and Andrews, 2012). Historically, a variety of approaches have been developed to characterize antigen-specific B cells in human infection and vaccination samples. The methods most frequently used include single-cell sorting with fluorescent antigen baits (Scheid et al., 2009; Wu et al., 2010), screens of immortalized B cells (Buchacher et al., 1994; Stiegler et al., 2001), and B cell culture (Bonsignori et al., 2018; Huang et al., 2014; Walker et al., 2009, 2011). However, these methods to couple functional screens with variable heavy (V_H) and variable light (V_L) immunoglobulin gene sequences are low throughput; generally, individual B cells can only be screened against a few antigens simultaneously.

Recent advances in next-generation sequencing (NGS) enable high-throughput interrogation of antibody repertoires at the

sequence level, including paired heavy and light chains (Busse et al., 2014; DeKosky et al., 2013; Tan et al., 2014). However, annotation of NGS antibody sequences for their cognate antigen partner(s) generally requires synthesis, production, and characterization of individual recombinant monoclonal antibodies (DeFalco et al., 2018; Setliff et al., 2018). Recent efforts to develop new antibody screening technologies have sought to overcome throughput limitations while still uniting antibody sequence and functional information. For example, natively paired human B cell receptor (BCR) heavy- and light-chain amplicons can be expressed and screened as Fab (Wang et al., 2018) or single-chain variable fragment (Adler et al., 2017a, 2017b) in a yeast display system. Although these various antibody discovery technologies have led to the identification of potentially neutralizing antibodies, they remain limited by the number of antigens against which single cells can simultaneously be screened efficiently.

Inspired by previous methods combining surface protein marker detection with single-cell RNA sequencing (Peterson et al., 2017; Stoeckius et al., 2017) and T cell epitope determination with T cell receptor sequence (Zhang et al., 2018), we developed LIBRA-seq (linking B cell receptor to antigen specificity through sequencing) to simultaneously recover both antigen specificity and paired heavy- and light-chain BCR sequence. LIBRA-seq is a NGS-based readout for BCR-antigen binding interactions that utilizes oligonucleotides (oligos) conjugated to recombinant antigens. Antigen barcodes are recovered during paired-chain BCR sequencing experiments and bioinformatically mapped to single cells. To demonstrate the utility of LIBRA-seq, we applied the method to peripheral blood mononuclear cell (PBMC) samples from two HIV-infected subjects, and from these, we successfully identified HIV- and influenza-specific antibodies, including both known and novel broadly neutralizing antibody (bNAb) lineages. LIBRA-seq is high-throughput, scalable, and applicable to many targets. This single integrated assay enables mapping of monoclonal antibody sequences to panels of diverse antigens, theoretically unlimited in number, and facilitates rapid identification of cross-reactive antibodies that may serve as therapeutic agents or vaccine templates.

RESULTS

LIBRA-seq Method and Validation

LIBRA-seq transforms antibody-antigen interactions into sequencing-detectable events by conjugating barcoded DNA oligos to each antigen in a screening library. All antigens are labeled with the same fluorophore, which enables sorting of antigen-positive B cells by fluorescence activated cell sorting (FACS) before encapsulation of single B cells via droplet microfluidics. Antigen barcodes and BCR transcripts are tagged with a common cell barcode from bead-delivered oligos, enabling direct mapping of BCR sequence to antigen specificity (Figure 1A).

To test the ability of LIBRA-seq to accurately unite BCR sequence and antigen specificity, we devised a proof-of-principle mapping experiment using two Ramos B cell lines with different BCR sequences and antigen specificities (Weaver

et al., 2016). These engineered B cell lines do not display endogenous BCR and, instead, express specific, user-defined surface immunoglobulin M (IgM) BCR sequences (Weaver et al., 2016). To that end, we chose two well-characterized BCRs: VRC01, a CD4-binding-site-directed HIV-1 bNAb (Wu et al., 2010), and Fe53, a bNAb recognizing the stem of group 1 influenza hemagglutinins (HAs) (Lingwood et al., 2012). We mixed these two populations of B cell lines at a 1:1 ratio and incubated them with three unique DNA-barcoded antigens: two stabilized trimeric HIV-1 Env proteins (SOSIP) from strains BG505 and CZA97 (Georgiev et al., 2015; Sanders et al., 2013; Ringe et al., 2017) and trimeric HA from strain H1 A/New Caledonia/20/1999 (Whittle et al., 2014; Figures 1B and S1A–S1C).

We recovered 2,321 cells with BCR sequence and antigen mapping information, highlighting the high throughput potential of LIBRA-seq (Figure S1D). For each cell, the LIBRA-seq scores for each antigen in the screening library were computed as a function of the number of unique molecular identifiers (UMIs) for the respective antigen barcode (STAR Methods). The LIBRA-seq scores of each individual antigen reliably categorized Ramos B cells by their specificity (Figure 1C). Overall, cells fell into two major populations based on their LIBRA-seq scores, and we did not observe cells that were cross-reactive for influenza HA and HIV-1 Env (Figure 1D). Further, VRC01 Ramos B cells bound both BG505 and CZA97 with a high correlation between the scores for these two antigens (Pearson's $r = 0.84$), demonstrating that LIBRA-seq readily identifies B cells that bind to multiple HIV-1 antigens (Figure 1E).

Isolation of Antibodies from a Known HIV bNAb Lineage from Donor NIAID45

We next used LIBRA-seq to analyze the antibody repertoire of donor NIAID45, who had been living with HIV-1 without antiretroviral therapy for approximately 17 years at the time of sample collection. This sample was selected as an appropriate target for LIBRA-seq analysis because a large lineage of HIV-1 bNAbs had been identified previously from this donor (Bonsignori et al., 2018; Wu et al., 2010, 2015). This lineage consists of the prototypical bNAb VRC01 as well as multiple clades of clonally related antibodies with diverse neutralization phenotypes (Wu et al., 2015). We used the same BG505, CZA97, and H1 A/New Caledonia/20/99 antigen screening library as in the Ramos B cell line experiment and recovered paired $V_H:V_L$ antibody sequences with antigen mapping for 866 cells (Figures 2A, S1D, and S2A). These B cells exhibited a variety of LIBRA-seq scores among the three antigens (Figure 2B), as can be expected from a polyclonal sample possessing a wide diversity of B cell specificities and antigen affinities. The cells displayed a few patterns based on their LIBRA-seq scores; we observed cells that were (1) $HA^{high}Env^{low}$ or (2) $HA^{low}Env^{high}$ (Figure 2B). Additionally, we observed cells that were double positive for both HIV Env variants, BG505 and CZA97, suggesting HIV-1 strain cross-reactivity of these B cells (Figure 2B).

To further validate the utility of LIBRA-seq in monoclonal antibody isolation, we next sought to identify members of

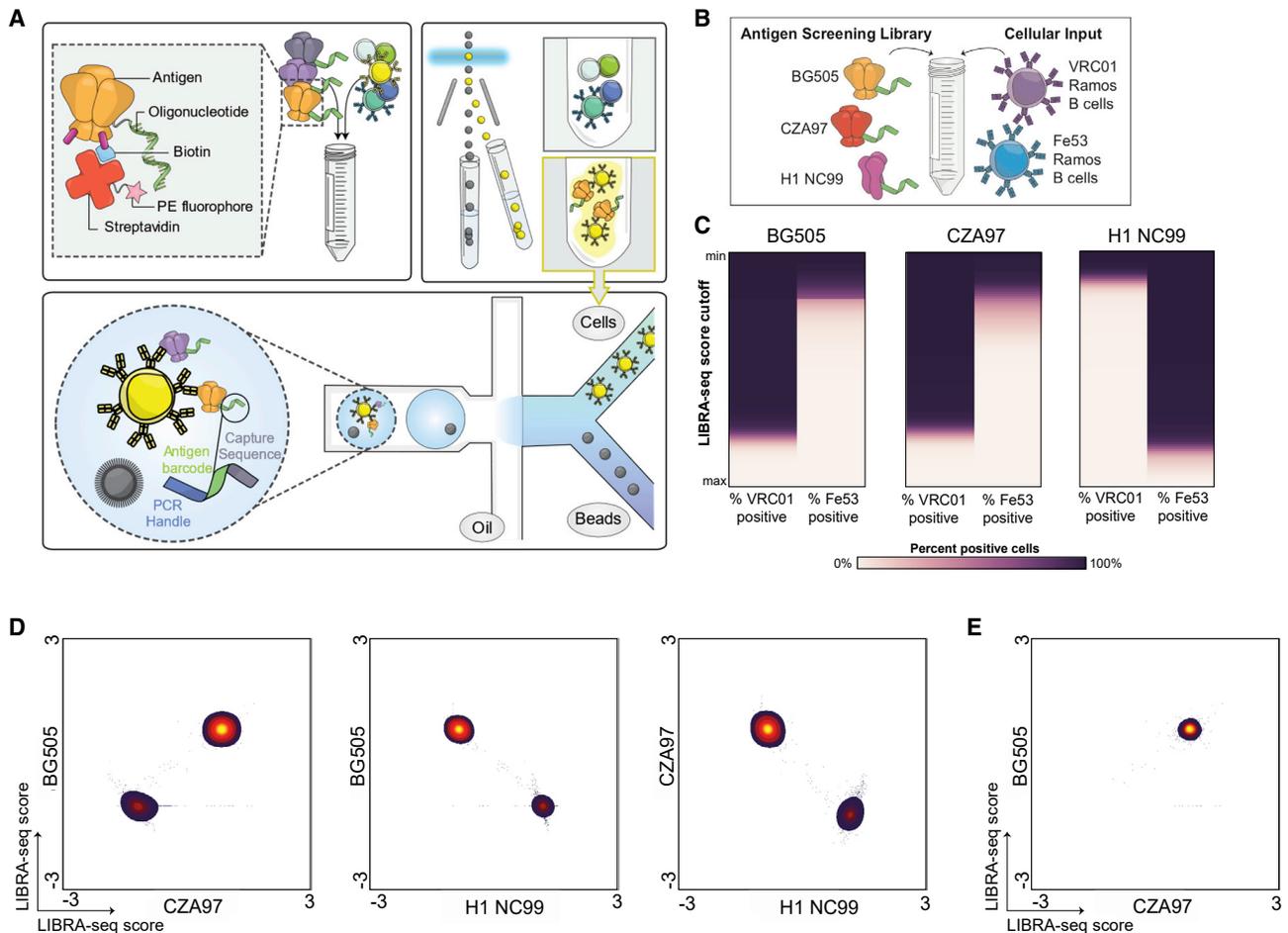


Figure 1. LIBRA-seq Assay Schematic and Validation

(A) Schematic of the LIBRA-seq assay: (top left) fluorescently labeled, DNA-barcoded antigens are used to (top right) sort antigen-positive B cells before (bottom) co-encapsulation of single B cells with bead-delivered oligos using droplet microfluidics. Bead-delivered oligos index both cellular BCR transcripts and antigen barcodes during reverse transcription, enabling direct mapping of BCR sequence to antigen specificity following sequencing. Elements of the depiction are not shown to scale, and the number and placement of oligos on each antigen can vary.

(B) The assay was initially validated on Ramos B cell lines expressing BCR sequences of the known neutralizing antibodies VRC01 and Fe53 with a three-antigen screening library: BG505, CZA97, and H1 A/New Caledonia/20/99.

(C) Between the minimum (y axis, top) and maximum (y axis, bottom) LIBRA-seq score for each antigen, different cutoffs were tested for their ability to classify each VRC01 cell and Fe53 cell as antigen-positive or -negative, where antigen-positive is defined as having a LIBRA-seq score greater than or equal to the cutoff being evaluated, and antigen-negative is defined as having a LIBRA-seq score below the cutoff. A series of 100 cutoff thresholds between the respective minimum and maximum antigen-specific LIBRA-seq scores was evaluated. At each cutoff, the percentage of total VRC01 cells (left column of each antigen subpanel) and percentage of total Fe53 cells (right columns) that were classified as positive for a given antigen are represented on a white (0%) to dark purple (100%) color scale.

(D) For each B cell, the LIBRA-seq scores for each pair of antigens were plotted. Each axis represents a range of LIBRA-seq scores for each antigen. Density of total cells is shown, with purple to yellow indicating lowest to highest number of cells, respectively.

(E) The LIBRA-seq score for BG505 (y axis) and CZA97 (x axis) for each VRC01 B cell was plotted. Each axis represents a range of LIBRA-seq scores for each antigen. Density of total cells is shown, with purple to yellow indicating lowest to highest number of cells, respectively.

See also [Figures S1](#) and [S6](#).

the VRC01 antibody lineage from the LIBRA-seq-identified antigen-specific B cells. We observed 29 BCRs that were clonally related to previously identified members of the VRC01 lineage ([Figure 2C](#)). All LIBRA-seq-identified BCRs had high levels of somatic hypermutation and utilized *IGHV1-2*02* along with the characteristic five-residue CDRL3 paired with *IGKV3-20* ([Figure 2C](#)). These B cells came from

multiple known clades of the VRC01 lineage, with sequences with high identity and phylogenetic relatedness to lineage variants VRC01, VRC02, VRC03, NIH45-46, and others ([Figure 2C](#)). Of these, 25 (86%) had a high LIBRA-seq score for at least 1 HIV-1 antigen, three (10%) had mid-range scores (between 0 and 1) for at least 1 HIV-1 antigen, and only one of the VRC01 lineage B cells had negative scores for both

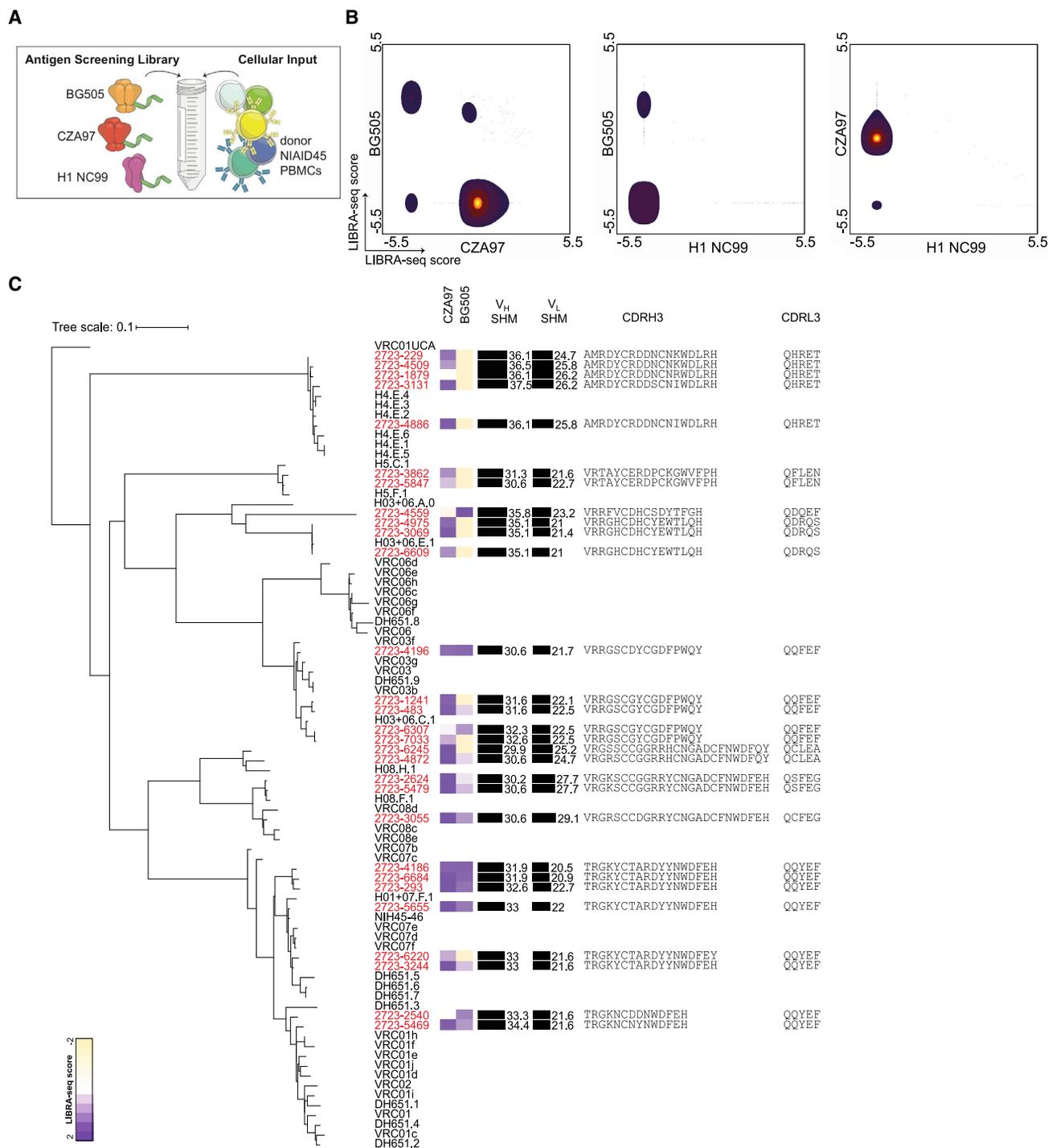


Figure 2. LIBRA-seq Applied to a Human B Cell Sample from HIV-Infected Subject NIAID45

(A) The LIBRA-seq experiment setup consisted of three antigens in the screening library: BG505, CZA97, and H1 A/New Caledonia/20/99, and the cellular input was donor NIAID45 PBMCs.

(B) After bioinformatics processing and filtering of cells recovered from single-cell sequencing, the LIBRA-seq score for each antigen was plotted (total, 866 cells). Each axis represents a range of LIBRA-seq scores for each antigen. Density of total cells is shown, with purple to yellow indicating lowest to highest number of cells, respectively.

(C) 29 VRC01 lineage B cells were identified and examined for phylogenetic relatedness to known lineage members and sequence features, with the phylogenetic tree showing the relatedness of previously identified VRC01 lineage members (black) and members identified using LIBRA-seq (red). Each row represents an antibody. Sequences were aligned using ClustalW, and a maximum likelihood tree was inferred using maximum likelihood inference. The resulting tree was

(legend continued on next page)

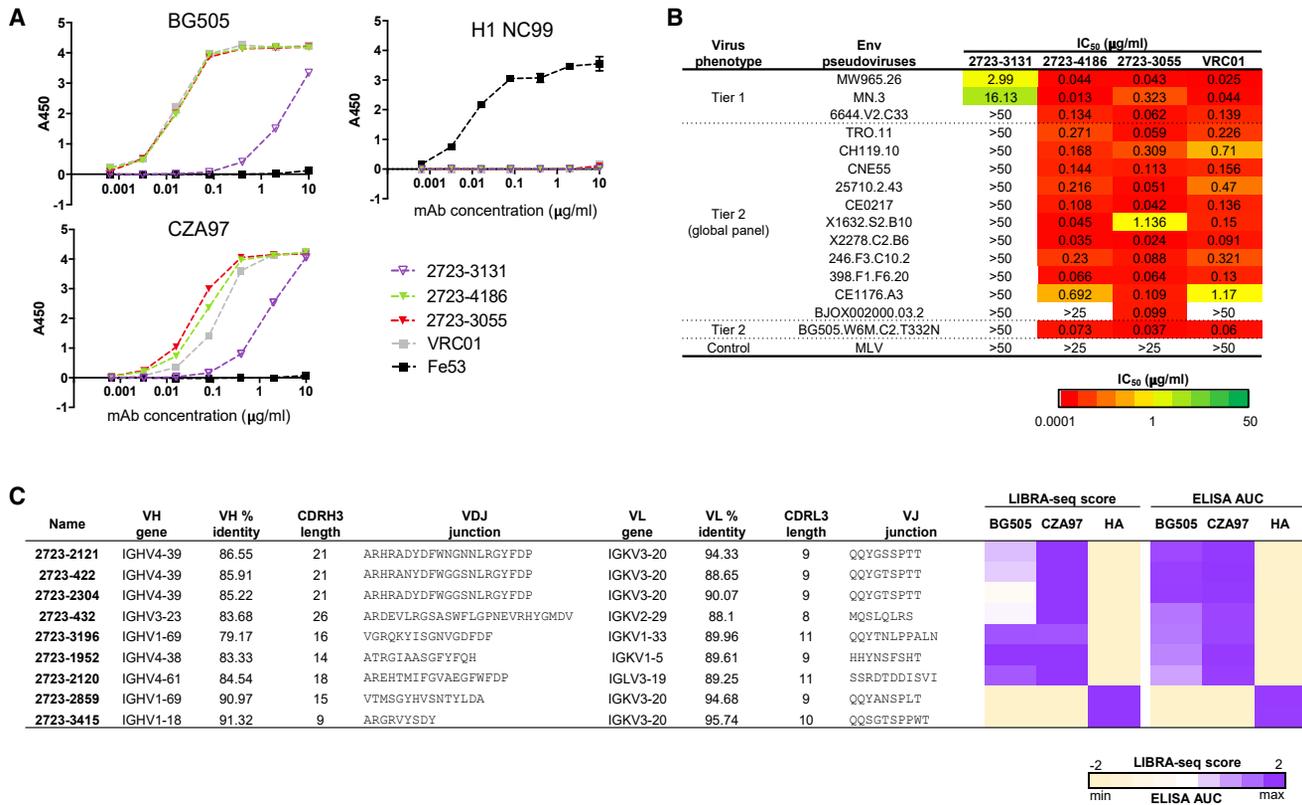


Figure 3. Characterization of LIBRA-seq-Identified Antibodies from Donor NIAID45

(A) Antigen specificity as predicted by LIBRA-seq was validated by ELISA for a subset of monoclonal antibodies belonging to the VRC01 lineage. Data are represented as mean ± SEM for one ELISA experiment. ELISA data are representative of at least two independent experiments.

(B) Neutralization of tier 1, tier 2, and control viruses by VRC01 and the LIBRA-seq-identified VRC01 lineage members 2723-3131, 2723-4186, and 2723-3055. IC₅₀ values are shown from high potency (0.0001 µg/ml, red) to low potency (50 µg/ml, green). Lack of neutralization IC₅₀ for concentrations tested is displayed as white.

(C) Sequence characteristics and antigen specificity of LIBRA-seq-identified antibodies from donor NIAID45. Percent identity is calculated at the nucleotide level, and CDRH3 and CDRL3 lengths and sequences are noted at the amino acid level. LIBRA-seq scores for each antigen are displayed as a heatmap, with a LIBRA-seq score of -2 displayed as light yellow, 0 as white, and 2 as purple; in this heatmap, scores lower or higher than that range are shown as -2 and 2, respectively. ELISA binding data against BG505, CZA97, and H1 A/New Caledonia/20/99 are displayed as a heatmap of the AUC analysis calculated from the data in Figure S3A, with AUC of 0 displayed as light yellow, 50% maximum as white, and maximum AUC as purple. ELISA data are representative of at least two independent experiments.

See also Figures S2 and S3.

HIV-1 antigens (Figures 2C and S2B). We recombinantly expressed three of the LIBRA-seq-identified lineage members, named 2723-3055, 2723-4186, and 2723-3131, to confirm the ability of these antibodies to bind the screening probes. Antibody 2723-3131 showed binding to CZA97 and BG505 by ELISA (Figure 3A) and neutralized two Tier 1 viruses but no viruses on a global panel of representative HIV-1 strains

(deCamp et al., 2014; Figure 3B). Both 2723-3055 and 2723-4186 bound to BG505 and CZA97 and potently neutralized 12 of 12 and 11 of 12 viruses on a global panel, respectively (Figures 3A and 3B). Together, the results from the donor NIAID45 analysis suggest that the LIBRA-seq platform can be used successfully to down-select cross-reactive bNAbs in prospective antibody discovery efforts.

visualized using an inferred VRC01 unmutated common ancestor (UCA) (GenBank: MK032222) as the root. For each antibody isolated from LIBRA-seq, a heatmap of the LIBRA-seq scores for each HIV antigen (BG505 and CZA97) is shown. A scale of tan-white-purple represents LIBRA-seq scores from -2 to 0 to 2; in this heatmap, scores lower or higher than that range are shown as -2 and 2, respectively. Levels of somatic hypermutation (SHM) at the nucleotide level for the heavy- and light-chain variable genes as reported by the international immunogenetics information system (IMGT) are displayed as bars, with the numerical percentage value listed to the right of the bar; the length of the bar corresponds to the level of SHM. Amino acid sequences of the complementarity determining region 3 for the heavy chain (CDRH3) and the light chain (CDRL3) for each antibody are displayed. The tree was visualized and annotated using iTol (Letunic and Bork, 2019).

See also Figures S1, S2, and S6.

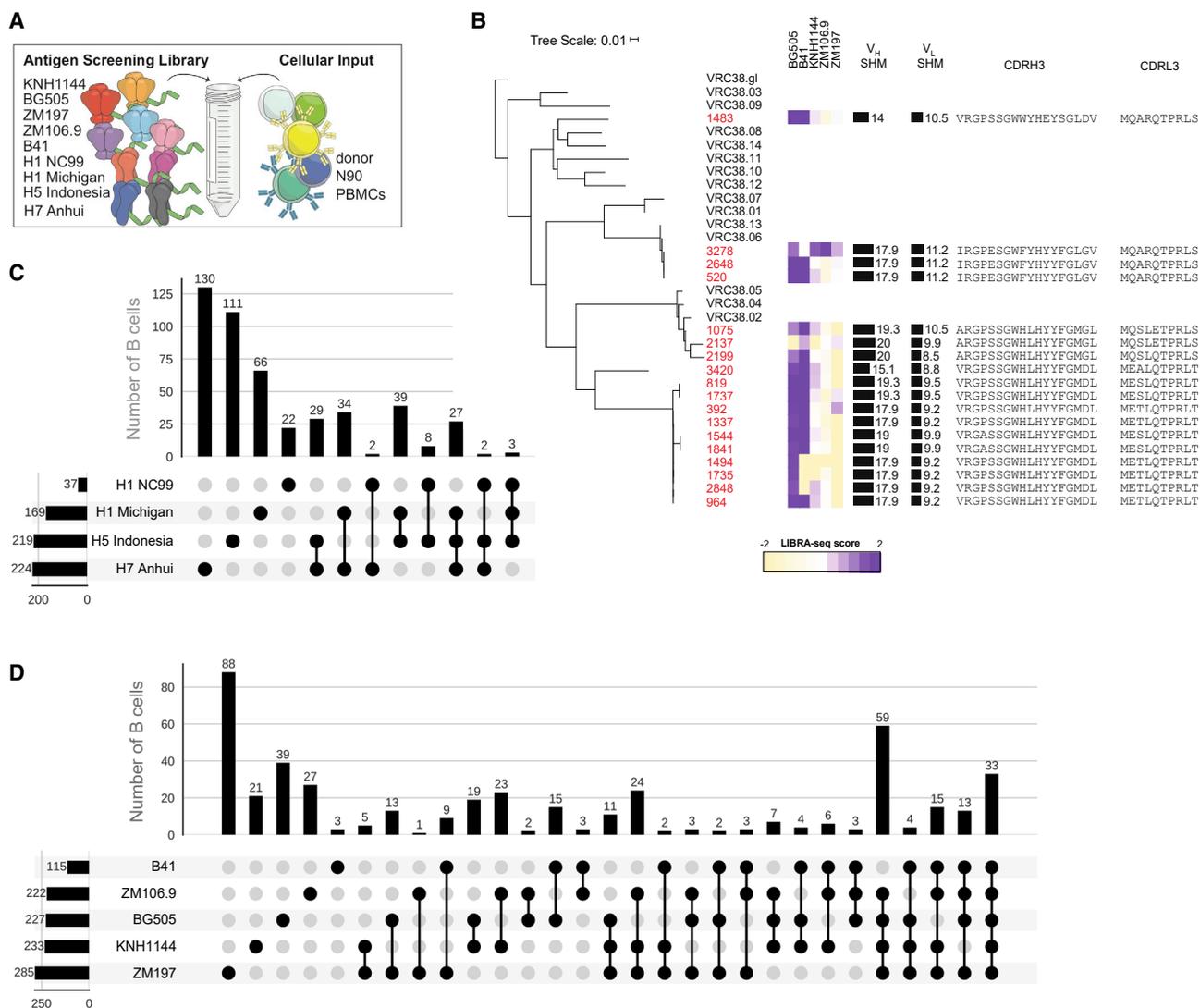


Figure 4. LIBRA-seq Applied to a Sample from NIAID Donor N90

(A) The LIBRA-seq experiment setup consisted of nine antigens in the screening library: 5 HIV-1 Env (KNH1144, BG505, ZM197, ZM106.9, and B41) and 4 influenza HA (H1 A/New Caledonia/20/99, H1 A/Michigan/45/2015, H5 Indonesia/5/2005, and H7 Anhui/1/2013); the cellular input was donor N90 PBMCs.

(B) 18 VRC38 lineage B cells were identified and examined for phylogenetic relatedness to known lineage members as well as for sequence features, with the phylogenetic tree showing the relatedness of previously identified VRC38 lineage members (black) and LIBRA-seq-identified members (red). Each row represents an antibody. Sequences were aligned using ClustalW, and a maximum likelihood tree was inferred using maximum likelihood inference. The resulting tree was visualized with a germline-reverted antibody from lineage VRC38 (STAR Methods) as the root. For each antibody isolated from LIBRA-seq, a heatmap of the LIBRA-seq scores for each HIV antigen is shown. Tan-white-purple represents LIBRA-seq scores from -2 to 0 to 2 ; in this heatmap, scores lower or higher than that range are shown as -2 and 2 , respectively. Levels of SHM at the nucleotide level for the heavy- and light-chain variable genes as reported by IMGT are displayed as bars, with the numerical percentage value listed to the right of the bar; the length of the bar corresponds to the level of SHM. Amino acid sequences of the CDRH3 and the CDRL3 for each antibody are displayed. The tree was visualized and annotated using iTol (Letunic and Bork, 2019).

(C and D) For each combination of (C) influenza HAs or (D) HIV SOSIPs, the number of B cells with high LIBRA-seq scores (≥ 1) is displayed as a bar graph. The combinations of antigens are displayed by filled circles, showing which antigens are part of a given combination. Each combination is mutually exclusive. The total number of B cells with high LIBRA-seq scores for each antigen is indicated as a horizontal bar at the bottom left of each subpanel. See also Figures S1, S4, S5, and S6.

Identification of Additional Anti-HIV and Anti-influenza Antibodies from Donor NIAID45

To further validate the ability of LIBRA-seq to accurately identify antigen-specific B cells, we produced a number of putative HIV-specific and influenza-specific monoclonal antibodies from

donor NIAID45 that did not belong to the VRC01 lineage. In particular, we recombinantly produced seven additional anti-HIV antibodies, three of which were clonally related (2723-2121, 2723-422, and 2723-2304) (Figure 3C). We selected these seven antibodies because all had high LIBRA-seq scores for at

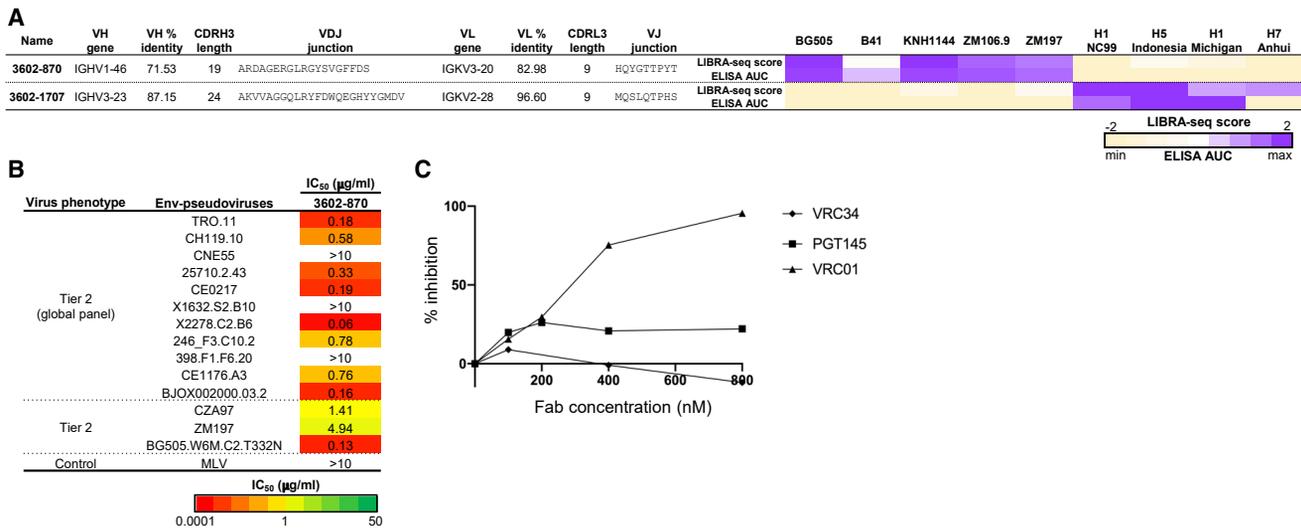


Figure 5. Characterization of LIBRA-seq-Identified Antibodies from NIAID Donor N90

(A) Sequence characteristics and antigen specificity of newly identified antibodies from donor N90. Percent identity is calculated at the nucleotide level, and CDR length and sequences are noted at the amino acid level. LIBRA-seq scores for each antigen are displayed as a heatmap, with a LIBRA-seq score of -2 displayed as light yellow, 0 as white, and 2 as purple; in this heatmap, scores lower or higher than that range are shown as -2 and 2, respectively. ELISA binding data are displayed as a heatmap of the AUC analysis calculated from the data in Figure S4B, with AUC of 0 displayed as light yellow, 50% maximum as white, and maximum AUC as purple. ELISA data are representative of at least two independent experiments.

(B) Neutralization of tier 2 and control viruses by antibody 3602-870. IC₅₀ values are shown from high potency (0.0001 µg/ml, red) to low potency (50 µg/ml, green). Lack of neutralization IC₅₀ for concentrations tested is displayed as white.

(C) Inhibition of BG505 DS-SOSIP/293F binding to 3602-870 IgG in the presence of VRC34 Fab (diamond), PGT145 Fab (square), and VRC01 Fab (triangle). See also Figures S4 and S5.

least one HIV-1 antigen. All seven antibodies bound the antigens by ELISA, as expected based on the respective LIBRA-seq scores, with high similarity between the patterns of LIBRA-seq scores and ELISA area under the curve (AUC) values (Figures 3C and S3A; STAR Methods). We further characterized one of these antibodies, 2723-2121, and determined that it bound to a stabilized BG505 trimer (DS-SOSIP) (Kwon et al., 2015) by surface plasmon resonance (SPR) (Figures S3B and S3C). Antibody 2723-2121 competed for trimer binding with VRC01 (Figure S3D), neutralized three tier 1 pseudoviruses and 2 of 11 tier 2 pseudoviruses from a global panel (Figure S3E), and mediated trogocytosis and antibody-dependent cellular phagocytosis (Figure S3F). In addition to the HIV-specific antibodies, we also characterized two antibodies predicted to have influenza specificity based on their LIBRA-seq scores for H1 A/New Caledonia/20/99 (Figure 3C). In agreement with the LIBRA-seq scores, antibodies 2723-2859 and 2723-3415 bound H1 A/New Caledonia/20/99 but not BG505 or CZA97 by ELISA, confirming the ability of LIBRA-seq to simultaneously isolate antibodies to multiple diverse antigens (Figures 3C and S3A).

Discovery of an HIV bNAb from Donor N90 Using a Nine-Antigen Screening Library

Having validated LIBRA-seq with three antigens on both Ramos B cell lines and primary B cells from a patient sample, we sought to increase the number of antigens in the screening library. To that end, we screened B cells from NIAID donor N90 against nine antigens (Figure 4A). We selected this sample because a

single bNAb lineage (VRC38) targeting the V1/V2 epitope was isolated previously from this donor; however, the neutralization breadth of the VRC38 lineage could not account for the full serum neutralization breadth (Cale et al., 2017; Wu et al., 2012). This suggested that there could be additional bNAb lineages present in the B cell repertoire of N90, and we reasoned that utilizing multiple SOSIP probes could help accelerate identification of such antibodies. Thus, we sought to determine whether LIBRA-seq could accomplish two goals: (1) to recover antigen-specific B cells from the VRC38 lineage and (2) to identify new bNAbs that could neutralize viruses that are resistant to neutralization by the VRC38 lineage.

To increase the number of antigens in our screening library, we utilized a panel that consisted of five HIV-1 Env trimers from a variety of clades, BG505 (clade A), B41 (clade B), ZM106.9 (clade C), ZM197 (clade C), and KNH1144 (clade A) (Sanders et al., 2013; Harris et al., 2011; Joyce et al., 2017; Julien et al., 2015; Pugach et al., 2015; Ringe et al., 2017), along with four diverse HA trimers (H1 A/New Caledonia/20/99, H1 A/Michigan/45/2015, H5 A/Indonesia/5/2005, and H7 A/Anhui/1/2013) (Figures 4A and S1A). After applying LIBRA-seq to donor N90 PBMCs, we recovered paired V_H:V_L antibody sequences with antigen mapping for 1,465 cells (Figures S1D and S4A). Within this set of cells, we identified 18 B cells that were members of the VRC38 lineage (Figure 4B). Of these, 17 had high LIBRA-seq scores for at least one HIV antigen, and one had no high LIBRA-seq scores but had a mid-range score for two SOSIPs (Figure 4B).

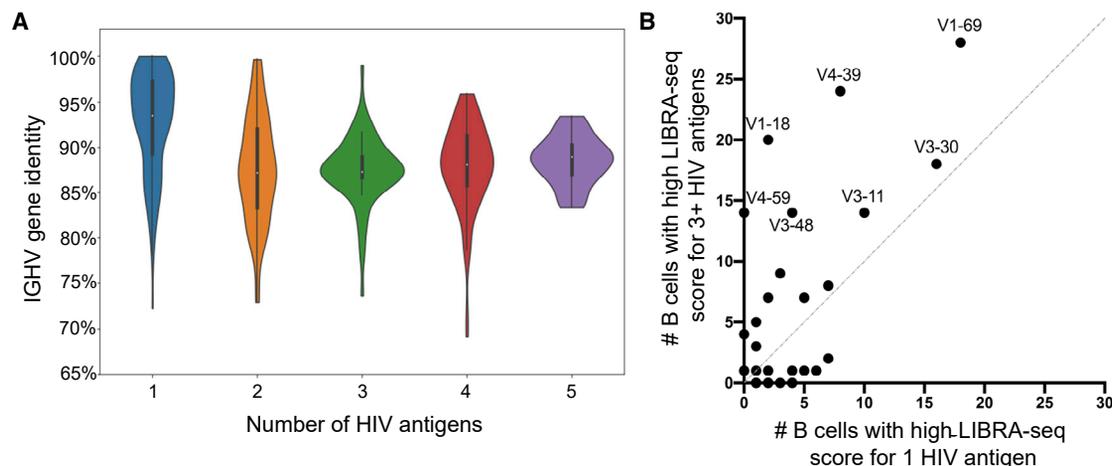


Figure 6. Sequence Properties of the Antigen-Specific B Cell Repertoire

(A) *IGHV* gene identity (y axis) is plotted for cells with high (≥ 1) LIBRA-seq scores for any combination of 1–5 HIV-1 SOSIP antigens (x axis). Each distribution is displayed as a kernel density estimation, where wider sections of a given distribution represent a higher probability that B cells possess a given germline identity percentage. The median of each distribution is displayed as a white dot, the interquartile range is displayed as a thick bar, and a thin line extends to 1.5 times the interquartile range. The violin ranges were limited to the observed data. Included are cells with IgG or IgA constant heavy genes as determined by Cell Ranger. (B) Each dot represents an *IGHV* germline gene, plotted based on the number of B cells reactive to only 1 HIV-1 SOSIP antigen (x axis) and the number of B cells reactive to 3 or more HIV-1 SOSIP antigens (y axis) that are assigned to that respective *IGHV* germline gene. Only B cells with high (≥ 1) LIBRA-seq scores for any HIV-1 antigen and with IgG or IgA constant heavy genes as determined by Cell Ranger are shown. See also Figure S5.

We next focused our analysis on the B cells with the highest LIBRA-seq scores in the N90 sample, focusing on cells that had LIBRA-seq scores for any antigen above one (901 cells) (Figures 4C, 4D, and S5A). We observed 32 cells that had high LIBRA-seq scores for three of the four influenza antigens (Figure 4C); we recombinantly produced one of these, 3602-1707, and confirmed broad influenza recognition, with high correlation between LIBRA-seq scores and ELISA AUC (Spearman correlation, 0.77; $p = 0.015$) (Figures 5A and S4B).

We also observed cells that had high LIBRA-seq scores for each of the different HIV-1 antigens, including 124 cells that had high scores for four or more SOSIPs (Figure 4D). We then down-selected SOSIP-high B cells based on having high LIBRA-seq scores to at least 3 SOSIP variants. In particular, we identified two members from the same antibody lineage that had high LIBRA-seq scores for BG505, KNH1144, ZM106.9, and ZM197. This lineage utilized the germline genes *IGHV1-46* and *IGKV3-20* and was highly mutated in both the heavy- and light-chain V genes. We recombinantly expressed one of the lineage members, 3602-870, which was 28.5% mutated in its heavy chain V gene and 17.0% mutated in its light chain V gene and had a 19-amino acid CDRH3 and 9-amino acid CDRL3 (Figure 5A). 3602-870 bound all SOSIP probes by ELISA (Spearman correlation of 0.97, $p < 0.001$ between LIBRA-seq scores and ELISA AUC) and neutralized 79% of tested tier 2 viruses (11 of 14), including several viruses that were not neutralized by VRC38.01 (Cale et al., 2017; Figures 5A, 5B, and S4B). Of note, 3602-870 neutralized BG505 and ZM197, both of which were used as probes in the antigen screening library (Figure 5B). 3602-870 bound BG505 DS-SOSIP by SPR and competed for BG505 DS-SOSIP binding with VRC01 Fab (Figures 5C and S4C).

In summary, LIBRA-seq enabled high-throughput, highly multiplexed screening of single B cells from an HIV-infected subject against a large antigen panel. This resulted in the identification of hundreds of antigen-specific monoclonal antibody leads from donor N90, with high-resolution antigen specificity mapping helping to facilitate rapid lead prioritization to identify a novel bNAb lineage.

DISCUSSION

Here, we developed a novel method to interrogate antibody-antigen interactions via a sequencing-based readout. After validating the approach on cell lines with known BCRs, we applied LIBRA-seq to prospective antibody discovery. We identified members of two known HIV-specific bNAb lineages from previously characterized human infection samples and a novel bNAb lineage. Additionally, we identified many other candidate broadly reactive HIV-specific antibodies and validated specificity for a subset of them. Within both HIV-1 infection samples, we also isolated influenza-specific antibodies using HA screening probes, highlighting the utility of LIBRA-seq for simultaneously screening B cell repertoires against multiple diverse antigen targets. In principle, NGS-based coupling of antibody sequence and specificity enables screening of potentially millions of single B cells for reactivity to a larger repertoire of epitopes than purely fluorescence-based methods because sequence space is not hindered by spectral overlap. Using LIBRA-seq may therefore help to maximize lead discovery per experiment, an important consideration when preserving limited samples.

Beyond LIBRA-seq's utility in antibody discovery, high-throughput coupling of antibody sequence and specificity can enable high-resolution immune profiling. For example, in donor

N90, we observed an increase in *IGHV* gene somatic hypermutation between B cells that had a high LIBRA-seq score for a single HIV-1 antigen versus B cells that had high LIBRA-seq scores for multiple HIV-1 antigens (Figure 6A). We also observed use of specific germline genes to be more frequent in B cells that exhibited broad as opposed to strain-specific HIV-1 antigen reactivity (Figures 6B and S5B). Elucidation of such relationships, enabled by the LIBRA-seq technology, may guide germline-targeting vaccine design efforts (Dosenovic et al., 2019; Jardine et al., 2013, 2016; Stamatatos et al., 2017) and can provide insights into the requirements for acquisition of HIV-1 antigen cross-reactivity. The application of LIBRA-seq to antibody discovery and immune profiling should translate into rapid accumulation of new data, leading to novel insights in basic and applied immunology.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **LEAD CONTACT AND MATERIALS AVAILABILITY**
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
 - Human Subjects
 - Cell Lines
- **METHOD DETAILS**
 - Antigen expression and purification
 - Oligonucleotide barcodes
 - Conjugation of oligonucleotide barcodes to antigens
 - Fluorescent labeling of antigens
 - Enrichment of antigen-specific B cells
 - 10X Genomics single cell processing and next generation sequencing
 - Processing of BCR sequence and antigen barcode reads
 - Determination of LIBRA-seq Score
 - Phylogenetic trees
 - Antibody expression and purification
 - Enzyme linked immunosorbent assay (ELISA)
 - TZM-bl Neutralization Assays
 - Surface Plasmon Resonance and Fab competition
 - ADCP, ADCD, Troglucytosis, ADCC Assays
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
- **DATA AND CODE AVAILABILITY**
 - Data Availability Statement
 - Code Availability Statement

ACKNOWLEDGMENTS

We thank Angela Jones, Latha Raju, and Jamie Robertson of Vanderbilt Technologies for Advanced Genomics for their expertise regarding NGS and library preparation; Rebecca Gillespie for help with producing the HA probe; Hannah Polikowsky, James Crowe, Jr., and Spyros Kalams for helpful discussions regarding experimental design; David Flaherty of the Vanderbilt Flow Cytometry Shared Resource for help with flow panel optimization; Carol Crowther of NICD for project management; Peter Kwong and John Mascola for providing the plasmid for BG505; and members of the Georgiev laboratory for comments on the manuscript. The Vanderbilt VANTAGE Core provided technical assistance for this work. VANTAGE is supported in part by CTSA

grant 5UL1 RR024975-03, the Vanderbilt Ingram Cancer Center (P30 CA68485), the Vanderbilt Vision Center (P30 EY08126), and NIH/NCRR (G20 RR030956). This work was conducted in part using the resources of the Advanced Computing Center for Research and Education at Vanderbilt University (Nashville, TN). Flow cytometry experiments were performed in the VUMC Flow Cytometry Shared Resource. The VUMC Flow Cytometry Shared Resource is supported by the Vanderbilt Ingram Cancer Center (P30 CA68485) and the Vanderbilt Digestive Disease Research Center (DK058404). For work described in this manuscript, I.S.G., I.S., A.R.S., N.R., and A.R.G. were supported by NIH grant R01 AI131722. A.R.S. was supported by NIH grant T32 GM008320-30. K.A.P. was supported by NIH grant T32 AI112541. A.A.M. and K.J.K. were supported by the Vanderbilt Trans-Institutional Program (TIPs; "Integrating Structural with Big Data for Next-Generation Vaccines"). M.K., B.S.G., and M.C. were supported by the NIH Intramural Research Program. L.M., C.O., and R.M. were supported by the South African Medical Research Council and H3A-U01 (NIH grant 1U01AI136677). P.A. and K.J. were supported by NIH grant R01 AI145687.

AUTHOR CONTRIBUTIONS

Conception of LIBRA-seq technology, I.S.; Conceptualization, I.S., A.R.S., and I.S.G.; Methodology, I.S., A.R.S., and I.S.G.; Investigation, I.S., A.R.S., K.A.P., A.A.M., R.M., S.R., C.O., N.R., A.R.G., K.J., K.J.K., J.S.Q., D.L., L.R., W.J.M., and I.S.G.; Software, I.S. and N.R.; Validation, I.S., A.R.S., and N.R.; Writing – Original Draft, I.S. and A.R.S.; Writing – Review & Editing, all authors; Funding Acquisition, I.S.G., D.L., L.M., P.A., M.C., B.S.G., and I.S.; Resources, M.K., P.A., B.S.G., M.C., D.L., L.R., L.M., and I.S.G.; Supervision, I.S.G., L.M., and P.A.

DECLARATION OF INTERESTS

Vanderbilt University has filed for patent protection of some of the technology and results presented in this study, with I.S., A.R.S., and I.S.G. listed as inventors. W.J.M. is an employee and shareholder of 10X Genomics.

Received: March 27, 2019

Revised: August 28, 2019

Accepted: October 31, 2019

Published: November 28, 2019

REFERENCES

- Ackerman, M.E., Moldt, B., Wyatt, R.T., Dugast, A.S., McAndrew, E., Tsoukas, S., Jost, S., Berger, C.T., Sciaranghella, G., Liu, Q., et al. (2011). A robust, high-throughput assay to determine the phagocytic activity of clinical antibody samples. *J. Immunol. Methods* 366, 8–19.
- Adler, A.S., Mizrahi, R.A., Spindler, M.J., Adams, M.S., Asensio, M.A., Edgar, R.C., Leong, J., Leong, R., and Johnson, D.S. (2017a). Rare, high-affinity mouse anti-PD-1 antibodies that function in checkpoint blockade, discovered using microfluidics and molecular genomics. *MAbs* 9, 1270–1281.
- Adler, A.S., Mizrahi, R.A., Spindler, M.J., Adams, M.S., Asensio, M.A., Edgar, R.C., Leong, J., Leong, R., Roalfe, L., White, R., et al. (2017b). Rare, high-affinity anti-pathogen antibodies from human repertoires, discovered using microfluidics and molecular genomics. *MAbs* 9, 1282–1296.
- Alamyar, E., Giudicelli, V., Li, S., Duroux, P., and Lefranc, M.P. (2012). IMGT/HighV-quest: The IMGT web portal for immunoglobulin (IG) or antibody and T cell receptor (TR) analysis from NGS high throughput and deep sequencing. *Immunome Res.* 8, 26.
- Bonsignori, M., Scott, E., Wiehe, K., Easterhoff, D., Alam, S.M., Hwang, K.-K., Cooper, M., Xia, S.-M., Zhang, R., Montefiori, D.C., et al. (2018). Inference of the HIV-1 VRC01 Antibody Lineage Unmutated Common Ancestor Reveals Alternative Pathways to Overcome a Key Glycan Barrier. *Immunity* 49, 1162–1174.e8.
- Brekke, O.H., and Sandlie, I. (2003). Therapeutic antibodies for human diseases at the dawn of the twenty-first century. *Nat. Rev. Drug Discov.* 2, 52–62.

- Briney, B., Inderbitzin, A., Joyce, C., and Burton, D.R. (2019). Commonality despite exceptional diversity in the baseline human antibody repertoire. *Nature* 566, 393–397.
- Buchacher, A., Predl, R., Strutzenberger, K., Steinfellner, W., Trkola, A., Purtscher, M., Gruber, G., Tauer, C., Steindl, F., Jungbauer, A., et al. (1994). Generation of human monoclonal antibodies against HIV-1 proteins; electrofusion and Epstein-Barr virus transformation for peripheral blood lymphocyte immortalization. *AIDS Res. Hum. Retroviruses* 10, 359–369.
- Busse, C.E., Czogiel, I., Braun, P., Arndt, P.F., and Wardemann, H. (2014). Single-cell based high-throughput sequencing of full-length immunoglobulin heavy and light chain genes. *Eur. J. Immunol.* 44, 597–603.
- Cale, E.M., Gorman, J., Radakovich, N.A., Crooks, E.T., Osawa, K., Tong, T., Li, J., Nagarajan, R., Ozorowski, G., Ambrozak, D.R., et al. (2017). Virus-like Particles Identify an HIV V1V2 Apex-Binding Neutralizing Antibody that Lacks a Protruding Loop. *Immunity* 46, 777–791.e10.
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890.
- deCamp, A., Hraber, P., Bailer, R.T., Seaman, M.S., Ochsenbauer, C., Kappes, J., Gottardo, R., Edlefsen, P., Self, S., Tang, H., et al. (2014). Global Panel of HIV-1 Env Reference Strains for Standardized Assessments of Vaccine-Elicited Neutralizing Antibodies. *J. Virol.* 88, 2489–2507.
- DeFalco, J., Harbell, M., Manning-Bog, A., Baia, G., Scholz, A., Millare, B., Sumi, M., Zhang, D., Chu, F., Dowd, C., et al. (2018). Non-progressing cancer patients have persistent B cell responses expressing shared antibody paratopes that target public tumor antigens. *Clin. Immunol.* 187, 37–45.
- DeKosky, B.J., Ippolito, G.C., Deschner, R.P., Lavinder, J.J., Wine, Y., Rawlings, B.M., Varadarajan, N., Giesecke, C., Dörner, T., Andrews, S.F., et al. (2013). High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. *Nat. Biotechnol.* 31, 166–169.
- Dosenovic, P., Pettersson, A.-K., Wall, A., Thientosapol, E.S., Feng, J., Weidle, C., Bhullar, K., Kara, E.E., Hartweg, H., Pai, J.A., et al. (2019). Anti-idiotypic antibodies elicit anti-HIV-1-specific B cell responses. *J. Exp. Med.* 216, 2316–2330.
- Georgiev, I.S., Joyce, M.G., Yang, Y., Sastry, M., Zhang, B., Baxa, U., Chen, R.E., Druz, A., Lees, C.R., Narpala, S., et al. (2015). Single-Chain Soluble BG505.SOSIP gp140 Trimers as Structural and Antigenic Mimics of Mature Closed HIV-1 Env. *J. Virol.* 89, 5318–5329.
- Georgiou, G., Ippolito, G.C., Beausang, J., Busse, C.E., Wardemann, H., and Quake, S.R. (2014). The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat. Biotechnol.* 32, 158–168.
- Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321.
- Gupta, N.T., Vander Heiden, J.A., Uduman, M., Gadala-Maria, D., Yaari, G., and Kleinstein, S.H. (2015). Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. *Bioinformatics* 31, 3356–3358.
- Harris, A., Borgnia, M.J., Shi, D., Bartesaghi, A., He, H., Pejchal, R., Kang, Y.K., Depetris, R., Marozsan, A.J., Sanders, R.W., et al. (2011). Trimeric HIV-1 glycoprotein gp140 immunogens and native HIV-1 envelope glycoproteins display the same closed and open quaternary molecular architectures. *Proc. Natl. Acad. Sci. USA* 108, 11440–11445.
- Huang, J., Kang, B.H., Pancera, M., Lee, J.H., Tong, T., Feng, Y., Imamichi, H., Georgiev, I.S., Chuang, G.-Y., Druz, A., et al. (2014). Broad and potent HIV-1 neutralization by a human antibody that binds the gp41-gp120 interface. *Nature* 515, 138–142.
- Hunter, J.D. (2007). Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* 9, 90–95.
- Jardine, J., Julien, J.-P., Menis, S., Ota, T., Kalyuzhnyi, O., McGuire, A., Sok, D., Huang, P.-S., MacPherson, S., Jones, M., et al. (2013). Rational HIV immunogen design to target specific germline B cell receptors. *Science* 340, 711–716.
- Jardine, J.G., Kulp, D.W., Havenar-Daughton, C., Sarkar, A., Briney, B., Sok, D., Sesterhenn, F., Ereno-Orbea, J., Kalyuzhnyi, O., Deresa, I., et al. (2016). HIV-1 broadly neutralizing antibody precursor B cells revealed by germline-targeting immunogen. *Science* 351, 1458–1463.
- Joyce, M.G., Georgiev, I.S., Yang, Y., Druz, A., Geng, H., Chuang, G.Y., Kwon, Y.D., Pancera, M., Rawi, R., Sastry, M., et al. (2017). Soluble Prefusion Closed DS-SOSIP.664-Env Trimers of Diverse HIV-1 Strains. *Cell Rep.* 21, 2992–3002.
- Julien, J.-P., Lee, J.H., Ozorowski, G., Hua, Y., Torrents de la Peña, A., de Taeye, S.W., Nieuwsma, T., Cupo, A., Yasmeen, A., Golabek, M., et al. (2015). Design and structure of two HIV-1 clade C SOSIP.664 trimers that increase the arsenal of native-like Env immunogens. *Proc. Natl. Acad. Sci. USA* 112, 11947–11952.
- Kong, R., Xu, K., Zhou, T., Acharya, P., Lemmin, T., Liu, K., Ozorowski, G., Soto, C., Taft, J.D., Bailer, R.T., et al. (2016). Fusion peptide of HIV-1 as a site of vulnerability to neutralizing antibody. *Science* 352, 828–833.
- Kotecha, N., Krutzik, P.O., and Irish, J.M. (2010). Web-based analysis and publication of flow cytometry experiments. *Curr. Protoc. Cytom. Chapter 10*, Unit 10.17.
- Kwon, Y.D., Pancera, M., Acharya, P., Georgiev, I.S., Crooks, E.T., Gorman, J., Joyce, M.G., Guttman, M., Ma, X., Narpala, S., et al. (2015). Crystal structure, conformational fixation and entry-related interactions of mature ligand-free HIV-1 Env. *Nat. Struct. Mol. Biol.* 22, 522–531.
- Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47 (W1), W256–W259.
- Lex, A., Gehlenborg, N., Strobel, H., Vuillemot, R., and Pfister, H. (2014). UpSet: Visualization of Intersecting Sets. *IEEE Trans Vis Comput Graph* 20, 1983–1992.
- Lingwood, D., McTamney, P.M., Yassine, H.M., Whittle, J.R.R., Guo, X., Boyington, J.C., Wei, C.J., and Nabel, G.J. (2012). Structural and genetic basis for development of broadly neutralizing influenza antibodies. *Nature* 489, 566–570.
- Mimitou, E.P., Cheng, A., Montalbano, A., Hao, S., Stoeckius, M., Legut, M., Roush, T., Herrera, A., Papalex, E., Ouyang, Z., et al. (2019). Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. *Nat. Methods* 16, 409–412.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12, 2825–2830.
- Peterson, V.M., Zhang, K.X., Kumar, N., Wong, J., Li, L., Wilson, D.C., Moore, R., McClanahan, T.K., Sadekova, S., and Klappenbach, J.A. (2017). Multiplexed quantification of proteins and transcripts in single cells. *Nat. Biotechnol.* 35, 936–939.
- Pollara, J., Hart, L., Brewer, F., Pickeral, J., Packard, B.Z., Hoxie, J.A., Komoriya, A., Ochsenbauer, C., Kappes, J.C., Roederer, M., et al. (2011). High-throughput quantitative analysis of HIV-1 and SIV-specific ADCC-mediating antibody responses. *Cytometry A* 79, 603–612.
- Pugach, P., Ozorowski, G., Cupo, A., Ringe, R., Yasmeen, A., de Val, N., Derking, R., Kim, H.J., Korzun, J., Golabek, M., et al. (2015). A native-like SOSIP.664 trimer based on an HIV-1 subtype B env gene. *J. Virol.* 89, 3380–3395.
- Rajewsky, K. (1996). Clonal selection and learning in the antibody system. *Nature* 381, 751–758.
- Richardson, S.I., Chung, A.W., Natarajan, H., Mavvakure, B., Mkhize, N.N., Garrett, N., Abdool Karim, S., Moore, P.L., Ackerman, M.E., Alter, G., and Morris, L. (2018a). HIV-specific Fc effector function early in infection predicts the development of broadly neutralizing antibodies. *PLoS Pathog.* 14, e1006987.
- Richardson, S.I., Crowther, C., Mkhize, N.N., and Morris, L. (2018b). Measuring the ability of HIV-specific antibodies to mediate trogocytosis. *J. Immunol. Methods* 463, 71–83.
- Ringe, R.P., Ozorowski, G., Yasmeen, A., Cupo, A., Cruz Portillo, V.M., Pugach, P., Golabek, M., Rantalainen, K., Holden, L.G., Cottrell, C.A., et al. (2017). Improving the Expression and Purification of Soluble, Recombinant Native-Like HIV-1 Envelope Glycoprotein Trimers by Targeted Sequence Changes. *J. Virol.* 91, e00264–e002617.

- Sanders, R.W., Derking, R., Cupo, A., Julien, J.P., Yasmeen, A., de Val, N., Kim, H.J., Blattner, C., de la Peña, A.T., Korzun, J., et al. (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.
- Sarzotti-Kelsoe, M., Bailer, R.T., Turk, E., Lin, C.L., Bilski, M., Greene, K.M., Gao, H., Todd, C.A., Ozaki, D.A., Seaman, M.S., et al. (2014). Optimization and validation of the TZM-bl assay for standardized assessments of neutralizing antibodies against HIV-1. *J. Immunol. Methods* **409**, 131–146.
- Scheid, J.F., Mouquet, H., Feldhahn, N., Seaman, M.S., Velinzon, K., Pietzsch, J., Ott, R.G., Anthony, R.M., Zebroski, H., Hurley, A., et al. (2009). Broad diversity of neutralizing antibodies isolated from memory B cells in HIV-infected individuals. *Nature* **458**, 636–640.
- Setliff, I., McDonnell, W.J., Raju, N., Bombardi, R.G., Murji, A.A., Scheepers, C., Ziki, R., Mynhardt, C., Shepherd, B.E., Mamchak, A.A., et al. (2018). Multi-Donor Longitudinal Antibody Repertoire Sequencing Reveals the Existence of Public Antibody Clonotypes in HIV-1 Infection. *Cell Host Microbe* **23**, 845–854.e6.
- Soto, C., Bombardi, R.G., Branchizio, A., Kose, N., Matta, P., Sevy, A.M., Sinkovits, R.S., Gilchuk, P., Finn, J.A., and Crowe, J.E., Jr. (2019). High frequency of shared clonotypes in human B cell receptor repertoires. *Nature* **566**, 398–402.
- Stamatatos, L., Pancera, M., and McGuire, A.T. (2017). Germline-targeting immunogens. *Immunol. Rev.* **275**, 203–216.
- Stiegler, G., Kunert, R., Purtscher, M., Wolbank, S., Voglauer, R., Steindl, F., and Katinger, H. (2001). A potent cross-clade neutralizing human monoclonal antibody against a novel epitope on gp41 of human immunodeficiency virus type 1. *AIDS Res. Hum. Retroviruses* **17**, 1757–1765.
- Stoeckius, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chattopadhyay, P.K., Swerdlow, H., Satija, R., and Smibert, P. (2017). Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868.
- Tan, Y.C., Kongpachith, S., Blum, L.K., Ju, C.H., Lahey, L.J., Lu, D.R., Cai, X., Wagner, C.A., Lindstrom, T.M., Sokolove, J., and Robinson, W.H. (2014). Barcode-enabled sequencing of plasmablast antibody repertoires in rheumatoid arthritis. *Arthritis Rheumatol.* **66**, 2706–2715.
- Walker, L.M., Phogat, S.K., Chan-Hui, P.-Y., Wagner, D., Phung, P., Goss, J.L., Wrinn, T., Simek, M.D., Fling, S., Mitcham, J.L., et al. (2009). Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. *Science* **326**, 285–289.
- Walker, L.M., Huber, M., Doores, K.J., Falkowska, E., Pejchal, R., Julien, J.P., Wang, S.K., Ramos, A., Chan-Hui, P.Y., Moyle, M., et al. (2011). Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature* **477**, 466–470.
- Wang, B., DeKosky, B.J., Timm, M.R., Lee, J., Normandin, E., Misasi, J., Kong, R., McDaniel, J.R., Delidakis, G., Leigh, K.E., et al. (2018). Functional interrogation and mining of natively paired human V_H:V_L antibody repertoires. *Nat. Biotechnol.* **36**, 152–155.
- Weaver, G.C., Villar, R.F., Kanekiyo, M., Nabel, G.J., Mascola, J.R., and Lingwood, D. (2016). In vitro reconstitution of B cell receptor-antigen interactions to evaluate potential vaccine candidates. *Nat. Protoc.* **11**, 193–213.
- Whittle, J.R.R., Wheatley, A.K., Wu, L., Lingwood, D., Kanekiyo, M., Ma, S.S., Narpala, S.R., Yassine, H.M., Frank, G.M., Yewdell, J.W., et al. (2014). Flow Cytometry Reveals that H5N1 Vaccination Elicits Cross-Reactive Stem-Directed Antibodies from Multiple Ig Heavy-Chain Lineages. *J. Virol.* **88**, 4047–4057.
- Wilson, P.C., and Andrews, S.F. (2012). Tools to therapeutically harness the human antibody response. *Nat. Rev. Immunol.* **12**, 709–719.
- Wu, X., Yang, Z.-Y., Li, Y., Hogerkerp, C.-M., Schief, W.R., Seaman, M.S., Zhou, T., Schmidt, S.D., Wu, L., Xu, L., et al. (2010). Rational Design of Envelope Identifies Broadly Neutralizing Human Monoclonal Antibodies to HIV-1. *Science* **329**, 856–861.
- Wu, X., Wang, C., O'Dell, S., Li, Y., Keele, B.F., Yang, Z., Imamichi, H., Doria-Rose, N., Hoxie, J.A., Connors, M., et al. (2012). Selection pressure on HIV-1 envelope by broadly neutralizing antibodies to the conserved CD4-binding site. *J. Virol.* **86**, 5844–5856.
- Wu, X., Zhang, Z., Schramm, C.A., Joyce, M.G., Kwon, Y.D., Zhou, T., Sheng, Z., Zhang, B., O'Dell, S., McKee, K., et al. (2015). Maturation and diversity of the VRC01-antibody lineage over 15 years of chronic HIV-1 infection. *Cell* **161**, 470–485.
- Zhang, S.-Q., Ma, K.-Y., Schonnesen, A.A., Zhang, M., He, C., Sun, E., Williams, C.M., Jia, W., and Jiang, N. (2018). High-throughput determination of the antigen specificities of T cell receptors in single cells. *Nat. Biotechnol.* **36**, 1156–1159.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
APC-Cy7 Mouse Anti-Human CD14	BD	Cat#561709; RRID: AB_10893806
FITC Anti-Human CD3 (OKT3)	Tonbo biosciences	Cat#35-0037; RRID: AB_2621662
PE-Cy5 Mouse Anti-Human IgG	BD	Cat#551497; RRID: AB_394220
V500 Mouse Anti-Human CD14	BD	Cat#561391; RRID: AB_10611856
BV711 Mouse Anti-Human CD19	BD	Cat#563036; RRID: AB_2737968
FITC Mouse Anti-Human IgG	BD	Cat#555786; RRID: AB_396121
APC-Cy7 Mouse Anti-Human CD3	BD	Cat#561800; RRID: AB_10551317
Anti-HIV-1 gp120 Monoclonal (VRC01)	NIH AIDS Reagent Program	Cat#12033; RRID: AB_2491019
Monoclonal anti-HIV-1 Env PGT145	NIH AIDS Reagent Program; https://www.hiv.lanl.gov/	Cat#12703; RRID: AB_2491054
Monoclonal anti-HIV-1 Env VRC34	VRC	Kong et al., 2016
Goat Anti-Human IgG (Fc Specific) Peroxidase	Sigma-Aldrich	Cat#A0170; RRID: AB_257868
Monoclonal anti-influenza Fe53	Daniel Lingwood	N/A
Monoclonal anti-HIV-1 Env 2723-2121	This paper	GenBank: MN580644 GenBank: MN580655
Monoclonal anti-HIV-1 Env 2723-2304	This paper	GenBank: MN580645 GenBank: MN580656
Monoclonal anti-HIV-1 Env 2723-422	This paper	GenBank: MN580646 GenBank: MN580657
Monoclonal anti-HIV-1 Env 2723-3055	This paper	GenBank: MN580560 GenBank: MN580589
Monoclonal anti-HIV-1 Env 2723-4186	This paper	GenBank: MN580550 GenBank: MN580579
Monoclonal anti-HIV-1 Env 2723-3131	This paper	GenBank: MN580558 GenBank: MN580587
Monoclonal anti-HIV-1 Env 2723-432	This paper	GenBank: MN580647 GenBank: MN580658
Monoclonal anti-HIV-1 Env 2723-3196	This paper	GenBank: MN580648 GenBank: MN580659
Monoclonal anti-HIV-1 Env 2723-1952	This paper	GenBank: MN580649 GenBank: MN580660
Monoclonal anti-HIV-1 Env 2723-2120	This paper	GenBank: MN580650 GenBank: MN580661
Monoclonal anti-influenza 2723-2859	This paper	GenBank: MN580651 GenBank: MN580662
Monoclonal anti-influenza 2723-3415	This paper	GenBank: MN580652 GenBank: MN580663
Monoclonal anti-influenza 3602-1707	This paper	GenBank: MN580653 GenBank: MN580664
Monoclonal anti-HIV-1 Env 3602-870	This paper	GenBank: MN580654 GenBank: MN580665
Anti-Human/Mouse C3/C3b/iC3b (FITC)	Cedarlane	Cat#CL7631F
Palivizumab	AstraZeneca	N/A
Anti-HIV Immune Globulin (HIVIG)	NIH AIDS Reagent Program	Cat#3957

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Goat Anti-Human IgG Antibody (Peroxidase)	Jackson ImmunoResearch	Cat#109-035-088; RRID: AB_2337584
Avi Tag Antibody, mAb, Mouse	GenScript	Cat#A01738-100
Bacterial and Virus Strains		
MW965.26	HIV Specimen Cryorepository (Germany)	201307035544P
MN.3	HIV Specimen Cryorepository (Germany)	201307035545P
6644.V2.C33	Duke University	N/A
TRO.11	NIH AIDS Reagent Program	Cat#11023
CH119.10	Duke University	N/A
CNE55	Duke University	N/A
25710.2.43	NIH AIDS Reagent Program	Cat#11505
CE0217 (CE703010217_B6)	Duke University	N/A
X1632.S2.B10	Duke University	N/A
X2278.C2.B6	Duke University	N/A
246.F3.C10.2	Duke University	N/A
398.F1.F6.20	Duke University	N/A
BJOX002000.03.2	Duke University	N/A
BG505.W6M.C2.T332N	NICD	N/A
CZA097 (97ZA012.12)	Duke University	N/A
ZM197	NIH AIDS Reagent Program	Cat#11309
MLV	NIH AIDS Reagent Program	Cat#1065
Biological Samples		
PBMC from Donor NIAID45	Mark Connors, NIH (Wu et al., 2015)	N/A
PBMC from NIAID Donor N90	Mark Connors, NIH (Wu et al., 2012)	N/A
PBMC from healthy donors	Simone Richardson	N/A
Complement from healthy donors	Simone Richardson	N/A
Chemicals, Peptides, and Recombinant Proteins		
BG505.SOSIP.664 gp140 trimer	Ivelin Georgiev	N/A
CZA97.SOSIP.664 gp140 trimer	Ivelin Georgiev	N/A
A/New Caledonia/20/99 (H1N1)	Barney Graham	GenBank: ACF41878
B41.SOSIP.664 gp140 trimer	Ivelin Georgiev	N/A
ZM197.SOSIP.664 gp140 trimer	Ivelin Georgiev	N/A
ZM106.9.SOSIP.664 gp140 trimer	Ivelin Georgiev	N/A
A/Michigan/45/2015 (H1N1) trimer	Barney Graham, Whittle et al., 2014	GenBank: AMA11475
A/Indonesia/5/2005 (H5N1) trimer	Barney Graham	GenBank: ABP51969
A/Anhui/1/2013 (H7N9) trimer	Barney Graham	GISAID EPI439507
BG505 DS-SOSIP trimer	VRC	Kwon et al., 2015
Streptavidin R-phycoerythrin (SA-PE)	Invitrogen	S866
Carboxyfluorescein diacetate N-succinimidyl ester (CFSE)	Sigma-Aldrich	Cat#21888
ConC gp120 WT monomer	Lynn Morris	N/A
ConC gp120 D368R monomer	Lynn Morris	N/A
Protein A Resin	GenScript	Cat# L00210
Polyethylenimine Linear MW 25000	Polysciences	23966-1
Gelatin Veronal buffer (0.1%)	Sigma-Aldrich	Cat#G6514
DMEM	Thermo Fisher Scientific	Cat#41966029
Hyclone Fetal Bovine Serum (FBS)	Thermo Fisher Scientific	Cat#SV30160.03

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Gentamycin	Sigma-Aldrich	Cat#G1272
HEPES Buffer	Thermo Fisher Scientific	Cat#15630056
Diethylaminoethyl-dextran hydrochloride	Sigma-Aldrich	Cat#D9885
Trypsin-EDTA	Thermo Fisher Scientific	Cat#25200056
Phosphate Buffered Saline (PBS)	Thermo Fisher Scientific	Cat#10010015
Trypan Blue Solution (0.4%)@	Sigma-Aldrich	Cat#T8154
Bright-Glo Luciferase Assay System	Promega	Cat# E2650
Ghost Dye Red 780	Tonbo biosciences	Cat#13-0865
Live/Dead Fixable Aqua Dead Cell Stain Kit	Thermo Fisher Scientific	Cat#L34957
Critical Commercial Assays		
PKH26 Red Fluorescent Cell Linker Mini Kit	Sigma-Aldrich	Cat#MINI26
FluoSpheres NeutrAvidin-Labeled Microspheres, 1.0 μm, yellow-green fluorescent	Thermo Fisher Scientific	Cat#F8776
GranToxiLux Kit	Oncolmmunin	N/A
EZ link Sulfo-NHS-LC-Biotin	Thermo Fisher Scientific	Cat#21327
BirA-500: Bir A biotin-protein ligase standard reaction kit	Avidity	Cat# BirA500
Solulink Protein-Oligonucleotide Conjugation Kit	TriLink Biotechnologies	Cat# S-9011
B cell Single Cell V(D)J solution	10X Genomics	N/A
Sensor Chip CM5	GE Healthcare	Cat # 29-1496-03
Human Antibody Capture Kit	GE Healthcare	Cat # BR-1008-39
Deposited Data		
NIAD45-2723 heavy chain sequences	This paper	GenBank: MN580550 - MN580578
NIAD45-2723 light chain sequences	This paper	GenBank: MN580579 - MN580607
N90-3602 heavy chain sequences	This paper	GenBank: MN580608 - MN580625
N90-3602 light chain sequences	This paper	GenBank: MN580626 - MN580643
Raw next-generation sequencing data	This paper	SRA: PRJNA578389
Experimental Models: Cell Lines		
Human: VRC01 Ramos B Cells	Daniel Lingwood	N/A
Human: Fe53 Ramos B Cells	Daniel Lingwood	N/A
Human CEM.NKR CCR5+ Cells	AIDS Reagent Program	Cat#4376
Human THP-1 Cells	AIDS Reagent Program	Cat#9942
Human: Freestyle 293F cells	ThermoFisher Scientific	Cat#A14528
Human: Expi293F cells	ThermoFisher Scientific	Cat#A14527
Human: Expi293 GnT1- cells	Invitrogen	N/A
TZM-bl Cell-line	NIH AIDS Reagent Program	Cat#8129
Oligonucleotides		
Oligonucleotides for Protein DNA-barcoding	This paper	N/A
Software and Algorithms		
Cell Ranger version 2.2.0	10X Genomics	https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/2.2
Fastp	Chen et al., 2018	https://github.com/OpenGene/fastp
HighV-Quest	Alamyar et al., 2012	http://www.imgt.org/IMGIndex/IMGTHighV-QUEST.php
ChangeO	Gupta et al., 2015	https://changeo.readthedocs.io/en/stable/
PhyML Maximum Likelihood Geneious plugin	Guindon et al., 2010	https://www.geneious.com/plugins/phyml-plugin/

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Geneious 11.1.5	https://www.geneious.com/	N/A
Cytobank	Kotecha et al., 2010	https://www.cytobank.org/
iTol	Letunic and Bork, 2019	https://itol.embl.de/
Flowjo v10	TreeStar	https://www.flowjo.com/
GraphPad Prism 8.0.0	https://www.graphpad.com	N/A
Biacore T200 control software	GE Healthcare	N/A
Biacore T200 evaluation software	GE Healthcare	N/A
Other		
Galanthus nivalis lectin	Vector Laboratories	Cat # AL-1243-5
1-Step Ultra TMB-ELISA Substrate Solution	ThermoFisher Scientific	Cat#34029
Costar High Binding Microplates	Corning	Cat#9018

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to the Lead Contact, Ivelin Georgiev (Ivelin.Georgiev@Vanderbilt.edu). Antibody plasmids generated in this study are available from the Lead Contact with a completed Materials Transfer Agreement.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human Subjects

Donor NIAID45: Peripheral blood mononuclear cells were collected from donor NIAID45 on July 12, 2007. Donor NIAID45 was enrolled in investigational review board approved clinical protocols at the National Institute of Allergy and Infectious Diseases (NIAID) and had been living with HIV without antiretroviral treatment for approximately 17 years at the time of sample collection. *Donor N90*: Peripheral blood mononuclear cells were collected from donor N90 on May 29, 2008. Donor N90 was enrolled in investigational review board approved clinical protocols at NIAID and had been living with HIV without antiretroviral treatment through the time point of sample collection since diagnosis in 1985 ([Wu et al., 2012](#)).

Cell Lines

Ramos B cell lines were engineered from a clone of Ramos Burkitt's lymphoma that do not display endogenous antibody, and they ectopically express specific surface IgM B cell receptor sequences. The B cell lines used expressed B cell receptor sequences for HIV-specific antibody VRC01 and influenza-specific antibody Fe53. The cells are cultured at 37°C with 5% CO₂ saturation in complete RPMI, made up of RPMI supplemented with 15% fetal bovine serum, 1% L-Glutamine, and 1% Penicillin/Streptomycin. Although endogenous heavy chains are scrambled, endogenous light chain transcripts remain and are detectable by sequencing. We thus identified and classified single Ramos Burkitt's B cells as either VRC01 or Fe53 based on their heavy chain sequences. These Ramos B cell lines were validated for binding to our antigen probes by FACS ([Figure S1B](#)).

METHOD DETAILS

Antigen expression and purification

For the different LIBRA-seq experiments, a total of six HIV-1 gp140 SOSIP variants from strains BG505 (clade A), CZA97 (clade C), B41 (clade B), ZM197 (clade C), ZM106.9 (clade C), KNH1144 (clade A) and four influenza hemagglutinin variants from strains A/New Caledonia/20/99 (H1N1) (GenBank ACF41878), A/Michigan/45/2015 (H1N1) (GenBank AMA11475), A/Indonesia/5/2005 (H5N1) (GenBank ABP51969), and A/Anhui/1/2013 (H7N9) (GISAID EPI439507) were expressed as recombinant soluble antigens.

The single-chain variants ([Georgiev et al., 2015](#)) of BG505, CZA97, B41, ZM197, ZM106.9, and KNH1144 each containing an AviTag, were expressed in FreeStyle 293F mammalian cells (ThermoFisher) using polyethylenimine (PEI) transfection reagent and cultured for 5-7 days. FreeStyle 293F were maintained in FreeStyle 293F medium or FreeStyle F17 expression medium supplemented with 1% of 10% Pluronic F-68 and 20% of 200 mM L-Glutamine. These cells were cultured at 37°C with 8% CO₂ saturation and shaking. After transfection and 5-7 days of culture, cultures were centrifuged at 6000 rpm for 20 minutes. Supernatant was filtered with Nalgene Rapid Flow Disposable Filter Units with PES membrane (0.45 μm), and then run slowly over an affinity column of agarose bound *Galanthus nivalis* lectin (Vector Laboratories cat no. AL-1243-5) at 4°C. The column was washed with PBS, and proteins were eluted with 30 mL of 1 M methyl-α-D-mannopyranoside. The protein elution was buffer exchanged 3X into PBS and

concentrated using 30kDa Amicon Ultra centrifugal filter units. Concentrated protein was run on a Superose 6 Increase 10/300 GL or Superdex 200 Increase 10/300 GL sizing column on the AKTA FPLC system, and fractions were collected on an F9-R fraction collector. Fractions corresponding to correctly folded antigen were analyzed by SDS-PAGE, and antigenicity by ELISA was characterized with known monoclonal antibodies specific for that antigen.

Recombinant HA proteins all contained the HA ectodomain with a point mutation at the sialic acid-binding site (Y98F), T4 fibrin foldon trimerization domain, AviTag, and hexahistidine-tag, and were expressed in Expi 293F mammalian cells using Expifectamine 293 transfection reagent (Thermo Fisher Scientific) cultured for 4-5 days. Culture supernatant was harvested and cleared as above, and then adjusted pH and NaCl concentration by adding 1M Tris-HCl (pH 7.5) and 5M NaCl to 50 mM and 500 mM, respectively. Ni Sepharose excel resin (GE Healthcare) was added to the supernatant to capture hexahistidine tag. Resin was separated on a column by gravity and captured HA protein was eluted by a Tris-NaCl (pH 7.5) buffer containing 300 mM imidazole. The eluate was further purified by a size exclusion chromatography with a HiLoad 16/60 Superdex 200 column (GE Healthcare). Fractions containing HA were concentrated, analyzed by SDS-PAGE and tested for antigenicity by ELISA with known antibodies. Proteins were frozen in LN₂ and stored at -80°C until use.

All HIV gp140 SOSIP variant antigens and all influenza hemagglutinin variant antigens included an AviTag modification at the C terminus of their sequence, and after purification, each AviTag labeled antigen was biotinylated using the BirA-500: BirA biotin-protein ligase standard reaction kit (Avidity LLC, cat no. BirA500).

Oligonucleotide barcodes

We used oligos that possess a 13-15 bp antigen barcode, a sequence capable of annealing to the template switch oligo that is part of the 10X bead-delivered oligos, and contain truncated TruSeq small RNA read 1 sequences in the following structure: 5'-CCTTGGCACCCGAGAATTCCANNNNNNNNNNNNNCCCATATAAGA*A*A-3', where Ns represent the antigen barcode. For the cell line and NIAID45 experiments, we used the following antigen barcodes: CATGATTGGCTCA (BG505), TGTCGGCAATAA (CZA97), GATCGTAATACCA (H1 A/New Caledonia/20/99). For the N90 experiment, we used longer antigen barcodes (15 bp), as follows: TCCTTTCCTGATAGG (ZM106.9), TAACTCAGGGCCTAT (KNH1144), GCTCCTTTACACGTA (ZM197), GCAGCGTATA AGTCA (B41), ATCGTCGAGAGCTAG (BG505), CAGGTCCCTTATTTT (A/Indonesia/5/2005), ACAATTTGTCTGCGA (A/Anhui/1/2013), TGACCTTCTCTCCT (A/Michigan/45/2015), AATCACGGTCCCTTGT (A/New Caledonia/20/99). Oligos were ordered from Sigma-Aldrich and IDT with a 5' amino modification and HPLC purified.

Conjugation of oligonucleotide barcodes to antigens

For each antigen, a unique DNA barcode was directly conjugated to the antigen itself. In particular, 5' amino-oligonucleotides were conjugated directly to each antigen using the Solulink Protein-Oligonucleotide Conjugation Kit (TriLink cat no. S-9011) according to manufacturer's instructions. Briefly, the oligo and protein were desalted, and then the amino-oligo was modified with the 4FB cross-linker, and the biotinylated antigen protein was modified with S-HyNic. Then, the 4FB-oligo and the HyNic-antigen were mixed together. This causes a stable bond to form between the protein and the oligonucleotide. The concentration of the antigen-oligo conjugates was determined by a BCA assay, and the HyNic molar substitution ratio of the antigen-oligo conjugates was analyzed using the NanoDrop according to the Solulink protocol guidelines. AKTA FPLC was used to remove excess oligonucleotide from the protein-oligo conjugates, which were also checked using SDS-PAGE with a silver stain.

Fluorescent labeling of antigens

After attaching DNA barcodes directly to a biotinylated antigen, the barcoded antigens were mixed with streptavidin labeled with fluorophore phycoerythrin (PE). The streptavidin-PE was mixed with biotinylated antigen at a 5X molar excess of antigen to streptavidin. 1/5 of the streptavidin-oligo conjugate was added to the antigen every 20 minutes with constant rotation at 4°C.

Enrichment of antigen-specific B cells

For a given sample, cells were stained and mixed with fluorescently labeled DNA-barcoded antigens and other antibodies, and then sorted using fluorescence activated cell sorting (FACS). First, cells were counted and viability was assessed using Trypan Blue. Then, cells were washed with DPBS supplemented with 1% Bovine serum albumin (BSA) through centrifugation at 300 g for 7 minutes. Cells were resuspended in PBS-BSA and stained with a variety of cell markers. For donor NIAID45 PBMCs, these markers included CD3-APCCy7, IgG-FITC, CD19-BV711, CD14-V500, and LiveDead-V500. Additionally, fluorescently labeled antigen-oligo conjugates (described above) were added to the stain. For donor N90 PBMCs, these markers included Ghost Red 780, CD14-APCCy7, CD3-FITC, CD19-BV711, and IgG-PECy5. Additionally, fluorescently labeled antigen-oligo conjugates were added to the stain. After staining in the dark for 30 minutes at room temperature, cells were washed 3 times with PBS-BSA at 300 g for 7 minutes. Then, cells were resuspended in DPBS and sorted on the cell sorter. Antigen positive cells were bulk sorted and then they were delivered to the Vanderbilt VANTAGE sequencing core at an appropriate target concentration for 10X Genomics library preparation and subsequent sequencing. FACS data were analyzed using Cytobank (Kotecha et al., 2010).

10X Genomics single cell processing and next generation sequencing

Single-cell suspensions were loaded onto the Chromium Controller microfluidics device (10X Genomics) and processed using the B cell Single Cell V(D)J solution according to manufacturer's suggestions for a target capture of 10,000 B cells per 1/8 10X cassette for B cell lines, 9,000 cells for B cells from donor NIAID45, and 4,000 for donor N90, with minor modifications in order to intercept, amplify and purify the antigen barcode libraries. The library preparation followed the CITE-seq protocol (available at <https://cite-seq.com>), with the exception of an increase in the number of PCR cycles of the antigen barcodes. Briefly, following cDNA amplification using an additive primer (5'-CCTTGGCACCCGAGAATT*C*C-3') to increase the yield of antigen barcode libraries (Stoeckius et al., 2017), SPRI separation was used to size separate antigen barcode libraries from cellular mRNA libraries, PCR amplified for 10-12 cycles, and purified using 1.6X purification. Sample preparation for the cellular mRNA library continued according to 10X Genomics-suggested protocols, resulting in Illumina-ready libraries. Following library construction, we sequenced both BCR and antigen barcode libraries on a NovaSeq 6000 at the VANTAGE sequencing core, dedicating ~2.5% of a flow cell to each experiment, with a target 10% of this fraction dedicated to antigen barcode libraries. This resulted in ~334.5 million reads for the cell line V(D)J libraries (~96,500 reads/cell), ~376.3 million reads for donor NIAID45 V(D)J libraries (~79,300 reads/cell), and ~272.4 million reads for the N90 V(D)J libraries (~151,400 reads/cell). The N90 antigen barcode libraries were also sequenced a second time.

Processing of BCR sequence and antigen barcode reads

We developed a pipeline that takes paired-end FASTQ files of oligo libraries as input, processes and annotates reads for cell barcode, UMI, and antigen barcode, and generates a cell barcode - antigen barcode UMI count matrix. BCR sequence reads were processed using Cell Ranger version 2.2.0 (10X Genomics) using GRCh38 as reference. For the antigen barcode libraries, initial quality and length filtering was carried out by fastp (Chen et al., 2018) using default parameters for filtering (Figure S6A). In a histogram of insert lengths, this resulted in a sharp peak of the expected insert size of 52-54 bp (Figures S6B-S6D). Fastx_collapser was then used to group identical sequences and convert the output to deduplicated FASTA files. We proceeded to process just the R2 sequences, as the entire insert is present in both R1 and R2. Each unique R2 sequence was processed one-by-one using the following steps: (1) The reverse complement of the R2 sequence was determined. (2) The sequence was screened for possessing an exact match to any of the valid 10X cell barcodes present in the filtered_contig.fasta file output by Cell Ranger during processing of BCR V(D)J FASTQ files. Sequences without a BCR-associated cell barcode were discarded. (3) The 10 bases immediately 3' to the cell barcode were annotated as the read UMI. (4) The remainder of the sequence 3' to the UMI was screened for a 13 or 15 bp sequence within a hamming distance of 2 to any of the antigen barcodes used in the screening library. Following this processing, only sequences with lengths of 51 to 58 were retained. After processing each sequence one-by-one, we screened for cell barcode - UMI - antigen barcode collisions. Any cell barcode - UMI combination that had multiple antigen barcodes associated with it was removed. We then constructed a cell barcode - antigen barcode UMI count matrix, which served as the basis of subsequent analysis. Additionally, we aligned the BCR contigs (filtered_contigs.fasta file output by Cell Ranger, 10X Genomics) to IMGT reference genes using HighV-Quest (Alamyar et al., 2012). The output of HighV-Quest was parsed using Change-O (Gupta et al., 2015), and merged with the UMI count matrix.

Determination of LIBRA-seq Score

Starting with the UMI count matrix, we set all counts of 1, 2, or 3 UMIs to 0, with the idea that these low counts could likely be attributed to noise. After this, the UMI count matrix was subset to contain only cells with a count of at least 4 UMIs for at least 1 antigen. We also removed cells that had only non-functional heavy chain sequences as well as cells with multiple functional heavy chain sequences using different *IGHV* genes, reasoning that these may be multiplets. We then calculated the centered-log ratios (CLR) of each antigen UMI count for each cell (Mimitou et al., 2019; Stoeckius et al., 2017). Because UMI counts were on different scales for each antigen, possibly due to differential oligo loading during oligo-antigen conjugation, we rescaled the CLR UMI counts using the StandardScaler method in *scikit learn* (Pedregosa and Varoquaux, 2011). Lastly, we performed a correction procedure to the scaled CLR from UMI counts of 0, setting them to the minimum for each antigen for donor NIAID45 and N90 experiments, and to -1 for the Ramos B cell line experiment. These CLR-transformed, scaled, corrected values served as the final LIBRA-seq scores. LIBRA-seq scores were visualized using Cytobank (Kotecha et al., 2010) and Matplotlib (Hunter, 2007). Cells with a LIBRA-seq score of 1 or greater for donor N90 data were also visualized using UpSet plots (Lex et al., 2014) using the *UpSetPlot* package in Python. Donor NIAID45 and N90 data were subsetted to include only cells with a functional light chain.

Phylogenetic trees

Phylogenetic trees of antibody heavy chain sequences were constructed in order to assess the relatedness of antibodies within a given lineage. For the VRC01 lineage, the 29 sequences identified by LIBRA-seq and 53 sequences identified from the literature were aligned using clustal within Geneious. We then used the PhyML maximum likelihood (Guindon et al., 2010) plugin in Geneious (available at <https://www.geneious.com/plugins/phyml-plugin/>) to infer a phylogenetic tree. The resulting tree was then rooted to the inferred unmutated common ancestor (Bonsignori et al., 2018) (accession GenBank: MK032222). Names for sequences and their accession include the following: H01+07.F.1 (GenBank: KP840594); H03+06.C.1 (GenBank: KP840597); H03+06.E.1 (GenBank: KP841560); H4.E.6 (GenBank: KP841696); H4.E.5 (GenBank: KP841700); H4.E.4 (GenBank: KP841639); H4.E.3 (GenBank: KP841608); H4.E.2 (GenBank: KP841609); H4.E.1 (GenBank: KP841701); H5.C.1 (GenBank: KP840607); H5.F.1 (GenBank: KP840608); H08.F.1 (GenBank: KP840603); H08.H.1 (GenBank: KP840835); VRC03b (GenBank: KP840671); VRC03f

(GenBank: KP840674); VRC03 g (GenBank: KP840675); DH651.1 (GenBank: MK032223); DH651.3 (GenBank: MK032225); DH651.9 (GenBank: MK032231); DH651.8 (GenBank: MK032230); VRC06c (GenBank: KP840678); VRC06d (GenBank: KP840679); VRC06e (GenBank: KP840680); VRC06f (GenBank: KP840681); VRC06 g (GenBank: KP840682); VRC06h (GenBank: KP840683); DH651.2 (GenBank: MK032224); DH651.4 (GenBank: MK032226); DH651.5 (GenBank: MK032227); DH651.6 (GenBank: MK032228); DH651.7 (GenBank: MK032229); VRC06 (GenBank: JX466923.1); VRC03 (GenBank: GU980706.1); NIH45-46 (GenBank: HE584543); VRC01 (GenBank: GU980702); VRC01c (GenBank: KP840658); VRC01d (GenBank: KP840659); VRC01e (GenBank: KP840660); VRC01f (GenBank: KP840661); VRC01h (GenBank: KP840663); VRC01i (GenBank: KP840664); VRC01j (GenBank: KP840665); VRC02 (GenBank: GU980704); VRC07b (GenBank: KP840666); VRC07c (GenBank: KP840667); VRC07d (GenBank: KP840668); VRC07e (GenBank: KP840669); VRC07f (GenBank: KP840670); VRC08c (GenBank: KP840685); VRC08d (GenBank: KP840686); VRC08e (GenBank: KP840687); H03+06.A.0 (GenBank: KP841501); VRC01UCA (GenBank: MK032222). A similar process was used to build a phylogenetic tree for the VRC38 lineage, with one exception. Rather than using an inferred germline precursor, we germline-reverted framework 1, CDR1, framework 2, CDR2, framework 3, and framework 4 and used the junction nucleotide sequence of the lineage member with the least *IGHV* somatic mutation (VRC38.03). Trees were annotated and visualized in iTol (Letunic and Bork, 2019). While trees were constructed based on heavy chains, all VRC01 and VRC38 B cells had a correct light chain transcript, although sometimes additional light chain transcripts were also observed. One LIBRA-seq-identified VRC38 lineage member, 3602-1544, contained a single nucleotide deletion in the Cell Ranger-determined contig sequence in framework 2; this was manually corrected prior to inferring the phylogenetic tree.

Antibody expression and purification

For each antibody, variable genes were inserted into plasmids encoding the constant region for the heavy chain (pFUSEss-CHlg-hG1, Invivogen) and light chain (pFUSE2ss-CLlg-hl2, Invivogen and pFUSE2ss-CLlg-hk Invivogen) and synthesized from GenScript. mAbs were expressed in FreeStyle 293F or Expi293F mammalian cells (ThermoFisher) by co-transfecting heavy chain and light chain expressing plasmids using polyethylenimine (PEI) transfection reagent and cultured for 5-7 days. FreeStyle 293F (ThermoFisher) and Expi293F (ThermoFisher) cells were maintained in FreeStyle 293F medium or FreeStyle F17 expression medium supplemented with 1% of 10% Pluronic F-68 and 20% of 200 mM L-Glutamine. These cells were cultured at 37°C with 8% CO₂ saturation and shaking. After transfection and 5-7 days of culture, cell cultures were centrifuged at 6000 rpm for 20 minutes. Supernatant was 0.45 μm filtered with Nalgene Rapid Flow Disposable Filter Units with PES membrane. Filtered supernatant was run over a column containing Protein A agarose resin that had been equilibrated with PBS. The column was washed with PBS, and then antibodies were eluted with 100 mM Glycine HCl at pH 2.7 directly into a 1:10 volume of 1 M Tris-HCl pH 8. Eluted antibodies were buffer exchanged into PBS 3 times using 10kDa Amicon Ultra centrifugal filter units.

Enzyme linked immunosorbent assay (ELISA)

For hemagglutinin ELISAs, soluble hemagglutinin protein was plated at 2 μg/ml overnight at 4°C. The next day, plates were washed three times with PBS supplemented with 0.05% Tween20 (PBS-T) and coated with 5% milk powder in PBS-T. Plates were incubated for one hour at room temperature and then washed three times with PBS-T. Primary antibodies were diluted in 1% milk in PBS-T, starting at 10 μg/ml with a serial 1:5 dilution and then added to the plate. The plates were incubated at room temperature for one hour and then washed three times in PBS-T. The secondary antibody, goat anti-human IgG conjugated to peroxidase, was added at 1:20,000 dilution in 1% milk in PBS-T to the plates, which were incubated for one hour at room temperature. Plates were washed three times with PBS-T and then developed by adding TMB substrate to each well. The plates were incubated at room temperature for ten minutes, and then 1 N sulfuric acid was added to stop the reaction. Plates were read at 450 nm.

For recombinant trimer capture for single-chain SOSIPs, 2 μg/ml of a mouse anti-AviTag antibody (GenScript) was coated overnight at 4°C in phosphate-buffered saline (PBS) (pH 7.5). The next day, plates were washed three times with PBS-T and blocked with 5% milk in PBS-T. After an hour incubation at room temperature and three washes with PBS-T, 2 μg/ml of recombinant trimer proteins diluted in 1% milk PBS-T were added to the plate and incubated for one hour at room temperature. Primary and secondary antibodies, along with substrate and sulfuric acid, were added as described above. Data are represented as mean ± SEM for one ELISA experiment. ELISAs were repeated 2 or more times. The area under the curve (AUC) was calculated using GraphPad Prism 8.0.0.

TZM-bl Neutralization Assays

Antibody neutralization was assessed using the TZM-bl assay as described (Sarzotti-Kelsoe et al., 2014). This standardized assay measures antibody-mediated inhibition of infection of JC53BL-13 cells (also known as TZM-bl cells) by molecularly cloned Env-pseudoviruses. Viruses that are highly sensitive to neutralization (Tier 1) and/or those representing circulating strains that are moderately sensitive (Tier 2) were included, plus additional viruses, including a subset of the antigens used for LIBRA-seq. Murine leukemia virus (MLV) was included as an HIV-specificity control and VRC01 was used as a positive control. Results are presented as the concentration of monoclonal antibody (in μg/ml) required to inhibit 50% of virus infection (IC₅₀).

Surface Plasmon Resonance and Fab competition

HIV-1 Env BG505 DS-SOSIP was produced either in GnT1- or 293F cells and purified as described previously (Kwon et al., 2015). The binding of antibodies 2723-2121 and 3602-870 to BG505 DS-SOSIP was assessed by surface plasmon resonance (SPR) on Biacore T-200 (GE-Healthcare) at 25°C with HBS-EP+ (10 mM HEPES, pH 7.4, 150 mM NaCl, 3 mM EDTA, and 0.05% surfactant P-20) as the running buffer. Antibodies VRC01 and PGT145 were tested as positive control, and antibody 17b was tested as negative control to confirm that the trimer was in the closed conformation. Antibodies 2723-2121 and 3602-870 were captured on a flow cell of CM5 chip immobilized with ~9000 RU of anti-human Fc antibody, and binding was measured by flowing over a 200 nM solution BG505-DS SOSIP in running buffer. Similar runs were performed with VRC01, PGT145 and 17b IgGs. To determine their epitopes, antibodies 2723-2121 IgG and 3602-870 were captured on a single flow cell of CM5 chip immobilized with anti-human Fc antibody. Next 200 nM BG505 DS-SOSIP, either alone or with different concentrations of antigen binding fragments (Fab) of VRC01 or PGT145 or VRC34 was flowed over the captured 2723-2121 or 3602-870 flow cell for 60 s at a rate of 10 μ l/min. The surface was regenerated between injections by flowing over 3M MgCl₂ solution for 10 s with flow rate of 100 μ l/min. Blank sensorgrams were obtained by injection of same volume of HBS-EP+ buffer in place of trimer with Fabs solutions. Sensorgrams of the concentration series were corrected with corresponding blank curves.

ADCP, ADCD, Trogocytosis, ADCC Assays

Antibody-dependent cellular phagocytosis (ADCP) was performed using gp120 ConC coated neutravidin beads as previously described (Ackerman et al., 2011). Phagocytosis score was determined as the percentage of cells that took up beads multiplied by the fluorescent intensity of the beads. Antibody-dependent complement deposition (ADCD) was performed as in Richardson et al. (2018a) where CEM.NKR.CCR5 gp120 ConC coated target cells were opsonized with mAb and incubated with complement from a healthy donor. C3b deposition was then determined by flow cytometry with complement deposition score determined as the percentage of C3b positive cells multiplied by the fluorescence intensity. Antibody-dependent cellular trogocytosis (ADCT) was measured as the percentage transfer of PKH26 dye of the surface of CEM.NKR.CCR5 target cells to CSFE stained monocytic cell line THP-1 cells in the presence of HIV specific mAbs as described elsewhere (Richardson et al., 2018b). Antibody-dependent cellular cytotoxicity (ADCC) was done using a GranToxiLux based assay (Pollara et al., 2011) with gp120 ConC coated CEM.NKR.CCR5 target cells and PBMCs from a healthy donor. The percentage of granzyme B present in target cells was measured by flow cytometry.

QUANTIFICATION AND STATISTICAL ANALYSIS

ELISA error bars (standard error of the mean) were calculated using GraphPad Prism version 8.0.0. The Pearson's *r* value comparing BG505 and CZA97 LIBRA-seq scores for Ramos B cell lines was calculated using Cytobank. Spearman correlations and associated *p* values were calculated using *SciPy* in Python.

DATA AND CODE AVAILABILITY

Data Availability Statement

Raw sequencing data used in this study are available on the Sequence Read Archive under BioProject accession number SRA: PRJNA578389.

Identified antibody sequences related to the VRC01 lineage have been deposited to GenBank under accession codes GenBank: MN580550 – MN580578 (heavy chain) and GenBank: MN580579 – MN580607 (light chain). Identified antibody sequences related to the VRC38 lineage have been deposited to GenBank under accession codes GenBank: MN580608 – MN580625 (heavy chain) and GenBank: MN580626 – MN580643 (light chain). Other sequences from antibodies identified and recombinantly produced as part of this study have been deposited to GenBank under accession codes GenBank: MN580644 – MN580654 (heavy chain) and GenBank: MN580655 – MN580665 (light chain).

Code Availability Statement

Custom scripts used to analyze data in this manuscript are available upon request to the Lead Contact.

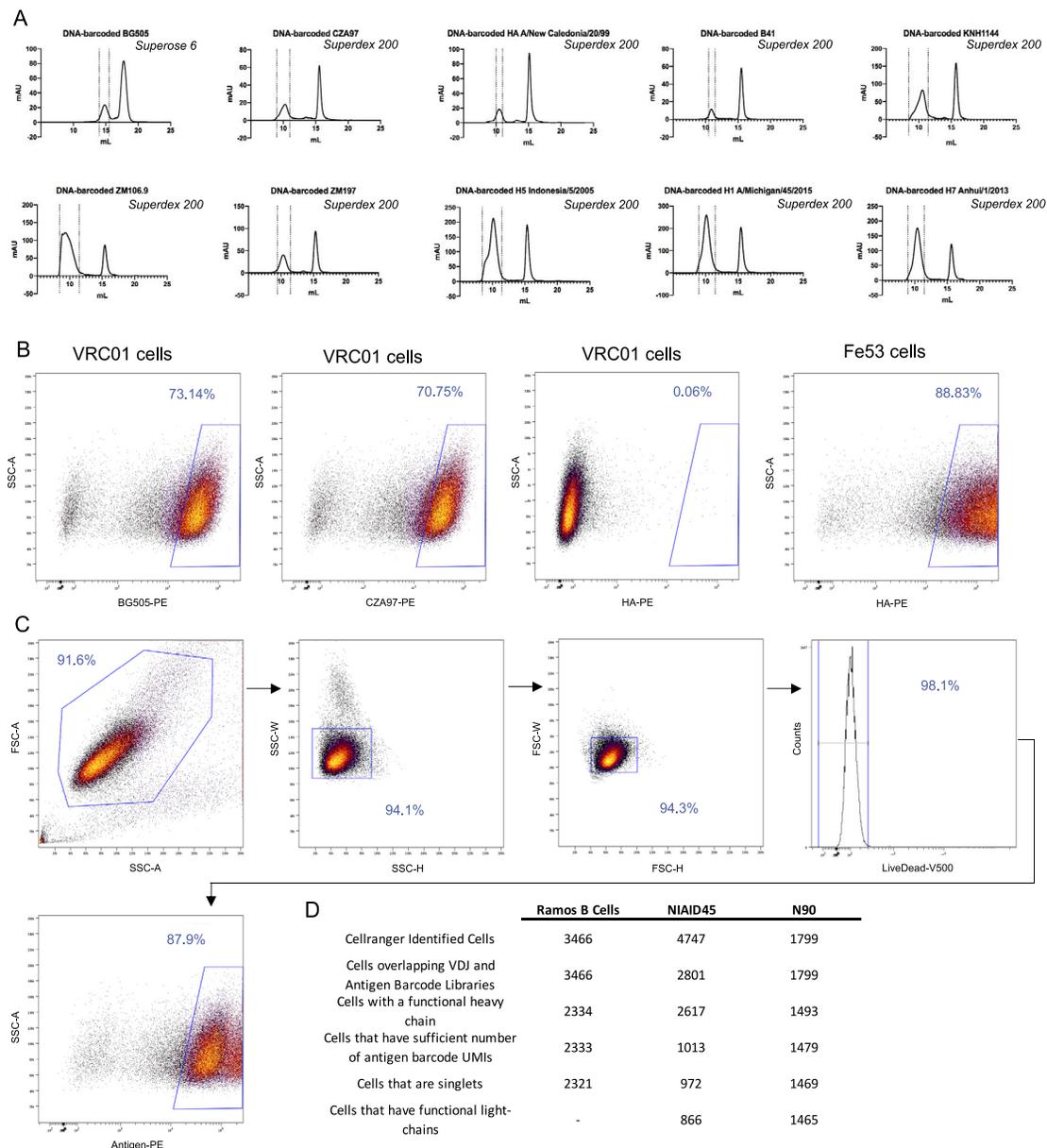


Figure S1. Purification of DNA-Barcoded Antigens and LIBRA-seq Validation Sorting Schematic on Ramos B Cell Lines, Related to Figures 1, 2, and 4 and STAR Methods

A. After barcoding each antigen with a unique oligonucleotide, antigen-oligo complexes are run on size exclusion chromatography to remove excess, unconjugated oligonucleotide from the reaction mixture. DNA-barcoded BG505 was run on the Superose 6 Increase 10/300 GL column and all other DNA-barcoded antigens were run on the Superdex 200 Increase 10/300 GL on the AKTA FPLC system. For size exclusion chromatography, dotted lines indicate DNA-barcoded antigens and fractions taken. The second peak indicates excess oligonucleotide from the conjugation reaction.

B. Binding of VRC01 or Fe53 Ramos B cell lines to DNA-barcoded, fluorescently labeled antigens via flow cytometry. VRC01 cells bound to DNA-barcoded BG505-PE, DNA-barcoded CZA97-PE, and not DNA-barcoded H1 A/New Caledonia/20/99-PE. Fe53 cells bound to DNA-barcoded H1 A/New Caledonia/20/99-PE.

C. Gating scheme for fluorescence activated cell sorting of Ramos B cell lines. VRC01 and Fe53 Ramos B cells were mixed in a 1:1 ratio and then stained with LiveDead-V500 and a DNA-barcoded antigen screening library consisting of BG505-PE, CZA97-PE, and H1 A/New Caledonia/20/99-PE. Gates as drawn are based on gates used during the sort, and percentages from the sort are listed.

D. For each experiment, the categorization of the number of Cell Ranger-identified (10X Genomics) cells after sequencing is shown. Each category (row) is a subset of cells of the previous category (row).

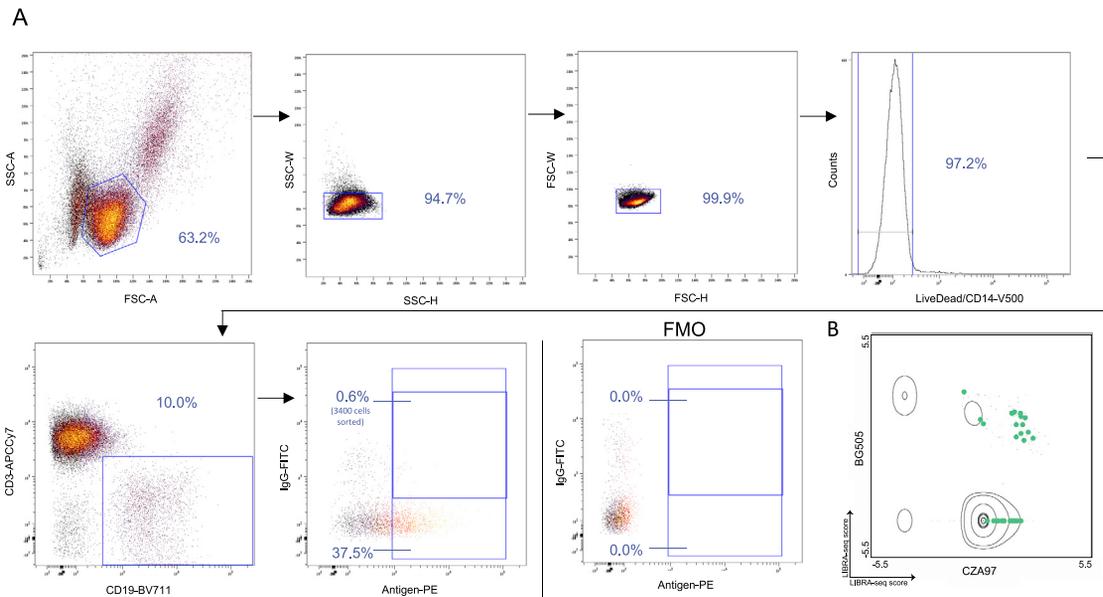


Figure S2. Identification of Antigen-Specific B Cells from Donor NIAID45 PBMCs, Related to Figures 2 and 3

A. Gating scheme for fluorescence activated cell sorting of donor NIAID45 PBMCs. Cells were stained with LiveDead-V500, CD14-V500, CD3-APCCy7, CD19-BV711, IgG-FITC, and a DNA-barcoded antigen screening library consisting of BG505-PE, CZA97-PE, and H1 A/New Caledonia/20/99-PE. Gates as drawn are based on gates used during the sort, and percentages from the sort are listed. These plots show a starting number of 50,187 total events. Due to the visualization parameters, 18 IgG-positive, antigen-positive cells are displayed, but 3400 IgG-positive, antigen-positive cells were sorted and supplemented with 13,000 antigen-positive B cells for single-cell sequencing. A small aliquot of donor NIAID45 PBMCs were used for fluorescence minus one (FMO) staining, and were stained with the same antibody panel as listed above with the exception of the HIV-1 and influenza antigens.

B. LIBRA-seq scores for BG505 (y axis) and CZA97 (x axis) are shown. Each axis represents the range of LIBRA-seq scores for each antigen. Density of total cells is shown. Overlaid on the density plot are the 29 VRC01 lineage members (dots) indicated in light green.

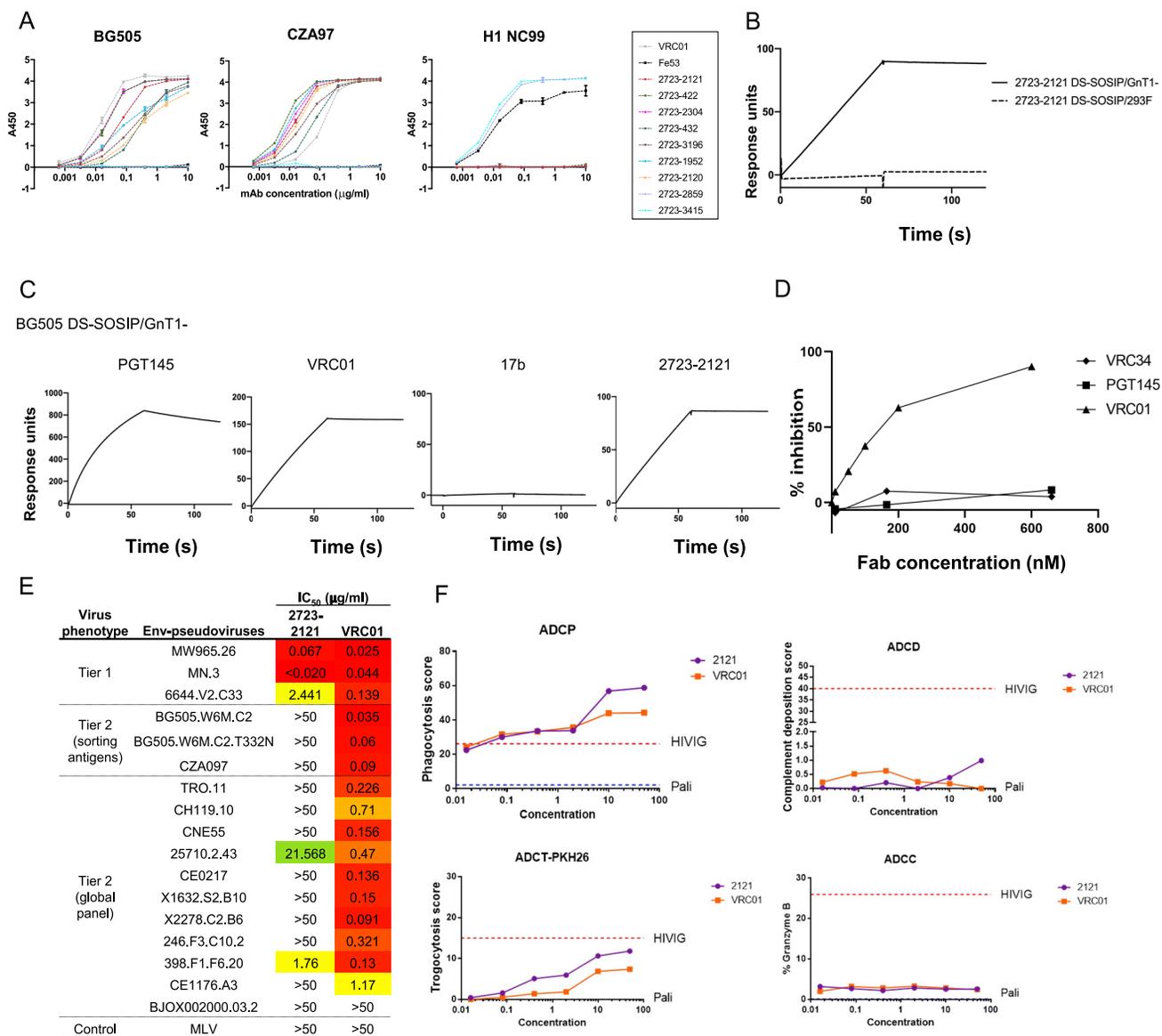


Figure S3. Characterization of Antibodies from Donor NIAID45, Related to Figure 3

A. Antigen specificity as predicted by LIBRA-seq was validated by ELISA for a variety of antibodies isolated from donor NIAID45. Antibodies were tested for binding to BG505, CZA97, and H1 A/New Caledonia/20/99. Data are represented as mean \pm SEM for one ELISA experiment. ELISAs were repeated 2 or more times.

B. Binding of BG505 DS-SOSIP/GnT1- (resulting in Man5-enriched glycans) or BG505 DS-SOSIP/293F cells (complex glycans) to 2723-2121 IgG.

C. Binding of BG505 DS-SOSIP/GnT1- trimer to PGT145 IgG, VRC01 IgG, 17b IgG, and 2723-2121 IgG.

D. Inhibition of BG505 DS-SOSIP/GnT1- binding to 2723-2121 IgG in presence of VRC34 Fab (diamond), PGT145 Fab (square) and VRC01 Fab (triangle).

E. Neutralization of Tier 1, Tier 2, and control viruses by antibody 2723-2121 and VRC01. Results are shown as the concentration of antibody (in μ g/ml) needed for 50% inhibition (IC₅₀).

F. Levels of ADCP, ADCC, ADCT-PKH26, and ADCC displayed by antibody 2723-2121 compared to VRC01. HIVIG was used as a positive control and the anti-RSV mAb Palivizumab as a negative control.

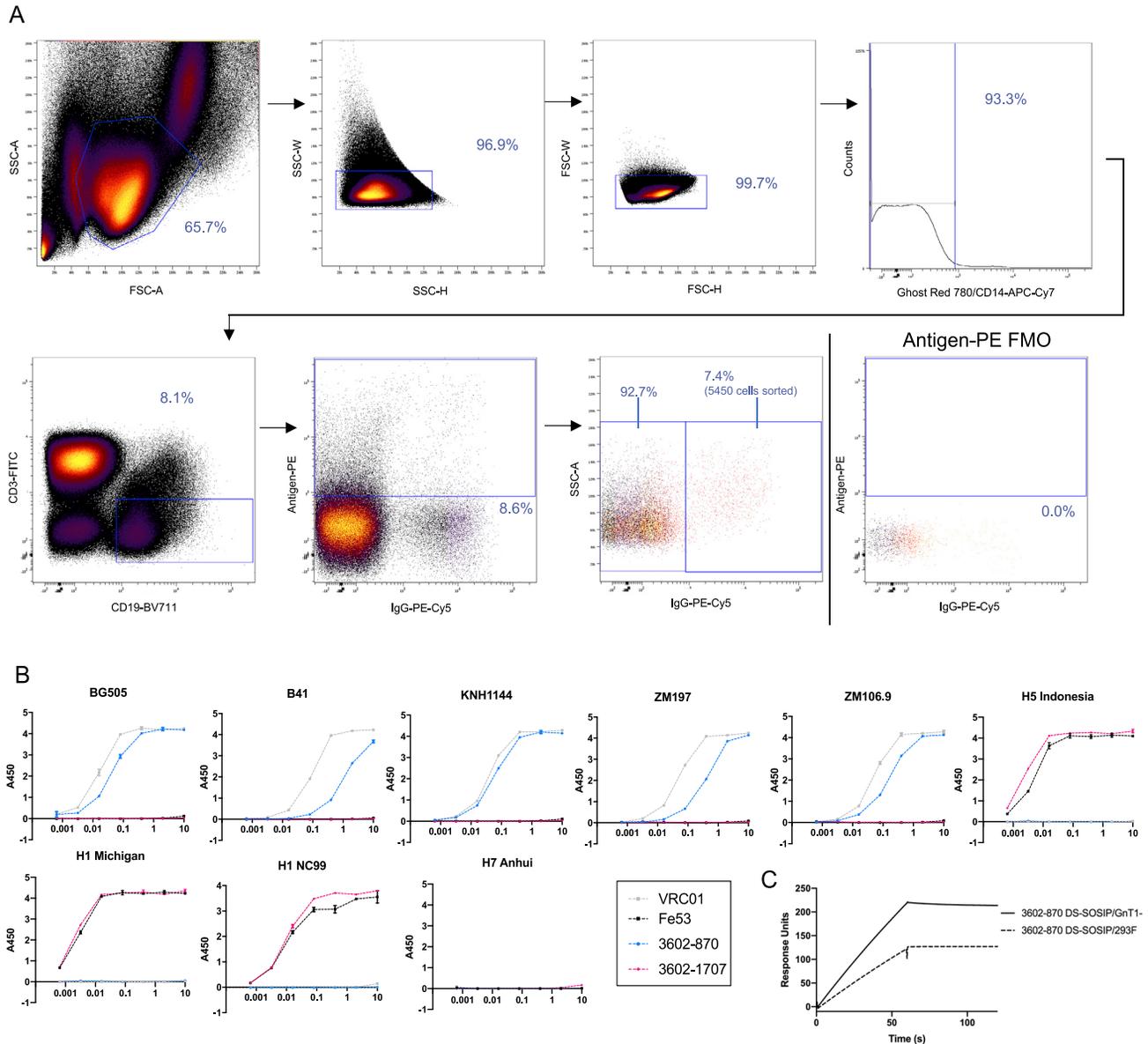
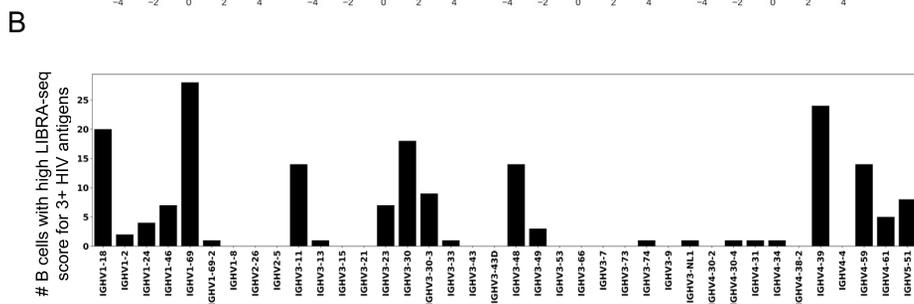
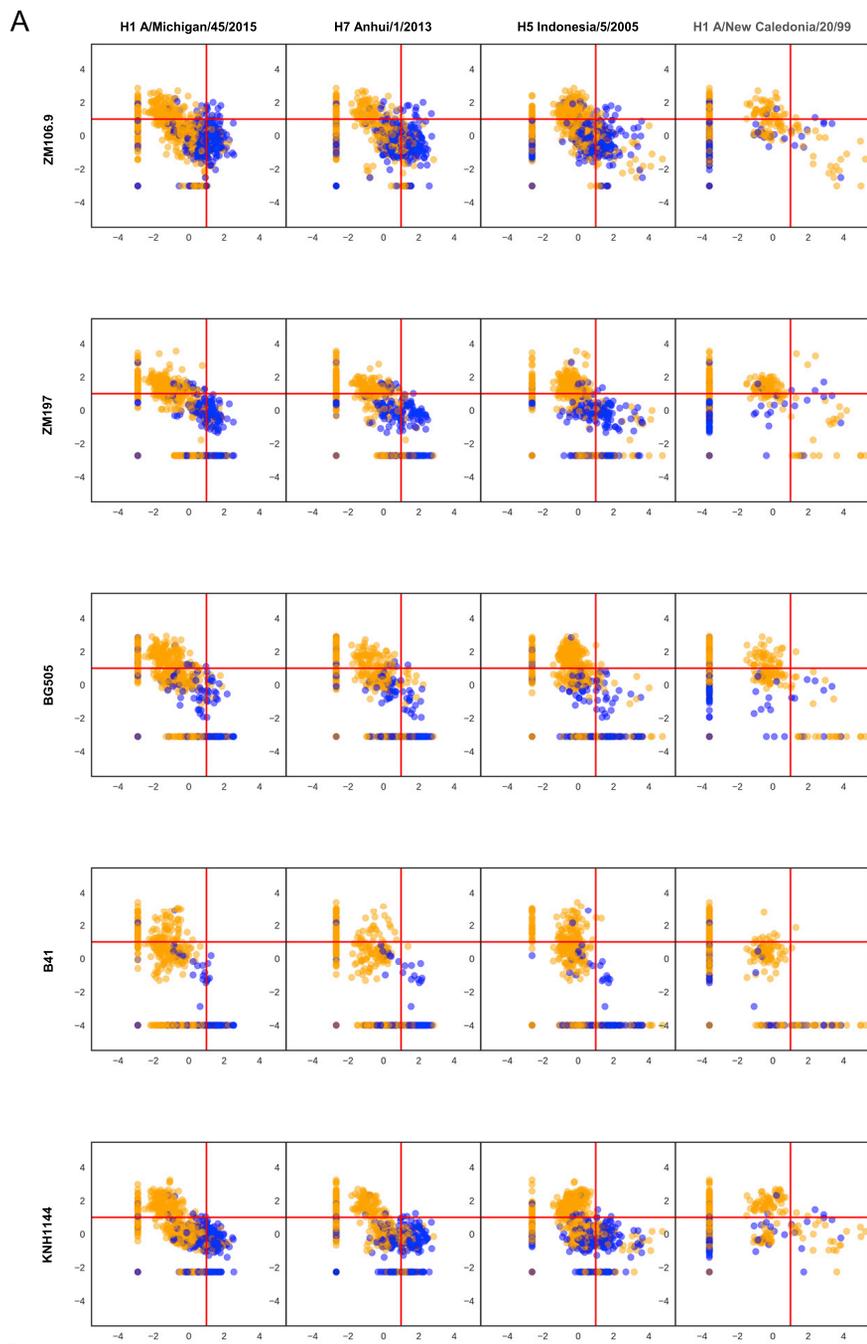


Figure S4. Identification of Antigen-Specific B Cells from Donor N90 PBMCs, Related to Figures 4 and 5

A. Gating scheme for fluorescence activated cell sorting of donor N90 PBMCs. Cells were stained Ghost Red 780, CD14-APCCy7, CD3-FITC, CD19-BV711, and IgG-PECy5 along with a DNA-barcoded antigen screening library consisting of BG505-PE, KNH1144-PE, ZM197-PE, ZM106.9-PE, B41-PE, H1 A/New Caledonia/20/99-PE, H1 A/Michigan/45/2015-PE, H5 Indonesia/5/2005-PE, H7 Anhui/1/2013-PE. Gates as drawn are based on gates used during the sort, and percentages from the sort are listed. 5450 IgG-positive, antigen-positive cells were sorted and supplemented with 1480 IgG-negative, antigen-positive B cells for single-cell sequencing. A small aliquot of donor N90 PBMCs were used for fluorescence minus one (FMO) staining, and were stained with the same antibody panel as listed above without the antigen screening library.

B. Antigen specificity as predicted by LIBRA-seq was validated by ELISA for two antibodies isolated from donor N90. Antibodies were tested for binding to all antigens from the screening library: 5 HIV-1 SOSIP (BG505, KNH1144, ZM197, ZM106.9, B41), and 4 influenza HA (H1 A/New Caledonia/20/99, H1 A/Michigan/45/2015, H5 Indonesia/5/2005, H7 Anhui/1/2013). Data are represented as mean \pm SEM for one ELISA experiment. ELISAs were repeated 2 or more times.

C. Binding of BG505 DS-SOSIP grown in GnT1- (resulting in Man5-enriched glycans) or 293F cells (complex glycans) to 3602-870 IgG.



(legend on next page)

Figure S5. Analysis of Antigen Reactivity for B Cells from Donor N90, Related to Figures 4, 5, and 6

A. Each graph shows the LIBRA-seq score for an HIV antigen (y-axis) versus an influenza antigen (x-axis) in the screening library. The 901 cells that had a LIBRA-seq score above one for at least one antigen are displayed as individual dots. IgG cells (591 of 901) are colored orange and cells of all other isotypes are colored blue. Red lines on each axis indicate a LIBRA-seq score of one. Only 9 of the 591 IgG cells displayed high LIBRA-seq scores for at least one HIV-1 antigen and one influenza antigen, confirming the ability of the technology to successfully discriminate between diverse antigen specificities.

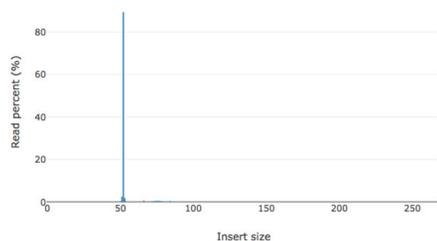
B. V gene usage of broadly HIV-reactive B cells. For each *IGHV* gene, the number of B cells with IgG or IgA constant heavy gene and high (> 1) LIBRA-seq scores for 3 or more HIV-1 SOSIP variants is displayed as a bar. The x axis shows only *IGHV* genes used by at least 1 B cell with a high LIBRA-seq score for at least 1 HIV-1 antigen and an IgG or IgA CH gene.

A

Ramos B Cell Lines		Donor NIH45		Donor N90	
Summary		Summary		Summary	
Sequencing	paired end (150 cycles + 150 cycles)	Sequencing	paired end (150 cycles + 150 cycles)	Sequencing	paired end (150 cycles + 150 cycles)
Insert size peak	52 bp	Insert size peak	52 bp	Insert size peak	54 bp
Before filtering		Before filtering		Before filtering	
Total reads	46,683,278 M	Total reads	39,610,554 M	Total reads	82,957,516 M
Total bases	7,002,492 G	Total bases	5,940,879 G	Total bases	8,653,64 G
Q20 bases	5,890,977 G (84.126864%)	Q20 bases	4,959,718 G (83.484582%)	Q20 bases	7,647,815 G (88.377127%)
Q30 bases	5,343,662 G (76.310871%)	Q30 bases	4,463,982 G (75.140098%)	Q30 bases	6,985,525 G (80.723789%)
After filtering		After filtering		After filtering	
Total reads	46,317,712 M	Total reads	36,922,686 M	Total reads	80,923,090 M
Total bases	2,473,195 G	Total bases	2,061,362 G	Total bases	6,994,344 G
Q20 bases	2,438,363 G (98.591592%)	Q20 bases	1,992,712 G (96.669638%)	Q20 bases	6,314,613 G (90.281709%)
Q30 bases	2,358,131 G (95.347568%)	Q30 bases	1,910,089 G (92.661463%)	Q30 bases	5,822,037 G (83.239223%)
Filtering results		Filtering results		Filtering results	
Reads passed filters	46,317,712 M	Reads passed filters	36,922,686 M	Reads passed filters	80,923,090 M
Reads with low quality	253,286,000 K (0.542563%)	Reads with low quality	1,506,144 K (3.802381%)	Reads with low quality	1,494,196 K (1.801158%)
Reads with too many N	94 (0.000201%)	Reads with too many N	158 (0.000399%)	Reads with too many N	4,434,000 K (0.005333%)
Reads too short	112,186,000 K (0.240313%)	Reads too short	1,181,566 M (2.982958%)	Reads too short	535,806,000 K (0.645880%)

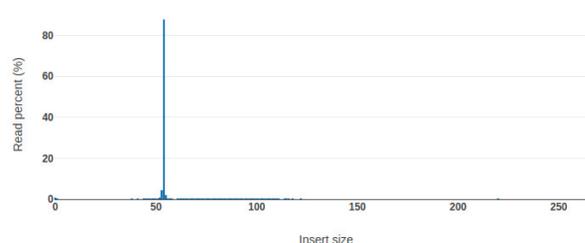
B

Insert size distribution (0.678262% reads are with unknown length)



D

Insert size distribution (1.572161% reads are with unknown length)



C

Insert size distribution (6.482794% reads are with unknown length)

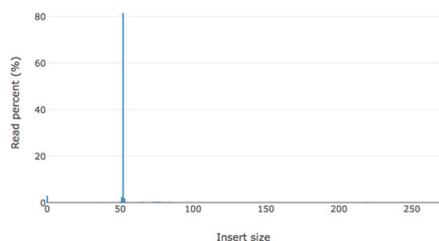


Figure S6. Sequencing Preprocessing and Quality Statistics, Related to Figures 1, 2, and 4 and STAR Methods

A. Quality filtering of the antigen barcode FASTQ files. Fastp (Chen et al., 2018) was used to trim adapters and remove low-quality reads using default parameters. Shown are read and base statistics generated from the output html report from each of the Ramos B cell experiment (left), primary B cell experiment from donor NIAID45 (middle), and primary B cell experiment from donor N90 (right).

B. Shown is a distribution of insert sizes of the antigen barcode reads from the Ramos B cell line experiment, as output from the fastp html report.

C. Shown is a distribution of insert sizes of the antigen barcode reads from the donor NIAID45 experiment, as output from the fastp html report.

D. Shown is a distribution of insert sizes of the antigen barcode reads from the donor NIH90 experiment, as output from the fastp html report.