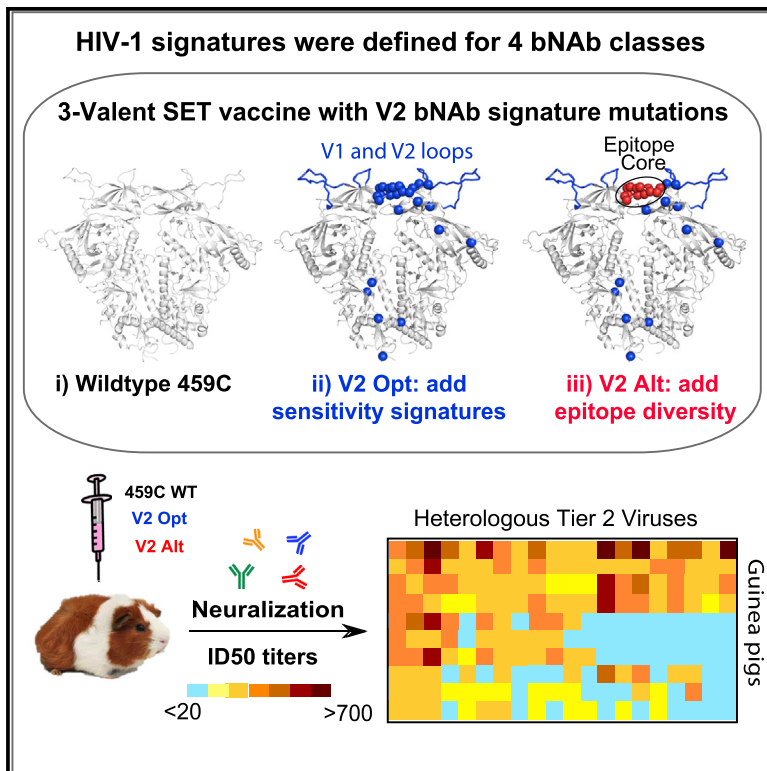


Cell Host & Microbe

HIV-1 Neutralizing Antibody Signatures and Application to Epitope-Targeted Vaccine Design

Graphical Abstract



Authors

Christine A. Bricault, Karina Yusim, Michael S. Seaman, ..., David C. Montefiori, Dan H. Barouch, Bette Korber

Correspondence

dbarouch@bidmc.harvard.edu (D.H.B.), btk@lanl.gov (B.K.)

In Brief

HIV-1 Env amino acid signatures associated with sensitivity to broadly neutralizing antibodies were systematically defined from large neutralization panels. V2 signatures were incorporated in a trivalent vaccine to enhance epitope exposure and to include common epitope variants, resulting in increased neutralization breadth against heterologous viruses in a guinea pig model.

Highlights

- HIV-1 bNAb sensitivity signatures from 4 large virus panels mapped across 4 Ab classes
- Non-contact hypervariable region characteristics are critical for bNAb sensitivity
- HIV-1 Env 459C used alone as a vaccine can elicit modest tier 2 NAb in guinea pigs
- V2 bNAb signature-guided modifications in 459C enhanced neutralization breadth



HIV-1 Neutralizing Antibody Signatures and Application to Epitope-Targeted Vaccine Design

Christine A. Bricault,^{1,17} Karina Yusim,^{2,3,17} Michael S. Seaman,¹ Hyejin Yoon,² James Theiler,^{2,3} Elena E. Giorgi,^{2,3} Kshitij Wagh,^{2,3} Maxwell Theiler,² Peter Hraber,² Jennifer P. Macke,² Edward F. Kreider,⁴ Gerald H. Learn,⁴ Beatrice H. Hahn,⁴ Johannes F. Scheid,^{5,6} James M. Kovacs,^{7,8,9} Jennifer L. Shields,¹ Christy L. Lavine,¹ Fadi Ghantous,¹ Michael Rist,¹ Madeleine G. Bayne,¹ George H. Neubauer,¹ Katherine McMahan,¹ Hanqin Peng,^{7,8} Coraline Chéneau,¹ Jennifer J. Jones,¹⁰ Jie Zeng,¹⁰ Christina Ochsenbauer,¹⁰ Joseph P. Nkolola,¹ Kathryn E. Stephenson,^{1,11} Bing Chen,^{7,8} S. Gnanakaran,^{2,3} Mattia Bonsignori,^{12,13} LaTonya D. Williams,¹² Barton F. Haynes,^{12,13,14} Nicole Doria-Rose,¹⁵ John R. Mascola,¹⁵ David C. Montefiori,^{12,16} Dan H. Barouch,^{1,11,18,19,*} and Bette Korber^{2,3,18,*}

¹Center for Virology and Vaccine Research, Beth Israel Deaconess Medical Center, Boston, MA 02215, USA

²Los Alamos National Laboratory, Los Alamos, NM 87545, USA

³New Mexico Consortium, Los Alamos, NM 87545, USA

⁴Departments of Medicine and Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

⁵Massachusetts General Hospital, Boston, MA 02114, USA

⁶Harvard Medical School, Boston, MA 02114, USA

⁷Division of Molecular Medicine, Children's Hospital, Boston, MA 02115, USA

⁸Department of Pediatrics, Harvard Medical School, Boston, MA 02115, USA

⁹Departments of Chemistry and Biochemistry, University of Colorado, Colorado Springs, CO 80918, USA

¹⁰Department of Medicine and CFAR, University of Alabama at Birmingham, Birmingham, AL 35294, USA

¹¹Ragon Institute of Massachusetts General Hospital, MIT, and Harvard, Boston, MA 02114, USA

¹²Duke Human Vaccine Institute, Duke University School of Medicine, Durham, NC 27710, USA

¹³Department of Medicine, Duke University School of Medicine, Durham, NC 27710, USA

¹⁴Department of Immunology, Duke University School of Medicine, Durham, NC 27710, USA

¹⁵Vaccine Research Center, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20814, USA

¹⁶Department of Surgery, Duke University School of Medicine, Durham, NC 27710, USA

¹⁷These authors contributed equally

¹⁸These authors contributed equally

¹⁹Lead Contact

*Correspondence: dbarouch@bidmc.harvard.edu (D.H.B.), btk@lanl.gov (B.K.)

<https://doi.org/10.1016/j.chom.2018.12.001>

SUMMARY

Eliciting HIV-1-specific broadly neutralizing antibodies (bNAbs) remains a challenge for vaccine development, and the potential of passively delivered bNAbs for prophylaxis and therapeutics is being explored. We used neutralization data from four large virus panels to comprehensively map viral signatures associated with bNAb sensitivity, including amino acids, hypervariable region characteristics, and clade effects across four different classes of bNAbs. The bNAb signatures defined for the variable loop 2 (V2) epitope region of HIV-1 Env were then employed to inform immunogen design in a proof-of-concept exploration of signature-based epitope targeted (SET) vaccines. V2 bNAb signature-guided mutations were introduced into Env 459C to create a trivalent vaccine, and immunization of guinea pigs with V2-SET vaccines resulted in increased breadth of NAb responses compared with Env 459C alone. These data demonstrate that bNAb signatures can be utilized to engineer HIV-1 Env vaccine immunogens capable of eliciting antibody responses with greater neutralization breadth.

INTRODUCTION

Vaccine induction of broadly neutralizing antibodies (bNAbs) against diverse global tier 2 HIV-1 strains remains an unsolved challenge for the HIV-1 vaccine field. There has been some progress in animal models (Escolano et al., 2016; Saunders et al., 2017), but human trials have yet to elicit bNAbs, although NAbs with varying levels of breadth arise during natural infection (Hraber et al., 2014b). bNAbs typically develop slowly during chronic infection as the virus diversifies under immune pressure and B cell lineages adapt to the evolving virus (Bonsignori et al., 2017b; Doria-Rose et al., 2014; Liao et al., 2013; Wu et al., 2015). bNAb breadth and potency are evaluated using large panels of HIV-1 Envelope (Env) pseudoviruses that sample global HIV-1 diversity (Hraber et al., 2014a) or the C clade diversity of Southern Africa (Rademeyer et al., 2016).

We used data from 4 large neutralization panels for a more comprehensive mapping of viral signatures associated with bNAb sensitivity than undertaken previously (Chuang et al., 2013; Evans et al., 2014; Ferguson et al., 2013; West et al., 2013). Signature sites were identified using a strategy that incorporates a phylogenetic correction (Gnanakaran et al., 2010) for amino acids (AAs) and potential N-linked glycosylation sites (PNGSs) (Crispin and Doores, 2015), and we also explored the impacts of hypervariable region characteristics



and clades. Recurrent signature patterns were found among bNAbs with shared specificities (Burton and Hangartner, 2016).

We next used variable V2 apex (V2) epitope bNAb signatures to inform HIV-1 Env immunogen design in a proof-of-concept exploration of an approach we call signature-based epitope targeted (SET) vaccines. Other vaccine design strategies include engaging bNAb germline precursors (Steichen et al., 2016), using polyvalent sets to capture diversity (Korber et al., 2017), lineage-based designs (Bonsignori et al., 2017b), and engineered native-like Envs (e.g., SOSIPs) (Sanders et al., 2013; Steichen et al., 2016). SOSIP vaccines elicit robust autologous NABs that have limited breadth in rabbits and non-human primates (NHPs) (Pauthner et al., 2017; Sanders et al., 2015).

Our SET vaccine design started with the Env 459C (Bricault et al., 2015), as 459C alone elicited modest neutralization of some tier 2 heterologous strains in guinea pigs. V2-SET immunogens are a trivalent combination of 459C wild-type (WT) plus two additional proteins designed by modifying 459C to include V2 bNAb signatures intended to both enhance V2 epitope exposure and include relevant variation. V2-SET vaccines expressed as either gp140 SOSIP trimers or foldon trimers elicited increased NAB breadth compared to 459C alone in guinea pigs, suggesting the potential utility of bNAb signatures in vaccine design.

RESULTS

Neutralization Data

Four datasets measuring the sensitivity of bNAbs against panels of HIV-1 Envs were analyzed. Three panels sampled global viral diversity (Hraber et al., 2014a), and the other sampled only C clade, which dominates in Southern Africa (Rademeyer et al., 2016). Table S1 summarizes bNAb dataset inclusion, relationships, and provenance. bNAbs are grouped by epitope class: V2, V3 glycan (V3), CD4 binding site (CD4bs), and membrane proximal external region (MPER) (Burton and Mascola, 2015). V2 and V3 bNAbs often have great potency but limited breadth, CD4bs have expanded breadth, and MPER has high breadth but low potency (Figures 1, S1, and S2). Heatmaps displaying inhibitory concentrations of 50% (IC_{50}) data illustrate shared pattern sensitivity across bNAb classes (Figures 1 and S1), enabling the definition of common bNAb class signatures.

Neutralization Sensitivity and HIV-1 Clades

bNAb sensitivity patterns are associated with HIV-1 clades, which have geographic associations (Korber et al., 2009). The V2 bNAb lineage CAP256.VRC26 is less potent against clade B viruses (Doria-Rose et al., 2016); diminished clade B potency extends across the V2 bNAb class (Figures 1, 2A, and S2). CD4bs bNAbs are more sensitive to clade A viruses, and some bNAbs are less sensitive to C clade, in particular 3BNC117 and VRC01 (Figures 1, 2A, and S2). MPER bNAbs are less potent against clade A. Circulating Recombinant Form 01 (CRF01) viruses, common in Southeast Asia, are extremely resistant to V3 bNAbs (Figures 1 and S2), which is driven by a PNGS shift from positions N332 to N334 found in 96% of CRF01 viruses (Figure 1; Table S2). The N332 glycan directly contacts V3 bNAbs (Kong et al., 2013) and is essential for some, although PGT128

can tolerate its loss (Figure 1). CRF01 was already highly diverse in central Africa when a founder seeded the Thai epidemic (Korber et al., 2000); African CRF01 viruses also lack the PNGS at N332. PNGS N332 frequencies vary between clades, correlating with overall clade sensitivity to V3 bNAbs (Table S2); e.g., ~40% of A clade viruses are resistant to most V3 bNAbs, and 43% lack the PNGS N332 (Figure 2A; Table S2). The N332 PNGS is under-represented among transmitted C clade viruses (Rademeyer et al., 2016), which may impact the efficacy of V3 bNAbs in southern Africa.

Amino Acid Signatures

Signature patterns were identified by testing each AA and PNGS across Env alignments for associations with sensitivity to each bNAb in each of the 4 neutralization panels (Figure S1), based on associations with either potency or detectable neutralization (Tables S3A–S3D; Figure 3). Key signature sites were initially identified using a method that corrects for phylogenetic artifacts (Gnanakaran et al., 2010) (Tables S3E–S3H). Once a site was deemed of interest by this stringent criterion, any significant associations within these key sites were listed (Tables S3I–S3L).

CD4bs bNAbs

While many CD4bs signatures are in contacts (Figures 3 and S3; Table S3) and known to influence sensitivity (Gao et al., 2014; Lynch et al., 2015), most are outside of contact surfaces. We tested two CD4bs resistance signatures directly by introducing them into the CH505 transmitted-founder (TF) backbone. The G458Y signature mutation conferred complete resistance ($IC_{50} > 25 \mu\text{g/mL}$) to VRC01 and 3BNC117, and both can neutralize the CH505 TF (IC_{50} of 0.14 and 0.03 $\mu\text{g/mL}$, respectively). The non-contact T234N signature substitution introduced a PNGS at N234 that increased resistance to CD4bs bNAbs 5- to 7-fold, and IC_{50} titers went from 0.36 to 1.82 $\mu\text{g/mL}$ for CH235 and from 0.12 to 0.87 $\mu\text{g/mL}$ for VRC01.

MPER bNAbs

As their breadth is so great, most 10E8, 4E10, and DH511 signatures are associated with potency, except for rare mutations associated with complete resistance: W672L, F673L, and W680G (Figures 3 and S3; Table S3). The 10E8-specific resistance signature N671T accounts for the increased breadth of DH511 lineage bNAbs over 10E8 (Williams et al., 2017). The MPER epitope signature K683R is also associated with resistance to CD4bs and V2 bNAbs (Figure 3; Table S3), consistent with MPER changes impacting overall neutralization sensitivity (Bradley et al., 2016). Clade C resistance to 2F5 may be explained by clade C not having an otherwise conserved Ala, A667 (Figure S3).

V3 bNAbs

Glycans that interact directly with V3 bNAbs (Behrens et al., 2016) and are positive signatures for neutralization are at positions N332, N301, and N295 (Figures 3 and S3; Table S3). Similar to CD4bs and MPER bNAbs, V3 bNAb sensitivity signatures in the epitope were relatively conserved, with positive signatures as the most common variant (Figure 3). In contrast, V3 bNAb signature sites between positions 336–442 were extremely variable and so may contribute to more nuanced levels of potency. In

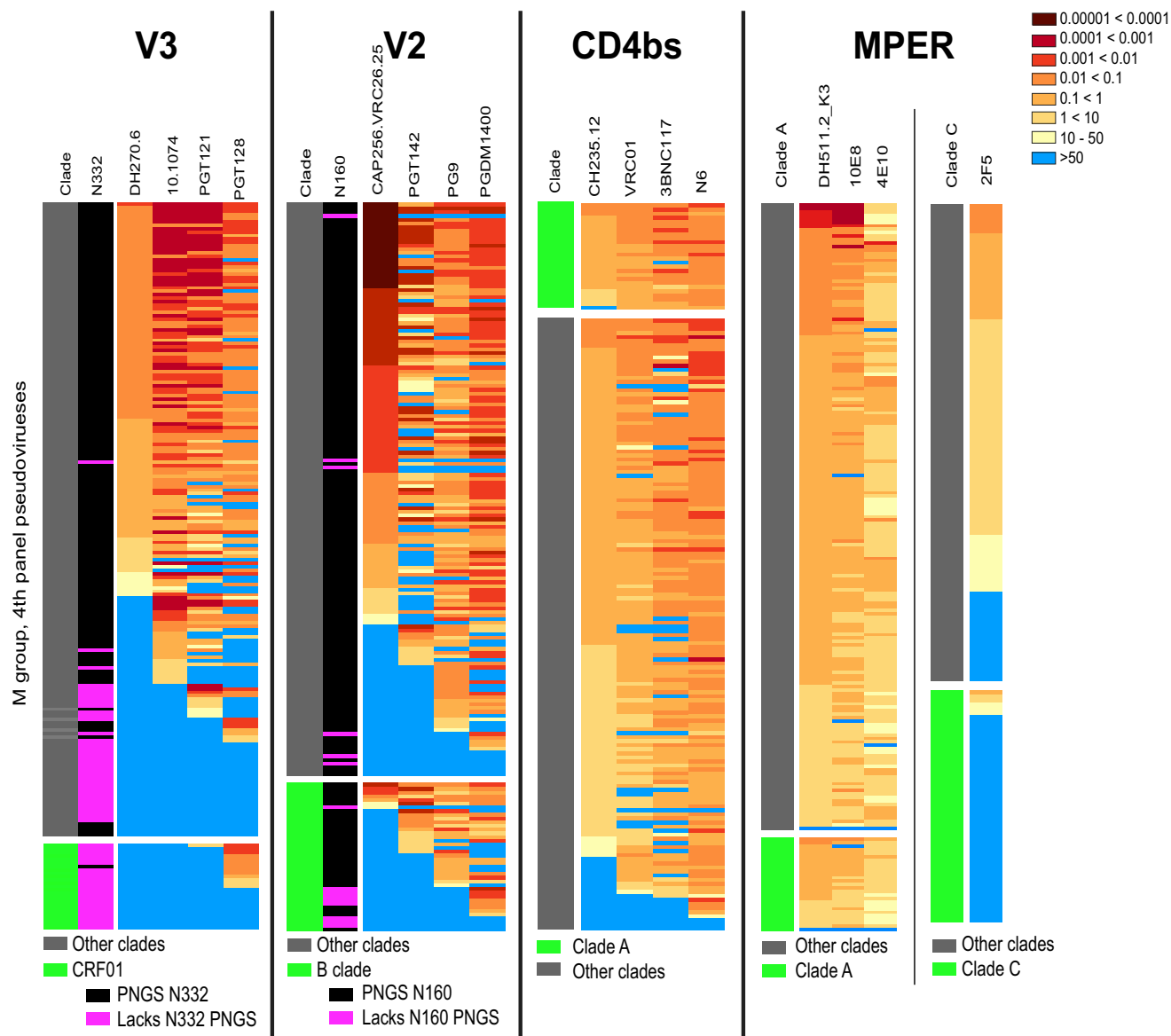


Figure 1. Heatmaps Showing IC_{50} Neutralization Titers for Dataset 4

Darker red hues indicate more potent neutralization and blue indicates undetected responses. Rows represent pseudoviruses, ordered differently in each panel to highlight commonalities in neutralization profiles across bNAbs in each class. The clade with the strongest clade effect is separated and indicated in green to the left. Key PNGSs are indicated by magenta. Among MPER bNAbs 2F5 is considered separately as it has a unique epitope.

the GDIR contact motif (Sok et al., 2016) only D325 is a signature, because the other 3 positions are nearly invariant.

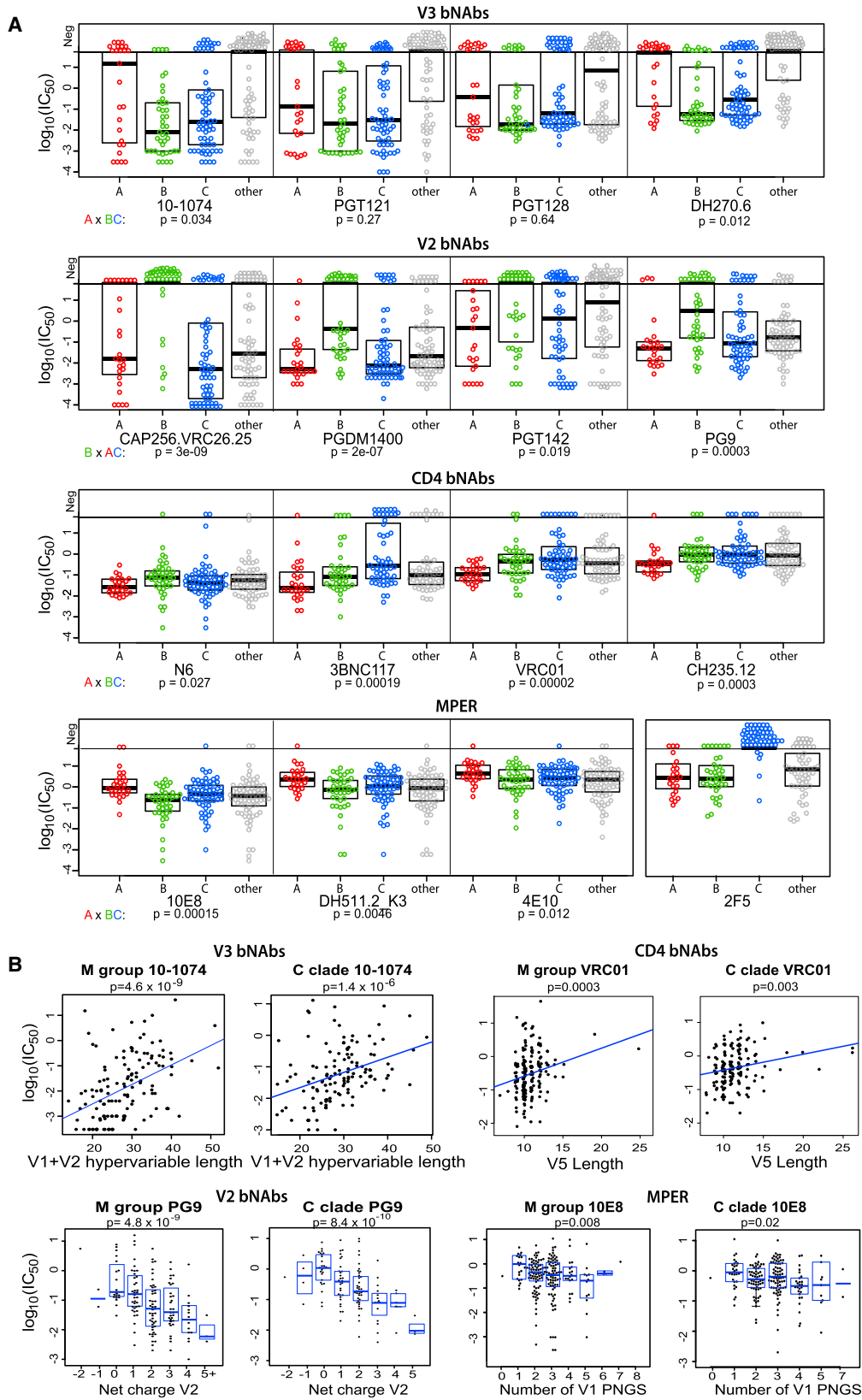
V2 bNAbs

V2 bNAbs contact glycans at positions N156 and N160. The PNGS N160 is critical for many V2 bNAbs, with a few exceptions that may be enabled by nearby glycans acting in compensatory role (Figure 1; Table S3) (McLellan et al., 2011). In contrast, 11 viruses of the 380 unique viruses in the combined datasets 3 and 4 lacked the PNGS at N156, and these 11 viruses had very similar distributions of IC_{50} titers to viruses that had the PNGS site for all V2 bNAbs tested. PNGSs at positions N130 and N332 are associated with V2

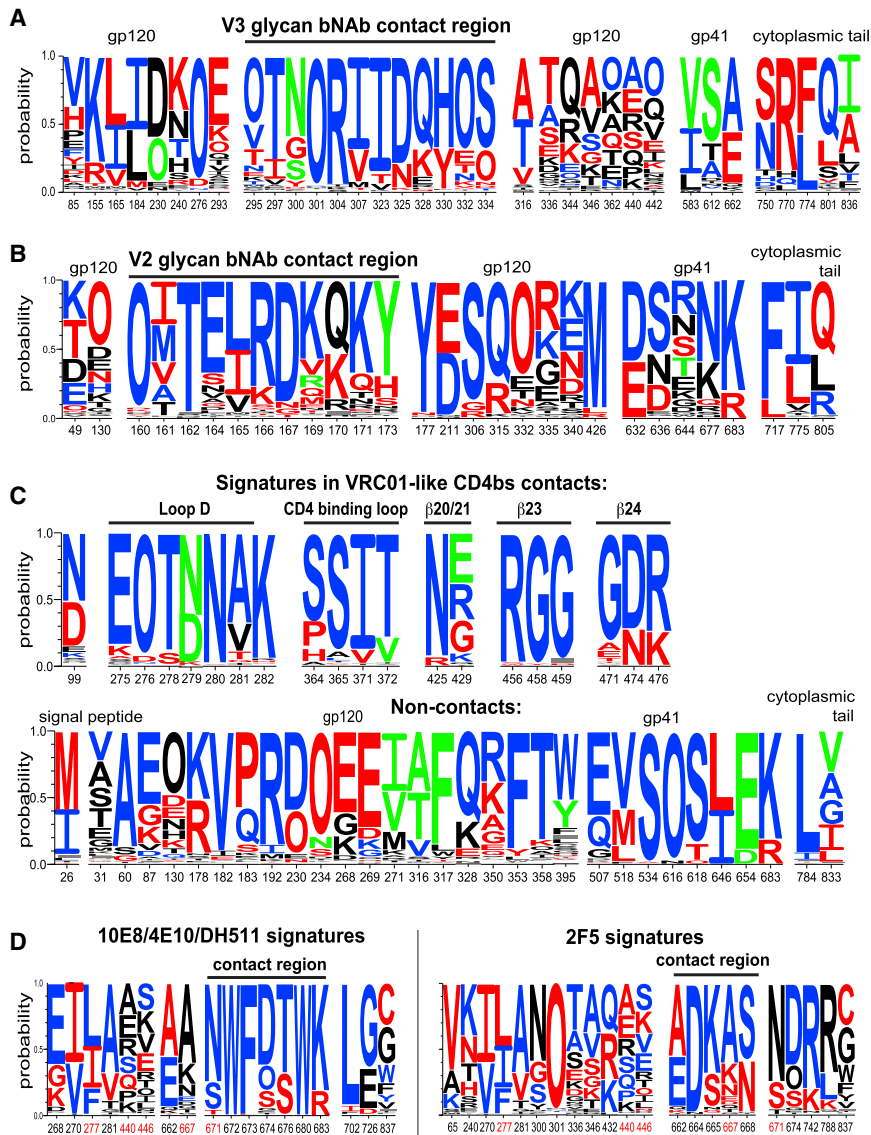
bNAb resistance. The glycan at N130 is near the CAP256.VRC26 contact surface (Figure S3) and it interacts with CH03 (Gorman et al., 2016), but glycans at N130 and N332 may also act indirectly through glycan dynamics (Stewart-Jones et al., 2016) or carbohydrate processing (Behrens et al., 2016). Many V2 bNAb contact signatures have been studied: K169 and K171 loss, and Q170K, confers resistance, while E/D164 increases CAP256.VRC26.25 sensitivity (Doria-Rose et al., 2012, 2016).

Evolutionary Counter-pressure

Some resistance signatures come with a fitness cost (Lynch et al., 2015). Also, some signatures are associated with resistance or



(legend on next page)



sensitivity depending on the bNAb class (Figure 3; Table S3), including the PNGS at N130, associated with V2 bNAb resistance and CD4bs VRC03 sensitivity; the PNGS at N332, associated with V3 bNAb sensitivity and V2 bNAb resistance; and L165, associated with V2 bNAb sensitivity and V3 bNAb resistance (Figure 3; Table S3). Antibodies within a class can also have contradictory signatures. Some MPER signature AAs have opposing associations for 2F5 versus 4E10/10E8/DH511 (Figures 3 and S3; Table S3). A negatively charged D279 is associated with VRC01 sensitivity and with 12A12 resistance, likely due to the local

Figure 3. Sequence LOGOs of AA Signatures by Antibody Class

This figure highlights the more robust signature sites in that they were supported by multiple lines of evidence—they either had phylogenetically corrected associations supported by at least 2 datasets, were a signature site in a contact residue, or both. Not all bNAbs in a class are associated with every signature. Complete lists with detailed statistics are provided in Table S3. Letter height represents AA frequencies in dataset 4. “O” represents an Asn in a PNGS motif. AAs associated with resistance and sensitivity are red and blue, respectively. AAs shown in green differ for different bNAbs within the class. (A) V3 bNAbs, (B) V2, (C) VH1-2 and VH1-46 CD4bs, and (D) MPER, with 10E8/4E10/DH511 on the left, 2F5 on the right, and red HXB2 position numbers highlighting opposing signatures between the two.

charge in the bNAb paratopes (Figure S3) (Klein et al., 2013). CD4bs signatures for CDRH3 bNAbs (Zhou et al., 2015) often had opposing signatures relative to VH1-2 or VH1-46 bNAbs (Table S3).

Clade Sensitivity

Signature sites offer hypotheses to explain bNAb clade sensitivities. Four contact site candidates were proposed to contribute to the reduced reactivity of CAP256.VRC26 bNAbs with B clade viruses (Doria-Rose et al., 2016). We found an additional 17 signatures that may limit V2 bNAb potency against B clade viruses (Figure S4A). In contrast, CAP256.VRC26 bNAbs are most potent against clade C viruses and 12 signatures may be relevant. CD4bs antibodies have enhanced potency against A clade viruses, and resistance signatures are relatively rare in A clade (Figure S4B); in contrast, 3BNC117 and VRC01 have reduced breadth and potency against C clade viruses (Figure 2).

Hypervariable Loops and Neutralization Sensitivity

HIV-1 Env hypervariable regions evolve rapidly *in vivo* by insertions and deletions, giving rise to extreme length variation and making alignments spanning these regions unreliable. These changes can mediate bNAb escape (Bonsignori et al., 2016; Gao et al., 2014; Korber et al., 2017). We use alignment-independent characteristics of these regions (length, net charge, and

Figure 2. Env Characteristics Associated with bNAb Class Sensitivity

(A) Clade associations. Circles illustrate IC_{50} titers from dataset 4, highlighting the 3 best represented clades: A in red, B in green, and C and CRF07 (which is clade C in Env) in blue. All others are gray. Boxplots show medians and quartiles. Patterns of relative clade sensitivity are consistent across bNAb classes. The p values are based on two-sided Wilcoxon tests comparing the most distinctive clade among A, B, and C to the other two clades. Points above the horizontal line were above the threshold of detection. The bolus of negative points in the “other” group for V3 bNAbs is primarily CRF01.

(B) Examples of hypervariable loop characteristic correlations with bNAb sensitivity, including one for each bNAb class (complete associations are in Tables S3N–S3Q). M-group and clade C data are from datasets 4 and 3, respectively. The p values are based on Kendall’s tau.

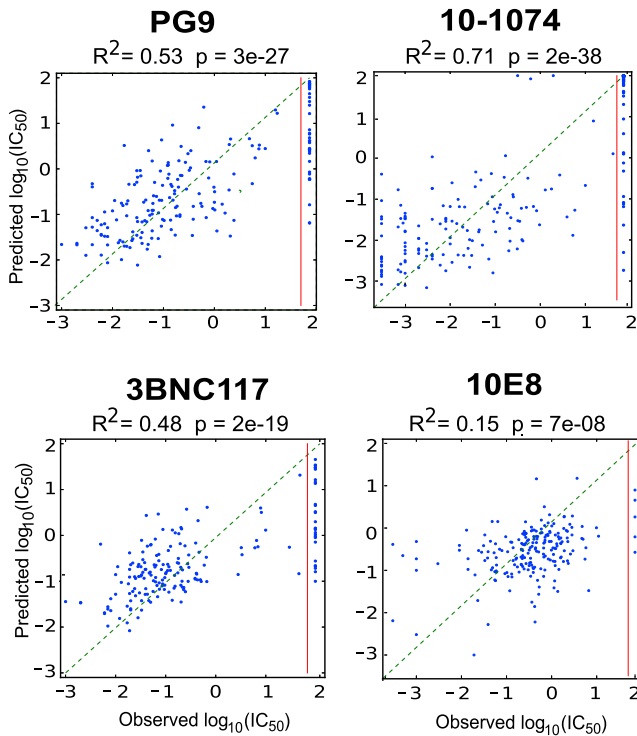


Figure 4. Random Forest Signature-Based Prediction Accuracy

Leave-one-out cross-validation regression predictions using dataset 4 for one bNAb from each bNAb class. R^2 is the standard coefficient of determination; the p value is from Kendall's tau. The red line marks the threshold of detection. R^2 values remained significant when negative points were excluded: PG9, $R^2 = 0.35$, $p = 4 \times 10^{-14}$; 10-1074, $R^2 = 0.27$, $p = 2 \times 10^{-9}$; 3BNC117, $R^2 = 0.30$, $p = 3 \times 10^{-9}$; and 10E8, $R^2 = 0.12$, $p = 9 \times 10^{-7}$.

number of PNGSs) to identify patterns associated with bNAb resistance profiles. The strongest of these associations are included in Tables S3M–S3P, and examples are shown in Figure 2B.

Combined V1+V2 length was the strongest correlate with V3 bNAb neutralization resistance—so insertions in either or both regions reduce sensitivity—followed by the number of PNGSs in V1+V2 and V1 length (Figure 2B; Table S3). V1 length variation played a critical role in the development of V3 bNAb DH270 lineage (Bonsignori et al., 2017a). A glycan in hypervariable V1 contacts PGT121-family bNAbs, enabling interactions when the key N332 glycan is absent. However, when N332 glycan is present, consistent with our signature predictions, removing this glycan enhances PGT121-family sensitivity (Garces et al., 2015).

The strongest variable region correlation with V2 bNAb sensitivity was V2 loop net-positive charge (Figure 2B; Table S6). The effect remained strong when only the V2 hypervariable region (positions 185–190) was considered. V2 bNAbs have long anionic CDRH3 loops (Doria-Rose et al., 2012; McLellan et al., 2011), which may drive the preference for positive charge. Also, V1+V2 hypervariable region length was inversely correlated with V2 bNAb sensitivity (Table S3).

Paradoxically, although CD4bs bNAbs can have great breadth, they bind near the V5 region which is subject to extreme length variation. Both V5 length and the number of PNGSs within

V5 are associated with reduced CD4bs bNAb potency (Figure 2B; Table S3). CD4bs bNAbs can be selected to tolerate changes in V5 length after they arise as escape mutations *in vivo* (Fera et al., 2014). Long V1 and V2 regions also correlate with CD4bs bNAb resistance (Table S3) and mediate *in vivo* escape (Gao et al., 2014). Hence, long V1 and V2 regions are associated with relative resistance to 3 major classes of bNAbs: CD4bs, V2, and V3. At the population level, there are evolutionary counter-pressures against larger loops, as smaller hypervariable regions tend to be selected at transmission (Derdeyn et al., 2004). Unexpectedly, increasing numbers of PNGSs in V1 correlated with enhanced sensitivity to MPER bNAbs, particularly 10E8 (Figure 2B; Table S3).

Machine Learning Predictions of Env Neutralization

We next explored using bNAb signatures for sequence-based machine learning predictions of bNAb sensitivity. Using a Random Forest (RF) for IC_{50} regression predictions, we compared prediction accuracies for 3 prefiltering strategies: (1) a standard prefilter, minimal-redundancy-maximal-relevance (mRMR) (Peng et al., 2005); (2) the full bNAb signatures for each antibody class, including AA, PNGS, clade, and variable loop characteristics; and (3) only signatures in contact sites. We evaluated predictions using leave-one-out cross validation for dataset 4, as well as for an independent C clade holdout set. The accuracy using the full signature was superior to mRMR ($p = 0.003$, paired Wilcoxon, C clade holdout comparison), and to the contact-region-only signature (also $p = 0.003$) (Table S4); examples are shown in Figure 4. We also tested positive/negative classification predictions using our prefilter strategies and RF, comparing them to a published method called IDEpi (Hepler et al., 2014); the methods were comparable (Table S5). Hypervariable region characteristics were consistently among the most important factors for predicting IC_{50} titers (Table S6).

V2-SET Vaccines

Immunogen Design and Expression

We next utilized V2 bNAb signatures to inform vaccine design, intending to increase V2 epitope exposure and represent relevant natural diversity within the epitope. V2-SET trivalent vaccines included 459C WT, which alone induced low-level neutralization of some tier 2 viruses plus two complementary immunogens, called Optimal (Opt) and Alternative (Alt) (Figures 5A and 5B; Table S7). Opt introduced V2 bNAb virus sensitivity signatures into the 459C WT backbone, to enhance epitope expression, exposure, affinity, or relevant carbohydrate processing (Figure 5A). Alt incorporated V2 bNAb sensitivity signatures outside of the contact region; however, within the epitope it captured natural diversity in V2 signature sites—including globally common AAs complementary to those found in 459C WT and Opt—even if associated with relative resistance (Figure 5B). V2-SET immunogens also incorporated modified hypervariable regions with characteristics favoring V2 bNAb sensitivity, including short V1 and V2 hypervariable regions with a positively charged V2 region (Figure 2B; Tables S3N–S3R). The WT Env T250_4 met these criteria and was also highly sensitive to both V2 and V3 bNAbs (Figure S1). Hypothesizing T250_4 hypervariable regions might improve polyclonal

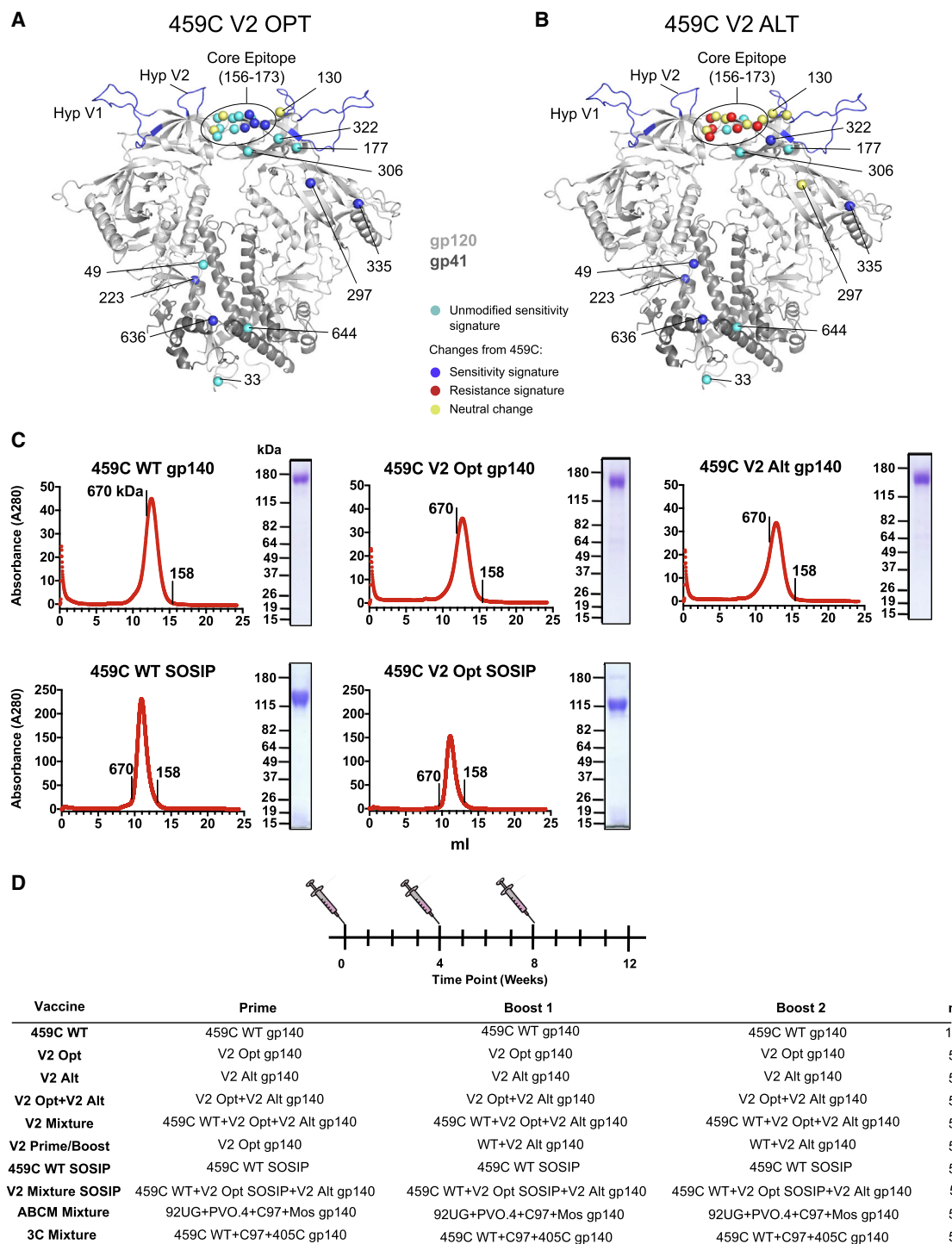


Figure 5. V2-SET Vaccine Design and Production

(A and B) Structural mapping (PDB: 5FYJ) of mutations introduced into 459C WT (Table S7) to create (A) Opt and (B) Alt V2-SET vaccine constructs. Spheres are color-coded to indicate AA modifications associated with sensitivity or resistance. Opt constructs uniformly carry sensitivity signatures. Alt constructs carry mutations that enhance sensitivity outside the core epitope, but in the epitope introduce signature mutations complementary to 459C WT and Opt to capture epitope diversity.

(C) Gel filtration chromatography traces of gp140 V2-SET immunogens run on a Superose 6 column (foldon) or Superdex 200 column (SOSIP). Coomassie stained SDS-PAGE of purified Envs are next to each trace with molecular weight standards noted.

(D) Guinea pig vaccination regimens. Animals were vaccinated intramuscularly in the quadriceps with 100 μ g total immunogen at weeks 0, 4, and 8; n is the group size.

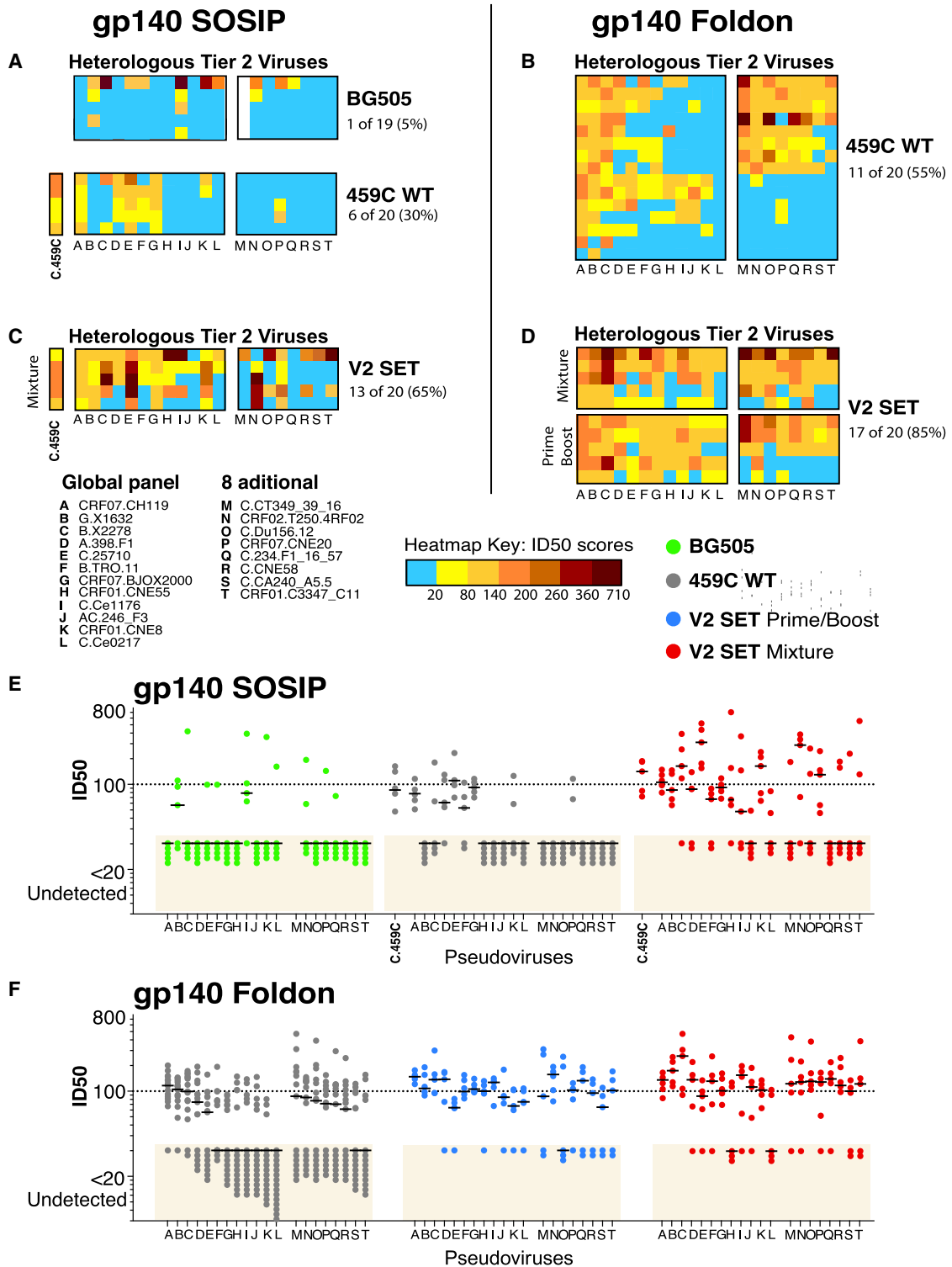


Figure 6. V2-SET Vaccines Improve the Breadth and the Magnitude of Tier 2 NAbs Compared to BG505 and 459C WT

(A–D) Heatmaps of neutralizing responses comparing groups of guinea pigs vaccinated with 459C WT, V2-SET, and BG505 vaccines. Monovalent vaccines: (A) BG505 and 459C WT SOSIP and (B) 459C WT gp140 foldon. Trivalent vaccines: (C) V2-SET SOSIP and (D) gp140 foldon vaccines delivered as either a mixture or prime boost. Columns represent tier 2 pseudoviruses (see key), ordered by sensitivity. Rows represent guinea pigs, organized by vaccine group. The potency of ID₅₀ responses increases from yellow to dark red, below threshold responses are blue. To compare the breadth of response between different vaccine regimens, the median number of detectable responses is reported for each vaccination regimen to the right of the heatmaps, and detectable responses per animal were

(legend continued on next page)

responses to both bNAb classes, we used T250_4's V1 and V2 regions in our V2-SET constructs (Table S7).

The V2-SET gp140s were produced in 293T cells by transient transfection as either SOSIP or foldon immunogens. The gp140 foldon contains a C-terminal T4-fibrin trimerization domain following the MPER and is not cleaved by furin (Frey et al., 2008). The gp140 SOSIP is a native-like trimer truncated before the MPER and cleaved by furin (Steichen et al., 2016). Each purified Env protein ran as a single symmetrical peak by size exclusion chromatography, and as a single band on SDS-PAGE (Figure 5C). We could not produce a stable V2 Alt SOSIP.

The antigenic properties of the V2-SET foldon immunogens were probed using surface plasmon resonance (Tables S5A–S5C). Soluble CD4 (Kwong et al., 1998) and V3 bNAb 10–1074 both bound to all three. V2 bNAb PG16 bound Opt and Alt immunogens more robustly than 459C WT, consistent with increased V2 exposure. Trimer-specific V2 bNABs PGT145 and PGDM1400 bound robustly to SOSIP but not to foldon gp140s, consistent with SOSIP gp140s being native-like (Sanders et al., 2013; Steichen et al., 2016).

Immunogenicity of V2-SET Env Vaccines

Guinea pigs were immunized three times at monthly intervals (Figure 5D), intramuscularly in the quadriceps, with a total of 100 μ g Env (divided equally among immunogens in cocktails) formulated with CpG/Emulsigen adjuvant. Animals were bled 4 weeks after each vaccination, with peak immunogenicity at week 12. Binding responses were assessed by ELISA using the immunogens and a small panel of additional Envs expressed as gp140s, and V1/V2 gp70 scaffolds (Kayman et al., 1994) (Figures S5D–S5E). All vaccination regimens elicited comparable high magnitude binding responses with similar kinetics. Furthermore, all vaccines elicited tier-1 NAb responses against easy-to-neutralize viruses (Figure S5F).

Trivalent V2-SET vaccines augmented the magnitude and breadth of neutralization against 20 tier 2 pseudoviruses (the standard global panel of 12 [deCamp et al., 2014], plus 8 additional) compared with 459C WT alone, using either the gp140 SOSIP or gp140 foldon vaccine platforms (Figure 6). Using the gp140 foldon platform, the V2-SET Mixture (the 3 vaccine components co-delivered) and V2-SET Prime/Boost (V2 Opt prime and V2 Alt+459C WT boost) both improved the magnitude of tier 2 NABs responses compared to 459C WT alone ($p = 0.006$ and $p = 0.008$, respectively, using a non-parametric permutation test). A median of 85% of the heterologous tier 2 viruses were neutralized in the V2-SET groups, compared to 55% in the 459C WT alone group ($p = 0.004$, two-sided Wilcoxon rank sum) (Figure 6B). Using the SOSIP platform the V2-SET Mixture similarly improved neutralization potency compared to SOSIP 459C WT alone ($p = 0.008$), and a median of 65% of the heterol-

ogous tier 2 viruses tested were neutralized in the V2-SET group, compared to 30% for the 459C WT group ($p = 0.008$) (Figure 6A). In contrast, the BG505 SOSIP gp140 induced a median of 5% of the heterologous tier 2 viruses tested, although one animal had more breadth (Figure 6C).

We next explored the generalizability these findings. First, tier 2 NAb responses with the V2-SET vaccine were similarly enhanced compared to 459C WT using a different adjuvant, MPLA (Nkolola et al., 2014b) (Figures S6A and S6B). Second, while V2 Opt and Alt delivered alone were immunogenic, they did not enhance neutralization breadth like the trivalent combination (Figure S6C). Third, the tier 2 NAb responses were mediated by purified IgG (Figure S7D). Finally, other previously studied multivalent Env cocktails involving natural sequence immunogens, a trivalent clade C vaccine (Bricault et al., 2015) and a tetravalent multiclade vaccine (Bricault et al., 2018), did not significantly enhance tier 2 breadth over 459C WT (Figure S6C). Together, these data suggest that the improved NAB breadth induced with the trivalent V2-SET vaccine was generalizable, dependent on the SET bioinformatic design, and could not be achieved by simple mixtures of WT Env immunogens.

We also assessed post-vaccination antibody responses for binding to linear peptides spanning Env (Stephenson et al., 2015). Despite similar overall ELISA titers (Figure S5), V2-SET vaccines binding responses to linear V3 peptides were markedly lower than 459C WT, suggesting they were redirected away from non-neutralizing linear V3 epitopes (Figure 7A). To assess whether the improved V2-SET tier 2 NAb responses resulted from V2-specific conformational antibodies, we constructed pseudoviruses that resulted in the loss of the PNGS at position N160—a critical sensitivity signature for V2 bNABs. The N160-PNGS-loss mutation pseudoviruses showed increased neutralization sensitivity to 459C WT vaccine-elicited NABs, suggesting that a region partly shielded by the N160 glycan was targeted, but they also abrogated the enhancement in NAB potency achieved with V2-SET vaccine compared to the 459C WT vaccine (Figure 7B), suggesting that the improved the performance of the V2-SET vaccine over 459C WT depended on the PNGS at N160.

DISCUSSION

We defined HIV-1 Env bNAb signatures using neutralization data from four large virus panels to provide an unprecedented level of bNAb signature mapping—including hundreds of AA and PNGS signatures, as well as critical hypervariable region characteristics—for over 50 NABs. The accuracy of Env sequence-based machine learning predictions of IC_{50} titers were generally improved by focusing on relevant signatures and hypervariable

compared by a two-sided Wilcoxon test. BG505 and 459C WT SOSIP vaccines were comparable ($p = 0.13$), and V2-SET SOSIP vaccine responses were broader than either BG505 ($p = 0.02$) or 459C WT ($p = 0.01$). Responses elicited by the V2-SET foldon vaccine were broader than responses elicited by the 459C WT foldon vaccine ($p = 0.006$).

(E and F) Magnitudes of tier 2 NAb responses to SOSIP and foldon vaccine groups, respectively. Responses that were at least 10 above the background are considered positive and are shown; the dotted line at an ID_{50} titer of 100 is added for visual emphasis. Colors represent the vaccine groups (see key). Horizontal black lines are the median response to each pseudovirus. Response magnitudes were compared with a nonparametric permutation test (STAR Methods). For SOSIP vaccines, there was no statistical difference in potency between BG505 and 459C WT vaccine groups ($p = 0.06$), but V2-SET responses were more potent than both BG505 ($p = 0.007$) and 459C WT ($p = 0.008$). For gp140 foldon delivery, V2-SET responses were more potent than 459C WT ($p = 0.002$).

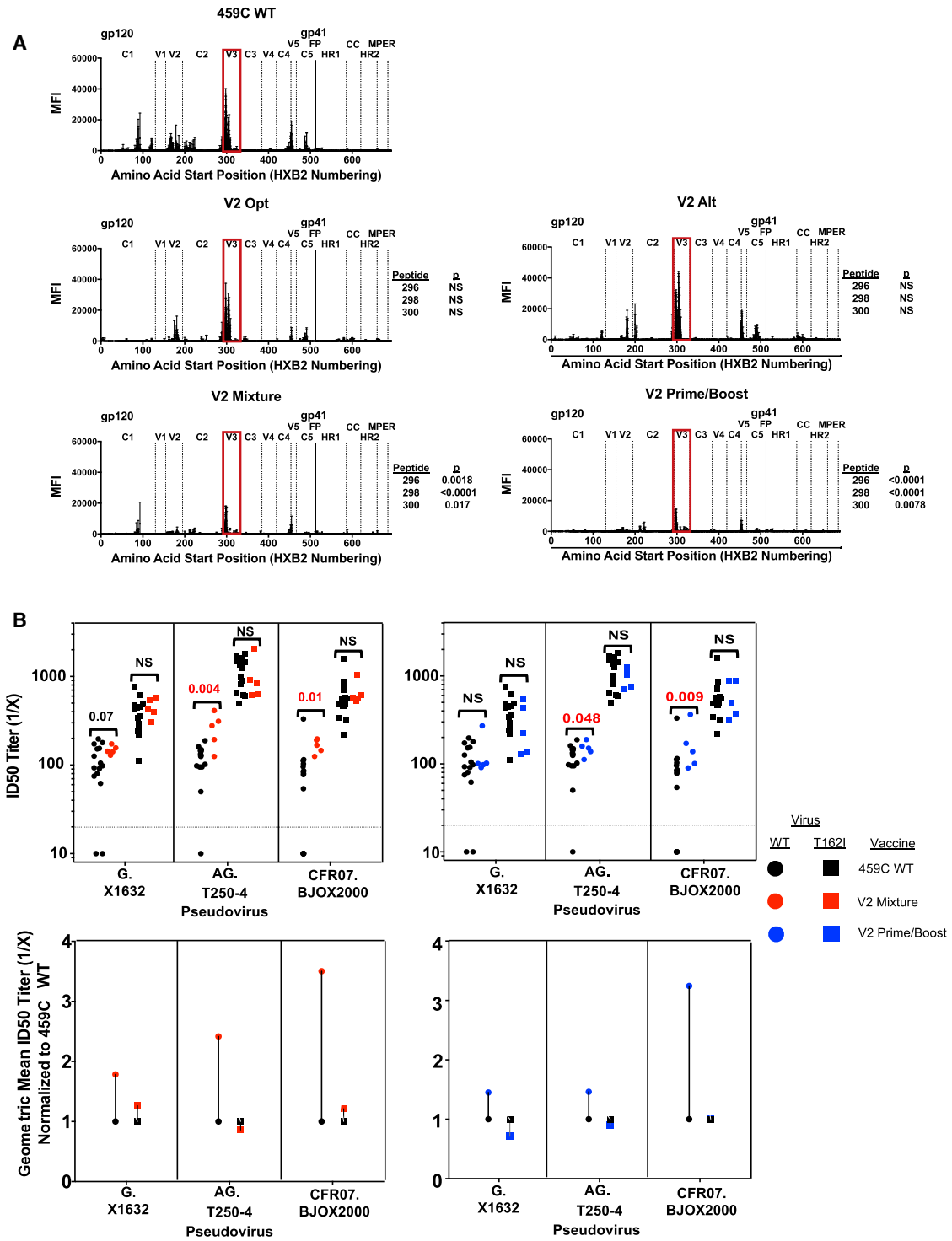


Figure 7. Mapping of Antibody Responses Elicited by V2-SET Vaccine

(A) Magnitude and position of binding antibody responses from guinea pig sera to linear 15-mer peptides on peptide microarrays from each gp140 foldon vaccine group. Each dot represents an average MFI (mean fluorescence intensity) per peptide that is positive for antibody binding within each vaccination group, standard deviation shown. Env regions are delineated by vertical lines, the V3 loop highlighted in red. Statistical differences for binding responses to peptides with starting positions in V3 as compared to 459C WT are shown. The p values are based on a Wilcoxon one-sided test; NS means not significant.

(B) Neutralizing titers against select pseudoviruses with a N160-dependent enhancement of V2-SET responses over 459C WT. ID₅₀ titers in guinea pigs vaccinated with 459C WT and V2-SET vaccines against the native pseudoviruses are shown by dots, and against the N160 glycan deletion mutant (T162I) (legend continued on next page)

region characteristics that were consistently highly ranked as key features for accurate predictions; thus supporting the consideration of these characteristics in vaccine design. Such machine-learning-based predictions of bNAb sensitivity levels across populations of Env sequences may ultimately be useful for modeling the relative sensitivity of a set of bNAbs across a regional epidemic targeted for treatment and for interpreting results of bNAb-based prevention and therapeutic clinical studies. Our data also highlight the importance of HIV-1 clades for both bNAb passive infusion studies and vaccine studies.

We also developed a signature-based approach to Env immunogen design using the V2 bNAb signature patterns to inform the design of a trivalent V2-SET vaccine. Induction of heterologous tier 2 NAbs has proven to be a major challenge for the HIV-1 vaccine field, and to date, the breadth of tier 2 NAbs induced by vaccines has been modest in both small and large animal models. The native-like BG505 SOSIP trimer induces potent autologous NAbs but with minimal tier 2 NAb breadth (Pauthner et al., 2017; Sanders et al., 2015). The trivalent V2-SET vaccine induced greater breadth of tier 2 NAb responses than the 459C SOSIP trimer alone, and both proved superior to the BG505 SOSIP trimer. The improved NAb breadth using V2-SET antigens was reproducible in guinea pig vaccination studies and generalizable using two common HIV-1 Env trimer platforms (SOSIP and foldon gp140s) and two adjuvants (CpG/Emulsigen and MPLA). Moreover, the enhanced NAb breadth was not observed with two cocktails of natural sequence HIV-1 Env immunogens. Although the magnitude of the tier 2 NAb responses remained low to moderate, these data demonstrate the proof-of-concept that bNAb signatures can contribute to the design of next-generation HIV-1 Env immunogens.

STAR★METHODS

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

- **KEY RESOURCES TABLE**
- **CONTACT FOR REAGENT AND RESOURCE SHARING**
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
 - Human Subjects
 - Cell Lines
 - Guinea Pig Vaccinations
- **METHOD DETAILS**
 - Experimental Methods for Vaccine Evaluation
 - Sequence and Signature Analysis
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
 - Sequence Analyses
 - Vaccine Immune Response Comparisons
- **DATA AND SOFTWARE AVAILABILITY**
 - GenSig
 - CATNAP Enhancement
 - FilteredForests
 - bNAb signature information access

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and seven tables and can be found with this article online at <https://doi.org/10.1016/j.chom.2018.12.001>.

ACKNOWLEDGMENTS

Supported by the Bill & Melinda Gates Foundation (OPP1032144, OPP1040741, OPP1032144, and OPP1169339), the NIH (AI096040, AI100645, A1095985, AI124377, AI126603, AI128751, and AI129797), and the Ragon Institute of MGH, MIT, and Harvard. We have no financial conflicts of interest. We thank Robert Bailer and Mark Louder, VRC, NIAID, for contributing large panel neutralization data. We thank Mark Connors, Lynn Morris, Penny Moore, Michel Nussenzweig, Michael Farzan, and Matthew Gardner for availability of antibodies and reagents; the HVTN Laboratory Program and Leonidas Stamatatos for envelope sequences; Nicholas Provine, Zi Han Kang, Alex Badamchi-Zadeh, Pablo Penaloza-MacMaster, Rafael Larocca, Sophia Rits-Vollach, Sandra Vertentes, and Helen DeCosta for generous assistance and advice; and Kelli Green and Hongmei Gao for their organizational efforts.

AUTHOR CONTRIBUTIONS

Conceptualization, B.K., C.A.B., D.H.B., and D.C.M.; Software, H.Y., J.T., B.K., M.T., K.W., and P.H.; Writing, C.A.B., K.Y., B.K., and D.H.B.; Review & Editing, J.T., G.H.L., B.H.H., K.W., J.P.M., E.F.K., M.S.S., D.C.M., B.F.H., and J.R.M.; Analysis, B.K., J.T., K.Y., E.E.G., K.W., M.T., P.H., C.A.B., D.H.B., and M.S.S.; Supervision, D.C.M., B.K., B.H.H., S.G., D.H.B., M.S.S., B.C., C.O., and K.E.S.; Bioinformatic Methodology, B.K., D.C.M., K.Y., and E.F.K.; Data Curation, J.P.M., K.Y., and G.H.L.; Immunological and Biochemical Assays, C.A.B., M.S.S., J.M.K., J.L.S., C.L.L., F.G., M.R., M.G.B., G.H.N., K.M., J.J.J., J.Z., C.O., H.P., C.C., J.P.N., K.E.S., and L.D.W.; Visualization, B.K., K.W., K.Y., and J.T.; Resources, N.D.-R., J.R.M., M.S.S., D.C.M., J.F.S., M.B., L.D.W., and B.F.H.; Funding Acquisition, D.C.M., D.H.B., B.F.H., and B.K.

DECLARATION OF INTERESTS

The authors declare no competing interests. A provisional patent application for the vaccine technology has been filed by C.A.B., K.Y., D.H.B., and B.K.

Received: April 14, 2018

Revised: July 6, 2018

Accepted: November 14, 2018

Published: January 9, 2019

REFERENCES

- Arlot, S., and Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statist. Surv.* 4, 40–79.
- Behrens, A.J., Vasiljevic, S., Pritchard, L.K., Harvey, D.J., Andev, R.S., Krumm, S.A., Struwe, W.B., Cupo, A., Kumar, A., Zitzmann, N., et al. (2016). Composition and antigenic effects of individual glycan sites of a trimeric HIV-1 envelope glycoprotein. *Cell Rep.* 14, 2695–2706.
- Bhattacharya, T., Daniels, M., Heckerman, D., Foley, B., Frahm, N., Kadie, C., Carlson, J., Yusim, K., McMahon, B., Gaschen, B., et al. (2007). Founder effects in the assessment of HIV polymorphisms and HLA allele associations. *Science* 315, 1583–1586.
- 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

pseudoviruses by squares. Colors represent the particular gp140 foldon vaccination regimen: black is 459C WT, red is V2-SET mixture, and blue is V2-SET prime/boost. The top plots show ID₅₀ titers for each guinea pig. The dotted line at 20 marks the limit of detection. The bottom plots show the geometric means of the NAb titers from the top plot (over the animals vaccinated by the same vaccine and tested on the same pseudovirus as in the top plot), normalized to 459C WT. The p values from Wilcoxon pairwise comparisons are shown in red. NS, not significant.

- broadly neutralizing antibodies and their inferred unmutated common ancestors. *J. Virol.* **85**, 9998–10009.
- Bonsignori, M., Kreider, E.F., Fera, D., Meyerhoff, R.R., Bradley, T., Wiehe, K., Alam, S.M., Aussedat, B., Walkowicz, W.E., Hwang, K.K., et al. (2017a). Staged induction of HIV-1 glycan-dependent broadly neutralizing antibodies. *Sci. Transl. Med.* **9**, <https://doi.org/10.1126/scitranslmed.aai7514>.
- Bonsignori, M., Liao, H.X., Gao, F., Williams, W.B., Alam, S.M., Montefiori, D.C., and Haynes, B.F. (2017b). Antibody-virus co-evolution in HIV infection: paths for HIV vaccine development. *Immunol. Rev.* **275**, 145–160.
- 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.
- Bradley, T., Trama, A., Tumba, N., Gray, E., Lu, X., Madani, N., Jahanbakhsh, F., Eaton, A., Xia, S.M., Parks, R., et al. (2016). Amino acid changes in the HIV-1 gp41 membrane proximal region control virus neutralization sensitivity. *EBioMedicine* **12**, 196–207.
- Breiman, L. (2001). Random forests. *Mach. Learn.* **45**, 5–32.
- Bricault, C.A., Kovacs, J.M., Badamchi-Zadeh, A., McKee, K., Shields, J.L., Gunn, B.M., Neubauer, G.H., Ghantous, F., Jennings, J., Gillis, L., et al. (2018). Neutralizing antibody responses following long-term vaccination with HIV-1 Env gp140 in guinea pigs. *J. Virol.* **92**, <https://doi.org/10.1128/JVI.00369-18>.
- Bricault, C.A., Kovacs, J.M., Nkolola, J.P., Yusim, K., Giorgi, E.E., Shields, J.L., Perry, J., Lavine, C.L., Cheung, A., Ellingson-Strouss, K., et al. (2015). A multivalent clade C HIV-1 Env trimer cocktail elicits a higher magnitude of neutralizing antibodies than any individual component. *J. Virol.* **89**, 2507–2519.
- 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. Retrovir.* **10**, 359–369.
- Burton, D.R., Barbas, C.F., 3rd, Persson, M.A., Koenig, S., Chanock, R.M., and Lerner, R.A. (1991). A large array of human monoclonal antibodies to type 1 human immunodeficiency virus from combinatorial libraries of asymptomatic seropositive individuals. *Proc. Natl. Acad. Sci. USA* **88**, 10134–10137.
- Burton, D.R., and Hangartner, L. (2016). Broadly neutralizing antibodies to HIV and their role in vaccine design. *Annu. Rev. Immunol.* **34**, 635–659.
- Burton, D.R., and Mascola, J.R. (2015). Antibody responses to envelope glycoproteins in HIV-1 infection. *Nat. Immunol.* **16**, 571–576.
- Cardoso, R.M., Zwick, M.B., Stanfield, R.L., Kunert, R., Binley, J.M., Katinger, H., Burton, D.R., and Wilson, I.A. (2005). Broadly neutralizing anti-HIV antibody 4E10 recognizes a helical conformation of a highly conserved fusion-associated motif in gp41. *Immunity* **22**, 163–173.
- Chuang, G.Y., Acharya, P., Schmidt, S.D., Yang, Y., Louder, M.K., Zhou, T., Kwon, Y.D., Pancera, M., Bailer, R.T., Doria-Rose, N.A., et al. (2013). Residue-level prediction of HIV-1 antibody epitopes based on neutralization of diverse viral strains. *J. Virol.* **87**, 10047–10058.
- Chuang, G.Y., Liou, D., Kwong, P.D., and Georgiev, I.S. (2014). NEP: web server for epitope prediction based on antibody neutralization of viral strains with diverse sequences. *Nucleic Acids Res.* **42**, W64–W71.
- Corti, D., Langedijk, J.P., Hinz, A., Seaman, M.S., Vanzetta, F., Fernandez-Rodriguez, B.M., Silacci, C., Pinna, D., Jarrossay, D., Balla-Jhaghoorsingh, 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**, e8805.
- Crispin, M., and Doores, K.J. (2015). Targeting host-derived glycans on enveloped viruses for antibody-based vaccine design. *Curr. Opin. Virol.* **11**, 63–69.
- Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190.
- Dabney, A., and Storey, J.D. (2013). Q-value estimation for false discovery rate control. R package version 2.14.0. <http://github.com/jdstorey/qvalue>.
- deCamp, A., Hraber, P., Bailer, R.T., Seaman, M.S., Ochsenbauer, C., Kappes, J., Gottardo, R., Edlefsen, P., Self, S., Tang, H., et al. (2014). Global panel of HIV-1 Env reference strains for standardized assessments of vaccine-elicited neutralizing antibodies. *J. Virol.* **88**, 2489–2507.
- Derdeyn, C.A., Decker, J.M., Bibollet-Ruche, F., Mokili, J.L., Muldoon, M., Denham, S.A., Heil, M.L., Kasolo, F., Musonda, R., and Hahn, B.H. (2004). Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. *Science* **303**, 2019–2022.
- Diskin, R., Scheid, J.F., Marcovecchio, P.M., West, A.P., Jr., Klein, F., Gao, H., Gnanapragasam, P.N., Abadir, A., Seaman, M.S., Nussenzweig, M.C., et al. (2011). Increasing the potency and breadth of an HIV antibody by using structure-based rational design. *Science* **334**, 1289–1293.
- Doria-Rose, N.A., Bhiman, J.N., Roark, R.S., Schramm, C.A., Gorman, J., Chuang, G.Y., Pancera, M., Cale, E.M., Erandes, 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., Georgiev, I., O'Dell, S., Chuang, G.Y., Staupe, R.P., McLellan, J.S., Gorman, J., Pancera, M., Bonsignori, M., Haynes, B.F., et al. (2012). A short segment of the HIV-1 gp120 V1/V2 region is a major determinant of resistance to V1/V2 neutralizing antibodies. *J. Virol.* **86**, 8319–8323.
- Doria-Rose, N.A., Schramm, C.A., Gorman, J., Moore, P.L., Bhiman, J.N., DeKosky, B.J., Erandes, 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.
- Escolano, A., Steichen, J.M., Dosenovic, P., Kulp, D.W., Golijanin, J., Sok, D., Freund, N.T., Gitlin, A.D., Oliveira, T., Araki, T., et al. (2016). Sequential immunization elicits broadly neutralizing anti-HIV-1 antibodies in Ig knockin mice. *Cell* **166**, 1445–1458.e12.
- Evans, M.C., Phung, P., Paquet, A.C., Parikh, A., Petropoulos, C.J., Wrin, T., and Haddad, M. (2014). Predicting HIV-1 broadly neutralizing antibody epitope networks using neutralization titers and a novel computational method. *BMC Bioinformatics* **15**, 77.
- Fera, D., Schmidt, A.G., Haynes, B.F., Gao, F., Liao, H.X., Kepler, T.B., and Harrison, S.C. (2014). Affinity maturation in an HIV broadly neutralizing B-cell lineage through reorientation of variable domains. *Proc. Natl. Acad. Sci. USA* **111**, 10275–10280.
- Ferguson, A.L., Falkowska, E., Walker, L.M., Seaman, M.S., Burton, D.R., and Chakraborty, A.K. (2013). Computational prediction of broadly neutralizing HIV-1 antibody epitopes from neutralization activity data. *PLoS One* **8**, e80562.
- Freeman, M.M., Seaman, M.S., Rits-Volloch, S., Hong, X., Kao, C.Y., Ho, D.D., and Chen, B. (2010). Crystal structure of HIV-1 primary receptor CD4 in complex with a potent antiviral antibody. *Structure* **18**, 1632–1641.
- Frey, G., Peng, H., Rits-Volloch, S., Morelli, M., Cheng, Y., and Chen, B. (2008). A fusion-intermediate state of HIV-1 gp41 targeted by broadly neutralizing antibodies. *Proc. Natl. Acad. Sci. USA* **105**, 3739–3744.
- Gao, F., Bonsignori, M., Liao, H.X., Kumar, A., Xia, S.M., Lu, X., Cai, F., Hwang, K.K., Song, H., Zhou, T., et al. (2014). Cooperation of B cell lineages in induction of HIV-1-broadly neutralizing antibodies. *Cell* **158**, 481–491.
- Garces, F., Lee, J.H., de Val, N., Torrents de la Pena, A.T., Kong, L., Puchades, C., Hua, Y., Stanfield, R.L., Burton, D.R., Moore, J.P., et al. (2015). Affinity maturation of a potent family of HIV antibodies is primarily focused on accommodating or avoiding glycans. *Immunity* **43**, 1053–1063.
- Geurts, P., Ernst, D., and Wehenkel, L. (2006). Extremely randomized trees. *Mach. Learn.* **63**, 3–42.
- Gnanakaran, S., Daniels, M.G., Bhattacharya, T., Lapedes, A.S., Sethi, A., Li, M., Tang, H., Greene, K., Gao, H., Haynes, B.F., et al. (2010). Genetic signatures in the envelope glycoproteins of HIV-1 that associate with broadly neutralizing antibodies. *PLoS Comp. Biol.* **6**, e1000955.
- Gorman, J., Soto, C., Yang, M.M., Davenport, T.M., Guttman, M., Bailer, R.T., Chambers, M., Chuang, G.Y., DeKosky, B.J., Doria-Rose, N.A., et al. (2016). Structures of HIV-1 Env V1V2 with broadly neutralizing antibodies reveal commonalities that enable vaccine design. *Nat. Struct. Mol. Biol.* **23**, 81–90.
- 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.

- Hepler, N.L., Scheffler, K., Weaver, S., Murrell, B., Richman, D.D., Burton, D.R., Poignard, P., Smith, D.M., and Kosakovsky Pond, S.L. (2014). IDEPI: rapid prediction of HIV-1 antibody epitopes and other phenotypic features from sequence data using a flexible machine learning platform. *PLoS Comput. Biol.* *10*, e1003842.
- Hraber, P., Korber, B.T., Lapedes, A.S., Bailer, R.T., Seaman, M.S., Gao, H., Greene, K.M., McCutchan, F., Williamson, C., Kim, J.H., et al. (2014a). Impact of clade, geography, and age of the epidemic on HIV-1 neutralization by antibodies. *J. Virol.* *88*, 12623–12643.
- Hraber, P., Seaman, M.S., Bailer, R.T., Mascola, J.R., Montefiori, D.C., and Korber, B.T. (2014b). Prevalence of broadly neutralizing antibody responses during chronic HIV-1 infection. *AIDS* *28*, 163–169.
- 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.
- 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.
- Julien, J.P., Sok, D., Khayat, R., Lee, J.H., Doores, K.J., Walker, L.M., Ramos, A., Diwanji, D.C., Pejchal, R., Cupo, A., et al. (2013). Broadly neutralizing antibody PGT121 allosterically modulates CD4 binding via recognition of the HIV-1 gp120 V3 base and multiple surrounding glycans. *PLoS Pathog.* *9*, e1003342.
- Kayman, S.C., Wu, Z., Revesz, K., Chen, H., Kopelman, R., and Pinter, A. (1994). Presentation of native epitopes in the V1/V2 and V3 regions of human immunodeficiency virus type 1 gp120 by fusion glycoproteins containing isolated gp120 domains. *J. Virol.* *68*, 400–410.
- Klein, F., Diskin, R., Scheid, J.F., Gaebler, C., Mouquet, H., Georgiev, I.S., Pancera, M., Zhou, T., Incesu, R.B., Fu, B.Z., et al. (2013). Somatic mutations of the immunoglobulin framework are generally required for broad and potent HIV-1 neutralization. *Cell* *153*, 126–138.
- Klein, F., Gaebler, C., Mouquet, H., Sather, D.N., Lehmann, C., Scheid, J.F., Kraft, Z., Liu, Y., Pietzsch, J., Hurley, A., et al. (2012). Broad neutralization by a combination of antibodies recognizing the CD4 binding site and a new conformational epitope on the HIV-1 envelope protein. *J. Exp. Med.* *209*, 1469–1479.
- Kong, L., Lee, J.H., Doores, K.J., Murin, C.D., Julien, J.P., McBride, R., Liu, Y., Marozsan, A., Cupo, A., Klasse, P.J., et al. (2013). Supersite of immune vulnerability on the glycosylated face of HIV-1 envelope glycoprotein gp120. *Nat. Struct. Mol. Biol.* *20*, 796–803.
- Korber, B., Hraber, P., Wagh, K., and Hahn, B.H. (2017). Polyvalent vaccine approaches to combat HIV-1 diversity. *Immunol. Rev.* *275*, 230–244.
- Korber, B., Muldoon, M., Theiler, J., Gao, F., Gupta, R., Lapedes, A., Hahn, B.H., Wolinsky, S., and Bhattacharya, T. (2000). Timing the ancestor of the HIV-1 pandemic strains. *Science* *288*, 1789–1796.
- Korber, B.T., Letvin, N.L., and Haynes, B.F. (2009). T-cell vaccine strategies for human immunodeficiency virus, the virus with a thousand faces. *J. Virol.* *83*, 8300–8314.
- Kwong, P.D., Wyatt, R., Robinson, J., Sweet, R.W., Sodroski, J., and Hendrickson, W.A. (1998). Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* *393*, 648–659.
- Li, M., Gao, F., Mascola, J.R., Stamatatos, L., Polonis, V.R., Koutsoukos, M., Voss, G., Goepfert, P., Gilbert, P., Greene, K.M., et al. (2005). Human immunodeficiency virus type 1 env clones from acute and early subtype B infections for standardized assessments of vaccine-elicited neutralizing antibodies. *J. Virol.* *79*, 10108–10125.
- Liao, H.X., Bonsignori, M., Alam, S.M., McLellan, J.S., Tomaras, G.D., Moody, M.A., Kozink, D.M., Hwang, K.K., Chen, X., Tsao, C.Y., et al. (2013). Vaccine induction of antibodies against a structurally heterogeneous site of immune pressure within HIV-1 envelope protein variable regions 1 and 2. *Immunity* *38*, 176–186.
- Lynch, R.M., Wong, P., Tran, L., O'Dell, S., Nason, M.C., Li, Y., Wu, X., and Mascola, J.R. (2015). HIV-1 fitness cost associated with escape from the VRC01 class of CD4 binding site neutralizing antibodies. *J. Virol.* *89*, 4201–4213.
- McLellan, J.S., Pancera, M., Carrico, C., Gorman, J., Julien, J.P., Khayat, R., Louder, R., Pejchal, R., Sastry, M., Dai, K., et al. (2011). Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. *Nature* *480*, 336–343.
- McLeod, A.I. (2011). Kendall: Kendall Rank Correlation and Mann-Kendall Trend Test. <https://cran.r-project.org/web/packages/Kendall/Kendall.pdf>.
- Mouquet, H., Scharf, L., Euler, Z., Liu, Y., Eden, C., Scheid, J.F., Halper-Stromberg, A., Gnanapragasam, P.N., Spencer, D.I., Seaman, M.S., et al. (2012). Complex-type N-glycan recognition by potent broadly neutralizing HIV antibodies. *Proc. Natl. Acad. Sci. USA* *109*, E3268–E3277.
- Nelson, J.D., Brunel, F.M., Jensen, R., Crooks, E.T., Cardoso, R.M., Wang, M., Hessel, A., Wilson, I.A., Binley, J.M., Dawson, P.E., et al. (2007). An affinity-enhanced neutralizing antibody against the membrane-proximal external region of human immunodeficiency virus type 1 gp41 recognizes an epitope between those of 2F5 and 4E10. *J. Virol.* *81*, 4033–4043.
- Nickle, D.C., Heath, L., Jensen, M.A., Gilbert, P.B., Mullins, J.I., and Kosakovsky Pond, S.L. (2007). HIV-specific probabilistic models of protein evolution. *PLoS One* *2*, e503.
- Nkolola, J.P., Bricault, C.A., Cheung, A., Shields, J., Perry, J., Kovacs, J.M., Giorgi, E., van Winsen, M., Apetri, A., Brinkman-van der Linden, E.C.M., et al. (2014a). Characterization and immunogenicity of a novel mosaic M HIV-1 gp140 trimer. *J. Virol.* *88*, 9538–9552.
- Nkolola, J.P., Cheung, A., Perry, J.R., Carter, D., Reed, S., Schuitemaker, H., Pau, M.G., Seaman, M.S., Chen, B., and Barouch, D.H. (2014b). Comparison of multiple adjuvants on the stability and immunogenicity of a clade C HIV-1 gp140 trimer. *Vaccine* *32*, 2109–2116.
- Nkolola, J.P., Peng, H., Settembre, E.C., Freeman, M., Grandpre, L.E., Devoy, C., Lynch, D.M., La Porte, A., Simmons, N.L., Bradley, R., et al. (2010). Breadth of neutralizing antibodies elicited by stable, homogeneous clade A and clade C HIV-1 gp140 envelope trimers in guinea pigs. *J. Virol.* *84*, 3270–3279.
- Ofek, G., Tang, M., Sambor, A., Katinger, H., Mascola, J.R., Wyatt, R., and Kwong, P.D. (2004). Structure and mechanistic analysis of the anti-human immunodeficiency virus type 1 antibody 2F5 in complex with its gp41 epitope. *J. Virol.* *78*, 10724–10737.
- Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* *20*, 289–290.
- Parrish, N.F., Gao, F., Li, H., Giorgi, E.E., Barbian, H.J., Parrish, E.H., Zajic, L., Iyer, S.S., Decker, J.M., Kumar, A., et al. (2013). Phenotypic properties of transmitted founder HIV-1. *Proc. Natl. Acad. Sci. USA* *110*, 6626–6633.
- Pauthner, M., Havenar-Daughton, C., Sok, D., Nkolola, J.P., Bastidas, R., Boopathy, A.V., Carnathan, D.G., Chandrashekar, A., Cirelli, K.M., Cottrell, C.A., et al. (2017). Elicitation of robust tier 2 neutralizing antibody responses in nonhuman primates by HIV envelope trimer immunization using optimized approaches. *Immunity* *46*, 1073–1088.e6.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* *12*, 2825–2830.
- Pejchal, R., Doores, K.J., Walker, L.M., Khayat, R., Huang, P.-S., Wang, S.-K., Stanfield, R.L., Julien, J.-P., Ramos, A., Crispin, M., et al. (2011). A potent and broad neutralizing antibody recognizes and penetrates the HIV glycan shield. *Science* *334*, 1097–1103.
- Peng, H., Long, F., and Ding, C. (2005). Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* *27*, 1226–1238.
- 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.
- Rudicell, R.S., Kwon, Y.D., 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.
- Sanders, R.W., Derking, R., Cupo, A., Julien, J.P., Yasmeen, A., de Val, N., Kim, H.J., Blattner, C., de la Pena, A.T., Korzun, J., et al. (2013). A next-generation cleaved, soluble HIV-1 Env trimer, BG505 SOSIP. 664 gp140, expresses multiple epitopes for broadly neutralizing but not non-neutralizing antibodies. *PLoS Pathog.* **9**, e1003618.
- Sanders, R.W., van Gils, M.J., Derking, R., Sok, D., Ketas, T.J., Burger, J.A., Ozorowski, G., Cupo, A., Simonich, C., Goo, L., et al. (2015). HIV-1 vaccines. HIV-1 neutralizing antibodies induced by native-like envelope trimers. *Science* **349**, aac4223.
- Sarzotti-Kelsoe, M., Bailer, R.T., Turk, E., Lin, C.L., 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.
- Saunders, K.O., Verkoczy, L.K., Jiang, C., Zhang, J., Parks, R., Chen, H., Housman, M., Bouton-Verville, H., Shen, X., Trama, A.M., et al. (2017). Vaccine induction of heterologous Tier 2 HIV-1 neutralizing antibodies in animal models. *Cell Rep.* **21**, 3681–3690.
- Scheid, J.F., Horwitz, J.A., Bar-On, Y., Kreider, E.F., Lu, C.L., Lorenzi, J.C., Feldmann, A., Braunschweig, M., Nogueira, L., Oliveira, T., et al. (2016). HIV-1 antibody 3BNC117 suppresses viral rebound in humans during treatment interruption. *Nature* **535**, 556–560.
- 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.
- Sok, D., Pauthner, M., Briney, B., Lee, J.H., Saye-Francisco, K.L., Hsueh, J., Ramos, A., Le, K.M., Jones, M., Jardine, J.G., et al. (2016). A prominent site of antibody vulnerability on HIV envelope incorporates a motif associated with CCR5 binding and its camouflaging glycans. *Immunity* **45**, 31–45.
- 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. USA* **111**, 17624–17629.
- Steichen, J.M., Kulp, D.W., Tokatlian, T., Escolano, A., Dosenovic, P., Stanfield, R.L., McCoy, L.E., Ozorowski, G., Hu, X., Kalyuzhnyi, O., et al. (2016). HIV vaccine design to target germline precursors of glycan-dependent broadly neutralizing antibodies. *Immunity* **45**, 483–496.
- Stephenson, K.E., Neubauer, G.H., Reimer, U., Pawlowski, N., Knaute, T., Zerweck, J., Korber, B.T., and Barouch, D.H. (2015). Quantification of the epitope diversity of HIV-1-specific binding antibodies by peptide microarrays for global HIV-1 vaccine development. *J. Immunol. Methods* **416**, 105–123.
- Stewart-Jones, G.B., Soto, C., Lemmin, T., Chuang, G.Y., Druz, A., Kong, R., Thomas, P.V., Wagh, K., Zhou, T., Behrens, A.J., et al. (2016). Trimeric HIV-1 Env structures define glycan shields from clades A, B, and G. *Cell* **165**, 813–826.
- Storey, J.D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* **100**, 9440–9445.
- Trkola, A., Purtscher, M., Muster, T., Ballaun, C., Buchacher, A., Sullivan, N., Srinivasan, K., Sodroski, J., Moore, J.P., and Katinger, H. (1996). Human monoclonal antibody 2G12 defines a distinctive neutralization epitope on the gp120 glycoprotein of human immunodeficiency virus type 1. *J. Virol.* **70**, 1100–1108.
- Wagh, K., Bhattacharya, T., Williamson, C., Robles, A., Bayne, M., Garrity, J., Rist, M., Rademeyer, C., Yoon, H., Lapedes, A., et al. (2016). Optimal combinations of broadly neutralizing antibodies for prevention and treatment of HIV-1 Clade C infection. *PLoS Pathog.* **12**, e1005520.
- 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.
- 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.
- West, A.P., Jr., Scharf, L., Horwitz, J., Klein, F., Nussenzweig, M.C., and Bjorkman, P.J. (2013). Computational analysis of anti-HIV-1 antibody neutralization panel data to identify potential functional epitope residues. *Proc. Natl. Acad. Sci. USA* **110**, 10598–10603.
- Williams, L.D., Ofek, G., Schätzle, S., McDaniel, J.R., Lu, X., Nicely, N.I., Wu, L., Loughheed, C.S., Bradley, T., Louder, M.K., et al. (2017). Potent and broad HIV neutralizing antibodies in memory B cells and plasma. *Sci. Immunol.* **2**, <https://doi.org/10.1126/sciimmunol.aal2200>.
- 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., Zhang, Z., Schramm, C.A., Joyce, M.G., Kwon, Y.D., Zhou, T., Sheng, Z., Zhang, B., O'Dell, S., McKee, K., et al. (2015). Maturation and diversity of the VRC01-antibody lineage over 15 years of chronic HIV-1 infection. *Cell* **161**, 470–485.
- 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.
- Yoon, H., Macke, J., West, A.P., Jr., 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.
- Zhou, T., Xu, L., Dey, B., Hessel, A.J., Van Ryk, D., Xiang, S.H., Yang, X., Zhang, M.Y., Zwick, M.B., Arthos, J., et al. (2007). Structural definition of a conserved neutralization epitope on HIV-1 gp120. *Nature* **445**, 732–737.
- Zhou, T., Georgiev, I., Wu, X., Yang, Z.Y., Dai, K., Finzi, A., Kwon, Y.D., 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**, 811–817.
- 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.
- 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.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
10-1074	Laboratory of Michel Nussenzweig, Rockefeller University	RRID: AB_2491062; Mouquet et al. (2012)
PG16	Polymun	RRID: AB_2491031; Cat#AB016
PGT145	Catalent	RRID: AB_2491054
PGDM1400	Catalent	Sok et al. (2014)
HRP-conjugated goat anti-guinea pig secondary antibody	Jackson ImmunoResearch Laboratories	Cat#106-035-003
Alexa Fluor 647-conjugated AffiniPure Goat Anti-Guinea Pig IgG (H+L)	Jackson ImmunoResearch Laboratories	Cat#706-605-148
Bacterial and Virus Strains		
X1632 T162l pseudovirus	Laboratory of Christina Ochsenbauer, University of Alabama Birmingham	N/A
T250-4 T162l pseudovirus	Laboratory of Christina Ochsenbauer, University of Alabama Birmingham	N/A
BJOX2000 T162l pseudovirus	Laboratory of Christina Ochsenbauer, University of Alabama Birmingham	N/A
Chemicals, Peptides, and Recombinant Proteins		
HIV-1 gp70 V1/V2 (ConC)	Immune Technology Corp	Cat#IT-001-213p
HIV-1 gp70 V1/V2 (Case A2)	Immune Technology Corp	Cat#IT-001-214p
HIV-1 gp70 V1/V2 (CN54)	Immune Technology Corp	Cat#IT-001-211p
HIV-1 gp70 V1/V2 (A244)	Immune Technology Corp	Cat#IT-001-212p
HBS-EP	GE Healthcare	Cat#BR100188
Soluble CD4	Laboratory of Bing Chen, Children's Hospital Boston	Freeman et al. (2010)
Amine Coupling Kit	GE Healthcare	Cat#BR100050
Pierce Recombinant Protein A	ThermoScientific	Cat#21184
Emulsigen	MVP Adjuvants	N/A
Monophosphoryl Lipid A from S. minnesota R595	InvivoGen	Cat#vac-mpla
SuperBlock T20 (TBS) Blocking Buffer	Thermo Scientific	Cat#37536
High-Capacity Protein A Agarose	Thermo Scientific	Cat# 89948
459C WT gp140 foldon	Bricault et al. (2015)	N/A
459C V2 Opt gp140 foldon	This paper	N/A
459C V2 Alt gp140 foldon	This paper	N/A
459C WT SOSIP	This paper	N/A
459C V2 Opt SOSIP	This paper	N/A
C97ZA012 gp140 foldon	Nkolola et al., (2010)	N/A
405C gp140 foldon	Bricault et al., (2015)	N/A
92UG037 gp140 foldon	Nkolola et al., (2010)	N/A
PVO.4 gp140 foldon	Li et al., (2005)	Accession number: AY835444
Mosaic gp140 foldon	Nkolola et al., (2014a)	N/A

(Continued on next page)

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Neutralizing antibody signatures will be deposited in the Los Alamos HIV database and accessible through the Genome Browser, the Neutralizing Antibody relational database, and the Env annotation tables.	Los Alamos HIV Database	https://www.hiv.lanl.gov/content/immunology/neutralizing_ab_resources.html , www.hiv.lanl.gov/components/sequence/HIV/featuredb/search/env_ab_search_pub.comp , www.hiv.lanl.gov/content/sequence/genome_browser/browser.html
Experimental Models: Cell Lines		
Human: 293T	ATCC	ATCC CRL-3216
Experimental Models: Organisms/Strains		
Hartley guinea pigs: Outbred	Elm Hill Labs	N/A
Oligonucleotides		
CpG: 5'-TCGTCGTTGTCGTTTTGTCGTT-3'	Midland Reagent Company	N/A
Recombinant DNA		
GeneArt	Life Technologies	Cat#817003DE
Software and Algorithms		
Softmax Pro-4.7.1	Molecular Devices	https://www.moleculardevices.com/systems/microplate-readers/softmax-pro-7-software
GenePix Pro 7 software	Molecular Devices	https://www.moleculardevices.com/en/asset/br/data-sheets/genepix-pro-software-datasheet-v7-rev-b
GenePix Array List	Stephenson et al., (2015)	N/A
GenSig, a signature analysis web interface	Los Alamos HIV Database	https://www.hiv.lanl.gov/content/sequence/GENETICSIGNATURES/gs.html
CATNAP, Neutralization data resource	Los Alamos HIV Database	https://www.hiv.lanl.gov/components/sequence/HIV/neutralization/
Filtered Forests	Los Alamos HIV Database, github	https://www.hiv.lanl.gov/content/sequence/FLTFORESTS/fltforests.html , https://github.com/hivdb-lanl/FilteredForests
Other		
RepliTope Antigen Collection HIV Ultra slides	JPT Peptide Technologies GmbH	Cat#RT-HD_HIV
CM5 Chips	GE Healthcare	Cat#BR100012

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Dan Barouch (dbarouch@bidmc.harvard.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human Subjects

The HIV-1 bNAbs used in this study were all isolated in the context of other studies (Table S1). The Env pseudoviruses are all part of widely used standard panels. Human specimens used to derive these reagents are de-identified and considered exempt by the Duke University IRB, and the exemption approved by the Los Alamos National Lab IRB.

Cell Lines

Human endothelial kidney 293T cells (ATCC) were used for transient transfection of HIV-1 Env expressing plasmids and stably transfected human endothelial kidney 293T cells (Codex Biosolutions) were utilized for the production of HIV-1 Env gp140 and SOSIP immunogens.

Guinea Pig Vaccinations

Healthy, outbred, research-naïve, female Hartley guinea pigs (bred at and purchased from Elm Hill) at between 350 and 500 grams and about 1 to 2 months of age were used for vaccination studies and housed at the Animal Research Facility of Beth Israel Deaconess Medical Center under approved Institutional Animal Care and Use Committee (IACUC) protocols. Animals were co-housed 2 to 5 animals per cage, based on animal weight. All animals were naïve at the initiation of the study. Guinea pigs (5-15/group) were immunized with Env gp140 immunogens intramuscularly in the quadriceps bilaterally at 4-week intervals (weeks 0, 4, 8) for a total of 3 injections. Vaccine formulations for each guinea pig consisted of a total of 100 μ g of immunogen per injection formulated in 15% Emulsigen (vol/vol) oil-in-water emulsion (MVP Laboratories) and 50 μ g CpG (Midland Reagent Company) or 10 μ g Monophosphoryl lipid A (MPLA) (InvivoGen) as adjuvants. We also tested the V2-SET immunogen sequences in the context of the gp140 MD39 SOSIP constructs (Steichen et al., 2016), using a lengthened schedule of vaccinations at weeks 0, 8, and 24. Serum samples were obtained from the vena cava of anesthetized animals four weeks after each immunization as well as prior to vaccination for week 0, naïve sera.

METHOD DETAILS

Experimental Methods for Vaccine Evaluation

Plasmids, Cell Lines, Protein Production, and Antibodies

Our baseline immunogen was the C clade Env 459C, initially selected because it elicited tier 1B NAb responses (Bricault et al., 2015), and subsequently found to induce low levels of select tier 2 NABs upon evaluation of larger tier 2 pseudovirus panels. The codon-optimized synthetic genes of the V2-SET HIV-1 Env gp140 foldon (gp140) and gp140 MD39 SOSIP (SOSIP) immunogens were produced by GeneArt (Life Technologies). All gp140 constructs contained a consensus leader signal sequence peptide, as well as a C-terminal foldon trimerization tag followed by a His-tag as described previously (Frey et al., 2008; Nkolola et al., 2010). Large scale production of HIV-1 Env gp140 foldon and SOSIPs were produced as described previously (Nkolola et al., 2010, 2014a; Steichen et al., 2016). Of note, the gp140 SOSIP immunogens were cleaved by furin and the gp140 foldon immunogens were not. Soluble two-domain CD4 was produced as described previously (Freeman et al., 2010). 10-1074 was generously provided by Michel Nussenzweig (Rockefeller University, New York, NY). PG16 was purchased from Polymun Scientific, PGT145 and PGDM1400 from Catalent, gp70 V1/V2 HIV-1 envelope scaffolds including ConC, Case A2, CN54, and A244 V1/V2 from Immune Technology Corp.

Surface Plasmon Resonance Binding Analysis

SPR experiments were conducted on a Biacore 3000 (GE Healthcare) at 25°C utilizing HBS-EP [10 mM Hepes (pH 7.4), 150 mM NaCl, 3 mM EDTA, 0.005% P20] (GE Healthcare) as the running buffer. Immobilization of CD4 (~1,000 response units (RU)) or protein A (ThermoScientific) to CM5 chips was performed following the standard amine coupling procedure as recommended by the manufacturer (GE Healthcare). Select protein-protein interactions were analyzed using single-cycle kinetics consisting of four cycles of a 1-min association phase and a 4-min dissociation phase without regeneration between injections, followed by an additional cycle of a 1-min association phase and a 15-min dissociation phase, at a flow rate of 50 μ L/min. Immobilized IgGs were captured at about 500 RU for 10-1074 and about 3,000 RU for PG16. Soluble gp140 foldon was then passed over the surface at increasing concentrations from 62.5 nM-1,000 nM. Regeneration was conducted with 35 mM NaOH, 1.3 M NaCl (pH 12) at 100 μ L/min followed by 5-min equilibration in the HBS-EP buffer. For experiments run with PGDM1400 and PGT145 IgG, immobilized PGDM1400 and PGT145 IgGs were captured at between 150-200 RU. Binding experiments were conducted with a flow rate of 50 μ L/min with a 2-minute associate phase and a 5-minute dissociation phase. Soluble gp140 foldon or gp140 SOSIP were then passed over the surface at increasing concentrations from 31.25 nM-500 nM. Regeneration was conducted with one injection (3 seconds) of 35 mM sodium hydroxide, 1.3 M sodium chloride at 100 μ L/min followed by a 3-minute equilibration phase in HBS-EP. Identical injections over blank surfaces were subtracted from the binding data for analysis. All samples were run in duplicate and yielded similar kinetic results. Single curves of the duplicates are shown in all figures.

Endpoint ELISAs

Serum binding antibodies against gp140 foldon and V1/V2 scaffolds were measured by endpoint enzyme-linked immunosorbent assays (ELISAs) as described previously (Nkolola et al., 2010). Briefly, ELISA plates (Thermo Scientific) were coated with individual gp140s or V1/V2 scaffolds (Immune Technology) and incubated overnight. Guinea pig sera were then added in serial dilutions and later detected with an HRP-conjugated goat anti-guinea pig secondary antibody (Jackson ImmunoResearch Laboratories). Plates were developed and read using the Spectramax Plus ELISA plate reader (Molecular Devices) and Softmax Pro-4.7.1 software. End-point titers were considered positive at the highest dilution that maintained an absorbance >2-fold above background values.

Peptide Microarrays

RepliTope Antigen Collection HIV Ultra slides (JPT Peptide Technologies GmbH) arrays were generated, conducted, and analyzed using methods as described previously (Stephenson et al., 2015). These slides contain linear 15-mer peptides designed utilizing the HIV-1 global sequence database to provide coverage of HIV-1 global sequences (Stephenson et al., 2015). Briefly, microarray slides were incubated with guinea pig sera diluted 1/200 in SuperBlock T20 (TBS) Blocking Buffer (Thermo Scientific). Binding antibody responses were detected with Alexa Fluor 647-conjugated AffiniPure Goat Anti-Guinea Pig IgG (H+L) (Jackson ImmunoResearch Laboratories). All batches of slides were run in parallel with a control slide incubated with the secondary antibody only for background subtraction. Slides were scanned with a GenePix 4300A scanner (Molecular Devices) and analyzed with GenePix Pro 7 software and

GenePix Array List (Stephenson et al., 2015). The threshold values for positivity were calculated as the point at which the chance that signal is noise as low as possible ($P < 10^{-16}$). The peak positive antibody binding responses to linear V3 Env peptides were further analyzed comparing the 459C WT and the V2-SET vaccines. Peptides with the highest magnitude binding responses were analyzed comparing geometric means over animals separately against each 15-mer peptide start position. Geometric means were calculated for each vaccination group resulting in a single point per vaccine per peptide sequence.

TZM.bl Neutralization Assay for Vaccine Sera

All IC_{50} data for the large neutralization panels were obtained using the validated luciferase-based TZM.bl assay (Sarzotti-Kelsoe et al., 2014); most antibodies to a maximum concentration of 50 $\mu\text{g/ml}$. For vaccine responses, 20 tier 2 pseudoviruses were used in the TZM.bl neutralization assay: the standardized global panel of 1 HIV-1 reference strains independently selected to be representative of larger global panels (deCamp et al., 2014) and a panel of 8 additional tier 2 pseudoviruses selected because they were relatively sensitive to human sera (falling in the top quartile of geometric mean serological reactivity of the tier 2 panel), were sensitive to the V2 bNAb monoclonals (Yoon et al., 2015), and were relatively close in sequence to the V2-SET vaccines in the neutralization signature positions. The 8 additional pseudoviruses were added as an *a priori* attempt to increase the chances of getting a positive signal, but when tested were found to be very comparable in sensitivity to the global panel. The rationally selected tier 2 pseudoviruses included clade C (Du156.12, CT349_39_16, 234_F1_15_57, CNE58, and CA240_A5.5), CRF 02_AG (T250_4), CRF 07_BC (CNE20), and CRF 01_AE (C3347_C11) strains. For purification of guinea pig polyclonal IgG from sera, High-Capacity Protein A Agarose (Thermo Scientific) was utilized following manufacturer's instructions. After purification by protein A, polyclonal IgG samples were buffer exchanged into 1X phosphate buffered saline, pH 7.4 (Gibco) utilizing a EMD Millipore Amicon Ultra-15 Centrifugal Filter Unit (Millipore) at 4°C. Mutant pseudoviruses were generated with point mutations in V2/glycans to map NAb responses targeting these epitopes. Point mutations aiming to abrogate V2 antibody neutralization were selected to minimize disruptions in the virus backbone by representing mutations that occur commonly in nature. A T162I mutation was introduced into X1632, T250-4, and BJOX2000 to knock out the glycan at position N160; this mutations is relatively common among natural isolates.

Sequence and Signature Analysis

Signature analysis

To systematically identify sites of interest, we used our phylogenetically corrected approach (Gnanakaran et al., 2010) to minimize false positives due to lineage effects (Bhattacharya et al., 2007), and q-values to constrain false positives due to multiple testing (Storey and Tibshirani, 2003). These sites are summarized by antibody in Tables S3A–S3D, and population frequencies of signatures are displayed as LOGOs in Figure 3. Statistical details for all phylogenetically corrected signatures that met the statistical cutoff ($q \leq 0.2$) are provided in a summary table organized by antibody (Tables S3E–S3H). To be more inclusive, within each bNAb class, for sites that have either have a phylogenetically corrected signature for any bNAb within that class or are in the epitope binding region for a crystalized representative of the class, we also list in Tables S3I–S3L all AA and PNGS associations with a q-value < 0.2 , without the constraint of a phylogenetic correction, organized by Env position (Tables S3I–S3L). For comparison, published bNAb signatures from previous studies (Chuang et al., 2014; Ferguson et al., 2013; West et al., 2013) are also included in Tables S3E–S3L). This comparison shows that our analysis provides more detailed mapping of sites that may be relevant to the overall bNAb sensitivity than has been previously assembled. Because the bNAb field is advancing rapidly and new data are continuously accruing, we have also integrated our signature code into the CATNAP bioinformatics tool into the Los Alamos HIV Immunology Database (Yoon et al., 2015), allowing signature analysis to be conducted on-the-fly as new bNAb data is entered into the database.

Phylogenetically corrected signature methods were described in detail in earlier publications (Bhattacharya et al., 2007; Gnanakaran et al., 2010). Briefly, for a simple uncorrected test, a 2 x 2 contingency table is generated where the data is divided about a phenotypic cutoff (e.g. “high” or “low” IC_{50} values split about the median) and whether or not a sequence has a given amino acid at a given position, and a Fisher's exact test is used to assess statistical significance of each such contingency table. All amino acids are tested in all positions, and a false discovery rate (FDR) adjusted q-value (Storey and Tibshirani, 2003) with a threshold of < 0.2 used to define sites of interest, to be inclusive but still limit false negatives. This simple test can also be used to test associations with PNGS.

Even with FDR, without a phylogenetic correction simple signatures can yield an extreme over-abundance of apparent results, and many associations will not be causative, but can be carried along by genetic linkage to a site where the variation has direct consequences. An example illustrating how can happen is provided in Figure S7. In this example, the CRF01 clade is highly resistant to V3 bNAbs, and this is likely to be primarily driven the loss of the critical PNGS at N332 throughout the entire clade. But given the lack of reactivity for V3 bNAbs among CRF01 sequences, any amino acid highly enriched among CRF01 sequences will be associated with V3 bNAb resistance, yet most are likely to not be causative. An association that is still statistically supported after a phylogenetic correction, which requires that the correlation between the amino acid and the phenotype recur in sequences in dispersed locations throughout the tree, is more likely to reveal direct causative associations with the phenotype.

For a phylogenetically corrected test, a maximum likelihood tree inferred by the signature code is used to estimate the most likely ancestral amino acids at branch points in the interior of the tree (Bhattacharya et al., 2007). For the branch point preceding each leaf node, the most likely amino acid is determined based on the most likely nucleotides at each position in the codon, which is translated to obtain the ancestral AA state of that leaf. A Fisher's exact test contingency table is based on whether the amino acid changed away from or stayed the same as the ancestral state, and whether the neutralizing phenotype is resistant or sensitive. Full statistics and

contingency tables are provided in [Table S3](#), including detailed examples about how the contingency tables are constructed and their interpretation. As above, this phylogenetically corrected association test can also be used to analyze PNGS associations.

Several cutoffs were used to define relative sensitivity and resistance: IC₅₀ titers being above (negative) or below (positive) the threshold of detection based on the highest concentration of Ab used, or partitioning the data about the median or the quartile responses. For a given amino acid at a given site, the results for phylogenetically corrected test with the lowest p-values are shown in [Table S3](#) E-H, with the test performed indicated in the table. If there are ties, they are broken by presenting the undetected vs detected responses (called PosNeg) when they are available, or by presenting the median over the quartile breakdowns if the tie is just between those two cutoffs. We also present uncorrected associations for all amino acids in *positions of interest* with bNAb sensitivity, again with a q-value cutoff of 0.2 ([Tables S3I–S3L](#)). This enabled us to explore the potential of amino acids at these interesting positions to contribute to levels of bNAb sensitivity, with a less stringent test than required surviving a phylogenetic correction. Sites were *deemed of interest* for this extended exploration by being either located within epitope, or by being found to be significantly associated with IC₅₀ titers using a phylogenetically corrected test for at least one bNAb in a class.

Throughout this study, signatures were generally defined using the Fisher's exact test method described above, but we also explored using a non-parametric Wilcoxon rank sum test. In this case, the distributions of IC₅₀ titers were compared when an amino acid or PNGS site was present or absent in a given position, in a phylogenetically corrected analysis. The Wilcoxon test was generally less sensitive than a Fisher's test, however some CD4bs signatures were best defined using this test, and these results are provided in [Table S3M](#).

Antibody references

Links between particular antibodies, references, and antibody provenance and relationships are provided in table format [Table S1](#). Fourteen V3 glycan bNAbs were studied ([Bonsignori et al., 2016](#); [Garces et al., 2015](#); [Julien et al., 2013](#); [Kong et al., 2013](#); [Mouquet et al., 2012](#); [Pejchal et al., 2011](#); [Trkola et al., 1996](#); [Walker et al., 2011](#)). Ten V2 bNAbs were analyzed ([Bonsignori et al., 2011](#); [Doria-Rose et al., 2014, 2016](#); [McLellan et al., 2011](#); [Sok et al., 2014](#); [Walker et al., 2009, 2011](#)). Twenty-six CD4bs bNAbs were studied ([Bonsignori et al., 2017a](#); [Burton et al., 1991](#); [Corti et al., 2010](#); [Diskin et al., 2011](#); [Gao et al., 2014](#); [Huang et al., 2016](#); [Klein et al., 2012](#); [Rudicell et al., 2014](#); [Scheid et al., 2011](#); [Wagh et al., 2016](#); [Wu et al., 2010, 2011, 2015](#); [Zhou et al., 2013, 2015](#)). These were grouped into 3 types ([Zhou et al., 2015](#)): VH1-2 restricted, VH1-46 restricted, and those with a CDR H3 dominant mode of binding ([Table S1](#)). The 4 MPER bNAbs studied were grouped by epitope, 2F5 or 4E10/10E8/DH511 ([Buchacher et al., 1994](#); [Huang et al., 2012](#); [Nelson et al., 2007](#); [Williams et al., 2017](#))

Phylogenetic trees

Maximum likelihood trees were generated based on amino acid sequences using PhyML ([Guindon et al., 2010](#)) using the HIVb model ([Nickle et al., 2007](#)) (<https://www.hiv.lanl.gov/content/sequence/PHYML/interface.html>), and represented using Rainbow Tree at the Los Alamos database (<https://www.hiv.lanl.gov/content/sequence/RAINBOWTREE/rainbowtree.html>), ([Paradis et al., 2004](#)).

Alignments

The signature analysis tool requires as input codon-aligned nucleotide alignments, which we generated using the Gene Cutter tool at https://www.hiv.lanl.gov/content/sequence/GENE_CUTTER/cutter.html, followed with hand editing. The complete dataset alignments for data sets 3 and 4 and TZM.bl neutralization assay IC₅₀ data used in this study will be in the Special Interest Alignments page of the Los Alamos upon publication (https://www.hiv.lanl.gov/content/sequence/HIV/SI_alignments/datasets.html), and all of the neutralization data is publicly available through web-based CATNAP tool.

Sequence Representation

Amino acids single-letter codes are used throughout. Standard HXB2 numbers is used throughout. The Los Alamos database Analyze Align tool (https://www.hiv.lanl.gov/content/sequence/ANALYZEALIGN/analyze_align.html) was used to generate sequence LOGOS ([Crooks et al., 2004](#)). LOGOS represent the frequency of amino acids in the illustrations included here, the measure of interest for this study, in the M group dataset 4, or in the C clade dataset 3. During the course of this study we have built convenient features into the Analyze Align LOGO generation tool: (1) when an N is embedded in a glycosylation site motif Nx[ST], we replace N with the letter O in the LOGO figure, otherwise we leave it as an N, (2) a grey box is used to indicate gaps inserted to maintain the alignment, (3) specific color schemes (e.g. our red/blue sensitivity/resistance color scheme) are now available, and (4) the tool can now make LOGOs of discontinuous sites by utilizing HXB2 numbering.

Hypervariable region characterization

Hypervariable regions were characterized using the Los Alamos database *Variable Region Characteristics* tool: https://www.hiv.lanl.gov/content/sequence/VAR_REG_CHAR/index.html. The variable loops V1, V2, V4, V5 each have hypervariable regions that frequently mutate by insertion and deletion. V3 loops have low levels of mutation by insertion and deletion, thus this region is readily aligned and was not considered hypervariable. The boundaries of these regions are shown in [Table S3](#) R. We systematically tested for correlations with each variable region characteristic (length, net charge, and number of PNGS) and Ab sensitivity for every Ab in each dataset, and calculated q-values to address multiple tests. If an Ab in a class had a q value of < 0.20 with a V-loop correlate, it was considered of interest, and all other Abs of that same class were tracked and included in [Tables S3N–S3Q](#). The hypervariable nature of these regions leads to rapid changes in them within the course of a given infection, so one would expect markedly diminished phylogenetic correlation with these hypervariable loop characteristics across a population, and thus a phylogenetic correction is not appropriate for these analyses. For each bNAb, we tested for correlations between all variable region characteristics and bNAb sensitivity, both including and excluding censored data ([Tables S3N–S3Q](#)).

QUANTIFICATION AND STATISTICAL ANALYSIS

Sequence Analyses

Signature Statistics

To address multiple tests false discovery rate q -values were calculated (Storey and Tibshirani, 2003). For all signature comparisons in Table S3, a q -value ≤ 0.2 was required for inclusion. We built our own Fisher's exact test code and q -value estimates into the signature analysis package (Bhattacharya et al., 2007). We use R (www.r-project.org/) to perform Wilcoxon rank sum comparisons of distributions, to perform Kendall's tau (McLeod, 2011) to test for correlations, and to calculate q -values for addressing multiple tests when evaluating variable loop characteristics (Dabney and Storey, 2013). Heatmaps were generated using the Los Alamos HIV database tool (<https://www.hiv.lanl.gov/content/sequence/HEATMAP/heatmap.html>).

Machine learning predictions

We used the Python package scikit-learn for machine learning predictions of bNAb sensitivity (Pedregosa et al., 2011). We initially compared several machine learning strategies (Random Forest, Support Vector Machine, and Linear Discriminant Analysis) using the M group cross-validation scores; the C clade holdout group was not considered during this selection process. Random Forest (RF) (Breiman, 2001) strategies performed best, in particular the ExtraTreesRegressor and ExtraTreesClassifier methods (Geurts et al., 2006) gave the highest accuracy overall, and so were used here as a basis for comparing the accuracy of sequence-based filtering strategies for obtaining input features for neutralization predictions. The overall RF result is obtained by combining the results from each of the individual trees in the forest. For our RF experiments, we fixed the size of the ensemble to 250 trees but otherwise used default values of the scikit-learn parameters. In particular, the depth of the trees (*i.e.*, the number of branches) was adaptively determined using a bootstrap approach available in scikit-learn.

As mentioned in the text, the three pre-filtering strategies used were: mRMR, the full signature set (including outside epitope signatures, hypervariable loop characteristics and clade associations), and only signatures in the bNAb epitopes. Input data files for signature-based prefiltering were created with columns of data translated so that clades, and signature amino acids and PNGSs in a given position, were assigned a 1 if associated with sensitivity, -1 if associated with resistance, or 0 if not associated. Quantitative values for correlated variable loop characteristics were also included.

For each pre-filter strategy, we obtained predictions for two scenarios. First, we predicted IC_{50} titers for the 207 Env sequences from the M group dataset 4, using leave-one-out cross-validation. We decided to use leave-one-out cross validation because the datasets were small enough that we were not computationally constrained, and this approach minimizes bias in small data sets (Arlot and Celisse, 2010). Second, we trained the RF on the M group data from datasets 1, 2 and 4, and evaluated the performance on a holdout C clade virus set from dataset 3. To maintain the independence of our holdout set, we used signature pre-filters that were defined only on the basis of the M group datasets 1, 2, and 4, and we excluded the 26 pseudoviruses from the 200 in the C clade set that were also found in M group data. Predictions were made for the 13 bNABs that were available in both dataset 3 and 4, and both regression (IC_{50} titers; Table S4; Figure 4) and classification (positive versus negative; Table S5) prediction accuracies were assessed. We compared the accuracy of results using different strategies to pre-filter Env sequence alignment data.

For regression, the aim is to predict the potency, and here we used three measures of performance to assess the quality of these predictions. The most direct measure is the mean absolute value of the prediction error (MAE). We also used the R^2 statistic (the coefficient of determination), whose variation is generally (but not strictly) bounded between zero and one, with larger values corresponding to better predictions. Finally, to assess the statistical significance of the predictions vis-a-vis a null hypothesis of no predictive power, we computed the p -value associated with a Kendall's tau test comparing the predictions to the true values. Ranked importance of different features from the RF analysis are provided in Table S6.

For classification, the goal is to predict a binary outcome of whether a bNAb will give detectable neutralization responses against a given sequence or not. The most intuitive measurement of performance is the accuracy, *i.e.*, the fraction of sequences for which the prediction is correct. In some cases, however, simply predicting all positives or all negatives will give a very high accuracy score (e.g. 10E8 neutralizes at some level 98% of the viruses tested), so machine learning prediction is highly accurate, but it is not much better than just predicting that all Envs are positive.

We tested 3 comparisons of particular interest to highlight the importance of signatures in enabling accurate predictions. First, we used only signature sites that were in contact residues versus the complete signatures; the complete list was favored for regression predictions (see Results). Second, we compared using signature sites sequence features as inputs, to using the mRMR approach to filter out the 100 most informative sites (Hepler et al., 2014; Peng et al., 2005). As noted in the Results, complete signatures yielded the most accurate predictions for regression, but there was no clear preference for classification. Before switching to our own mRMR-RF code, to make sure our approach was at least comparable in prediction accuracy to the previously published IDEpi classification code (Hepler et al., 2014), we compared the prediction accuracy of the two methods using 10-fold cross-validation for M group analysis, and also comparing the accuracy for the C clade holdout. Our implementation of the mRMR-RF approach was generally comparable to IDEpi (Table S5), although for a small number of antibodies our error was substantially lower (e.g. VRC01 and 10E8). As a final comparison, because most published computational studies present only a very small number of amino acid signatures for each Ab, we sought to determine whether reducing the number of features to only the strongest features improved the scores, so we limited the Random Forest to include only the 3 most informative features. When comparing this restricted set to the full signature pattern, we found the restricted set not only did not improve classification or regression scores, it often made them much worse.

Vaccine Immune Response Comparisons

Analysis of SET vaccines neutralization data

Neutralization data were analyzed using the R package (www.r-project.org) and GraphPad Prism version 6.00 software (GraphPad Software, San Diego California USA, www.graphpad.com”).

We considered ID₅₀ titers positive if they were at least 10 above background:

Cutoff 1: Response = Post, if Post > MuLV + 10; 10 otherwise, where ‘Post’ is post-vaccination sera ID₅₀ (4 weeks-post last vaccination), ‘MuLV’ is the background level for an animal-matched MuLV negative control (4 weeks-post last vaccination) (Bricault et al, 2015), and “10” considered below the level of detection. We compared the statistical results presented in Figure 6 with the outcome using alternative cutoffs 2 and 3:

Cutoff 2: Response = Post - MuLV, if Post-MuLV > 10, 10 otherwise,

Cutoff 3: Response = Post, if Post > 3*MuLV, 10 otherwise,

and found the results obtained using Cutoff 2 and Cutoff 3 were consistent with the results obtained with the cutoff 1 when comparing vaccine groups, so Cutoff 1 is shown.

The breadth of neutralization response (detected vs not-detected) was assessed by counting for each animal the proportion of pseudoviruses with detectable neutralization and then applying the two-sided Wilcoxon rank-sum test to compare the differences in distributions of responses per animal between the 459C WT and the V2-SET vaccines.

The differences in the magnitude of responses between V2-SET vaccines and the 459C WT alone were assessed by a nonparametric permutation test following the strategy described in (Parrish et al., 2013). Briefly, this test compares the medians of responses elicited by the 459C WT and the given V2-SET vaccine in the observed data and in the 10,000 randomized sets of resampled data where the vaccine category is randomly reassigned between vaccinated animals. The fraction of occurrences of median differences in the randomized data that are equal to or less than that observed median differences in the actual data provides an estimate of the probability for observing the actual results by the chance alone.

DATA AND SOFTWARE AVAILABILITY

In this study we have created a catalog of new signature sites, also including those that were defined previously, and created 3 web-based tools to facilitate future analyses: *GenSig* enables users to implement their own phylogenetically corrected signature analysis, *FilteredForests* enables machine learning predictions using either bNAb signature-based or mRMR prefilters, and we have automated neutralization signature predictions for new bNAb neutralization panels as they are incorporated into the Los Alamos HIV database CATNAP NAb interface (Yoon et al., 2015).

GenSig

We have developed a Signature Tool web interface, *GenSig*, at the Los Alamos HIV database: <https://www.hiv.lanl.gov/content/sequence/GENETICSIGNATURES/gS.html>. It can work on any phenotype file given in conjunction with a codon-aligned nucleotide alignment of a protein coding region of moderate size (<1000 gene sequences) – the tool is not specific for HIV-1 and neutralization data. If, however, an input alignment is an HIV-1 gene alignment with the HXB2 reference sequence is included, the numbering of the output will be given according to HIV standardized HXB2 numbering.

CATNAP Enhancement

The HIV-1 pseudovirus sequence data for the viral panels and previously published GenBank accession numbers are all already available through the Los Alamos HIV Database CATNAP tool, which we maintain. All new neutralization data used in this study will be integrated into the CATNAP tool at the time of publication: <https://www.hiv.lanl.gov/catnap>. Since new HIV-1 bNAbs are continuously being added to the literature, and new neutralization panel data is regularly entered into the Los Alamos database CATNAP tool (Yoon et al., 2015) for comparative analysis (<https://www.hiv.lanl.gov/components/sequence/HIV/neutralization/index.html>), we have added an automated signature analysis feature to update signatures for new data as it accrues.

FilteredForests

A web interface to run our sequence-based prefilters for machine learning predictions of bNAb sensitivity automatically coupled to RF code from the Python scikit-learn package, (Pedregosa et al., 2011) is called *FilteredForests* code. One can generate their own signature-based prefilters or use mRMR (Peng et al., 2005) to generate from a sequence alignment input files for the RF machine learning codes ExtraTreesRegressor and ExtraTreesClassifier. This web interface is available at: <https://www.hiv.lanl.gov/content/sequence/FLTFORESTS/fitforests.html>

The code is available at: <https://github.com/hivdb-lanl/FilteredForests>

bNAb signature information access

To enable comparisons to sites of interest for particular bNAbs identified here, to sites identified in the previously published literature, all signatures identified in this study have been incorporated into the Los Alamos HIV Immunology Databases. The signature

information will be included and accessible through three Los Alamos database tools: a simple spread sheet that is an overview of many of the key findings from the literature, that allows comparisons of findings for all sites (rows) in Env across many antibody studies, organized by paper and/or antibody (columns) (https://www.hiv.lanl.gov/content/immunology/neutralizing_ab_resources.html). The signatures will also be accessible through the relational database we have built for searching bNAb characteristics (Neutralizing Antibody Contexts and Features, (www.hiv.lanl.gov/components/sequence/HIV/featuredb/search/env_ab_search_pub.comp); and the Genome Browser, which allows users to interactively explore functional domains and sites relevant to antibodies across Env (www.hiv.lanl.gov/content/sequence/genome_browser/browser.html).”