HIGH-RESOLUTION INTERROGATION OF ANTIBODY REPERTOIRES USING NEXT-GENERATION SEQUENCING

By

Ian Setliff

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Chemical & Physical Biology

January 31, 2020

Nashville, Tennessee

Approved:

James E. Crowe, Jr., M.D.
Kristen Ogden, Ph.D.
Jens Meiler, Ph.D.
Spyros A. Kalams, M.D.

ACKNOWLEDGEMENTS

The work presented in this thesis would not have been possible without the incredible leadership and guidance of my adviser, Dr. Ivelin Georgiev. I will forever remain grateful for his support, trust, advice, patience, wisdom, cool-headedness, and encouragement in my scientific endeavors. He always placed more trust in me than I deserved, which motivated me to continue pushing forward with my science.

I would additionally like to thank the past and present members of the Georgiev laboratory: Amyn Murji, Nagarajan Raju, Kevin Kramer, Kelsey Pilewski, Andrea Shiakolas, Emilee Fechter, Sanjay Mishra, Lauren Walker, Rohit Venkat, Allie Greenplate, Wyatt McDonnell, Surya SarillaBennett, Steven Wall, and Juliana Qin. Beyond their friendship, I also have greatly benefited from their willingness to collaborate and provide scientific input to my work. In particular, I owe large debts of gratitude to Andrea Shiakolas, with whom I collaborated especially closely on the second half of this thesis, and to Amyn Murji and Wyatt McDonnell, who contributed large amounts of time and effort to my studies.

I am also eternally grateful to my thesis committee: James Crowe, Jr, Jens Meiler, Kristen Ogden, and Spyros Kalams. My committee was always incredibly encouraging, pushed me to ‘think big’, provided invaluable scientific input, and dedicated significant amounts of their valuable time to my research. I am also grateful for my committee’s support for working on science in a team setting, in close collaboration with experimental scientists.

Thank you to those who provided valuable input and collaborated with me from afar: Lynn Morris, Rutendo Mapengo, Simone Richardson, Charissa Oosthuysen, Cathrine Scheepers, Priyamvada Acharya, Katarzyna Janowska, Masaru Kanekiyo, Barney Graham, Mark Connors, Daniel Lingwood, Larance Ronsard, and Salim Abdool Karim.

I am additionally grateful for the institutional support and infrastructure provided by Vanderbilt. Without the VANTAGE sequencing core and the Vanderbilt Flow Cytometry Core, I would not have been able to acquire the data used in this thesis. In particular, I would like to thank Latha Raju, Jamie Roberson, Angela Jones, and David Flaherty.

I also thank the NIH and the Vanderbilt Institute for Clinical and Translational Research for funding. In particular, I thank Dr. Walter Chazin for supporting me via the Molecular Biophysics Training Program T32.

Finally, I would like to thank my family. My parents have provided unending support and encouragement in my educational career since before primary school. Their ability to encourage me to persevere through difficult points of graduate school without themselves being scientists is a primary reason I was able to make it to the point where I could write this thesis.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Antibodies and Antibody Repertoire Development

Antibodies are Y-shaped proteins produced by the immune systems of many species in response to foreign pathogens. Human antibodies are composed of four polypeptides: two heavy chains and two light chains. The ends of these proteins are referred to as the variable regions due to the high sequence variability in these regions among different antibody proteins. These variable regions, typically encoded by 110-140 amino acids of each of the heavy and light chains in humans, typically give each antibody a high level of specificity for its cognate antigen (Figure 1).

Each of the variable heavy region and the variable light region can be further divided into 4 framework (FR) regions and 3 complementarity-determining regions (CDRs). The particular amino acid sequences encoding the 4 FRs and 3 CDRs are determined from a number of processes that contribute to antibody sequence diversity, beginning with a gene recombination event. That is, for the heavy chain, a V gene segment is selected from a library of variable (V*)* gene segments, a D gene segment is selected from a library of diversity (D) gene segments, and a J gene segment is selected from a library of joining (J) segments (Tonegawa, 1983), which are recombined via recombination activating genes (RAG-1 and RAG-2) (Janeway et al., 2001). A

similar process occurs in the variable light region; however, no D segment is involved. The many combinations of V, D, and J gene segments in the heavy chain and V and J segments in the light chain contribute to a large combinatorial diversity; the approximately 65 functional *IGHV* gene segments, 27 *IGHD* gene segments, and 6 *IGHJ* gene segments form approximately 11,000 different possible V_H regions (Janeway et al., 2001). This number is further amplified by the many possible pairings of heavy and light chains; assuming equal gene usage frequencies and no pairing biases, each of the ~11,000 possible heavy chains may be paired with the ~320 different light chain recombinations, for a total of ~3.5 x 10^6 different antibody specificities (Janeway et al., 2001).

In addition to combinatorial diversity, antibody sequence diversity is increased by the junctional diversity occurring at the joints between different gene segments. During RAG-mediated recombination, p-nucleotides may be added at single-stranded DNA breakage sites, which may result in additional codons in the junction region (Janeway et al., 2001). Exonucleases may also trim nucleotides encoded by germline gene segments during recombination. Additionally, terminal deoxynucleotidyl transferase (TdT) may add non-templated nucleotides at the V-D and D-J junctions, further increasing junctional diversity. The high expression of TdT during heavy chain formation results in a large variation in CDRH3 length (which include V-D and D-J junctions) (DeKosky, 2015; Janeway et al., 2001). This RAG- and TdT-mediated addition and subtraction of nucleotides during gene recombination further increases the theoretical antibody sequence diversity to approximately 10^{11} different antigen receptors (Janeway

et al., 2001; Trepel, 1974), though recent studies suggest biases in B cell development reduce this number (Soto et al., 2019).

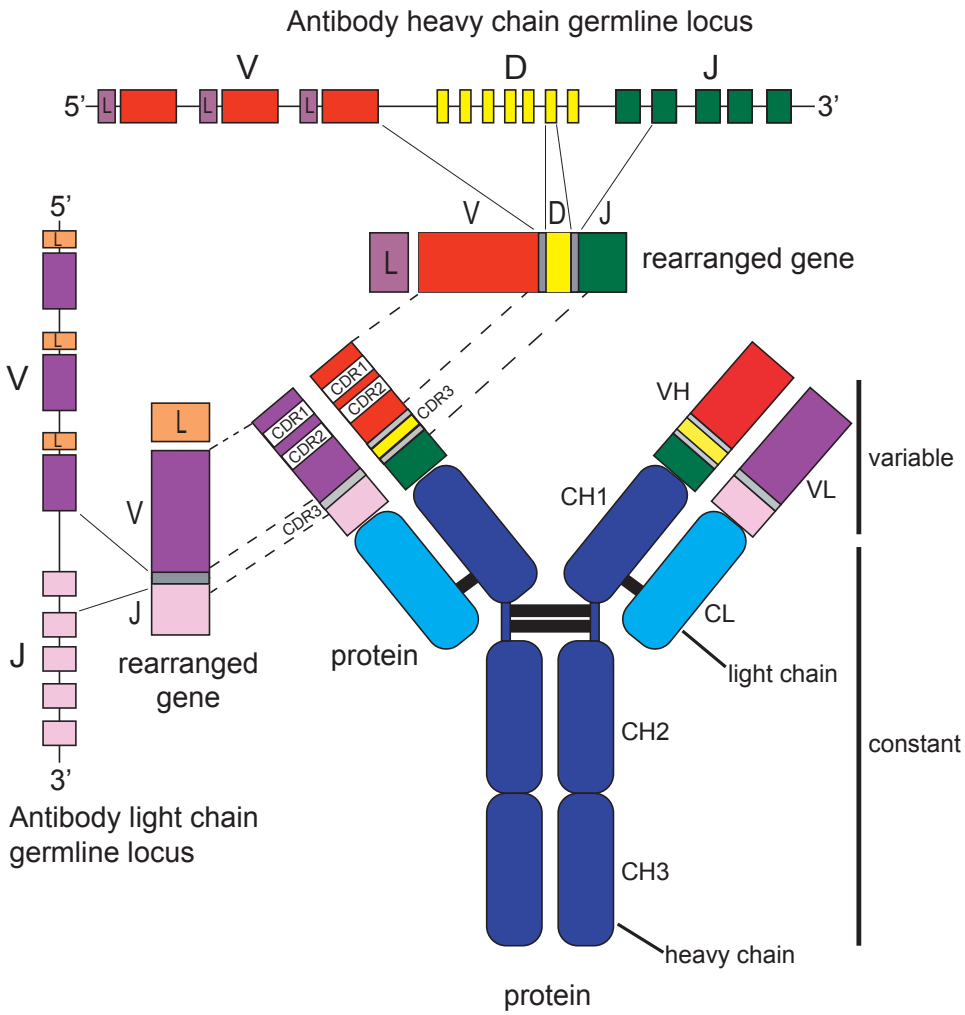


Figure 1. Antibody gene rearrangement, reproduced with permission from (Boyd and Joshi, 2014). The unrearranged genomic DNA configuration of the heavy chain locus is shown at the top of the figure, with the rearranged heavy gene depicted below it; a similar schematic of the unrearranged and rearranged light chain gene is shown to the left of the figure. CDRs 1 and 2 are encoded by the heavy/light V gene, while CDRH3 and CDRL3 are the located at the junction of the V(D)J gene segments. The constant regions of the heavy and light chains (CH1, CH2, CH3, CL) are encoded by downstream exons that are joined to the rearranged V(D)J gene during mRNA splicing.

Following gene segment rearrangement, antigen receptors are expressed on the surface of naïve B-lymphocytes during initial development in the central lymphoid organs. B-lymphocytes may then migrate to the peripheral lymphoid organs, where the sequences of the functional immunoglobulin genes can be fine-tuned for antigen recognition and further diversified. During this process, known as somatic hypermutation, point mutations are introduced into the variable regions of the rearranged heavy- and light-chain genes (Figure 2) (Janeway et al., 2001). Some of the mutant immunoglobulins bind cognate antigen with higher affinity than the original rearranged sequence; these affinity-matured B-lymphocytes are preferentially selected to mature into antibody-secreting cells. While mutations can and do occur in the FR regions, the CDRs – often referred to as hypervariable loops – have much higher levels of variation across different antibodies using the same germline gene segments.

This process of gene recombination, combinatorial chain pairing, junctional editing, and somatic hypermutation occurs across billions of B lymphocytes in humans, resulting in a vast and highly diverse antibody repertoire. This antibody repertoire is influenced by exposure, and, sometimes, re-exposure to diverse pathogens and vaccinations, and thus is incredibly dynamic. While also responding to threats, the antibody-mediated immune system must also be able to distinguish “self” antigens so as not to mount an auto-immune response. The antibody repertoire system thus collectively functions as a molecular classifier (Lasserson, 2012), leveraging its vast sequence diversity to classify and make life-or-death decisions to an almost infinite number of molecular shapes; if stored as digital data, the amount of information needed to encode the number of molecular shapes to which antibody repertoires must respond

would certainly surpass all information stored in all genomes of living individuals (Lasserson, 2012). Further, responses by the antibody repertoire must be made in a timely manner, and against targets that rapidly alter their sequence, and, therefore, their molecular shapes. And, beyond contemporaneous responses to these molecular shapes, the antibody repertoire maintains memory of previous responses. As such, efforts to decode the information stored in the antibody repertoire in order to identify new therapeutic agents or diagnostics are well-motivated.

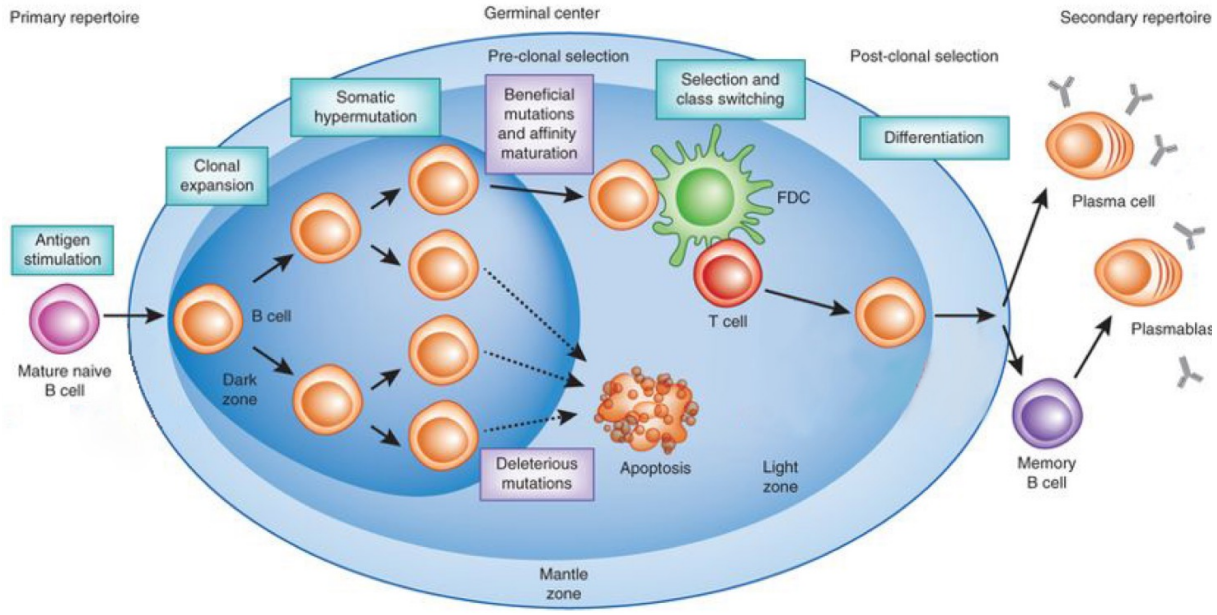


Figure 2. Affinity maturation in the germinal center, adapted with permission from (Georgiou et al., 2014). After being generated in the bone marrow, B cells migrate to the periphery as mature naïve B cells, where they may come into contact with cognate antigen. Upon activation by antigen, they enter a germinal center where they rapidly proliferate. During this proliferation, mutations are introduced into the variable region of the B cell receptor (BCR). Mutations that result in decreased affinity for cognate antigen ultimately apoptose, while B cells bearing BCRs with increased affinity for antigen survive the germinal center reaction. These cells can undergo class switching to IgG, IgA, or IgE isotypes, and ultimately differentiate into memory B cells, antibody-secreting plasmablasts or plasma cells.

1.2 Understanding Antibody Responses to Infection and Vaccination is Important for Vaccine and Therapeutics Development

1.2.1 Human Monoclonal Antibodies Are Powerful Therapeutics

Since Köhler and Milstein succeeded in producing the first hybridoma in 1975 (Köhler and Milstein, 1975), monoclonal antibodies have been pursued as human therapeutics. Köhler and Milstein's technology enabled the discovery of the first FDA-approved therapeutic monoclonal antibody, Muromonab, a mouse anti-CD3 monoclonal antibody approved in 1986 for use in preventing kidney transplant rejection (Singh et al., 2017). Following a lull in approval of monoclonal antibodies, in part due to concerns of immunogenicity risks to patients of murine monoclonal antibodies, Rituximab, a chimeric antibody containing human and murine components, was approved for the treatment of low-grade B cell lymphoma in 1997 (Singh et al., 2017). To overcome risks of immunogenicity, new technologies were developed for the generation of predominately or entirely human monoclonal antibodies, such as human hybridoma technology (Olsson and Kaplan, 1980) and phage display (McCafferty et al., 1990). Such technologies led to the development of the first fully human therapeutic monoclonal antibody, Humira (Adalimumab), approved in 2004 for the treatment of rheumatoid arthritis (Singh et al., 2017); in a large phase 3 continuation study, over a third of patients who completed 10 years of Adalimumab treatment had sustained clinical and functional responses (Furst et al., 2015).

In-depth studies of human immunogenetics have resulted in antibody display libraries composed of primarily human antibody sequences (Chen et al., 2010; Glanville et al., 2009) and in mice transgenic for the human Ig locus (Green, 1999). And while immune libraries and immunized transgenic mice can be rich sources of human monoclonal antibodies, another way to obtain fully human monoclonal antibodies is to directly sample the B cell repertoire of convalescent humans. Indeed, potent broadly neutralizing antibodies (bNAbs) have been isolated directly from humans for many global public health threats, including Ebola virus (Corti et al., 2016; Flyak et al., 2016; Gilchuk et al., 2018), human immunodeficiency virus (HIV) (Bonsignori et al., 2011, 2012; Corti et al., 2010; Doria-Rose et al., 2014; Falkowska et al., 2014; Huang et al., 2012, 2014; Mouquet et al., 2012; Scheid et al., 2011; Sok et al., 2014; Walker et al., 2009, 2011; Wu et al., 2010, 2011), malaria (Kisalu et al., 2018; Tan et al., 2018), influenza virus (Andrews et al., 2017; Bangaru et al., 2019; Joyce et al., 2016; Stevens et al., 2008), Zika virus (Robbiani et al., 2017; Sapparapu et al., 2016), and many others. The utility of some of these candidate bNAbs in treatment as an immunotherapy has been evaluated in human clinical trials. For example, a combination of bNAbs 10-1074 and 3BNC117 proved effective in maintaining low viral loads in non-viremic HIV-infected subjects (Mendoza et al., 2018) and in reducing viral load in viremic patients (Bar-On et al., 2018). Similarly, phase 2/3 human clinical trials for therapeutic antibodies against Ebola virus disease are ongoing, but mAb114 protects macaques from lethal Ebola virus infection when administered as late as 5 days after challenge (Corti et al., 2016) and is well-tolerated in humans (Gaudinski et al., 2019).

Beyond infectious disease settings, transplant rejection, lymphoma, and rheumatoid arthritis, monoclonal antibodies are used therapeutically in almost every other indication; to date, approximately 100 monoclonal antibodies have been approved in the US and/or Europe or are currently in review, and the rate of approvals is rapidly increasing (Ecker et al., 2015; Kaplon and Reichert, 2018, 2019; Reichert, 2012, 2016, 2017). This number is sure to rise in coming years, as the ~225 antibody therapeutics in Phase 2 studies transition to later-stage studies (Kaplon and Reichert, 2019).

1.2.2 Human Monoclonal Antibodies Can Prevent Disease

Beyond their utility in treating active disease, antibodies are also effective in preventing disease. Indeed, polyclonal antibodies have been used for decades in prevention of viral diseases. For example, immunoglobulin has been used extensively for prevention of hepatitis A infection (Stokes et al., 1951). Similarly, hepatitis B immune globulin, prepared from donors with high anti-HBSAg antibody titers, is effective in preventing hepatitis B infection (Keller and Stiehm, 2000). In the case of respiratory syncytial virus (RSV), a blinded, randomized study showed significantly decreased intensive care unit stays and severe RSV infection in infants with intravenous immunoglobulin (RSV-IGIV) (Groothuis et al., 1995).

Improvements in recombinant antibody technology and monoclonal antibody manufacturing have led to increased production capacities, bringing the cost of monoclonal antibodies down to around that of blood-derived polyclonal antibodies, while also increasing supply globally (Kelley, 2009; Sparrow et al., 2017). This has led to the use of monoclonal antibodies displacing intravenous immunoglobulin in some

indications. For example, palivizumab, a monoclonal antibody targeting the fusion protein of RSV, was approved in 1998 for prevention of RSV disease in high-risk infants as the first monoclonal antibody commercially available for the prevention of an infectious disease (Keller and Stiehm, 2000). Palivizumab was successful in reduction in hospitalization due to RSV infection in two multicenter randomized controlled clinical trials (Connor, 1998; Feltes et al., 2003).

Since palivizumab, there have been a number of passive immunization studies for other infectious diseases, including HIV. This began with studies of first-generation bNAbs 2F5 (gp41 MPER-directed) and 2G1G (gp120 glycan-directed), that were used together with HIVIG in macaques (Pegu et al., 2017). Mascola and colleagues observed that while individual bNAbs resulted in partial protection, the combination of 2G12, 2F5, and HIVIG was most protective (Mascola et al., 1999, 2000; Pegu et al., 2017). Later, a single bNAb, b12 (CD4-bs-directed), was shown to protect against mucosal simian/human immunodeficiency virus (SHIV) (Parren et al., 2001), with similar results observed for other single bNAbs (Hessell et al., 2009, 2010).

Following the establishment of large HIV-infection cohorts and the standardization of virus neutralization assays, a set of next-generation bNAbs with increased potencies was identified (Doria-Rose et al., 2016; Falkowska et al., 2014; Huang et al., 2012, 2014; Kong et al., 2016; Moore et al., 2011; Mouquet et al., 2012; Pegu et al., 2017; Scheid et al., 2009; Sok et al., 2014; Walker et al., 2009, 2011; Wu et al., 2010; Zhou et al., 2013, 2015). In a study testing the potential clinical use of one of these next generation bNAbs, PGT121 (V3-glycan-directed), Burton and colleagues achieved sterilizing immunity in all $SHIV_{SF162P3}$-challenged macaques administered at

least 1 mg/kg, and three of five macaques given 0.2 mg/kg (Moldt et al., 2012; Pegu et al., 2017). This study demonstrated that bNAbs could provide *in vivo* protection at lower serum concentrations than shown in previous studies (Moldt et al., 2012; Pegu et al., 2017). Many similar studies with other next generation bNAbs have yielded similar results (Gautam et al., 2016; Pegu et al., 2017; Rudicell et al., 2014; Saunders et al., 2015).

1.2.3 Human Monoclonal Antibodies Guide Vaccine Design

Most viral vaccines provide protection from future infection through the generation of neutralizing antibodies (nAbs) (Havenar-Daughton et al., 2018; Plotkin, 2010). Highly mutable pathogens, such as HIV and influenza virus, have proven especially difficult to neutralize. To tackle these difficult targets, efforts collectively referred to as 'reverse vaccinology' have been employed (Burton, 2017; Havenar-Daughton et al., 2018; Rappuoli et al., 2016). Generally, such efforts seek to identify protective nAbs isolated from the antibody repertoires of infected subjects and study: (1) the molecular interactions between nAb and antigen, and (2) the molecular evolution of nAb in response to antigen. With this information, reverse vaccinology seeks to use the nAb as a template, working backwards to coax the BCR expressed on a naïve B lymphocyte down a pathway of affinity maturation to ultimately produce a nAb, or, alternatively, to directly elicit epitope-specific nAbs.

A key first challenge to reverse vaccinology-based vaccine design is that naïve B lymphocytes expressing BCRs that recognize the correct epitope may be rare, effectively buried deep in the large antibody repertoire (Havenar-Daughton et al., 2018).

This can be further complicated by the fact that B cells specific for non-protective but nearby epitopes may outcompete these rare, low-affinity precursor B cells. To overcome this, many groups have focused on rigorously characterizing the genetic and structural characteristics of nAbs and their precursors. For example, the bNAb VRC01, which is specific for the HIV envelope glycoprotein (Env), as well as several other similar antibodies, form a class of V_H-gene-restricted bNAbs (Zhou et al., 2015). These antibodies primarily contact Env through their CDRH2 domains, utilize either IGHV1-2 (VRC01 class) or IGHV1-46 (8ANC131-class), have a particular angle of approach to Env (Zhou et al., 2015), and have extraordinarily high levels of somatic hypermutation (Wu et al., 2010). Additionally, VRC01-class antibodies are derived from one of the possible five human IGHV1-2 alleles, the IGHV1-2*02 allele, use one of very few V_L genes (κ3-20, κ3-15, κ1-33, or λ2-14), and have unusually short, five-amino-acid-long CDRL3s (Stamatatos et al., 2017). This gene restriction is due to three germline-encoded amino acids ($Trp50_{HC}$, $Asn58_{HC}$, and $Arg71_{HC}$) in their CDRH2 regions that make conserved contacts with Env (Scharf et al., 2016; Stamatatos et al., 2017; West et al., 2012; Zhou et al., 2010, 2013). It has also been observed that the CDRH3 domains of VRC01-class antibodies participate in binding the CD4-binding site of Env through a $Trp100B_{HC}$, located exactly five amino acids before the framework 4 region, which expands the overall surface area between Env and VRC01-class antibodies through interaction with Asn279 in Loop D of gp120 (Stamatatos et al., 2017; West et al., 2012; Yacoob et al., 2016).

Such in-depth structural, sequence, and functional studies of antibodies like VRC01 have been fed forward into vaccine design strategies. For example, germline-

targeting immunogen efforts by the Stamatatos and Schief groups, among others, have designed candidate immunogens specifically engineered to engage the inferred germline form of VRC01. These strategies generally consist of a priming phase, designed to select and expand clones possessing the desired genetic characteristics (i.e., *IGHV1-2*02* allele usage, short CDRL3, etc.), and a boosting phase, designed to select for expanded B cells from the priming phase that have acquired mutations that mediate recognition of native-like Envs. Such immunization schemes have shown promise; immunizations of inferred germline (iGL) heavy-chain (HC) 3BNC60 (a VRC01-like antibody) mice were conducted with eOD-GT8 (a 60-meric nanoparticle germline-targeting immunogen) and 426c core (a dextramer-based multimeric germline-targeting immunogen) (Dosenovic et al., 2015; Stamatatos et al., 2017). One or two immunizations with these germline-targeting immunogens elicited serum antibodies, some of which were CD4-bs-specific, while multiple immunizations with recombinant native-like Env failed to elicit a significant serum antibody response, presumably due to lack of binding to the native-like Env immunogen (Stamatatos et al., 2017). CD4-bs-specific B cells were isolated from eOD-GT8-immunized animals, with one of the 8 clonal B-cell lineages displaying the classical CDRL3 signature; no isolated antibody displayed neutralizing activity (Stamatatos et al., 2017). While still far from an effective vaccine, these studies are good examples of how isolating human monoclonal antibodies can directly inform vaccine design efforts.

A major milestone in antibody-specific vaccine design came with the development of a fusion peptide-based immunization scheme developed by Kwong and colleagues (Xu et al., 2018). A previous study identified a bNAb, N123-VRC34.01, from

an HIV-infected subject that targets the conserved N-terminal region of the HIV-1 fusion peptide (Kong et al., 2016). Immunization with FP8 peptide (residues 512-519 of the fusion peptide) conjugated to keyhole limpet hemocyanin (KLH) elicited cross-clade neutralizing responses when followed by Env trimer boosts (Combadière et al., 2019; Xu et al., 2018). Initial studies showed an elicited neutralization breadth of 31% of a cross-clade panel of 208 HIV-1 strains in mice, with sporadic and lower neutralization breadth in guinea pigs and rhesus macaques (Combadière et al., 2019; Xu et al., 2018). A follow-up study showed that consistent elicitation of cross-clade neutralizing responses could be achieved in guinea pigs through repetitive Env trimer boosting (Cheng et al., 2019; Combadière et al., 2019).

While many more studies are likely needed to find the correct combination of priming immunogen, boosting immunogen(s), and adjuvant, the above studies demonstrate the power that in-depth studies of monoclonal antibodies can have for vaccine design. As eOD-GT8 and a fusion peptide-based candidate vaccine make their ways through clinical trials in the coming years, we will continue to learn about the translatability of these vaccine concepts into humans.

1.3 Methods to Sample the Natural B Cell Repertoire

In recognition of the power of natural immune systems to produce potent antibodies, a number of methods have been developed to sample the natural B cell repertoire (Table 1). Conceptually, these methods seek to ultimately obtain two categories of information for sampled B cells: (1) antibody sequence, and (2) functional antibody information.

The sequence of the antibody provides important information about its development and relatedness to other antibodies. For example, alignment of the antibody's sequence to a database of germline V, D, and J genes allows for identification of gene usage and annotation of junction sequences (heavy and light chain) (Alamyar et al., 2012). This information allows for detection of mutations from germline, thus providing insight to the affinity maturation pathway leading to the antibody's function.

Additionally, with the full sequence of the heavy and light chains, it is possible to produce the antibody as recombinant IgG in order to perform various functional assays. This generally includes first determining the antibody's cognate antigen specificity. After determining specificity, it may be desirable to determine additional functional features, such as residue-level epitope, kinetics, or mechanistic function (e.g., blocking receptor binding).

B cell repertoire sampling methods generally can do either one of these two categories of information (sequence or function) at scale, while obtaining the second category may rely on low-throughput methods. One example of this is antigen-specific B

cell sorting using fluorescence activated cell sorting (FACS) (Figure 3). With antigen-specific B cell sorting, peripheral blood mononuclear cells (PBMCs) are stained with fluorescent antibodies, as well as fluorescent antigen baits, and sorted into individual wells in order to identify antigen-specific B cells (Boonyaratanakornkit and Taylor, 2019; Gaebler et al., 2013; Gilman et al., 2016; Gross et al., 2015; Hamilton et al., 2015; Krishnamurty et al., 2016; Moody and Haynes, 2008; Moody et al., 2012; Taylor et al., 2012). This method enables screening against millions of cells rather quickly but has several limitations. Generally, a maximum of 2-3 antigens can fit on a flow panel before encountering spectral overlap. Moreover, uniting the fluorescence-based specificity readout with sequence is low in throughput; multiple rounds of PCR are required to amplify the heavy- and light-chain transcripts from each individual well, and sequencing is generally performed using Sanger sequencing. The lack of parallelizability in Sanger sequencing results in a linear increase in cost as the number of cells increases. Though each well can be individually barcoded enabling amplicons to be pooled and sequenced on an Illumina platform (DeFalco et al., 2018; Setliff et al., 2018; Tan et al., 2014), the need to prepare amplicons from each well individually remains.

Limiting dilution is another technique often used to isolated antigen-specific B cells. In this approach, primary cells are diluted serially until individual B cells are separated in microwell plates (Blanchard-Rohner et al., 2010; Boonyaratanakornkit and Taylor, 2019). The B cells can then be cultured and expanded *ex vivo* and/or immortalized so that each well contains a monoclonal antibody (Köhler and Milstein, 1975; Traggiai et al., 2004; Yu et al., 2008). A major advantage of this approach is that the culture supernatants containing monoclonal antibodies can be screened in

neutralization or other functional assays, rather than just binding. However, the requirement for *ex vivo* culture prior to screening is time-consuming and laborious, and results in loss of many clones due to imperfect fusion efficiencies or cell death during culture. Similar to single-cell sorting, heavy and light chain antibody sequences have to be acquired one well at a time.

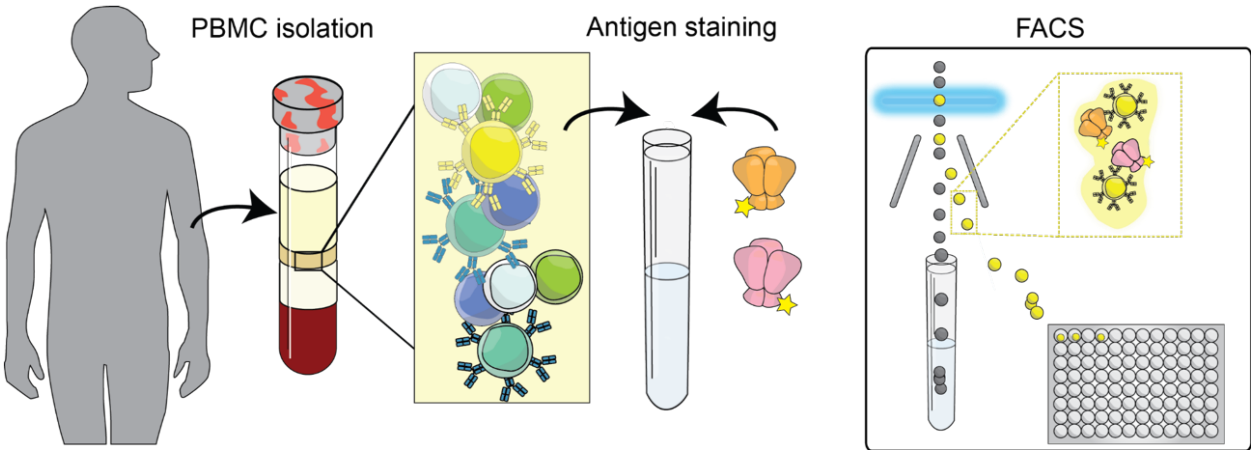


Figure 3. Antigen-specific B cell sorting. Following isolation of PBMCs from an antigen-experienced host, PBMCs are stained with fluorescent antibodies and antigen baits and single-cell sorted via FACS.

In addition to FACS-based and limiting dilution-based methods to isolate antigen-specific B cells, which offer potential throughput advantages for acquiring functional information, next-generation sequencing-based (NGS-based) methods have been recently employed in B cell repertoire sampling studies. Generally, such approaches facilitate a significant increase in antibody sequence acquisition throughput when compared to FACS-based and limiting dilution-based methods. However, generally, identified monoclonal antibody sequences need to be expressed as recombinant IgG and tested individually in order to acquire functional properties (DeFalco et al., 2018;

Goldstein et al., 2019; Setliff et al., 2018). NGS-based methods are reviewed in-depth in section 1.4.

Technique	Number B cells	Paired V_H:V_L	Number antigens	Advantages	Limitations
Limiting dilution	10^1-10^3	Yes	Separate assay for each antigen	Allows for functional screens beyond binding	(1) Labor-intensive; (2) Requires antibody secretion; (3) Sequences recovered one-by-one
FACS sorting	10^2-10^3	Yes	1-3	(1) Retains cells for potential downstream analysis; (2) Detection of low affinity antigen-specific B cells	(1) Potential for fluorochrome-specific, streptavidin-specific, or linker-specific B cells (Boonyaratanakornkit and Taylor, 2019); (2) Spectral overlap limits number of antigens; (3) Requires soluble, stable antigens; (4) Sequences recovered one-by-one
Display technologies	10^2-10^7	Yes	Sequential sorting	(1) Can use affinity gates; (2) High-throughput for both screening and sequence acquisition	(1) Sorts of large display libraries are time intensive; (2) In sequential sorts against many antigens, single cells are not indexed for each antigen; (3) Antibody fragments expressed in yeast are not authentic immunoglobulin
Bulk BCR sequencing	10^8-10^9	No	N/A	(1) Deep sequencing can identify many somatic variants; (2) Enables large-scale characterization	(1) Antigen specificity of observed sequences is unknown; (2) Lack of maintaining heavy and light chain pairing precludes

				of the B cell repertoire	production as monoclonal antibodies
Paired chain BCR sequencing	10^2-10^5	Yes	N/A	(1) Maintains VH:VL pairing, enabling expression as recombinant IgG; (2) Many single-cell cDNA libraries can be prepared in parallel	(1) Antigen specificity of observed sequences is unknown, requiring expression of recombinant IgG; (2) Relatively low number of cells means likely only a few antigen-specific cells

Table 1. Methods to sample the natural B cell repertoire.

1.4 High-Throughput Antibody Sequencing

Major decreases in the cost of next-generation DNA sequencing technologies have greatly accelerated the pace of biological discovery. In addition to many other areas of biological research, these technologies have been applied to study antibody repertoires (Bonsignori et al., 2016; Briney et al., 2019; Greiff et al., 2015; Jackson et al., 2014; Parameswaran et al., 2013; Setliff et al., 2018; Soto et al., 2019; Weinstein et al., 2009). High-throughput antibody repertoire sequencing protocols begin with collection of 10^3 – 10^9 lymphocytes, cell lysis, and cellular mRNA extraction (DeKosky, 2015). mRNA is then reverse transcribed and amplified using primers specific to antibody variable genes, and, possibly, constant genes. Following construction of antibody amplicon cDNA libraries, high-throughput sequencing and bioinformatic analysis are used to characterize each observed sequence. After alignment and annotation of each individual antibody sequence, repertoire-level summary statistics can

be calculated, such as gene usage frequencies, clonality, diversity, CDR3 network architecture, CDR3 sequence motif usage and biophysical properties, and others.

Until recently, it was not possible to perform high-throughput sequencing of paired heavy- and light-chain sequences; rather, native V_H:V_L pairs could be identified only after single-cell cloning by limiting dilution and Sanger sequencing (Georgiou et al., 2014). Instead, the utility of next-generation sequencing for antibody discovery was confined to inferring pairing from oligoclonal populations (Reddy et al., 2010), inference of V_H:V_L pairs from topologies of V_H and V_L phylogenetic trees (Zhu et al., 2013a, 2013b), and use as a reference library for proteomic identification of antibodies (Cheung et al., 2012; Sato et al., 2012; Wine et al., 2013).

New approaches for sequencing endogenous V_H:V_L pairs have increased throughput to beyond that enabled by single-cell cloning by limiting dilution and Sanger sequencing. For example, Wardemann and colleagues (Busse et al., 2014) used a barcoded primer matrix to combine single-cell V_H and V_L gene amplification with next-generation sequencing (Georgiou et al., 2014). A key advantage of this advance is the ability to clone IgH and Igκ/Igλ sequences into expression vectors for further antibody characterization (Georgiou et al., 2014). Georgiou and colleagues (Dekosky et al., 2013) developed a V_H:V_L pairing approach that genetically links V_H and V_L transcripts by emulsion overlap extension PCR after capturing RNA on poly-dT beads. Alternatively, single B cells can be encapsulated in individual emulsions together with uniquely barcoded beads using a microfluidic device, as made commercially available by 10X Genomics. With this method, V_H and V_L transcripts originating from the same cell are indexed by a shared, bead-delivered cell barcode during reverse transcription and PCR.

PCR products are then sequenced on an Illumina platform, with V_H:V_L pairs identified bioinformatically by their shared cell barcodes. This method was recently combined with recombinant antibody expression to identify antigen-reactive clones (Goldstein et al., 2019).

While these advances have facilitated high-throughput identification of V_H:V_L pairs, the process of obtaining antibody sequences is still fundamentally decoupled from functional information. While high-throughput antibody synthesis or expression cloning can be used to unite observed antibody sequence with antigen specificity (Busse et al., 2014; Goldstein et al., 2019), this process is generally time-intensive, laborious, and expensive. Recently, next-generation sequencing of antibody heavy and light chain transcripts has been combined with display technologies in order to unite sequence with antigen specificity. With such methods, genetically-linked V_H and V_L transcripts are expressed as yeast single-chain variable fragment (scFv) libraries (Adler et al., 2017a, 2017b) or Fab libraries (Wang et al., 2018), screened for binders via FACS, and sequenced. While these technologies directly unite sequence and binding information, screening large display libraries can be time-consuming and screening against high numbers of antigens requires sequential sorts.

1.5 Summary of Thesis Work

This dissertation utilizes next-generation sequencing to interrogate antibody repertoires perturbed by infection, with a focus on HIV. In Chapter Two (*Longitudinal High-Throughput Sequencing of the Heavy Chain Repertoire During HIV Infection*), we utilize current bulk BCR sequencing protocols to explore features of antibody repertoires

of HIV-infected subjects that are observed across multiple subjects. In particular, we identify the existence of public antibodies that are shared by multiple donors and HIV-specific. In so doing, we also explore the limitations of bulk sequencing in guiding antibody discovery. This chapter is based on Setliff et al., 2018.

Next, in Chapter Three (*High-Throughput Mapping of B-cell Receptor Sequences to Antigen Specificity for High-Dimensional Screens of Single B Cells*), we present a method, LIBRA-seq, to address important limitations of previous B cell repertoire sampling methods. We describe a technology in detail that interrogates antigen-antibody interactions using a next-generation sequencing-based readout, and present validation on a simple system composed of B cell lines expressing known BCRs. Further, we apply the technology to infection samples of HIV-infected subjects and demonstrate the method's ability to recover previously-identified bNAb lineages as well as identify new bNAbs. This chapter is based on Setliff et al., 2019.

In Chapter Four (*High-Throughput Epitope Binning Using LIBRA-seq)*, we show an extension of LIBRA-seq that facilitates the acquisition of residue-level epitope information for thousands of single B cells in parallel, thus showing that functional information beyond just binding is also accessible using a sequencing-based readout. No data in this chapter have been published previously.

We believe the technological developments presented in this thesis pave the way for a new approach to B cell repertoire analysis and antibody discovery and anticipate major biological discoveries, beyond the ones presented within this thesis, will be made using this thesis as a foundation.

CHAPTER 2

LONGITUDINAL HIGH-THROUGHPUT SEQUENCING OF THE HEAVY CHAIN REPERTOIRE DURING HIV INFECTION

Most of the text in this chapter is reproduced from Setliff et al., 2018. Author contributions are indicated in Appendix 1.

2.1 Rationale

The HIV-1 envelope glycoprotein mediates receptor recognition and viral fusion and serves as the sole target of the neutralizing antibody response (Pancera et al., 2014; Ward and Wilson, 2015). The developmental pathway of Env-specific antibodies has been probed previously using high-throughput sequencing (Bonsignori et al., 2016; Doria-Rose et al., 2014; Huang et al., 2016; Liao et al., 2013; Wu et al., 2011), but such analyses have focused on single broadly neutralizing antibody (bNAb) lineages after infection. However, bNAbs comprise only a fraction of the antibody response within a given individual, which also includes antibodies with limited or no breadth. These diverse antibodies are subject to viral selection pressures and host constraints, target a variety of epitopes on Env, and potentially possess functions other than neutralization (Ackerman et al., 2016; Burton and Mascola, 2015; Corey et al., 2015; Horwitz et al., 2017). More generally, thorough and large-scale profiling of the repertoire-wide antibody response during the course of natural infection remains a predominantly unexplored area of investigation and an unmet need in HIV-1 research. Indeed, the extensive

evidence of the global effects that HIV-1 has on the adaptive immune system, including hypergammaglobulinemia (De Milito et al., 2004), CD4+ T-cell abnormalities (Kaufmann et al., 2007; Palmer et al., 2004; Zhang et al., 2004), and defective CD8+ T-cell function (Harrer et al., 1996; Rinaldo et al., 1995) motivates efforts to understand the dynamics of the antibody repertoires of HIV-infected individuals.

Although putative bNAb precursors have been discovered in HIV-naïve repertoires (Jardine et al., 2016; Yacoob et al., 2016), it is unclear how the antibody repertoires of HIV-infected individuals change from the time before infection through different stages of infection. Furthermore, while ontogeny and structural studies of HIV-reactive antibodies have revealed convergence at the structural level in multiple donors (Scheid et al., 2011; Wu et al., 2011; Zhou et al., 2015), the overall differences and similarities in the antibody repertoires of HIV-infected donors have not been characterized. Due to the diversity of potential target epitopes on Env, as well as the potentially infinite antibody sequence space resulting from gene recombination and affinity maturation, it could be expected that the antibody repertoire of each individual might be unique. Yet, public antibody clonotypes that are shared among multiple individuals have been observed previously for dengue virus infection (Parameswaran et al., 2013), after influenza vaccination (Jackson et al., 2014), and in other immune settings (Arentz et al., 2012; Henry Dunand and Wilson, 2015; Pieper et al., 2017; Trück et al., 2015). However, in the context of HIV-1 infection the potential for public antibodies has not been explored.

To better understand antibody repertoire dynamics throughout HIV-1 infection, we performed antibody repertoire sequence analysis to examine distinctive

characteristics of the pre- and post-infection repertoires of multiple donors. To that end, we longitudinally sequenced the global immunoglobulin heavy chain repertoires of six South African donors from the Centre for the AIDS Programme of Research in South Africa (CAPRISA) from before infection through acute and chronic infection. We also performed paired heavy and light chain sequencing of the Env-specific post-infection repertoires of two additional CAPRISA donors. The resulting analysis provides insights into how antibody repertoires of different individuals are reshaped during the course of HIV-1 infection.

2.2 Results

2.2.1 CAPRISA Donor Samples

Antibody variable genes in peripheral blood cell samples from three timepoints, categorized as pre-infection, 6 months post-infection (6mpi), or 3 years post-infection (3ypi), were sequenced for each of six CAPRISA donors (Table 2). The pre-infection timepoints ranged from 30 to 2 weeks before infection, with the exception of donor CAP322, for whom the earliest available sample was at 2 weeks post-infection. All CAPRISA donors were infected with clade C viruses (Rademeyer et al., 2016), but exhibited diverse neutralization phenotypes, including substantial variation in neutralization breadth between 0 and 61% on a representative panel of diverse HIV-1 strains (Table 2). For the three donors with demonstrable neutralization breadth (CAP287, CAP312, and CAP322), we also performed neutralization fingerprinting analysis (Doria-Rose et al., 2017; Georgiev et al., 2013) to delineate the epitope

specificities of broadly neutralizing antibodies at the 3ypi timepoints. These three donors were predicted to possess different types of antibody specificities (Table 3). Taken together, the observed differences in neutralization phenotypes for the six donors indicated diversity in the types of antibody specificities present in each donor.

Donor ID	Date of Infection	Visit Classification	Visit weeks post infection (wpi)	Visit Date	Medical Events	Viral Load	Neutralization Breadth (%)
CAP301	31-May-08	Pre-infection	-9	31-Mar-08	(1) Hep B core IgG positive at enrolment on 25-Jun-08. (2) 5 episodes of documented STIs. (3) Other minor ailments reported: tonsillitis, axillary abscess, scleral lesions.	—	
		6mpi	28	11-Dec-08		6210	
		3ypi	171	13-Sep-11		< 20	11
CAP351	14-May-09	Pre-infection	-18	8-Jan-09	(1) PTB in 2007-2008. Completed treatment in Aug 2008. (2) TB episode prior to HIV acquisition. (3) Herpes zoster 24-Mar-11. (4) Recurrent upper respiratory tract infections.	—	
		6mpi	22	15-Oct-09		33600	
		3ypi	144	13-Feb-12		38026	6
CAP335	24-Dec-08	Pre-infection	-30	26-May-08	Generally well throughout	—	
		6mpi	23	3-Jun-09		11500	
		3ypi	172	10-Apr-12		16955	0
CAP287	1-Dec-07	Pre-infection	-2	15-Nov-07	(1) Generally well. (2) Trichomas vaginalis July 2008. (3) Chronic hypertension. (4) Mild anaemia with B12 defiiency in pregnancy. (5) Pregnancy Jan to Oct 2009.	—	
		6mpi	28	12-Jun-08		39500	
		3ypi	162	11-Jan-11		4140	61
CAP312	29-Jul-08	Pre-infection	-26	29-Jan-08	(1) PV discharge and Lymphadenopathy noted at numerous Phase 2-4 physical exam CRFs (2008-2011). (2) Bartholin's abcess. Admitted to hospital for marsupialisation of abscess. (3) T. vaginalis 05-Aug-10. (4) Miscarriage 15-Jul-11. History of passing clots and pregnancy test positive, outcome unknown. (5) Recurrent diarrhoea in 2014.	—	
		6mpi	25	20-Jun-09		10400	
		3ypi	157	2-Aug-11		546	61
CAP322	1-Jul-08	(Pre-infection)	2	15-Jul-08	(1) Generally well. (2) HIV related cervical lymphadenopathy Sept 2010 that resolved spontaneously. (3) Hep B cAb positive, Hep B sAg negative May 2010. (4) Hep B cAb negative Apr 2011. (5) Normal liver function tests throughout.	—	
		6mpi	27	7-Jan-09		34300	
		3ypi	161	4-Aug-11		5386	39
CAP248	23-Mar-05	5.9ypi	298	8-Feb-11	Generally well throughout	26509	59 (at 3ypi)
CAP314	9-Jun-08	2ypi	115	24-Aug-10	(1) History of TB. (2) Pneumonia in 2010. (3) Herpes Zoster in 2010. (4) History of Candidiasis.	25805	44

Table 2. Clinical features and timepoints of sample collection for CAPRISA donors. Data generously provided by Nigel Garrett, Fathima Sayed, Duduzile Nkosi, Nonhlanhla Yende, and Precious Radebe.

serum	VRC01-like	b12-like	HJ16-like	8ANC195-like	PG9-like	PGT128-like	2F5-like	10E8-like	35O22-like	PGT151-like	Median of delineation scores
CAP287		0.17	0.18			0.64					0.0000
CAP312	0.03	0.33	0.21		0.28					0.14	0.0175
CAP322	0.24	0.22				0.23		0.09		0.21	0.0485

Table 3. Neutralization fingerprinting for CAP287, CAP312, and CAP322 indicating differing antibody specificities among subjects with broadly neutralizing serum.

2.2.2 High Turnover of Antibody Repertoires During the Course of HIV-1 Infection

After sequencing the antibody heavy chain variable gene regions from all three timepoints for each of the six donors (Materials and Methods), we first investigated how repertoire composition changed over time. We began by determining clonal family membership for each observed V(D)J sequence using V-gene assignment, J-gene assignment, junction length, and junction identity (Materials and Methods). All timepoints from each donor were included during clonal family assignment so that closely related sequences from different longitudinal samples could be assigned to the same clonotype. Clonal family assignment revealed the number of clonotypes (groups of sequences resulting from clonal family assignment) in each donor to range from 9141 to 26777, with a total of 103475 unique clonotypes across all 6 donors (Table 4). Few clonotypes spanned multiple timepoints within a donor, with the majority of clonotypes belonging to a single timepoint (Figure 4A). While some clonotypes were present in all three timepoints within individual donors, this finding was rare, typically representing

only ~0.08% to 1.16% of clonotypes (Figure 4A). Clonotypes spanning any two adjacent timepoints were also rare, although somewhat more frequent than those that spanned all three timepoints, representing ~0.52% to 1.78% of all clonotypes (Figure 4A). Interestingly, although sequence turnover was high, antibody heavy chain variable gene usage distributions remained predominantly constant over the course of infection (Figure 4B, Figure 5). While some genes appeared to be used preferentially within each sample (Figure 5), there was no discernable pattern across all donors of longitudinal enrichment of particular genes. Similarly, the length distributions for the third heavy chain complementarity-determining region (CDRH3) remained relatively unchanged throughout infection for each donor ($P > 0.1$ for all CDRH3 lengths in all samples, *Z*-test with Benjamini-Hochberg correction) (Figure 4C). Taken together, these results indicated that while systems-level repertoire features were conserved over time, antibody sequence retention over the course of HIV-1 infection was low; rather, each donor was associated with virtually non-overlapping repertoires at the three different timepoints.

PID	Timepoint	Sample #	Sum of Filtered Read Count	# Unique VDJ Sequences	# Clonotypes (70%)	# Public Clonotypes	# Private Clonotypes	# Unique VDJ Sequences in Public Clonotypes
CAP301	pre-infection	12	1435042	85957	5017	356	4661	20252
	6mpi	11	1430714	83750	4487	380	4107	2539
	3ypi	7	784491	48700	6420	610	5810	4526
CAP351	pre-infection	16	1180638	65297	4712	253	4459	695
	6mpi	8	1709720	101426	6204	329	5875	2774
	3ypi	6	1085531	70554	13832	916	12916	3277
CAP335	pre-infection	14	525443	28110	3527	252	3275	5538
	6mpi	5	802391	50823	7349	524	6825	4206
	3ypi	1	1095128	68986	6609	470	6139	22722
CAP287	pre-infection	10	1536756	53242	1872	129	1743	12324
	6mpi	15	1105900	57990	5150	518	4632	6181
	3ypi	2	1178014	58283	2214	147	2067	6116
CAP312	pre-infection	9	1265235	79683	3839	297	3542	32392
	6mpi	13	1134878	54980	2369	142	2227	10426
	3ypi	3	1808428	106429	8438	420	8018	3122
CAP322	(pre-infection)	17	1523989	96935	9245	602	8643	6588
	6mpi	18	1155960	60456	3004	202	2802	21487
	3ypi	4	1464792	94790	14920	1005	13915	8501
Total			**22223050**	**1266391**	**103475**	**3515**	**99960**	**172755**

Table 4. Sequencing depth information and counts of public and private clonotypes for each sample, for the 70% junction region identity threshold.

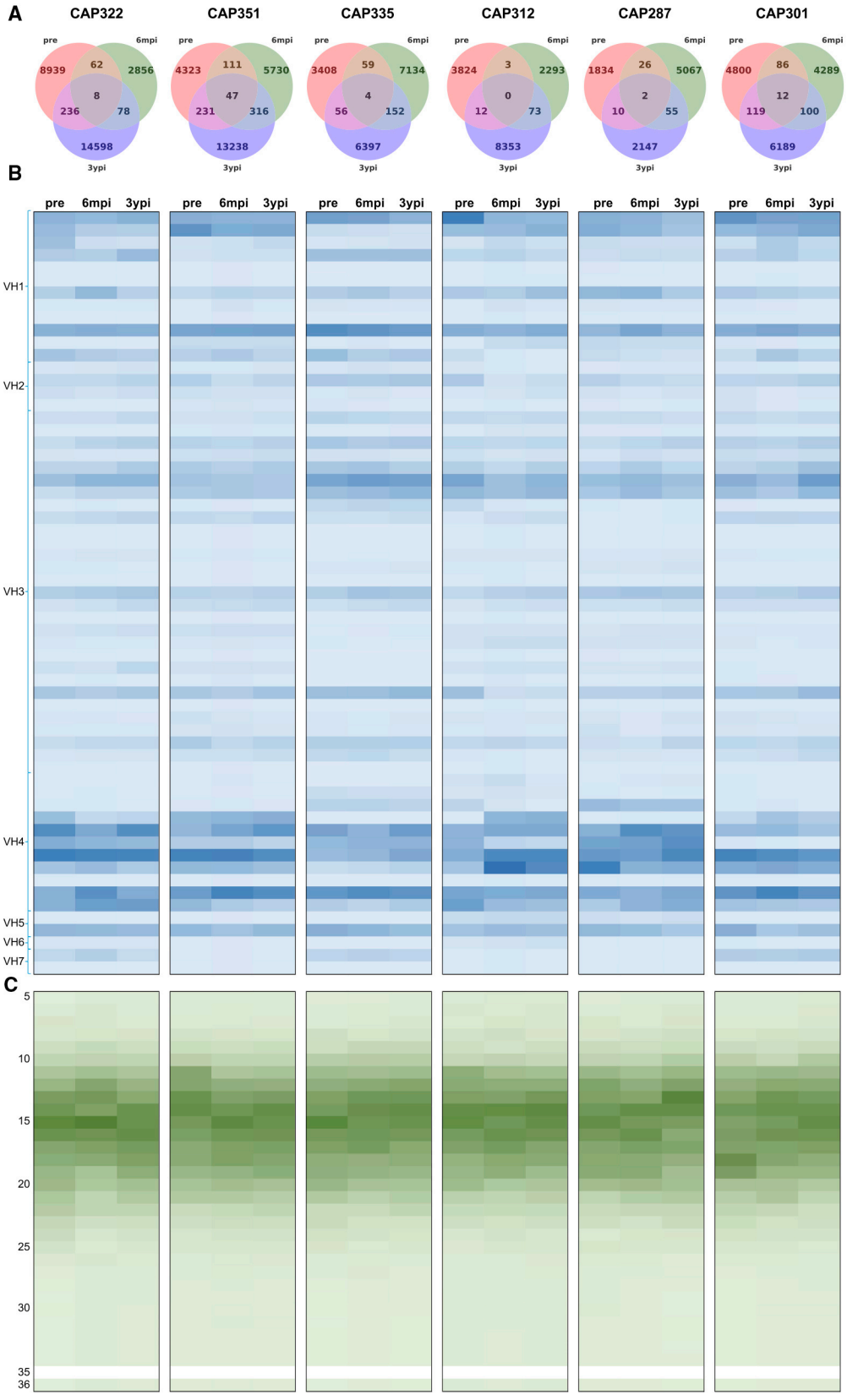
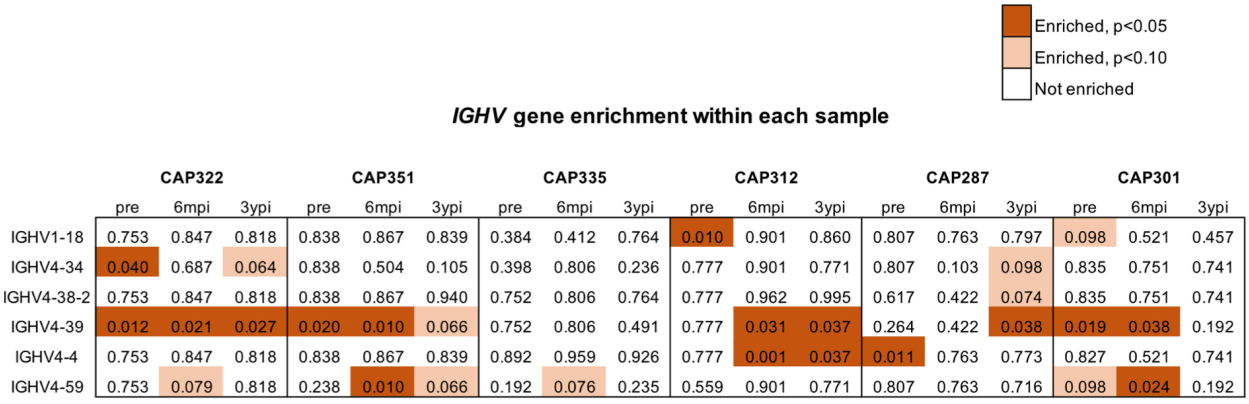

A
CAP322
pre
6mpi
3ypi
8939
62
2856
8
236
78
14598
CAP351
4323
111
5730
47
231
316
13238
CAP335
3408
59
7134
4
56
152
6397
CAP312
3824
3
2293
0
12
73
8353
CAP287
1834
26
5067
2
10
55
2147
CAP301
4800
86
4289
12
119
100
6189
B
VH1
VH2
VH3
VH4
VH5
VH6
VH7
C
5
10
15
20
25
30
35
36

Figure 4. Within-Donor Longitudinal Antibody Repertoire Analysis from Pre-infection through Chronic HIV-1 Infection.
(A) For each donor, the number of clonotypes unique to each time point is shown, as well as the clonotypes shared between two or all three time points.
(B) Heatmap of V-gene usage by donor and time point. For each time point of each donor, the number of clonotypes using each VH gene (excluding orphan genes) was summed and the *Z* score was calculated. *Z* scores range from −0.81 (light blue) to 4.26 (dark blue).
(C) Heatmap of CDRH3 amino acid length by donor and time point. For each time point of each donor, the number of clonotypes of each CDRH3 length was summed and the *Z* score was calculated. No sequences had a CDRH3 length of 35 in any sample in this study. *Z* scores range from −0.90 (light green) to 2.54 (dark green).



	CAP322			CAP351			CAP335			CAP312			CAP287			CAP301		
	pre	6mpi	3ypi	pre	6mpi	3ypi	pre	6mpi	3ypi	pre	6mpi	3ypi	pre	6mpi	3ypi	pre	6mpi	3ypi
IGHV1-18	0.753	0.847	0.818	0.838	0.867	0.839	0.384	0.412	0.764	0.010	0.901	0.860	0.807	0.763	0.797	0.098	0.521	0.457
IGHV4-34	0.040	0.687	0.064	0.838	0.504	0.105	0.398	0.806	0.236	0.777	0.901	0.771	0.807	0.103	0.098	0.835	0.751	0.741
IGHV4-38-2	0.753	0.847	0.818	0.838	0.867	0.940	0.752	0.806	0.764	0.777	0.962	0.995	0.617	0.422	0.074	0.835	0.751	0.741
IGHV4-39	0.012	0.021	0.027	0.020	0.010	0.066	0.752	0.806	0.491	0.777	0.031	0.037	0.264	0.422	0.038	0.019	0.038	0.192
IGHV4-4	0.753	0.847	0.818	0.838	0.867	0.839	0.892	0.959	0.926	0.777	0.001	0.037	0.011	0.763	0.773	0.827	0.521	0.741
IGHV4-59	0.753	0.079	0.818	0.238	0.010	0.066	0.192	0.076	0.235	0.559	0.901	0.771	0.807	0.763	0.716	0.098	0.024	0.192

Figure 5. Longitudinal analysis of gene usage enrichment. Clonotype abundance of each V gene was normalized by timepoint for each donor as described in (Cheadle et al., 2003). These standard scores were then tested with Z-tests, the *P* values of which were then adjusted using adaptive Benjamini-Hochberger correction for an FDR of 0.05. Genes with at least one $P < 0.1$ are shown, and all *P* values lower than, respectively, 0.10 and 0.05 after correction are colored.

2.2.3 Identification of Public Antibody Clonotypes in HIV-1 Infection

Since recent studies have indicated the presence of public antibody clonotypes shared among individuals, including in the antigen-directed response (Jackson et al., 2014; Parameswaran et al., 2013), we set out to explore the presence of public antibody sequences in the case of HIV-1 infection. To identify public antibodies, we performed clonal family assignment on all samples from all donors simultaneously (Materials and

Methods). As expected, this approach resulted in most clonotypes consisting of sequences from single donors. However, we also identified a number of clonotypes with sequences from multiple donors (Figure 6), which we designated public clonotypes. Although these sequence groups are not technically biological clones (as they were derived from multiple individuals and thus are not derived from a single B cell), public clonotypes were defined as groups of sequences with the same V_H gene, the same J_H gene, the same junction length, and CDRH3 amino acid sequences of high identity between donors (Materials and Methods).

We analyzed the antibody sequences using a range of junction region identity thresholds during the clonal assignment procedure (Figure 19). Of note, public clonotypes were identified for all threshold values, including at 100% junction identity. Although identities between members of known bNAb lineages can be as low as ~30% (Table 5), for further analysis we selected a conservative threshold of 70%, in which all members of a clonotype would have at least 70% junction region identity to all other members of the clonotype. This threshold aimed at allowing reasonable inclusion of intra-clonal evolution without allowing highly divergent sequences from different donors to be grouped together.

Comparison of the antibody repertoires between all pairs of donors revealed that public antibody clonotypes existed at each of the pre-infection, 6mpi and 3ypi timepoints (Figure 6A). Intriguingly, at the 70% junction identity threshold used for this analysis, the number of public clonotypes significantly increased at 3ypi (average number of public clonotypes from pairwise comparisons of 69.1) relative to pre-infection (average number of public clonotypes from pairwise comparisons of 21.5) ($P < 0.0001$, linear mixed

effects model) (Figure 6A-B). Although the difference between the numbers of public clonotypes at pre-infection and 6mpi was not significant (P = 0.69, linear mixed effects model), the difference between 6mpi and 3ypi was also significant (P = 0.0002, linear mixed effects model). These public clonotypes were not restricted to germline sequences and exhibited a wide distribution of V-gene deviation from germline (Figure 7).

To determine whether public clonotypes could be identified within larger subsets of donors, we compared the antibody repertoires for all combinations of two to six donors for all three timepoints (Figure 6C). Smaller numbers of clonotypes were detected as the number of donors between whom a clonotype was shared increased, but 27 clonotypes shared by at least four of the six donors also were identified (Figure 6C, Figure 19B-C). Interestingly, public antibodies present 3ypi were found predominantly only after infection, with at most ~7.5% of 3ypi antibodies shared by any pair of donors present in the pre-infection repertoire of either donor (Figure 19D). Taken together, these results suggested that a non-negligible fraction of public antibodies appeared to emerge after HIV-1 infection (Figure 19D-E) alongside the highly diverse private antibody response.

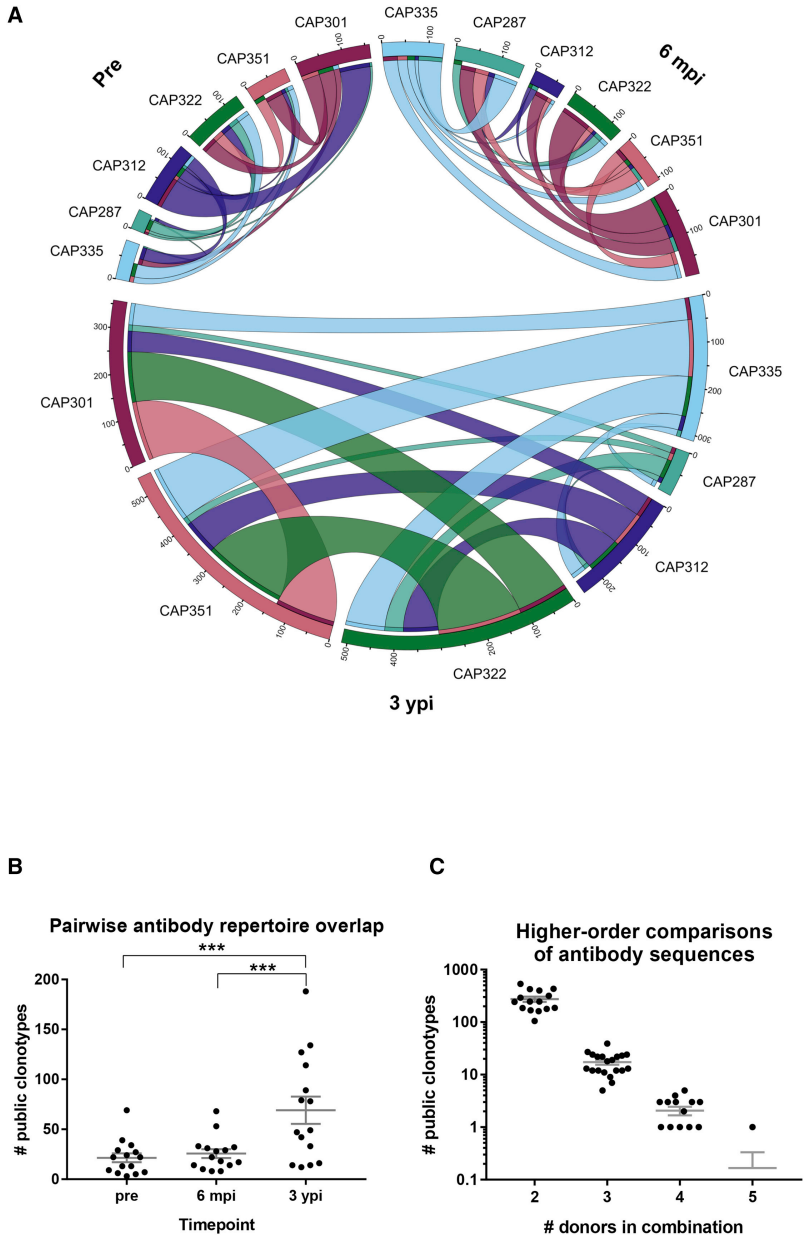


Figure 6. Identification of Public Antibody Clonotypes after Infection with HIV-1.
(A) Clonal overlap between pre-infection (top left), 6 mpi (top right), and 3 ypi (bottom) samples. The width of the curved bands connecting each pair of samples is proportional to the numbers (tick labels) of antibody clonotypes shared by that donor pair at the given time point.
(B) For each of the three time points (x axis), the numbers of public antibody clonotypes (y axis) are plotted for each pair (dots) of donors. ***p < 0.001. Error bars: mean ± SEM.
(C) For all combinations (dots) of two, three, four, and five donors (x axis), the corresponding numbers of public antibody clonotypes (y axis) across all time points are shown. Error bars: mean ± SEM.

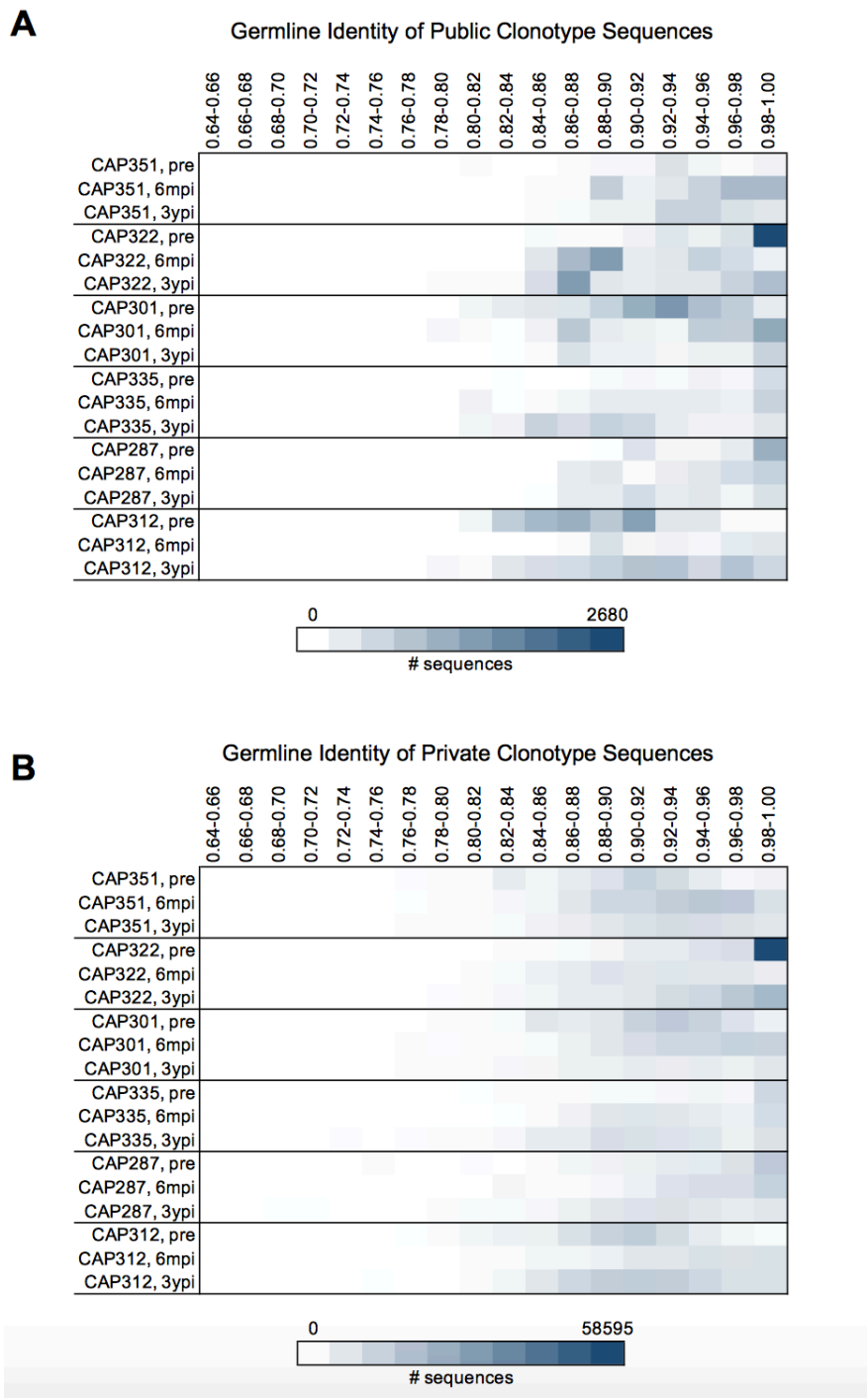


Figure 7. Somatic hypermutation distributions of the public (A) and private (B) clonotypes for each donor (row), binned (columns by identity to germline in increments of 2%, which each bin being right-open, except for 0.98-1.00, which is inclusive of 1.00. Heatmap color intensity is proportional to the number of sequences in each bin, for each sample.

2.2.4 Confirmation of HIV-1 Reactivity of a Public Antibody Clonotype

We performed paired heavy-light chain sequencing of the antigen-specific repertoire of two additional donors, CAP248 and CAP314 (Section 2.4.5, Table 2), to explore the potential of public antibodies to be HIV-1 reactive. Comparison of these sequences with the sequences from each of the other six CAPRISA donors revealed a large public clonotype between CAP248, CAP314 and CAP351 (Figure 8). Examples of representative antibodies from this public clonotype, including CAP351 CDRH3 sequences within a 1-2 amino acid difference from each of antibodies CAP248_#30 and CAP314_#30, are shown in Figure 8A. The maximum junction difference among these sequences (Figure 8A, Figure 19F) was three amino acids, or ~81% identity, for CAP248_#30 and CAP314_#30. In addition to the high CDRH3 identity, 4 of 13 somatic mutation changes from the *IGHV1-69* germline gene in the CAP248 antibody were identical in one of the representative CAP351 antibodies. Remarkably, CAP314_#30 and CAP248_#30 each used the same light chain germline gene, *IGKV1-27*, with CDRL3 sequences differing by just one amino acid (Figure 19F). As paired heavy and light chain sequence information was available for CAP248_#30 and CAP314_#30, we produced these antibodies as recombinant IgG proteins and tested their reactivity against HIV-1. The two antibodies bound ConC gp120 (Figure 8B) and neutralized tier 1 viruses MN.3 and MW965 but did not neutralize tier 2 viruses (Figure 19G). Epitope mapping revealed that antigen binding by both antibodies was affected by the D368R mutation, but not by the N332A mutation (Figure 8C), potentially suggesting a CD4-binding site epitope specificity. Overall, these results demonstrate that public HIV-reactive antibodies can emerge during the course of natural HIV-1 infection.

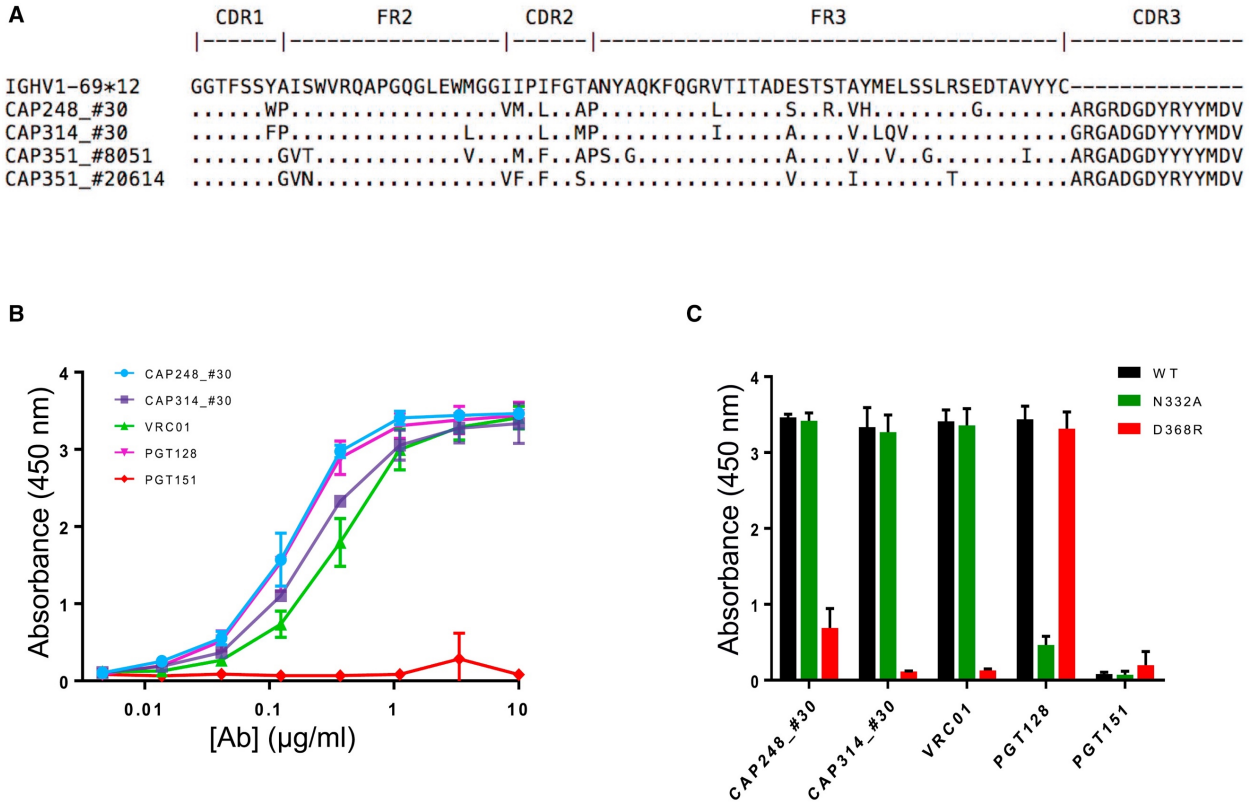


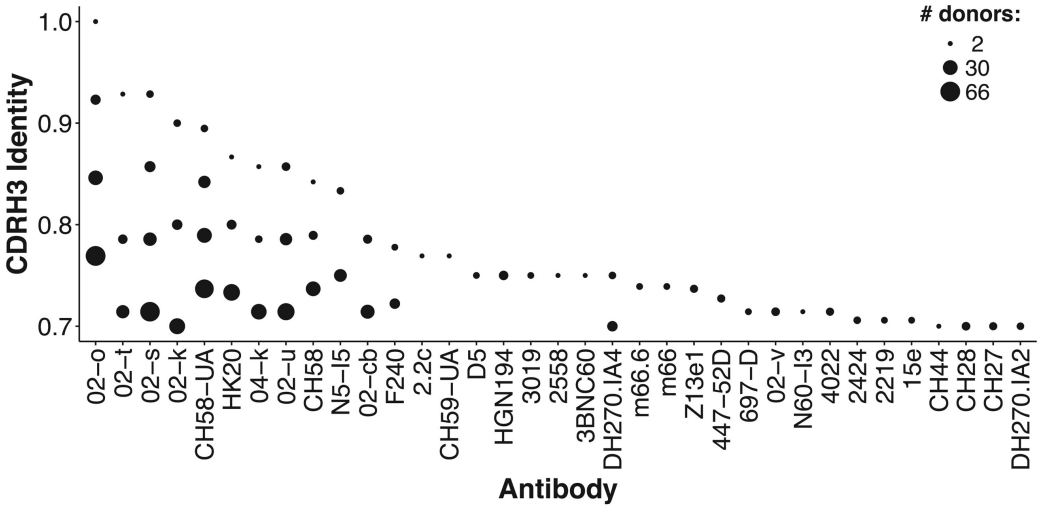
Figure 8. Characterization of a Public HIV-Reactive Antibody Clonotype Shared by Three HIV-Infected Donors.
(A) Multiple sequence alignment of the CDR1-CDR3 regions of the heavy chain sequences from a three-donor public clonotype. Included are antibodies CAP248_#30 and CAP314_#30, as well as representative CAP351 antibodies CAP351_#8051 and CAP351_#20614, along with *IGHV1-69∗12*, a top germline allele assignment for all antibodies shown. Dots within the V-gene show identity to germline, while letters show mutations from germline.
(B) ELISA binding of CAP248_#30 and CAP314_#30 to ConC gp120 at increasing antibody concentrations (x axis), with antibodies VRC01 and PGT128 as positive controls, and PGT151 as a negative control. Error bars: mean ± SD.
(C) ELISA binding at a concentration of 1.11 µg/mL of antibodies CAP248_#30 and CAP314_#30 to a wild-type ConC gp120 protein, ConC gp120 with a N332A mutation, and ConC gp120 with a D368R mutation. Control antibodies are VRC01 (D368R sensitive), PGT128 (N332A sensitive), and PGT151 (negative control). Error bars: mean ± SD.

2.2.5 Existence of Public Antibody Clonotypes With High Sequence Identity to Known HIV-1 Antibodies

In general, while HIV-1 reactivity for an antibody cannot be determined solely from its sequence, it is informative to determine whether antibodies that are similar to known HIV-reactive sequences exist at the population level, as a way to assess the uniqueness of the space of antibody HIV-1 reactivity. We compared over 18 million heavy chain antibody sequences from approximately 250 publicly available next-generation sequencing samples to the heavy chain sequences of known HIV-reactive antibodies. To that end, we collected and curated a set of heavy chain sequences from HIV-reactive antibodies found in the Protein Data Bank and CATNAP (Yoon et al., 2015) (Materials and Methods). The sequencing data were derived from infection, vaccination, and autoimmunity studies, as well as from CAPRISA samples sequenced as part of this study. We then compared the CDRH3 amino acid sequences and V-gene assignments from all sequences in each sample to CDRH3 sequences and V-gene assignments encoding HIV-reactive antibodies (Figure 9). Specific sequence features known to be important for HIV-1 recognition, such as a requisite tryptophan in the fifth position preceding the framework 4 region for 3BNC60 (Scheid et al., 2011) and gp120-binding putative VRC01 precursors (Yacoob et al., 2016), were included as additional constraints in the comparison. As expected, many observed sequences had low identity to known HIV-1 antibodies. However, thirty-five known HIV-reactive antibodies had matches with at least 70% CDRH3 identity, in addition to matching V-gene and sequence feature requirements, to antibodies in the examined samples (Figure 9A). With these search parameters, as many as ~70 donors were matched to a given HIV-

reactive antibody. Additionally, nine antibodies had matches within a difference of no more than two amino acids in the CDRH3 region, including an exact match for antibody 02-o (Yacoob et al., 2016) (Figure 9B). The nine known HIV-reactive antibodies (Figure 9B) included antibodies with extra-neutralization functions (Acharya et al., 2014), weakly neutralizing antibodies (Sabin et al., 2010), putative vaccine-elicited antibody precursors (Nicely et al., 2015), and putative bNAb precursors (Yacoob et al., 2016), and spanned a range of CDRH3 lengths from 10 to 19 amino acids. These results suggest that sequences within a short distance from HIV-reactive antibodies, including some that could be possible vaccine templates, exist at the population level.

A



B

Known Antibody *Query Antibody Source*	Reference Publication	V-gene	V-gene Identity to Germline	CDRH3 Length	CDRH3 Alignment
02-k	Yacoob *et al*	IGHV1-2	99.3%	10	ARGWGWYFDL
SRR3990873	Galson *et al*		100%		...D.....
02-o	Yacoob *et al*	IGHV1-2	99.7%	13	ARGGYSSSWYFDY
SRR5928556	de Bourcy *et al*		100%		.....
02-s	Yacoob *et al*	IGHV1-2	100%	14	ARDGGYSSGWYFDY
CAP351	This study		100%		.....D...
02-t	Yacoob *et al*	IGHV1-2	99.7%	14	ARDQADSSGWSFDY
SRR5928486	de Bourcy *et al*		100%		..A.....
02-u	Yacoob *et al*	IGHV1-2	100%	14	ARDSGYSSGWFFDY
CAP351	This study		100%		...G.....D...
04-k	Yacoob *et al*	IGHV1-2	99.7%	14	ARDSGATSDWYFDL
SRR4431788	Gupta et al		98.6%		...GW.....
CH58 UA	Nicely *et al*	IGHV5-51	100%*	19	ARLGGRYYYDSSG-YYYFDY
SRR5928502	de Bourcy *et al*		100%		..-.....Y.....
HK20	Sabin *et al*	IGHV1-69	85.4%*	15	ARASYSSSSPYAFDI
SRR2151499	Parameswaran *et al*		98.5%		..GY.....
N5-l5	Acharya et al	IGHV3-23	89.8%*	12	AKDLRLGGGSDY
SRR5928602	de Bourcy *et al*		96.8%		.....HF..

*Identity to germline determined using amino acid sequence

Figure 9. Comparison of Published Antibody Repertoires to Known HIV-Reactive Antibody Sequences.
(A) Number of donors (size of dots) with antibody heavy chain sequences with identical V-gene assignment, signature sequence features, and CDRH3 identity (y axis) of at least 70% to a set of known HIV-reactive antibodies (x axis).
(B) Antibody heavy chain sequences with identical V-gene assignment, signature sequence features, and CDRH3 distance of at most two amino acids. For each pair of known/query antibody sequences, shown are the source dataset, references, V-gene assignment, percentage identity to germline, and CDRH3 length and alignment. V-gene deviation from germline for sequences obtained from the PDB was determined from amino acid sequence using IMGT/DomainGapAlign (Brochet et al., 2008).

2.3 Discussion

Here, we interrogated the antibody heavy chain repertoires of multiple HIV-infected donors at multiple timepoints through next-generation sequencing and systems immunology analysis. Antibody sequence repertoires in chronic HIV-1 infection appeared to diverge from their pre-infection repertoires. While we cannot exclude sampling depth as a possible explanation for the observed repertoire turnover, virus-induced turnover or a more general property of B cell repertoire dynamics could also explain this observation.

The discovery of public antibody clonotypes in the setting of HIV-1 infection is intriguing, especially given the rapid viral evolution that occurs during HIV-1 infection. While previous studies (Parameswaran et al., 2013) have observed public antibodies following acute viral infections, these antigens are far less diverse than HIV-1. It is plausible that some of the identified public antibody clonotypes are not HIV-specific. In particular, existence of public clonotypes with affinity-matured sequences in pre-infection repertoires (Figure 6) could indicate shared modes of recognition at the antibody sequence level in response to common, previously encountered antigens. However, public clonotypes that are HIV-reactive also exist (Figure 8). Furthermore, public clonotypes with high sequence identity to known HIV-reactive antibodies also were identified in a variety of immune settings, including in HIV-naïve individuals (Figure 9). Our analysis identified public clonotypes for potential HIV-1 bNAb precursors, but less frequently for public mature bNAbs (Figure 9), suggesting the possibility that bNAbs may evolve from virtually identical recombination events that subsequently take

divergent evolutionary pathways in different individuals. Taken together, our findings suggest that a number of HIV-reactive antibody sequences may be readily accessible in multiple individuals, or even at the population level. Future vaccine development strategies may therefore benefit from specifically avoiding immunogen engagement with non-neutralizing public antibody clonotypes, such as the three-donor *IGHV1-69* public clonotype characterized here and aim to target engagement of bNAb or bNAb-precursor public clonotypes.

Recent studies have established the existence of public clonotypes as a common feature of the B cell repertoire (Briney et al., 2019; Soto et al., 2019). Generally, these studies used an even more stringent definition of public clonotype than used here, requiring 100% amino acid identity in the CDRH3 sequences (Briney et al., 2019; Soto et al., 2019). While we, too, explored this definition of public clonotype (Figure 19A), we concluded that this ignores somatic variants from the same lineage. We thus sought a logical medium that includes somatic variants while not being overly liberal so as to include cross-donor sequences with low levels of CDRH3 sequence identity to other sequences of a given clonotype. To implement this, we used complete linkage clustering, which requires all members of a given cluster of sequences to be within a minimum CDRH3 amino acid sequence identity threshold of all other members (rather than just to any 1 member as in single linkage clustering). These, as well as other parameters (such as requiring V allele match, using nucleotide sequence during sequence identity calculations, using alternate clustering methods, and others) will need to continually be revisited and reconsidered as new sequencing data and accompanying functional validation data accumulate.

The complexity of the humoral response to HIV-1 infection warrants large-scale analyses, such as those described here. Future juxtaposition of the antibody repertoires of broad and weak neutralizers, and from before and after infection or vaccination, could prove to be critical for identifying key correlates of protection and for iterative improvement of vaccine strategies. Such vaccine strategies could seek to recapitulate repertoire features observed to be conducive to protection or aim to interact with specific and common features of the antibody repertoire, such as public sequences (Crowe and Koff, 2015). Further exploration of these possibilities with larger infection cohorts representative of global HIV-1 infection diversity is therefore well-motivated.

2.4 Methods

2.4.1 Library Preparation and Repertoire Sequencing

For sequencing of global antibody repertoires, total RNA was extracted from PBMCs and an RT-PCR reaction was performed using a primer mix (BIOMED2 to framework 1 and framework 4) designed to amplify all heavy-chain antibody variable gene regions in an unbiased fashion. The resulting amplicons were purified, and agarose gel electrophoresis was used to confirm complete primer removal and appropriate amplicon size. The amplicon sample was quantified prior to submission to the Vanderbilt Technologies for Advanced Genomics (VANTAGE) core for paired-end (2 x 300 bp) Illumina MiSeq sequencing. The 18 samples from 6 donors were multiplexed across 3 MiSeq runs. FASTQ files from Illumina MiSeq sequencing served as the main input for subsequent data analysis. As a control, a second sequencing run was

conducted with all 18 samples on one chip (Figure 20A-B), from which we identified ~0.55 million V(D)J sequences across the 18 samples. Public sequences from run 1 were frequently recovered in the second run (Figure 20B).

2.4.2 Sequencing Data Preprocessing and Clonal Analysis

Preprocessing was carried out using pRESTO (Vander Heiden et al., 2014) as follows: *1)* Paired-end reads were interleaved and reads with a mean Phred quality score below 20 were removed. *2)* Orientation of sequences was corrected to the forward orientation (V to J) as necessary. *3)* V-region primers were masked and C-region primers were cut. Sequences with no match to primers were discarded. *4)* Duplicate sequences were removed and a duplication count of each sequence was annotated. *5)* All sequences with duplication count of 1 were removed. A wide distribution of duplication counts was observed (Figure 20C). *6)* Each sequence was annotated for V, D, and J gene usage using IgBLAST (Ye et al., 2013), using reference sequences from IMGT (Lefranc et al., 2015). After removing non-functional sequences and sequences with CDRH3 lengths of under 5 amino acids, clonal clustering of all 1316148 V(D)J sequences from all donors was performed using Change-O (Gupta et al., 2015) (Figure 20D). Functional V_H V(D)J sequences were assigned to clonal groups by first grouping sequences based on common *IGHV* gene annotation, *IGHJ* gene annotation and junction region lengths. *IGHV* and *IGHJ* gene annotations for each group of sequences were determined by the first gene assignment of gene assignments within each junction length. Within these larger groups, sequences differing from one another by a threshold distance within the junction region were defined as clonotypes by

complete-linkage clustering. Distance was determined using an amino acid Hamming distance normalized to the length of the junction. We used amino acid distance during clustering because this is the determinant of antibody molecular recognition; however, nucleotide and amino acid identities were generally very similar (Figure 20E).

2.4.4 Single-Cell Sorting

Paired chain antibody sequencing for CAP248 and CAP314 was carried out by Atreca (Redwood City, CA) on IgG cells sorted into microtiter plates at one cell per well by FACS. Briefly, cryopreserved PBMCs were stained with the following antibodies: CD14-FITC (HCD14), CD3-FITC (UCHT1), IgM-A488 (MHM-88), IgD-A488 (IA6-2), CD19-BV421 (HIB19) or CD19-PE (SJ25C1), CD20-PECy7 or CD20-BV711 (2H7), CD38-PECy7 or CD38-PerCPCy5.5 (HIT2), CD27-BV510 (O323) from BioLegend and IgA-FITC (IS11-8E10) from Miltenyi. For CAP248 PBMCs, antigen-specific cells were isolated using CAP45 SOSIP.664 trimeric protein. For CAP314 PBMCs, antigen-specific cells were isolated using ConC gp120. The sorted antigen-specific B cells were cultured for 4 days in IMDM medium (Invitrogen) in the presence of FBS, Normocin, IL-2 (PeproTech), IL-21 (PeproTech), rCD40 ligand (R&D Systems), and His-Tag antibodies (R&D Systems), prior to single cell sequencing.

2.4.5 Paired Chain Antibody Sequencing

Generation of barcoded cDNA, PCR amplification, and next-generation sequencing of paired IgG heavy & light chains were performed as described in (DeFalco et al., 2018), with the following modifications: desthiobiotinylated oligo (dT) and RT

maxima H- (Fisher Scientific Company) were used for reverse transcription, DynaBeads™ MyOne™ Streptavidin C1 (Life Technologies) was used to isolate desthiobiotinylated cDNA , PCR amplicon concentrations were determined using qPCR (KAPA SYBR® FAST qPCR Kit for Illumina, Kapabiosystems), and amplicons were sequenced on an Illumina MiSeq instrument.

2.4.6 Barcode Assignment, Sequence Assembly, Assignment of V(D)J and Identification of Mutations in Paired-Chain Sequencing

Fastq output files were grouped and parsed into separate Fastq files on the basis of their compound ID (plate-ID + well-ID). We used Atreca proprietary software to assemble paired end reads into consensus sequences, requiring a minimum coverage of 30 reads for each heavy and each light chain assembly. Wells with more than one contig for a given chain type were rejected from consideration unless one of the contigs included at least 90% of the reads. V(D)J assignment and mutation identification was performed using IgBLAST (Ye et al., 2013). Allele assignments (Figure 8A, Figure 19) were determined using IMGT/V-Quest (Brochet et al., 2008).

2.4.7 Comparison of Antibody Repertoire Sequencing Data to Known HIV-Reactive Antibodies

A list of HIV-reactive antibodies was curated manually from the Protein Databank and CATNAP (Yoon et al., 2015), and CDRH3 sequence and V-gene usage of each antibody was determined using ANARCI (Dunbar and Deane, 2016) with IMGT numbering. Publicly available antibody sequencing datasets collected from the Short

Read Archive and the European Nucleotide Archive, in addition to the samples presented in this study, were processed via MiXCR (Bolotin et al., 2015). Briefly, reads in each sample were aligned with the *mixcr align* command and clustered by junction sequence and V-gene assignment using the *mixcr assemble* command with the –*OseparateByV=true* flag. First and last residues of the junction were trimmed to match the IMGT CDRH3 definition, and the first V-gene assignment in the alignment score-ranked list of possible gene assignments was used for comparison to the list of HIV-reactive antibodies. Within each sample, only sequences with a matching V-gene assignment and CDRH3 length to each HIV-1 antibody were compared for CDRH3 identity. Sequence identities were calculated using the *editdistance* library in Python and normalized to the length of the CDRH3 sequence. For antibodies 3BNC60, 02-cb, 02-o, 02-s, 02-u, 02-k, 02-t, 02-v, and 04-k, the requirement of a tryptophan in the fifth position preceding the framework 4 region also was imposed. Each accession number used in this study was manually annotated for donor ID in order to count the number of donors with matches to each HIV-reactive antibody (Figure 9). To retrieve identity to germline of MiXCR-processed data, alignments were exported using the *exportAlignments* command including the *-vBestIdentityPercent* flag. This calculates the fraction of matching nucleotides with the germline V-gene divided by the alignment length, which varied among publicly available samples due to differing sample preparation methods and sequencing strategies.

2.4.8 Neutralization Fingerprinting

Neutralization fingerprinting analysis of donor sera was performed as described previously (Doria-Rose et al., 2017). Briefly, epitope-specific neutralization fingerprints were constructed for ten antibody specificities against a panel of 21 diverse HIV-1 strains (Georgiev et al., 2013). For each serum, the serum-virus neutralization data against the same panel of HIV-1 strains was compared to the epitope-specific antibody fingerprints, in order to estimate the relative contribution of each of these reference antibody specificities to the polyclonal serum neutralization. For each serum, the estimated neutralization contributions by each of the ten reference antibody specificities were reported on a scale of 0 to 1, with all specificities adding up to 1.

2.4.9 Quantification and Statistical Analysis

To determine statistical significance of pairwise donor overlap across timepoints (Figure 6), a linear mixed effects model was fit with timepoints designated as fixed effects and donors treated as random effects, thereby accounting for any correlation resulting from each donor being part of 5 pairwise comparisons per timepoint. Briefly, each donor was represented by a 45-dimensional vector, with each dimension having a value of 1 or 0 based on if that donor was part of each of the 45 pairwise comparisons that occurred across the 3 timepoints. The linear mixed effects model was then fit using the *lmer* function from the lme4 package (Douglas et al., 2015) in R. P values were determined using the lmerTest package (Kuznetsova et al., 2014) in R. Other statistical tests were performed using R version 3.4.1.

CHAPTER 3

HIGH-THROUGHPUT MAPPING OF B CELL RECEPTOR SEQUENCES TO ANTIGEN SPECIFICITY FOR HIGH-DIMENSIONAL SCREENS OF SINGLE B CELLS

The text in this chapter is a reproduction of Setliff et al., 2019. Author contributions are indicated in Appendix 2.

3.1 Rationale

The antibody repertoire – the collection of antibodies present in an individual – responds efficiently to invading pathogens due to 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 therapeutics, 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 bait (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 sequences of the variable heavy and variable light immunoglobulin genes 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 BCR heavy and light chain amplicons can be expressed and screened as Fab (Wang et al., 2018) or scFV (Adler et al., 2017a, 2017b) in a yeast display system. Though these various antibody discovery technologies have identified a large number of potently 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 next-generation sequencing-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 PBMC samples from two HIV-infected subjects, and from these, we successfully identified HIV- and influenza-specific antibodies, including both known broadly neutralizing antibody (bNAb) lineages and a new bNAb. LIBRA-seq is high-throughput, scalable, and applicable to many targets. This single, integrated assay enables the mapping of monoclonal antibody sequences to panels of diverse antigens theoretically unlimited in number and facilitates the rapid identification of cross-reactive antibodies that may serve as therapeutics or vaccine templates.

3.2 Results

3.2.1 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 10A).

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 differing 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 IgM BCR sequences (Weaver et al., 2016). To that end, we chose two well-characterized BCRs: VRC01, a CD4-binding site-directed HIV-1 broadly neutralizing antibody (bNAb) (Wu et al., 2010), and Fe53, a bNAb recognizing the stem of group 1 influenza hemagglutinins (HA) (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 variants of the trimeric HIV-1 Env protein from strains BG505 and CZA97 (Georgiev et al., 2015; van Gils et al., 2013; Ringe et al., 2017), and trimeric hemagglutinin from strain H1 A/New Caledonia/20/1999 (Whittle et al., 2014) (Figure 10B, Figure 21A-C).

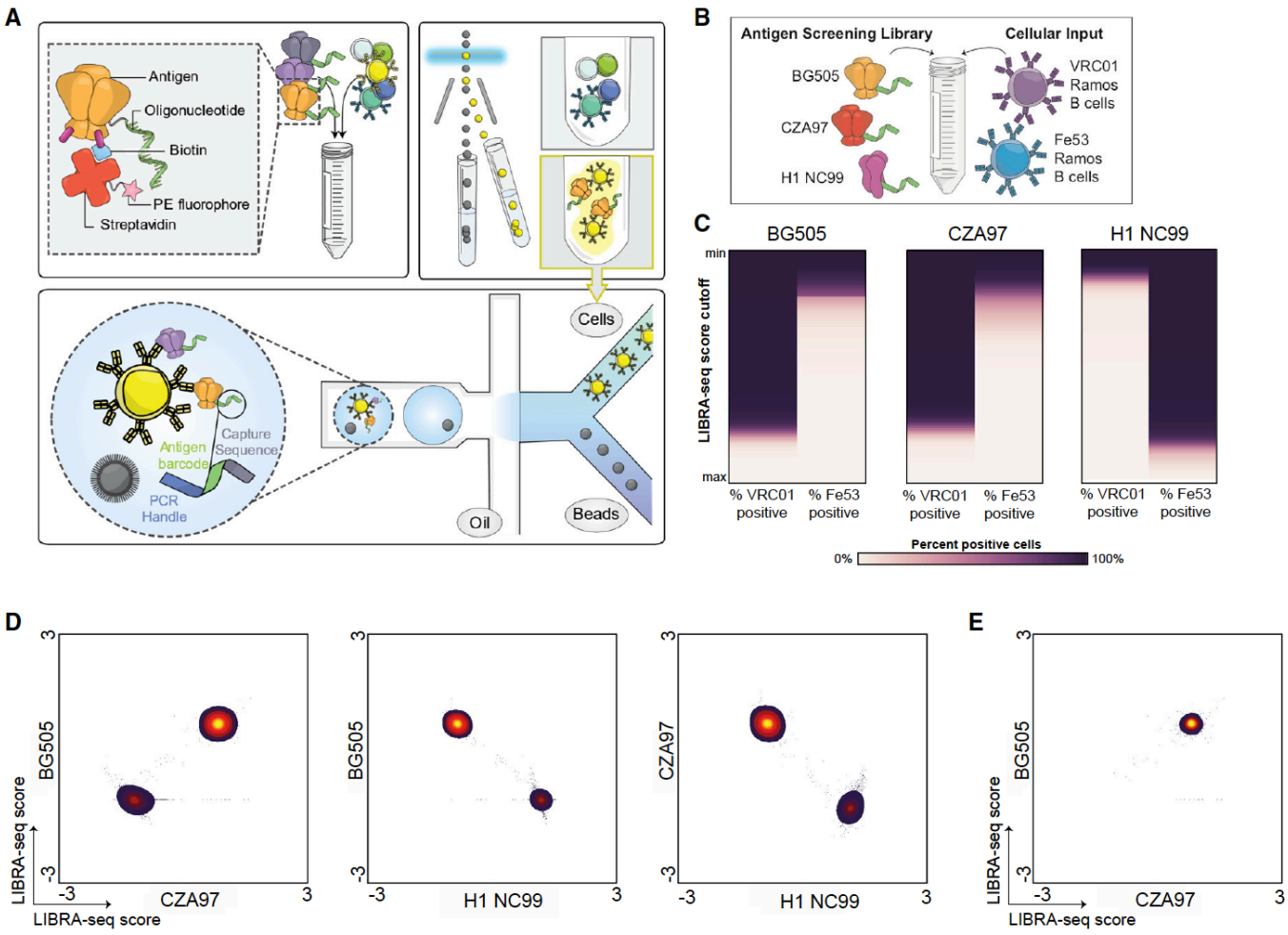


Figure 10. LIBRA-seq Assay Schematic and Validation.
(A) Schematic of the LIBRA-seq assay: (top left) fluorescently labeled, DNA-barcoded antigens are used to (top right) sort antigen-positive B cells before (bottom) co-encapsulation of single B cells with bead-delivered oligos using droplet microfluidics. 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.

We recovered 2321 cells with BCR sequence and antigen mapping information, highlighting the high-throughput potential of LIBRA-seq (Figure 22D). 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; therefore, scores serve as a proxy for the relative amount of bound antigen (Methods). The LIBRA-seq scores of each individual antigen reliably categorized Ramos B cells by their specificity (Figure 10C). 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 10D). Further, VRC01 Ramos B cells bound both BG505 and CZA97 and had a high Pearson correlation (Pearson's r=0.84), demonstrating that LIBRA-seq readily identifies B cells that bind to multiple HIV-1 antigens (Figure 10E).

3.2.2 Isolation of Antibodies From a Known HIV bNAb Lineage

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 bNAbs 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 experiments and recovered paired V_H:V_L antibody sequences with antigen mapping for 866 cells (Figure 11A, Figure 22D, Figure 12A). These B cells exhibited a variety of LIBRA-seq scores among the three antigens (Figure 11B), as can be expected from a polyclonal sample possessing a wide variety of B cell specificities and antigen affinities. The cells displayed a few discrete patterns based on their LIBRA-seq scores; we observed cells that were (1) $HA^{high}Env^{low}$ or (2) $HA^{low}Env^{high}$ (Figure 11B). 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 11B).

To further validate the utility of LIBRA-seq in monoclonal antibody isolation, we next sought to identify new members of 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 11C). All newly identified BCRs had high levels of somatic hypermutation and utilized IGHV1-2*02 along with the characteristic five-residue CDRL3 paired with IGVK3-20 (Figure 11C). These B cells came from multiple known clades of the VRC01 lineage, with sequences with high identity and phylogenetic relatedness to lineage members VRC01, VRC02, VRC03, VRC07, VRC08, NIH45-46, and others (Figure 11C). Of these, 25 (86%) had a high LIBRA-seq score for at least 1 HIV-1 antigen, three (10%) had mid-range scores

(between mean and one standard deviation) for at least 1 HIV-1 antigen, and only one of the VRC01 lineage B cells had scores less than the mean for both HIV-1 antigens (Figure 11C, Supplemental Figure 12B). These results suggest that the LIBRA-seq platform can be successfully used to down-select cross-reactive bNAbs in prospective antibody discovery efforts.

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 13A) and neutralized two Tier 1 viruses but no viruses on a global panel of representative HIV-1 strains (deCamp et al., 2014) (Figure 13B). 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 12A and 12B). 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.

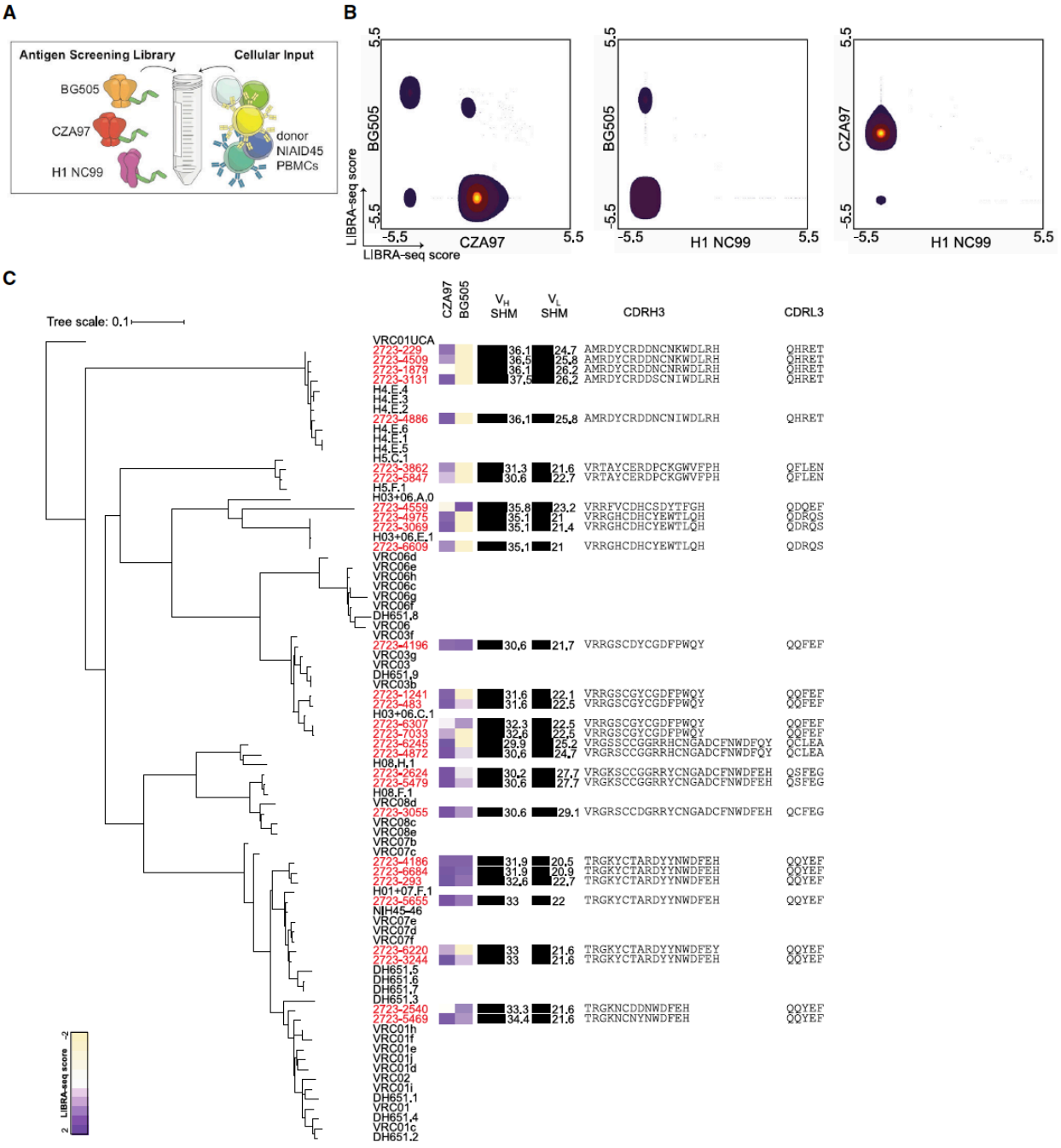
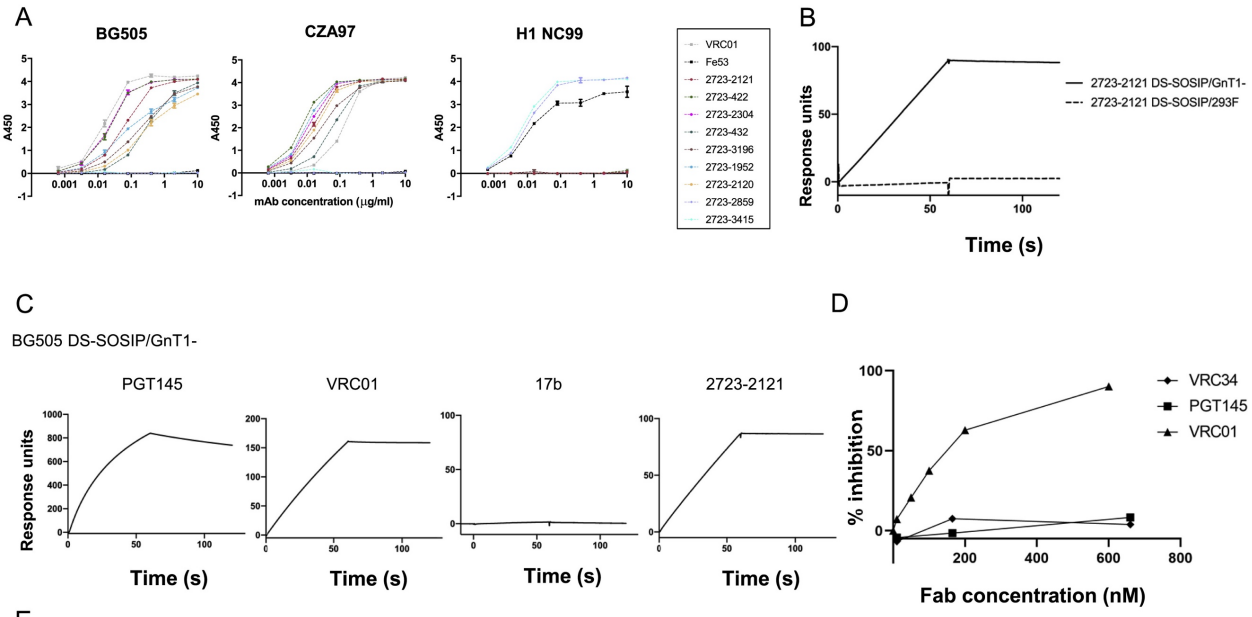


Figure 11. LIBRA-seq Applied to a Human B Cell Sample from HIV-Infected Subject NIAID45.

(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 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).



E

Virus phenotype	Env-pseudoviruses	IC_{50} (µg/ml) 2723-2121	VRC01
Tier 1	MW965.26	0.067	0.025
	MN.3	<0.020	0.044
	6644.V2.C33	2.441	0.139
Tier 2 (sorting antigens)	BG505.W6M.C2	>50	0.035
	BG505.W6M.C2.T332N	>50	0.06
	CZA097	>50	0.09
Tier 2 (global panel)	TRO.11	>50	0.226
	CH119.10	>50	0.71
	CNE55	>50	0.156
	25710.2.43	21.568	0.47
	CE0217	>50	0.136
	X1632.S2.B10	>50	0.15
	X2278.C2.B6	>50	0.091
	246.F3.C10.2	>50	0.321
	398.F1.F6.20	1.76	0.13
	CE1176.A3	>50	1.17
	BJOX002000.03.2	>50	>50
Control	MLV	>50	>50

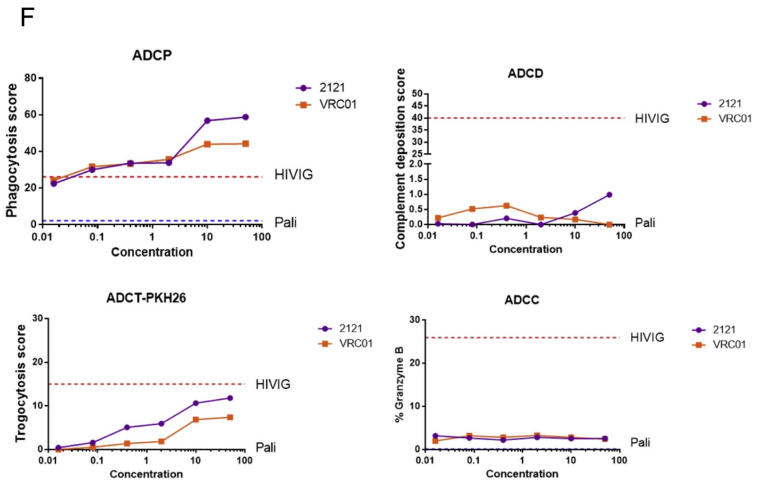


Figure 12. Characterization of Antibodies from Donor NIAID45.
(A) Antigen specificity as predicted by LIBRA-seq was validated by ELISA for a variety of antibodies isolated from donor NIAID45. Antibodies were tested for binding to BG505, CZA97, and H1 A/New Caledonia/20/99. Data are represented as mean ± 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 µg/ml) needed for 50%

inhibition (IC_{50}).
(F) Levels of ADCP, ADCD, 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.

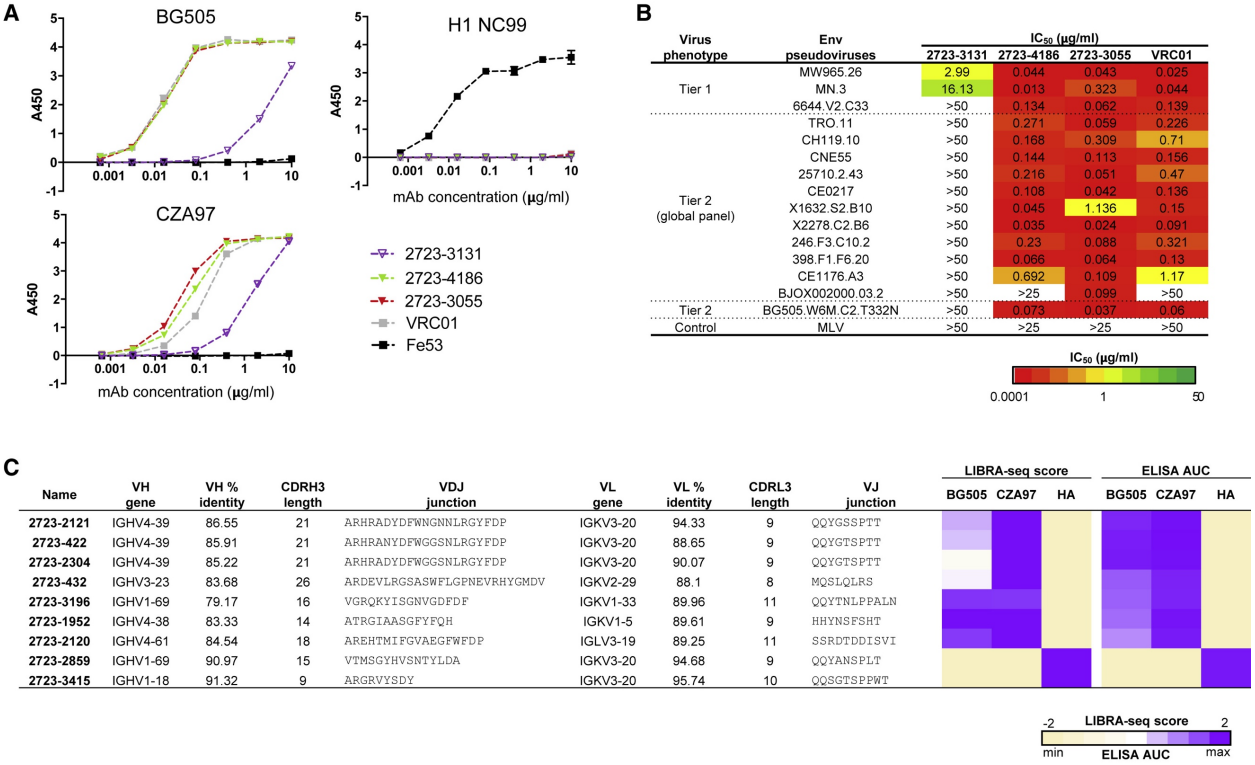


Figure 13. Characterization of LIBRA-seq-Identified Antibodies from Donor NIAID45.
(A) Antigen specificity as predicted by LIBRA-seq was validated by ELISA for a subset of monoclonal antibodies belonging to the VRC01 lineage. Data are represented as mean ± SEM for one ELISA experiment. ELISA data are 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. IC_{50} values are shown from high potency (0.0001 µg/ml, red) to low potency (50 µg/ml, green). Lack of neutralization IC_{50} 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 11A, 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.

3.2.3 Identification of 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 13C). We selected these seven antibodies because all had high LIBRA-seq scores for at 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 (Figure 13C, Figure 22A). We further characterized one of these antibodies, 2723-2121, and determined that it bound to a stabilized BG505 trimer (DS-SOSIP) (Do Kwon et al., 2015) by surface plasmon resonance (SPR) (Figure 12B, Figure 12C). Antibody 2723-2121 competed for trimer binding with VRC01 (Figure 12D), neutralized three tier 1 pseudoviruses and 2 of 11 tier 2 pseudoviruses from a global panel (Figure 12E), and mediated trogocytosis and antibody-dependent cellular phagocytosis (Figure 12F). 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 13C). 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 (Figure 13C, Figure 12A).

3.2.4 Discovery of an HIV bNAb 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 14A). 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) (van Gils 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) (Figure 14A,

Figure 21). After applying LIBRA-seq to donor N90 PBMCs, we recovered paired V_H:V_L antibody sequences with antigen mapping for 1,465 cells (Figure 21D, Figure 23A). Within this set of cells, we identified 18 B cells that were members of the VRC38 lineage (Figure 14B). 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 14B).

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) (Figure 14C, Figure 14D, Figure 15A). We observed 32 cells that had high LIBRA-seq scores for three of the four influenza antigens (Figure 14C); 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; $p = 0.015$) (Figure 15A, Figure 23B).

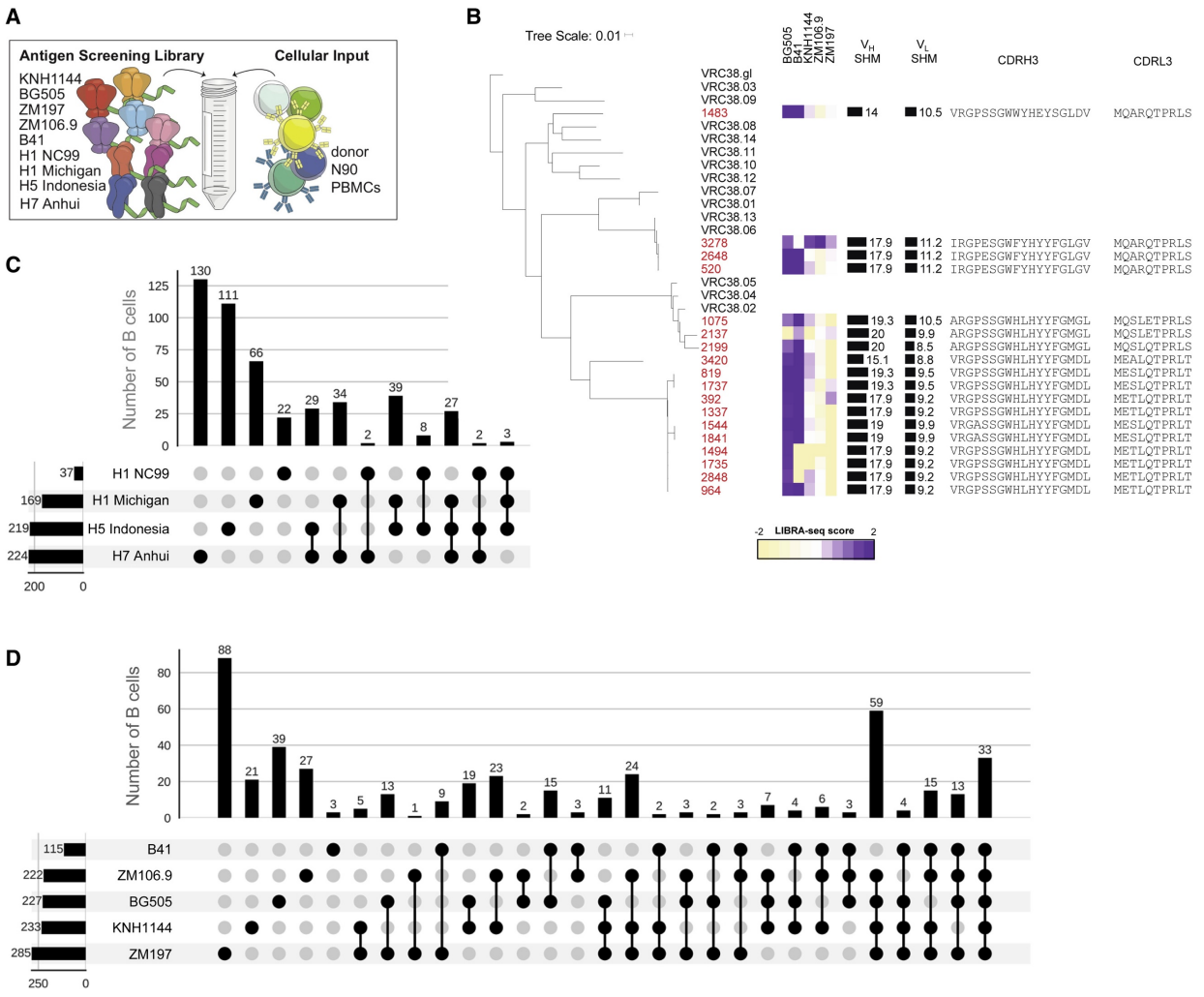


Figure 14. LIBRA-seq Applied to a Sample from NIAID Donor N90.
(A) The LIBRA-seq experiment setup consisted of nine antigens in the screening library: 5 HIV-1 Env (KNH1144, BG505, ZM197, ZM106.9, and B41) and 4 influenza HA (H1 A/New Caledonia/20/99, H1 A/Michigan/45/2015, H5 Indonesia/5/2005, and H7 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 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.

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 14D). 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 16A). 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) (Figure 16A, Figure 16B, Figure 23B). Of note, 3602-870 neutralized BG505 and ZM197, both of which were used as probes in the antigen screening library (Figure 15B). 3602-870 bound BG505 DS-SOSIP by SPR and competed for BG505 DS-SOSIP binding with VRC01 Fab (Figure 15C, Figure 23C).

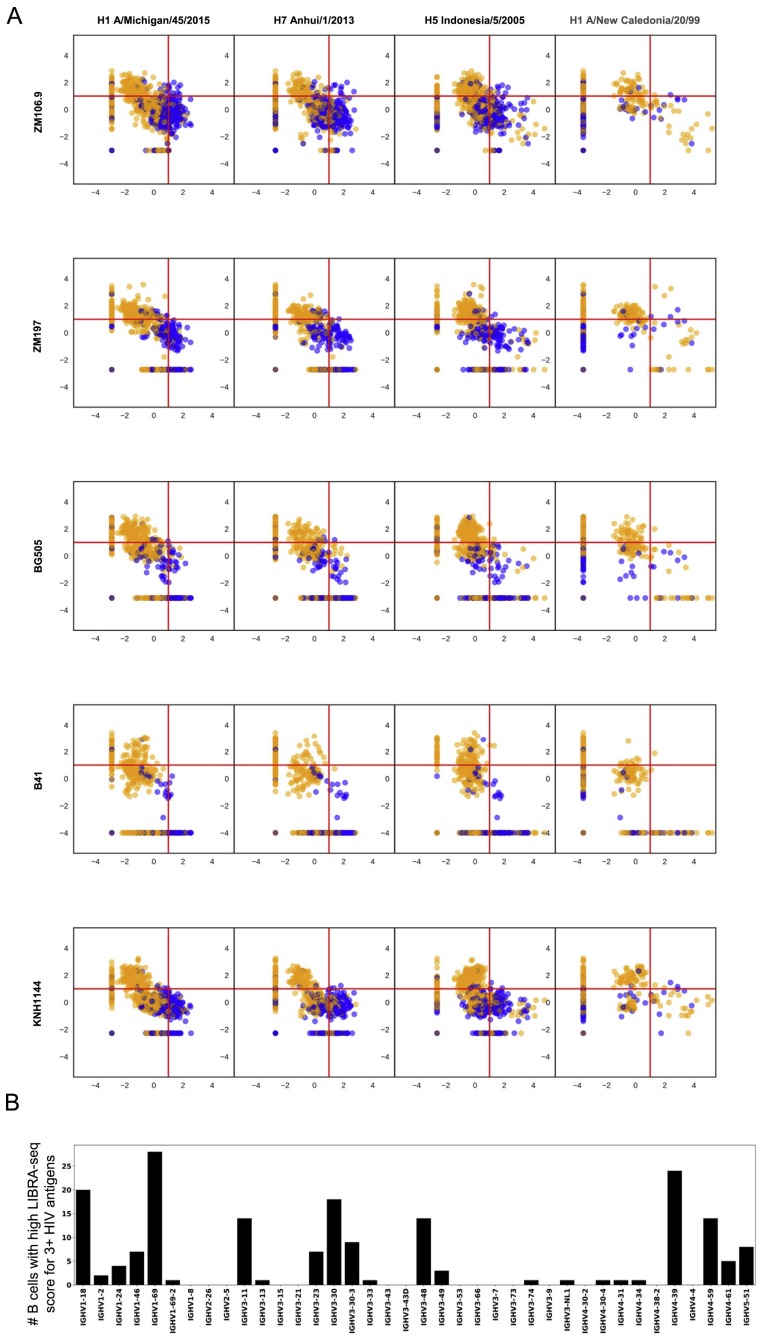


Figure 15. Analysis of Antigen Reactivity for B Cells from Donor N90.
(A) Each graph shows the LIBRA-seq score for an HIV antigen (y-axes) versus an influenza antigen (x-axes) in the screening library. The 901 cells that had a LIBRA-seq score above one for at least one antigen are displayed as individual 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.

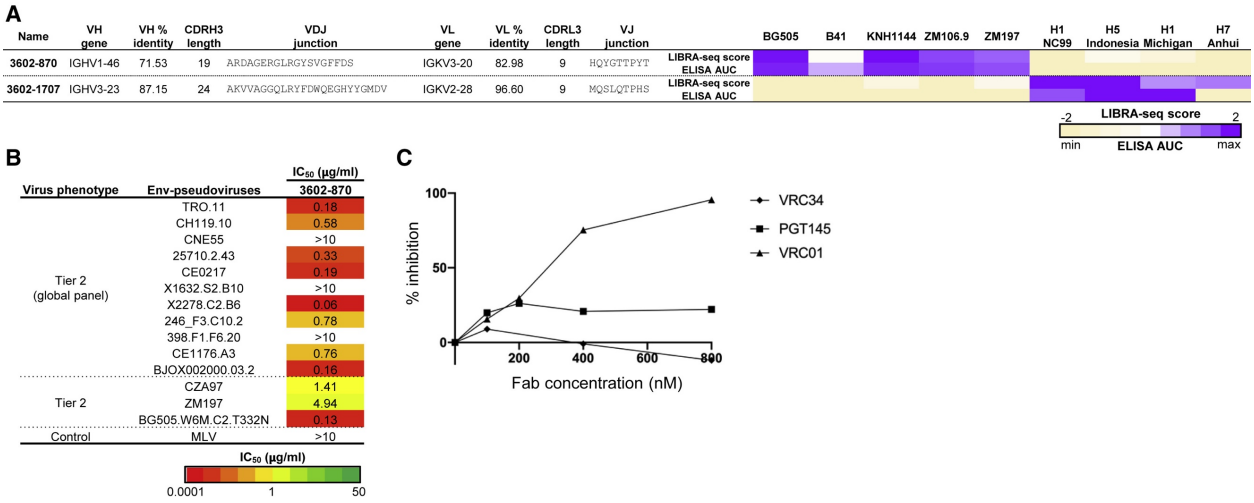


Figure 16. Characterization of LIBRA-seq-Identified Antibodies from NIAID Donor N90. (A) Sequence characteristics and antigen specificity of newly identified antibodies from donor N90. Percent identity is calculated at the nucleotide level, and CDR length and sequences are noted at the amino acid level. 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 23B, 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. IC_{50}values are shown from high potency (0.0001 μg/ml, red) to low potency (50 μg/ml, green). Lack of neutralization IC_{50} 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).

3.3 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 17A). 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 (Figure 17B, Figure 14B).

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; Sok et al., 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.

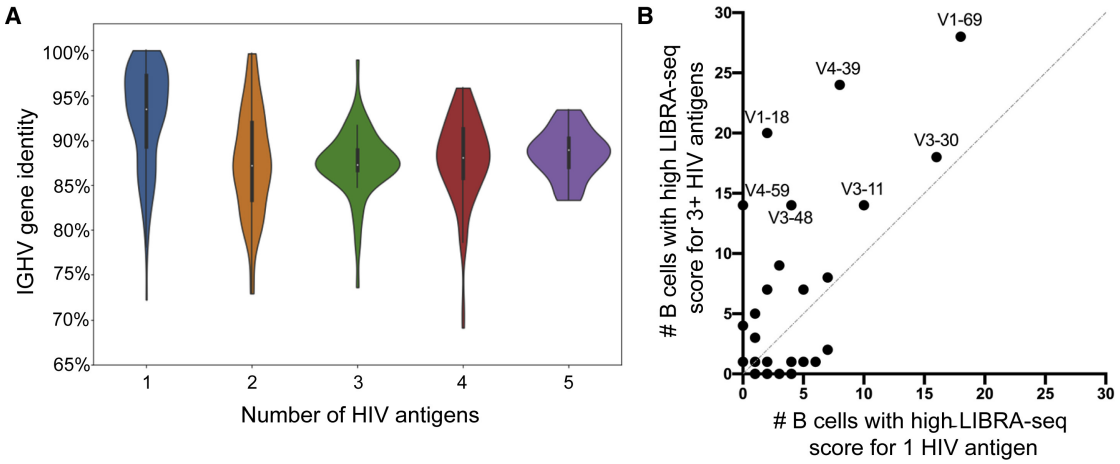


Figure 17. Sequence Properties of the Antigen-Specific B Cell Repertoire
(A) *IGHV* gene identity (y axis) is plotted for cells with high (≥1) LIBRA-seq scores for any combination of 1–5 HIV-1 SOSIP antigens (x axis). Each distribution is displayed as a kernel density estimation, where wider sections 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.

3.4 Methods

3.4.1 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% CO_2 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 fibritin 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 LN_2 and stored at –80C° 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).

3.4.2 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′-CCTTGGCACCCGAGAATTCCANNNNNNNNNNNNCCCATATAAGA*A*A-3′, where Ns represent the antigen barcode. For the cell line and NIAID45 experiments, we used the following antigen barcodes: CATGATTGGCTCA (BG505), TGTCCGGCAATAA (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), GCAGCGTATAAGTCA (B41), ATCGTCGAGAGCTAG (BG505), CAGGTCCCTTATTTC (A/Indonesia/5/2005), ACAATTTGTCTGCGA (A/Anhui/1/2013), TGACCTTCCTCTCCT (A/Michigan/45/2015), AATCACGGTCCTTGT (A/New Caledonia/20/99). Oligos were ordered from Sigma-Aldrich and IDT with a 5′ amino modification and HPLC purified.

3.4.3 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 crosslinker, 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.

3.4.4 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.

3.4.5 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).

3.4.6 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.

3.4.7 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 24A). In a histogram of insert lengths, this resulted in a sharp peak of the expected insert size of 52-54 bp (Figure 24B-D). 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.

3.4.8 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 CLRs 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.

3.4.9 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.

3.4.10 Antibody Expression and Purification

For each antibody, variable genes were inserted into plasmids encoding the constant region for the heavy chain (pFUSEss-CHIg-hG1, Invivogen) and light chain (pFUSE2ss-CLIg-hl2, Invivogen and pFUSE2ss-CLIg-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% CO_2 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.

3.4.11 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.

3.4.12 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 (IC_{50}).

3.4.13 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 (Do 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 µl/min. The surface was regenerated between injections by flowing over 3M $MgCl_2$ solution for 10 s with flow rate of 100 µl/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.

3.4.15 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.

3.4.16 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.

CHAPTER 4

HIGH-THROUGHPUT EPITOPE BINNING USING LIBRA-SEQ

4.1 Rationale

In some cases, depending on the goals of a given antibody discovery program, discovery efforts may focus on the identification of antibodies with a particular epitope specificity. For example, in the case of HIV, a particular antibody discovery program may seek antibodies specific for the CD4-binding site, V1V2 apex, V3-glycan, or fusion peptide, or membrane proximal external region. Alternatively, rather than isolating antibodies to a known epitope, the identification of antibodies targeting novel epitopes may help in the development of new antibody-based therapeutics and vaccine templates. As such, substantial effort has focused on developing approaches for characterizing the different types of antibody specificities that are present in the polyclonal B cell response to infection or vaccination.

Previous methods include computational delineation of individual specificities from functional or biophysical assays (Chung et al., 2015; Georgiev et al., 2013), and deconvolution of the polyclonal secreted antibody pool using electron microscopy (Bianchi et al., 2018). None of these technologies, however, can provide simultaneous information about both antibody sequence and target epitope for single B cells in a polyclonal sample.

Thus, we sought to conduct a pilot study to determine if the LIBRA-seq technology (described in Chapter 3) could be extended to enable high-throughput epitope mapping for thousands of single B cells in parallel. The combination of this extension with the initial implementation of LIBRA-seq could, in principle, allow for: (1) screening thousands (or more) of single B cells against many diverse antigen specificities, (2) determination of residue-level epitope information for each B cell, and (3) identification of the monoclonal antibody sequence, all during the initial screening step. The many dimensions of data obtained for each single B cell would enable rapid prioritization of monoclonal antibody leads, and potentially greatly accelerate the identification of therapeutic molecules.

4.2 Results

4.2.1 Production of an Epitope Screening Library

To initially test the ability to map epitope specificity for single B cells, we created an antigen library composed of BG505 single-chain variants, including a wild-type protein and 3 mutants: D36R, N160K, and K169E. Each of the mutant proteins differed from the wild-type protein by a single residue. We reasoned that in a mixture of HIV-specific B cells and Env proteins, where interactions between the B cells and Env proteins are governed by the law of mass action, sensitivity to mutagenesis of a key residue in a particular epitope may be detectible in our sequencing-based binding readout. Further, lack of sensitivity for a given B cell to other mutations should result in similar binding signals among variants with wild-type sequences in the B cell's epitope.

In addition, trimeric hemagglutinin from strain H1 A/New Caledonia/20/1999 was included in the screening library as overall negative control.

4.2.2 Detection of Sensitivity to D36R Mutation by Thousands of CD4-Binding Site-Directed B Cells

To test the ability of LIBRA-seq to accurately map BCR sequences to their epitope specificities, we devised a proof-of-principle mapping experiment using two Ramos B-cell lines with differing BCR sequences and antigen specificities (Weaver et al., 2016). Similar to the validation experiment described in Chapter 3, we chose two well-characterized BCRs: VRC01, a CD4-binding site-directed HIV-1 broadly neutralizing antibody (bNAb) (Wu et al., 2010), and Fe53, a bNAb recognizing the stem of group 1 influenza hemagglutinins (HA) (Lingwood et al., 2012). We mixed these two populations of B-cell lines at a 80:20 ratio in favor of VRC01 and incubated them with the DNA-barcoded antigens in the screening library. We then performed LIBRA-seq as described in (Setliff et al., 2019).

We recovered 7047 cells that passed quality control filtering. Of these, 5595 (79.4%) were VRC01 cells as determined by the BCR sequence; we focused subsequent analysis on this subset of cells. Upon examining the distributions of LIBRA-seq scores for the 4 BG505 antigen variants, we observed consistently lower scores for the D368R mutant when comparing to other BG505 variants, in agreement with the known reduced VRC01 affinity for D368R (Wu et al., 2010) (Figure 18). We also observed highly similar distributions for the BG505 variants with wild-type CD4-bs sequences (Figure 18). Together, these observations suggest that LIBRA-seq can be

used to successfully identify residue-level epitope information for many single B cells in parallel.

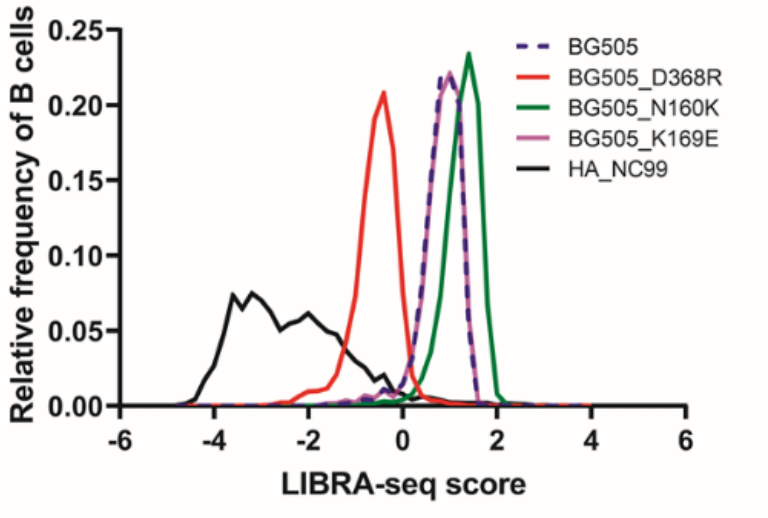


Figure 18. High-throughput epitope mapping with LIBRA-seq. Displayed are distributions of frequencies of B cells (y-axis) of LIBRA-seq scores (x-axis) for 4 BG404 variants and one influenza hemagglutinin.

4.3 Discussion

Here, we demonstrated in a preliminary pilot study that through customization of the antigen screening library, LIBRA-seq can be used to extract residue-level epitope information from a simple sequencing-based readout. Though follow-up studies are certainly necessary, the ability to detect differential binding to single point mutations of a given protein at the single-cell level represents an important aspect of antibody discovery. Indeed, the ability to down-select monoclonal antibody leads based on epitope specificity may accelerate antibody discovery efforts moving forward.

In this initial study, we used a limited set of epitope knockouts. Using an expanded medium or large antigen set should provide substantially more

comprehensive epitope mapping for large numbers of single B cells simultaneously. Further, by screening all residues on the surface of a given protein, such as through the creation of an alanine screening library, it may be possible to identify novel epitopes. An alanine screening approach could be complemented by considering the naturally occurring sequence variation in *Env,* and optimized using computational protein modelling suites, such as OSPREY (Gainza et al., 2013; Georgiev and Donald, 2007; Georgiev et al., 2006, 2008b, 2008a).

Moving forward, it will be important to take into account several considerations that may impact the readout of larger LIRBA-seq experiments seeking to simultaneously consider multiple pathogen targets, multiple antigen variants for a given pathogen, and residue-level epitope information. For example, the effect of number of epitope mutant proteins for a given antigen variant in a screening library on the LIBRA-seq score will eventually need to be systematically explored. For this, we may consider small (~5 epitope-knockout mutations), medium (~30) and large (~100) screening libraries. Such studies will then need to be combined other within-pathogen antigen variants to explore the impact of competition of antigen variants for a given BCR.

Though there is still much work to be done, the ability to query antibody epitope through next-generation sequencing represents an exciting direction for antibody science moving forward. LIBRA-seq's ability to rapidly generate large amounts of antibody – epitope information should enable fundamental discoveries in antibody science and serve as a rich source of future computational data mining studies.

4.4 Methods

4.4.1 Production of Antigen Screening Library

The single-chain variant (Georgiev et al., 2015) of wild-type BG505, along with point mutants D368R, N160K and K169E (point mutants produced by mutagenesis), were each expressed containing an AviTag 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% CO_2 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 H1 A/New Caledonia/20/1999 contained the HA ectodomain with a point mutation at the sialic acid-binding site (Y98F), T4 fibritin 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 LN_2 and stored at –80C° until use.

4.4.2 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, with minor modifications in order to intercept, amplify and purify the antigen barcode libraries (Setliff et al., 2019; Stoeckius et al., 2017). Following library construction, we sequenced both BCR and antigen barcode libraries on approximately

2.5% of a NovaSeq 6000 flow cell at the VANTAGE sequencing core, with a target 10% of this fraction dedicated to antigen barcode libraries.

4.4.3 Processing of BCR Sequence and Antigen Barcode Reads and Determination of LIBRA-seq Score

We utilized the computational pipeline explained in Chapter 3 and (Setliff et al., 2019) to process the antigen barcode reads. Briefly, the pipeline 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 as described in (Setliff et al., 2019), resulting in a cell barcode - antigen barcode UMI count matrix. Additionally, we aligned the BCR contigs 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. LIBRA-seq scores were calculated as described in (Setliff et al., 2019).

4.4.4 Quantification and Statistical Analysis

LIBRA-seq score distributions were plotted using GraphPad Prism version 8.0.0.

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

5.1 Conclusions and Future Directions

We developed and validated computational and experimental methods to profile immune responses and guide antibody discovery using molecular genomics technologies. Our search for repertoire-level biomarkers (e.g., V gene usage) of immunological status led to the identification of antibody sequence convergence among HIV-infected subjects, as well as the identification of public putative bNAb precursors in HIV-naïve individuals, including VRC01 precursors (Setliff et al., 2018). This raises the possibility of taking advantage of these seemingly readily available sequences by targeting public clonotypes with priming immunogens. However, our studies using deep sequencing of the heavy chain repertoire also highlighted a fundamental problem with traditional B cell repertoire sequencing: sequence does not beget function. We thus sought to overcome this by coupling sequence and functional information. To do so, we devised a sequencing-based readout for antibody specificity determination (Setliff et al., 2019). The method, LIBRA-seq, is applicable to many targets, accurate, facilitates the rapid identification of cross-reactive antibodies, and can be performed as a single, integrated assay within a few hours (plus sequencing time).

The extension of LIBRA-seq to additional functional properties, such as residue-level epitope, will be an important evolution of the technology. This will require

screening B cells expressing known antibodies against large libraries of single- and double-residue knockouts, in various background antigens from diverse pathogens. Using a similar approach, it may also be possible to determine a relative affinity of each antibody for antigen using LIBRA-seq's sequencing-based readout.

An additional important future extension is related to the types of antigens that can be used in the screening library. So far, we have focused on recombinant soluble antigens. However, it may be the case that the precise antigen is unknown. For viruses, this may be surmountable by using the whole virus. Multiple strains of whole virus may be used to stain a B cell population of interest, with the viral genetic material serving as the barcode (Russell et al., 2019). Future extensions should also endeavor to enable screening of multi-pass membrane proteins, perhaps through the use of virus-like particles or membrane mimetics.

Even more custom applications should be able to be devised using custom microfluidics. For example, by using microfluidics solutions other than those offered by 10X Genomics, it will be possible to assay antibody-secreting cells (ASCs). By capturing single ASCs in individual droplets, without immediate lysis, a number of functional assays can be devised in-droplet using sequencing as the primary readout. For example, receptor-blocking assays may be able to be implemented by using barcoded ligands. This would allow LIBRA-seq to probe not just specificity, but also the mechanism of action of the antibody.

Even when sticking to standard microfluidics solutions, as the cell throughput of standard commercial microfluidics solutions increases, we anticipate much of the AIRR-seq field to transition to protocols that include functional annotation at single-cell

resolution, such as LIBRA-seq. We have therefore focused on making the technology largely accessible by using commercially available solutions (10X Genomics) for single-cell isolation via droplet microfluidics and single-cell barcoding. However, the method is general and can be used in other commercially available or custom homebrew set-ups.

Here, we have focused on characterizing the immune response to HIV infection. However, because our technologies enable rapid, high-resolution interrogation of the immune response to any antigen(s) conducive to DNA barcoding, they may be used in many diverse disease settings or in antibody engineering efforts. For example, in vaccine studies, LIBRA-seq may be used to iteratively improve multi-valent vaccine designs or monitor the development of breadth. Alternatively, if attempting to develop a single antibody therapeutic molecule that can be used in both pre-clinical and clinical studies, it may be desirable to screen against multiple species orthologs (mouse, cynomolgus macaque, human, etc) and/or allelic variants; LIBRA-seq, in principle, readily facilitates such screens. Indeed, the rapid isolation of antibody 3602-870 using a nine-antigen screen via LIBRA-seq demonstrates the utility of high-dimensional screens for antibody discovery.

No matter the application, the ability the leverage next-generation sequencing-based technologies to measure functional aspects of B cells with single-cell resolution should result in a large accumulation of antibody sequence : binding specificity information. These data will be useful in large-scale data mining efforts, which may subsequently uncover more fundamental properties of antibody-antigen interactions that can be fed forward into future immunogen design strategies.

PUBLICATION LIST

Peer-Reviewed Articles Related to Dissertation

Setliff, I., Shiakolas, A.R., Pilewski, K.A., Murji, A.A., Mapengo, R.E., Janowska, K., Richardson, S., Oosthuysen, C., Raju, N., Ronsard, L., et al. (2019). High-Throughput Mapping of B Cell Receptor Sequences to Antigen Specificity. *Cell 179*, P1636-1646.E15. doi: 10.1016/j.cell.2019.11.003.

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. doi: 10.1016/j.chom.2018.05.001.

Peer-Reviewed Co-Authored Articles

Dussupt V, Sankhala RS, Gromowski GD, Donofrio G, De La Barrera R, Larocca R, Zaky W, Mendez-Rivera L, Choe M, Davidson E, Abbink P, Bai H, Bias CH, Berry IM, Botero N, Cook T, Doria-Rose N, Geretz A, Hernandez M, Jian N, Kabra K, Leggat D, Liu J, Rutvisuttinunt W, **Setliff I**, Tran U, Townsley S, Doranz BJ, Rolland M, McDermott AB, Georgiev IG, Thomas R, Robb ML, Eckels K, Jarman R, Stephenson KE, Barouch DH, Modjarrad K, Michael NL, Joyce MG, Krebs SJ. “Potent Zika and dengue cross-neutralizing antibodies induced by Zika vaccination in a dengue-experienced donor.” Accepted, *Nature Medicine*.

Keating CL, Kuhn E, Bals J, Cocco A, Yousif A, Matysiak C, Sangesland M, Ronsard L, Smoot M*,* Thalia Bracamonte M, Okonkwo V, **Setliff I**, Georgiev I, Balazs AB, Carr SA, Lingwood D. Spontaneous glycan reattachment following N-glycanase treatment of influenza and HIV vaccine antigens. *Journal of Proteome Research,* 2020. doi: 10.1021/acs.jproteome.9b00620.

Magaret CA, Benseker DC, Williamson BD, Boratel BR, Carpp LN, Georgiev IS, **Setliff I**, Dingens AS, Simon N, Carone M, Simpkins C, Montefiori D, Alter G, Yu W, Juraska M, Edlefsen PT, Karunal S, Mgodi NM, Edugupanti S, Gilbert PB. "Prediction of VRC01 neutralization sensitivity by HIV-1 gp160 sequence features." *PLOS Computational Biology*, 2019. doi: 10.1371/journal.pcbi.1006952.

Raju N, **Setliff I**, Georgiev IS. "NFPws: A web server for delineating broadly neutralizing antibody specificities from serum HIV-1 neutralization data." *Bioinformatics*, 2019. doi: 10.1093/bioinformatics/btz097.

Koethe J, McDonnell WJ, Kennedy A, Abana C, Pilkington M, **Setliff I**, Barnett L, Hager C, Smith R, Kalams S, Hasty A, Mallal SA. "Adipose tissue is enriched for activated and senescent CD8+ T cells, and shows distinct CD8+ receptor usage

compared to blood in HIV-infected persons." *JAIDS*, 2017. doi: 10.1097/QAI.0000000000001573.

Patents Related to Dissertation

Full Patent Applications

Setliff I, Georgiev IS, Shiakolas AR. "System and Methods for Simultaneous Detection of Antigens and Antigen Specific Antibodies." PCT/US19/43570.

Setliff I, Georgiev IS, Morris L. "Conserved HIV Antibody Clonotypes and Methods of Use." PCT/US19/14121.

Provisional Patent Applications

Setliff I, Georgiev IS. "Methods for Identification of Antigen Binding Specificity of Antibodies." 62/913,432.

Setliff I, Georgiev IS. "Methods for Identification of Antigen Binding Specificity of Antibodies." 62/895,687.

Setliff I, Georgiev IS, Shiakolas AR. "System and Methods for Simultaneous Detection of Antigens and Antigen Specific Antibodies." 62/854,437.

Setliff I, Georgiev IS, Shiakolas AR. "System and Methods for Simultaneous Detection of Antigens and Antigen Specific Antibodies." 62/818,864.

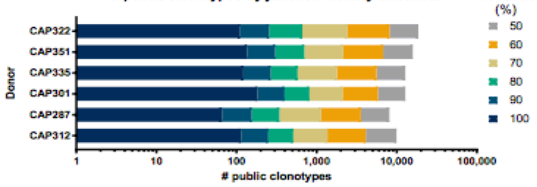
Setliff I, Georgiev IS, Shiakolas AR. "Simultaneous Recovery of Antigen Receptor Sequence and Specificity in Single B Cells." 62/716,013.

Setliff I, Georgiev IS, Morris L. "Conserved HIV Antibody Clonotypes and Methods of Use." 62/619,266.

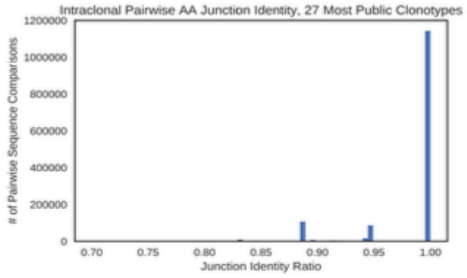
Setliff I, Georgiev IS, Morris L. "Conserved HIV Antibody Clonotypes and Methods of Use." 62/645,964.

APPENDIX 1

A

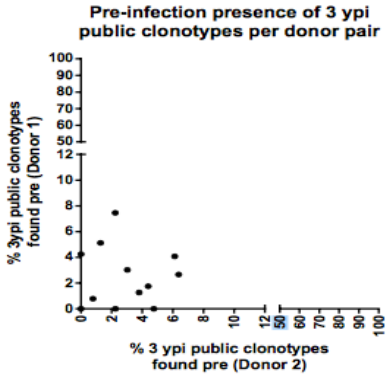


B



C

D



E

CAP322 — *P* = 1.000

3 ypi present pre-infection	Public (2+ donors) Yes	Public (2+ donors) No
Yes	43	201
No	962	13714

CAP312 — *P* = 0.981

3 ypi present pre-infection	Public (2+ donors) Yes	Public (2+ donors) No
Yes	2	10
No	418	8008

CAP351 — *P* = 0.841

3 ypi present pre-infection	Yes	No
Yes	22	256
No	894	12660

CAP301 — *P* = 0.978

3 ypi present pre-infection	Yes	No
Yes	19	112
No	591	5698

CAP335 — *P* = 0.747

3 ypi present pre-infection	Yes	No
Yes	5	55
No	465	6084

CAP287 — *P* = 0.813

3 ypi present pre-infection	Yes	No
Yes	1	11
No	146	2056

F

```
                FR1                 CDR1        FR2         CDR2                FR3                                 CDR3        FR4
                |-------------------|-----|-----------|-----|-----------------------------------|-----------|---------|

IGHV1-69*12     QVQLVQSGAEVKKPGSSVKVSCKASGGTFSSYAISWVRQAPGQGLEWMGGIIPIFGTANYAQKFQGRVTITADESTSTAYMELSSLRSEDTAVYYC----------------------
CAP248_#30      ..............................WP.................VM.L..AP.........L.....S..R.VH.......G.......ARGRDGDYRYYMDVWGEGTPVTVSS
CAP314_#30      .......................R.........FP..............L.....L..MP.........I.....A....V.LQV.............GRGADGDYYYYMDVWGDGTTVIVSA

                FR1                 CDR1        FR2         CDR2                FR3                       CDR3        FR4
                |-------------------|-----|-----------|-----|-----------------------------------|---------|---------|

IGKV1-27*01     DIQMTQSPSSLSASVGDRVTITCRASQGISNYLAWYQQKPGKVPKLLIYAASTLQSGVPSRFSGSGSGTDFTLTISSLQPEDVATYYC--------------------
CAP314_#30.lc   ................K...........RH.................G....H.............H...........F.....QKYDSGPRTFGPGTKVDFK
CAP248_#30.lc   ................................T.................G.....................Y...........F.....QKYDSAPRTFGQGTNVDLK
```

G

Antibody	Pseudovirus MN.3	Pseudovirus MW965
CAP248_#30	0.1086	0.0133
CAP314_#30	0.1651	0.0278

Figure 19. Analysis of public antibody clonotypes in HIV-1 infection.
(A) Number of public clonotypes (x-axis) for different junction identity clustering thresholds (colors) for each donor (y-axis).
(B) Number of pairwise comparisons (y-axis) with given amino acid junction identities (x-axis) between pairs of members within each of the 27 most public clonotypes.
(C) Logo plots of 27 public clonotypes identified in 4 or more of the 6 CAPRISA donors whose global antibody repertoires were sequenced. Plots were generated using unique donor-deduplicated CDRH3 sequences and WebLogo as described in (Crooks et al., 2004). Briefly, the height of each amino acid is proportional to sequence conservation at that position, and the amino acids are colored by physicochemical properties. Member CDRH3 sequences in a given public clonotype sequence were deduplicated to unique CDRH3s and then used to generate a consensus sequence. These consensus sequences were manually curated to ensure that each sequence was present in its donor, and then used to build the final inter-donor consensus sequences displayed here.
(D) Retention of public clonotypes during chronic infection with HIV-1. For each pair of donors (axes), shown is the fraction of shared clonotypes at 3ypi that were also found pre-infection.
(E) Contingency tables testing for presence of 3ypi public/private clonotypes (x-axis) in pre-infection repertoires (y-axis) of each donor; *P* values are from Fisher's exact test with simulated *P* value from 1000 bootstraps.
(F) Multiple sequence alignment of the heavy and light chain sequences from antibodies CAP248_#30 and CAP314_#30, along with the respective germline genes. Dots within the V-gene indicate no changes from germline, while letters show mutations from germline.
(G) Neutralization data for antibodies CAP248_#30 and CAP314_#30. Displayed are the IC_{50} values of each antibody against tier 1 viruses MN.3 and MW965.

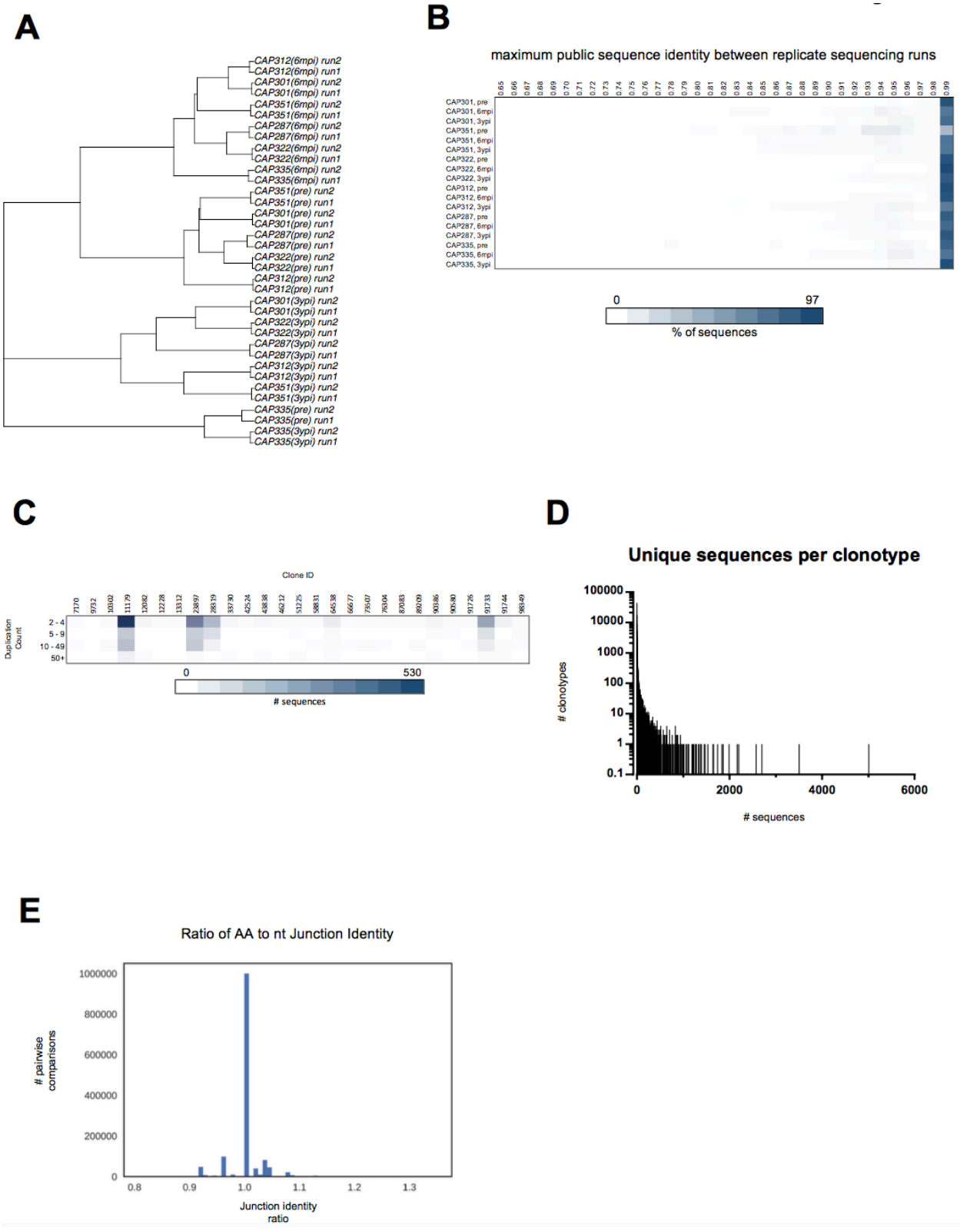

A
CAP312(6mpi) run2
CAP312(6mpi) run1
CAP301(6mpi) run2
CAP301(6mpi) run1
CAP351(6mpi) run2
CAP351(6mpi) run1
CAP287(6mpi) run2
CAP287(6mpi) run1
CAP322(6mpi) run2
CAP322(6mpi) run1
CAP335(6mpi) run2
CAP335(6mpi) run1
CAP351(pre) run2
CAP351(pre) run1
CAP301(pre) run2
CAP301(pre) run1
CAP287(pre) run2
CAP287(pre) run1
CAP322(pre) run2
CAP322(pre) run1
CAP312(pre) run2
CAP312(pre) run1
CAP301(3ypi) run2
CAP301(3ypi) run1
CAP322(3ypi) run2
CAP322(3ypi) run1
CAP287(3ypi) run2
CAP287(3ypi) run1
CAP312(3ypi) run2
CAP312(3ypi) run1
CAP351(3ypi) run2
CAP351(3ypi) run1
CAP335(pre) run2
CAP335(pre) run1
CAP335(3ypi) run2
CAP335(3ypi) run1
B
maximum public sequence identity between replicate sequencing runs
CAP301, pre
CAP301, 6mpi
CAP301, 3ypi
CAP351, pre
CAP351, 6mpi
CAP351, 3ypi
CAP322, pre
CAP322, 6mpi
CAP322, 3ypi
CAP312, pre
CAP312, 6mpi
CAP312, 3ypi
CAP287, pre
CAP287, 6mpi
CAP287, 3ypi
CAP335, pre
CAP335, 6mpi
CAP335, 3ypi
0
97
% of sequences
C
Clone ID
Duplication Count
2 - 4
5 - 9
10 - 49
50+
0
530
sequences
D
Unique sequences per clonotype
clonotypes
100000
10000
1000
100
10
1
0.1
0
2000
4000
6000
sequences
E
Ratio of AA to nt Junction Identity
pairwise comparisons
1000000
800000
600000
400000
200000
0
0.8
0.9
1.0
1.1
1.2
1.3
Junction identity ratio

Figure 20. Analysis of sequencing data properties.

(A) Repertoire similarity of technical replicates for the same donor was very high, as assessed using the repertoire similarity metric reported by VDJtools as described in (Shugay et al., 2015). An all-vs-all comparison of repertoires was performed, and hierarchical clustering was performed based on repertoire similarity as measured by overlap of identical CDRH3 sequences. In all cases, replicates were most similar to each other.

(B) For each public sequence (CDR1 through CDR3) in each sample (rows) of the original sequencing run, the maximum identity (columns) found in the corresponding sample of the duplicate sequencing run was calculated. Heatmap color intensity is proportional to the density of antibody clonotypes for each binned identity value for each sample. Bin labels indicate the left bound, with all bins right-open except 0.99, which is inclusive of 1.00.

(C) Duplication count (rows) per VDJ sequence in each of the 27 most public clonotypes (columns).

(D) Number of clonotypes (y-axis) with given number of member unique VDJ sequences (x-axis), for the 70% junction identity threshold plus matched V- and J-gene, using complete linkage clustering.

(E) Number of pairwise comparisons of sequences (y-axis) with given ratios of amino acid junction identity and nucleotide junction identity (x-axis) between pairs of members of the 27 most public clonotypes. Only member sequences belonging to the same clonotype were compared to one another.

A

CAP256-VRC26	VRC26.12	VRC26.11	VRC26.10	VRC26.09	VRC26.08	VRC26.07	VRC26.06	VRC26.05	VRC26.04	VRC26.03	VRC26.02	VRC26.01
VRC26.12		0.615	0.615	0.609	0.634	0.769	0.725	0.692	0.743	0.743	0.615	0.384
VRC26.11	0.615		0.743	0.609	0.585	0.641	0.575	0.692	0.666	0.666	0.666	0.333
VRC26.10	0.615	0.743		0.658	0.609	0.666	0.625	0.769	0.692	0.692	0.897	0.435
VRC26.09	0.609	0.609	0.658		0.926	0.78	0.547	0.658	0.804	0.804	0.634	0.341
VRC26.08	0.634	0.585	0.609	0.926		0.731	0.571	0.682	0.756	0.756	0.585	0.317
VRC26.07	0.769	0.641	0.666	0.78	0.731		0.65	0.641	0.948	0.923	0.692	0.384
VRC26.06	0.725	0.575	0.625	0.547	0.571	0.65		0.65	0.65	0.65	0.65	0.4
VRC26.05	0.692	0.692	0.769	0.658	0.682	0.641	0.65		0.666	0.666	0.743	0.487
VRC26.04	0.743	0.666	0.692	0.804	0.756	0.948	0.65	0.666		0.974	0.692	0.384
VRC26.03	0.743	0.666	0.692	0.804	0.756	0.923	0.65	0.666	0.974		0.692	0.384
VRC26.02	0.615	0.666	0.897	0.634	0.585	0.692	0.65	0.743	0.692	0.692		0.435
VRC26.01	0.384	0.333	0.435	0.341	0.317	0.384	0.4	0.487	0.384	0.384	0.435	

B

CH235	CH235.10	CH235.11	CH235.12	CH235.13	CH235.6	CH235.7	CH235.8	CH235.9
CH235.10		0.588	0.705	0.705	0.47	0.647	0.529	0.705
CH235.11	0.588		0.588	0.588	0.352	0.529	0.411	0.588
CH235.12	0.705	0.588		1	0.529	0.882	0.647	1
CH235.13	0.705	0.588	1		0.529	0.882	0.647	1
CH235.6	0.47	0.352	0.529	0.529		0.588	0.529	0.529
CH235.7	0.647	0.529	0.882	0.882	0.588		0.647	0.882
CH235.8	0.529	0.411	0.647	0.647	0.529	0.647		0.647
CH235.9	0.705	0.588	1	1	0.529	0.882	0.647	

Table 5. Intra-clonal junction region identity of known broadly neutralizing HIV antibodies.

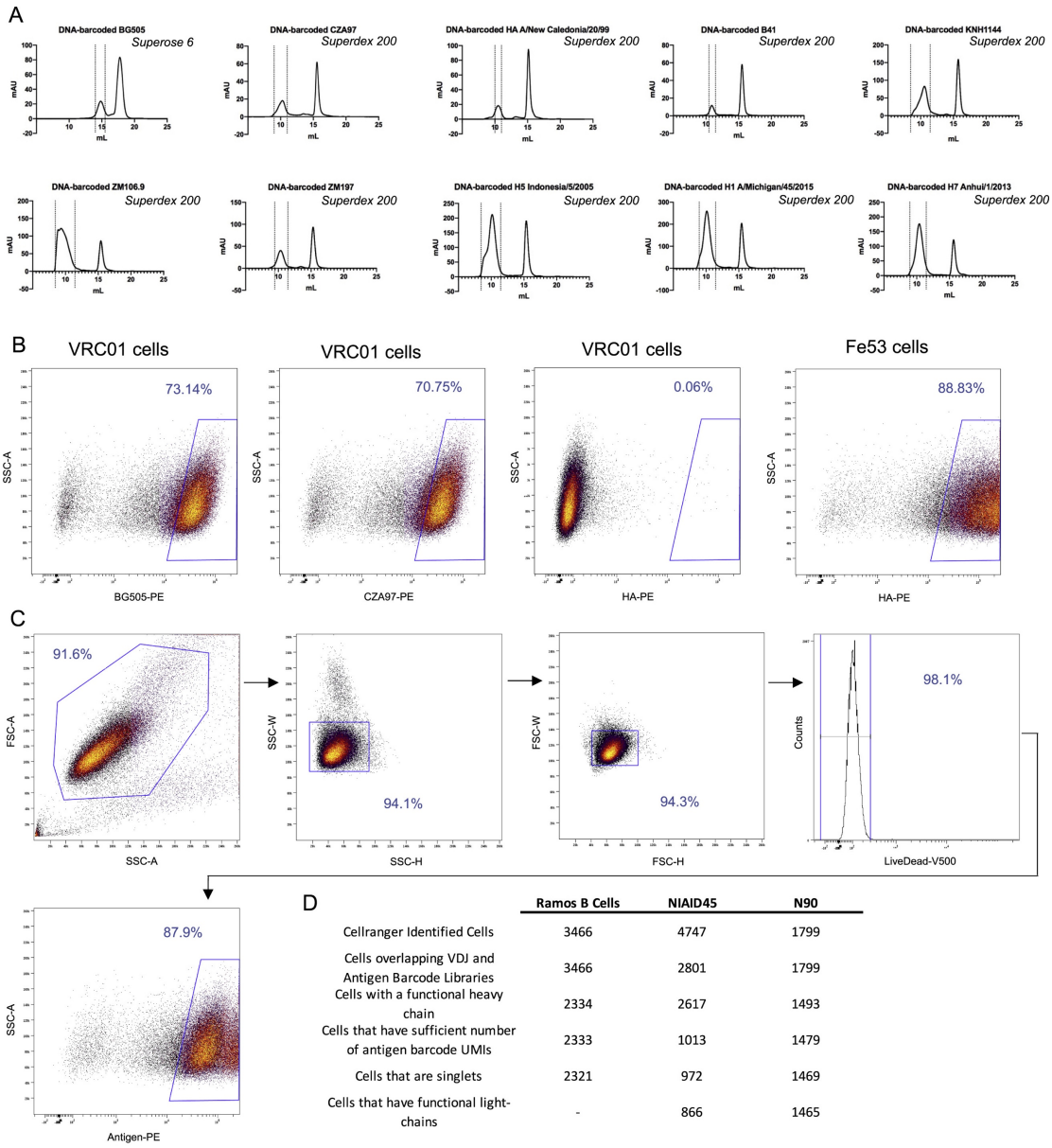
(A) Pairwise junction identity matrix of members of the VRC26 broadly neutralizing antibody lineage (Doria-Rose et al., 2014).

(B) Pairwise junction identity matrix of members of the CH235 broadly neutralizing antibody lineage (Bonsignori et al., 2016).

Contributions to Chapter 2

Conceptualization:	Ivelin Georgiev, Lynn Morris
Methodology:	Ivelin Georgiev, Lynn Morris, Ian Setliff, Wyatt McDonnell, Nagarajan Raju
Samples:	Nigel Garrett, Salim Abdool Karim
Software:	Ian Setliff, Nagarajan Raju, Wyatt McDonnell
Validation:	Ian Setliff, Wyatt McDonnell, Nagarajan Raju, Bryan Shepherd
Investigation:	Ian Setliff, Wyatt McDonnell, Nagarajan Raju, Robin Bombardi, Cathrine Scheepers, Charissa Oosthuysen, Lynn Morris, Ivelin Georgiev, Alusha Mamchak
Resources:	Ivelin Georgiev, Lynn Morris, Simon Mallal, James Crowe
Data Curation:	Ian Setliff, Nagarjan Raju, Wyatt McDonnell
Writing – Original Draft:	Ian Setliff, Wyatt McDonnell, Ivelin Georgiev
Writing – Review & Editing:	Ian Setliff, Wyatt McDonnell, Nagarajan Raju, Robin Bombardi, Amyn Murji, Cathrine Scheepers, Rutendo Mapengo, Charissa Oosthuysen, Bryan Shepherd, Alusha Mamchak, Nigel Garrett, Salim Abdool Karim, Simon Mallal, James Crowe, Lynn Morris, Ivelin Georgiev
Visualization:	Ian Setliff, Wyatt McDonnell, Ivelin Georgiev, Amyn Murji
Supervision:	Ivelin Georgiev, Lynn Morris
Project Administration:	Ivelin Georgiev, Lynn Morris
Funding Acquisition:	Ivelin Georgiev, Ian Setliff, Lynn Morris, Simon Mallal, James Crowe

APPENDIX 2



D

	Ramos B Cells	NIAID45	N90
Cellranger Identified Cells	3466	4747	1799
Cells overlapping VDJ and Antigen Barcode Libraries	3466	2801	1799
Cells with a functional heavy chain	2334	2617	1493
Cells that have sufficient number of antigen barcode UMIs	2333	1013	1479
Cells that are singlets	2321	972	1469
Cells that have functional light-chains	-	866	1465

Figure 21. Purification of DNA-barcoded antigens and LIBRA-seq validation sorting schematic on Ramos B cell lines.

(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).

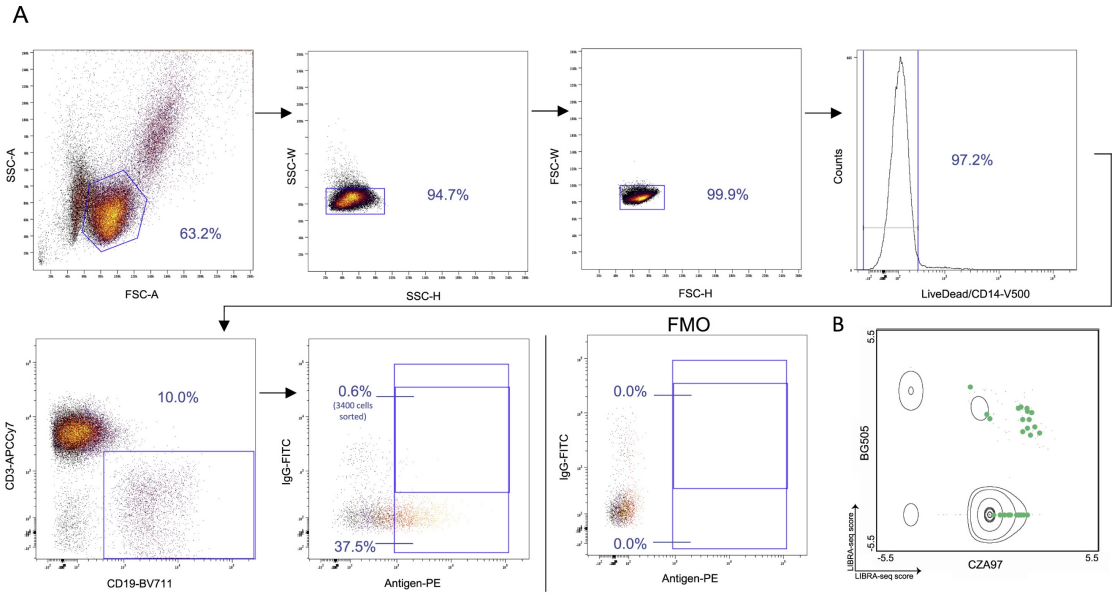


Figure 22. Identification of antigen-specific B cells from Donor NIAID45 PBMCs.
(A) Gating scheme for fluorescence activated cell sorting of donor NIAID45 PBMCs. Cells were stained with LiveDead-V500, CD14-V500, CD3-APCCy7, CD19-BV711, IgG-FITC, and a DNA-barcoded antigen screening library consisting 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.

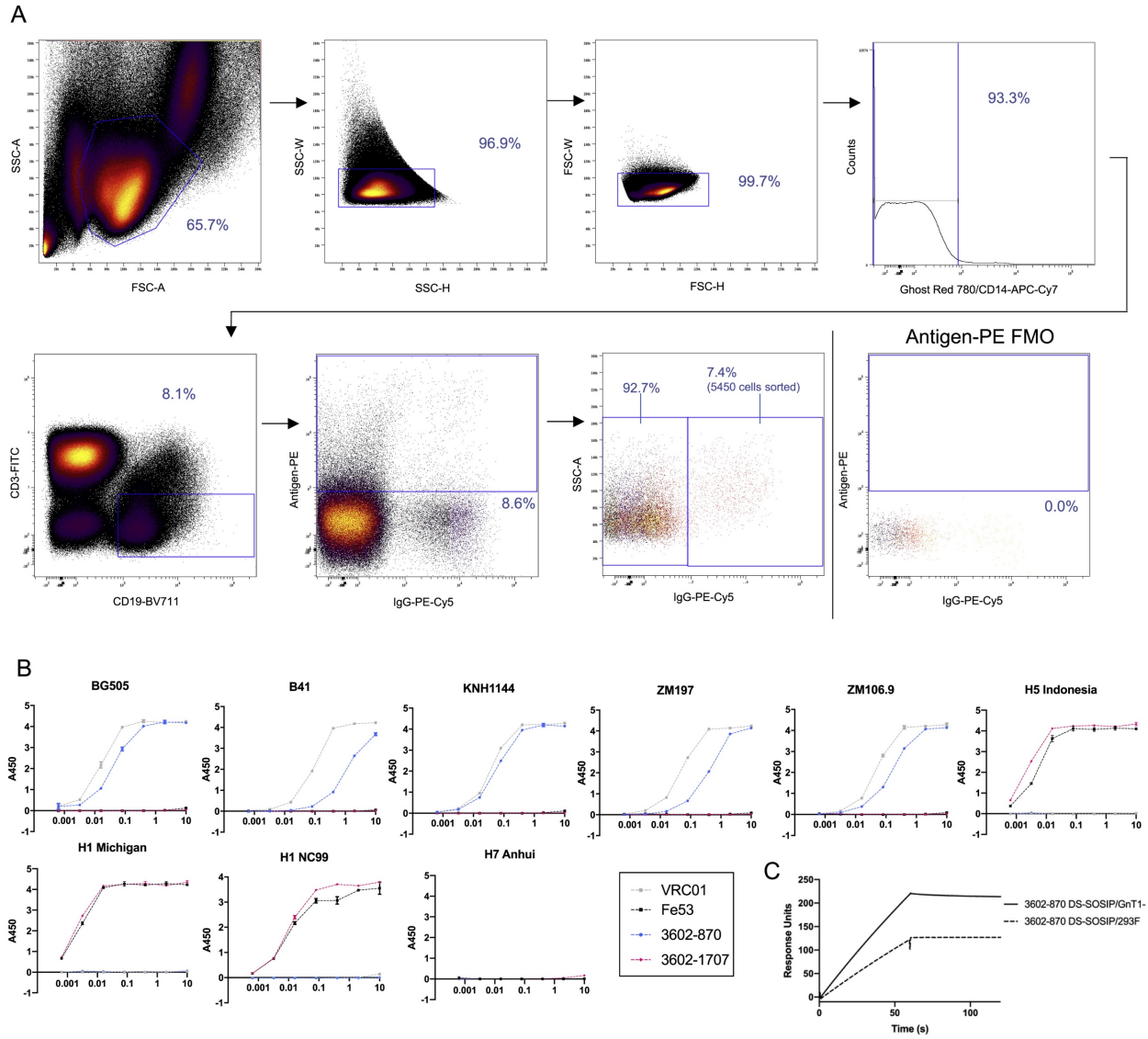


Figure 23. LIBRA-seq applied to a sample from Donor N90.

(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 ± 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.

A

Ramos B Cell Lines

Summary	
Sequencing	paired end (150 cycles + 150 cycles)
Insert size peak	52 bp
Before filtering	
Total reads	46.683278 M
Total bases	7.002492 G
Q20 bases	5.890977 G (84.126864%)
Q30 bases	5.343662 G (76.310871%)
After filtering	
Total reads	46.317712 M
Total bases	2.473195 G
Q20 bases	2.438363 G (98.591592%)
Q30 bases	2.358131 G (95.347568%)
Filtering results	
Reads passed filters	46.317712 M
Reads with low quality	253.286000 K (0.542563%)
Reads with too many N	94 (0.000201%)
Reads too short	112.186000 K (0.240313%)

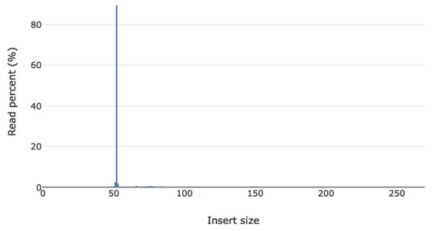
Donor NIH45

Summary	
Sequencing	paired end (150 cycles + 150 cycles)
Insert size peak	52 bp
Before filtering	
Total reads	39.610554 M
Total bases	5.940879 G
Q20 bases	4.959718 G (83.484582%)
Q30 bases	4.463982 G (75.140098%)
After filtering	
Total reads	36.922686 M
Total bases	2.061362 G
Q20 bases	1.992712 G (96.669638%)
Q30 bases	1.910089 G (92.661463%)
Filtering results	
Reads passed filters	36.922686 M
Reads with low quality	1.506144 K (3.802381%)
Reads with too many N	158 (0.000399%)
Reads too short	1.181566 M (2.982958%)

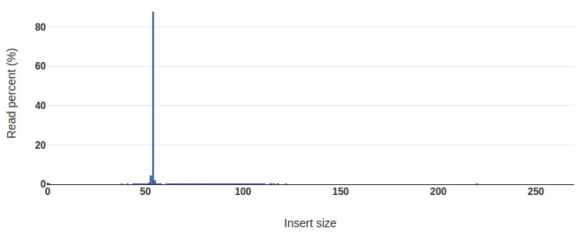
Donor N90

Summary	
Sequencing	paired end (150 cycles + 150 cycles)
Insert size peak	54 bp
Before filtering	
Total reads	82.957516 M
Total bases	8.65364 G
Q20 bases	7.647815 G (88.377127%)
Q30 bases	6.985525 G (80.723789%)
After filtering	
Total reads	80.923090 M
Total bases	6.994344 G
Q20 bases	6.314613 G (90.281709%)
Q30 bases	5.822037 G (83.239223%)
Filtering results	
Reads passed filters	80.923090 M
Reads with low quality	1.494196 M (1.801158%)
Reads with too many N	4.434000 K (0.005333%)
Reads too short	535.806000 K (0.645880%)

B



D



C

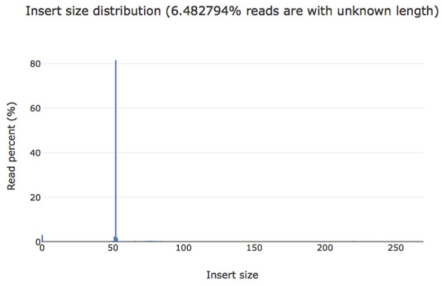


Figure 24. Sequencing preprocessing and quality statistics.

(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.

Contributions to Chapter 3

Conception of technology:	Ian Setliff
Conceptualization:	Ian Setliff, Andrea Shiakolas, Ivelin Georgiev
Methodology:	Ian Setliff, Andrea Shiakolas, Ivelin Georgiev
Investigation:	Ian Setliff, Andrea Shiakolas, Kelsey Pilewski, Amyn Murji, Rutendo Mapengo, Simone Richardson, Charissa Oosthuysen, Nagarajan Raju, Allison Greenplate, Katarzyna Janowska, Kevin Kramer, Juliana Qin, Daniel Lingwood, Larance Ronsard, Wyatt McDonnell, Ivelin Georgiev
Software:	Ian Setliff, Nagarajan Raju
Validation:	Ian Setliff, Andrea Shiakolas, Nagarajan Raju
Writing – Original Draft:	Ian Setliff, Andrea Shiakolas
Writing – Review & Editing:	Ian Setliff, Andrea Shiakolas, Kelsey Pilewski, Amyn Murji, Rutendo Mapengo, Simone Richardson, Charissa Oosthuysen, Nagarajan Raju, Allison Greenplate, Katarzyna Janowska, Kevin Kramer, Juliana Qin, Daniel Lingwood, Larance Ronsard, Wyatt McDonnell, Ivelin Georgiev, Priyamvada Acharya, Masaru Kanekiyo, Barney Graham, Marc Connors, Lynn Morris
Funding Acquisition:	Ivelin Georgiev, Daniel Lingwood, Lynn Morris, Priyamvada Acharya, Marc Connors, Barney Graham, Ian Setliff
Resources:	Masaru Kanekiyo, Priyamvada Acharya, Barney Graham, Marc Connors, Daniel Lingwood, Lynn Morris, Ivelin Georgiev
Supervision:	Ivelin Georgiev, Lynn Morris, Priyamvada Acharya

APPENDIX 3

Contributions to Chapter 4

Conception of technology:	Ian Setliff
Conceptualization:	Ian Setliff, Andrea Shiakolas, Amyn Murji, Ivelin Georgiev
Methodology:	Ian Setliff, Andrea Shiakolas, Amyn Murji, Ivelin Georgiev
Investigation:	Ian Setliff, Andrea Shiakolas, Amyn Murji, Juliana Qin, Daniel Lingwood, Larance Ronsard, Ivelin Georgiev
Software:	Ian Setliff
Validation:	Ian Setliff, Andrea Shiakolas, Amyn Murji
Writing – Original Draft:	Ian Setliff
Writing – Review & Editing:	Ian Setliff, Ivelin Georgiev
Funding Acquisition:	Ivelin Georgiev, Daniel Lingwood, Ian Setliff
Resources:	Masaru Kanekiyo, Barney Graham, Daniel Lingwood, Ivelin Georgiev
Supervision:	Ivelin Georgiev

REFERENCES

Acharya, P., Tolbert, W.D., Gohain, N., Wu, X., Yu, L., Liu, T., Huang, W., Huang, C. -c., Kwon, Y.D., Louder, R.K., et al. (2014). Structural Definition of an Antibody-Dependent Cellular Cytotoxicity Response Implicated in Reduced Risk for HIV-1 Infection. J. Virol. *88*, 12895–12906.

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.

Ackerman, M.E., Mikhailova, A., Brown, E.P., Dowell, K.G., Walker, B.D., Bailey-Kellogg, C., Suscovich, T.J., and Alter, G. (2016). Polyfunctional HIV-Specific Antibody Responses Are Associated with Spontaneous HIV Control. PLoS Pathog. *12*.

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. (2017a). Rare, high-affinity anti-pathogen antibodies from human repertoires, discovered using microfluidics and molecular genomics. MAbs *9*, 1282–1296.

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. (2017b). Rare, high-affinity mouse anti-PD-1 antibodies that function in checkpoint blockade, discovered using microfluidics and molecular genomics. MAbs *9*, 1270–1281.

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*.

Andrews, S.F., Gordon Joyce, M., Chambers, M.J., Gillespie, R.A., Kanekiyo, M., Leung, K., Yang, E.S., Tsybovsky, Y., Wheatley, A.K., Crank, M.C., et al. (2017). Preferential induction of cross-group influenza A hemagglutinin stem–specific memory B cells after H7N9 immunization in humans. Sci. Immunol. *2*.

Arentz, G., Thurgood, L.A., Lindop, R., Chataway, T.K., and Gordon, T.P. (2012). Secreted human Ro52 autoantibody proteomes express a restricted set of public clonotypes. J. Autoimmun. *39*, 466–470.

Bangaru, S., Lang, S., Schotsaert, M., Vanderven, H.A., Zhu, X., Kose, N., Bombardi, R., Finn, J.A., Kent, S.J., Gilchuk, P., et al. (2019). A Site of Vulnerability on the Influenza Virus Hemagglutinin Head Domain Trimer Interface. Cell *177*, 1136-1152.e18.

Bar-On, Y., Gruell, H., Schoofs, T., Pai, J.A., Nogueira, L., Butler, A.L., Millard, K., Lehmann, C., Suárez, I., Oliveira, T.Y., et al. (2018). Safety and antiviral activity of combination HIV-1 broadly neutralizing antibodies in viremic individuals. Nat. Med. *24*,

1701–1707.

Bianchi, M., Turner, H.L., Nogal, B., Cottrell, C.A., Oyen, D., Pauthner, M., Bastidas, R., Nedellec, R., McCoy, L.E., Wilson, I.A., et al. (2018). Electron-Microscopy-Based Epitope Mapping Defines Specificities of Polyclonal Antibodies Elicited during HIV-1 BG505 Envelope Trimer Immunization. Immunity *49*, 288-300.e8.

Blanchard-Rohner, G., Galli, G., Clutterbuck, E.A., and Pollard, A.J. (2010). Comparison of a limiting dilution assay and ELISpot for detection of memory B-cells before and after immunisation with a protein-polysaccharide conjugate vaccine in children. J. Immunol. Methods *358*, 46–55.

Bolotin, D.A., Poslavsky, S., Mitrophanov, I., Shugay, M., Mamedov, I.Z., Putintseva, E. V, and Chudakov, D.M. (2015). MiXCR: software for comprehensive adaptive immunity profiling. Nat Meth *12*, 380–381.

Bonsignori, M., Hwang, K.-K., Chen, X., Tsao, C.-Y., Morris, L., Gray, E., Marshall, D.J., Crump, J.A., Kapiga, S.H., Sam, N.E., et al. (2011). Analysis of a Clonal Lineage of HIV-1 Envelope V2/V3 Conformational Epitope-Specific Broadly Neutralizing Antibodies and Their Inferred Unmutated Common Ancestors. J. Virol. *85*, 9998–10009.

Bonsignori, M., Montefiori, D.C., Wu, X., Chen, X., Hwang, K.-K., Tsao, C.-Y., Kozink, D.M., Parks, R.J., Tomaras, G.D., Crump, J.A., et al. (2012). Two Distinct Broadly Neutralizing Antibody Specificities of Different Clonal Lineages in a Single HIV-1-Infected Donor: Implications for Vaccine Design. J. Virol. *86*, 4688–4692.

Bonsignori, M., Zhou, T., Sheng, Z., Chen, L., Gao, F., Joyce, M.G., Ozorowski, G., Chuang, G.-Y., Schramm, C.A., Wiehe, K., et al. (2016). Maturation Pathway from Germline to Broad HIV-1 Neutralizer of a CD4-Mimic Antibody. Cell *165*, 449–463.

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.

Boonyaratanakornkit, J., and Taylor, J.J. (2019). Techniques to Study Antigen-Specific B Cell Responses. Front. Immunol. *10*, 1694.

Boyd, S.D., and Joshi, S.A. (2014). High-Throughput DNA Sequencing Analysis of Antibody Repertoires. In Microbiology Spectrum, pp. 345–362.

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.

Brochet, X., Lefranc, M.P., and Giudicelli, V. (2008). IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. Nucleic Acids Res. *36*, W503-8.

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.

Burton, D.R. (2017). What are the most powerful immunogen design vaccine strategies?: Reverse vaccinology 2.0 shows great promise. Cold Spring Harb. Perspect. Biol. *9*.

Burton, D.R., and Mascola, J.R. (2015). Antibody responses to envelope glycoproteins in HIV-1 infection. Nat. Immunol. *16*, 571–576.

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.

Cheadle, C., Vawter, M.P., Freed, W.J., and Becker, K.G. (2003). Analysis of microarray data using Z score transformation. J. Mol. Diagn. *5*, 73–81.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, i884–i890.

Chen, W., Zhu, Z., Feng, Y., and Dimitrov, D.S. (2010). A large human domain antibody library combining heavy and light chain CDR3 diversity. Mol. Immunol. *47*, 912–921.

Cheng, C., Xu, K., Kong, R., Chuang, G.Y., Corrigan, A.R., Geng, H., Hill, K.R., Jafari, A.J., O'Dell, S., Ou, L., et al. (2019). Consistent elicitation of cross-clade HIV-neutralizing responses achieved in Guinea pigs after fusion peptide priming by repetitive envelope trimer boosting. PLoS One *14*.

Cheung, W.C., Beausoleil, S.A., Zhang, X., Sato, S., Schieferl, S.M., Wieler, J.S., Beaudet, J.G., Ramenani, R.K., Popova, L., Comb, M.J., et al. (2012). A proteomics approach for the identification and cloning of monoclonal antibodies from serum. Nat. Biotechnol. *30*, 447–452.

Chung, A.W., Kumar, M.P., Arnold, K.B., Yu, W.H., Schoen, M.K., Dunphy, L.J., Suscovich, T.J., Frahm, N., Linde, C., Mahan, A.E., et al. (2015). Dissecting Polyclonal Vaccine-Induced Humoral Immunity against HIV Using Systems Serology. Cell *163*, 988–998.

Combadière, B., Beaujean, M., Chaudesaigues, C., and Vieillard, V. (2019). Peptide-Based Vaccination for Antibody Responses Against HIV. Vaccines (Basel) *7*.

Connor, E.M. (1998). Palivizumab, a humanized respiratory syncytial virus monoclonal antibody, reduces hospitalization from respiratory syncytial virus infection in high-risk infants. Pediatrics *102*, 531–537.

Corey, L., Gilbert, P.B., Tomaras, G.D., Haynes, B.F., Pantaleo, G., and Fauci, A.S. (2015). Immune correlates of vaccine protection against HIV-1 acquisition. Sci. Transl. Med. *7*, 310rv7.

Corti, D., Langedijk, J.P.M., Hinz, A., Seaman, M.S., Vanzetta, F., Fernandez-Rodriguez, B.M., Silacci, C., Pinna, D., Jarrossay, D., Balla-Jhagjhoorsingh, S., et al. (2010). Analysis of memory B cell responses and isolation of novel monoclonal antibodies with neutralizing breadth from HIV-1-infected individuals. PLoS One *5*.

Corti, D., Misasi, J., Mulangu, S., Stanley, D.A., Kanekiyo, M., Wollen, S., Ploquin, A., Doria-Rose, N.A., Staupe, R.P., Bailey, M., et al. (2016). Protective monotherapy against lethal Ebola virus infection by a potently neutralizing antibody. Science. *351*, 1339–1342.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: A sequence logo generator. Genome Res. *14*, 1188–1190.

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 LP – 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.

DeKosky, B. (2015). Decoding the Antibody Repertoire: High Throughput Sequencing of Multiple Transcripts from Single B Cells. Univ. Texas Austin PhD Diss.

Doria-Rose, N.A., Schramm, C.A., Gorman, J., Moore, P.L., Bhiman, J.N., DeKosky, B.J., Ernandes, M.J., Georgiev, I.S., Kim, H.J., Pancera, M., et al. (2014).

Developmental pathway for potent V1V2-directed HIV-neutralizing antibodies. Nature *509*, 55–62.

Doria-Rose, N.A., Bhiman, J.N., Roark, R.S., Schramm, C.A., Gorman, J., Chuang, G.-Y., Pancera, M., Cale, E.M., Ernandes, M.J., Louder, M.K., et al. (2016). New Member of the V1V2-Directed CAP256-VRC26 Lineage That Shows Increased Breadth and Exceptional Potency. J. Virol. *90*, 76–91.

Doria-Rose, N.A., Altae-Tran, H.R., Roark, R.S., Schmidt, S.D., Sutton, M.S., Louder, M.K., Chuang, G.Y., Bailer, R.T., Cortez, V., Kong, R., et al. (2017). Mapping Polyclonal HIV-1 Antibody Responses via Next-Generation Neutralization Fingerprinting. PLoS Pathog. *13*, e1006148.

Dosenovic, P., Von Boehmer, L., Escolano, A., Jardine, J., Freund, N.T., Gitlin, A.D., McGuire, A.T., Kulp, D.W., Oliveira, T., Scharf, L., et al. (2015). Immunization for HIV-1 Broadly Neutralizing Antibodies in Human Ig Knockin Mice. Cell *161*, 1505–1515.

Dosenovic, P., Pettersson, A.-K., Wall, A., Thientosapol, E.S., Feng, J., Weidle, C., Bhullar, K., Kara, E.E., Hartweger, H., Pai, J.A., et al. (2019). Anti-idiotypic antibodies elicit anti-HIV-1–specific B cell responses. J. Exp. Med. jem.20190446.

Douglas, B., Maechler, M., Ben, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. J. Stat. Softw. *67*, 1–48.

Dunbar, J., and Deane, C.M. (2016). ANARCI: Antigen receptor numbering and receptor classification. Bioinformatics *32*, 298–300.

Ecker, D.M., Jones, S.D., and Levine, H.L. (2015). The therapeutic monoclonal antibody market. MAbs *7*, 9–14.

Falkowska, E., Le, K.M., Ramos, A., Doores, K.J., Lee, J., Blattner, C., Ramirez, A., Derking, R., vanGils, M.J., Liang, C.H., et al. (2014). Broadly neutralizing HIV antibodies define a glycan-dependent epitope on the prefusion conformation of gp41 on cleaved envelope trimers. Immunity *40*, 657–668.

Feltes, T.F., Cabalka, A.K., Meissner, H.C., Piazza, F.M., Carlin, D.A., Top, F.H., Connor, E.M., and Sondheimer, H.M. (2003). Palivizumab prophylaxis reduces hospitalization due to respiratory syncytial virus in young children with hemodynamically significant congenital heart disease. J. Pediatr. *143*, 532–540.

Flyak, A.I., Shen, X., Murin, C.D., Turner, H.L., David, J.A., Fusco, M.L., Lampley, R., Kose, N., Ilinykh, P.A., Kuzmina, N., et al. (2016). Cross-Reactive and Potent

Neutralizing Antibody Responses in Human Survivors of Natural Ebolavirus Infection. Cell *164*, 392–405.

Furst, D.E., Kavanaugh, A., Florentinus, S., Kupper, H., Karunaratne, M., and Birbara, C.A. (2015). Final 10-year effectiveness and safety results from study DE020: Adalimumab treatment in patients with rheumatoid arthritis and an inadequate response to standard therapy. Rheumatol. (United Kingdom) *54*, 2188–2197.

Gaebler, C., Gruell, H., Velinzon, K., Scheid, J.F., Nussenzweig, M.C., and Klein, F. (2013). Isolation of HIV-1-reactive antibodies using cell surface-expressed gp160δcBaL. J. Immunol. Methods *397*, 47–54.

Gainza, P., Roberts, K.E., Georgiev, I., Lilien, R.H., Keedy, D.A., Chen, C.Y., Reza, F., Anderson, A.C., Richardson, D.C., Richardson, J.S., et al. (2013). Osprey: Protein design with ensembles, flexibility, and provable algorithms. In Methods in Enzymology, pp. 87–107.

Gaudinski, M.R., Coates, E.E., Novik, L., Widge, A., Houser, K. V., Burch, E., Holman, L.S.A., Gordon, I.J., Chen, G.L., Carter, C., et al. (2019). Safety, tolerability, pharmacokinetics, and immunogenicity of the therapeutic monoclonal antibody mAb114 targeting Ebola virus glycoprotein (VRC 608): an open-label phase 1 study. Lancet *393*, 889–898.

Gautam, R., Nishimura, Y., Pegu, A., Nason, M.C., Klein, F., Gazumyan, A., Golijanin, J., Buckler-White, A., Sadjadpour, R., Wang, K., et al. (2016). A single injection of anti-HIV-1 antibodies protects against repeated SHIV challenges. Nature *533*, 105–109.

Georgiev, I., and Donald, B.R. (2007). Dead-End Elimination with backbone flexibility. Bioinformatics *23*.

Georgiev, I., Lilien, R.H., and Donald, B.R. (2006). Improved pruning algorithms and divide-and-conquer strategies for Dead-End Elimination, with application to protein design. Bioinformatics *22*.

Georgiev, I., Lilien, R.H., and Donald, B.R. (2008a). The minimized dead-end elimination criterion and its application to protein redesign in a hybrid scoring and search algorithm for computing partition functions over molecular ensembles. J. Comput. Chem. *29*, 1527–1542.

Georgiev, I., Keedy, D., Richardson, J.S., Richardson, D.C., and Donald, B.R. (2008b). Algorithm for backrub motions in protein design. Bioinformatics *24*.

Georgiev, I.S., Doria-Rose, N.A., Zhou, T., Do Kwon, Y., Staupe, R.P., Moquin, S., Chuang, G.-Y., Louder, M.K., Schmidt, S.D., Altae-Tran, H.R., et al. (2013). Delineating Antibody Recognition in Polyclonal Sera from Patterns of HIV-1 Isolate Neutralization. Science. *340*, 751–756.

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.

Gilchuk, P., Kuzmina, N., Ilinykh, P.A., Huang, K., Gunn, B.M., Bryan, A., Davidson, E., Doranz, B.J., Turner, H.L., Fusco, M.L., et al. (2018). Multifunctional Pan-ebolavirus Antibody Recognizes a Site of Broad Vulnerability on the Ebolavirus Glycoprotein. Immunity *49*, 363-374.e10.

Gilman, M.S.A., Castellanos, C.A., Chen, M., Ngwuta, J.O., Goodwin, E., Moin, S.M., Mas, V., Melero, J.A., Wright, P.F., Graham, B.S., et al. (2016). Rapid profiling of RSV antibody repertoires from the memory B cells of naturally infected adult donors. Sci. Immunol. *1*.

van Gils, M.J., Moore, J.P., de Val, N., Derking, R., Cupo, A., Blattner, C., Julien, J.-P., Klasse, P.J., Kim, H.J., Golabek, M., 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.

Glanville, J., Zhai, W., Berka, J., Telman, D., Huerta, G., Mehta, G.R., Ni, I., Mei, L., Sundar, P.D., Day, G.M.R., et al. (2009). Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. Proc. Natl. Acad. Sci. U. S. A. *106*, 20216–20221.

Goldstein, L.D., Chen, Y.J.J., Wu, J., Chaudhuri, S., Hsiao, Y.C., Schneider, K., Hoi, K.H., Lin, Z., Guerrero, S., Jaiswal, B.S., et al. (2019). Massively parallel single-cell B-cell receptor sequencing enables rapid discovery of diverse antigen-reactive antibodies. Commun. Biol. *2*.

Green, L.L. (1999). Antibody engineering via genetic engineering of the mouse: XenoMouse strains are a vehicle for the facile generation of therapeutic human monoclonal antibodies. J. Immunol. Methods *231*, 11–23.

Greiff, V., Bhat, P., Cook, S.C., Menzel, U., Kang, W., and Reddy, S.T. (2015). A bioinformatic framework for immune repertoire diversity profiling enables detection of immunological status. Genome Med. *7*.

Groothuis, J.R., Simoes, E.A.F., and Hemming, V.G. (1995). Respiratory syncytial virus (RSV) infection in preterm infants and the protective effects of RSV immune globulin (RSVIG). Pediatrics *95*, 463–467.

Gross, A., Schoendube, J., Zimmermann, S., Steeb, M., Zengerle, R., and Koltay, P. (2015). Technologies for single-cell isolation. Int. J. Mol. Sci. *16*, 16897–16919.

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.

Hamilton, J.A., Li, J., Wu, Q., Yang, P., Luo, B., Li, H., Bradley, J.E., Taylor, J.J., Randall, T.D., Mountz, J.D., et al. (2015). General Approach for Tetramer-Based Identification of Autoantigen-Reactive B Cells: Characterization of La- and snRNP-Reactive B Cells in Autoimmune BXD2 Mice. J. Immunol. *194*, 5022–5034.

Harrer, T., Harrer, E., Kalams, S., Elbeik, T., Staprans, S.I., Feinberg, M.B., Cao, Y., Ho, D.D., Yilma, T., Caliendo, A.M., et al. (1996). Strong Cytotoxic T Cell and Weak Neutralizing Antibody Responses in a Subset of Persons with Stable Nonprogressing HIV Type 1 Infection. AIDS Res. Hum. Retroviruses *12*, 585–592.

Harris, A., Borgnia, M.J., Shi, D., Bartesaghi, A., He, H., Pejchal, R., Kang, Y., 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. *108*, 11440–11445.

Havenar-Daughton, C., Abbott, R.K., Schief, W.R., and Crotty, S. (2018). When designing vaccines, consider the starting material: the human B cell repertoire. Curr. Opin. Immunol. *53*, 209–216.

Vander Heiden, J.A., Yaari, G., Uduman, M., Stern, J.N.H., O'connor, K.C., Hafler, D.A., Vigneault, F., and Kleinstein, S.H. (2014). PRESTO: A toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. Bioinformatics *30*, 1930–1932.

Henry Dunand, C.J., and Wilson, P.C. (2015). Restricted, canonical, stereotyped and convergent immunoglobulin responses. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *370*, 20140238.

Hessell, A.J., Rakasz, E.G., Poignard, P., Hangartner, L., Landucci, G., Forthal, D.N., Koff, W.C., Watkins, D.I., and Burton, D.R. (2009). Broadly neutralizing human anti-HIV antibody 2G12 is effective in protection against mucosal SHIV challenge even at low serum neutralizing titers. PLoS Pathog. *5*.

Hessell, A.J., Rakasz, E.G., Tehrani, D.M., Huber, M., Weisgrau, K.L., Landucci, G.,

Forthal, D.N., Koff, W.C., Poignard, P., Watkins, D.I., et al. (2010). Broadly Neutralizing Monoclonal Antibodies 2F5 and 4E10 Directed against the Human Immunodeficiency Virus Type 1 gp41 Membrane-Proximal External Region Protect against Mucosal Challenge by Simian-Human Immunodeficiency Virus SHIVBa-L. J. Virol. *84*, 1302–1313.

Horwitz, J.A., Bar-On, Y., Lu, C.L., Fera, D., Lockhart, A.A.K., Lorenzi, J.C.C., Nogueira, L., Golijanin, J., Scheid, J.F., Seaman, M.S., et al. (2017). Non-neutralizing Antibodies Alter the Course of HIV-1 Infection In Vivo. Cell *170*, 637-648.e10.

Huang, J., Ofek, G., Laub, L., Louder, M.K., Doria-Rose, N.A., Longo, N.S., Imamichi, H., Bailer, R.T., Chakrabarti, B., Sharma, S.K., et al. (2012). Broad and potent neutralization of HIV-1 by a gp41-specific human antibody. Nature *491*, 406–412.

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.

Huang, J., Kang, B.H., Ishida, E., Zhou, T., Griesman, T., Sheng, Z., Wu, F., Doria-Rose, N.A., Zhang, B., McKee, K., et al. (2016). Identification of a CD4-Binding-Site Antibody to HIV that Evolved Near-Pan Neutralization Breadth. Immunity *45*, 1108–1121.

Hunter, J.D. (2007). Matplotlib: A 2D graphics environment. Comput. Sci. Eng. *9*, 99–104.

Jackson, K.J.L., Liu, Y., Roskin, K.M., Glanville, J., Hoh, R.A., Seo, K., Marshall, E.L., Gurley, T.C., Moody, M.A., Haynes, B.F., et al. (2014). Human responses to influenza vaccination show seroconversion signatures and convergent antibody rearrangements. Cell Host Microbe *16*, 105–114.

Janeway, C.A., Travers, P., Walport, M., and Shlomchik, M.J. (2001). Immunobiology: The immune system in health and disease. J. Allergy Clin. Immunol.

Jardine, J., Julien, J.-P., Menis, S., Ota, T., Kalyuzhniy, 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., Kalyuzhniy, 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., Wheatley, A.K., Thomas, P. V., Chuang, G.Y., Soto, C., Bailer, R.T., Druz, A., Georgiev, I.S., Gillespie, R.A., Kanekiyo, M., et al. (2016). Vaccine-Induced

Antibodies that Neutralize Group 1 and Group 2 Influenza A Viruses. Cell *166*, 609–623.

Joyce, M.G., Georgiev, I.S., Yang, Y., Druz, A., Geng, H., Chuang, G.Y., Kwon, Y. Do, 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., Nieusma, 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. *112*, 11947–11952.

Kaplon, H., and Reichert, J.M. (2018). Antibodies to watch in 2018. MAbs *10*, 183–203.

Kaplon, H., and Reichert, J.M. (2019). Antibodies to watch in 2019. MAbs *11*, 219–238.

Kaufmann, D.E., Kavanagh, D.G., Pereyra, F., Zaunders, J.J., Mackey, E.W., Miura, T., Palmer, S., Brockman, M., Rathod, A., Piechocka-Trocha, A., et al. (2007). Upregulation of CTLA-4 by HIV-specific CD4+ T cells correlates with disease progression and defines a reversible immune dysfunction. Nat. Immunol. *8*, 1246.

Keller, M.A., and Stiehm, E.R. (2000). Passive immunity in prevention and treatment of infectious diseases. Clin. Microbiol. Rev. *13*, 602–614.

Kelley, B. (2009). Industrialization of mAb production technology: The bioprocessing industry at a crossroads. MAbs *1*, 440–449.

Kisalu, N.K., Idris, A.H., Weidle, C., Flores-Garcia, Y., Flynn, B.J., Sack, B.K., Murphy, S., Schön, A., Freire, E., Francica, J.R., et al. (2018). A human monoclonal antibody prevents malaria infection by targeting a new site of vulnerability on the parasite. Nat. Med. *24*, 408–416.

Köhler, G., and Milstein, C. (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. Nature *256*, 495–497.

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.

Krishnamurty, A.T., Thouvenel, C.D., Portugal, S., Keitany, G.J., Kim, K.S., Holder, A., Crompton, P.D., Rawlings, D.J., and Pepper, M. (2016). Somatically Hypermutated Plasmodium-Specific IgM+ Memory B Cells Are Rapid, Plastic, Early Responders upon Malaria Rechallenge. Immunity *45*, 402–414.

Kuznetsova, A., Brockhoff, P.B., and Christensen, R.H.B. (2014). Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). R package version 2.0-3. R Packag.

Do Kwon, Y., 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. &Amp; Mol. Biol. *22*, 522.

Lasserson, U. (2012). High-Throughput Methods for Characterizing the Immune Repertoire. Harvard-MIT Div. Heal. Sci. Technol. PhD Diss.

Lefranc, M.P., Giudicelli, V., Duroux, P., Jabado-Michaloud, J., Folch, G., Aouinti, S., Carillon, E., Duvergey, H., Houles, A., Paysan-Lafosse, T., et al. (2015). IMGT, the international ImMunoGeneTics information system 25 years on. Nucleic Acids Res. *43*, D413–D422.

Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. *47*, W256–W259.

Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., and Pfister, H. (2014). UpSet: Visualization of intersecting sets. IEEE Trans. Vis. Comput. Graph. *20*, 1983–1992.

Liao, H.-X., Lynch, R., Zhou, T., Gao, F., Alam, S.M., Boyd, S.D., Fire, A.Z., Roskin, K.M., Schramm, C.A., Zhang, Z., et al. (2013). Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. Nature *496*, 469–476.

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.

Mascola, J.R., Lewis, M.G., Stiegler, G., Harris, D., VanCott, T.C., Hayes, D., Louder, M.K., Brown, C.R., Sapan, C. V, Frankel, S.S., et al. (1999). Protection of Macaques against pathogenic simian/human immunodeficiency virus 89.6PD by passive transfer of neutralizing antibodies. J. Virol. *73*, 4009–4018.

Mascola, J.R., Stiegler, G., Vancott, T.C., Katinger, H., Carpenter, C.B., Hanson, C.E., Beary, H., Hayes, D., Frankel, S.S., Birx, D.L., et al. (2000). Protection of macaques against vaginal transmission of a pathogenic HIV- 1/SIV chimeric virus by passive infusion of neutralizing antibodies. Nat. Med. *6*, 207–210.

McCafferty, J., Griffiths, A.D., Winter, G., and Chiswell, D.J. (1990). Phage antibodies: filamentous phage displaying antibody variable domains. Nature *348*, 552–554.

Mendoza, P., Gruell, H., Nogueira, L., Pai, J.A., Butler, A.L., Millard, K., Lehmann, C., Suárez, I., Oliveira, T.Y., Lorenzi, J.C.C., et al. (2018). Combination therapy with anti-

HIV-1 antibodies maintains viral suppression. Nature *561*, 479–484.

De Milito, A., Nilsson, A., Titanji, K., Thorstensson, R., Reizenstein, E., Narita, M., Grutzmeier, S., Sönnerborg, A., and Chiodi, F. (2004). Mechanisms of hypergammaglobulinemia and impaired antigen-specific humoral immunity in HIV-1 infection. Blood *103*, 2180 LP – 2186.

Mimitou, E.P., Cheng, A., Montalbano, A., Hao, S., Stoeckius, M., Legut, M., Roush, T., Herrera, A., Papalexi, E., Ouyang, Z., et al. (2019). Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. Nat. Methods *16*, 409–412.

Moldt, B., Rakasz, E.G., Schultz, N., Chan-Hui, P.-Y., Swiderek, K., Weisgrau, K.L., Piaskowski, S.M., Bergman, Z., Watkins, D.I., Poignard, P., et al. (2012). Highly potent HIV-specific antibody neutralization in vitro translates into effective protection against mucosal SHIV challenge in vivo. Proc. Natl. Acad. Sci. U. S. A. *109*, 18921–18925.

Moody, M.A., and Haynes, B.F. (2008). Antigen-specific B cell detection reagents: Use and quality control. Cytom. Part A *73*, 1086–1092.

Moody, M.A., Yates, N.L., Amos, J.D., Drinker, M.S., Eudailey, J.A., Gurley, T.C., Marshall, D.J., Whitesides, J.F., Chen, X., Foulger, A., et al. (2012). HIV-1 gp120 Vaccine Induces Affinity Maturation in both New and Persistent Antibody Clonal Lineages. J. Virol. *86*, 7496–7507.

Moore, P.L., Gray, E.S., Sheward, D., Madiga, M., Ranchobe, N., Lai, Z., Honnen, W.J., Nonyane, M., Tumba, N., Hermanus, T., et al. (2011). Potent and Broad Neutralization of HIV-1 Subtype C by Plasma Antibodies Targeting a Quaternary Epitope Including Residues in the V2 Loop. J. Virol.

Mouquet, H., Scharf, L., Euler, Z., Liu, Y., Eden, C., Scheid, J.F., Halper-Stromberg, A., Gnanapragasam, P.N.P., Spencer, D.I.R., Seaman, M.S., et al. (2012). Complex-type N-glycan recognition by potent broadly neutralizing HIV antibodies. Proc. Natl. Acad. Sci. U. S. A. *109*, E3268-77.

Nicely, N.I., Wiehe, K., Kepler, T.B., Jaeger, F.H., Dennison, S.M., Rerks-Ngarm, S., Nitayaphan, S., Pitisuttithum, P., Kaewkungwal, J., Robb, M.L., et al. (2015). Structural analysis of the unmutated ancestor of the HIV-1 envelope V2 region antibody CH58 isolated from an RV144 vaccine efficacy trial vaccinee. EBioMedicine *2*, 713–722.

Olsson, L., and Kaplan, H.S. (1980). Human-human hybridomas producing monoclonal antibodies of predefined antigenic specificity. Proc. Natl. Acad. Sci. U. S. A. *77*, 5429–5431.

Palmer, B.E., Boritz, E., and Wilson, C.C. (2004). Effects of Sustained HIV-1 Plasma Viremia on HIV-1 Gag-Specific CD4 T Cell Maturation and Function. J. Immunol. *172*,

3337 LP – 3347.

Pancera, M., Zhou, T., Druz, A., Georgiev, I.S., Soto, C., Gorman, J., Huang, J., Acharya, P., Chuang, G.-Y., Ofek, G., et al. (2014). Structure and immune recognition of trimeric pre-fusion HIV-1 Env. Nature *514*, 455–461.

Parameswaran, P., Liu, Y., Roskin, K.M., Jackson, K.K.L., Dixit, V.P., Lee, J.-Y., Artiles, K., Zompi, S., Vargas, M.J., Simen, B.B., et al. (2013). Convergent antibody signatures in human dengue. Cell Host Microbe *13*, 691–700.

Parren, P.W.H.I., Marx, P.A., Hessell, A.J., Luckay, A., Harouse, J., Cheng-Mayer, C., Moore, J.P., and Burton, D.R. (2001). Antibody Protects Macaques against Vaginal Challenge with a Pathogenic R5 Simian/Human Immunodeficiency Virus at Serum Levels Giving Complete Neutralization In Vitro. J. Virol. *75*, 8340–8347.

Pedregosa, F., and Varoquaux, G. (2011). Scikit-learn: Machine learning in Python.
Pegu, A., Hessell, A.J., Mascola, J.R., and Haigwood, N.L. (2017). Use of broadly neutralizing antibodies for HIV-1 prevention. Immunol. Rev. *275*, 296–312.

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.

Pieper, K., Tan, J., Piccoli, L., Foglierini, M., Barbieri, S., Chen, Y., Silacci-Fregni, C., Wolf, T., Jarrossay, D., Anderle, M., et al. (2017). Public antibodies to malaria antigens generated by two LAIR1 insertion modalities. Nature *548*, 597–601.

Plotkin, S.A. (2010). Correlates of protection induced by vaccination. Clin. Vaccine Immunol. *17*, 1055–1065.

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. Cytom. Part A *79 A*, 603–612.

Pugach, P., Cupo, A., Ringe, R., Yasmeen, A., Kim, H.J., Korzun, J., Golabek, M., de los Reyes, K., Ketas, T.J., Sanders, R.W., et al. (2015). A native-like SOSIP.664 trimer based on an HIV-1 subtype B env gene. J. Virol. *89*, 3380–3395.

Rademeyer, C., Korber, B., Seaman, M.S., Giorgi, E.E., Thebus, R., Robles, A., Sheward, D.J., Wagh, K., Garrity, J., Carey, B.R., et al. (2016). Features of Recently Transmitted HIV-1 Clade C Viruses that Impact Antibody Recognition: Implications for Active and Passive Immunization. PLoS Pathog. *12*, e1005742.

Rajewsky, K. (1996). Clonal selection and learning in the antibody system. Nature *381*, 751–758.

Rappuoli, R., Bottomley, M.J., D'Oro, U., Finco, O., and De Gregorio, E. (2016). Reverse vaccinology 2.0: Human immunology instructs vaccine antigen design. J. Exp. Med. *213*, 469–481.

Reddy, S.T., Ge, X., Miklos, A.E., Hughes, R.A., Kang, S.H., Hoi, K.H., Chrysostomou, C., Hunicke-Smith, S.P., Iverson, B.L., Tucker, P.W., et al. (2010). Monoclonal antibodies isolated without screening by analyzing the variable-gene repertoire of plasma cells. Nat. Biotechnol. *28*, 965–969.

Reichert, J.M. (2012). Marketed therapeutic antibodies compendium. MAbs *4*, 413–415.

Reichert, J.M. (2016). Antibodies to watch in 2016. MAbs *8*, 197–204.

Reichert, J.M. (2017). Antibodies to watch in 2017. MAbs *9*, 167–181.

Richardson, S.I., Chung, A.W., Natarajan, H., Mabvakure, B., Mkhize, N.N., Garrett, N., Abdool Karim, S., Moore, P.L., Ackerman, M.E., Alter, G., et al. (2018a). HIV-specific Fc effector function early in infection predicts the development of broadly neutralizing antibodies. PLoS Pathog. *14*.

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.

Rinaldo, C., Huang, X.L., Fan, Z.F., Ding, M., Beltz, L., Logar, A., Panicali, D., Mazzara, G., Liebmann, J., and Cottrill, M. (1995). High levels of anti-human immunodeficiency virus type 1 (HIV-1) memory cytotoxic T-lymphocyte activity and low viral load are associated with lack of disease in HIV-1-infected long-term nonprogressors. J. Virol. *69*, 5838–5842.

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*, e00264-17.

Robbiani, D.F., Bozzacco, L., Keeffe, J.R., Khouri, R., Olsen, P.C., Gazumyan, A., Schaefer-Babajew, D., Avila-Rios, S., Nogueira, L., Patel, R., et al. (2017). Recurrent Potent Human Neutralizing Antibodies to Zika Virus in Brazil and Mexico. Cell *169*, 597-609.e11.

Rudicell, R.S., Kwon, Y. Do, Ko, S.-Y., Pegu, A., Louder, M.K., Georgiev, I.S., Wu, X., Zhu, J., Boyington, J.C., Chen, X., et al. (2014). Enhanced potency of a broadly neutralizing HIV-1 antibody in vitro improves protection against lentiviral infection in vivo. J. Virol. *88*, 12669–12682.

Russell, A.B., Elshina, E., Kowalsky, J.R., te Velthuis, A.J.W., and Bloom, J.D. (2019). Single-Cell Virus Sequencing of Influenza Infections That Trigger Innate Immunity. J. Virol. *93*.

Sabin, C., Corti, D., Buzon, V., Seaman, M.S., Hulsik, D.L., Hinz, A., Vanzetta, F., Agatic, G., Silacci, C., Mainetti, L., et al. (2010). Crystal structure and size-dependent neutralization properties of HK20, a human monoclonal antibody binding to the highly conserved heptad repeat 1 of gp41. PLoS Pathog. *6*, e1001195.

Sapparapu, G., Fernandez, E., Kose, N., Bin Cao, Fox, J.M., Bombardi, R.G., Zhao, H., Nelson, C.A., Bryan, A.L., Barnes, T., et al. (2016). Neutralizing human antibodies prevent Zika virus replication and fetal disease in mice. Nature *540*, 443–447.

Sarzotti-Kelsoe, M., Bailer, R.T., Turk, E., Lin, C., Bilska, 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.

Sato, S., Beausoleil, S.A., Popova, L., Beaudet, J.G., Ramenani, R.K., Zhang, X., Wieler, J.S., Schieferl, S.M., Cheung, W.C., and Polakiewicz, R.D. (2012). Proteomics-directed cloning of circulating antiviral human monoclonal antibodies. Nat. Biotechnol. *30*, 1039–1043.

Saunders, K.O., Pegu, A., Georgiev, I.S., Zeng, M., Joyce, M.G., Yang, Z.-Y., Ko, S.-Y., Chen, X., Schmidt, S.D., Haase, A.T., et al. (2015). Sustained Delivery of a Broadly Neutralizing Antibody in Nonhuman Primates Confers Long-Term Protection against Simian/Human Immunodeficiency Virus Infection. J. Virol. *89*, 5895–5903.

Scharf, L., West, A.P., Sievers, S.A., Chen, C., Jiang, S., Gao, H., Gray, M.D., McGuire, A.T., Scheid, J.F., Nussenzweig, M.C., et al. (2016). Structural basis for germline antibody recognition of HIV-1 immunogens. Elife *5*.

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.

Scheid, J.F., Mouquet, H., Ueberheide, B., Diskin, R., Klein, F., Oliveira, T.Y.K., Pietzsch, J., Fenyo, D., Abadir, A., Velinzon, K., et al. (2011). Sequence and Structural Convergence of Broad and Potent HIV Antibodies That Mimic CD4 Binding. Science. *333*, 1633–1637.

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.

Setliff, I., Shiakolas, A.R., Pilewski, K.A., Murji, A.A., Mapengo, R.E., Janowska, K., Richardson, S., Oosthuysen, C., Raju, N., Ronsard, L., et al. (2019). High-Throughput Mapping of B Cell Receptor Sequences to Antigen Specificity. Cell *179*, P1636-1646.E15.

Shugay, M., Bagaev, D. V, Turchaninova, M.A., Bolotin, D.A., Britanova, O. V, Putintseva, E. V, Pogorelyy, M. V, Nazarov, V.I., Zvyagin, I. V, Kirgizova, V.I., et al. (2015). VDJtools: Unifying Post-analysis of T Cell Receptor Repertoires. PLOS Comput. Biol. *11*, e1004503.

Singh, S., Tank, N.K., Dwiwedi, P., Charan, J., Kaur, R., Sidhu, P., and Chugh, V.K. (2017). Monoclonal Antibodies: A Review. Curr. Clin. Pharmacol. *13*, 85–99.

Sok, D., van Gils, M.J., Pauthner, M., Julien, J.-P., Saye-Francisco, K.L., Hsueh, J., Briney, B., Lee, J.H., Le, K.M., Lee, P.S., et al. (2014). Recombinant HIV envelope trimer selects for quaternary-dependent antibodies targeting the trimer apex. Proc. Natl. Acad. Sci. *111*, 17624–17629.

Sok, D., Briney, B., Jardine, J.G., Kulp, D.W., Menis, S., Pauthner, M., Wood, A., Lee, E.C., Le, K.M., Jones, M., et al. (2016). Priming HIV-1 broadly neutralizing antibody precursors in human Ig loci transgenic mice. Science. *353*, 1557–1560.

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. (2019). High frequency of shared clonotypes in human B cell receptor repertoires. Nature *566*, 398–402.

Sparrow, E., Friede, M., Sheikh, M., and Torvaldsen, S. (2017). Therapeutic antibodies for infectious diseases. Bull. World Health Organ. *95*, 235–237.

Stamatatos, L., Pancera, M., and McGuire, A.T. (2017). Germline-targeting immunogens. Immunol. Rev. *275*, 203–216.

Stevens, J., Crowe, J.E., Martinez, O., McGraw, P.A., House, F.S., Pappas, C., Hicar, M.D., Tumpey, T.M., Keefer, C.J., Yu, X., et al. (2008). Neutralizing antibodies derived from the B cells of 1918 influenza pandemic survivors. Nature *455*, 532–536.

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.

Stokes, J., Farquhar, J.A., Drake, M.E., Capps, R.B., Ward, C.S., Mills, O., and Kitts, A.W. (1951). Infectious hepatitis: Length of protection by immune serum globulin (gamma globulin) during epidemics. J. Am. Med. Assoc. *147*, 714–719.

Tan, J., Sack, B.K., Oyen, D., Zenklusen, I., Piccoli, L., Barbieri, S., Foglierini, M., Fregni, C.S., Marcandalli, J., Jongo, S., et al. (2018). A public antibody lineage that potently inhibits malaria infection through dual binding to the circumsporozoite protein. Nat. Med. *24*, 401–407.

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., et al. (2014). Barcode-enabled sequencing of plasmablast antibody repertoires in rheumatoid arthritis. Arthritis Rheumatol. *66*, 2706–2715.

Taylor, J.J., Martinez, R.J., Titcombe, P.J., Barsness, L.O., Thomas, S.R., Zhang, N., Katzman, S.D., Jenkins, M.K., and Mueller, D.L. (2012). Deletion and anergy of polyclonal B cells specific for ubiquitous membrane-bound self-antigen. J. Exp. Med. *209*, 2065–2077.

Tonegawa, S. (1983). Somatic generation of antibody diversity. Nature *302*, 575–581.

Traggiai, E., Becker, S., Subbarao, K., Kolesnikova, L., Uematsu, Y., Gismondo, M.R., Murphy, B.R., Rappuoli, R., and Lanzavecchia, A. (2004). An efficient method to make human monoclonal antibodies from memory B cells: Potent neutralization of SARS coronavirus. Nat. Med. *10*, 871–875.

Trepel, F. (1974). Number and distribution of lymphocytes in man. A critical analysis. Klin. Wochenschr. *52*, 511–515.

Trück, J., Ramasamy, M.N., Galson, J.D., Rance, R., Parkhill, J., Lunter, G., Pollard, A.J., and Kelly, D.F. (2015). Identification of Antigen-Specific B Cell Receptor Sequences Using Public Repertoire Analysis. J. Immunol. *194*, 252–261.

Walker, L.M., Phogat, S.K., Chan-Hui, P.-Y., Wagner, D., Phung, P., Goss, J.L., Wrin, 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.

Ward, A.B., and Wilson, I.A. (2015). Insights into the trimeric HIV-1 envelope glycoprotein structure. Trends Biochem. Sci. *40*, 101–107.

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.

Weinstein, J.A., Jiang, N., White, R.A., Fisher, D.S., and Quake, S.R. (2009). High-throughput sequencing of the zebrafish antibody repertoire. Science. *324*, 807–810.

West, A.P., Diskin, R., Nussenzweig, M.C., and Bjorkman, P.J. (2012). Structural basis for germ-line gene usage of a potent class of antibodies targeting the CD4-binding site of HIV-1 gp120. Proc. Natl. Acad. Sci. *109*, E2083–E2090.

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 LP – 4057.

Wilson, P.C., and Andrews, S.F. (2012). Tools to therapeutically harness the human antibody response. Nat. Rev. Immunol. *12*, 709–719.

Wine, Y., Boutz, D.R., Lavinder, J.J., Miklos, A.E., Hughes, R.A., Hoi, K.H., Jung, S.T., Horton, A.P., Murrin, E.M., Ellington, A.D., et al. (2013). Molecular deconvolution of the monoclonal antibodies that comprise the polyclonal serum response. Proc. Natl. Acad. Sci. U. S. A. *110*, 2993–2998.

Wu, X., Yang, Z.-Y., Li, Y., Hogerkorp, 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., Zhou, T., Zhu, J., Zhang, B., Georgiev, I., Wang, C., Chen, X., Longo, N.S., Louder, M., McKee, K., et al. (2011). Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. Science *333*, 1593–1602.

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., Do Kwon, Y., 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.

Xu, K., Acharya, P., Kong, R., Cheng, C., Chuang, G.Y., Liu, K., Louder, M.K., O'Dell,

S., Rawi, R., Sastry, M., et al. (2018). Epitope-based vaccine design yields fusion peptide-directed antibodies that neutralize diverse strains of HIV-1. Nat. Med. *24*, 857–867.

Yacoob, C., Pancera, M., Vigdorovich, V., Oliver, B.G., Glenn, J.A., Feng, J., Sather, D.N., McGuire, A.T., and Stamatatos, L. (2016). Differences in Allelic Frequency and CDRH3 Region Limit the Engagement of HIV Env Immunogens by Putative VRC01 Neutralizing Antibody Precursors. Cell Rep. *17*, 1560–1570.

Ye, J., Ma, N., Madden, T.L., and Ostell, J.M. (2013). IgBLAST: an immunoglobulin variable domain sequence analysis tool. Nucleic Acids Res. *41*, W34–W40.

Yoon, H., Macke, J., West, A.P., Foley, B., Bjorkman, P.J., Korber, B., and Yusim, K. (2015). CATNAP: A tool to compile, analyze and tally neutralizing antibody panels. Nucleic Acids Res. *43*, W213–W219.

Yu, X., McGraw, P.A., House, F.S., and Crowe, J.E. (2008). An optimized electrofusion-based protocol for generating virus-specific human monoclonal antibodies. J. Immunol. Methods *336*, 142–151.

Zhang, R., Fichtenbaum, C.J., Hildeman, D.A., Lifson, J.D., and Chougnet, C. (2004). CD40 Ligand Dysregulation in HIV Infection: HIV Glycoprotein 120 Inhibits Signaling Cascades Upstream of CD40 Ligand Transcription. J. Immunol. *172*, 2678–2686.

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.

Zhou, T., Georgiev, I., Wu, X., Yang, Z.-Y., Dai, K., Finzi, A., Do Kwon, Y., Scheid, J.F., Shi, W., Xu, L., et al. (2010). Structural Basis for Broad and Potent Neutralization of HIV-1 by Antibody VRC01. Science. *329*, 811 LP – 817.

Zhou, T., Zhu, J., Wu, X., Moquin, S., Zhang, B., Acharya, P., Georgiev, I.S., Altae-Tran, H.R., Chuang, G.Y., Joyce, M.G., et al. (2013). Multidonor analysis reveals structural elements, genetic determinants, and maturation pathway for HIV-1 neutralization by VRC01-class antibodies. Immunity *39*, 245–258.

Zhou, T., Lynch, R.M., Chen, L., Acharya, P., Wu, X., Doria-Rose, N.A., Joyce, M.G., Lingwood, D., Soto, C., Bailer, R.T., et al. (2015). Structural repertoire of HIV-1-neutralizing antibodies targeting the CD4 supersite in 14 donors. Cell *161*, 1280–1292.

Zhu, J., Wu, X., Zhang, B., McKee, K., O'Dell, S., Soto, C., Zhou, T., Casazza, J.P., Mullikin, J.C., Kwong, P.D., et al. (2013a). De novo identification of VRC01 class HIV-1-neutralizing antibodies by next-generation sequencing of B-cell transcripts. Proc. Natl. Acad. Sci. U. S. A. *110*.

Zhu, J., Ofek, G., Yang, Y., Zhang, B., Louder, M.K., Lu, G., McKee, K., Pancera, M., Skinner, J., Zhang, Z., et al. (2013b). Mining the antibodyome for HIV-1-neutralizing antibodies with next-generation sequencing and phylogenetic pairing of heavy/light chains. Proc. Natl. Acad. Sci. U. S. A. *110*, 6470–6475.